Statistics in Psychology (PSY516)

Table of Content

Lesson 1

INTRODUCTION TO STATISTICS IN PSYCHOLOGY-I

Introduction to Statistics

The study of statistics is gaining recognition in a great many fields. In particular, researchers in the social, behavioral and health sciences note its importance for problem solving and its practical importance in their areas. Statistics is a science that enables us to draw conclusions about various phenomena on the basis of real data collected on sample-basis. It is a tool for data-based research and also known as Quantitative Analysis. It is the science of collecting and learning from data. It is a branch of mathematics concerned with the collection, classification, analysis, and interpretation of numerical facts, for drawing inferences on the basis of their quantifiable likelihood (probability). Statistics also allows the researcher to interpret grouped data, too large to be intelligible by ordinary observation.

Statistics has a lot of application in a wide variety of disciplines i.e. Agriculture, Anthropology, Astronomy, Biology, Economic, Engineering, Environment, Geology, Genetics, Medicine, Physics, Psychology, Sociology, Zoology…. Virtually every single subject from Anthropology to Zoology …. A to Z!

Any scientific enquiry in which you would like to base your conclusions and decisions on real-life data, you need to employ statistical techniques. Now a days, in the developed countries of the world, there is an active movement for Statistical Literacy.

Defining Statistics

The term **statistics** refers to a set of mathematical procedures for organizing, summarizing, and interpreting information. One goes through the four stages of statistics:

1. Collection of data,
2. Organizing and summarizing the data,
3. Analysis of data, and
4. Making inferences, or decisions and predictions.

Stage four is the core objective of statistics in psychology, i.e., making inferences about a population based on information contained in a representative sample taken from that population. Statistics consist of facts and figures such as average income, crime rate, birth rate, baseball batting averages, and so on. These statistics are usually informative and time saving because they condense large quantities of information into a few simple figures.

Statistics is the science of learning from data. Statistical procedures help to ensure that the information or observations are presented and interpreted in an accurate and informative way. In somewhat grandiose terms, statistics help researchers bring order out of chaos. Specifically, statistics serve two general purposes:

1. It is used to organize and summarize the information so that the researcher can see what happened in the research study and can communicate the results to others.
2. It helps the researcher to answer the questions that initiated the research by determining exactly what general conclusions are justified based on the specific results that were obtained.

Research in psychology (and other fields) involves gathering information. To determine, for example, whether violence on TV has any effect on children's behavior, you would need to gather information about children's behaviors and the TV programs they watch. When researchers finish the task of gathering information, they typically find themselves with pages and pages of measurements such as IQ scores, personality scores, reaction time scores, and so on.

Importance Of Statistics

As it is such an important area of knowledge, it is definitely useful to have a fairly good idea about the way in which it works, and this is exactly the purpose of this introductory course.

The following points indicate some of the main functions of this science:

- Statistics assists in summarizing the larger set of data in a form that is easily understandable.
- Statistics assists in the efficient design of laboratory and field experiments as well as surveys.
- Statistics assists in a sound and effective planning in any field of inquiry.
- Statistics assists in drawing general conclusions and in making predictions of how much of a thing will happen under given conditions.

Importance Of Statistics In various fields

As stated earlier, Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly.

- A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision.
- A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles.
- A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments.
- A social scientist uses statistical methods in various areas of socioeconomic life of a nation. It is sometimes said that "*a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there*".

Myths About Statistics

When we are studying statistics, several questions and myths arises in our mind. Following are some common myths about statistics:

- Statistics is very Hard.
- Statistics is Math.
- Why it is necessary to study Statistics in Psychology?
- You Lie with Statistics.
- If something is not statistically significant, it is not important.

Lesson 2

INTRODUCTION TO STATISTICS IN PSYCHOLOGY-II

Terms and Definitions

The field of statistics is subdivided into *descriptive* statistics and *inferential* statistics. Following are some statistical jargon used in psychology:

Descriptive statistics

Descriptive statistics are statistical procedures used to summarize, organize, and simplify data. Descriptive statistics are techniques that take raw scores and organize or summarize them in a form that is more manageable. Often the scores are organized in a table or a graph so that it is possible to see the entire set of scores. Another common technique is to summarize a set of scores by computing an average. Typically, there are two general types of statistics that are used to describe data"

- Measures of central tendency
- Measure of spread/dispersion

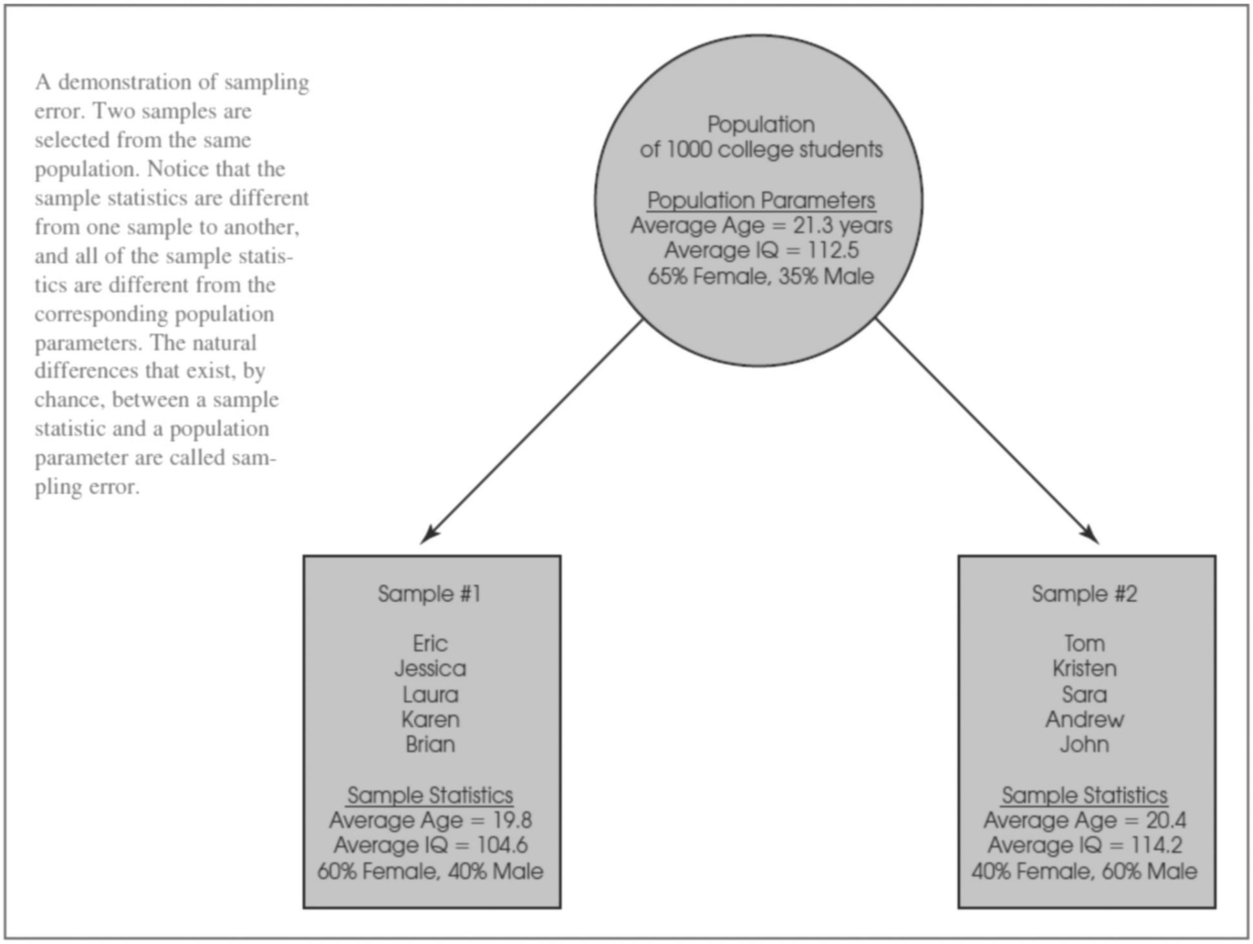
Inferential statistics

Whereas descriptive statistics is the branch of statistics that involves organizing, displaying, and describing data, inferential statistics is the branch of statistics that involves drawing conclusions about a population based on information contained in a sample taken from that population. Inferential statistics consist of techniques that allow us to study samples and then make generalizations about the populations from which they were selected. Most research uses statistical models called the Generalized Linear model and include Student's t-tests, ANOVA (Analysis of Variance), regression analysis.

Sampling Error

Sampling error is the naturally occurring discrepancy, or error, that exists between a sample statistic and the corresponding population parameter.

Because populations are typically very large, it usually is not possible to measure everyone in the population. Therefore, a sample is selected to represent the population. By analyzing the results from the sample, we hope to make general statements about the population. One problem with using samples, however, is that a sample provides only limited information about the population. Although samples are generally representative of their populations, a sample is not expected to give a perfectly accurate picture of the whole population. There usually is some discrepancy between a sample statistic and the corresponding population parameter. This discrepancy is called sampling error.

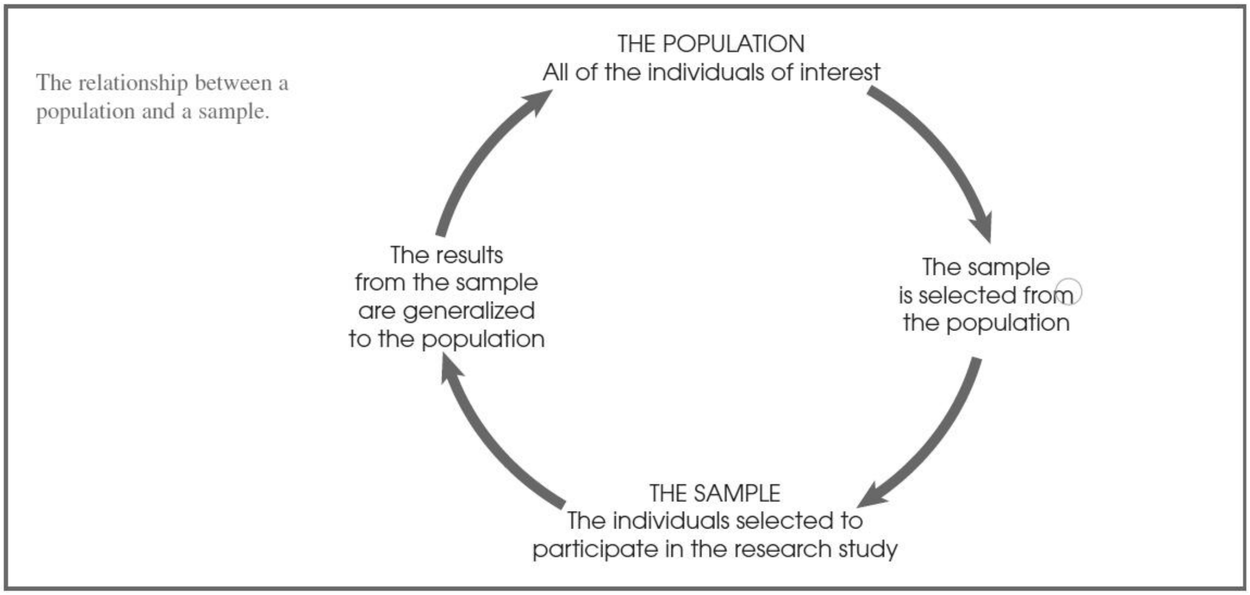


Population

A population is the set of all the individuals of interest in a particular study. For example, a researcher may be interested in the effect of divorce on the self-esteem of preteen children. Or a researcher may want to examine the amount of time spent in the bathroom for men compared to women. In the first example, the researcher is interested in the group of preteen children. In the second example, the researcher wants to compare the group of men with the group of women. In statistical terminology, the entire group that a researcher wishes to study is called a population.

Sample

A sample is a set of individuals selected from a population, usually intended to represent the population in a research study. Because populations tend to be very large, it usually is impossible for a researcher to examine every individual in the population of interest. Therefore, researchers typically select a smaller, more manageable group from the population and limit their studies to the individuals in the selected group. In statistical terms, a set of individuals selected from a population is called a sample. A sample is intended to be representative of its population, and a sample should always be identified in terms of the population from which it was selected.



Population Parameter and Sample Statistics

When describing data, it is necessary to distinguish whether the data come from a population or a sample. A characteristic that describes a population—for example, the average score for the population—is called a parameter. A characteristic that describes a sample is called a statistic. A **parameter** is a value, usually a numerical value, that describes a population. A parameter is usually derived from measurements of the individuals in the population. A **statistic** is a value, usually a numerical value, that describes a sample. A statistic is usually derived from measurements of the individuals in the sample.

Every population parameter has a corresponding sample statistic, and most research studies involve using statistics from samples as the basis for answering questions about population parameters.

Variables and their Data

Typically, researchers are interested in specific characteristics of the individuals in the population (or in the sample), or they are interested in outside factors that may influence the individuals. For example, a researcher may be interested in the influence of the weather on people's moods. As the weather changes, do people's moods also change? Something that can change or have different values is called a variable. A **variable** is a characteristic or condition that changes or has different values for different individuals.

To demonstrate changes in variables, it is necessary to make measurements of the variables being examined. The measurement obtained for each individual is called a datum or, more commonly, a score or raw score. The complete set of scores is called the data set or simply the data. **Data** (plural) are measurements or observations. A data set is a collection of measurements or observations. A datum (singular) is a single measurement or observation and is commonly called a score or raw score.

Before we move on, we should make one more point about samples, populations, and data. Earlier, we defined populations and samples in terms of individuals. For example, we discussed a population of college sophomores and a sample of preschool children. Be forewarned, however, that we will also refer to populations or samples of scores. Because research typically involves measuring each individual to obtain a score, every sample (or population) of individuals produces a corresponding sample (or population) of scores.

What to Measure?

Some variables, such as height, weight, and eye color are well-defined, concrete entities that can be observed and measured directly. On the other hand, many variables studied by behavioral scientists are internal characteristics that people use to help describe and explain behavior. Variables like intelligence, anxiety, and hunger are called constructs. The external behaviors can then be used to create an operational definition for the construct.

Constructs are internal attributes or characteristics that cannot be directly observed but are useful for describing and explaining behavior.

An **operational definition** identifies a measurement procedure (a set of operations) for measuring an external behavior and uses the resulting measurements as a definition and a measurement of a hypothetical construct. Note that an operational definition has two components: First, it describes a set of operations for measuring a construct. Second, it defines the construct in terms of the resulting measurements.

Discrete Variable

A discrete variable consists of separate, indivisible categories. No values can exist between two neighboring categories. Also known as a categorical variable, because it has separate, invisible categories.

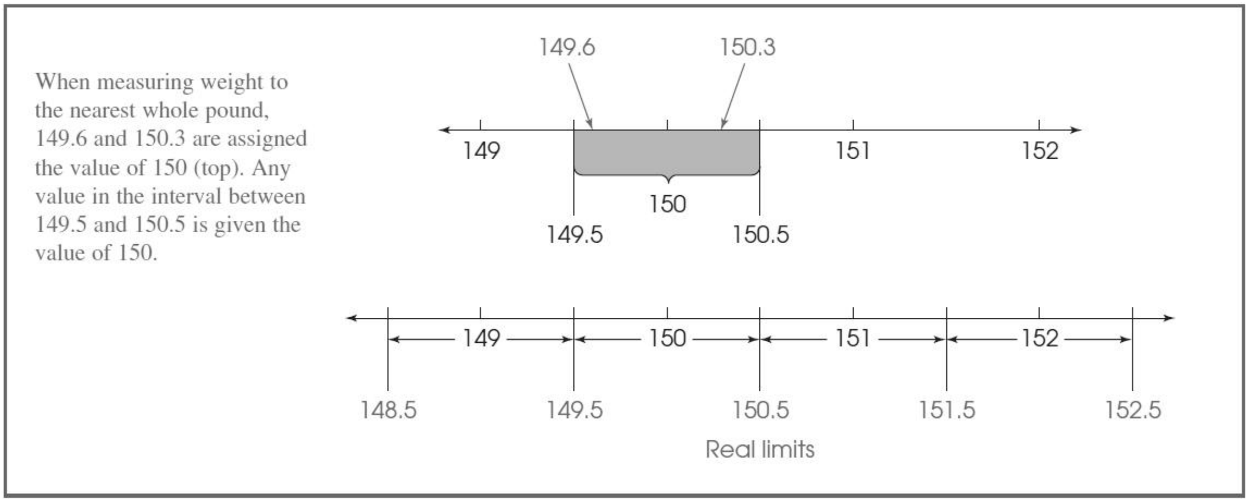
Discrete variables are commonly restricted to whole, countable numbers—for example, the number of children in a family or the number of students attending class. If you observe class attendance from day to day, you may count 18 students one day and 19 students the next day. However, it is impossible ever to observe a value between 18 and 19. A discrete variable may also consist of observations that differ qualitatively.

Continuous Variable

For a continuous variable, there are an infinite number of possible values that fall between any two observed values. A continuous variable is divisible into an infinite number of fractional parts. Also known as a ratio/interval variable, consists of ordered categories that are all intervals of exactly the same size with or without absolute zero. Variables such as time, height, and weight are not limited to a fixed set of separate, indivisible categories. You can measure time, for example, in hours, minutes, seconds, or fractions of seconds. Two other factors apply to continuous variables:

1. When measuring a continuous variable, it should be very rare to obtain identical measurements for two different individuals. Because a continuous variable has an infinite number of possible values, it should be almost impossible for two people to have exactly the same score.
2. When measuring a continuous variable, each measurement category is actually an interval that must be defined by boundaries. These boundaries are called real limits and are positioned exactly halfway between adjacent scores. Thus, a score of X 150 pounds is actually an interval bounded by a lower real limit of 149.5 at the bottom and an upper real limit of 150.5 at the top. Any individual whose weight falls between these real limits will be assigned a score of X 150.

Real Limits are the boundaries of intervals for scores that are represented on a continuous number line. The real limit separating two adjacent scores is located exactly halfway between the scores. Each score has two real limits. The upper real limit is at the top of the interval, and the lower real limit is at the bottom.



Independent Variable

A variable thought to be the cause of some effect. The independent variable is the variable that is manipulated by the researcher. In behavioral research, the independent variable usually consists of the two (or more) treatment conditions to which subjects are exposed. The independent variable consists of the antecedent conditions that were manipulated prior to observing the dependent variable. E.g. two conditions, or control and experimental group or pre/post test. Note that the independent variable always consists of at least two values. (Something must have at least two different values before you can say that it is "variable."). Predictor variable: A variable thought to predict an outcome variable. This is basically another term for independent variable.

Dependent Variable

The dependent variable is the variable that is observed to assess the effect of the treatment. A variable thought to be affected by changes in an independent variable. You can think of this variable as an outcome. The variable that is observed and measured to obtain scores is the dependent variable. Outcome variable: A variable thought to change as a function of changes in a predictor variable. This term could be synonymous with 'dependent variable'.

Control Variables

Variables that are held constant throughout the experiment. The temperature and light in the room the plants are kept in, and the volume of water given to each plant.

Extraneous Variables

Extraneous variables are any variables that you are not intentionally studying in your experiment or test but they effect the results of the study.

Extraneous Variables can be: Demand characteristics: environmental clues which tell the participant how to behave, like features in the surrounding or researcher's non-verbal behavior. Experimenter / Investigator Effects: where the researcher unintentionally affects the outcome by giving clues to the participants about how they should behave. Participant variables, like prior knowledge, health status or any other individual characteristic that could affect the outcome. Situational variables, like noise, lighting or temperature in the environment.

Confounding Variables

A variable that hides the true effect of another variable in your experiment. This can happen when another variable is closely related to a variable you are interested in, but you haven't controlled it in your experiment.

Scales of Measurement

It should be obvious by now that data collection requires that we make measurements of our observations. Measurement involves assigning individuals or events to categories. The categories can simply be names such as male/female or employed/unemployed, or they can be numerical values such as 68 inches or 175 pounds. The categories used to measure a variable make up a scale of measurement, and the relationships between the categories determine different types of scales.

The Nominal Scale

The word nominal means "having to do with names." A nominal scale consists of a set of categories that have different names. Measurements on a nominal scale label and categorize observations, but do not make any quantitative distinctions between observations.

The measurements from a nominal scale allow us to determine whether two individuals are different, but they do not identify either the direction or the size of the difference. A sub-type of nominal scale with only two categories (e.g. male/female) is called "dichotomous." If you are a student, you can use that to impress your teacher. Other sub-types of nominal data are "nominal with order" (like cold, warm, hot, very hot) and nominal without order (like male/female).

Although the categories on a nominal scale are not quantitative values, they are occasionally represented by numbers. For example, the rooms or offices in a building may be identified by numbers. You should realize that the room numbers are simply names and do not reflect any quantitative information. Room 109 is not necessarily bigger than Room 100 and certainly not 9 points bigger.

The Ordinal Scale

An ordinal scale consists of a set of categories that are organized in an ordered sequence. Measurements on an ordinal scale rank observation in terms of size or magnitude.

With measurements from an ordinal scale, you can determine whether two individuals are different and you can determine the direction of difference. However, ordinal measurements do not allow you to determine the size of the difference between two individuals. For example, if Billy is placed in the low-reading group and Tim is placed in the high-reading group, you know that Tim is a better reader, but you do not know how much better.

Ordinal scale consists of a series of ranks (first, second, third, and so on) like the order of finish in a horse race. Occasionally, the categories are identified by verbal labels like small, medium, and large drink sizes at a fast-food restaurant. In either case, the fact that the categories form an ordered sequence means that there is a directional relationship between categories.

Interval Scale

An interval scale consists of ordered categories that are all intervals of exactly the same size. Equal differences between numbers on a scale reflect equal differences in magnitude. However, the zero point on an interval scale is arbitrary and does not indicate a zero amount of the variable being measured.

For example, a temperature of 0° Fahrenheit does not mean that there is no temperature, and it does not prohibit the temperature from going even lower. Furthermore, you know that a measurement of 80° Fahrenheit is higher than a measure of 60°, and you know that it is exactly 20° higher.

Interval scales are numeric scales. Interval scales don’t have a “true zero.” Hence, zero and negative numbers also have meaning. Without a true zero, it is impossible to compute ratios.

The Ratio Scale

A ratio scale is an interval scale with the additional feature of an absolute zero point. With a ratio scale, ratios of numbers do reflect ratios of magnitude.

For example, a score of 16 on an anxiety scale means that the person is, in reality, twice as anxious as someone scoring 8.

A ratio scale is anchored by a zero point that is not arbitrary but rather is a meaningful value representing none (a complete absence) of the variable being measured. The existence of an absolute, non-arbitrary zero point means that we can measure the absolute amount of the variable; that is, we can measure the distance from 0. This makes it possible to compare measurements in terms of ratios.

Due to absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied. These variables can be meaningfully added, subtracted, multiplied, divided (ratios). Central tendency can be measured by mode, median, or mean; measures of dispersion, such as standard deviation and coefficient of variation can also be calculated from ratio scales.

Lesson 3

CORRELATION METHODS

Most research, are intended to examine relationships between two or more variables. For example, is there a relationship between the amount of violence that children see on television and the amount of aggressive behavior they display? Is there a relationship between the quality of breakfast and level of academic performance for elementary school children? To establish the existence of a relationship, researchers must make observations—that is, measurements of the two variables. The resulting measurements can be classified into two distinct data structures that also help to classify different research methods and different statistical techniques. In the following section we identify and discuss these two data structures.

The Correlational Method

One method for examining the relationship between variables is to observe the two variables as they exist naturally for a set of individuals. That is, simply measure the two variables for each individual. Most common analysis used in correlational studies is Pearson Product Moment Correlation (r).

In the **correlational method**, two different variables are observed to determine whether there is a relationship between them.

Correlational research allows researchers to:

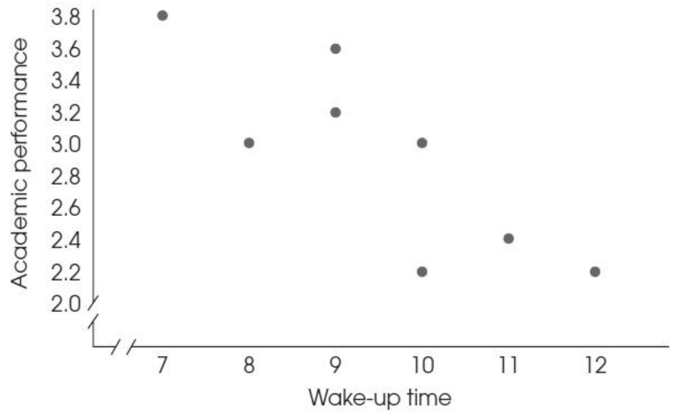
- establish reliability and validity
- provide converging evidence
- describe relationships
- make predictions

For example, research has demonstrated a relationship between sleep habits, especially wake-up time, and academic performance for college students (Trockel, Barnes, and Egget, 2000). The researchers used a survey to measure wake-up time and school records to measure academic performance for each student. The researchers then look for consistent patterns in the data to provide evidence for a relationship between variables. For example, as wake-up time changes from one student to another, is there also a tendency for academic performance to change?

(a)

Student	Wake-up Time	Academic Performance
A	11	2.4
B	9	3.6
C	9	3.2
D	12	2.2
E	7	3.8
F	10	2.2
G	10	3.0
H	8	3.0

(b)



One of two data structures for studies evaluating the relationship between variables. Note that there are two separate measurements for each individual (wake-up time and academic performance). The same scores are shown in a table (a) and in a graph (b).

Characteristics of the Relationship of Correlation:

- A ***positive relationship*** depicts, where low (or high) scores on one variable relate to low (or high) scores on a second variable.
- A ***negative relationship*** results, where low scores on one variable relate to high scores on the other variable.
- A ***zero relationship*** *of scores:* In this distribution, the variables are independent of each other. A particular score on one variable does not predict or tell us any information about the possible score on the other variable.

Degree Of Association means that the association between two variables or sets of scores is a correlation coefficient of –1.00 to +1.00, with 0.00 indicating no linear association at all.

Types of Correlational Designs:

Cross-Sectional Designs - One on more samples are drawn from a population, which are studied at one time on the same variables. E.g. effect of gender on depression of people born in 60s and 90s. (year makes two different samples).

Longitudinal Design - Same respondents or samples are surveyed over a period of time and is useful for assessing changes in behavior seen in individuals over time. The design is also the best form of survey for studying the effect of a naturally occurring event or phenomena.

Limitations of the Correlational Method:

The results from a correlational study can demonstrate the existence of a relationship between two variables, but they do not provide an explanation for the relationship. In particular, a correlational study cannot demonstrate a cause-and-effect relationship. In the above example a systematic relationship between wake-up time and academic performance for a group of college students; those who sleep late tend to have lower performance scores than those who wake early. However, there are many possible explanations for the relationship and we do not know exactly what factor (or factors) is responsible for late sleepers having lower grades.

Lesson 4

EXPERIMENTAL AND NON-EXPERIMENTAL METHOD

The Experimental Method

One specific research method that involves comparing groups of scores is known as the experimental method or the experimental research strategy. The goal of an experimental study is to demonstrate a cause-and-effect relationship between two variables. Specifically, an experiment attempts to show that changing the value of one variable causes change to occur in the second variable.

In the **experimental method**, one variable is manipulated while another variable is observed and measured. To establish a cause-and-effect relationship between the two variables, an experiment attempts to control all other variables to prevent them from influencing the results.

Three conditions needed to make causal inferences

1. ***Covariation:*** Covariation should exist i.e. when a relationship is seen between the IV and DV.
2. ***Time order relationship:*** Time order relationship: when researchers manipulate an independent variable and then see a difference in the subsequent behavior or DV. The cause should always come before the effect.
3. ***Elimination of confounding variables*** or alternative causes that can affect the outcome of experiment of DV.

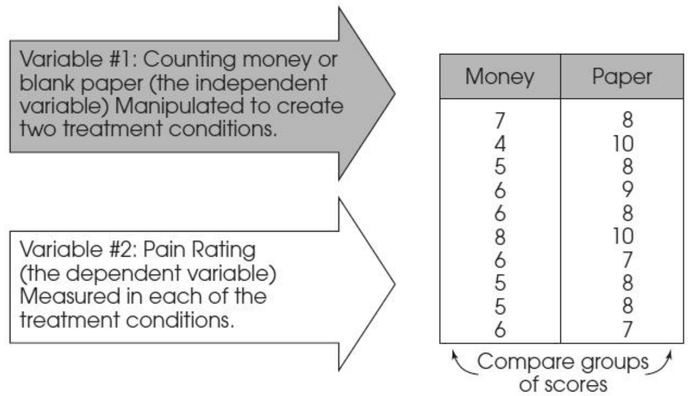
Two methods can be used to accomplish the goal of establishing cause-and-effect relationship in experimental method:

1. ***Manipulation***: The researcher manipulates one variable by changing its value from one level to another. A second variable is observed (measured) to determine whether the manipulation causes changes to occur.
2. ***Control***: The researcher must exercise control over the research situation to ensure that other, extraneous variables do not influence the relationship being examined.

To demonstrate these two characteristics, consider an experiment in which researchers demonstrate the pain-killing effects of handling money (Zhou & Vohs, 2009). In the experiment, a group of college students was told that they were participating in a manual dexterity study. The researcher then manipulated the treatment conditions by giving half of the students a stack of money to count and the other half a stack of blank pieces of paper. After the counting task, the

participants were asked to dip their hands into bowls of painfully hot water (122° F) and rate how uncomfortable it was. Participants who had counted money rated the pain significantly lower than those who had counted
paper.

The structure of an experiment. Participants are randomly assigned to one of two treatment conditions: counting money or counting blank pieces of paper. Later, each participant is tested by placing one hand in a bowl of hot (122° F) water and rating the level of pain. A difference between the ratings for the two groups is attributed to the treatment (paper versus money).



There are two general categories of variables that researchers must consider:

1. ***Participant Variables***: These are characteristics such as age, gender, and intelligence that vary from one individual to another. Whenever an experiment compares different groups of participants (one group in treatment A and a different group in treatment B), researchers must ensure that participant variables do not differ from one group to another.
2. ***Environmental Variables***: These are characteristics of the environment such as lighting, time of day, and weather conditions. A researcher must ensure that the individuals in treatment A are tested in the same environment as the individuals in treatment B.

Researchers typically use three basic techniques to control other variables:

1. Random assignment, which means that each participant has an equal chance of being assigned to each of the treatment conditions. The goal of random assignment is to distribute the participant characteristics evenly between the two groups so that neither group is noticeably smarter (or older, or faster) than the other. Random assignment can also be used to control environmental variables.

2. Second, the researcher can use matching to ensure equivalent groups or equivalent environments. For example, the researcher could match groups by ensuring that every group has exactly 60% females and 40% males.
3. Finally, the researcher can control variables by holding them constant. For example, if an experiment uses only 10-yearold children as participants (holding age constant), then the researcher can be certain that one group is not noticeably older than another.

Types of Experimental Research Design:

- ***Between Subjects or Independent Measures Design*** - The two or more groups of participants take part in different treatment conditions, a between-subjects design allows only one score per participant (every score represents a separate, unique participant).

    Main Components: Manipulation: It is used to check whether the change in the DV is because of the change or manipulation of the IV and not due to any other factor. *Holding* ***Conditions Constant:*** is a control techniques experimenter use to eliminate the effects of confounding factors. ***Balancing:*** The main assumption of the experiment method is to form comparable or similar groups.

    Block randomization: works by randomizing participants within blocks such that an equal number are assigned to each treatment to avoid selection bias.

- ***Within Subjects or Repeated Measures Design*** - Repeated measures design is the one in which each participant is exposed to all treatment conditions unlike independent groups design in which different participants take part in different treatment conditions.

Within subjects or repeated measures design is used:

- When there are small number of participants
- For controlling confounding
- For increasing the sensitivity of an experiment
- Less time consuming and more convenient to arrange
- No need for a separate control group

However, there are certain ***threats to internal validity*** of such experiments e.g. practice effect. These effects can be reduced through counterbalancing or giving participants the treatment conditions by changing their order like ABBA, ABCCAB, ABCCBA, BACABC etc. Practice effects can also be balanced using block randomization.

Non-Experimental Methods

Non-experimental designs encompass all the designs that does not come in true experimental designs. From such, Correlational Designs has already been discussed. Quasi-Experiments, Case Study, Observational Designs.

Quasi-Experiments

Such experiments involve a manipulation of an independent variable or variables but there is no random assignment of participants to different treatment conditions. It is used in contexts when randomization is not possible. In a non-experimental study, the "independent variable" that is used to create the different groups of scores is often called the quasi-independent variable. Following are the two sub types of quasi experiments:

Nonequivalent Groups: study comparing boys and girls. Notice that this study involves comparing two groups of scores (like an experiment). This type of research compares preexisting groups, the researcher cannot control the assignment of participants to groups and cannot ensure equivalent groups.

Nonequivalent group studies include comparing 8-year-old children and 10-year-old children, people with an eating disorder and those with no disorder, and comparing children from a single-parent home and those from a two-parent home. Because it is impossible to use techniques like random assignment to control participant variables and ensure equivalent groups, this type of research is not a true experiment.

Pre–Post Study: The two groups of scores are obtained by measuring the same variable twice for each participant; once before and again after applying treatment. In a pre–post study, however, the researcher has no control over the passage of time. The "before" scores are always measured earlier than the "after" scores. Although a difference between the two groups of scores may be caused by the treatment, it is always possible that the scores simply change as time goes by.

Single Case Study

These involve intensive and detailed descriptions and analysis of a single case. Multiple methods for data collection including: interviews, psychological tests, observations etc. So, the data obtained can be both quantitative and qualitative in nature and it differs from experiments due to lack of control. ***Advantages:*** Case studies can provide new ideas and hypothesis. It is best method for studying rare and individual phenomena or personalities. Opportunity to try out new therapeutic interventions on patients. For challenging theories and formulating new theories.

Disadvantages: Difficult to make cause and effect assumptions. Findings are often subjective because of the subjective bias of the experimenter and hence cannot be generalized.

Advantages of Non-Experimental Methods

- Can provide new ideas and hypothesis
- For studying rare phenomena
- Opportunity to try out new therapeutic interventions on patients
- Can be used for studying rare phenomena
- For challenging theories
- Formulating new theories

Disadvantages of Non-Experimental Methods

- Difficult to make cause and effect assumptions
- Findings are often subjective
- Problems in generalizing the results
- Selection biases

Observational Designs

They involve data collection using direct or indirect observations. *Direct observations* involve looking at a behavior or phenomena directly. These include: observations without intervention and observations with intervention. *Indirect observations* looking at evidence of past behavior using archival records like birth certificates, Facebook entries, marriage licenses, college degrees etc. and physical traces like drawings, textbooks, products used by an individual etc.

Direct observations have 3 variations:

1. ***Participant observation*** *(without Intervention):* Researcher is present physically at the scene and observing from far. ***Merits:*** It allows researchers to gain a closer look at a behavior or phenomena and record information. ***De-merits:*** But it might negatively effect the behavior of participants because when people are aware that they are being watched, their behaviors change
2. ***Structured Observations or controlled observations*** *(without Intervention):* The researchers do not get involved with the participants. The behaviors to be recorded and analyzed are ore-determined. Data collection and analysis techniques are structured. The researcher decides where, when and how the observation will occur.

Advantages: Are reliable, Easy to replicate, less time consuming and can explore multiple behaviors. ***Disadvantages:*** Can lack validity if participants become aware that they are being observed. The researcher's biases can influence the results

3. ***Field Experiment or social experiments*** *(direct observation with intervention):* Researcher mix among the participants and try to manipulate circumstance to produce some effect in a natural setting. ***Merits:*** It gives opportunity to document real life behaviors, the results are generalizable and the findings have mundane realism (having real life application). ***De-merits:*** low internal validity, no control on confounding factors, results can be biased etc.

Lesson 5

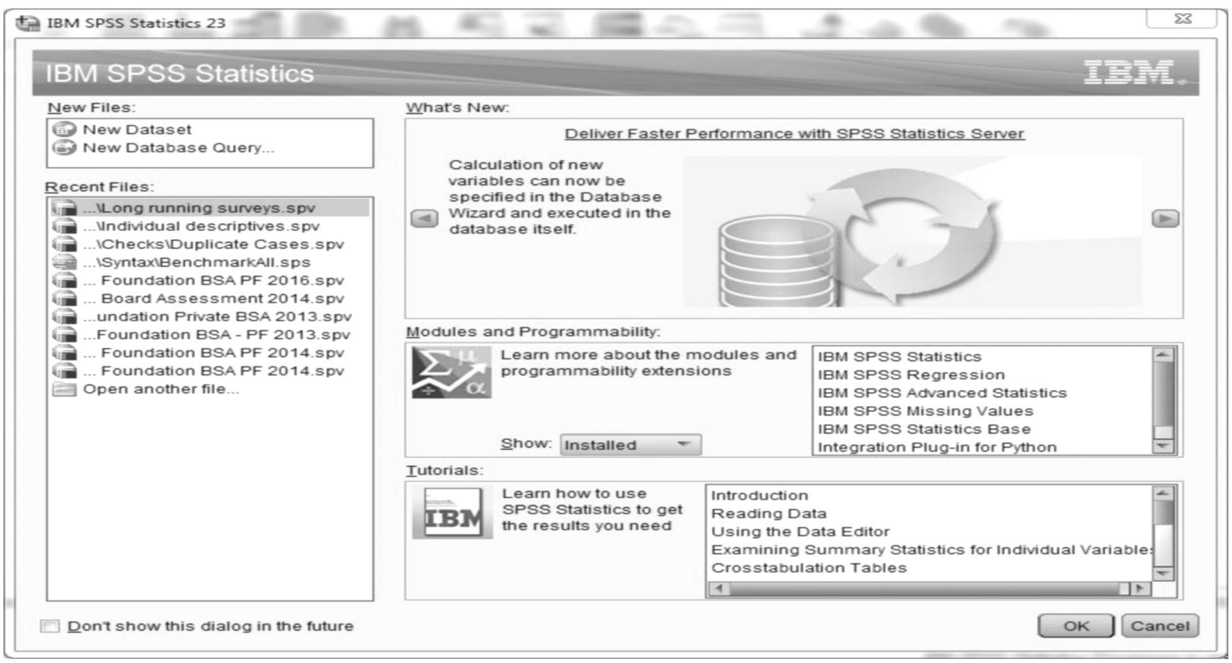
INTRODUCTION TO SPSS

Introduction to SPSS - Statistical Package for Social Science

SPSS for Windows is a popular and comprehensive data analysis package containing a multitude of features designed to facilitate the completion of a wide range of statistical analyses. It was developed for the analysis of data in the social sciences.

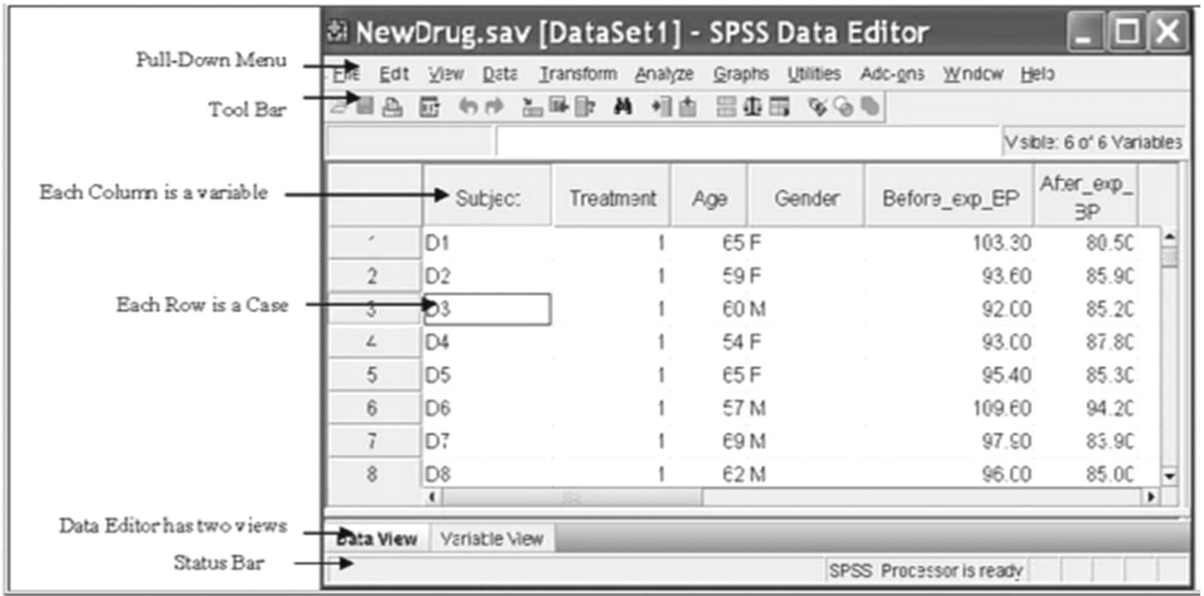
SPSS datasets always have 2-dimensional table structure where the rows typically represent cases (such as individuals or households) and the columns or variables represent measurements (such as age, sex or household income). SPSS can read and write data from ASCII text files, other statistics packages, spreadsheets and databases.

Once SPSS has been activated, a start-up window will appear, which allows you to select various options. You can open and create new file or open an existing file. It also shows the previously opened files.



After opening the file main screen shows spread sheet with two bars i.e. Menu Bar and Tool icon bar. Also, at the right bottom of the sheet show two views for data entry i.e. Data view and Variable view.

SPSS has three windows for working with data. 1. The Data Editor Window (.sav) i.e. Data view and Variable view, 2. The Output Viewer Window

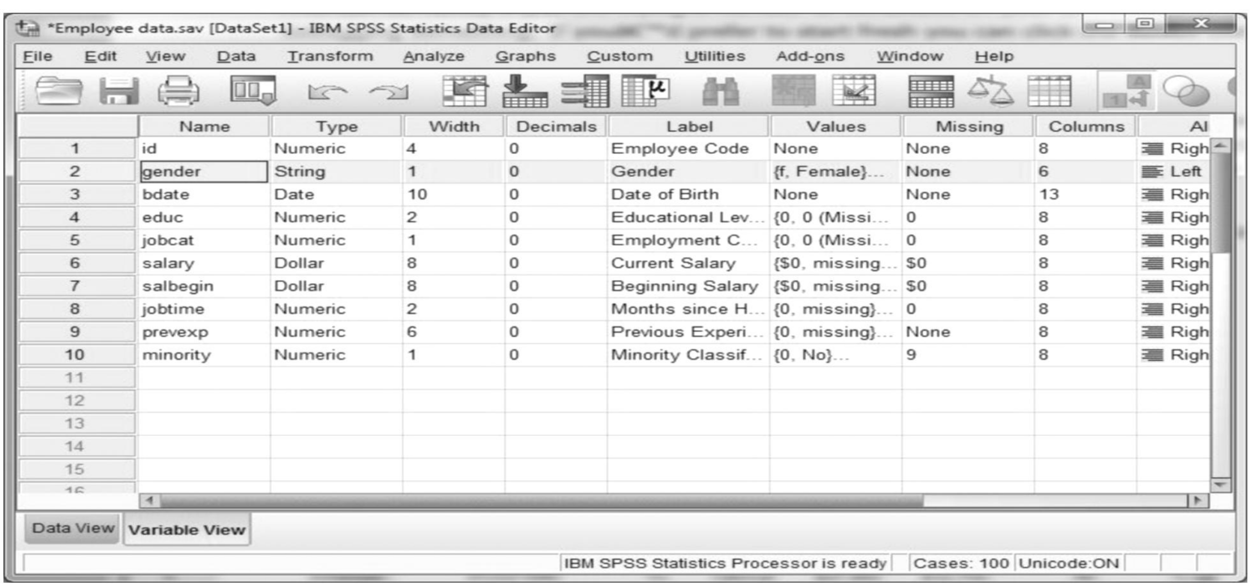


Variable View

Rows define the variable characteristics: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure.

The Measure of variables in the data set is important:

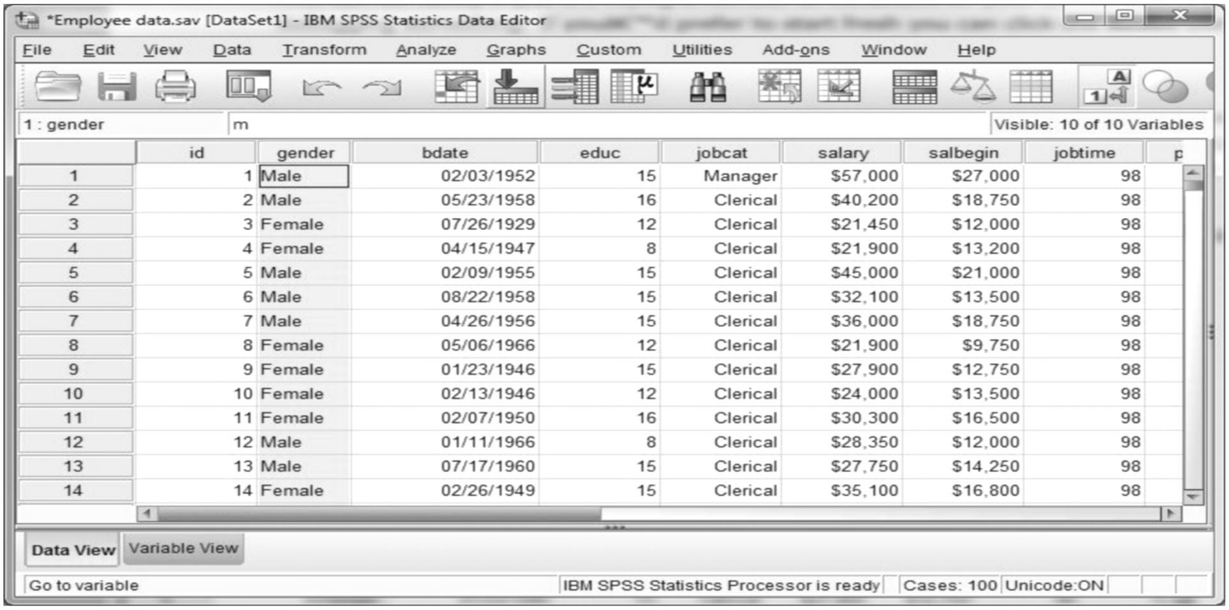
- Scale = “continuous” – e.g., age, weight, income
- Ordinal = “ordered” – categories that can be ranked (e.g., level of satisfaction or agreement, Likert-type scales)
- Nominal = “names” – categories that cannot be ranked (e.g., ID number)
- String = Letter plus numbers like Type 2.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Al
1	id	Numeric	4	0	Employee Code	None	None	8	Righ
2	gender	String	1	0	Gender	{f, Female}...	None	6	Left
3	bdate	Date	10	0	Date of Birth	None	None	13	Righ
4	educ	Numeric	2	0	Educational Lev...	{0, 0 (Missi...	0	8	Righ
5	jobcat	Numeric	1	0	Employment C...	{0, 0 (Missi...	0	8	Righ
6	salary	Dollar	8	0	Current Salary	{$0, missing...	$0	8	Righ
7	salbegin	Dollar	8	0	Beginning Salary	{$0, missing...	$0	8	Righ
8	jobtime	Numeric	2	0	Months since H...	{0, missing}...	0	8	Righ
9	prevexp	Numeric	6	0	Previous Experi...	{0, missing}...	None	8	Righ
10	minority	Numeric	1	0	Minority Classif...	{0, No}...	9	8	Righ
11									
12									
13									
14									
15									

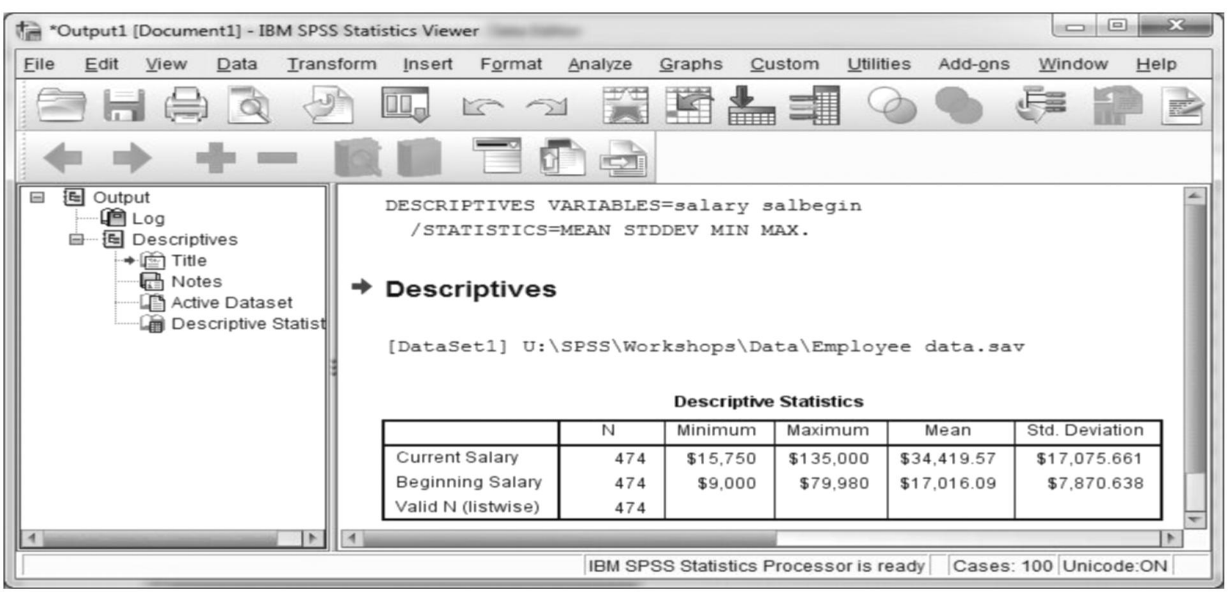
Data View

- Rows are cases (participants in the study).
- Columns are variables. A variable could be the answer to a question or any other piece of information recorded on each case.



The Output Viewer Window (.spv)

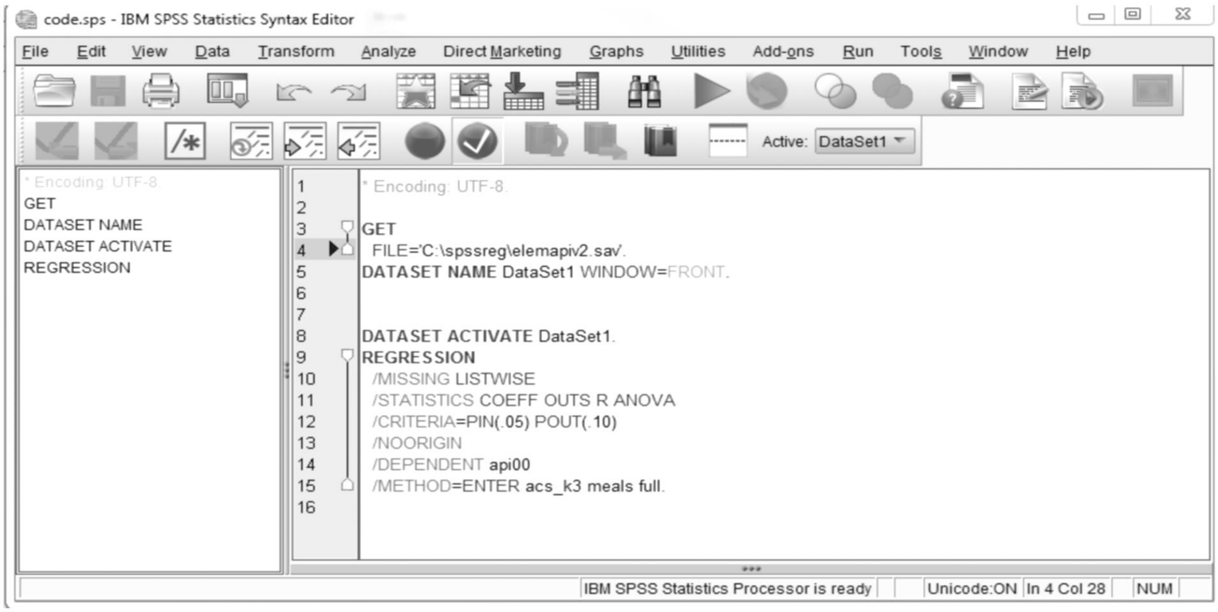
shows results of data analysis. The left-hand side is an outline of all of the output in the file. The right side is the actual output. To shrink or enlarge either side put your cursor on the line that divides them. When the double headed arrow appears, hold the left mouse button and move the line in either direction. Release the button and the size will be adjusted.



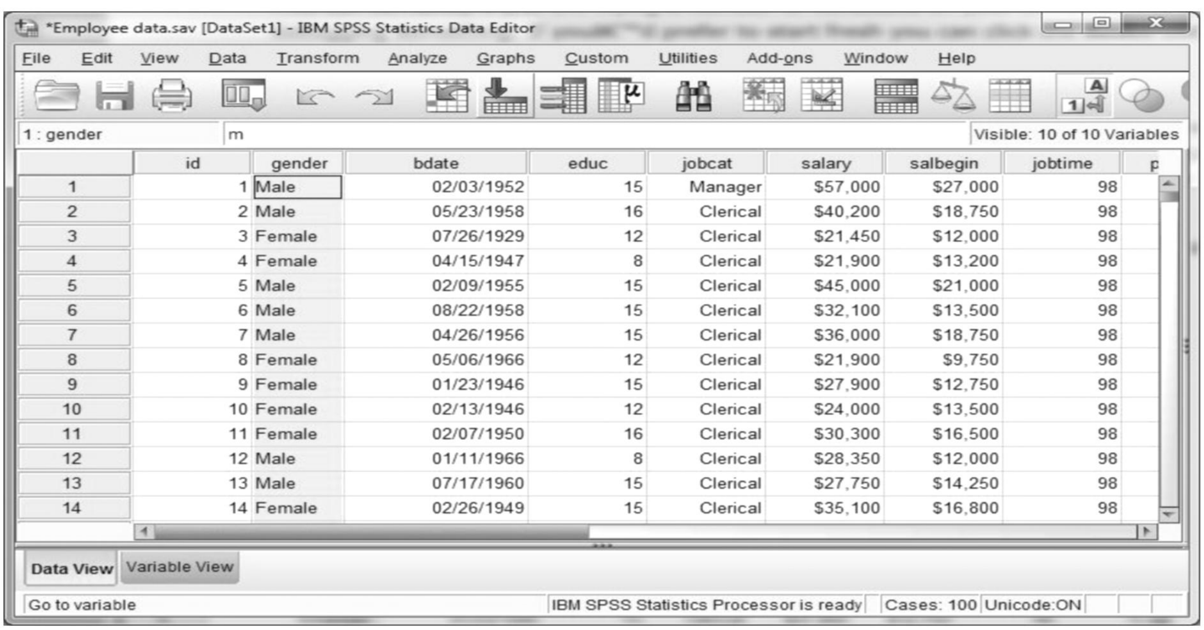
The Syntax Editor Window (.sps)

shows the syntax command script. This is also where you can type and run your own syntax commands. The Syntax window would be activated if you pasted the commands from the dialog box to it, or if you wrote you own syntax--something we will not focus on here. Syntax files end in the

extension .sps.



Data view in SPSS



Menu Bar Commands

Below is a brief reference guide to each of the menus and some of the options that they contain

- *File* includes all of the options you typically use in other programs, such as open, save, exit.
- *Edit* includes the typical cut, copy, and paste commands, and allows you to specify various options for displaying data and output.
- *View:* The option most frequently used is value labels.
- *Data* allows you to select several options ranging from displaying data that is sorted by a specific variable to selecting certain cases for subsequent analyses. For example, sorting cases, merging or aggregating files, and selecting or weighting cases.
- *Transform* includes several options to change current variables. It is used for recoding, computing new variables and dealing with missing values. example, you can change continuous variables to categorical variables, change scores into rank scores, add a constant to variables, etc.
- *Analyze* includes all of the commands to carry out statistical analyses and to calculate descriptive statistics.
- *Graphs* includes the commands to create various types of graphs including box plots, histograms, line graphs, and bar charts.
- *Utilities* allows you to list file information which is a list of all variables, there labels, values, locations in the data file, and type.
- *Add-ons* are programs that can be added to the base SPSS package. You probably do not have access to any of those.
- *Window* can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax). Since we have a data file and an output file open.
- *Help* has many useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide. Using topics, you can use the index option to type in any key word and get a list of options, or you can view the categories and subcategories available under contents. This is an excellent tool and can be used to troubleshoot most problems.

Toolbar Icons

The 19 Icons directly under the Menu bar provide shortcuts to many common commands that are available in specific menus.

 This icon gives you the option to open a previously saved file (if you are in the data editor, SPSS assumes you want to open a data file; if you are in the output viewer, it will offer to open a viewer file).

 This icon allows you to save files. It will save the file you are currently working on (be it data or output). If the file hasn’t already been saved it will produce the Save Data As dialog box.

 This icon activates a dialog box for printing whatever you are currently working on (either the data editor or the output). By default, SPSS will print everything in the output window, so a useful way to save trees is to print only a selection of the output.

 Clicking on this icon will activate a list of the last 12 dialog boxes that you used. You can select any box from the list and it will appear on the screen. This icon makes it easy for you to repeat parts of an analysis.

 This icon enables you to go directly to a case (a row in the data editor). This button is useful if you are working on large data files: if you were analyzing a survey with 3000 respondents it would get pretty tedious scrolling down the data sheet to find the responses of participant 2407. By clicking on this icon, you can skip straight to the case by typing the case number required.

 Similar to the previous icon, clicking this button activates a function that enables you to go directly to a variable (i.e., a column in the data editor).

 Clicking on this icon opens a dialog box that shows you the variables in the data editor and summary information about each one.

 Click this button to search for words or numbers in your data file and output window. In the data editor it will search within the variable (column) that is currently active.

 Clicking on this icon inserts a new case in the data editor (so it creates a blank row at the point that is currently highlighted in the data editor). This function is very useful if you need to add new data at a particular point in the data editor.

 Clicking on this icon creates a new variable to the left of the variable that is currently active (to activate a variable simply click once on the name at the top of the column).

 Clicking on this icon is a shortcut to the function Split-1file. There are often situations in which you might want to analyze groups of cases separately. In SPSS we differentiate groups of cases by using a coding variable, and this function lets us divide our output by such a variable.

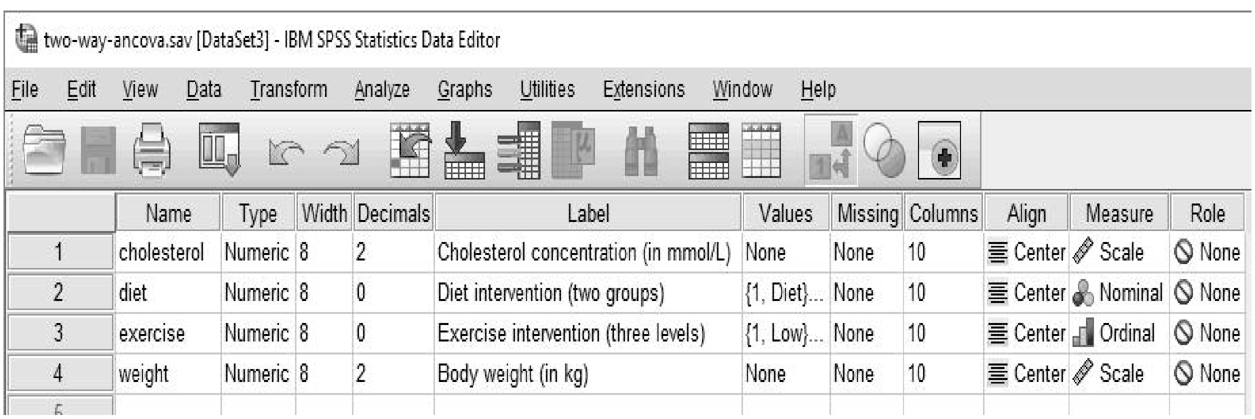
 This icon shortcuts to the function. This function is necessary when we come to input frequency data (see Section 18.5.2.2) and is useful for some advanced issues in survey sampling.

 This icon is a shortcut to the function. If you want to analyze only a portion of your data, this is the option for you. This function allows you to specify what cases you want to include in the analysis.

 Clicking on this icon will either display or hide the value labels of any coding variables. For example, if we coded gender as 1 = female, 0 = male then the computer knows that every time it comes across the value 1 in the Gender column, that person is a female. If you press this icon, the coding will appear on the data editor rather than the numerical values; so, you will see the words male and female in the Gender column rather than a series of numbers.

Variable View in SPSS

Every row of the variable view represents a variable, and you set characteristics of a particular variable by entering information into the following labelled columns (play around and you'll get the hang of it):


two-way-ancova.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	cholesterol	Numeric	8	2	Cholesterol concentration (in mmol/L)	None	None	10	Center	Scale	None
2	diet	Numeric	8	0	Diet intervention (two groups)	{1, Diet}...	None	10	Center	Nominal	None
3	exercise	Numeric	8	0	Exercise intervention (three levels)	{1, Low}...	None	10	Center	Ordinal	None
4	weight	Numeric	8	2	Body weight (in kg)	None	None	10	Center	Scale	None

You can enter a name in this column for each variable. This name will appear at the top of the corresponding column in the data view, and helps you to identify variables in the data view.

Type You can have different types of data. Mostly you will use numeric variables (which means that the variable contains numbers and is the default). You will come across string variables, which consist of strings of letters.

Width By default, when a new variable is created, SPSS sets it up to be numeric and to store 8 digits/characters, but you can change this value by typing a new number in this column in the dialog box. For numeric variables 8 digits is fine (unless you have very large numbers), but for string variables you will often make this value bigger (you can't write a lot in only 8 characters).

Decimals Another default setting is to have 2 decimal places displayed. (You'll notice that if you don't change this option then when you type in whole numbers to the data editor SPSS adds a decimal place with two zeros after it, which can be disconcerting.) If you want to change the number of decimal places for a given variable then replace the 2 with a new value or increase or decrease the value using.

Label The name of the variable (see above) has some restrictions on characters, and you also wouldn't want to use huge long names at the top of your columns (they become hard to read). Therefore, you can write a longer variable description in this column.

Values This column is for assigning numbers to represent groups of people.

Missing This column is for assigning numbers to missing data.

Columns Enter a number into this column to determine the width of the column, that is, how many characters are displayed in the column. (This characteristic differs from, which determines the width of the variable itself – you could have a variable of 10 characters but by setting the column width to 8 you would see only 8 of the 10 characters of the variable in the data editor.) It can be useful to increase the column width if you have a string variable that exceeds 8 characters, or a coding variable with value labels that exceed 8 characters.

Align You can use this column to select the alignment of the data in the corresponding column of the data editor. You can choose to align the data to the right, left or center.

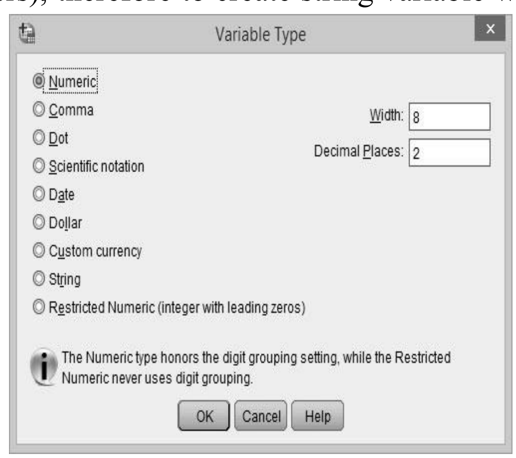
Measure This is where you define the level at which a variable was measured (Nominal, Ordinal or Scale).

There are some procedures in SPSS that attempt to run analyses automatically without you needing to think about what you're doing (one example is the Automatic Linear Modeling option in the Regression part of the Analyze menu). To think on your behalf, SPSS needs to know whether a variable is a predictor an outcome both, a variable that splits the analysis by different groups a variable that selects out part of the data or a variable that has no predefined role. These roles can be useful if you're chugging out huge numbers of analyses and want to automate them.

Computing And Recording Techniques

Creating a String Variable

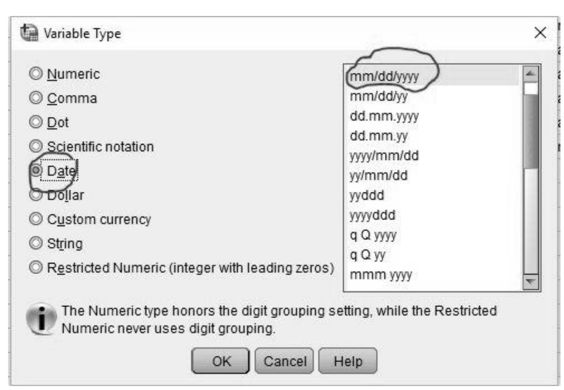
1. Click in the first white cell in the column labeled Name.
2. Type the word 'Name'.
3. Move off this cell using the arrow keys on the keyboard (you can also just click on a different cell, but this is a very slow way of doing it).
4. Move into the column labeled TYPE. Click on it and a dialogue box will open.
5. Choose the STRING option and click OK to make the variable string. SPSS assume that we want a numeric variable (i.e., numbers), therefore to create string variable we have to change the variable type.



6. Finally, to specify Measure and selecting either Nominal, Ordinal or Scale from the drop-down list.

Creating a Date Variable

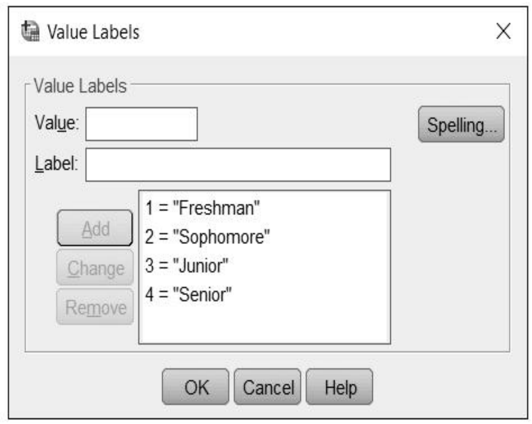
1. To enter date variables into SPSS we use the same procedure as with the string variable, except that we need to change the variable type.
2. Move into the column labelled TYPE. Click on it and a dialogue box will open.



3. Choose the DATE option and click OK to make the variable string.

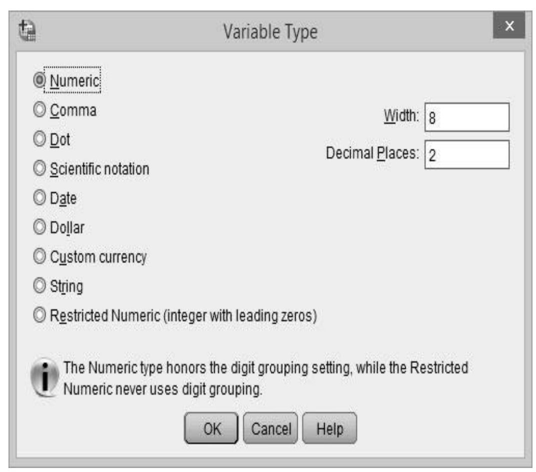
Creating Coding Variables

1. A coding variable (also known as a grouping variable) uses numbers to represent different groups of data. As such, it is a numeric variable, but these numbers represent names (i.e., it is a nominal variable).
2. For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3).
3. Example: In the sample data set, the variable Rank represents the student's class rank. The values 1, 2, 3, 4 represent the categories Freshman, Sophomore, Junior, and Senior, respectively.



Creating a Numeric Variable

1. Numeric variables are the easiest ones to create because SPSS assumes this format for data. e.g. Age, No. of Friends, total income etc.
2. Click in the first white cell in the column labeled Name and type Age.
3. Move into the column labeled TYPE. Click on it and a dialogue box will open. Choose Numeric and click OK.



4. Specify the level at which a variable was measured by going to the column labeled Measure and selecting SCALE.

Lesson 6

DESCRIPTIVE STATISTICS

Although researchers have developed a variety of different statistical procedures to organize and interpret data, these different procedures can be classified into two general categories: descriptive and inferential. Each of these segments is important, offering different techniques that accomplish different objectives. The first category, descriptive statistics, consists of statistical procedures that are used to simplify and summarize data.

Descriptive Statistics

Descriptive statistics are statistical procedures used to summarize, organize, and simplify data. Descriptive statistics are techniques that take raw scores and organize or summarize them in a form (graphically or numerically) that is more manageable. Often the scores are organized in a table or a graph so that it is possible to see the entire set of scores. Another common technique is to summarize a set of scores by computing an average. Note that even if the data set has hundreds of scores, the average provides a single descriptive value for the entire set.

Two objectives for descriptive statistics are:

We want to choose a statistic that shows how different units seems similar i.e. Central tendency.

We want to choose another statistic that shows how they differ i.e. Variability.

For example, you want to study the status of different leisure activities by gender. You distribute a survey and ask participants how many times they did each of the following in the past year:

- Go to a library
- Watch a movie at a theater
- Visit a national park

Your data set is the collection of responses to the survey. Now you can use descriptive statistics to find out the overall frequency of each activity (distribution), the averages for each activity (central tendency), and the spread of responses for each activity (variability).

Descriptive statistics usually include:

- Tables/Distributions
- Central Tendency
- Measures of Dispersion
- Graphs/Charts

Presenting The Data: Graphs

Descriptive statistics are very important because if we simply presented our raw data it would be hard to visualize what the data was showing, especially if there was a lot of it. Descriptive statistics therefore enables us to present the data in a more meaningful way, which allows simpler interpretation of the data. For example, if we had the results of 100 pieces of students' coursework, we may be interested in the overall performance of those students. We would also be interested in the distribution or spread of the marks.

When we use descriptive statistics, it is useful to summarize the group of data using a combination of tabulated description (i.e., tables), graphical description (i.e., graphs and charts) and statistical commentary (i.e., a discussion of the results).

Histograms

A histogram is a graphical display of a distribution. It presents the same information as a frequency table but in a way that is even quicker and easier to grasp.

Lesson 7

FREQUENCY DISTRIBUTION-I

The results from a research study usually consist of pages of numbers corresponding to the measurements, or scores, collected during the study. The immediate problem for the researcher is to organize the scores into some comprehensible form so that any patterns in the data can be seen easily and communicated to others. This is the job of descriptive statistics: to simplify the organization and presentation of data. One of the most common procedures for organizing a set of data is to place the scores in a frequency distribution.

Frequency Distribution

A frequency distribution is an organized tabulation of the number of individuals located in each category on the scale of measurement.

A frequency distribution takes a disorganized set of scores and places them in order from highest to lowest, grouping together individuals who all have the same score. If the highest score is *X* _ 10, for example, the frequency distribution groups together all the 10s, then all the 9s, then the 8s, and so on. Thus, a frequency distribution allows the researcher to see "at a glance" the entire set of scores. It shows whether the scores are generally high or low, whether they are concentrated in one area or spread out across the entire scale, and generally provides an organized picture of the data.

A frequency distribution can be structured either as a table or as a graph, but in either case, the distribution presents the same two elements:

1. The set of categories that make up the original measurement scale.

2. A record of the frequency, or number of individuals in each category.

Thus, a frequency distribution presents a picture of how the individual scores are distributed on the measurement scale—hence the name *frequency distribution*.

Frequency Distribution Tables

The simplest frequency distribution table presents the measurement scale by listing the different measurement categories (*X* values) in a column from highest to lowest. Beside each *X* value, we indicate the frequency, or the number of times that particular measurement occurred in the data.

It is customary to use an X as the column heading for the scores and an f as the column heading for the frequencies.

It is customary to list categories from highest to lowest, but this is an arbitrary arrangement. Many computer programs list categories from lowest to highest.

Proportions And Percentages

In addition to the two basic columns of a frequency distribution, there are other measures that describe the distribution of scores and can be incorporated into the table. The two most common are proportion and percentage. Proportion measures the fraction of the total group that is associated with each score. In Example 2.2, there were two individuals with $X = 4$. Thus, 2 out of 10 people had $X = 4$, so the proportion would be 2/10= 0.20. In general, the proportion associated with each score is

$$\textit{Proportion} = P = f/N$$

Because proportions describe the frequency (f) in relation to the total number (N), they often are called *relative frequencies*. Although proportions can be expressed as fractions (for example, 2/10), they more commonly appear as decimals. A column of proportions, headed with a *p,* can be added to the basic frequency distribution table.

In addition to using frequencies (f) and proportions (p), researchers often describe a distribution of scores with percentages. For example, an instructor might describe the results of an exam by saying that 15% of the class earned *A*s, 23% earned *B*s, and so on. To compute the percentage associated with each score, you first find the proportion (p) and then multiply by 100:

$\textit{Percentage} = p(100) = f/N\ (100)$

Percentages can be included in a frequency distribution table by adding a column headed with %.

X	f	$p = f/N$	$\% = p(100)$
5	1	1/10 = 0.10	10%
4	2	2/10 = 0.20	20%
3	3	3/10 = 0.30	30%
2	3	3/10 = 0.30	30%
1	1	1/10 = 0.10	10%

Grouped Frequency Distribution

If we were to list all of the individual scores from $X = 96$ down to $X = 41$, it would take 56 rows to complete the frequency distribution table. Although this would organize the data, the table would be long and cumbersome. Remember: The purpose for constructing a table is to obtain a relatively simple, organized picture of the data. This can be accomplished by grouping the scores into intervals and then listing the intervals in the table instead of listing each individual score. For example, we could construct a table showing the number of students who had scores in the 90s, the number with scores in the 80s, and so on. The result is called a *grouped frequency distribution table* because we are presenting groups of scores rather than individual values. The groups, or intervals, are called *class intervals*.

Grouped frequency distribution tables—group the scores into intervals and list these intervals in the frequency distribution table. Remember, when the scores are whole numbers, the number of rows is determined by highest scores– lowest scores+ 1

Rule of thumb is to have 5 to 10 intervals or less (of equal width)

Example

An instructor has obtained the set of N=25 exam scores shown here. To help organize these scores, we will place them in a frequency distribution table. The scores are:

82, 75, 88, 93, 53, 84, 87, 58, 72, 94, 69, 84, 61, 91, 64, 87, 84, 70, 76, 89, 75, 80, 73, 78, 60

The first step is to determine the range of scores. For these data, the smallest score is X=53 and the largest score is X=94, so a total of 42 rows would be needed for a table that lists each individual score. Because 42 rows would not provide a simple table, we have to group the scores into class intervals.

The best method for finding a good interval width is a systematic trial-and-error approach that uses guidelines 1 and 2 simultaneously. Specifically, we want about 10 intervals and we want the interval width to be a simple number. For this example, the scores cover a range of 42 points, so we will try several different interval widths to see how many intervals are needed to cover this range. For example, if each interval is 2 points wide; it would take 21 intervals to cover a range of 42 points. This is too many, so we move on to an interval width of 5 or 10 points. The following table shows how many intervals would be needed for these possible widths:

Width	Number of Intervals Needed to Cover a Range of 42 Points	
2	21	(too many)
5	9	(OK)
10	5	(too few)

Notice that an interval width of 5 will result in about 10 intervals, which is exactly what we want. The next step is to actually identify the intervals. The lowest score for these data is X=53, so the lowest interval should contain this value. Because the interval should have a multiple of 5 as its bottom score, the interval should begin at 50. The interval has a width of 5, so it should contain 5 values: 50, 51, 52, 53, and 54. Thus, the bottom interval is 50–54. The next interval would start at 55 and go to 59. Note that this interval also has a bottom score that is a multiple of 5, and contains exactly 5 scores (55, 56, 57, 58, and 59). The complete frequency distribution table showing all of the class intervals is presented in Table below:

X	*f*
90–94	3
85–89	4
80–84	5
75–79	4
70–74	3
65–69	1
60–64	3
55–59	1
50–54	1

This grouped frequency distribution table shows the data from Example 2.4. The original scores range from a high of X=94 to a low of X=53. This range has been divided into 9 intervals with each interval exactly 5 points wide. The frequency column (f) lists the number of individuals with scores in each of the class intervals.

Frequency Distribution Graphs/Charts

When the data consist of numerical scores that have been measured on an interval or ratio scale, there are two options for constructing a frequency distribution graph. The two types of graphs are called *histograms* and *polygons.*

Histograms To construct a histogram, you first list the numerical scores (the categories of measurement) along the *X*-axis. Then you draw a bar above each *X* value so that

a. The height of the bar corresponds to the frequency for that category.
b. For continuous variables, the width of the bar extends to the real limits of the category. For discrete variables, each bar extends exactly half the distance to the adjacent category on each side.

Polygons The second option for graphing a distribution of numerical scores from an interval or ratio scale of measurement is called a polygon. To construct a polygon, you begin by listing the numerical scores (the categories of measurement) along the *X*-axis. Then,

a) A dot is centered above each score so that the vertical position of the dot corresponds to the frequency for the category.
b) A continuous line is drawn from dot to dot to connect the series of dots.
c) The graph is completed by drawing a line down to the *X*-axis (zero frequency) at each end of the range of scores. The final lines are usually drawn so that they reach the *X*-axis at a point that is one category below the lowest score on the left side and one category above the highest score on the right side.

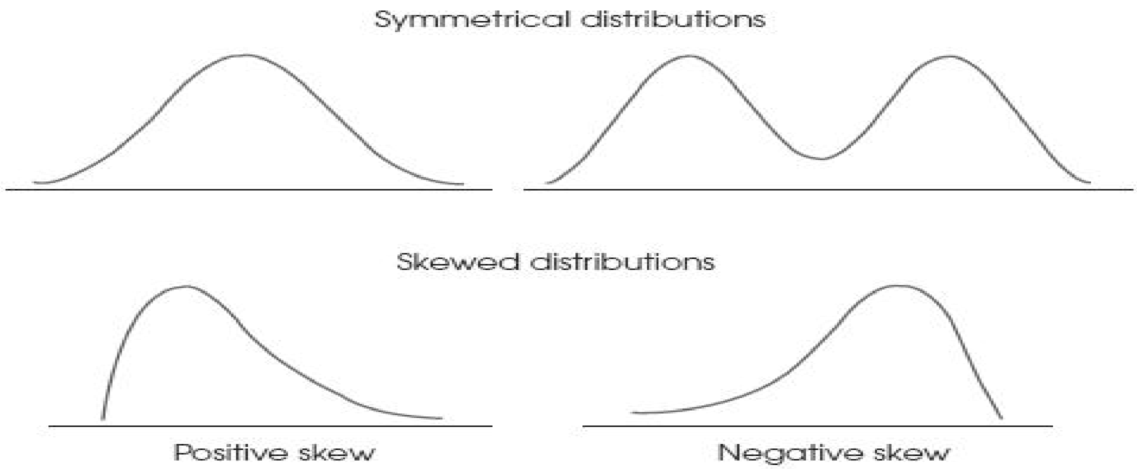
When the scores are measured on a nominal or ordinal scale (usually non-numerical values), the frequency distribution can be displayed in a *bar graph.*

Bar Graphs A bar graph is essentially the same as a histogram, except that spaces are left between adjacent bars. For a nominal scale, the space between bars emphasizes that the scale consists of separate, distinct categories. For ordinal scales, separate bars are used because you cannot assume that the categories are all the same size.

To construct a bar graph, list the categories of measurement along the *X*-axis and then draw a bar above each category so that the height of the bar equals the frequency for the category.

Nearly all distributions can be classified as being either ***symmetrical*** **or** ***skewed*****.** In a symmetrical distribution, it is possible to draw a vertical line through the middle so that one side of the distribution is a mirror image of the other.

In a skewed distribution, the scores tend to pile up toward one end of the scale and taper off gradually at the other end.


Symmetrical distributions
Skewed distributions
Positive skew
Negative skew

Lesson 8

FREQUENCY DISTRIBUTION-II

Percentiles and Percentile Ranks

Although the primary purpose of a frequency distribution is to provide a description of an entire set of scores, it also can be used to describe the position of an individual within the set. Individual scores, or *X* values, are called *raw scores*. By themselves, raw scores do not provide much information. For example, if you are told that your score on an exam is X _ 43, you cannot tell how well you did relative to other students in the class. To evaluate your score, you need more information, such as the average score or the number of people who had scores above and below you. With this additional information, you would be able to determine your relative position in the class. Because raw scores do not provide much information, it is desirable to transform them into a more meaningful form. One transformation that we consider changes raw scores into *percentiles*.

The rank or percentile rank of a particular score is defined as the percentage of individuals in the distribution with scores equal to or less than the particular value.

When a score is identified by its percentile rank, the score is called a percentile.

Suppose, for example, that you have a score of $X = 43$ on an exam and that you know that exactly 60% of the class had scores of 43 or lower. Then your score $X = 43$ has a percentile rank of 60%, and your score would be called the 60th percentile. Notice that *percentile rank* refers to a percentage and that *percentile* refers to a score. Also notice that your rank or percentile describes your exact position within the distribution.

To determine percentiles or percentile ranks, the first step is to find the number of individuals who are located at or below each point in the distribution. This can be done most easily with a frequency distribution table by simply counting the number who are in or below each category on the scale. The resulting values are called ***cumulative frequencies*** because they represent the accumulation of individuals as you move up the scale.

The cumulative frequencies show the number of individuals located at or below each score. To find percentiles, we must convert these frequencies into percentages. The resulting values are called ***cumulative percentages*** because they show the percentage of individuals who are accumulated as you move up the scale.

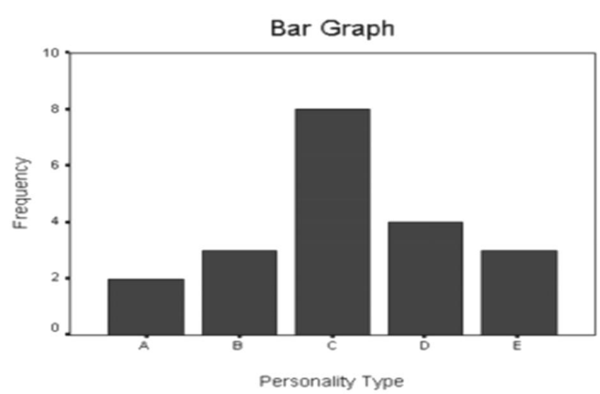
Stem and Leaf Displays

In 1977, J.W. Tukey presented a technique for organizing data that provides a simple alternative to a grouped frequency distribution table or graph (Tukey, 1977). This technique, called a *stem and leaf display*, requires that each score be separated into two parts: The first digit (or digits) is called the *stem*, and the last digit is called the *leaf*. For example, $X = 85$ would be separated into a stem of 8 and a leaf of 5. Similarly, $X = 42$ would have a stem of 4 and a leaf of 2.

Data			Stem and Leaf Display	
83	82	63	3	23
62	93	78	4	26
71	68	33	5	6279
76	52	97	6	283
85	42	46	7	1643846
32	57	59	8	3521
56	73	74	9	37
74	81	76		

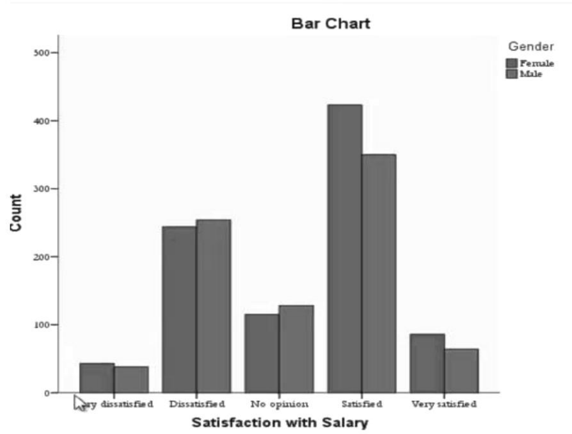
Art of Presenting Data

Bar Graph— like a histogram, a bar is drawn above each X value, so that the height of the bar corresponds to the frequency of the score. Usually is from discrete (nominal or ordinal level data).



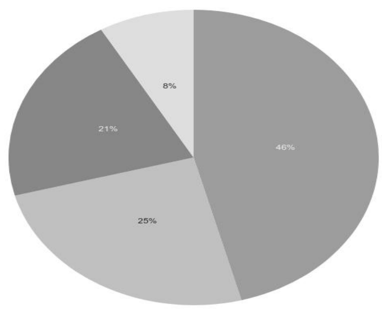
Clustered Bar Graph— can be used when you have two or more nominal or ordinal variables and want to illustrate the differences in the categories of these two variables based on some statistic.

e.g. comparing satisfaction with salary in males in females.



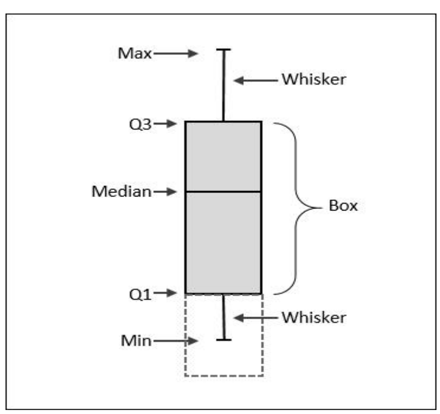
Pie Chart— is a special chart that uses Pie slices to show the relative sizes of the data.

Uses Nominal data; e.g. which movie types are most liked, (you can see in a glance that Romantic movies (blue slice) are liked more.



Romantic	Action	Horror	Drama
46%	25%	21%	8%

Box Plot— A boxplot is a graph that gives you a good indication of how the values in the data are spread out. They useful when comparing distributions between many groups or datasets.



Exercise: Constructing Frequency Distribution

For Frequency Distribution:

Enter data in data view→analyze→descriptive statistics→frequencies→drag scores in variable(s) box→ok. So here you can find frequency distribution table

For the grouped distribution got to transform→ visual binning→ send scores in variables to bin box→continue→name variable in binned variable box which will add a variable in your data set (I.e. groups), to make cut points click→ make cut points→ define value for first cut point (I.e. 5 for data starting from 2), define width (it will automatically define number of cut points)→apply→make labels→ok→ok

If you want exact frequency distribution that we calculated manually, click analyze→descriptive statistics→frequencies→drag binned variable(s) box→ok. So here you can find frequency distribution table of grouped data

For Graphs:

analyze→descriptive statistics→frequencies→ select variable (scores)→statistics tab→define percentile, minimum, maximum→continue→ click on charts→histograms→continue→ok

For Cumulative Percentage Curve/Ogive Percentile Graph

Graphs→legacy dialogue→histogram→drop variable in variable box→ok

You can also make it through legacy chart, for this click on Graphs→chart builder→ok→select histogram→select, hold and drag to the chart preview box→define variable on X-Axis→ok

For Polygon

Graphs→legacy dialogues→line chart→simple→define→drag score to category Axis→ok

For Ogive

Graphs→legacy dialogues→line chart→simple→define→drag score to category Axis→select cum%→OK

For Bar Diagram

Graphs→legacy dialogues→bar→simple→define→drag score to category Axis→ok

Similar procedure can be used for nominal or ordinal data

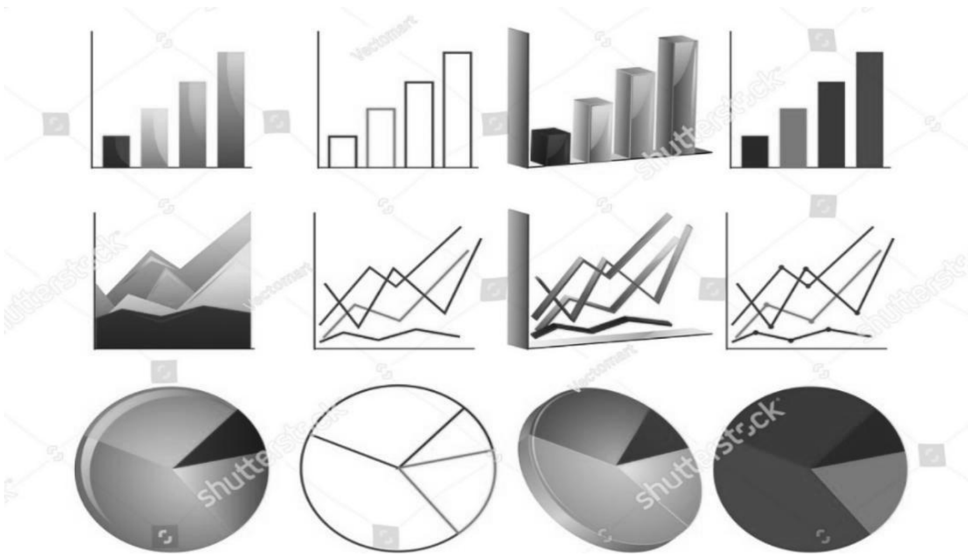
For Pie Chart

Graphs→legacy dialogues→pie→define→drag binned data to “define slices by”box→% of cases

Chart can also be edit by double clicking on chart, values and titles can also be placed.

For Cluster Bar Chart

First add categorical data, Graphs→legacy dialogues→bar→clustered→define→define clusters by gender→drag binned score in category axis→ok



Lesson 9

MEASURE OF CENTRAL TENDENCY-I

Central tendency is a statistical measure to determine a single score that defines the center of a distribution. The goal of central tendency is to find the single score that is most typical or most representative of the entire group.

In everyday language, central tendency attempts to identify the "average" or "typical" individual. This average value can then be used to provide a simple description of an entire population or a sample. In addition to describing an entire distribution, measures of central tendency are also useful for making comparisons between groups of individuals or between sets of figures. For example, weather data indicate that for Seattle, Washington, the average yearly temperature is 53° and the average annual precipitation is 34 inches. By comparison, the average temperature in Phoenix, Arizona, is 71° and the average precipitation is 7.4 inches. The point of these examples is to demonstrate the great advantage of being able to describe a large set of data with a single, representative number.

Unfortunately, there is no single, standard procedure for determining central tendency. To deal with these problems, statisticians have developed three different methods for measuring central tendency: the mean, the median, and the mode.

MEAN

The mean for a distribution is the sum of the scores divided by the number of scores.

The *mean*, also known as the arithmetic average, is computed by adding all the scores in the distribution and dividing by the number of scores. The mean for a population is identified by the Greek letter mu, μ (pronounced "mew"), and the mean for a sample is identified by *M* or $\bar{x}$, (read "x-bar").

The formula for the *population mean* is

$$\mu = \Sigma X/N$$

First, add all the scores in the population, and then divide by *N*. For a sample, the computation is exactly the same, but the formula for the *sample mean* uses symbols that signify sample values:

$$\text{Sample mean} = M = \Sigma x/n$$

Example:

For a population of $N = 4$ scores,

3, 7, 4, 6

the mean is

$$\mu = \frac{\Sigma X}{N} = \frac{20}{4} = 5$$

Weighted Mean

When we have two or more sets of data and want to find overall mean for the combined group.

Sample 1= 4, 5, 6, 7, 8

Sample 2= 4, 6, 8, 10, 12

$$\text{overall mean} = M = \frac{\Sigma X \text{ (overall sum for the combined group)}}{n \text{ (total number in the combined group)}}$$

$$= \frac{\Sigma X_1 + \Sigma X_2}{n_1 + n_2}$$

Mean from Grouped Data

The first step is to determine the midpoint of each interval, or class. These midpoints must then be multiplied by the frequencies of the corresponding classes. The sum of the products divided by the total number of values will be the value of the **mean.**

$$\bar{x} = \frac{\sum f x}{n}$$

MEDIAN

The second measure of central tendency we consider is called the *median*. The goal of the median is to locate the midpoint of the distribution. Unlike the mean, there are no specific symbols or notation to identify the median. Instead, the median is simply identified by the word *median*. In addition, the definition and the computations for the median are identical for a sample and for a population.

If the scores in a distribution are listed in order from smallest to largest, the **median** is the midpoint of the list. More specifically, the median is the point on the measurement scale below which 50% of the scores in the distribution are located. Defining the median as the *midpoint* of a distribution means that the scores are divided into two equal-sized groups.

This example demonstrates the calculation of the median when *n* is an odd number. With an odd number of scores, you list the scores in order (lowest to highest), and the median is the middle score in the list. Consider the following set of *N* =5 scores, which have been listed in order:

3, 5, 8, 10, 11

The middle score is *X* =8, so the median is equal to 8. Using the counting method, with *N* =5 scores, the 50% point would be 2 ½ scores. Starting with the smallest scores, we must count the 3, the 5, and the 8 before we reach the target of at least 50%. Again, for this distribution, the median is the middle score, *X*=8.

Finding Median with Grouped Data

$$M_m = l + \left(\frac{\frac{n}{2} - cf}{f} \right) h$$

Where

l= Lower limit of median class,

n= number of observations,

cf= cumulative frequency of class preceding the median class,

f= frequency of median class,

h= class size (assuming class size to be equal)

MODE

The final measure of central tendency that we consider is called the *mode*. In its common usage, the word *mode* means "the customary fashion" or "a popular style." The statistical definition is similar in that the mode is the most common observation among a group of scores.

In a frequency distribution, the mode is the score or category that has the greatest frequency.

As with the median, there are no symbols or special notation used to identify the mode or to differentiate between a sample mode and a population mode. In addition, the definition of the mode is the same for a population and for a sample distribution. The mode is a useful measure of central tendency because it can be used to determine the typical or average value for any scale of measurement, including a nominal scale.

Extreme Scores or Skewed Distributions

When a distribution has a few extreme scores, scores that are very different in value from most of the others, then the mean may not be a good representative of the majority of the distribution.

The problem comes from the fact that one or two extreme values can have a large influence and cause the mean to be displaced. In this situation, the fact that the mean uses all of the scores equally can be a disadvantage. Consider, for example, the distribution of n =10 scores. For this sample, the mean is

$$M = \Sigma x/n = 203/10 = 20.3$$

Notice that the mean is not very representative of any score in this distribution. Although most of the scores are clustered between 10 and 13, the extreme score of X =100 inflates the value of ΣX and distorts the mean.

The median, on the other hand, is not easily affected by extreme scores. For this sample, n _ 10, so there should be five scores on either side of the median. The median is 11.50. Notice that this is a very representative value. Also note that the median would be unchanged even if the extreme score were 1000 instead of only 100.

Because it is relatively unaffected by extreme scores, the median commonly is used when reporting the average value for a skewed distribution. For example, the distribution of personal incomes is very skewed, with a small segment of the population earning incomes that are astronomical. These extreme values distort the mean, so that it is not very representative of the salaries that most of us earn.

The median is the preferred measure of central tendency when extreme scores exist.

Lesson 10

MEASURE OF CENTRAL TENDENCY-II

When To Use Which Measure Of Central Tendency?

The goal of central tendency is to find out one single value that best represents the entire distribution. Mean is preferred measure as it uses every score in the distribution and is related to standard deviation so is valuable measure in inferential statistics. But then when and why to use median and mode?

When To Use The Median?

- When there are extreme scores in data/ skewed distribution
- For undetermined or incomplete data points
- Open ended distribution (e.g. 5 or less, 20 or more)
- When data is ordinal scale

When To Use The Mode?

We consider three situations in which the mode is commonly used as an alternative to the mean, or is used in conjunction with the mean to describe central tendency.

1. **Nominal Scales** The primary advantage of the mode is that it can be used to measure and describe central tendency for data that are measured on a nominal scale. Recall that the categories that make up a nominal scale are differentiated only by name. Because nominal scales do not measure quantity (distance or direction), it is impossible to compute a mean or a median for data from a nominal scale. Therefore, the mode is the only option for describing central tendency for nominal data.
2. **Discrete Variables** Recall that discrete variables are those that exist only in whole, indivisible categories. Often, discrete variables are numerical values, such as the number of children in a family or the number of rooms in a house. When these variables produce numerical scores, it is possible to calculate means. In this situation, the calculated means are usually fractional values that cannot actually exist. For example, computing means generates results such as “the average family has 2.4 children and a house with 5.33 rooms.” On the other hand, the mode always identifies the most typical case and, therefore, it produces more sensible measures of central tendency. Using the mode, our conclusion would be “the typical, or modal, family has 2 children and a house with 5

rooms." In many situations, especially with discrete variables, people are more comfortable using the realistic, whole-number values produced by the mode.

3. **Describing Shape** Because the mode requires little or no calculation, it is often included as a supplementary measure along with the mean or median as a no-cost extra. The value of the mode (or modes) in this situation is that it gives an indication of the shape of the distribution as well as a measure of central tendency. Remember that the mode identifies the location of the peak (or peaks) in the frequency distribution graph. For example, if you are told that a set of exam scores has a mean of 72 and a mode of 80, you should have a better picture of the distribution than would be available from the mean alone.

Central Tendency And The Shape Of The Distribution

For a ***symmetrical distribution,*** the right-hand side of the graph is a mirror image of the left-hand side. If a distribution is perfectly symmetrical, the median is exactly at the center because exactly half of the area in the graph is on either side of the center. The mean also is exactly at the center of a perfectly symmetrical distribution because each score on the left side of the distribution is balanced by a corresponding score (the mirror image) on the right side. As a result, the mean (the balance point) is located at the center of the distribution. Thus, for a perfectly symmetrical distribution, the mean and the median are the same (Figure 10.1). If a distribution is roughly symmetrical, but not perfect, the mean and median are close together in the center of the distribution.

If a symmetrical distribution has only one mode, it is also in the center of the distribution. Thus, for a perfectly symmetrical distribution with one mode, all three measures of central tendency, the mean, the median, and the mode, have the same value. For a roughly symmetrical distribution, the three measures are clustered together in the center of the distribution. On the other hand, a bimodal distribution that is symmetrical [see Figure 10.1(b)] has the mean and median together in the center with the modes on each side. A rectangular distribution [see Figure 10.1(c)] has no mode because all X values occur with the same frequency. Still, the mean and the median are in the center of the distribution.

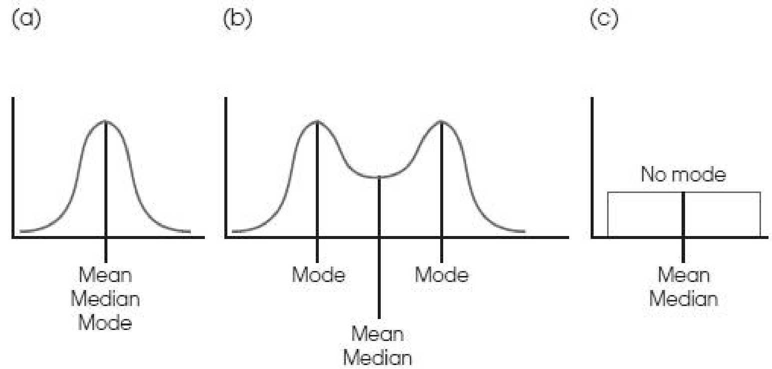


Figure 10.1 Measures of central tendency for three symmetrical distributions: normal, bimodal, and rectangular

In ***skewed distributions,*** especially distributions for continuous variables, there is a strong tendency for the mean, median, and mode to be located in predictably different positions. Figure 10.2(a), for example, shows a positively skewed distribution with the peak (highest frequency) on the left-hand side. This is the position of the mode. However, it should be clear that the vertical line drawn at the mode does not divide the distribution into two equal parts. To have exactly 50% of the distribution on each side, the median must be located to the right of the mode. Finally, the mean is located to the right of the median because it is the measure influenced most by the extreme scores in the tail and is displaced farthest to the right toward the tail of the distribution. Therefore, in a positively skewed distribution, the order of the three measures of central tendency from smallest to largest (left to right) is the mode, the median, and the mean.

Negatively skewed distributions are lopsided in the opposite direction, with the scores piling up on the right-hand side and the tail tapering off to the left. The grades on an easy exam, for example, tend to form a negatively skewed distribution [see Figure 10.2(b)]. For distribution with negative skew, the mode is on the right-hand side (with the peak), whereas the mean is displaced toward the left by the extreme scores in the tail. As before, the median is located between the mean and the mode. In order from the smallest value to the largest value (left to right), the three measures of central tendency for a negatively skewed distribution are the mean, the median, and the mode.

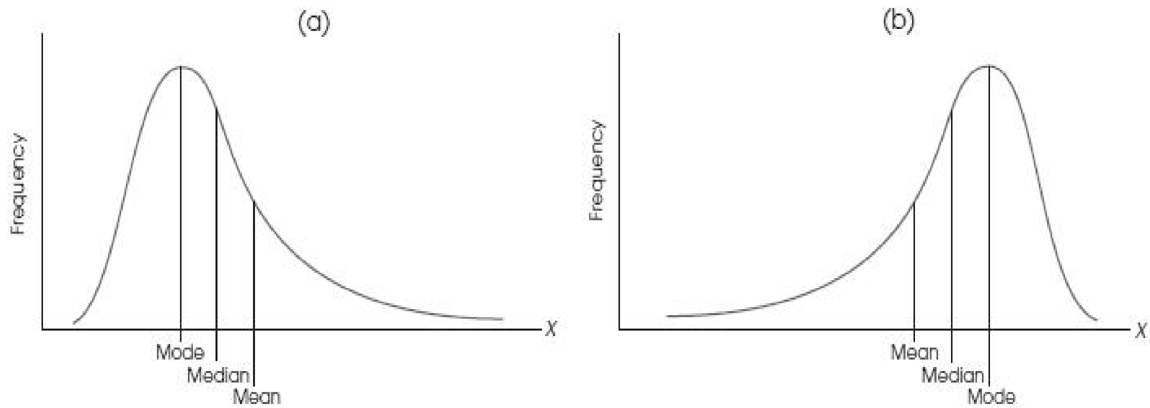


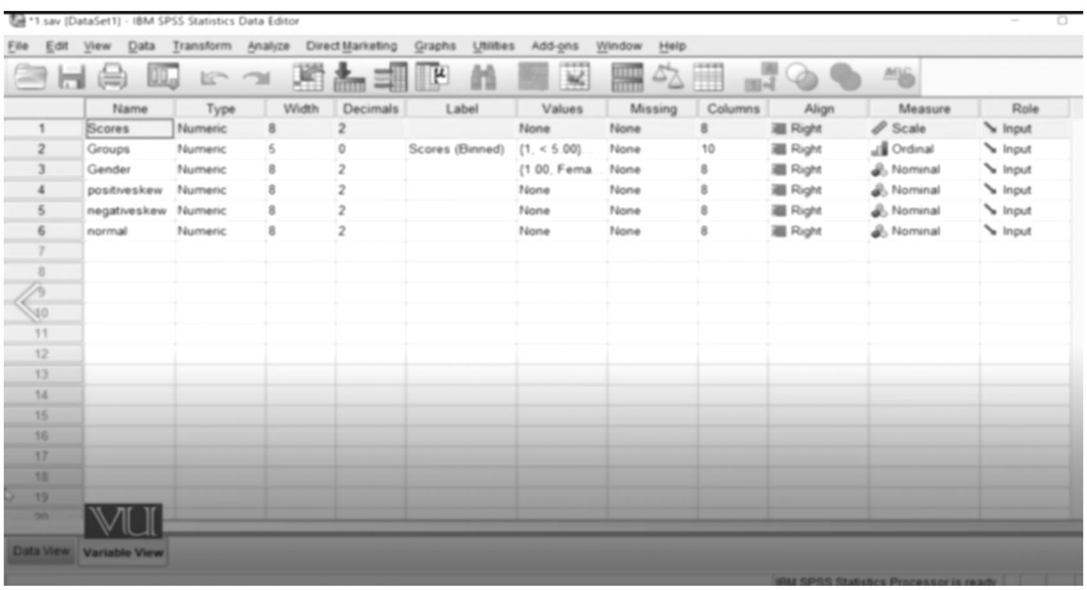
Figure 10.2 Measures of central tendency for skewed distribution

Exercise: Measure Of Central Tendency in SPSS

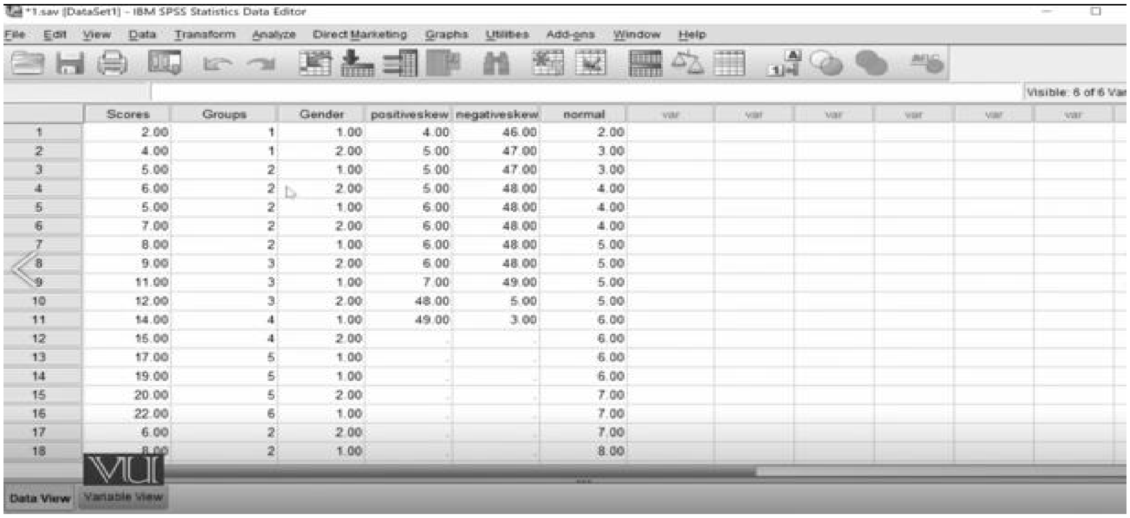
Analyze → Descriptive statistics→ Descriptive→ calculate mean, Standard Deviation, Minimum, Maximum

To calculate all together, go to Analyze → Frequencies→ Drag score in variables(s) → click statistics tab→ tick mean, median & mode→continue

Then select chart→ histogram→ tick show normal curve in histogram→continue→ok



Lets go to Graphs→legacy dialogue→line chart→simple→Define→send variable to category axis→OK



	Scores	Groups	Gender	positiveskew	negativeskew	normal
1	2.00	1	1.00	4.00	46.00	2.00
2	4.00	1	2.00	5.00	47.00	3.00
3	5.00	2	1.00	5.00	47.00	3.00
4	6.00	2	2.00	5.00	48.00	4.00
5	5.00	2	1.00	6.00	48.00	4.00
6	7.00	2	2.00	6.00	48.00	4.00
7	8.00	2	1.00	6.00	48.00	5.00
8	9.00	3	2.00	6.00	48.00	5.00
9	11.00	3	1.00	7.00	49.00	5.00
10	12.00	3	2.00	48.00	5.00	5.00
11	14.00	4	1.00	49.00	3.00	6.00
12	15.00	4	2.00	.	.	6.00
13	17.00	5	1.00	.	.	6.00
14	19.00	5	1.00	.	.	6.00
15	20.00	5	2.00	.	.	7.00
16	22.00	6	1.00	.	.	7.00
17	6.00	2	2.00	.	.	7.00
18	8.00	2	1.00	.	.	8.00

Lesson 11

MEASURE OF VARIABILITY-I

The term variability has much the same meaning in statistics as it has in everyday language; to say that things are variable means that they are not all the same. In statistics, our goal is to measure the amount of variability for a particular set of scores, a distribution. In simple terms, if the scores in a distribution are all the same, then there is no variability. If there are small differences between scores, then the variability is small, and if there are large differences between scores, then the variability is large.

Measure of Dispersion/Variability

Variability provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.

1. Variability describes the distribution. Specifically, it tells whether the scores are clustered close together or are spread out over a large distance. Usually, variability is defined in terms of *distance.* It tells how much distance to expect between one score and another, or how much distance to expect between an individual score and the mean. For example, we know that the heights for most adult males are clustered close together, within 5 or 6 inches of the average. Although more extreme heights exist, they are relatively rare.

2. Variability measures how well an individual score (or group of scores) represents the entire distribution. This aspect of variability is very important for inferential statistics, in which relatively small samples are used to answer questions about populations. For example, suppose that you selected a sample of one person to represent the entire population. Because most adult males have heights that are within a few inches of the population average (the distances are small), there is a very good chance that you would select someone whose height is within 6 inches of the population mean. On the other hand, the scores are much more spread out (greater distances) in the distribution of weights. In this case, you probably would *not* obtain someone whose weight was within 6 pounds of the population mean. Thus, variability provides information about how much error to expect if you are using a sample to represent a population.

Types of Measures of Dispersion

- Range
- Inter Quartile Range
- Variance

- Standard Deviation

The Range

The *range* is the distance covered by the scores in a distribution, from the smallest score to the largest score. When the scores are measurements of a continuous variable, the range can be defined as the difference between the upper real limit (URL) for the largest score (Xmax) and the lower real limit (LRL) for the smallest score (Xmin).

$$\text{Range} = \text{URL for } X\text{max} - \text{LRL for } X\text{min}$$

If the scores have values from 1 to 5, for example, the range is 5.5 – 0.5 = 5 points. When the scores are whole numbers, the range is also a measure of the number of measurement categories. If every individual is classified as either 1, 2, 3, 4, or 5, then there are five measurement categories and the range is 5 points. Defining the range as the number of measurement categories also works for discrete variables that are measured with numerical scores. For example, if you are measuring the number of children in a family and the data produce values from 0 to 4, then there are five measurement categories (0, 1, 2, 3, and 4) and the range is 5 points.

A commonly used alternative definition of the range simply measures the difference between the largest score (Xmax) and the smallest score (Xmin), without any reference to real limits.

range = Xmax – Xmax

Advantages and Disadvantages of Range

Advantages: Quick and give rough estimate of dispersion in the data

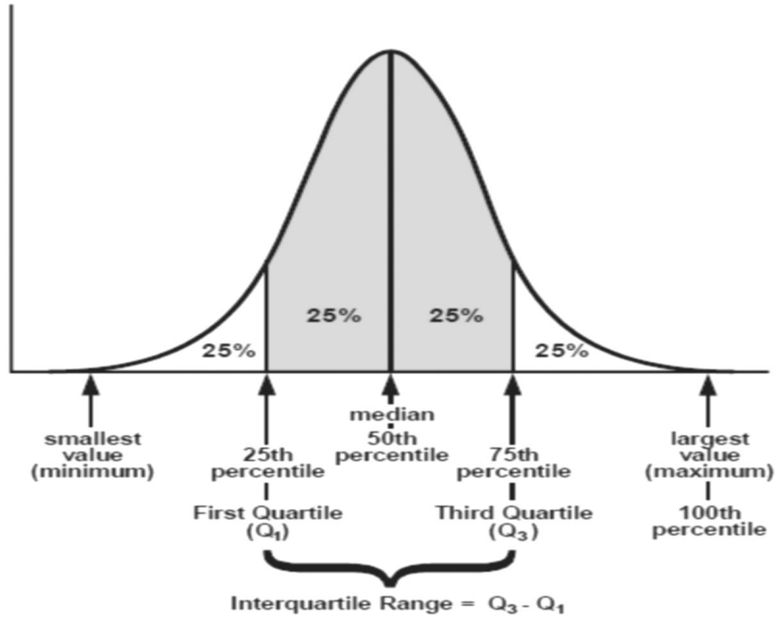
Disadvantages: Considers only extreme values and ignore all the middle values, so it's rough and imprecise and unreliable measure

Inter Quartile Range

The inter-quartile range (IQR) is a measure of variability, based on dividing a data set into quartiles. Quartiles divide a rank-ordered data set into four equal parts. The values that divide each part are called the first, second, and third quartile; and they are denoted by Q1, Q2, and Q3, respectively.

- Q1 is the "middle" value in the first half of the rank-ordered data set.
- Q2 is the median value in the set.
- Q3 is the "middle" value in the second half of the rank-ordered data set.

IQR = Q3-Q1



Pros of Inter Quartile Range

- Easy to calculate
- Eliminates influence of extreme scores

Cons of Inter Quartile Range

- Discard much of the data

Standard Deviation and Variance For a Population

The standard deviation is the most commonly used and the most important measure of variability. Standard deviation uses the mean of the distribution as a reference point and measures variability by considering the distance between each score and the mean. In simple terms, the standard deviation provides a measure of the standard, or average, distance from the mean, and describes whether the scores are clustered closely around the mean or are widely scattered. The fundamental definition of the standard deviation is the same for both samples and populations, but the calculations differ slightly. Although the concept of standard deviation is straightforward, the actual equations appear complex. Therefore, we begin by looking at the logic that leads to these equations. If you remember that our goal is to measure the standard, or typical, distance from the mean, then this logic and the equations that follow should be easier to remember.

The first step in finding the standard distance from the mean is to determine the *deviation,* or distance from the mean, for each individual score. By definition, the deviation for each score is the difference between the score and the mean.

Deviation is distance from the mean:

deviation score = $X - \mu$

For a distribution of scores with $\mu = 50$, if your score is $X = 53$, then your *deviation score* is

$X - \mu = 53 - 50 = 3$

If your score is $X = 45$, then your deviation score is

$X - \mu = 45 - 50 = -5$

Notice that there are two parts to a deviation score: the sign (_ or –) and the number. The sign tells the direction from the mean—that is, whether the score is located above (+) or below (–) the mean. The number gives the actual distance from the mean. For example, a deviation score of –6 corresponds to a score that is below the mean by a distance of 6 points.

Because our goal is to compute a measure of the standard distance from the mean, the obvious next step is to calculate the mean of the deviation scores. To compute this mean, you first add up the deviation scores and then divide by *N*. This process is demonstrated in the following example.

We start with the following set of $N = 4$ scores. These scores add up to $\Sigma X = 12$, so the mean is $\mu = \frac{12}{4} = 3$. For each score, we have computed the deviation.

X	$X - \mu$
8	+5
1	–2
3	0
0	–3
	$0 = \Sigma(X - \mu)$

Note that the deviation scores add up to zero. This should not be surprising if you remember that the mean serves as a balance point for the distribution. The total of the distances above the mean is exactly equal to the total of the distances below the mean. Thus, the total for the positive deviations is exactly equal to the total for the negative deviations, and the complete set of deviations always adds up to zero. Because the sum of the deviations is always zero, the mean of the deviations is also zero and is of no value as a measure of variability. The mean of the deviations is zero if the scores are closely clustered and it is zero if the scores are widely scattered. (You should note, however, that the constant value of zero can be useful in other ways. Whenever you are working with deviation scores, you can check your calculations by making sure that the deviation scores add up to zero.)

The average of the deviation scores does not work as a measure of variability because it is always zero. Clearly, this problem results from the positive and negative values canceling each other out. The solution is to get rid of the signs (+ and –). The standard procedure for accomplishing this is to square each deviation score. Using the squared values, you then compute the *mean squared deviation,* which is called *variance.*

Population variance equals the mean squared deviation. Variance is the average squared distance from the mean.

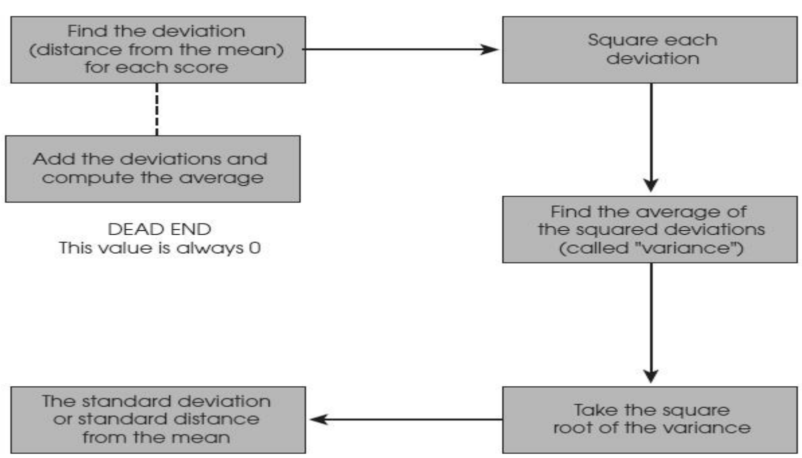
Note that the process of squaring deviation scores does more than simply get rid of plus and minus signs. It results in a measure of variability based on *squared* distances. Although variance is valuable for some of the *inferential* statistical methods covered later, the concept of squared distance is not an intuitive or easy to understand *descriptive* measure. For example, it is not particularly useful to know that the squared distance from New York City to Boston is 26,244 miles squared. The squared value becomes meaningful, however, if you take the square root. Therefore, we continue the process with one more step.

Remember that our goal is to compute a measure of the standard distance from the mean. Variance, which measures the average squared distance from the mean, is not exactly what we want. The final step simply takes the square root of the variance to obtain the *standard deviation,* which measures the standard distance from the mean.

Standard deviation is the square root of the variance and provides a measure of the standard, or average, distance from the mean.

$$\textbf{Standard deviation} = \sqrt{Variance}$$

Figure given below shows the overall process of computing variance and standard deviation.



Formulas For Population Variance And Standard Deviation

The concepts of standard deviation and variance are the same for both samples and populations. However, the details of the calculations differ slightly, depending on whether you have data from a sample or from a complete population.

The sum of squared deviations (*SS*) Recall that variance is defined as the mean of the squared deviations. This mean is computed in exactly the same way you compute any mean: First find the sum, and then divide by the number of scores.

variance = mean squared deviation = sum of squared deviations\number of scores

The value in the numerator of this equation, the sum of the squared deviations, is a basic component of variability, and we focus on it. To simplify things, it is identified by the notation *SS* (for sum of squared deviations), and it generally is referred to as the *sum of squares.*

SS, or sum of squares, is the sum of the squared deviation scores.

You need to know two formulas to compute *SS*. These formulas are algebraically equivalent (they always produce the same answer), but they look different and are used in different situations.

The first of these formulas is called the definitional formula because the symbols in the formula literally define the process of adding up the squared deviations:

Definitional formula: $SS = \Sigma(X - \mu)^2$

Although the definitional formula is the most direct method for computing *SS*, it can be awkward to use. In particular, when the mean is not a whole number, the deviations all contain decimals or fractions, and the calculations become difficult. In addition, calculations with decimal values introduce the opportunity for rounding error, which can make the result less accurate. For these reasons, an alternative formula has been developed for computing *SS*. The alternative, known as the computational formula, performs calculations with the scores (not the deviations) and therefore minimizes the complications of decimals and fractions.

$$\text{computational formula: } SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

With the definition and calculation of *SS* behind you, the equations for variance and standard deviation become relatively simple. Remember that variance is defined as the mean squared deviation. The mean is the sum of the squared deviations divided by *N*, so the equation for the *population variance* is

variance = SS\N

Standard deviation is the square root of variance, so the equation for the *population standard deviation* is

standard deviation =√SS \ N

To emphasize the relationship between standard deviation and variance, we use σ^2 as the symbol for population variance (standard deviation is the square root of the variance). Thus,

$$\text{population standard deviation} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{SS}{N}}$$

$$\text{population variance} = \sigma^2 = \frac{SS}{N}$$

Lesson 12

MEASURE OF VARIABILITY-II

Measure of Variability and Inferential Statistics

The goal of inferential statistics is to use the limited information from samples to draw general conclusions about populations. The basic assumption of this process is that samples should be representative of the populations from which they come. This assumption poses a special problem for variability because samples consistently tend to be less variable than their populations. Notice that a few extreme scores in the population tend to make the population variability relatively large. However, these extreme values are unlikely to be obtained when you are selecting a sample, which means that the sample variability is relatively small. The fact that a sample tends to be less variable than its population means that sample variability gives a *biased* estimate of population variability. This bias is in the direction of underestimating the population value rather than being right on the mark. Fortunately, the bias in sample variability is consistent and predictable, which means it can be corrected.

The calculations of variance and standard deviation for a sample follow the same steps that were used to find population variance and standard deviation. Except for minor changes in notation, the first three steps in this process are exactly the same for a sample as they were for a population. That is, calculating the sum of the squared deviations, *SS*, is the same for a sample as it is for a population. The changes in notation involve using *M* for the sample mean instead of μ, and using *n* (instead of *N*) for the number of scores. Thus, to find the *SS* for a sample:

1. Find the deviation from the mean for each score: deviation= $X\text{-}M$
2. Square each deviation: squared deviation= $(X\text{-}M)^2$
3. Add the squared deviations: $SS = \Sigma\,(X\text{-}M)^2$

These three steps can be summarized in a defintional formula for *SS:*

Definitional formula: $SS = \Sigma\,(X\text{-}M)^2$

The value of *SS* also can be obtained using a computational formula. Except for one minor difference in notation (using *n* in place of *N*), the computational formula for *SS* is the same for a sample as it was for a population (see Equation 4.2). Using sample notation, this formula is:

$$\text{Computational formula: } SS = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

Again, calculating *SS* for a sample is exactly the same as for a population, except for minor changes in notation. After you compute *SS*, however, it becomes critical to differentiate between samples and populations. To correct for the bias in sample variability, it is necessary to make an adjustment in the formulas for sample variance and standard deviation. With this in mind, *sample variance* (identified by the symbol *s*2) is defined as

$$\text{sample variance} = s^2 = \frac{SS}{n-1}$$

Sample standard deviation (identified by the symbol *s*) is simply the square root of the variance.

$$\text{sample standard deviation} = s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}}$$

Notice that the sample formulas divide by $n - 1$, unlike the population formulas, which divide by *N*. This is the adjustment that is necessary to correct for the bias in sample variability. The effect of the adjustment is to increase the value that you obtain. Dividing by a smaller number ($n - 1$ instead of *n*) produces a larger result and makes sample variance an accurate and unbiased estimator of population variance.

For a sample of *n* scores, the degrees of freedom, or *df,* for the sample variance are defined as $df = n - 1$. The degrees of freedom determine the number of scores in the sample that are independent and free to vary.

The $n - 1$ degrees of freedom for a sample is the same $n - 1$ that is used in the formulas for sample variance and standard deviation.

Earlier we noted that sample variability tends to underestimate the variability in the corresponding population. To correct for this problem, we adjusted the formula for sample variance by dividing by $n - 1$ instead of dividing by *n.* The result of the adjustment is that sample variance provides a much more accurate representation of the population variance. Specifically, dividing by $n - 1$ produces a sample variance that provides an *unbiased* estimate of the corresponding population variance. This does not mean that each individual sample variance is exactly equal to its population variance. In fact, some sample variances overestimate the population value and some underestimate it. However, the average of all the sample variances produces an accurate estimate of the population variance. This is the idea behind the concept of an unbiased statistic.

A sample statistic is unbiased if the average value of the statistic is equal to the population parameter. (The average value of the statistic is obtained from all the possible samples for a specific sample size, *n*.)

A sample statistic is biased if the average value of the statistic either underestimates or overestimates the corresponding population parameter.

Standard Deviation and Descriptive Statistics

Standard deviation is primarily a descriptive measure; it describes how variable, or how spread out, the scores are in a distribution. Standard deviation describes variability by measuring distance from the mean. In any distribution, some individuals are close to the mean, and others are relatively far from the mean. Standard deviation provides a measure of the typical, or standard, distance from the mean.

Transformations of Scale

Occasionally a set of scores is transformed by adding a constant to each score or by multiplying each score by a constant value. This happens, for example, when exposure to a treatment adds a fixed amount to each participant's score or when you want to change the unit of measurement (to convert from minutes to seconds, multiply each score by 60).

What happens to the standard deviation when the scores are transformed in this manner?

The easiest way to determine the effect of a transformation is to remember that the standard deviation is a measure of distance. If you select any two scores and see what happens to the distance between them, you also find out what happens to the standard deviation.

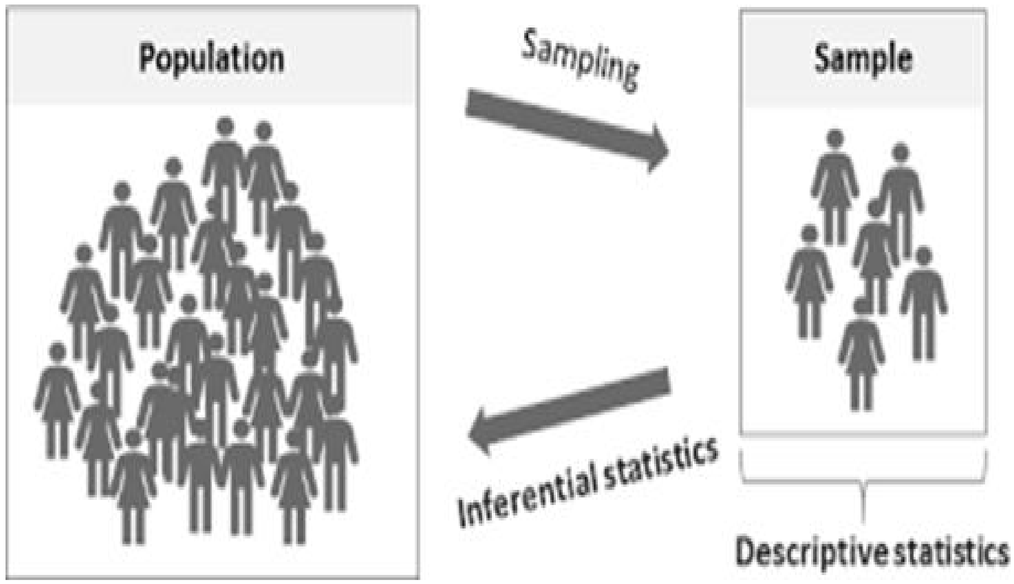
1. Adding a constant to each score does not change the standard deviation.
2. Multiplying each score by a constant causes the standard deviation to be multiplied by the same constant

Reporting the Standard Deviation

In reporting the results of a study, the researcher often provides descriptive information for both central tendency and variability. The dependent variables in psychology research are often numerical values obtained from measurements on interval or ratio scales. With numerical scores, the most common descriptive statistics are the mean (central tendency) and the standard deviation (variability), which are usually reported together. In many journals, especially those following APA style, the symbol SD is used for the sample standard deviation.

Inferential Statistics

Inferential statistics consist of techniques that allow us to study samples and then make generalizations about the populations from which they were selected.



There are *two main areas* of inferential statistics:

1. **Estimating Parameters**. This means taking a statistic from your sample data (for example the sample mean) and using it to say something about a population parameter (i.e. the population mean).

2. **Hypothesis Testing**. This is where you can use sample data to answer research questions. For example, you might be interested in knowing if a new cancer drug is effective. Or if breakfast helps children perform better in schools.

Most research uses statistical models called the *Generalized Linear model* and include Student's t-tests, ANOVA (Analysis of Variance), regression analysis. One problem with using samples, however, is that a sample provides only limited information about the population. Although samples are generally representative of their populations, a sample is not expected to give a perfectly accurate picture of the whole population. There usually is some discrepancy between a sample statistic and the corresponding population parameter. This discrepancy is called **sampling error**, and it creates the fundamental problem that inferential statistics must always address.

Lesson 13

Z-scores

Introduction to Z score

A score *by itself* does not necessarily provide much information about its position within a distribution. These original, unchanged scores that are the direct result of measurement are called *raw scores*. To make raw scores more meaningful, they are often transformed into new values that contain more information. This transformation is one purpose for *z*-scores. In particular, we transform *X* values into *z*-scores so that the resulting *z*-scores tell exactly where the original scores are located.

A second purpose for *z*-scores is to *standardize* an entire distribution. A common example of a standardized distribution is the distribution of IQ scores. Although there are several different tests for measuring IQ, the tests usually are standardized so that they have a mean of 100 and a standard deviation of 15. Because all the different tests are standardized, it is possible to understand and compare IQ scores even though they come from different tests. For example, we all understand that an IQ score of 95 is a little below average, *no matter which IQ test was used*. Similarly, an IQ of 145 is extremely high, *no matter which IQ test was used*. In general terms, the process of standardizing takes different distributions and makes them equivalent. The advantage of this process is that it is possible to compare distributions even though they may have been quite different before standardization. In summary, the process of transforming *X* values into *z*-scores serves two useful purposes:

1. Each *z*-score tells the exact location of the original *X* value within the distribution.

2. The *z*-scores form a standardized distribution that can be directly compared to other distributions that also have been transformed into *z*-scores.

One of the *primary purposes* of a *z*-score is to describe the exact location of a score within a distribution. The *z*-score accomplishes this goal by transforming each *X* value into a signed number (+ or –) so that

1. The *sign* tells whether the score is located above (+) or below (–) the mean, and

2. The *number* tells the distance between the score and the mean in terms of the number of standard deviations.

Thus, in a distribution of IQ scores with μ = 100 and σ =15, a score of X = 130 would be transformed into z =+2.00. The z value indicates that the score is located above the mean (+) by a distance of 2 standard deviations (30 points).

A z-score specifies the precise location of each X value within a distribution. The sign of the z-score (+ or –) signifies whether the score is above the mean (positive) or below the mean (negative). The numerical value of the z-score specifies the distance from the mean by counting the number of standard deviations between X and μ.

z-Scores And Location In A Distribution

The z-score definition is adequate for transforming back and forth from X values to z-scores as long as the arithmetic is easy to do in your head. For more complicated values, it is best to have an equation to help structure the calculations. Fortunately, the relationship between X values and z-scores is easily expressed in a formula. The formula for transforming scores into z-scores is

$$z = \frac{X - \mu}{\sigma}$$

The numerator of the equation, $X - \mu$, is a *deviation score*; it measures the distance in points between X and μ and indicates whether X is located above or below the mean. The deviation score is then divided by _ because we want the z-score to measure distance in terms of standard deviation units. The formula performs exactly the same arithmetic that is used with the z-score definition, and it provides a structured equation to organize the calculations when the numbers are more difficult.

Computing and Interpreting Z Scores

The following examples demonstrate the use of the z-score formula.

A distribution of scores has a mean of $\mu = 100$ and a standard deviation of $\sigma = 10$. What z-score corresponds to a score of $X = 130$ in this distribution?

According to the definition, the z-score has a value of +3 because the score is located above the mean by exactly 3 standard deviations. Using the z-score formula, we obtain

$$z = \frac{X - \mu}{\sigma} = \frac{130 - 100}{10} = \frac{30}{10} = 3.00$$

The formula produces exactly the same result that is obtained using the z-score definition.

Using Z-Scores To Standardize A Distribution

It is possible to transform every X value in a distribution into a corresponding z-score. The result of this process is that the entire distribution of X values is transformed into a distribution of z-

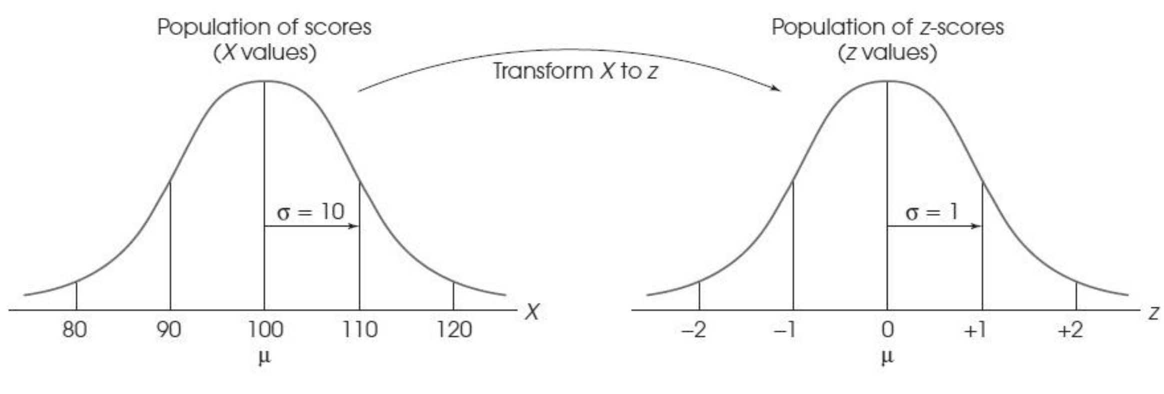
scores. The new distribution of *z*-scores has characteristics that make the *z-score transformation* a very useful tool. Specifically, if every *X* value is transformed into a *z*-score, then the distribution of *z*-scores will have the following properties:

1. Shape. The distribution of *z*-scores will have exactly the same shape as the original distribution of scores. If the original distribution is negatively skewed, for example, then the *z*-score distribution will also be negatively skewed. If the original distribution is normal, the distribution of *z*-scores will also be normal. Transforming raw scores into *z*-scores does not change anyone's position in the distribution. For example, any raw score that is above the mean by 1 standard deviation will be transformed to a *z*-score of +1.00, which is still above the mean by 1 standard deviation. Transforming a distribution from *X* values to *z* values does not move scores from one position to another; the procedure simply relabels each score. Because each individual score stays in its same position within the distribution, the overall shape of the distribution does not change.

2. The Mean. The *z*-score distribution will *always* have a mean of zero. In Figure given below, the original distribution of *X* values has a mean of 100. When this value, *X* = 100, is transformed into a *z*-score, the result is

$$z = \frac{X - \mu}{\sigma} = \frac{100 - 100}{10} = 0$$

Thus, the original population mean is transformed into a value of zero in the *z*-score distribution. The fact that the *z*-score distribution has a mean of zero makes the mean a convenient reference point.



An entire population of scores is transformed into *z*-scores. The transformation does not change the shape of the population, but the mean is transformed into a value of 0 and the standard deviation is transformed to a value of 1.

3. The Standard Deviation. The distribution of *z*-scores will *always* have a standard deviation of 1. In Figure 5.5, the original distribution of *X* values has mean 100 and standard deviation 10. In this distribution, a value of $X = 110$ is above the mean by exactly 10 points or 1 standard deviation. When $X = 110$ is transformed, it becomes $z = +1.00$, which is above the mean by exactly 1 point in the *z*-score distribution. Thus, the standard deviation corresponds to a 10-point distance in the *X* distribution and is transformed into a 1-point distance in the *z*-score distribution. The advantage of having a standard deviation of 1 is that the numerical value of a *z*-score is exactly the same as the number of standard deviations from the mean. For example, a *z*-score of $z + 1.50$ is exactly 1.50 standard deviations from the mean.

A standardized distribution is composed of scores that have been transformed to create predetermined values for μ and σ. Standardized distributions are used to make dissimilar distributions comparable.

Computing Z-Scores For a Sample

Although *z*-scores are most commonly used in the context of a population, the same principles can be used to identify individual locations within a sample. The definition of a *z*-score is the same for a sample as for a population, provided that you use the sample mean and the sample standard deviation to specify each *z*-score location. Thus, for a sample, each *X* value is transformed into a *z*-score so that

1. The sign of the *z*-score indicates whether the *X* value is above (+) or below (–) the sample mean, and

2. The numerical value of the *z*-score identifies the distance from the sample mean by measuring the number of sample standard deviations between the score (*X*) and the sample mean (*M*).

Expressed as a formula, each *X* value in a sample can be transformed into a *z*-score as follows:

$$z = \frac{X - M}{s}$$

Similarly, each Z-score can be transformed back into an X value, as follows:

$X = M + zs$

If all the scores in a sample are transformed into *z*-scores, the result is a sample of *z*-scores. The transformed distribution of *z*-scores will have the same properties that exist when a population of *X* values is transformed into *z*-scores. Specifically,

- The sample of z-scores will have the same shape as the original sample of scores.
- The sample of z-scores will have a mean of $M_z = 0$.
- The sample of z-scores will have a standard deviation of $s_z = 1$.

Lesson 14

INTRODUCTION TO PROBABILITY-I

Introduction to Probability

Suppose, for example, that you are selecting a single marble from a jar that contains 50 black and 50 white marbles. (In this example, the jar of marbles is the *population* and the single marble to be selected is the *sample*.) Although you cannot guarantee the exact outcome of your sample, it is possible to talk about the potential outcomes in terms of probabilities. In this case, you have a 50-50 chance of getting either color.

Now consider another jar (population) that has 90 black and only 10 white marbles.

Again, you cannot predict the exact outcome of a sample, but now you know that the sample probably will be a black marble. By knowing the makeup of a population, we can determine the probability of obtaining specific samples. In this way, probability gives us a connection between populations and samples, and this connection is the foundation for the inferential statistics.

Probability is a huge topic that extends far beyond the limits of introductory statistics, and we do not attempt to examine it all here. Instead, we concentrate on the few concepts and definitions that are needed for an introduction to inferential statistics. We begin with a relatively simple definition of probability.

For a situation in which several different outcomes are possible, the probability for any specific outcome is defined as a fraction or a proportion of all the possible outcomes. If the possible outcomes are identified as A, B, C, D, and so on, then:

$$\text{probability of } A = \frac{\text{number of outcomes classified as } A}{\text{total number of possible outcomes}}$$

Probability Values The probability of a specific outcome is expressed with a *p* (for probability) followed by the specific outcome in parentheses. Probability fractions can be expressed as either decimals or percentages

E.g., the probability of obtaining heads for a coin toss is written as:

$$p(\text{heads})=1/2=0.50=50\%$$

All probability values are contained in a range from 0 to 1

Probability in Inferential Statistics

- Relationship between sample and population is defined in terms of probability
- When we draw a sample, what is the probability that sample taken is right? or probability of making a wrong decision?
- Probability is used to predict what kind of samples are likely to be obtained from a population. (e.g.2 jars of marbles)

Random Sampling

For the preceding definition of probability to be accurate, it is necessary that the outcomes be obtained by a process called *random sampling*.

A **random sample** requires that each individual in the population has an *equal chance* of being selected.

A second requirement, necessary for many statistical formulas, states that if more than one individual is being selected, the probabilities must *stay constant* from one selection to the next. Adding this second requirement produces what is called *independent random sampling*. The term *independent* refers to the fact that the probability of selecting any particular individual is independent of those individuals who have already been selected for the sample. For example, the probability that you will be selected is constant and does not change even when other individuals are selected before you are.

An **independent random sample** requires that each individual has an equal chance of being selected and that the probability of being selected stays constant from one selection to the next if more than one individual is selected.

Probability and Frequency Distributions

The situations in which we are concerned with probability usually involve a population
of scores that can be displayed in a frequency distribution graph. If you think of the graph as representing the entire population, then different proportions of the graph represent different proportions of the population. Because probabilities and proportions are equivalent, a particular proportion of the graph corresponds to a particular probability in the population. Thus, whenever a population is presented in a frequency distribution graph, it is possible to represent probabilities as proportions of the graph.

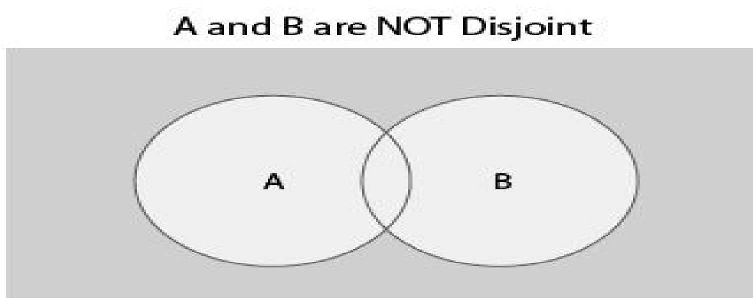
Rules and Basics Laws of Probability

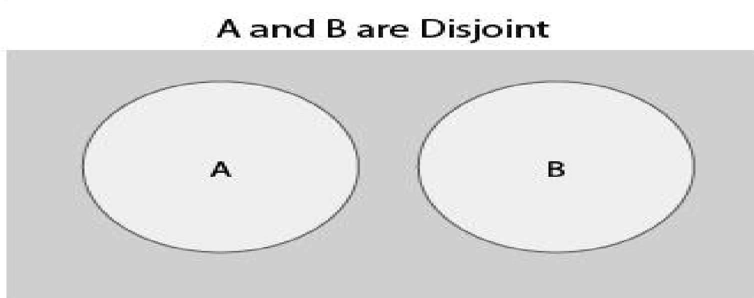
Basic Properties of Probability

- Every probability is between zero and one. If A is an event, then **0≤P(A)≥ 1.**
- The sum of the probabilities of all possible outcomes is 1.
- Impossible events have probability zero. That is, if event A is impossible, then **P(A)=0.**

An example of such an event is rolling a 7 on a standard six-sided die.

- Two events that cannot occur at the same time are called *disjoint or mutually exclusive.*





Consider the following two events:

1. A-a randomly chosen person has blood type A, and
2. B-a randomly chosen person has blood type B.

We are going to assume that each person can have only one blood type. Therefore, it is impossible for the events A and B to occur together. **Events A and B are DISJOINT**

Consider the following two events:

1. A-a randomly chosen person has blood type A
2. B-a randomly chosen person is a woman.

In this case, it is possible for events A and B to occur together. **Events A and B are NOT DISJOINT.**

Basics Laws Of Probability

Additive Law of Probability-Given a set of mutually exclusive events, the probability of the occurrence of one event or another is equal to the sum of their separate probabilities.

Multiplicative Law of Probability-The probability of the joint occurrence of two or more independent events is the product of their individual probabilities.

Joint and Conditional Probabilities

Two types of probabilities play an important role in discussions of probability: joint probabilities and conditional probabilities.

A ***joint probability*** is defined simply as the probability of the co-occurrence of two or more events. If those two events are independent, then the probability of their joint occurrence can be found by using the multiplicative law. If they are not independent, the probability of their joint occurrence is more complicated to compute.

A ***conditional probability*** is the probability that one event will occur, given that some other event has occurred. The probability that a person will contract AIDS, given that he or she is an intravenous drug user, is a conditional probability.

Lesson 15

INTRODUCTION TO PROBABILITY-II

Understanding Probability

According to Stanford University's Blood Center these are the probabilities of human blood types in the United State

Blood Type	O	A	B	AB
Probability	0.44	0.42	0.10	0.04

What is the probability that a randomly chosen person is a potential donor for a person with blood type A?

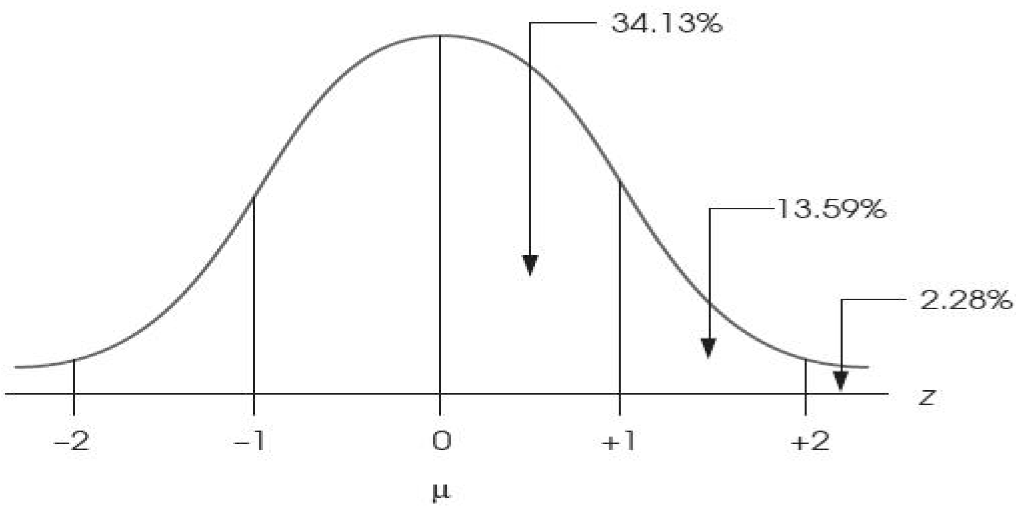
- Suppose we roll two dice. What is the probability that both dice show a 5?
- Sample Space
- You toss a coin 3 times, what is the probability of at least two heads

Probability and Normal Distribution

Note that the normal distribution is symmetrical, with the highest frequency in the middle and frequencies tapering off as you move toward either extreme. Although the exact shape for the normal distribution is defined by an equation (see Figure 6.3), the normal shape can also be described by the proportions of area contained in each section of the distribution. Statisticians often identify sections of a normal distribution by using *z*-scores. Figure 6.4 shows a normal distribution with several sections marked in *z*-score units. You should recall that *z*-scores measure positions in a distribution in terms of standard deviations from the mean. (Thus, $z = +1$ is 1 standard deviation above the mean, $z = +2$ is 2 standard deviations above the mean, and so on.) The graph shows the percentage of scores that fall in each of these sections. For example, the section between the mean ($z = 0$) and the point that is 1 standard deviation above the mean ($z = 1$) contains 34.13% of the scores. Similarly, 13.59% of the scores are located in the section between 1 and 2 standard deviations above the mean. In this way it is possible to define a normal distribution in terms of its proportions; that is, a distribution is normal if and only if it has all the right proportions.

There are two additional points to be made about the distribution shown in Figure 6.4. First, you should realize that the sections on the left side of the distribution have exactly the same areas as the corresponding sections on the right side because the normal distribution is symmetrical.

Second, because the locations in the distribution are identified by z-scores, the percentages shown in the figure apply to *any normal distribution* regardless of the values for the mean and the standard deviation. Remember: When any distribution is transformed into z-scores, the mean becomes zero and the standard deviation becomes one.

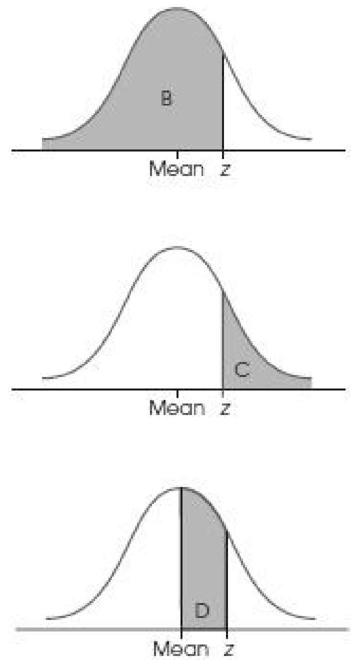


The Unit Normal Table

A more complete listing of z-scores and proportions is provided in the *unit normal table*. This table lists proportions of the normal distribution for a full range of possible z-score values. The complete unit normal table is provided in Appendix B Table B.1, and part of the table is reproduced in Figure given below. Notice that the table is structured in a four-column format. The first column (A) lists z-score values corresponding to different positions in a normal distribution. If you imagine a vertical line drawn through a normal distribution, then the exact location of the line can be described by one of the z-score values listed in column A. You should also realize that a vertical line separates the distribution into two sections: a larger section called the *body* and a smaller section called the *tail*.

Columns B and C in the table identify the proportion of the distribution in each of the two sections. Column B presents the proportion in the body (the larger portion), and column C presents the proportion in the tail. Finally, we have added a fourth column, column D, that identifies the proportion of the distribution that is located *between* the mean and the z-score.

(A) z	(B) Proportion in body	(C) Proportion in tail	(D) Proportion between mean and z
0.00	.5000	.5000	.0000
0.01	.5040	.4960	.0040
0.02	.5080	.4920	.0080
0.03	.5120	.4880	.0120
0.21	.5832	.4168	.0832
0.22	.5871	.4129	.0871
0.23	.5910	.4090	.0910
0.24	.5948	.4052	.0948
0.25	.5987	.4013	.0987
0.26	.6026	.3974	.1026
0.27	.6064	.3936	.1064
0.28	.6103	.3897	.1103
0.29	.6141	.3859	.1141
0.30	.6179	.3821	.1179
0.31	.6217	.3783	.1217
0.32	.6255	.3745	.1255
0.33	.6293	.3707	.1293
0.34	.6331	.3669	.1331



A portion of the unit normal table. This table lists proportions of the normal distribution corresponding to each z-score value. Column A of the table lists z-scores. Column B lists the proportion in the body of the normal distribution up to the z-score value. Column C lists the proportion of the normal distribution that is located in the tail of the distribution beyond the z-score value. Column D lists the proportion between the mean and the z-score value.

Probabilities and Proportions for Scores from a Normal Distribution

In most situations, however, it is necessary to find probabilities for specific *X* values. Consider the following example:

It is known that IQ scores form a normal distribution with $\mu=100$ and $\sigma=15$. Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?

This problem is asking for a specific probability or proportion of a normal distribution.

However, before we can look up the answer in the unit normal table, we must first transform the IQ scores (*X* values) into *z*-scores. Thus, to solve this new kind of probability problem, we must add one new step to the process. Specifically, to answer probability questions about scores (*X* values) from a normal distribution, you must use the following two-step procedure:

1. Transform the *X* values into *z*-scores.

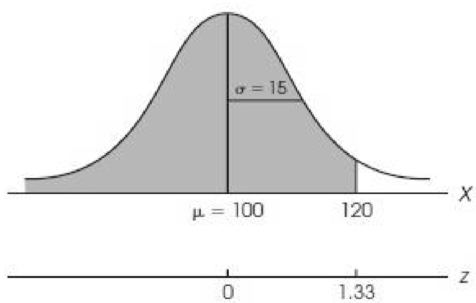
2. Use the unit normal table to look up the proportions corresponding to the *z*-score values.

This process is demonstrated in the following examples. Once again, we suggest that you sketch the distribution and shade the portion you are trying to find to avoid careless mistakes.

We now answer the probability question about IQ scores that we presented earlier. Specifically, what is the probability of randomly selecting and individual with an IQ score less than 120? Restated in terms of proportions, we want to find the proportion of IQ distribution that corresponds to scores less than 120. The distribution is drawn in given Figure, and the portion we want has been shaded. The first step is to change the X values into z-scores. In particular, the score of X=120 is changed to

$$z = \frac{X-\mu}{\sigma} = \frac{120-100}{15} = \frac{20}{15} = 1.33$$

The distribution of IQ scores. The problem is to find the probability or proportion of the distribution corresponding to scores less than 120.



Thus, an IQ score of $X = 120$ corresponds to a z-score of $z = 1.33$, and IQ scores less than 120 correspond to z-scores less than 1.33.

Next, look up the z-score value in the unit normal table. Because we want the proportion of the distribution in the body to the left of $X = 120$ (see Figure 6.10), the answer is in column B. Consulting the table, we see that a z-score of 1.33 corresponds to a proportion of 0.9082. The probability of randomly selecting an individual with an IQ less than 120 is $p = 0.9082$. In symbols,

$p(X <\_ 120) =\_ p(z \_< 1.33) =\_ 0.9082$ (or 90.82%)

Finding Proportions/Probabilities Located Between Two Scores The next example demonstrates the process of finding the probability of selecting a score that is located *between* two specific values. Although these problems can be solved using the proportions of columns B and C (body and tail), they are often easier to solve with the proportions listed in column D.

The highway department conducted a study measuring driving speeds on a local section of interstate highway. They found an average speed of μ = 58 miles per hour with a standard deviation of 10. The distribution was approximately normal.

Given this information, what proportion of the cars are traveling between 55 and 65 miles per hour? Using probability notation, we can express the problems as

$p(55 < X < 65) = ?$

The distribution of driving speeds is shown in figure below with the appropriate area shared. The first step is to determine the z-score corresponding to the *X* value at each end of the interval.

$$\text{For } X = 55\text{: } z = \frac{X-\mu}{\sigma} = \frac{55-58}{10} = \frac{-3}{10} = -0.30$$

$$\text{For } X = 65\text{: } z = \frac{X-\mu}{\sigma} = \frac{65-58}{10} = \frac{7}{10} = 0.70$$

Looking again at figure, we see the the proportion we are seeking can be divided into two sections: (1) the area left of the mean, and (2) the area right of the mean. The first area is the proportion between the mean and z= -0.30, and the second is the proportion between the mean and z= +0.70. using column D of the unit normal table, these two proportions are 0.1179 and 0.2580. The total proportion is obtained by adding these two sections:

$p(55 < X < 65) = p(-0.30 < z < +0.70) = 0.1179 + 0.2580= 0.3759$

Review of Probability

- Introduction to Probability
- Rules and Basic laws of Probability
- Probability and Normal Distribution

Lesson 16

INTRODUCTION TO PROBABILITY-III

A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range. It is a statistical model that shows the possible outcomes of a particular event or course of action as well as the statistical likelihood of each event.

Probability and Binomial Distribution

When a variable is measured on a scale consisting of exactly two categories, the resulting data are called binomial. The term *binomial* can be loosely translated as "*two names*," referring to the two categories on the measurement scale.

Binomial data can occur when a variable naturally exists with only two categories. It is further explained by an example, people can be classified as male or female, and a coin toss results in either heads or tails. It also is common for a researcher to simplify data by collapsing the scores into two categories. For example, a psychologist may use personality scores to classify people as either high or low in aggression.

In binomial situations, the researcher often knows the probabilities associated with each of the two categories. With a balanced coin, for example, $p(\text{heads}) = p(\text{tails}) = 1/2$. The question of interest is the number of times each category occurs in a series of trials or in a sample of individuals. For example:

- What is the probability of obtaining 15 heads in 20 tosses of a balanced coin?
- What is the probability of obtaining more than 40 introverts in a sampling of 50 college freshmen?

To answer probability questions about binomial data, we must examine the binomial distribution. To define and describe this distribution, we first introduce some notation.

1. The two categories are identified as *A* and *B*.

2. The probabilities (or proportions) associated with each category are identified as:

p* =*p*(*A*) = the probability of *A

q* =*p*(*B*) = the probability of *B

Notice that $p = q = 1.00$ because *A* and *B* are the only two possible outcomes.

3. The number of individuals or observations in the sample is identified by *n*.

4. The variable *X* refers to the number of times category *A* occurs in the sample.

Notice that *X* can have any value from 0 (none of the sample is in category *A*) to *n* (all of the sample is in category *A*).

Using the notation presented here, the binomial distribution shows the probability associated with each value of *X* from $X = 0$ to $X = n$.

The binomial distribution tends toward a normal shape, especially when the sample size (n) is relatively large. In binomial distribution, probability of success and failure in each and every trial is equal to one, p+q=1

The fact that the binomial distribution tends to be normal in shape means that we can compute probability values directly from *z*-scores and the unit normal table.

The Normal Approximation To The Binomial Distribution

Binomial distribution tends to approximate a normal distribution, particularly when n is large. To be more specific,The binomial distribution is nearly perfect normal distribution when pn and qn are both equal to or greater than 10. Under these circumstances, the binomial distribution approximates a normal distribution with the following parameters:

Mean: $\mu = pn$

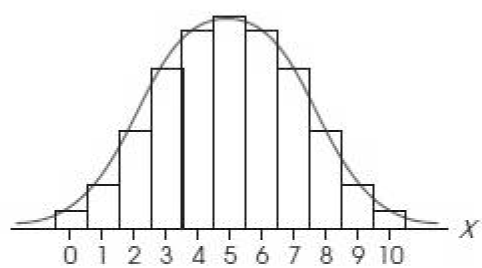
standard deviation: $\sigma = \sqrt{npq}$

Within this normal distribution, each value of X has a corresponding z-score,

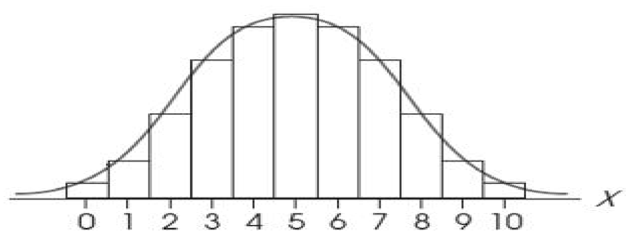
$$z = \frac{X - \mu}{\sigma} = \frac{X - pn}{\sqrt{npq}}$$

It is important to remember that the normal distribution is only an approximation of a true binomial distribution. Binomial values, such as the number of heads in a series of coin tosses, are *discrete*. The normal distribution is *continuous*. However, the *normal approximation* provides an extremely accurate model for computing binomial probabilities in many situations. Figure given below shows the difference between the true binomial distribution, the discrete histogram, and the normal curve that approximates the binomial distribution. Although the two distributions are slightly different, the area under the distributions is nearly equivalent. *Remember, it is the area under the distribution that is used to find probabilities.*

The relationship between the binomial distribution and the normal distribution. The binomial dis bution is always a discrete histogram, and the normal distribution is a continuous, smooth curve. Each *X* value is represented by a bar in the histogram or a section of the normal distribution.



In binomial distribution if the value of p is smaller or lesser than 0.5 then the binomial distribution is skewed to right. Binomial values, such as the number of heads in a series of coin tosses, are discrete. The normal distribution is continuous. However, the normal approximation provides an extremely accurate model for computing binomial probabilities. Each X value in the binomial distribution actually corresponds to a bar in the histogram. For example, if the score X=6 is represented by a bar it is bounded by real limits of 5.5 and 6.5. The actual probability of X = 6 is determined by the area contained in this bar.



Example:

To make it more understandable, it is explained through an example. In the game Rock-Paper-Scissors, the probability that both players will select the same response and tie is p=1/3 and the probability that they will pick different responses is p=2/3. If two people play 72 rounds of the game and choose their responses randomly, what is the probability that they will choose the same response (tie) more than 28 times?

Using Binomial Distribution To Test Hypothesis

In a **directional hypothesis test,** or a **one-tailed test,** the statistical hypotheses (*H*0 and *H*1) specify either an increase or a decrease in the population mean. That is, they make a statement about the direction of the effect.

Because a specific direction is expected for the treatment effect, it is possible for the researcher to perform a directional test. The first step (and the most critical step) is to state the statistical hypotheses. Remember that the null hypothesis states that there is no treatment effect and the alternative hypothesis says that there is an effect. For this example, the predicted effect is that the blueberry supplement will increase test scores. Thus, the two hypotheses would state:

H0: Test scores are not increased. (The treatment does not work.)

H1: Test scores are increased. (The treatment works as predicted.)

To express directional hypothesis in symbols, t usually is easier to begin with the alternative hypothesis (H1). Again, we know that the general population has an average test score of μ=80, and H1 states that the test scores will be increased by the blueberry supplement. Therefore, expressed in symbols, H1 states,

H1: μ>80 (With the supplement, the average score is greater than 80.)

The null hypothesis states that the supplement does not increase scores. In symbols,

Ho:$\mu \leq 80$ (With the supplement, the average score is not greater than 80.)

Note again that the two hypothesis are mutually exclusive and cover all of the possibilities.

The major distinction between one-tailed and two-tailed tests is the criteria that they use for rejecting *H*0. A one-tailed test allows you to reject the null hypothesis when the difference between the sample and the population is relatively small, provided that the difference is in the specified direction. A two-tailed test, on the other hand, requires a relatively large difference independent of direction. This point is illustrated in the following example.

To test hypothesis with the binomial distribution, we must calculate probability p, of the observed event. We compare this to the level of significance α.

If $p > \alpha$ then we do not reject the null hypothesis.

If $p < \alpha$ then we accept the alternative hypothesis.

A coin is tossed 20 times (n=20) landing on head more than 6 times (X=6.5). Perform a hypothesis test at a 5% (0.05) significance level to see if the coin is biased.

Ho: the coin is not biased

H1 : the coin is biased in favor of tails

Lesson 17

SAMPLING DISTRIBUTION

A sampling distribution is a statistic that is arrived out through repeated sampling from a larger population. It describes a range of possible outcomes that of a statistic, such as the mean or mode of some variable, as it truly exists a population.

The distribution of sample means is the collection of sample means for all of the possible random samples of a particular size (n) that can be obtained from a population. Because statistics are obtained from samples, a distribution of statistics is referred to as a sampling distribution. Sample is a subset of population.

A sampling distribution is a distribution of statistics obtained by selecting all of the possible samples of a specific size from a population. Thus, the distribution of sample means is an example of a sampling distribution. In fact, it often is called the sampling distribution of M.

The Distribution Of Sample Means

It is the collection of sample means for all of the possible random samples of a particular size (n) that can be obtained from a population. The distribution of sample means contains all of the possible samples. It is necessary to have all of the possible values to compute probabilities.

For example, if the entire set contains exactly 100 samples, then the probability of obtaining any specific sample is 1 out of 100: $p=1/100$

General Characteristics Of a Distribution

Sampling distribution is a statistic that determines the probability of an event based on data from a small group within a large population. Its primary purpose is to establish *representative results* of small samples of a comparatively larger population.

1. Sample means should be relatively close to the population mean.
2. The pile of sample means should tend to form a normal-shaped distribution.
3. The larger the sample size, the closer the sample means should be to the population mean, μ.

Central Limit Theorem

A mathematical proposition known as the central limit theorem provides a precise description of the distribution that would be obtained if you selected every possible sample, calculated every sample mean, and constructed the distribution of the sample mean. This important and useful theorem serves as a cornerstone for much of inferential statistics. Following is the essence of the theorem. For any population with mean μ and standard deviation , the distribution of sample means for sample size n will have a mean of μ and a standard deviation of n and will approach a normal distribution as n approaches infinity.

Central limit theorem describes the distribution of sample means by identifying the three basic characteristics that describe any distribution: shape, central tendency, and variability.The central limit theorem states that if the sample size increases sampling distribution must approach normal distribution. Generally a sample size more than 30 us considered as large enough.

The Shape Of The Distribution Of Sample Means

The distribution of sample means tends to be a normal distribution. In fact, this distribution is almost perfectly normal if;

1. The population from which the samples are selected is a normal distribution.
2. The number of scores (n) in each sample is relatively large, around 30 or more.

The mean of the distribution of sample means is equal to the mean of the population of scores, μ, and is called the expected value of M.

Standard Error Of Sampling Distribution

Sampling error is the natural discrepancy, or amount of error, between a sample statistic and its corresponding population parameter.

The standard deviation (measure of variability) for the distribution of sample means is identified by the symbol M and is called the standard error of M.

The standard error serves the two purposes for the distribution of sample means.

The Standard Error Of M

1. The standard error describes the distribution of sample means. It provides a measure of how much difference is expected from one sample to another.

2. Standard error measures how well an individual sample mean represents the entire distribution.

The symbol for the standard error is σM. The σ indicates that this value is a standard deviation, and the subscript M indicates that it is the standard deviation for the distribution of sample means.

The magnitude of the standard error is determined by two factors:

(1) The size of the sample and

(2) The standard deviation of the population from which the sample is selected.

As the sample size increases, the error between the sample mean and the population mean should decrease. This rule is also known as the law of large numbers.

The law of large numbers states that the larger the sample size (n), the more probable it is that the sample mean is close to the population mean.

The Population Standard Deviation

When n=1 the standard deviation for the distribution of sample means, which is the standard error, is identical to the standard deviation for the distribution of scores.

When n=1, $\sigma M = \sigma$ (standard error = standard deviation).

Standard error = $\sigma M = \sigma/\sqrt{n}$

As sample size (n) increases, the size of the standard error decreases. (Larger samples are more accurate.) When the sample consists of a single score (n=1), the standard error is the same as the standard deviation (M).

Lesson 18

CONFIDENCE INTERVAL-I

A *confidence interval* is defined as the range of values that we observe in our sample and for which we expect to find the value that accurately reflects the population. A confidence interval is an interval, or range of values, centered around a sample statistic. The logic behind a confidence interval is that a sample statistic, such as a sample mean, should be relatively near to the corresponding population parameter. Therefore, we can confidently estimate that the value of the parameter should be located in the interval.

Confidence interval is an interval of values computed from sample data that is almost sure to cover the true population parameter. We are estimating population parameter from sample statistics. Point estimate will be at the center of a confidence interval. The most common level of confidence used is 95%.

For example, a confidence interval for the population mean could be calculated with data obtained from a sample and would provide an estimated range of values within which the actual population mean is believed to lie. A confidence interval often is reported in addition to the point estimate of a population parameter. We can have 90% and 99% confidence intervals. It is impossible to construct an interval in which we could be 100% confident unless we actually measure the entire population.

For an approximately normal data set, the values within one standard deviation of the mean account for about 68% of the set; while within two standard deviations account for about 95%; and within three standard deviations account for about 99.7%.

Confidence Interval For Population Proportion

Estimating population proportion from a sample proportion. The most commonly reported information that can be used to construct a confidence interval is the **margin of error.**

- To construct a 95% confidence interval for a population proportion, simply add and subtract the margin of error to the sample proportion.
- The margin of error is often reported using the symbol "$\pm$".

- The formula for a 95% confidence interval can thus be expressed as Sample proportion $\pm$ margin of error.

Constructing a Confidence Interval For a Proportion

If numerous samples are taken, the frequency curve made from proportions from the various samples will be approximately bell-shaped. The mean will be the true proportion from the population. The Standard deviation will be;

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

To be exact, we would actually add and subtract 1.96(SD) instead of 2(SD) because 95% of the values for a bell-shaped curve fall within 1.96 standard deviations of the mean. However, in most practical applications, rounding 1.96 off to 2.0 will not make much difference and this is common practice.

Applying the reasoning we used to construct the formula for a 95% confidence interval and using the information about bell-shaped curves we can construct for instance, a 68% confidence interval. Relationship between a 95 confidence interval and a 99 confidence interval from the same sample is that 99% interval will be wider. For example by simply adding and subtracting 1 standard deviation to the sample proportion instead of 2.

Lesson 19

CONFIDENCE INTERVAL- II

In previous lesson we have discussed confidence interval for population proportion as Estimating population proportion from a sample proportion. In this lesson we will discuss about confidence interval for population mean.

Confidence Interval For Population Means

We can try to estimate population mean when all we have available is a sample of measurements from the population. All we need from the sample are its mean, standard deviation, and number of observations. Sample mean do not affect the width of the confidence interval.

Normally-distributed data forms a bell shape when plotted on a graph, with the sample mean in the middle and the rest of the data distributed fairly evenly on either side of the mean. The confidence interval for data which follows a standard normal distribution is:

$$CI = \overline{X} \pm Z^* \frac{\sigma}{\sqrt{n}}$$

Where:

CI = the confidence interval

$\overline{X}$ = the population mean

Z^* = the critical value of the z-distribution

σ = the population standard deviation

√n = the square root of the population size

The confidence interval for the t-distribution follows the same formula, but replaces the Z^* with the t^*.

In real life, you never know the true values for the population (unless you can do a complete census). Instead, we replace the population values with the values from our sample data, so the formula becomes:

$$CI = \hat{x} \pm Z^* \frac{s}{\sqrt{n}}$$

Where:

ˆx = the sample mean

s = the sample standard deviation

The Rule for Sample Means

- If numerous samples of the same size are taken, the frequency curve of means from the various samples will be approximately bell-shaped. The mean of this collection of sample means will be the same as the mean of the population. Population standard deviation/square root of sample size = $\sigma/\sqrt{n}$
- The standard deviation for the possible sample means is called the standard error of the mean.
- In other words, SEM = standard error = population standard deviation$\sqrt{n}$
- To construct a 95% confidence interval for a mean we can use the same reasoning we used for proportions. In 95% of all samples, the sample mean will fall within 2 standard errors of the true population mean.The values associated with a two-sided 95% confidence interval of the standard normal distribution are ± 1.96.

The formula for a 95% confidence interval for a population mean becomes:

Sample mean ± 2 standard errors

Where Standard error = $\sigma\sqrt{n}$

This formula should be used only if there are at least 30 observations in the sample.

To compute a 95% confidence interval for the population mean based on smaller samples, a multiplier larger than 2 is used, which is found from a "t-distribution."

Examples:

Ali got a sample of 30 graduate students and found that mean age for that sample was = 33 years with SD of 4.3 years. What is the 95% confidence interval for average age of entire university graduate students.

Wood and colleagues (1988), studied a group of 89 sedentary men for a year. 42 men were placed on a diet; the remaining 47 were put on an exercise routine. The group on a diet lost an average of 7.2 kg, with a standard deviation of 3.7 kg. The men who exercised lost an average of 4.0 kg, with a standard deviation of 3.9 kg.

Notice that these intervals are trying to capture the true mean or average value for the population. They do not encompass the full range of weight loss that would be experienced by most individuals. Also, remember that these intervals could be wrong. Ninety-five percent of intervals constructed this way will contain the correct population mean value, but 5% will not.

Confidence Interval For Between Two Means

Instead of separately comparing the two groups from the population the efficient way is to construct a single confidence interval for the difference in the population means for the two groups or conditions.

1. Collect a large sample of observations (at least 30), independently, under each condition or from each group. Compute the mean and the standard deviation for each sample.

2. Compute the standard error of the mean (SEM) for each sample by dividing the sample standard deviation by the square root of the sample size.

3. Square the two SEMs and add them together. Then take the square root. This will give you the necessary "measure of variability," which is called the standard error of the difference in two means. In other words:

measure of variability=square root of [(SEM1)2 + (SEM1)2]

4. A 95% confidence interval for the difference in the two population means is

difference in sample means ± 2×measure of variability

or

difference in sample means ± 2×square root of [(SEM1)2 + (SEM1)2]

This method is valid only when independent measurements are taken from the two groups. For instance, if matched pairs are used and one treatment is randomly assigned to each half of the pair, the measurements would not be independent.

To construct a formula for constructing a confidence interval for the difference between two population means includes:

- A point estimate of the difference between the population means.
- The standard error of the sampling distribution of the sample means.
- The confidence level.

Reporting Confidence Interval

Confidence intervals are sometimes reported in papers, though researchers more often report the standard deviation of their estimate. If you are asked to report the confidence interval, you should include the upper and lower bounds of the confidence interval.

The first confidence interval compares the mean educational levels for the smokers and nonsmokers. The result tells us that, the average educational level for nonsmokers was 0.67 year higher than for smokers.

The interval tells us that the difference in the population is probably between 0.15 and 1.19 years of education. In other words, mothers who did not smoke were also likely to have had more education.

	Sample Means		
	0 Cigarettes	10+ Cigarettes	Difference (95% CI)
Maternal education, grades	11.57	10.89	0.67 (0.15,1.19)
Stanford-Binet (IQ), 48 mo	113.28	103.12	10.16 (5.04,15.30)
Birthweight, g	3416	3035	381.0 (167.1,594.9)

Lesson 20

HYPOTHESIS TESTING-I

Hypothesis testing is one of the most commonly used inferential procedures. In fact, most of the remainder of this book examines hypothesis testing in a variety of different situations and applications. Although the details of a hypothesis test change from one situation to another, the general process remains constant. In this chapter, we introduce the general procedure for a hypothesis test. You should notice that we use the statistical techniques that is, we combine the concepts of *z*-scores, probability, and the distribution of sample means to create a new statistical procedure known as a *hypothesis test*.

A **hypothesis test** is a statistical method that uses sample data to evaluate a hypothesis about a population. In very simple terms, the logic underlying the hypothesis-testing procedure is as follows:

- First, we state a hypothesis about a population. Usually the hypothesis concerns the value of a population parameter. For example, we might hypothesize that American adults gain an average of 7 pounds between Thanksgiving and New Year's Day each year.
- Before we select a sample, we use the hypothesis to predict the characteristics that the sample should have. For example, if we predict that the average weight gain for the population is 7 pounds, then we would predict that our sample should have a mean *around* 7 pounds. Remember: The sample should be similar to the population, but you always expect a certain amount of error.
- Next, we obtain a random sample from the population. For example, we might select a sample of n = 200 American adults and measure the average weight change for the sample between Thanksgiving and New Year's Day.
- Finally, we compare the obtained sample data with the prediction that was made from the hypothesis. If the sample mean is consistent with the prediction, then we conclude that the hypothesis is reasonable. But if there is a big discrepancy between the data and the prediction, then we decide that the hypothesis is wrong.

Logic of Hypothesis Testing

Hypothesis is an assumption made about a population (which is to be tested).

Before we select a sample, we use the hypothesis to predict the characteristics that the sample should have.Next, we obtain a random sample from the population. Finally, we compare the obtained sample data with the prediction that was made from the hypothesis. The researcher begins with a known population (a sample). This is the set of individuals as they exist.

The basic idea is to avoid having to reason about the real world by setting up a hypothetical world that is completely understood. The observed patterns of the data are then compared to what would be generated in the hypothetical world. If they don't match, then there is reason to doubt that the data support the hypothesis

Basic Steps For Hypothesis Testing

Step 1: State the hypothesis

As the name implies, the process of hypothesis testing begins by stating a hypothesis about the unknown population. Actually, we state two opposing hypotheses. Notice that both hypotheses are stated in terms of population parameters. The first, and most important, of the two hypotheses is called the *null hypothesis*. The null hypothesis states that the treatment has no effect. In general, the null hypothesis states that there is no change, no effect, no difference nothing happened, hence the name *null*. The null hypothesis is identified by the symbol $H0$. (The H stands for *hypothesis*, and the zero subscript indicates that this is the *zero-effect* hypothesis.)

For the study in Example 8.1, the null hypothesis states that the blueberry supplement has no effect on cognitive functioning for the population of adults who are more than 65 years old. In symbols, this hypothesis is:

Ho : μ(with supplement) = 80 (Even with the supplement, the mean test score is still 80.)

The **null hypothesis** ($H0$) states that in the general population there is no change, no difference, or no relationship. In the context of an experiment, $H0$ predicts that the independent variable (treatment) *has no effect* on the dependent variable (scores) for the population.

The second hypothesis is simply the opposite of the null hypothesis, and it is called the *scientific,* or *alternative, hypothesis* ($H1$). This hypothesis states that the treatment has an effect on the dependent variable. The **alternative hypothesis** (*H*1) states that there is a change, a difference, or a relationship for the general population. In the context of an experiment, *H*1 predicts that the independent variable (treatment) *does have an effect* on the dependent variable.

For this example, the alternative hypothesis states that the supplement does have an effect on cognitive functioning for the population and will cause a change in the mean score. In symbols, the alternative hypothesis is represented as:

H1: μ(with supplement) ≠ 80 (Even with the supplement, the mean test score is different from 80.)

Notice that the alternative hypothesis simply states that there will be some type of change. It does not specify whether the effect will be increased or decreased test scores. In some circumstances, it is appropriate for the alternative hypothesis to specify the direction of the effect. For example, the researcher might hypothesize that the supplement will increase neuropsychological test scores ($\mu > 80$). This type of hypothesis results in a directional hypothesis test. For now we concentrate on non-directional tests, for which the hypotheses simply state that the treatment has no effect (*H*0) or has some effect (*H*1).

Step 2: Set the criteria for a decision

Eventually the researcher uses the data from the sample to evaluate the credibility of the null hypothesis. The data either provide support for the null hypothesis or tend to refute the null hypothesis. In particular, if there is a big discrepancy between the data and the null hypothesis, then we conclude that the null hypothesis is wrong.

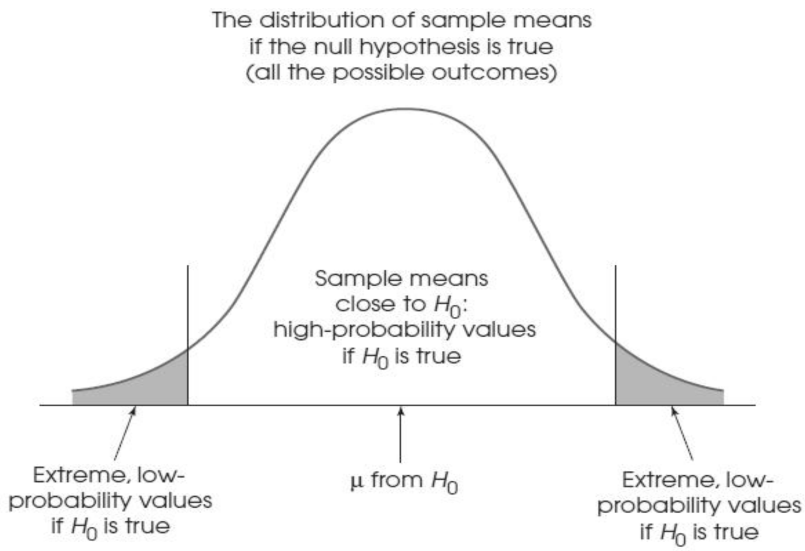
To formalize the decision process, we use the null hypothesis to predict the kind of sample mean that ought to be obtained. Specifically, we determine exactly which sample means are consistent with the null hypothesis and which sample means are at odds with the null hypothesis.

The Alpha Level

To find the boundaries that separate the high-probability samples from the low-probability samples, we must define exactly what is meant by “low” probability and “high” probability.

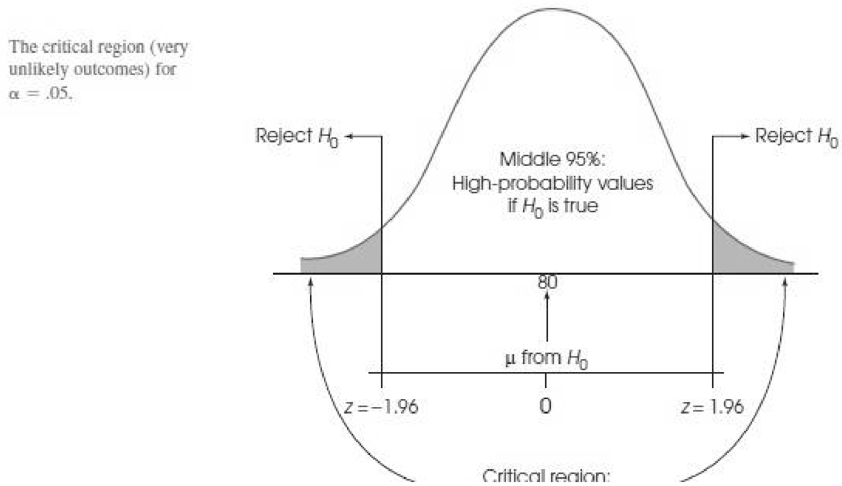
*The **alpha level,** or the **level of significance,** is a probability value that is used to define the concept of “very unlikely” in a hypothesis test.*

As in the graph given below, this is accomplished by selecting a specific probability value, which is known as the *level of significance,* or the *alpha level,* for the hypothesis test. The alpha (α) value is a small probability that is used to identify the low probability samples. By convention, commonly used alpha levels are $\alpha = .05$ (5%), $\alpha = .01$ (1%), and $\alpha = .001$ (0.1%). For example, with $\alpha = .05$, we separate the most unlikely 5% of the sample means (the extreme values) from the most likely 95% of the sample means (the central values).



The extremely unlikely values, as defined by the alpha level, make up what is called the *critical region*. These extreme values in the tails of the distribution define outcomes that are not consistent with the null hypothesis; that is, they are very unlikely to occur if the null hypothesis is true. Whenever the data from a research study produce a sample mean that is located in the critical region, we conclude that the data are not consistent with the null hypothesis, and we reject the null hypothesis.

The below graph shows the critical region. Critical region is composed of the extreme sample values that are very unlikely (as defined by the alpha level) to be obtained if the null hypothesis is true. The boundaries for the critical region are determined by the alpha level. If sample data fall in the critical region, the null hypothesis is rejected.



Step 3: Collect data and compute sample statistics

At this time, we select a sample of adults who are more than 65 years old and give each one a daily dose of the blueberry supplement. After 6 months, the neuro-psychological test is used to measure cognitive function for the sample of participants. Notice that the data are collected *after* the researcher has stated the hypotheses and established the criteria for a decision. This sequence of events helps to ensure that a researcher makes an honest, objective evaluation of the data and does not tamper with the decision criteria after the experimental outcome is known.

Next, the raw data from the sample are summarized with the appropriate statistics: For this example, the researcher would compute the sample mean. Now it is possible for the researcher to compare the sample mean (the data) with the null hypothesis. This is the heart of the hypothesis test: comparing the data with the hypothesis. The comparison is accomplished by computing a *z*-score that describes exactly where the sample mean is located relative to the hypothesized population mean from *Ho*. In step 2, we constructed the distribution of sample means that would be expected if the null hypothesis were true—that is, the entire set of sample means that could be obtained if the treatment has no effect. Now we calculate a *z*-score that identifies where our sample mean is located in this hypothesized distribution. The *z*-score formula for a sample mean is:

$$z = \frac{M - \mu}{\sigma_M}$$

In the formula, the value of the sample mean (M) is obtained from the sample data, and the value of μ is obtained from the null hypothesis. Thus, the z-score formula can be expressed in words as follows:

z = sample mean-hypothesized population mean/ standard error between M and μ.

Notice that the top of the z-score formula measures how much difference there is between the data and the hypothesis. The bottom of the formula measures the standard distance that ought to exist between a sample mean and the population mean.

Step 4: Make a decision

In the final step, the researcher uses the *z*-score value obtained in step 3 to make a decision about the null hypothesis according to the criteria established in step 2. There are two possible outcomes:

1. The sample data are located in the critical region. By definition, a sample value in the critical region is very unlikely to occur if the null hypothesis is true. Therefore, we conclude that the sample is not consistent with *H*0 and our decision is to *reject the null hypothesis*. Remember, the null hypothesis states that there is no treatment effect, so rejecting *H*0 means that we are concluding that the treatment did have an effect.

For the example we have been considering, suppose that the sample produced a mean of M = 92 after taking the supplement for 6 months. The null hypothesis states that the population mean is μ= 80 and, with n = 25 and σ=20, the standard error for the sample mean is

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = \frac{20}{5} = 4$$

Thus. a sample mean of M = 92 produces a z-score of :

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = \frac{20}{5} = 4$$

With an alpha level of a = .05, this z-score is far beyond the boundary of 1.96. Because the sample z-score is in the critical region, we reject the null hypothesis and conclude that the blueberry supplement did have an effect on cognitive functioning.

2. The second possibility is that the sample data are not in the critical region. In this case, the sample mean is reasonably close to the population mean specified in the null hypothesis (in the center of the distribution). Because the data do not provide strong evidence that the null hypothesis is wrong, our conclusion is to *fail to reject the null hypothesis*. This conclusion means that the treatment does not appear to have an effect.

For the research study examining the-blueberry supplement, suppose our sample produced a mean test score of M = 84. As before, the standard error for a sample of n = 25 is σ_M = 4, and the null hypothesis states that is μ = 80. These values produce a z-score of

$$z = \frac{M - \mu}{\sigma_M} = \frac{84 - 80}{4} = \frac{4}{4} = 1.00$$

The z-score of 1.00 is not in the critical region. Therefore, we would fail to reject the null hypothesis and conclude that the blueberry supplement does not appear to have an effect on cognitive functioning.

Types Of Hypothesis

Depending on the population distribution, you can classify the statistical hypothesis into two types:

Simple Hypothesis: A simple hypothesis specifies an exact value for the parameter. *Composite Hypothesis:* A composite hypothesis specifies a range of values.

One-Tailed And Two-Tailed Hypothesis

Hypothesis can be One-tailed and two-tailed.

In a **directional hypothesis test,** or a **one-tailed test,** the statistical hypotheses ($H0$ and $H1$) specify either an increase or a decrease in the population mean. That is, they make a statement about the direction of the effect.

Because a specific direction is expected for the treatment effect, it is possible for the researcher to perform a directional test. The first step (and the most critical step) is to state the statistical hypotheses. Remember that the null hypothesis states that there is no treatment effect and the alternative hypothesis says that there is an effect. For this example, the predicted effect is that the blueberry supplement will increase test scores. Thus, the two hypotheses would state:

Ho: Test scores are not increased. (The treatment does not work.)

H1:Test scores are increased. (The treatment works as predicted.)

To express directional hypotheses in symbols, it usually is easier to begin with the alternative hypothesis (H1). Again we know that the general population has an average test score of $\mu=80$, and H1 states that scores will be increased by the blueberry supplement. Therefore, expressed in symbols, H1 states,

H1:$\mu > 80$ (with the supplement, the average score is greater than 80.)

The null hypothesis states that the supplement does not increase scores. In symbols,

H0:$\mu \leq 80$ (With the supplement, the average score is not greater than 80.)

Note again that the two hypothesis are mutually exclusive and cover all of the possibilities.

Two-tailed hypothesis tests are also known as non directional and two-sided tests. It can test for effects in both directions. When a two-tailed test is performed the significance level/critical region is split between both tails of the distribution.

Ho: there is no training effect on performance; $M \leq \mu$

H1: there is training effect on performance $M \neq \mu$

Comparison Of One-Tailed Versus Two-Tailed Tests

The major distinction between one-tailed and two-tailed tests is the criteria that they use for rejecting *H*0. A one-tailed test allows you to reject the null hypothesis when the difference between the sample and the population is relatively small, provided that the difference is in the specified direction. A two-tailed test, on the other hand, requires a relatively large difference independent of direction.

Lesson 21

HYPOTHESIS TESTING-II

Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution. First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H0.

In very simple terms, the logic underlying the hypothesis-testing procedure is we state a hypothesis about a population. Before we select a sample, we use the hypothesis to predict the characteristics that the sample should have. Next, we obtain a random sample from the population. Finally, we compare the obtained sample data with the prediction that was made from the hypothesis. If the sample mean is consistent with the prediction, then we conclude that the hypothesis is reasonable. But if there is a big discrepancy between the data and the prediction, then we decide that the hypothesis is wrong.

Assumptions For Hypothesis Tests

Assumptions underlying the hypothesis testing is as follows.

1. **Random Sampling**

It is assumed that the participants used in the study were selected randomly.

2. **Independent Observations**

The values in the sample must consist of independent observations.

Two events (or observations) are independent if the occurrence of the first event has no effect on the probability of the second event.

3. **The value Of (σ) Is Unchanged By The Treatment**

We assume that the standard deviation for the unknown population (after treatment) is the same as it was for the population before treatment.

4. **Normal Sampling Distribution**

For hypotheses with z-scores, the unit normal table is used to identify the critical region. This table can be used only if the distribution of sample means is normal.

Errors In Hypothesis Testing

In the framework of hypothesis tests there are two types of errors: Type I error and type II error. A type I error occurs if a true null hypothesis is rejected (a "false positive"), while a type II error occurs if a false null hypothesis is not rejected (a "false negative").

- **Type I Error**

A Type I error occurs when a researcher rejects a null hypothesis that is actually true.

It means that the researcher concludes that a treatment does have an effect when, in fact, it has no effect. A Type I error is more likely to occur when a researcher unknowingly obtains an extreme, non representative sample. The alpha level determines the probability of a Type I error.

- **Type II Error**

A Type II error occurs when a researcher fails to reject a null hypothesis that is really false. It means that the hypothesis test has failed to detect a real treatment effect.

In this case the treatment does influence the sample, but the magnitude of the effect is not big enough to move the sample mean into the critical region. Because the sample is not substantially different from the original population, the statistical decision is to fail to reject the null hypothesis.

The probability of a Type II error is represented by the symbol β, the Greek letter beta and it depends on variety of factors rather than a specific number.

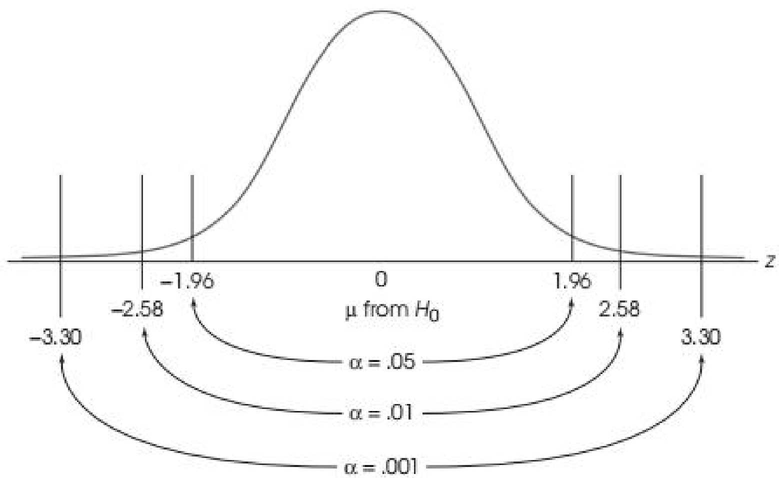
		No Effect, H_0 True	Effect Exists, H_0 False
Experimenter's Decision	Reject H_0	Type I error	Decision correct
	Retain H_0	Decision correct	Type II error

Selecting An Alpha Level

Alpha is a threshold value used to judge whether a test statistic is statistically significant. It is chosen by the researcher. Alpha represents an acceptable probability of a Type I error in a statistical test. Because alpha corresponds to a probability, it can range from 0 to 1. The alpha level for a hypothesis test serves two very important functions.

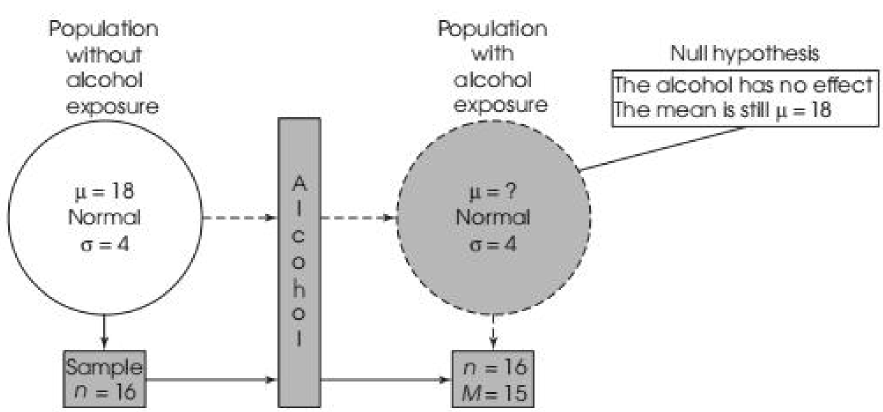
- First, the alpha level helps to determine the boundaries for the critical region
- Secondly, the alpha level determines the probability of a Type I error.

The primary concern when selecting an alpha level is to minimize the risk of a Type I error. The consequences of a Type I error can be relatively serious, therefore, many researchers and prefer to use an alpha level such as .01 or .001 to reduce the risk.



Exercise: How To Do Hypothesis Testing

For practice of students, hypothesis testing with another example is mentioned that is a researcher would like to investigate the effect of prenatal alcohol exposure on birth weight. In diagram below, a random sample of n=16 pregnant rats is obtained. The mother rats are given daily doses of alcohol. At birth, one pup is selected from each litter to produce a sample of n=16 newborn rats. The average weight for the sample is M=15 grams. The researcher would like to compare the sample with the general population of rats. It is known that regular newborn rats (not exposed to alcohol) have an average weight of μ=18 grams. The distribution of weights is normal with σ=4.



The structure of a research study to determine whether prenatal alcohol effects birth weight. A sample is selected from the original population and is exposed to alcohol. The is what would happen if the entire population were exposed to alcohol. The treated sample provides information about the unknown treated population.

1. State hypothesis

The null hypothesis states that exposure to alcohol has no effect on birth weight

Ho: µalcohol exposure =18

The alternative hypothesis states that alcohol exposure does affect birth weight

H1 : µ alcohol exposure ≠18

2. Select alpha level

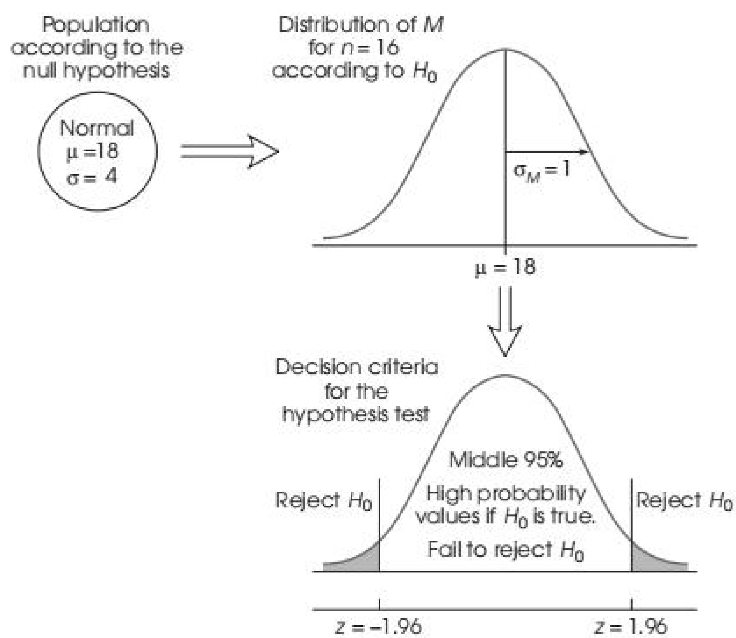
An alpha level of .05 will be used

That is, we are taking a 5% risk of committing a Type I error.

3. ***Set the decision criteria by locating the critical region***

Begin with the population of scores

Construct the distribution of sample means for the sample size



Use z-scores to separate the extreme outcomes (as defined by the alpha level) from the high-probability outcomes.

4. Collect the data, and compute the test statistic

Researcher would select one newborn pup from each of the n=16 mothers that received alcohol during pregnancy. The birth weight is recorded for each pup and the sample mean is computed; M= 15grams

The sample mean is then converted to a z-score

5. Make a decision

The z-score computed has a value of 3.00, which is beyond the boundary of -1.96.

Therefore, the sample mean is located in the critical region.

Lesson 22

HYPOTHESIS TESTING-III

The final decision in a hypothesis test is determined by the value obtained for the z-score statistic. If the z-score is large enough to be in the critical region, then we reject the null hypothesis and conclude that there is a significant treatment effect. Otherwise, we fail to reject H0 and conclude that the treatment does not have a significant effect.

Factors Influencing Hypothesis Test

The most obvious factor influencing the size of the z-score is the difference between the sample mean and the hypothesized population mean from H0. A big mean difference indicates that the treated sample is noticeably different from the untreated population and usually supports a conclusion that the treatment effect is significant. In addition to the mean difference, however, there are other factors that help determine whether the z-score is large enough to reject H0.

In this section we examine two factors that can influence the outcome of a hypothesis test.

1. The variability of the scores, which is measured by either the standard deviation or the variance. The variability influences the size of the standard error in the denominator of the z-score.

2. The number of scores in the sample. This value also influences the size of the standard error in the denominator.

Statistical Power Of The Test

The power of a test is defined as the probability that the test will reject the null hypothesis if the treatment really has an effect.

The power of a statistical test is the probability that the test will correctly reject a false null hypothesis. That is, power is the probability that the test will identify a treatment effect if one really exists.

There are only two possible outcomes for a hypothesis test: either fail to reject Ho or reject Ho.

- The first outcome, failing to reject Ho when there is a real effect is a Type II error with a probability identified as p= β.
- The second outcome have a probability of 1 – β. Hence, rejecting Ho when there is a real effect this and is the power of the test.

Factors Affecting Power Of The Test

There are many factors that can influence the power of a test. Such as power is to detect the effect when it is actually present, power is to avoid/lower the probability of committing type II error. Beta is the probability of Type II error. Beta is equal to proportion of alternate distribution that falls below the critical value. Factors that can effect the power of a test is mentioned below:

- **Sample Size**

When the sample size is reduced, power decreases to less than 50%. In general, a larger sample produces greater power for a hypothesis test.

- **Alpha Level**

Reducing the alpha level for a hypothesis test also reduces the power of the test. For example, lowering from .05 to .01 lowers the power of the hypothesis test.

- **One-Tailed Versus Two-Tailed Test**

Changing from a regular two-tailed test to a one-tailed test increases the power of the hypothesis test.

Lesson 23

HYPOTHESIS TESTING-IV

Role Of Sample Size In Statistical Significance

Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Thus, it increases the power of the test. The effect size is not affected by sample size. Higher sample size allows the researcher to increase the significance level of the findings, since the confidence of the result are likely to increase with a higher sample size. This is to be expected because larger the sample size, the more accurately it is expected to mirror the behavior of the whole group.

Statistical Significance

If 95% confidence intervals do not overlap then we can could conclude that the means come from different populations, and, therefore, that they are significantly different.

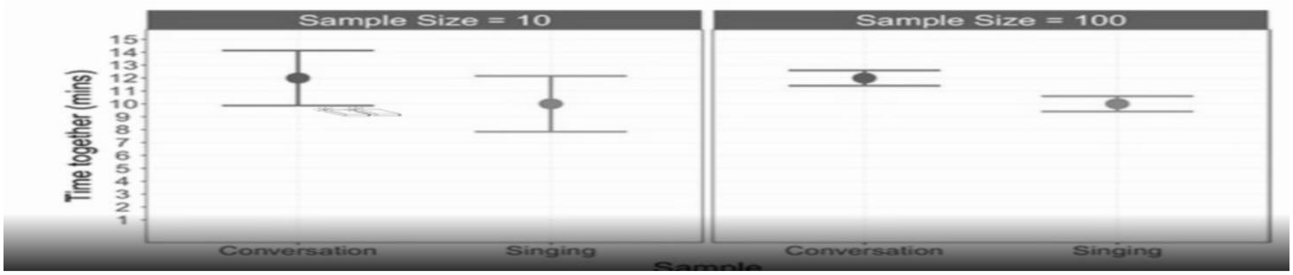
- Whether the results of a study are statistically significant can depend on the sample size.
- A larger sample size tends to yield a statistically significant results.
- On the other hand, if the sample size is too small, an important relationship or difference can go undetected.
- In that case, the power of the test is too low.

Given that power is the ability of a test to find an effect that genuinely exists, and the effect is found by having a statistically significant result (i.e., $p < 0.05$).

Hence, there is also a connection between the sample size and the p-value associated with a test statistic.We clear this concept by giving an example that is study got two groups of 10 heterosexual young men and got them to go up to a woman that they found attractive and either engage them in conversation (group 1) or sing them a song (group 2). Researchers measured how long it was before the woman ran away. Imagine the experiment was also repeated using 100 men in each group.

Results showed that in both cases the singing group had a mean of 10 and a standard deviation of 3, and the conversation group had a mean of 12 and a standard deviation of 3. The only difference between the two experiments is that one collected 10 scores per sample, and the other 100 scores per sample.

The confidence intervals become much narrower when the samples contain 100 scores than when they contain only 10 scores.



The sample size affects whether a difference between samples is deemed significant or not.

In large samples, small differences can be significant, and in small samples large differences can be non-significant. Even a difference of practically zero can be deemed 'significant' if the sample size is big enough.

Calculating Sample Size

We calculate sample size by the following way. The sample size necessary to achieve a given level of power can be calculated by setting the value of α, 1_β and effect size. Normally we have alpha value 0.05, power 0.8 and moderate effect size to determine the sample size. Using power to calculate the necessary sample size is the more common and more useful thing to do.

The actual computations are very complicated. Therefore, we usually prefer to use a software for this purpose.

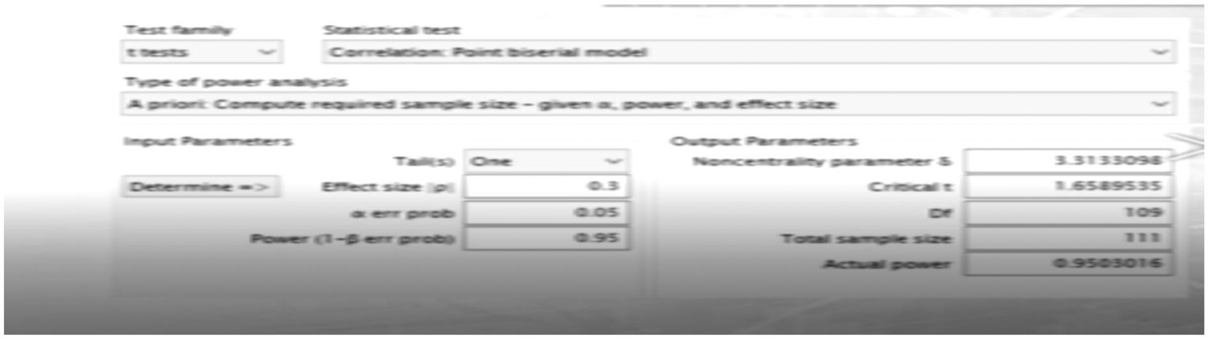
Calculating Sample Size G' Power

G* power is a free to use software or a general power analysis programme, which is used for determining the sample size and analysis power in research studies. To do power analysis to estimate the sample size, we need;

- Hypothesis
- Number of independent variables or number of groups
- What statistical test (one of the inferential statistics) we will be using.

Below picture will show the alpha level (usually .05), power of the test (.80), and effect size (small, moderate or large based on the test being used).

Statistics are showing that total sample size required for the study is 111.



Hypothesis Testing Review

Hypothesis testing is a statistical procedure that allows researchers to use sample data to draw inferences about the population of interest.

Hypothesis testing includes:

Stating a hypothesis, either null or alternative. One says that effect does not exist (the null hypothesis, Ho) and the other says that an effect exists (the alternative hypothesis, H1).

Example:

A researcher begins with a known population, with $\mu=65$ and $\sigma=15$. The researcher suspects that special training in reading skills will produce a change in the scores for individuals in the population. A sample of n=25 individuals is selected, and the treatment is given to this sample. Following treatment, the average score for this sample is M=70.

We then, (ii) Set criteria for a decision determined by: Alpha level (level of significance that separates the high probability samples from the low-probability samples) and Critical region (extreme sample values that are very unlikely to be obtained if the null hypothesis is true).

Next, (iii) data is collected data and sample statistics are computed

By comparing the data with hypothesis

Comparison is accomplished by computing a z-score

$z=(M-\mu)/\sigma_M$

Lastly , (iv) decision is made

A sample value in the critical region is very unlikely to occur if the null hypothesis is true. Therefore, decision is to reject the null hypothesis. If the sample data are not in the critical region then, the sample mean is reasonably close to the population mean (in the center of the distribution) conclusion is to fail to reject the null hypothesis.

Assumptions for hypothesis testing include;

- Random selection of participants

- The values in the sample must consist of independent observations
- The value of population standard deviation (σ) is unchanged by the treatment
- Sample should be normally distributed.

In hypothesis testing a ***Type I error*** occurs when a researcher rejects a null hypothesis that is actually true. It means that the researcher concludes that a treatment does have an effect when, in fact, it has no effect.

A ***Type II error*** occurs when a researcher fails to reject a null hypothesis that is really false.

It means that the hypothesis test has failed to detect a real treatment effect.

Hypothesis test is influenced by;

The **variability of the scores** (standard deviation). Higher variability can reduce the chances of finding a significant treatment effect.

The **number of score in the sample**. Increasing the sample size can increase the chances of finding a significant treatment effect.

The statistical power of a test is defined as the probability that the test will reject the null hypothesis if the treatment really has an effect.

In hypothesis testing, the first outcome, failing to reject Ho when there is a real effect is a Type II error with a probability identified as $p= \beta$. The second outcome have a probability of $1 - \beta$. Hence, rejecting Ho when there is a real effect this and is the power of the test.

Factors Affecting Power Of The Test:

1. Power of the test is affected by sample size. In general, a larger sample produces greater power for a hypothesis test.

2. Power of the test is affected by alpha level. Reducing the alpha level for a hypothesis test also reduces the power of the test.

3. Power of the test is affected by one-tailed versus two-tailed test. Changing from a regular two-tailed test to a one-tailed test increases the power of the hypothesis test.

Lesson 24

T-TEST-I

The t Statistic: An Alternative To z

1. A sample mean (*M*) is expected to approximate its population mean (μ). This permits us to use the sample mean to test a hypothesis about the population mean.

2. The standard error provides a measure of how well a sample mean approximates the population mean. Specifically, the standard error determines how much difference is reasonable to expect between a sample mean (*M*) and the population mean (μ).

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

To quantify our inferences about the population, we compare the obtained sample mean (*M*) with the hypothesized population mean (μ) by computing a *z*-score test statistic.

$$z = \frac{M - \mu}{\sigma_M} = \frac{\text{obtained difference between data and hypothesis}}{\text{standard distance between } M \text{ and } \mu}$$

The Problem With *z*-Scores

The shortcoming of using a *z*-score for hypothesis testing is that the *z*-score formula requires more information than is usually available. Specifically, a *z*-score requires that we know the value of the population standard deviation (or variance), which is needed to compute the standard error. In most situations, however, the standard deviation for the population is not known. In fact, the whole reason for conducting a hypothesis test is to gain knowledge about an *unknown* population. This situation appears to create a paradox, you want to use a *z*-score to find out about an unknown population, but you must know about the population before you can compute a *z*-score. Fortunately, there is a relatively simple solution to this problem. When the variability for the population is not known, we use the sample variability in its place.

Introducing The t Statistic

When the variability for the population is not known, we use the sample variability in its place.

Sample variance as an unbiased estimate of the corresponding population variance is

$s^2 = \sum(X-M)^2/n-1$

$= SS/n-1$

$=SS/df$

The sample variance was developed specifically to provide an unbiased estimate of the corresponding population variance. Recall that the formulas for sample variance and sample standard deviation are as follows:

$$\text{sample variance} = s^2 = \frac{SS}{n-1} = \frac{SS}{df}$$

$$\text{sample standard deviation} = s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}}$$

Using the sample values, we can now *estimate* the standard error. Recall that the value of the standard error can be computed using either standard deviation or variance:

$$\text{standard error} = \sigma_M = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

Now we estimate the standard error by simply substituting the sample variance or standard deviation in place of the unknown population value:

$$\text{estimated standard error} = s_M = \frac{s}{\sqrt{n}} \quad \text{or} \quad s_M = \sqrt{\frac{s^2}{n}}$$

Notice that the symbol for the *estimated standard error of* M is *sM* instead of *σM*, indicating that the estimated value is computed from sample data rather than from the actual population parameter.

The **estimated standard error** (***sM***) is used as an estimate of the real standard error, *σM,* when the value is unknown. It is computed from the sample variance or sample standard deviation and provides an estimate of the standard distance between a sample mean, *M*, and the population mean, μ.

The estimated standard error of *M* typically is presented and computed using:

$$s_M = \sqrt{\frac{s^2}{n}}$$

There are two reasons for making this shift from standard deviation to variance:

1. We saw that the sample variance is an *unbiased* statistic; on average, the sample variance (s2) provides an accurate and unbiased estimate of the population variance (σ2). Therefore, the most accurate way to estimate the standard error is to use the sample variance to estimate the population variance.

2. In future chapters we encounter other versions of the *t* statistic that require variance (instead of standard deviation) in the formulas for estimated standard error. To maximize the similarity from one version to another, we use variance in the formula for *all* of the different *t* statistics. Thus, whenever we present a *t* statistic, the estimated standard error is computed as:

$$\text{estimated standard error} = \sqrt{\frac{\text{sample variance}}{\text{sample size}}}$$

Now we can substitute the estimated standard error in the denominator of the z score formula. The result is a new test statistic called a t statistics.

$$t = \frac{M - \mu}{s_M}$$

The *t* statistic is used to test hypotheses about an unknown population mean, μ, when the value of σ is unknown. The formula for the *t* statistic has the same structure as the *z*-score formula, except that the *t* statistic uses the estimated standard error in the denominator.

The only difference between the *t* formula and the *z*-score formula is that the *z*-score uses the actual population variance, $\boldsymbol{\sigma}$ 2 (or the standard deviation), and the *t* formula uses the corresponding sample variance (or standard deviation) when the population value is not known.

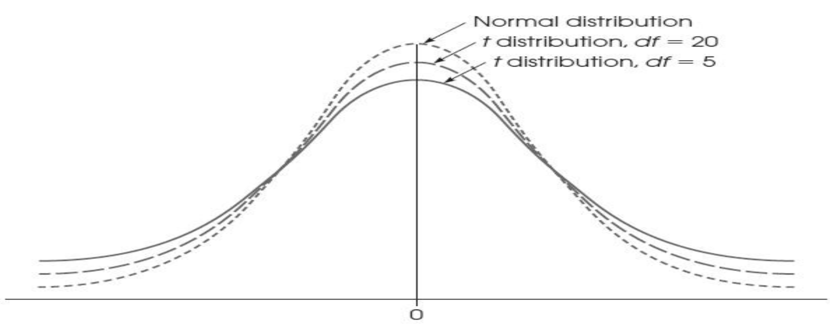
$$z = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\sqrt{\sigma^2 / n}} \qquad t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\sqrt{s^2 / n}}$$

Reviewing briefly, you must know the sample mean before you can compute sample variance. This places a restriction on sample variability such that only $n - 1$ scores in a sample are independent and free to vary. The value $n - 1$ is called the *degrees of freedom* (or *df*) for the sample variance.

degrees of freedom = *df* = *n* – 1

Degrees of freedom describe the number of scores in a sample that are independent and free to vary. Because the sample mean places a restriction on the value of one score in the sample, there are $n - 1$ degrees of freedom for a sample with n scores.

A *t* distribution is the complete set of *t* values computed for every possible random sample for a specific sample size (*n*) or a specific degrees of freedom (*df*). The *t* distribution approximates the shape of a normal distribution.



The exact shape of a t distribution changes with degrees of freedom. In fact, statisticians speak of a “family” of t distributions. That is, there is a different sampling distribution of t (a distribution of all possible sample t values) for each possible number of degrees of freedom. As *df* gets very large, the t distribution gets closer in shape to a normal z-score distribution. A quick glance at Figure reveals that distributions of t are bell-shaped and symmetrical and have a mean of zero. However, the t distribution has more variability than a normal z distribution, especially when *df* values are small. The t distribution tends to be flatter and more spread out, whereas the normal z distribution has more of a central peak. The reason that the t distribution is flatter and more variable than the normal z-score distribution becomes clear if you look at the structure of the formulas for z and t. The t distribution has more variability than a normal z distribution, especially when *df* values are small. The t distribution tends to be flatter and more spread out, whereas the normal z distribution has more of a central peak.

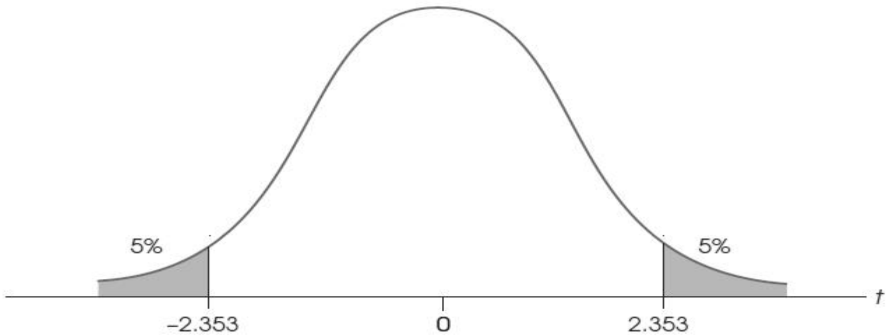
Determining Proportions And Probabilities For t Distributions

Just as we used the unit normal table to locate proportions associated with *z*-scores, we use a *t* distribution table to find proportions for *t* statistics. The complete *t* distribution table is presented below. The two rows at the top of the table show proportions of the *t* distribution contained in either one or two tails, depending on which row is used. The first column of the table lists degrees of freedom for the *t* statistic. Finally, the numbers in the body of the table are the *t* values that mark the boundary between the tails and the rest of the *t* distribution.

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707

For example, with *df* = 3, exactly 5% of the t distribution is located in the tail beyond t = 2.353. The process of finding this value is highlighted in table above. Begin by locating *df* = 3 in the first column of the table. Then locate a pro-portion of 0.05 (5%) in the one-tail proportion row. When you line up these two values in the table, you should find *t* = 2.353. Similarly, 5% of the t distribution is located in the tail beyond t = ±2.353. Finally, notice that a total of 10% (or 0.10) is contained in the two tails beyond t = ±2.353 (check the proportion value in the "two-tails combined" row at the top of the table).

A close inspection of the *t* distribution table demonstrates a point we made earlier: As the value for *df* increases, the t distribution becomes more similar to a normal distribution. For example, examine the column containing *t* values for a 0.05 proportion in two tails. You will find that when *df* = 1, the *t* values that separate the extreme 5% (0.05) from the rest of the distribution are t = ± 12.706.



Hypothesis Tests With The t Statistic

As always, the null hypothesis states that the treatment has no effect; specifically, *H*0 states that the population mean is unchanged. Thus, the null hypothesis provides a specific value for the unknown population mean. The sample data provide a value for the sample mean. Finally, the variance and estimated standard error are computed from the sample data. When these values are used in the *t* formula, the result becomes

$$t = \frac{\begin{array}{c}\text{sample mean} \\ \text{(from the data)}\end{array} - \begin{array}{c}\text{population mean} \\ \text{(hypothesized from } H_0\text{)}\end{array}}{\begin{array}{c}\text{estimated standard error} \\ \text{(computed from the sample data)}\end{array}}$$

The following research situation demonstrates the procedures of hypothesis testing with the *t* statistic. Note that this is another example of a null hypothesis that is founded in logic rather than prior knowledge of a population mean.

Constructing Null And Alternative Hypothesis

t statistic permits hypothesis testing in situations for which a population mean is not known

The t test does not require any prior knowledge about the population mean or the population variance.

The null hypothesis would state

Ho: $\mu=4$

The alternative hypothesis would state that

H1: $\mu\neq4$

Infants, even newborns, prefer to look at attractive faces compared to less attractive faces (Slater, et al., 1998). In the study, infants from 1 to 6 days old were shown two photographs of women's faces. Previously, a group of adults had rated one of the faces as significantly more attractive than the other. The babies were positioned in front of a screen on which the photographs were presented. The pair of faces remained on the screen until the baby accumulated a total of 20 seconds of looking at one or the other. The number of seconds looking at the attractive face was recorded for each infant. Suppose that the study used a sample of $n = 9$ infants and the data produced an average of $M = 13$ seconds for the attractive face with $SS = 72$. Note that all of the available information comes from the sample. Specifically, we do not know the population mean or the population standard deviation.

Step 1

State the hypotheses and select an alpha level. Although we have no information about the population of scores, it is possible to form a logical hypothesis about the value of μ. In this case, the null hypothesis states that the infants have no preference for either face. That is, they should average half of the 20 seconds looking at each of the two faces. In symbols, the null hypothesis states

$H0$: μ attractive = 10 seconds

The alternative hypothesis states that there is a preference and one of the faces is preferred over the other. A directional, one-tailed test would specify which of the two faces is preferred, but the non-directional alternative hypothesis is expressed as follows:

$H1$: μ attractive $\neq$ 10 seconds

We set the level of significance at $\alpha = .05$ for two tails.

Step 2

Locate the critical region. The test statistic is a *t* statistic because the population variance is not known. Therefore, the value for degrees of freedom must be determined before the critical region can be located. For this sample

$df = n - 1 = 9 - 1 = 8$

For a two-tailed test at the .05 level of significance and with 8 degrees of freedom, the critical region consists of t values greater than +2.306 or less than –2.306.

Step 3

Calculate the test statistic. The t statistic typically requires much more computation than is necessary for a z-score. Therefore, we recommend that you divide the calculations into a three-stage process as follows:

a. First, calculate the sample variance. Remember that the population variance is unknown, and you must use the sample value in its place. (This is why we are using a t statistic instead of a z-score.)

$$s^2 = \frac{SS}{n-1} = \frac{SS}{df} = \frac{72}{8} = 9$$

b. Next, use the sample variance ($s2$) and the sample size (n) to compute the estimated standard error. This value is the denominator of the t statistic and measures how much difference is reasonable to expect by chance between a sample mean and the corresponding population mean.

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{9}{9}} = \sqrt{1} = 1$$

Finally, compute the t statistic for the sample data.

$$t = \frac{M - \mu}{s_M} = \frac{13-10}{1} = 3.00$$

Step 4

Make a decision regarding H0. The obtained t statistic of 3.00 falls into the critical region on the right-hand side of the t distribution. Our statistical decision is to reject $H0$ and conclude that babies do show a preference when given a choice between an attractive and an unattractive face. Specifically, the average amount of time that the babies spent looking at the attractive face was significantly different from the 10 seconds that would be expected if there were no preference.

As indicated by the sample mean, there is a tendency for the babies to spend more time looking at the attractive face.

Assumptions Of The t Test

Two basic assumptions are necessary for hypothesis tests with the *t* statistic.

1. The values in the sample must consist of *independent* observations. In everyday terms, two observations are independent if there is no consistent, predictable relationship between the first observation and the second. More precisely, two events (or observations) are independent if the occurrence of the first event has no effect on the probability of the second event.

2. The population that is sampled must be normal. This assumption is a necessary part of the mathematics underlying the development of the *t* statistic and the *t* distribution table. However, violating this assumption has little practical effect on the results obtained for a *t* statistic, especially when the sample size is relatively large. With very small samples, a normal population distribution is important. With larger samples, this assumption can be violated without affecting the validity of the hypothesis test. If you have reason to suspect that the population distribution is not normal, use a large sample to be safe.

The Influence Of Sample Size And Sample Variance

The number of scores in the sample and the magnitude of the sample variance both have a large effect on the t statistic and thereby influence the statistical decision.

As the estimated standard error appears in the denominator of the formula, a larger value for sM produces a smaller value (closer to zero) for *t*.

The estimated standard error is directly related to the sample variance so that the larger the variance, the larger the error.

The estimated standard error is inversely related to the number of scores in the sample. The larger the sample is, the smaller the error is.

If all other factors are held constant, large samples tend to produce bigger t statistics and therefore are more likely to produce significant results.

The estimated standard error is directly related to the sample variance so that the larger the variance, the larger the error. Thus, large variance means that you are less likely to obtain a significant treatment effect. In general, large variance is bad for inferential statistics. Large variance means that the scores are widely scattered, which makes it difficult to see any consistent patterns or trends in the data. In general, high variance reduces the likelihood of rejecting the null hypothesis.

On the other hand, the estimated standard error is inversely related to the number of scores in the sample. The larger the sample is, the smaller the error is. If all other factors are held constant, large samples tend to produce bigger *t* statistics and therefore are more likely to produce significant results.

$$t = \frac{M - \mu}{s_M} \qquad \text{where } s_M = \sqrt{\frac{s^2}{n}}$$

The effect size fort-test becomes:

$$\text{estimated } d = \frac{\text{mean difference}}{\text{sample standard deviation}} = \frac{M - \mu}{s}$$

And where $s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}}$

Lesson 25

T-TEST-II

Calculating T statistics

The one-sample t-test is used to determine whether a sample comes from a population with a specific mean. This population mean is not always known, but is sometimes hypothesized.

For example, you want to show that a new teaching method for pupils struggling to learn English grammar can improve their grammar skills to the national average. Your sample would be pupils who received the new teaching method and your population mean would be the national average score. Alternately, you believe that doctors that work in accident and emergency (A & E) departments work 100 hour per week despite the dangers (e.g., tiredness) of working such long hours. You sample 1000 doctors in A & E departments and see if their hours differ from 100 hours.

In two-tailed test, a sample of n=4 individuals is selected from a population with a mean of μ=40. A treatment is administered to the individuals in the sample and, after treatment, the sample has a mean of M=44 and a variance of s^2=16. In this type of data to know "Is this sample sufficient to conclude that the treatment has a significant effect?" We have to use a two-tailed test with alpha .05. In one-tailed test, in a study, the research question is whether attractiveness affects the behavior of infants looking at photographs of women's faces. The researcher predicts that the infants will spend more than half of the 20-second period looking at the attractive face.

The researcher tested a sample of n=9 infants and obtained a mean of M=13 seconds looking at the attractive face with SS=72. We use one tail test.

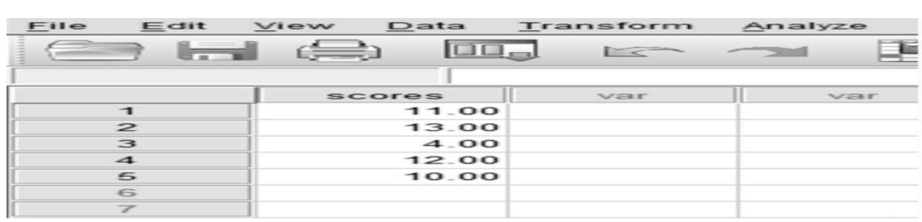
One sample T-test using SPSS

Data Entry

A random sample of n=5 participants with scores on a test (11,13,4,12,10) is obtained from a population with a mean of μ=13. A treatment is administered to the individuals in the sample.

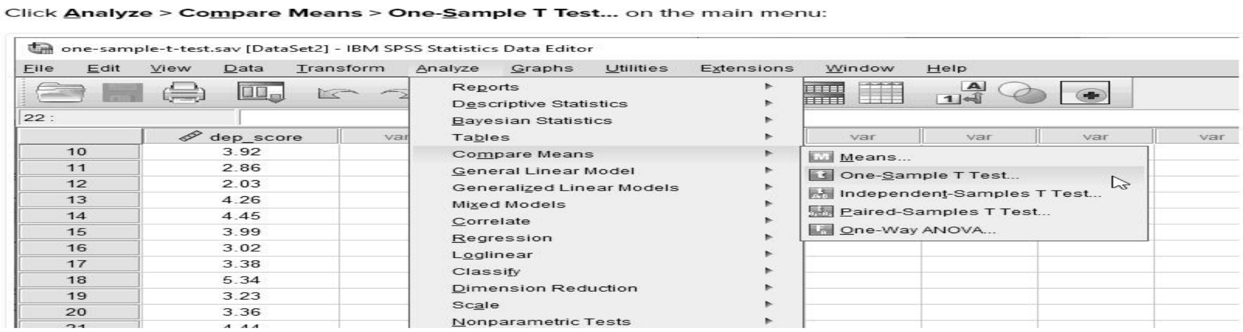
A researcher wants to examine if the sample means differ from population mean after treatment.

Enter all of the scores from the sample in one column of the data editor.



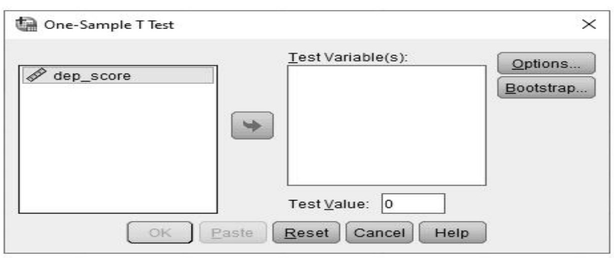
Data Analysis

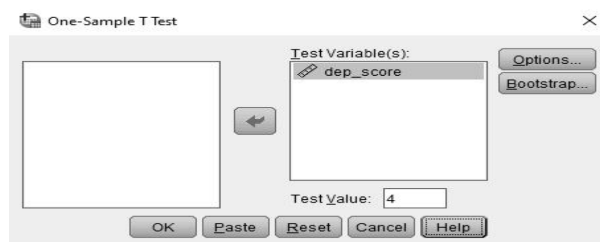
1. Click Analyze on the tool bar, select Compare Means, and click on One-Sample T Test.



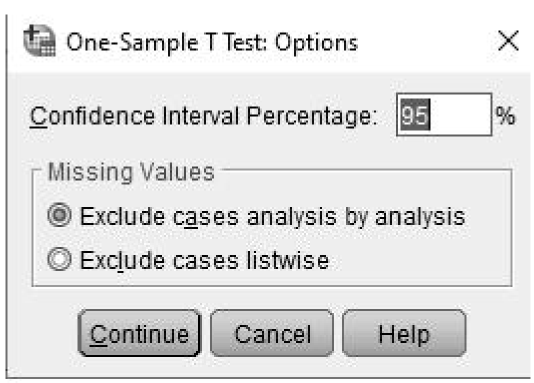
2. Highlight the column label for the set of scores in the left box and click the arrow to move it into the Test Variable(s) box.

3. In the Test Value box at the bottom of the One-Sample t Test window, enter the hypothesized value for the population mean from the null hypothesis (μ=13)





4. In addition to performing the hypothesis test, the program computes a confidence interval for the population mean difference. The confidence level is automatically set at 95%, but you can select Options and change the percentage. Click on the Continue button. You will be returned to the One-Sample T Test dialogue box.Click on the OK button to generate the output. It will generate SPSS output.



Reporting Results Of T-Test

SPSS Statistics generates two main tables of output for the one-sample t-test that contains all the information you require to interpret the results of a one-sample t-test.

A scientific report typically uses the term significant to indicate that the null hypothesis has been rejected and the term not significant to indicate failure to reject Ho.

There is a prescribed format for reporting the calculated value of the test statistic, degrees of freedom, and alpha level for a t test.

You can make an initial interpretation of the data using the One-Sample Statistics table, which presents relevant descriptive statistics. It is more common than not to present your descriptive statistics using the mean and standard deviation ("Std. Deviation" column) rather than the standard error of the mean ("Std. Error Mean" column), although both are acceptable.

However, by running a one-sample t-test, you are really interested in knowing whether the sample you have (dep_score) comes from a 'normal' population (which has a mean of 4.0). This is discussed in the next section.

The One-Sample Test table reports the result of the one-sample t-test. The top row provides the value of the known or hypothesized population mean you are comparing your sample data to, as highlighted below:

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

In this example, you can see the 'normal' depression score value of "4" that you entered in earlier. You now need to consult the first three columns of the One-Sample Test table, which provides information on whether the sample is from a population with a mean of 4 (i.e., are the means statistically significantly different).

Moving from left-to-right, you are presented with the observed t-value ("t" column), the degrees of freedom ("df"), and the statistical significance (p-value) ("Sig. (2-tailed)") of the one-sample t-test. In this example, $p < .05$ (it is $p = .022$). Therefore, it can be concluded that the population means are statistically significantly different. If $p > .05$, the difference between the sample-estimated population mean and the comparison population mean would not be statistically significantly different.

SPSS Statistics also reports that t = -2.381 ("t" column) and that there are 39 degrees of freedom ("df" column). You need to know these values in order to report your results, which you could do as follows:

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

Depression score was statistically significantly lower than the population normal depression score, t(39) = -2.381, p = .022.

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

This section of the table shows that the mean difference in the population means is -0.28 ("Mean Difference" column) and the 95% confidence intervals (95% CI) of the difference are -0.51 to -0.04 ("Lower" to "Upper" columns). For the measures used, it will be sufficient to report the values to 2 decimal places. You could write these results as:

Depression score was statistically significantly lower than the population normal depression score, t(39) = -2.381, p = .022.

You can also include measures of the difference between the two population means in your written report. This information is included in the columns on the far-right of the One-Sample Test table. This section of the table shows that the mean difference in the population means is -0.28 ("Mean Difference" column) and the 95% confidence intervals (95% CI) of the difference are -0.51 to -0.04 ("Lower" to "Upper" columns). For the measures used, it will be sufficient to report the values to 2 decimal places. You could write these results as:

Depression score was statistically significantly lower by 0.28 (95% CI, 0.04 to 0.51) than a normal depression score of 4.0, t(39) = -2.381, p = .022

Lesson 26

INDEPENDENT SAMPLE T-TEST-I

The t Test for Two Independent Samples

Although these single sample techniques are used occasionally in real research, most research studies require the comparison of two (or more) sets of data. For example, a social psychologist may want to compare men and women in terms of their political attitudes, an educational psychologist may want to compare two methods for teaching mathematics, or a clinical psychologist may want to evaluate a therapy technique by comparing depression scores for patients before therapy with their scores after therapy. The two sets of data could come from two completely separate groups of participants. This is called an independent measures research design or between-subject design.

In each case, the research question concerns a mean difference between two sets of data.

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. For example, you could use an independent t-test to understand whether first year graduate salaries differed based on gender (i.e., your dependent variable would be "first year graduate salaries" and your independent variable would be "gender", which has two groups: "male" and "female"). Alternately, you could use an independent t-test to understand whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates").

There are two general research designs that can be used to obtain the two sets of data to be compared:

1. The two sets of data could come from two completely separate groups of participants. For example, the study could involve a sample of men compared with a sample of women. Or the study could compare grades for one group of freshmen who are given laptop computers with grades for a second group who are not given computers.

2. The two sets of data could come from the same group of participants. For example, the researcher could obtain one set of scores by measuring depression for a sample of patients before they begin therapy and then obtain a second set of data by measuring the same individuals after 6 weeks of therapy.

The first research strategy, using completely separate groups, is called an independent measures research design or a between-subjects design. These terms emphasize the fact that the design involves separate and independent samples and makes a comparison between two groups of individuals.

A research design that uses a separate group of participants for each treatment condition (or for each population) is called an independent-measures research design or a between-subjects research design.

Assumptions Underlying The Independent-Measures t Formula

There are three assumptions that should be satisfied before you use the independent measures t formula for hypothesis testing:

1. The observations within each sample must be independent.

2. The two populations from which the samples are selected must be normal.

3. The two populations from which the samples are selected must have equal variances.

The third assumption is referred to as homogeneity of variance and states that the two populations being compared must have the same variance. You may recall a similar assumption for the z-score hypothesis test. For that test, we assumed that the effect of the treatment was to add a constant amount to (or subtract a constant amount from) each individual score. As a result, the population standard deviation after treatment was the same as it had been before treatment. We now are making essentially the same assumption, but phrasing it in terms of variances.

The t statistic for an independent-measures research design involves two separate samples, we need some special notation to help specify which data go with which sample. This notation involves the use of subscripts, which are small numbers written beside a sample statistic. For example, the number of scores in the first sample would be identified by n1; for the second sample, the number of scores is n2. The sample means would be identified by M1 and M2. The sums of squares would be SS1 and SS2.

Hypothesis Testing

The goal of an independent-measures research study is to evaluate the mean difference between two populations (or between two treatment conditions). Using subscripts to differentiate the two populations, the mean for the first population is $\mu 1$, and the second population mean is μ 2. The difference between means is simply μ 1 - μ 2. As always, the null hypothesis states that there is no change, no effect, or, in this case, no difference. Thus, in symbols, the null hypothesis for the independent-measures test is

***H*0: *μ* 1 - *μ* 2 = 0 (No difference between the population means)**

The alternative hypothesis states that there is a mean difference between the two populations

H: $\mu - \mu_2 \neq 0$

Calculation Of Independent Sample t-Test

1. The basic structure of the *t* statistic is the same for both the independent-measures and the single-sample hypothesis tests. The overall t formula is given below:

t=sample mean difference−population mean difference /estimated standard error

$t=[(M_1-M_2)-(\mu_1-\mu_2)] / S(M_1-M_2)$

2. The independent-measures *t* is basically a *two-sample* t *that doubles all the elements of the single-sample* t *formulas*. To demonstrate the second point, we examine the two *t* formulas piece by piece.

The single-sample *t* uses one sample mean to test a hypothesis about one population mean. The sample mean and the population mean appear in the numerator of the *t* formula, which measures how much difference there is between the sample data and the population hypothesis.

t= sample mean-population mean/ estimated standard error = $M - \mu / s_M$

The independent measure t test uses difference between two sample means to evaluate a hypothesis about the difference between population means. Thus, the independent measure t formula is:

t= sample mean difference -population mean difference / estimated standard error

$t=[(M_1-M_2)-(\mu_1-\mu_2)] / S(M_1-M_2)$

In this formula, the value of *M*1 - *M*2 is obtained from the sample data and the value of $\mu_1-\mu_2$ comes from the null hypothesis.

The overall t formula The single-sample t uses one sample mean to test a hypothesis about one population mean. The sample mean and the population mean appear in the numerator of the t formula, which measures how much difference there is between the sample data and the population hypothesis.

Interpreting The Estimated Standard Error

The estimated standard error of M1 - M2 that appears in the bottom of the independent-measures t statistic can be interpreted in two ways. First, the standard error is defined as a measure of the standard, or average, distance between a sample statistic (M1 - M2) and the corresponding population parameter (1 - 2). As always, samples are not expected to be perfectly accurate and the standard error measures how much difference is reasonable to expect between a sample statistic and the population parameter.

This produces a second interpretation for the estimated standard error. Specifically, the standard error can be viewed as a measure of how much difference is reasonable to expect between two sample means if the null hypothesis is true. The second interpretation of the estimated standard error produces a simplified version of the independent-measures t statistic.

For the independent-measures t statistic, we want to know the total amount of error involved in using two sample means to approximate two population means. To do this, we find the error from each sample separately and then add the two errors together. The resulting formula for standard error is

For the independent-measures t statistic, there are two SS values and two df values (one from each sample). The values from the two samples are combined to compute what is called the pooled variance.

With one sample, the variance is computed as SS divided by df. With two samples, the pooled variance is computed by combining the two SS values and then dividing by the combination of the two df values. As we mentioned earlier, the pooled variance is actually an average of the two sample variances, but the average is computed so that the larger sample carries more weight in determining the final value. The following examples demonstrate this point.

When computing the pooled variance, the weight for each of the individual sample variances is determined by its degrees of freedom. Because the larger sample has a larger df value, it carries more weight when averaging the two variances. This produces an alternative formula for computing *pooled variance*.

For the independent-measures t formula, the standard error measures the amount of error that is expected when you use a sample mean difference (M_1-M_2) to represent a population mean difference ($\mu_1-\mu_2$). The standard error for the sample mean difference is represented by the symbol $S(M_1 - M_2)$.

Each of the two sample means represents it own population mean, but in each case there is some error.

M_1 approximates $\mu_1$with some error.

M_2 approximates μ_2 with some error.

For the independent-measures t statistic, we want to know the total amount of error involved in using two sample means to approximate two population means.

$S(M_1 - M_2) = \sqrt{(s_1^2/n_1)+ (s_2^2/n_2)}$

The *estimated standard error of M*1 - *M*2 that appears in the bottom of the independent-measures *t* statistic can be interpreted in two ways. First, the standard error is defined as a measure of the standard, or average, distance between a sample statistic (*M*1 - *M*2) and the corresponding population parameter (μ1 - μ 2). As always, samples are not expected to be perfectly accurate and the standard error measures how much difference is reasonable to expect between a sample statistic and the population parameter.

This produces a second interpretation for the estimated standard error. Specifically, the standard error can be viewed as a measure of how much difference is reasonable to expect between two sample means if the null hypothesis is true. The second interpretation of the estimated standard error produces a simplified version of the independent-measures *t* statistic.

For the independent-measures *t* statistic, we want to know the total amount of error involved in using *two* sample means to approximate *two* population means. To do this, we find the error from each sample separately and then add the two errors together.

Pooled Variance

For correcting the bias in the standard error is to combine the two sample variances into a single value called the pooled variance.

The pooled variance is obtained by averaging or "pooling" the two sample variances using a procedure that allows the bigger sample to carry more weight in determining the final value.

For the independent-measures t statistic, there are two SS values and two df values. The values from the two samples are combined to compute what is called the pooled variance.

The pooled variance is identified by the symbol ***S^2p***

$S^2p=(SS_1+SS_2)/df_1+df_2$

With one sample, the variance is computed as *SS* divided by *df*. With two samples, the pooled variance is computed by combining the two *SS* values and then dividing by the combination of the two *df* values. As we mentioned earlier, the pooled variance is actually an average of the two sample variances, but the average is computed so that the larger sample carries more weight in determining the final value. The following examples demonstrate this point.

Note that the pooled variance is exactly halfway between the two sample variances. Because the two sample are exactly the same size, the pooled variance is simply the average of the two sample variance.

This time the pooled variance is not located halfway between the two sample variance. Instead, the pooled value is closer to the variance for the larger sample ($n=9$ and $s^2=6$) than to the variance for the smaller sample ($n=3$ and $s^2=10$). the larger sample carries more weght when the pooled variance s computed.

When computing the pooled variance, the weight for each of the individual sample variances is determined by its degrees of freedom. Because the larger sample has a larger *df* value, it carries more weight when averaging the two variances. This produces an alternative formula for computing pooled variance.

Estimated Standard Error

Using the pooled variance in place of the individual sample variances, we can now obtain an unbiased measure of the standard error for a sample mean difference. The resulting formula for the independent-measures estimated standard error is

$M_1 - M_2 = S(M_1 - M_2) = \sqrt{}(Sp^2/n_1) + (Sp^2/n_2)$

Confidence intervals for estimating $\mu_1-\mu_2$

For the independent-measures t, we use a sample mean difference, M_1-M_2, to estimate the population mean difference, $\mu_1-\mu_2$.

$\mu_1-\mu_2 = M_1-M_2 \pm t \times S(M_1-M_2)$

Conceptually, this standard error measures how accurately the difference between two sample means represents the difference between the two population means. In a hypothesis test, $H0$ specifies that $\mu 1 - \mu 2 = 0$, and the standard error also measures how much difference is expected, on average, between the two sample means. In either case, the formula combines the error for the first sample mean with the error for the second sample mean. Also note that the pooled variance from the two samples is used to compute the standard error for the sample mean difference.

The basic structure of the t statistic is the same for both the independent-measures and the single-sample hypothesis tests.

t=sample statistic −*h*hypothesized population mean /estimated standard error

The overall t formula is given below:

t=sample mean difference−population mean difference /estimated standard error

$t = [(M_1 - M_2) - (\mu_1 - \mu_2)] / S(M_1 - M_2)$

The degrees of freedom for the independent-measures *t* statistic are determined by the *df* values for the two separate samples:

df for the t statistic= *df* for the first sample + *df* for the first sample

$= df_1 + df_2$

$= (n_1 - 1) + (n_2 - 1)$

Equivalently, the *df* value for the independent measure *t* statistic can be expressed as:

$df = n_1 + n_2 - 2$

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample *t* statistic	M	μ	$\sqrt{\frac{s^2}{n}}$	$s^2 = \frac{SS}{df}$
Independent-measures *t* statistic	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$

Reporting The Results

SPSS Statistics generates two main tables of output for the independent t-test. Group statistics table provides useful descriptive statistics for the two groups that you compared, including the mean and standard deviation.independent samples test table provides the actual results from the independent t-test.

The participants who were informed about the withheld evidence gave significantly longer sentences than those participants who were not informed about the withhold evidence. The mean difference was significant, $t(14)= -3.53$, $p<.05$.

If an exact probability is available from a computer analysis, it should be reported.

If a confidence interval is reported to describe effect size, it appears immediately after the results from the hypothesis test.

Lesson 27

INDEPENDENT SAMPLE T-TEST-II

Independent Sample t-Test In SPSS

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. For example, you could use an independent t-test to understand whether first year graduate salaries differed based on gender (i.e., your dependent variable would be "first year graduate salaries" and your independent variable would be "gender", which has two groups: "male" and "female"). Alternately, you could use an independent t-test to understand whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates").

Independent sample t-test is used when the mean scores of two different groups of people or conditions are compared. It is also called between-subject design.This is a parametric test because it requires assumptions about parameters. Parametric tests require a numerical score for each individual in the sample. The scores then are added, squared, averaged, and otherwise manipulated using basic arithmetic.

This test tells whether there is a statistically significant difference in the mean scores for the two groups, for instance, whether males and females differ significantly in terms of their self-esteem levels. In statistical terms, it tests the probability that the two sets of scores (for males and females) came from the same population.

To run independent sample t-test in SPSS, we need two set of variables;

1. The ***Continuous /Dependent Variable***, whose mean score is compared between the two groups. SPSS defines it as the test Variable having scale measurement.

2. The ***Categorical/Independent Variable*** with two categories that defines which two samples will be compared in the t-test. In SPSS, for t-test it is defined as the Grouping Variable.

The grouping variable must have at least two categories (groups); it may have more than two categories but a t-test can only compare two groups, so there is need to specify which two groups to compare. We can also use a continuous variable by specifying a cut point to create two groups (i.e., values at or above the cut point and values below the cut point).

Downs and Abwender (2002) evaluated athletes (soccer players and swimmers) to determine whether the blows to the head experienced by both groups produced long-term neurological deficits. In the study, neurological tests were administered to mature soccer players and swimmers and a researcher obtained the following data presented in the table.

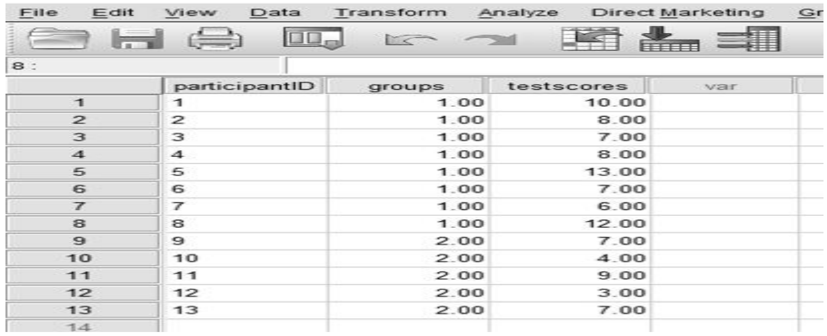
Are there any significant differences in the neurological test scores?

Swimmers	Soccer Players
10	7
8	4
7	9
8	3
13	7
7	
6	
12	

Running Independent Sample t-Test In SPSS

Entering Data

- In keeping the above mentioned example, first of all, the variables must be setup in the variable view tab of SPSS data editor, although it is not necessary but it is good practice to give each person a unique ID.
- We also need to provide codes that distinguish between the two groups (i.e., swimmers and soccer players).
- By clicking in the box of the 'values' field in the row containing the groups variable, we get a pop-up dialog that allows us to code the categorical variable. The swimmer is coded as '1' and the soccer player as '2'.

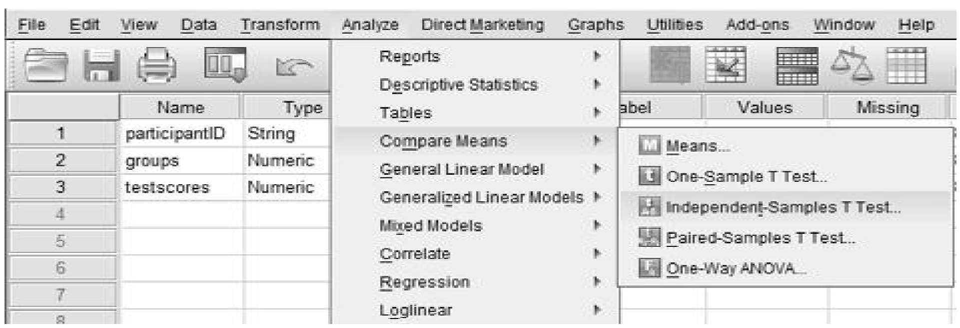


	participantID	groups	testscores	var
1	1	1.00	10.00	
2	2	1.00	8.00	
3	3	1.00	7.00	
4	4	1.00	8.00	
5	5	1.00	13.00	
6	6	1.00	7.00	
7	7	1.00	6.00	
8	8	1.00	12.00	
9	9	2.00	7.00	
10	10	2.00	4.00	
11	11	2.00	9.00	
12	12	2.00	3.00	
13	13	2.00	7.00	

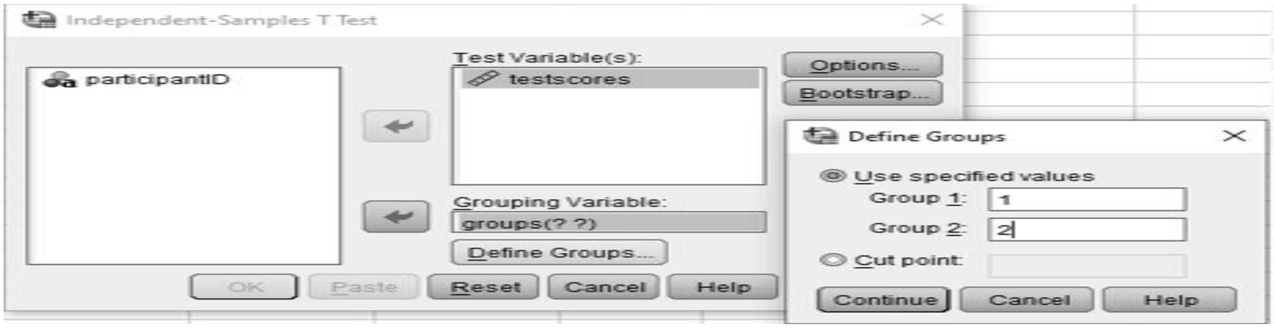
- The data is entered in the data view tab of the SPSS editor.
- We need to enter data for both the independent(categorical, i.e., blows to the head experienced by swimmers and soccer players)and dependent (continuous, i.e., test scores) variables.
- A code is entered for the both groups as they were on (1 = swimmer, 2 = soccer player) and the test score of each is entered.

Running Analysis

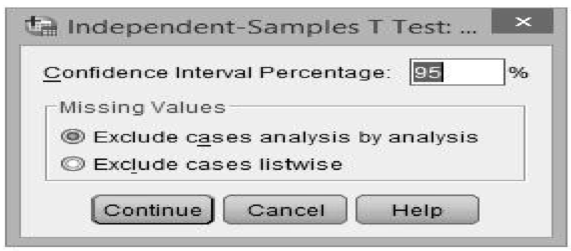
- To run independent sample t-test in SPSS, Click Analyze > Compare Means > Independent-Samples T Test, on the top menu, as shown below:



- The Independent-Samples T Test window opens where we transfer the dependent variable, tests cores, into the Test Variable(s): box, and the independent variable, groups, into the Grouping Variable: box, by highlighting the relevant variables and pressing the SPSS right arrow button.
- We then need to define the groups (athletes, swimmers and soccer players).
- By clicking on the button we will be presented with the Define Groups dialogue box, as shown below.
- Enter 1 into the Group 1: box and enter 2 into the Group 2: box.
- Remember that we labeled the swimmer group as 1 and the soccer player group as 2.
- Click the button Continue.



- The Options section is where we can set our desired confidence interval/level for the mean difference, and specify how SPSS should handle missing values. In SPSS by default confidence interval is set at 95% and we do not need to change it.
- The Missing Values section is not relevant if we have only specified one dependent variable; it only matters if we are entering more than one dependent (continuous numeric) variable. By clicking to the exclude values case by case.



- Then, click Continue and OK when finished.

Interpreting Results:

In the Group Statistics box, SPSS gives us the mean and standard deviation for each of your groups (in this case, swimmers and soccer players).

Group Statistics

	groups	N	Mean	Std. Deviation	Std. Error Mean
testscores	swimmers	8	8.8750	2.53194	.89518
	soccer players	5	6.0000	2.44949	1.09545

It also gives the number of people in each group (N).

The second section, Independent Samples Test, displays the results most relevant to the Independent Samples t-test.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
testscores	Equal variances assumed	.022	.886	2.015	11	.069	2.87500	1.42651	-.26474	6.01474
	Equal variances not assumed			2.032	8.867	.073	2.87500	1.41469	-.33260	6.08260

There are two parts that provide different pieces of information:

(A) Levene's Test for Equality of Variances

(B) t-test for Equality of Means.

In B part (t-test for Equality of Means) if the value in the Sig. (2-tailed) column is equal to or less than .05, there is a significant difference in the mean scores on dependent variable for each of the two groups.

The independent t-test assumes the variances of the two groups you are measuring are equal in the population. If your variances are unequal, this can affect the Type I error rate. The assumption of homogeneity of variance can be tested using Levene's Test of Equality of Variances, which is produced in SPSS Statistics when running the independent t-test procedure. If you have run Levene's Test of Equality of Variances in SPSS Statistics. This test for homogeneity of variance provides an F-statistic and a significance value (p-value). We are primarily concerned with the significance value – if it is greater than 0.05 (i.e., $p > .05$), our group variances can be treated as equal. However, if $p < 0.05$, we have unequal variances and we have violated the assumption of homogeneity of variances.

If the value is above .05, there is no significant difference between the two groups.

In the example presented here, the Sig. (2-tailed) value is .069. As this value is above the required cut-off of .05, we can conclude that there s not a statistically significant difference in the mean neurological test scores for Swimmers and soccer players.

Mean Difference is the difference between the sample means; it also corresponds to the numerator of the test statistic. Std. Error Difference is the standard error; it also corresponds to the denominator of the test statistic. Confidence interval of the Difference part of the t-test output complements the significance test results.

Typically, if the CI for the mean difference contains 0, the results are not significant at the chosen significance level. In this example, the CI is [-.265, +6.015], which does contain zero; this agrees with the p-value of the significance test.

Assumptions Of Independent Sample t-Test

There are some general assumptions that apply to all of the parametric tests (e.g. t-tests, analysis of variance), and additional assumptions associated with specific techniques.

These general assumption include;

1. Level of Measurement

Each of the parametric approaches assumes that the dependent variable is measured at the interval or ratio level; that is, using a continuous scale rather than discrete categories.

2. Random Sampling

The parametric techniques assume that the scores are obtained using a random sample from the population.

3. Independence of Observations

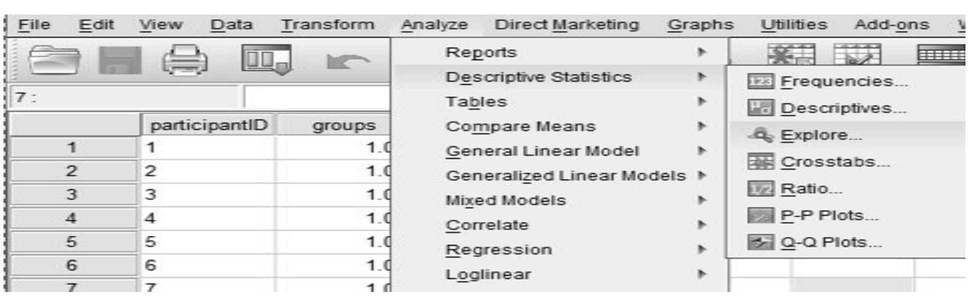
The observations that make up data must be independent of one another; that is, each observation or measurement must not be influenced by any other observation or measurement.

4. Normal Distribution

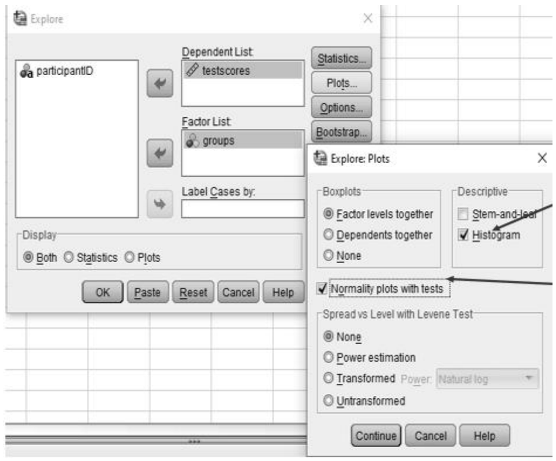
For parametric techniques, it is assumed that the populations from which the samples are taken are normally distributed.

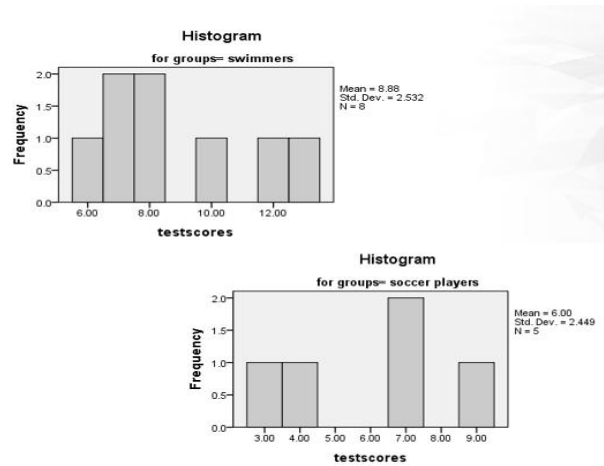
The technique available in SPSS is using the Explore option of the Descriptive Statistics menu.

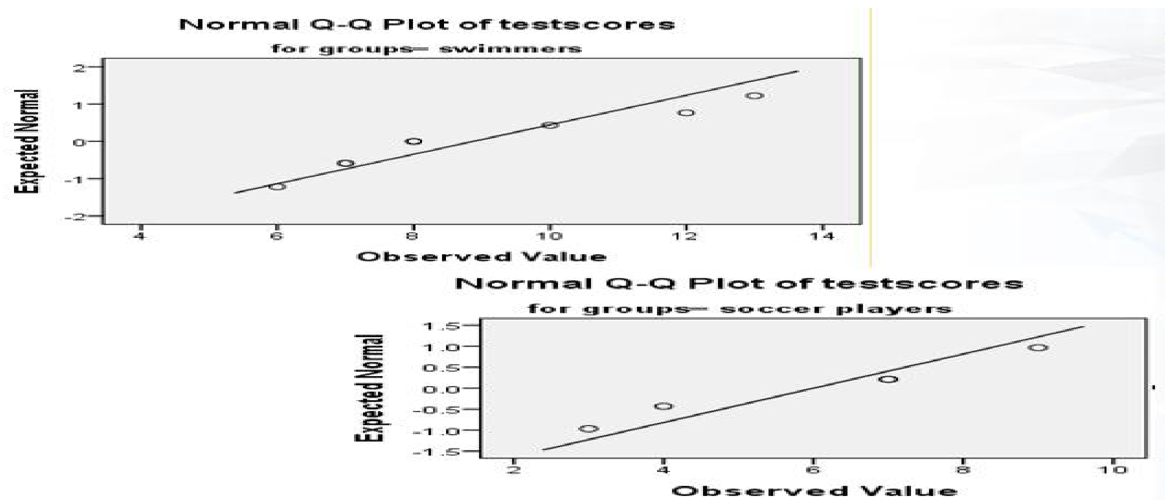
- Click on the variable(s) (test score) and move it into the Dependent List box.
- In the Factor list section add groups.
- In the Display section, make sure that Both is selected.
- Click on the Statistics button and click on Descriptives and Outliers. Click on Continue.



- Click on the Plots button. Under Descriptive, click on Histogram to select it.







- Click on Normality plots with tests. Click on Continue.
- Click on the Options button. In the Missing Values section, click on Exclude cases pairwise.
- The actual shape of the distribution for each group can be seen in the Histograms.
- This is also supported by an inspection of the normal probability plots (labeled Normal Q-Q Plot).
- In this plot, the observed value for each score is plotted against the expected value from the normal distribution. A reasonably straight line suggests a normal distribution.The Skewness value provides an indication of the symmetry of the distribution. Positive skewness values suggest that scores are clustered to the left at the low values. Negative skewness values indicate a clustering of scores at the high end (right-hand side of a graph). Kurtosis, on the other hand, provides information about the 'peakedness' of the distribution. Positive kurtosis values indicate that the distribution is rather peaked (clustered in the center), with

long thin tails. Kurtosis values below 0 indicate a distribution that is relatively flat (too many cases in the extremes).

- In the table labeled Tests of Normality, we are given the results of the Kolmogorov-Smirnov statistic. This assesses the normality of the distribution of scores. A non-significant result (Sig. value of more than .05) indicates normality. Other than mean and SD, descriptive statistics also provide some information concerning the distribution of scores on continuous variables (skewness and kurtosis). This information is needed if the variables are to be used in parametric statistical techniques (e.g. t-tests, analysis of variance).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	groups	Statistic	df	Sig.	Statistic	df	Sig.
testscores	swimmers	.260	8	.118	.896	8	.268
	soccer players	.258	5	.200*	.925	5	.563

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

5. Homogeneity of Variance

This assumption states that samples are obtained from populations of equal variances. This means that the variability of scores for each of the groups is similar. To test this, SPSS performs ***Levene's test for equality of variances*** as part of the t-test and analysis of variance analyses.

The first section of the Independent Samples Test output box gives the results of Levene's test for equality of variances.

This tests whether the variance (variation) of scores for the two groups (swimmers and soccer players) is the same. The outcome of this test determines which of the t-values that SPSS provides is the correct one to use.

If the Sig. value for Levene's test is larger than .05 we should use the first line in the table, which refers to Equal variances assumed. If the significance level of Levene's test is p=.05 or less, this means that the variances for the two groups are not the same. Therefore, refer to second line equal variances not assumed.

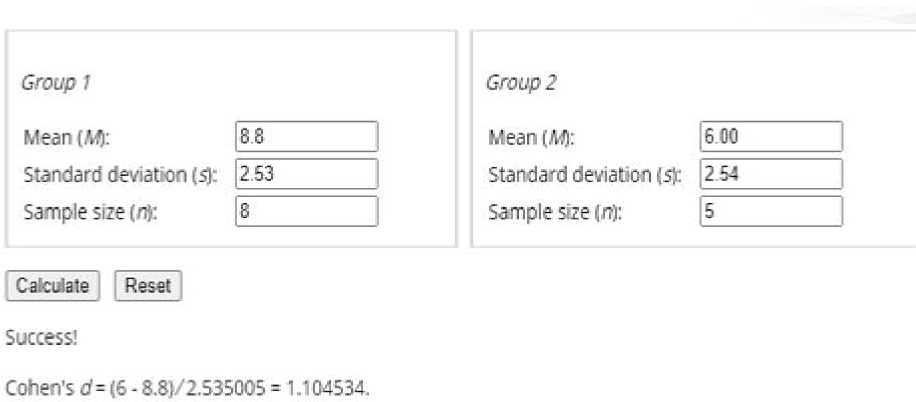
Effect Size

Effect size is the strength of the difference between groups, or the influence of the independent variable.

There are a number of different effect size statistics. The most commonly used to compare groups are partial eta squared and Cohen's d. Partial eta squared measures effect size for analysis of variance. Cohen's d, presents difference between groups for t-test.

SPSS calculates partial eta squared as part of the output (e.g. analysis of variance). However, it does not provide effect size statistics for t-tests.

We can use the information provided in the SPSS to calculate whichever effect size statistic we need preferably when the results are significant. For the independent samples T-test, Cohen's d is determined by calculating the mean difference between two groups, and then dividing the result by the pooled standard deviation.The magnitude of the differences in the means as calculated by The online Chen's d calculator was 1.105 as shown in the figure.



Reporting Results In Apa Format

An independent-samples t-test was conducted to compare the neurological test scores for swimmers and soccer players. There were no significant differences in scores for swimmers (M = 8.8, SD = 2.53) and soccer players (M = 6.00, SD = 2.54); t (11) = 2.02, p = .069, two-tailed). The magnitude of the differen**ces** in the means (mean difference = 2.86, 95% CI: -.265 to 6.015) was (Cohen's d= 1.105).

Lesson 28

REPEATED MEASURE T-TEST

The independent-measures design is characterized by the fact that two separate samples are used to obtain the two sets of scores that are to be compared. In this lesson, we examine an alternative strategy known as a repeated-measures design, or a within-subjects design. With a repeated-measures design, two separate scores are obtained for each individual in the sample.

For example, a group of patients could be measured before therapy and then measured again after therapy. Or, response time could be measured in a driving simulation task for a group of individuals who are first tested when they are sober and then tested again after two alcoholic drinks. In each case, the same variable is being measured twice for the same set of individuals; that is, we are literally repeating measurements on the same sample.

A repeated-measures design, or a within-subject design, is one in which the dependent variable is measured two or more times for each individual in a single sample. The same group of subjects is used in all of the treatment conditions.

In a matched-subjects design, each individual in one sample is matched with an individual in the other sample. The matching is done so that the two individuals are equivalent (or nearly equivalent) with respect to a specific variable that the researcher would like to control.

Repeated measure can be make understandable by this example. A research indicates that the color red increases men's attraction to women (Elliot & Niesta, 2008). In the original study, men were shown women's photographs presented on either a white or a red background. Photographs presented on red were rated significantly more attractive than the same photographs mounted on white. In a similar study, a researcher prepares a set of 30 women's photographs, with 15 mounted on a white background and 15 mounted on red. One picture is identified as the test photograph, and appears twice in the set, once on white and once on red. Research question is "are the ratings for the test photograph significantly different when it is presented on a red background compared to a white background?"Each male participant looks through the entire set of photographs and rates the attractiveness of each woman on a 12-point scale.

The T Statistic For A Repeated-Measures Research Design

In repeated-measures research design we use difference scores. Table below shows an example of data from a study that examines this phenomenon. Note that there is one sample of n =4 participants, and that each individual is measured twice. The first score for each person (X1) is a measurement of reaction time before the medication was administered. The second score (X2) measures reaction time 1 hour after taking the medication. Because we are interested in how the medication affects reaction time, we have computed the difference between the first score and the second score for each individual. The difference scores, or D values, are shown in the last column of the table.

Person	Before Medication (X_1)	After Medication (X_2)	Difference D
A	215	210	−5
B	221	242	21
C	196	219	23
D	203	228	25
			$\Sigma D = 64$

$$M_D = \frac{\Sigma D}{n} = \frac{64}{4} = 16$$

Typically, the difference scores are obtained by subtracting the first score (before treatment) from the second score (after treatment) for each person:

Difference score = D = X2 - X1

Hypothesis Testing

In repeated-measures research design, the researcher's goal is to use the sample of difference scores to answer questions about the general population. In particular, the researcher would like to know whether there is any difference between the two treatment conditions for the general population. Note that we are interested in a population of *difference scores*.

That is, we would like to know what would happen if every individual in the population were measured in two treatment conditions (*X*1 and *X*2) and a difference score (*D*) were computed for everyone. Specifically, we are interested in the mean for the population of difference scores. We identify this population mean difference with the symbol μ*D* (using the subscript letter *D* to indicate that we are dealing with *D* values rather than *X* scores). As always, the null hypothesis states that, for the general population, there is no effect, no change, or no difference.

For a repeated-measures study, the null hypothesis states that the mean difference for the general population is zero.

H_o:μD=0

The alternative hypothesis states that there is a treatment effect that causes the scores in one treatment condition to be systematically higher (or lower) than the scores in the other condition.

H_1: $\mu D \neq 0$

For the repeated-measures design, the sample data are difference scores and are identified by the letter *D,* rather than *X*. Therefore, we use *D*s in the formula to emphasize that we are dealing with difference scores instead of *X* values. Also, the population mean that is of interest to us is the population mean difference (the mean amount of change for the entire population), and we identify this parameter with the symbol μD. With these simple changes, the *t* formula for the repeated-measures design becomes:

t= (MD - μD)/SMD

To calculate the estimated standard error, the first step is to compute the variance (or the standard deviation) for the sample of D scores.

s=√SS/n−1= √SS/df

After calculating the variance, calculate the SMD just dividing the variance by taking under root of the sample size of the sample.

SMD= s/√n=√s^2/n

Running Repeated Measure T-test in SPSS

We will run SPSS repeated measure T-test in SPSS by considering the following example.

Researcher wants to examine the effect of a treatment on depression by measuring a group of n=8 participants before and after they receive the treatment. Is there a significant treatment effect? Use alpha .05, two tails.

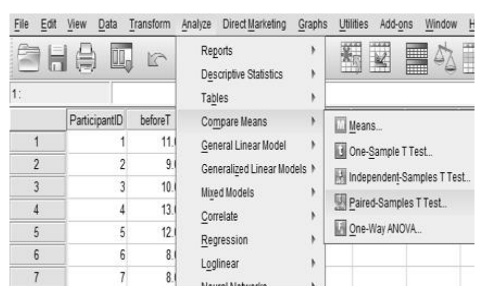
Participants	Before Treatment	After Treatment
1	11	7
2	9	6
3	10	9
4	13	5
5	12	4
6	8	7
7	8	8
8	7	3

We need two set of variables to run paired sample t-test in SPSS:

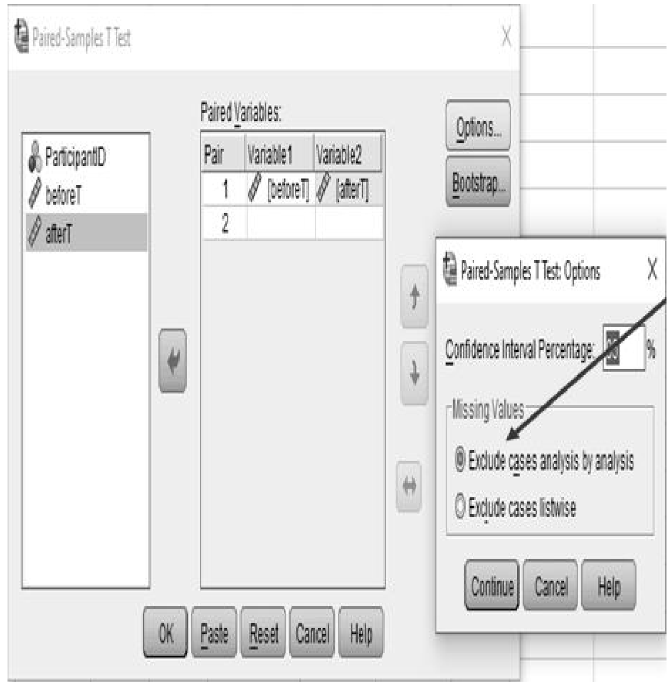
- One categorical independent variable (in this case it is treatment with two different levels before treatment, after treatment)
- One continuous, dependent variable (e.g. Fear of Statistics Test scores measured on two different occasions or under different conditions.

Entering Data

- First of all, the variables must be setup in the variable view tab of SPSS data editor such as before treatment and after treatment.
- We use participant ID as identifier instead of participant names, as this allows us to collect data while keeping the participants anonymous.
- The data is entered in the data view tab of the SPSS editor.
- Before T variable represents the depression scores of participants before treatment.
- After T variable represents the depression scores of participants after treatment.
- To run repeated measure t-test in SPSS, Click Analyze > Compare Means > Paired-Samples T Test on the top menu.



Running Analysis



1. The Paired-Samples T Test window opens where we select the both of these conditions one at a time and click the arrow to move them into the Paired Variables box.
2. The Options section is where we can set our desired confidence interval/level for the mean difference, and specify how SPSS should handle missing values.
3. In SPSS by default confidence interval is set at 95% and we do not need to change it.
4. The Missing Values section is not relevant if we have only specified one dependent variable; it only matters if we are entering more than one dependent (continuous numeric) variable.
5. Then, click Continue and OK button to run the analysis.

Assumptions of Repeated Measure T-test

The related-samples t statistic requires following basic assumptions:

1. Level of Measurement

Just like independent sample t-test, repeated measures also assumes that the dependent variable is measured at the interval or ratio level; that is, using a continuous scale.

2. Random Sampling

Another assumption includes that the scores are obtained using a random sample from the population.

3. Independence of Observations

The observations within each treatment condition must be independent.

This assumption of independence refers to the scores within each treatment.

Inside each treatment, the scores are obtained from different individuals and should be independent of one another.

4. Normal Distribution

This assumption states that the population distribution of difference scores (D values) must be normal.

The normality assumption is not a cause for concern unless the sample size is relatively small.

To test this assumption in SPSS we need to create a new variable that will represent the difference between before treatment and after treatment scores.

- To do this we will go to Transform> Compute Variable.
- Name the target variable (i.e., difference D). The numeric expression will be before treatment – after treatment then click OK.
- Go to Explore option of the Descriptive Statistics menu.
- Click on the newly generated variable (difference D) and move it into the Dependent List box.
- Click on the Statistics button and click on *Descriptives* and Outliers. Click on Continue.
- Click on the Plots button. Under Descriptive, click on Histogram to select it.
- Click on Normality plots with tests. Click on Continue and OK.
- The positive skewness value is suggesting the data is slightly skewed.
- Kurtosis values below 0 indicate a distribution that is relatively flat (too many cases in the extremes).
- The actual shape of the distribution for each group can be seen in the Histograms.
- In Q-Q plot, a reasonably straight line suggests a normal distribution.
- In the table labelled *Tests of Normality*, we are given the results of the Kolmogorov-Smirnov and Shapiro-Wilk statistics.
- The p value is ***0.2 a*** non-significant result (Sig. value of more than .05) would indicates the data is normally distributed.

Interpreting The Output

The Paired Samples Statistics box displays the descriptive statistics for two conditions.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 before T	9.7500	8	2.12132	.75000
after T	6.1250	8	2.03101	.71807

We can see from the two means that participants have lower depression scores after treatment (M=6.13) than before treatment (M=9.75).

SD also indicates the scores in both conditions are similarly dispersed.

The second table in the SPSS Output tells us how the two conditions relate to one another. This table, however, is not necessary for the interpretation of the t-test results.

Paired Samples Correlation

	N	Correlation	Sig
Pair 1 before T &after T	8	-.091	.830

The Paired Samples Test box helps us decide whether there is a statistically significant difference between the conditions.

We need to look in the final column, labelled Sig. (2-tailed). this is the probability (p) value. If this value is less than .05 (in this case .012), we can conclude that there is a significant difference between two scores.

Paired Samples

	Paired Differences					t	*df*	Sig.(2-tailed)
	Mean	**Std. Deviation**	**Std. Error Mean**	**95%CI of the difference**				
				Lower	**Upper**			
Pair1 before T& after T	**3.6350**	**3.06769**	**1.08459**	**1.06035**	**6.18965**	**3.342**	**7**	**.012**

Effect size

In case of repeated measures t-test, we calculate effect size from Cohen's d (measure of effect size). Effect size will be calculated from simply divides the sample mean difference by the sample standard deviation.

Cohen's d= mean difference/ SD

d= 3.63/3.07 = 1.18

Reporting results of repeated measure T-test in APA format

A paired-samples t-test was conducted to evaluate the impact of the intervention on participants' scores on depression. There was a statistically significant decrease in depression scores from before treatment ($M = 9.75$, $SD = 2.12$) to after treatment ($M = 6.13$, $SD = 2.03$), $t(7) = 3.34$, $p < .05$ (two-tailed) and Cohen's $d = 1.18$. The mean decrease in depression scores was 3.62 with a 95% confidence interval ranging from 1.06 to 6.19.

Lesson 29

ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (ANOVA) is a hypothesis-testing procedure that is used to evaluate mean differences between two or more treatments (or populations).

While running ANOVA we must decide between two interpretations:

1. There really are no differences between the populations (or treatments).

2. The populations (or treatments) really do have different means.

When a researcher manipulates a variable to create the treatment conditions in an experiment, the variable is called an independent variable. When a researcher uses a non-manipulated variable to designate groups, the variable is called a quasi-independent variable.

In ANOVA, the variable (independent or quasi-independent) that designates the groups being compared is called a ***factor.*** The individual groups or treatment conditions that are used to make up a factor are called the *levels of the factor*. ANOVA can be used to evaluate the results from a research study that involves more than one factor.

For example, a researcher may want to compare two different therapy techniques, examining their immediate effectiveness as well as the persistence of their effectiveness over time.

A study that combines two factors, is called a *two-factor design or a factorial design* (Two way ANOVA). A study that examines only one independent variable is called *single-factor design* (or One way ANOVA).

Hypotheses for (ANOVA)

In analysis of variance the hypothesis is as follows:

The null hypothesis states that there is no treatment effect.

H_o: $\mu_1 = \mu_2 = \mu_3$

The alternative hypothesis states that there is at least one mean difference among the populations.

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$

OR

H_1: $\mu_1 = \mu_3$ but μ_2 is different

In ANOVA we use variance to measure sample mean differences when there are two or more samples. The test statistic for ANOVA uses this fact to compute an F-ratio with the following structure:

F=Variance between treatments/Error Variance (within group variance)

Logic Behind Anova And Its Assumptions

The first step in ANOVA is to determine the total variability for the entire set of data. (one general measure of variability for the complete experiment).

The word analysis means dividing into smaller parts. Because we are going to analyze variability, the process is called analysis of variance.

The analysis process divides the total variability into two basic components.

1. Between-Treatments Variance

This simply measures how much difference exists between the treatment conditions.

There are two possible explanations for these between-treatment differences:

The differences are the result of sampling error.

The differences between treatments have been caused by the treatment effects.

2. ***Within-Treatment Variance***

Inside each treatment condition, a set of individuals receive exactly the same treatment but their scores differ.

The differences that exist within a treatment represent random and unsystematic differences when there are no treatment effects

Thus, the within-treatments variance provides a measure of how big the differences are when H_o is true.

The F-ratio: The Test Statistic For ANOVA

Once the total variability is analyzed into two basic components (between treatments and within treatments), we simply compare them.

The comparison is made by computing an F-ratio as:

F=Variance between treatments/Variance wit*h*in treatments

For ANOVA, the denominator of the F-ratio is called the error term. The error term provides a measure of the variance caused by random, unsystematic differences. Because ANOVA typically is used to examine data from more than two treatment conditions (and more than two samples), we need a notational system to keep track of all the individual scores and totals.

Assumptions For The ANOVA

- The observations within each sample must be independent.
- The populations from which the samples are selected must be normal.
- The populations from which the samples are selected must have equal variances (homogeneity of variance).
- IV is always categorical and DV is continues
- The assumption of homogeneity of variance is an important one.
- If a researcher suspects that it has been violated, it can be tested by Welch and Brown-Forsythe tests.
- If one of the assumptions for the independent-measures ANOVA has been violated, an alternative statistical analysis known as the Kruskal-Wallis test can be used.

Calculating ANOVA by Hand

In calculating ANOVA, the letter *k* is used to identify the number of treatment conditions, that is, the number of levels of the factor. For an independent-measures study, k also specifies the number of separate samples.The number of scores in each treatment is identified by a lowercase letter *n*.

For the final F-ratio we need an MS (variance) between treatments for the numerator and an MS (variance) within treatments for the denominator. In each case

$F = s^2 between/s^2 within$

F= MS between/MS within

$MS = s^2 = SS/df$

In ANOVA, each of the two variances in the F-ratio is calculated using the basic formula for sample variance.

$s^2 = SS/df$. SS total is the sum of squares for the entire set of N scores and calculated as:

SS total$=\sum X^2-(\sum X)^2/N$

Firstly we will calculate within-treatments sum of squares (SS within treatments). SS within is the variability inside each treatment condition. To find the overall within-treatment sum of squares, we simply add these values together:

SS within treatment= $\sum$ SS inside each treatment

Secondly we will calculate between-treatments sum of squares (SS between treatments). The value for SS between treatments can be found simply by subtraction:

SS between treatments= SStotal - SSwithin

After calculating the sum of squares, let calculate the total degrees of freedom (*df* total). To find the df associated with SS total we compute *df* as:

When all of these individual treatment values are added together, we obtain within-treatments degrees of freedom (df within)

df within= ∑(n-1)= ∑ df in each treatment

OR

df within= N-k

To find df between, simply count the number of treatments and subtract 1.

df between = K - 1

ANOVA Summary Tables

It is useful to organize the results of the analysis in one table called an ANOVA summary table. The table shows the source of variability (between treatments, within treatments, and total variability), SS, df, MS, and F and it is constructed as shown:

Source	SS	df	MS	
Between treatments	30	2	15	F=11.28
Within treatments	16	12	1.33	
Total	46	14		

Effect Size

For ANOVA, the simplest and most direct way to measure effect size is to compute the percentage of variance accounted for by the treatment conditions. We determine how much of the total SS is accounted for by the SS between treatments. The percentage of variance accounted for by the treatment effect is usually called η^2 (the Greek letter eta squared) instead of using r^2.

For this example:

η^2 = SS between treatments /SS total

Post Hoc Tests

Post hoc tests are additional hypothesis tests that are done after an ANOVA to determine exactly which mean differences are significant and which are not. As the name implies, post hoc tests are done after an ANOVA. More specifically, these tests are done after ANOVA when, you reject Ho and there are three or more treatments.

In general, a post hoc test enables us to go back through the data and compare the individual treatments two at a time. In statistical terms, this is called *making pairwise comparisons*. T***ukey's HSD*** test is a most commonly used post hoc test in psychological research. Tukey's test allows to compute a single value that determines the minimum difference between treatment means that is necessary for significance.

Lesson 30

ONE-WAY ANOVA

One-Way ANOVA in SPSS

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups).

For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels among students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an omnibus test statistic and cannot tell you which specific groups were statistically significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important. You can do this using a post hoc test.

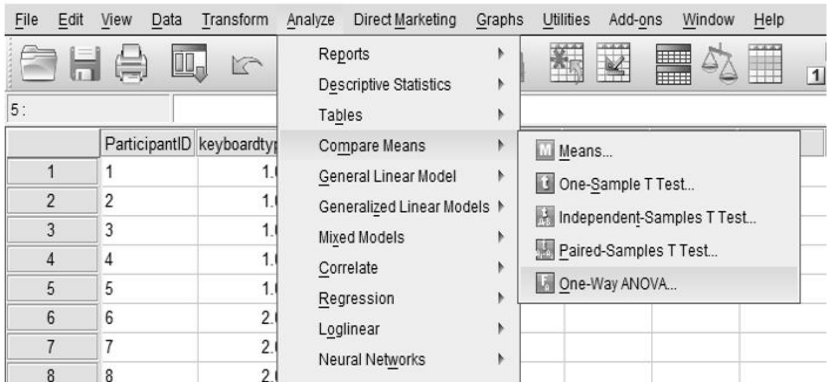
Entering Data in SPSS

Let understand the entering data into SPSS through keyboard example. We separate the groups for analysis by creating a grouping variable called Keyboard type (i.e., the independent variable. Number of errors committed is entered under the variable name errors (i.e., the dependent variable).

- By clicking in the box of the 'values' field in the row containing the groups variable, we get a pop-up dialog that allows us to code the categorical variable. Three types of keyboard are coded.
- The data is entered in the data view tab of the SPSS editor.
- A code is entered for the categories of independent variable as (1 = keyboard A, 2 = keyboard B, and 3 = keyboard C) and the number of committed errors in errors column.

Running One-Way ANOVA in SPSS

- To run one-way ANOVA in SPSS, Click Analyze > Compare Means > One-Way ANOVA on the top menu.



- The One-Way ANOVA window opens.
- Click on your dependent (continuous) variable (e.g. errors). Move this into the box marked Dependent List by clicking on the arrow button.
- Click on your independent, categorical variable (e.g. keyboard type). Move this into the box labelled Factor.
- Click the Options button and click on Descriptive, Homogeneity of variance test, Brown-Forsythe, Welch and Means Plot.
- For Missing values, make sure there is a dot in the option marked Exclude cases analysis by analysis. Click on Continue.
- Click on the button marked Post Hoc. Click on Tukey.
- Click on Games-Howell under the section Equal Variances Not Assumed
- Click on Continue and then OK.

Assumptions of One-Way ANOVA

ANOVA has the same general assumptions that apply to all of the parametric tests with an additional assumption of homogeneity of variances.

The general assumptions include;

1. Level of measurement (DV measured on interval or ratio scale, i.e., continuous)
2. Random sampling
3. Independence of observations
4. Normal distribution (by assessing skewness, kurtosis, histogram, Q-Q plots, Shapiro-Wilk and Kolmogorov-Smirnov tests)

5. Homogeneity of Variance

The homogeneity of variance option gives us the ***Levene's test*** for homogeneity of variances, which tests whether the variance in scores is the same for each of the three groups.

Check the significance value (Sig.) for Levene's test. If this number is greater than .05, the assumption is not violated.

If the assumption is found to be violated then the F test of the ANOVA is not robust enough to be used. In this case, there is need to consult the table in the output headed Robust Tests of Equality of Means.

The two tests there ***(Welch and Brown-Forsythe)*** are preferable when the assumption of the homogeneity of variance is violated.

Interpreting the Output

The descriptive table gives us the information about each group (number in each group, means, standard deviation, minimum and maximum, etc.)

The ANOVA table gives both between-groups and within-groups sums of squares, degrees of freedom etc.

The main thing is the Sig. (the p value). If the Sig. value is ***less than or equal to .05,*** there is a significant difference somewhere among the mean scores on dependent variable for the three groups.

ANOVA

typing perfomance in terms of errors

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	2	35.000	9.130	.004
Within Groups	46.000	12	3.833		
Total	116.000	14			

In this example the overall Sig. value is .004, which is less than .05, indicating a statistically significant result somewhere among the groups with an F-ratio of 9.13.

ANOVA

typing perfomance in terms of errors

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	2	35.000	9.130	.004
Within Groups	46.000	12	3.833		
Total	116.000	14			

The post-hoc tests table tells us exactly where the differences among the groups occur.

In the column labelled *Mean Difference* look for any asterisks (*) next to the values listed. If there is an (*), this means that the two groups being compared are significantly different from one another at the p<.05 level.

The exact significance value is given in the column labelled Sig.

There is a significant difference in errors (typing performance) between the groups that used the keyboard A and keyboard B (p =.018), as well as between the keyboard A and keyboard C (p=.004).

However, there were no differences between the groups that used keyboard B and keyboard C (p=.706).

Post Hoc Tests

Multiple Comparisons

Dependent Variable: typing perfomance in terms of errors

	(I) keyboardtype	(J) keyboardtype	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	keyboard A	keyboard B	-4.00000*	1.23828	.018	-7.3036	-.6964
		keyboard C	-5.00000*	1.23828	.004	-8.3036	-1.6964
	keyboard B	keyboard A	4.00000*	1.23828	.018	.6964	7.3036
		keyboard C	-1.00000	1.23828	.706	-4.3036	2.3036
	keyboard C	keyboard A	5.00000*	1.23828	.004	1.6964	8.3036
		keyboard B	1.00000	1.23828	.706	-2.3036	4.3036
Games-Howell	keyboard A	keyboard B	-4.00000*	1.26491	.034	-7.6623	-.3377
		keyboard C	-5.00000*	1.14018	.006	-8.2620	-1.7380
	keyboard B	keyboard A	4.00000*	1.26491	.034	.3377	7.6623
		keyboard C	-1.00000	1.30384	.733	-4.7501	2.7501
	keyboard C	keyboard A	5.00000*	1.14018	.006	1.7380	8.2620
		keyboard B	1.00000	1.30384	.733	-2.7501	4.7501

*. The mean difference is significant at the 0.05 level.

Effect Size

Although SPSS does not generate effect size for this analysis, it is possible to determine the effect size for this result.

The information that is needed to calculate eta squared, is provided in the ANOVA table.

η^2 = SSbetween/SStotal

η^2 = 70/116= 0.603

Reporting Results in APA

A one-way between-groups analysis of variance was conducted to explore the impact of keyboard type on typing performance, as measured by the number of errors committed. Participants were divided into three groups according to their use of keyboard (Group 1: keyboard A; Group 2: keyboard B; Group 3: keyboard C). The analysis of variance indicates that there are significant differences among the three keyboard types, $F(2,12) = 9.13$, $p=.004$, $\eta^2 = 0.60$.

Post-hoc comparisons using the Tukey HSD test indicated that the number of errors were statistically significantly higher after using the keyboard B ($M = 5.00$, $SD = 2.24$, $p=.018$) and keyboard C ($M = 6.00$, $SD = 1.87$, $p=.004$) as compared to keyboard A ($M = 1.00$, $SD = 1.73$). There were no statistically significant differences between the keyboard B and keyboard C ($p = .706$).

Lesson 31

TWO-WAY ANOVA-I

Introduction to Two-Way ANOVA

Previously we discussed one way ANOVA which is an experimental design having only one independent variable. In this chapter, we will extend the analysis of variance to cover experimental designs involving two or more independent variables.

In real life, variable rarely exists or relate to other variables in isolation. To examine more complex situations, researchers often design research studies that include more than one independent variable. Researchers systematically change two (or more) variables and then observe how the changes influence another (dependent) variable. When a research study involves more than one factor, it is called a *factorial design*. And when a research study examines ANOVA having exactly two factors, it is called Two-Way ANOVA. Tow-way ANOVA means groups are defined by two independent variables. The two independent variables in a two-factor experiment are identified as factor A and factor B. the goal of the factorial design is to evaluate the mean differences that may be produced by either of the factors acting independently or by the two factors working together. This technique allows us to look at the individual and joint effect of two independent variables on one dependent variable

Main Effects and Interactions

The mean differences among the levels of one factor are referred to as the main effect of that factor. When the design of the research study is represented as a matrix with one factor determining the rows and the second factor determining the columns, then the mean differences among the rows describe the main effect of one factor, and the mean differences among the columns describe the main effect for the second factor.

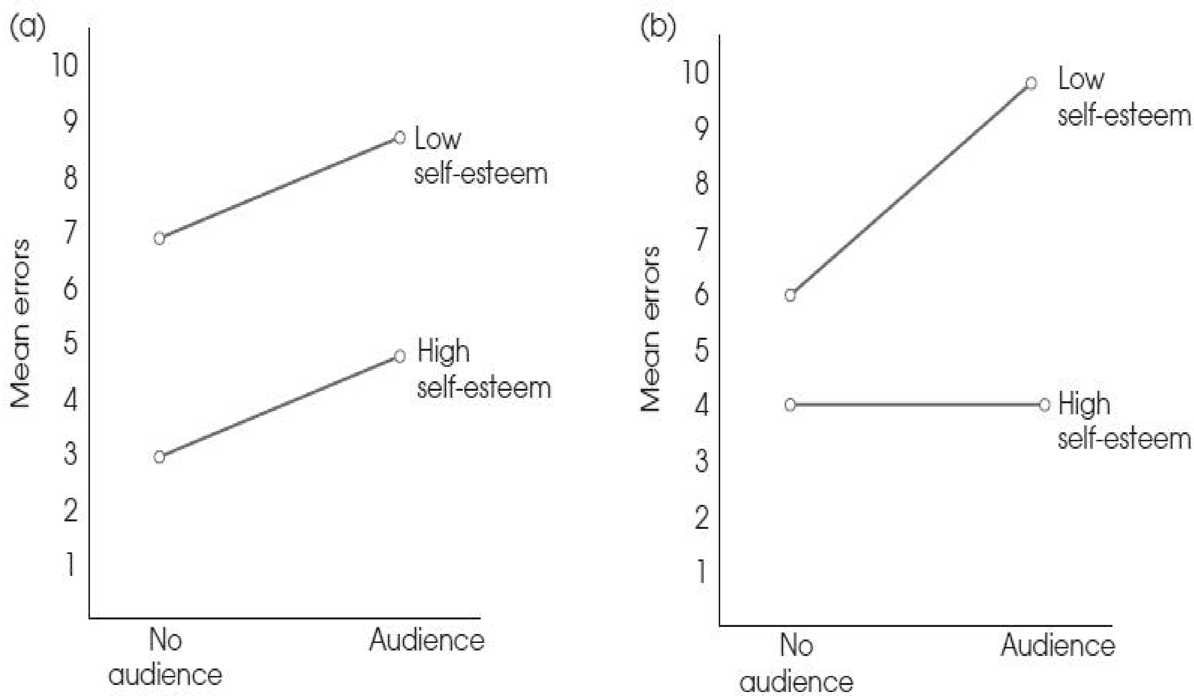
For factorial ANOVA, we took an example of Shrauger's study. Shrauger (1972) tested participants on a concept-formation task. Half of the participants worked alone (no audience), and half worked with an audience of people who claimed to be interested in observing the experiment. Shrauger also divided the participants into two groups on the basis of personality: those high in

Self-esteem and those low in self-esteem. The dependent variable for this experiment was the number of errors on the concept formation task. Data similar to those obtained by Shrauger are

shown in Figure below. Notice that the audience had no effect on the high-self-esteem participants. However, the low-self-esteem participants made nearly twice as many errors with an audience
as when working alone. Note that the study involves two separate factors: One factor is manipulated by the researcher, changing from no-audience to audience, and the second factor is self-esteem, which varies from high to low. The two factors are used to create a *matrix* with the different levels of self-esteem defining the rows and the different audience conditions defining the columns. The resulting two-by-two matrix shows four different combinations of the variables, producing four different conditions. Thus, the research study would require four separate samples, one for each *cell*, or box, in the matrix. The dependent variable for the study is the number of errors on the concept formation task for people observed in each of the four conditions.

	No Audience	Audience	
Low	$M = 7$	$M = 9$	$M = 8$
High	$M = 3$	$M = 5$	$M = 4$
	$M = 5$	$M = 7$	

The real advantage of combining two factors within the same study is the ability to examine not only the main effects but the unique effects caused by an interaction. The concept of interaction can be defined in terms of the pattern displayed in the graph.



To construct this figure, we select one of the factors to be displayed on the horizontal axis; in this case, the different levels of the audience factor. The dependent variable, the number of errors, is shown on the vertical axis. The figure actually contains two separate graphs: The top line shows the relationship between the audience factor and errors for the low-self-esteem and the bottom line shows the relationship for the high-self-esteem participants.

The *A* × *B* interaction typically is called the "*A* by *B*" interaction. If there is an interaction between an audience and self-esteem, it may be called the "audience by self-esteem" interaction.

Hypotheses for Two-Way (ANOVA)

In ANOVA, the variable (independent or quasi-independent) that designates the groups being compared is called a f**actor**. The individual conditions or values that make up a factor are called the **levels** of the factor. It consists of three hypothesis tests, two for main effects and one for interaction.

- H_o: No main effect of A
- H_1: There exists a main effect of factor A
- H_o: No main effect of B
- H_1: There exists a main effect of factor B

- H_o: No interaction
- H_1: There exists an interaction between factor A and factor B

Degrees of Freedom (df)

For two-way ANOVA, we calculate three kinds of degrees of freedom.

1. Total Degrees of Freedom (*df*total)
2. Within-Treatments Degrees of Freedom (*df*within)
3. Between-Treatments Degrees of Freedom (*df*between)

Assumptions for the Independent- Measures ANOVA

The independent-measures ANOVA requires the same three assumptions that were necessary for the independent-measures t hypothesis test:

1. The observations within each sample must be independent.
2. The populations from which the samples are selected must be normal.
3. The populations from which the samples are selected must have equal variances (homogeneity of variance).

Two-Way ANOVA in SPSS

To run two-way ANOVA we need three variables:

- Two categorical independent variables (e.g. sex: males/females; age group: young, middle, old)
- One continuous dependent variable (e.g. optimism).

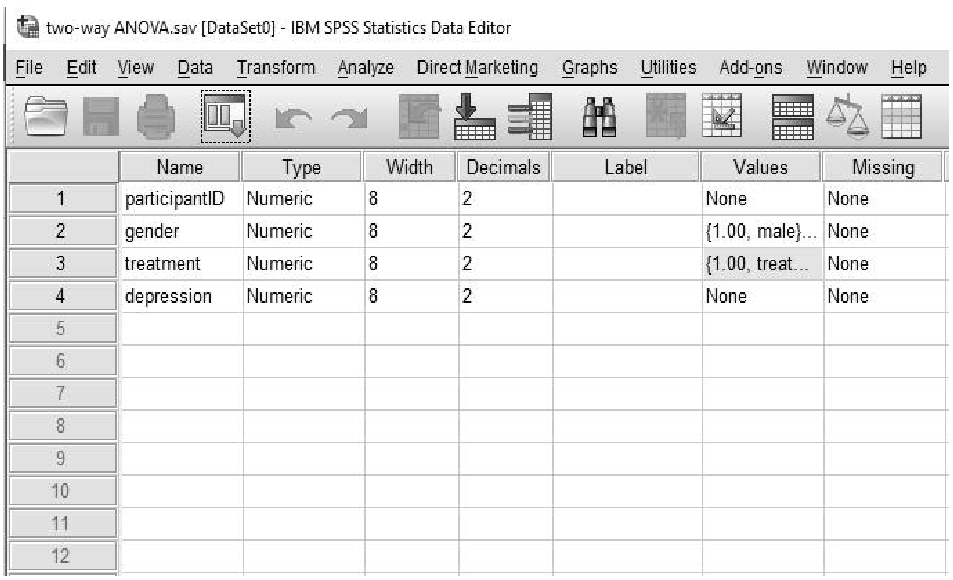
<table>
<tr><td rowspan="8">Factor A
Gender</td><td colspan="3">Factor B
Treatments</td></tr>
<tr><td rowspan="5">Male</td><td>I</td><td>II</td></tr>
<tr><td>3</td><td>2</td></tr>
<tr><td>8</td><td>8</td></tr>
<tr><td>9</td><td>7</td></tr>
<tr><td>4</td><td>7</td></tr>
<tr><td rowspan="2"></td><td>0</td><td>12</td></tr>
<tr><td>0</td><td>6</td></tr>
</table>

	Female	**2**	**9**
		6	**13**

The following data are from a two-factor study examining the depression scores across gender and two treatment conditions. Use an ANOVA with α=.05 for all tests to evaluate the significance of the main effects and the interaction.

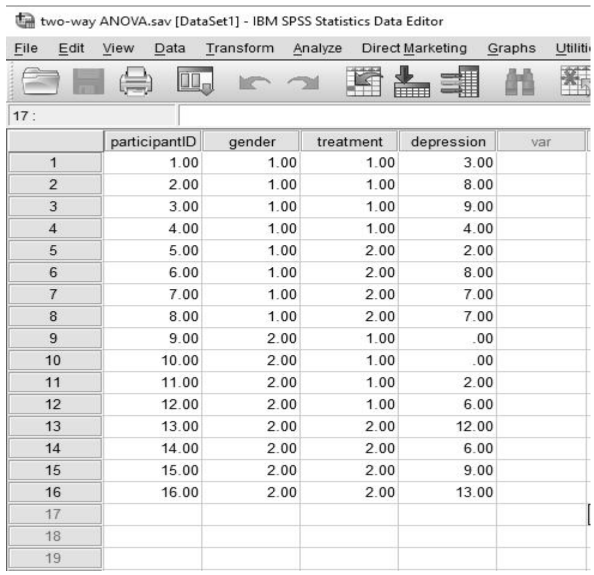
We need to create two grouping variables in the variable view of SPSS as, **treatment and gender** (i.e., the independent variables).

Depression scores are entered under the variable name **depression** (i.e., the dependent variable).



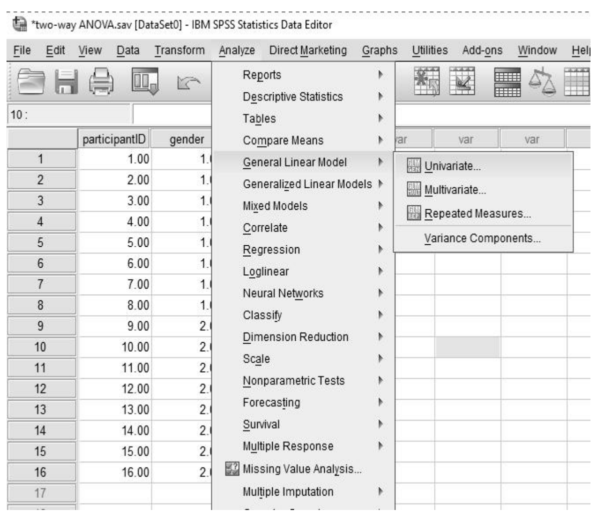
In the box of the 'values' we get a pop-up dialog that allows us to code the categorical variable. Gender and Treatment conditions are coded as shown in the value labels window.

While entering data A code is entered for the categories of independent variables as (**Gender**: 1=male, 2=female; **Treatment**: 1=treatment1, 2=treatment2) and the **depression** scores in the depression column.



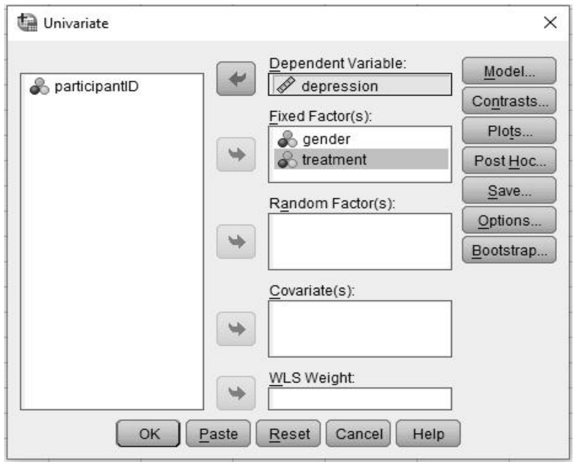
	participantID	gender	treatment	depression	var
1	1.00	1.00	1.00	3.00	
2	2.00	1.00	1.00	8.00	
3	3.00	1.00	1.00	9.00	
4	4.00	1.00	1.00	4.00	
5	5.00	1.00	2.00	2.00	
6	6.00	1.00	2.00	8.00	
7	7.00	1.00	2.00	7.00	
8	8.00	1.00	2.00	7.00	
9	9.00	2.00	1.00	.00	
10	10.00	2.00	1.00	.00	
11	11.00	2.00	1.00	2.00	
12	12.00	2.00	1.00	6.00	
13	13.00	2.00	2.00	12.00	
14	14.00	2.00	2.00	6.00	
15	15.00	2.00	2.00	9.00	
16	16.00	2.00	2.00	13.00	
17					
18					
19					

To run one-way ANOVA in SPSS, from the menu at the top of the screen, click on **Analyze** >**General Linear Model** >**Univariate**.

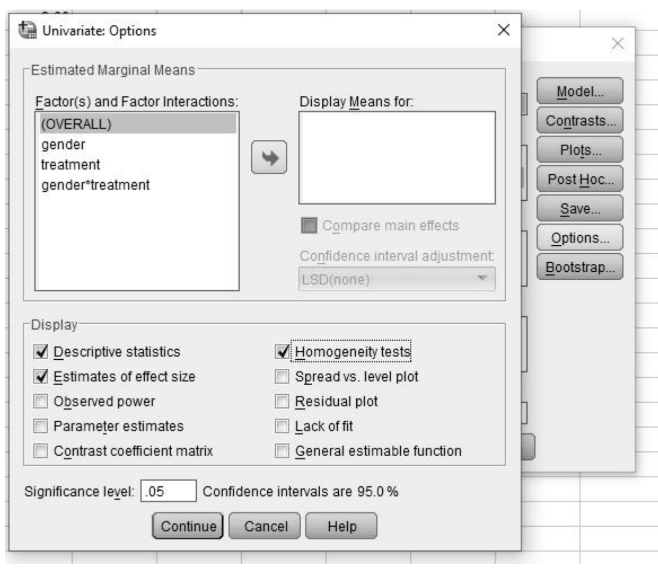


- A **Univariate** window opens.
- Click on the dependent, continuous variable (e.g. depression) and move it into the box labelled **Dependent Variable**.

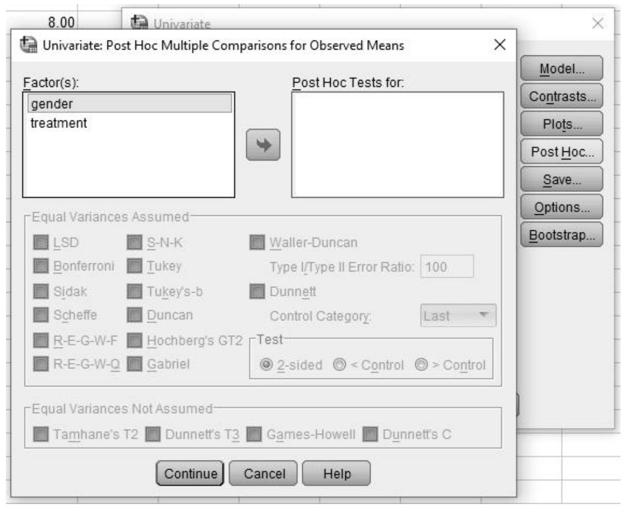
- Click on two independent, categorical variables (gender, treatment) and move these into the box labelled **Fixed Factors**.
-



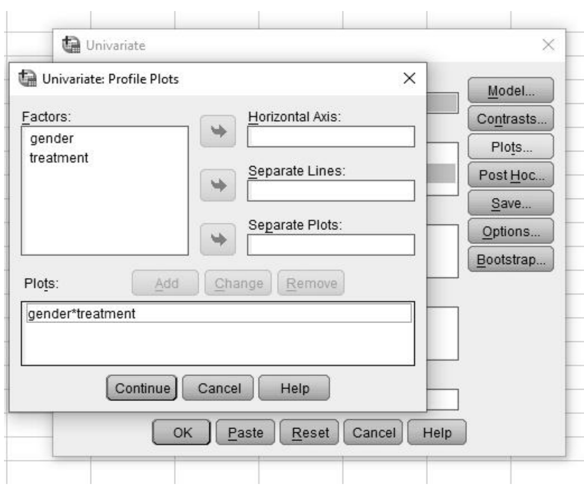
- Click on the **Options** button.\Click on **Descriptive Statistics**, **Estimates of effect size** and **Homogeneity tests** and then **Continue** when finished.



- Click on the **Post Hoc** button.
- From the **Factors** listed on the left-hand side, choose the independent variable(s) you are interested in (only if the variables have more than two levels).
- Click on the arrow button to move it into the **Post Hoc Tests for** section.
- Choose the test **Tukey** and click on **Continue**.



- Click on the **Plots** button.
- In the **Horizontal Axis** box, put the independent variable that has the most groups (e.g. gender).
- In the box labelled **Separate Lines**, put the other independent variable (e.g. treatment).
- Click on **Add**.
- In the section labelled **Plots**, you will see the two variables listed as gender*treatment
- Click on **Continue** and then **OK.**



Assumptions of Two-Way ANOVA

The general assumptions include;

1. Random sampling
2. Independence of observations

3. Normal distribution (by assessing skewness, kurtosis, histogram, Q-Q plots, Shapiro-Wilk and Kolmogorov-Smirnov tests).

Homogeneity of Variance

- **Levene's test for homogeneity of variances** allows us to use the sample variances from data to determine whether there is evidence for any differences among the population variances.
- The **Levene's test for homogeneity of variances** tests whether the variance in scores is the same for each of the groups.
- For the assumption to be fulfilled the **Sig**. value should be *greater* than .05 and therefore *non*-significant results. A significant result (Sig. value less than .05) suggests that the variance of dependent variable across the groups is not equal.

Levene's Test of Equality of Error Variances[a]

Dependent Variable: depression

F	df1	df2	Sig.
.200	3	12	.894

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + gender + treatment + gender * treatment

- The **Descriptive Statistics** table provides the **Mean** scores, **Std. deviations** and **N** for each subgroup.
- The main output from two-way ANOVA is a table labelled Tests of Between-Subjects Effects.
- This table gives us the *main effects* and *interaction effects.*
- The first thing to do is to check for the possibility of an interaction effect (e.g., that the influence of treatments on depression scores depends on whether the individual is male or female).

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- If the interaction effect is significant, the main effects cannot be interpreted easily.
- This is because, in order to describe the influence of one of the independent variables, there is need to specify the level of the other independent variable.
- The output we need to look at is labelled **gender*treatment**.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- In this example, the interaction effect is significant (F= 7.529, p= .018).
- This indicates that there is significant difference in the effect of treatment on depression males and females.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- To determine whether there is a main effect for each independent variable, in the left-hand column, find the variable (e.g., gender) and check in the column marked Sig. next to each variable. If the value is less than or equal to .05 there is a significant main effect for that independent variable.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- In the output shown in the table, there is no significant main effect for gender (p= 1.00) but a significant main effect for treatment (p= .018).
- This means that males and females do not differ in terms of their depression scores, but there is a difference in scores for treatment I and treatment II.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- The effect size for the independent variables is provided in the column labelled **Partial Eta Squared** (e.g., treatment, p= .386).

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000[a]	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

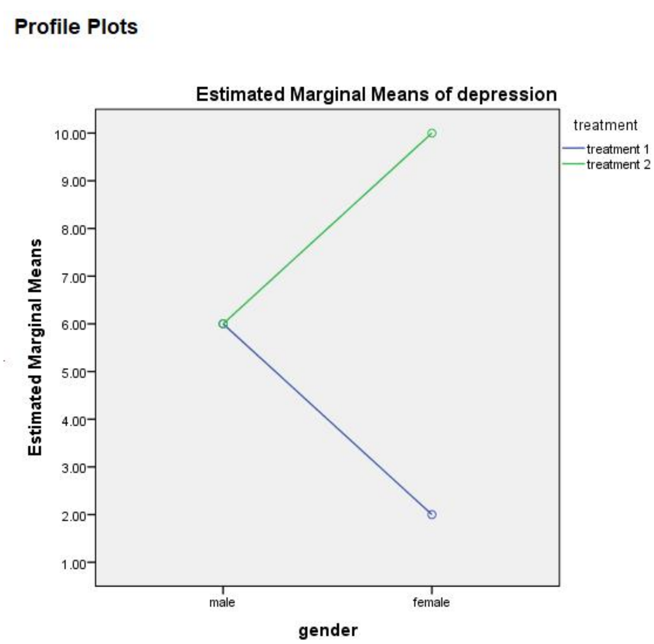
- ANOVA output also provides Post-hoc tests if the post hoc option is checked in the analysis procedure since these are relevant only if we have more than two levels (groups) of an independent variable.
- However, there is no need to look at them until you find a significant main effect or interaction effect in the overall (omnibus) analysis of variance test.

Lesson 32

TWO-WAY ANOVA-II

Interpreting the Plots

The output gives a plot of the depression scores for males and females, across the treatments. This plot is very useful for allowing us to visually inspect the relationship among variables. Although presented last, the plots are often useful to inspect first to help us better understand the impact of two independent variables. The main effect for gender is non-significant and the plot indicates that the mean score of depression for both males and females is same. The main effect for treatment is based on the tendency for the depression scores to increase from treatment I to treatment II as shown in graph.



The plots show the interdependence of both factors. Treatment I is likely to produce a decrease in depression scores for females than males. On the other hand, treatment II influences females more than males. That is, females are more likely to have increase in depression scores after treatment II.

Anova Is Robust Test

Each time we do a hypothesis test, we select an alpha level that determines the risk of a Type I error. With alpha.05, for example, there is a 5%, or a 1-in-20, risk of a Type I error.

Often a single experiment requires several hypothesis tests to evaluate all the mean differences. However, each test has a risk of a Type I error, and the more tests we do, the more risk there is. For this reason, researchers often make a distinction between the *test wise alpha level* and the *experiment wise alpha level*. The test wise alpha level is simply the alpha level that is selected select for each individual hypothesis test. The experiment wise alpha level is the total probability of a Type I error accumulated from all of the separate tests in the experiment.

For example, an experiment involving three treatments would require three separate *t* tests to compare all of the mean differences:

- Test 1 compares treatment I with treatment II.
- Test 2 compares treatment I with treatment III.
- Test 3 compares treatment II with treatment III.

The advantage of ANOVA is that it performs all three comparisons simultaneously in one hypothesis test. Thus, no matter how many different means are being compared, ANOVA uses one test with one alpha level to evaluate the mean differences, and thereby avoids the problem of an inflated experiment wise alpha level. ANOVA or F-test is also robust to moderate departures from normality when sample sizes are reasonably large and are equal.

The ANOVA, therefore, can tolerate data that is non-normal (skewed or kurtotic distributions) with only a small effect on the Type I error rate.

Reporting Results

A two-way between-groups analysis of variance was conducted to explore the impact of gender and treatment on depression. The interaction effect between gender and treatment was statistically significant, $\boldsymbol{F}$ **(1, 12) = 7.53,** $\boldsymbol{p}$ **< .05**. There was a statistically significant main effect for treatment (Treatment I, M=4, SD=3.42; Treatment II, M=8, SD=3.46), F (1, 12) = 7.53, $p < .05$, $\eta^2 = 0.37$. The main effect for gender was not significant (Male, M=6, SD=2.61; Female, M=6, SD=5.09), F (1, 12) =.00, p =1.00.

Lesson 33

CORRELATION ANALYSIS-I

Correlation is a statistical technique that is used to measure and describe the relationship between two variables. Usually the two variables are simply observed as they exist naturally in the environment—there is no attempt to control or manipulate the variables. For example, a researcher could check high school records (with permission) to obtain a measure of each student's academic performance, and then survey each family to obtain a measure of income. The resulting data could be used to determine whether there is relationship between high school grades and family income. Notice that the researcher is not manipulating any student's grade or any family's income, but is simply observing what occurs naturally.

You also should notice that a correlation requires two scores for each individual (one score from each of the two variables). These scores normally are identified as *X* and *Y*. The pairs of scores can be listed in a table, or they can be presented graphically in a scatter plot (Figure 15.2). In the scatter plot, the values for the *X* variable are listed on the horizontal axis and the *Y* values are listed on the vertical axis. Each individual is then represented by a single point in the graph so that the horizontal position corresponds to the individual's *X* value and the vertical position corresponds to the *Y* value. The value of a scatter plot is that it allows you to see any patterns or trends that exist in the data.

Person	Family Income (in $1000)	Student's Average Grade
A	31	72
B	38	86
C	42	81
D	44	78
E	49	85
F	56	80
G	58	91
H	65	89
I	70	94
J	90	83
K	92	90
L	106	97
M	135	89
N	174	95

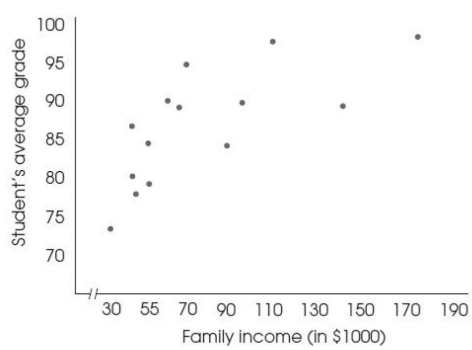


FIGURE 15.2

Correlational data showing the relationship between family income (X) and student grades (Y) for a sample of $n = 14$ high school students. The scores are listed in order from lowest to highest family income and are shown in a scatter plot.

Applications of Correlation

Although correlations have a number of different applications, a few specific examples include:

1. **Correlation is used for prediction:** In general, the squared correlation (r^2) measures the gain in accuracy that is obtained from using the correlation for prediction. The squared correlation measures the proportion of variability in the data that is explained by the relationship between *X* and *Y*. It is sometimes called the ***coefficient of determination***.
2. Correlation is used to demonstrate the **validity** of the test. Validity refers to whether or not the test measures what it claims to measure.
3. Correlations are also used to determine **reliability**. A measurement procedure is considered reliable to the extent that it produces stable, consistent measurements.

Furthermore, the prediction of the theory could be tested by determining the correlation between the two variables. The value of a correlation can be affected greatly by the range of scores represented in the data. The value of a correlation can be affected greatly by the range of scores represented in the data. One or two extreme data points, often called outliers, can have a dramatic effect on the value of a correlation.

The Characteristics of a Relationship

A correlation is a numerical value that describes and measures three characteristics of the relationship between *X* and *Y*. These three characteristics are as follows:

1. The Direction of the Relationship. The sign of the correlation, positive or negative, describes the direction of the relationship.

In a **positive correlation,** the two variables tend to change in the same direction: As the value of the *X* variable increases from one individual to another, the *Y* variable also tends to increase; when the *X* variable decreases, the *Y* variable also decreases.

In a **negative correlation,** the two variables tend to go in opposite directions. As the *X* variable increases, the *Y* variable decreases. That is, it is an inverse relationship.

Suppose you run the drink concession at the football stadium. After several seasons, you begin to notice a relationship between the temperature at game time and the beverages you sell. Specifically, you have noted that when the temperature is low, you sell relatively little beer. However, as the temperature goes up, beer sales also go up (Figure 15.3). This is an example of a positive correlation. You also have noted a relationship between temperature and coffee sales:

On cold days you sell a lot of coffee, but coffee sales go down as the temperature goes up. This is an example of a negative relationship.

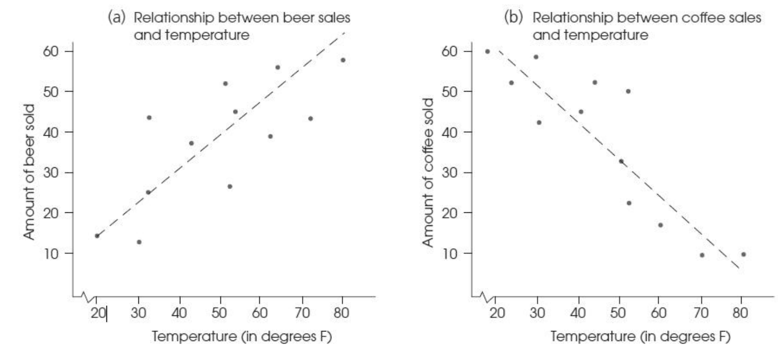


FIGURE 15.3

Examples of positive and negative relationships. (a) Beer sales are positively related to temperature. (b) Coffee sales are negatively related to temperature.

2. The Form of the Relationship. In the preceding coffee and beer examples, the relationships tend to have a linear form; that is, the points in the scatter plot tend to cluster around a straight line. We have drawn a line through the middle of the data points in each figure to help show the relationship. The most common use of correlation is to measure straight-line relationships. However, other forms of relationships do exist and there are special correlations used to measure them.

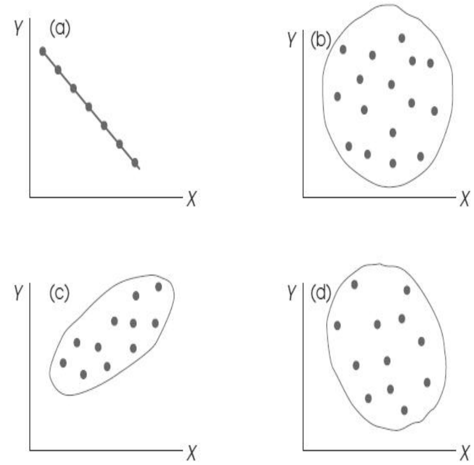
3. The Strength or Consistency of the Relationship. Finally, the correlation measures the consistency of the relationship. For a linear relationship, for example, the data points could fit perfectly on a straight line. Every time X increases by one point, the value of Y also changes by a consistent and predictable amount.

Figure 15.4(a) shows an example of a perfect linear relationship. However, relationships are usually not perfect. Although there may be a tendency for the value of Y to increase whenever X increases, the amount that Y changes is not always the same, and occasionally, Y decreases when X increases. In this situation, the data points do not fall perfectly on a straight line. The consistency of the relationship is measured by the numerical value of the correlation. A *perfect correlation* always is identified by a correlation of 1.00 and indicates a perfectly consistent relationship. For a correlation of 1.00 (or –1.00), each change in X is accompanied by a perfectly predictable change in Y. At the other extreme, a correlation of 0 indicates no consistency at all.

For a correlation of 0, the data points are scattered randomly with no clear trend [see Figure 15.4(b)]. Intermediate values between 0 and 1 indicate the degree of consistency.

FIGURE 15.4

Examples of different values for linear correlations: (a) a perfect negative correlation, −1.00; (b) no linear trend, 0.00; (c) a strong positive relationship, approximately +0.90; (d) a relatively weak negative correlation, approximately −0.40.



- At the other extreme, a correlation of 0 indicates no consistency at all. For a correlation of 0, the data points are scattered randomly with no clear trend. The sign (+ or -) and the strength of a correlation are independent. A correlation of 1.00 indicates a **perfectly consistent relationship** whether it is positive (+1.00) or negative (-1.00). Similarly, correlations of +0.80 and -0.80 are **equally consistent relationships**. One of the most common errors in interpreting correlations is to assume that a correlation necessarily implies a cause-and-effect relationship between the two variables.

Correlation and Causation

Although there may be a causal relationship between two variables, the simple existence of a correlation does not prove it. To establish a cause-and-effect relationship, it is necessary to conduct a true experiment in which one variable is manipulated by a researcher and other variables are rigorously controlled. Although there may be a causal relationship between two variables, the simple existence of a correlation does not prove it. To establish a cause-and-effect relationship, it is necessary to conduct a true experiment in which one variable is manipulated by a researcher and other variables are rigorously controlled. A study shows a positive correlation between number or churches and number of crimes in towns and cities. It is reasonable that small towns would have less crime and fewer churches and that large cities would have large values for both variables. Although a strong correlation exists between number of churches and crime, **the real cause of the relationship is the size of the population.**

Lesson 34

CORRELATION ANALYSIS-II

The Pearson Correlation

By far the most common correlation is the *Pearson correlation* (or the Pearson product– moment correlation) which measures the degree of straight-line relationship. The **Pearson correlation** measures the degree and the direction of the linear relationship between two variables. The Pearson correlation is identified by the letter ***r***.

Conceptually, this correlation is computed by:

$$r = \frac{\text{degree to which } X \text{ and } Y \text{ vary together}}{\text{degree to which } X \text{ and } Y \text{ vary separately}}$$
$$= \frac{\text{covariability of } X \text{ and } Y}{\text{variability of } X \text{ and } Y \text{ separately}}$$

When there is a perfect linear relationship, every change in the *X* variable is accompanied by a corresponding change in the *Y* variable. In this case, the covariability (*X* and *Y* together) is identical to the variability of *X* and *Y* separately, and the formula produces a correlation with a magnitude of +1.00 or –1.00. When there is no linear relationship, a change in the *X* variable does not correspond to any predictable change in the *Y* variable. In this case, there is no covariability, and the resulting correlation is zero.

l

EXAMPLE 15.2 The same set of $n = 4$ pairs of scores is used to calculate *SP*, first using the definitional formula and then using the computational formula.

For the definitional formula, you need deviation scores for each of the *X* values and each of the *Y* values. Note that the mean for the *X*s is $M_X = 3$ and the mean for the *Y*s is $M_Y = 5$. The deviations and the products of deviations are shown in the following table:

Caution: The signs (+ and –) are critical in determining the sum of products, *SP*.

Scores		Deviations		Products
X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)(Y - M_Y)$
1	3	−2	−2	+4
2	6	−1	+1	−1
4	4	+1	−1	−1
5	7	+2	+2	+4
				+6 = *SP*

For these scores, the sum of the products of the deviations is *SP*=6. For the computational formula, you need the *X* value, the *Y* value, and the *XY* product for each individual. Then you find the sum of the *X*s, the sum of the *Y*s, and the sum of the *XY* products. These values are as follows:

X	*Y*	*XY*	
1	3	3	
2	6	12	
4	4	16	
5	7	35	
12	20	66	Totals

Substituting the totals in the formula gives

$$SP = \Sigma XY - \frac{\Sigma X \Sigma Y}{n}$$

$$= 66 - \frac{12(20)}{4}$$

$$= 66–60$$

$$= 6$$

Both formulas produce the same result, $SP = 6$.

As noted earlier, the Pearson correlation consists of a ratio comparing the covariability of *X* and *Y* (the numerator) with the variability of *X* and *Y* separately (the denominator). In the formula for the Pearson *r*, we use *SP* to measure the covariability of *X* and *Y*. The variability of *X* is measured by computing *SS* for the *X* scores and the variability of *Y* is measured by *SS* for the *Y* scores. With these definitions, the formula for the Pearson correlation becomes

$$r = \frac{SP}{\sqrt{SS_X SS_Y}}$$

Lesson 35

CORRELATION ANALYSIS-III

Hypothesis Testing with the Pearson Correlation

The Pearson correlation is generally computed for sample data. A sample correlation is often used to answer questions about the corresponding population correlation. The basic question for this hypothesis test is whether a correlation exists in the population.

The null hypothesis is 'No. There is no correlation in the population,' or 'The population correlation is zero.'

The alternative hypothesis is 'Yes. There is a real, nonzero correlation in the population.'

The population correlation is traditionally represented by **ρ (the Greek letter rho).** Hypotheses would be stated in symbols as,

$$H_o: \rho = 0 \text{ (There is no population correlation)}$$

$$H_1: \rho \neq 0 \text{ (There is real correlation)}$$

When there is a specific prediction about the direction of the correlation, it is possible to do a directional, or one-tailed, test. For example,

$$H_o: \rho \leq 0 \text{ (The population correlation is not positive)}$$

$$H_1: \rho < 0 \text{ (The population correlation is positive)}$$

Samples are not expected to be identical to the populations from which they come; there is some discrepancy (sampling error) between a sample statistic and the corresponding population parameter.

Degrees of freedom

The hypothesis test for the Pearson correlation has degrees of freedom defined by $df = n - 2$. An intuitive explanation for this value is that a sample with only $n = 2$ data points has no degrees of freedom. Specifically, if there are only two points, they will fit perfectly on a straight line, and the sample produces a perfect correlation of $r = +1.00$ or $r = -1.00$. Because the first two points always produce a perfect correlation, the sample correlation is free to vary only when the data set contains more than two points. Thus, $df = n - 2$. The computations for evaluating r have already been completed and are summarized in **Table B.6 in Appendix B.** The table is based on the concept that a sample is expected to be representative of the population from which it was obtained.

	Level of Significance for One-Tailed Test			
	.05	.025	.01	.005
	Level of Significance for Two-Tailed Test			
$df = n - 2$	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641

Correlation in SPSS

The Pearson correlation (r) is designed for **interval level (continuous)** variables. It can also be used if you have one continuous variable (e.g. scores on a measure of self-esteem) and one dichotomous variable (e.g. gender: Male/Female)-that is the point-biserial correlation. To run correlation in SPSS we need two variables: both continuous, or one continuous and the other dichotomous (two values).

Assumptions

- The scale of measurement for the variables should be interval or ratio (continuous).
- One exception to this is if we have one dichotomous independent variable (with only two values e.g., gender) and one continuous dependent variable, we can still run correlation.
- Each subject must provide a score on both variable X and variable Y **(related pairs).**
- Both pieces of information must be from the same subject.
- The observations that make up the data must be **independent** of one another. That is, each observation or measurement must not be influenced by any other observation or measurement.

- The observations that make up the data must be independent of one another. That is, each observation or measurement must not be influenced by any other observation or measurement.
- Scores on each variable should be normally distributed
- The relationship between the two variables should be linear.

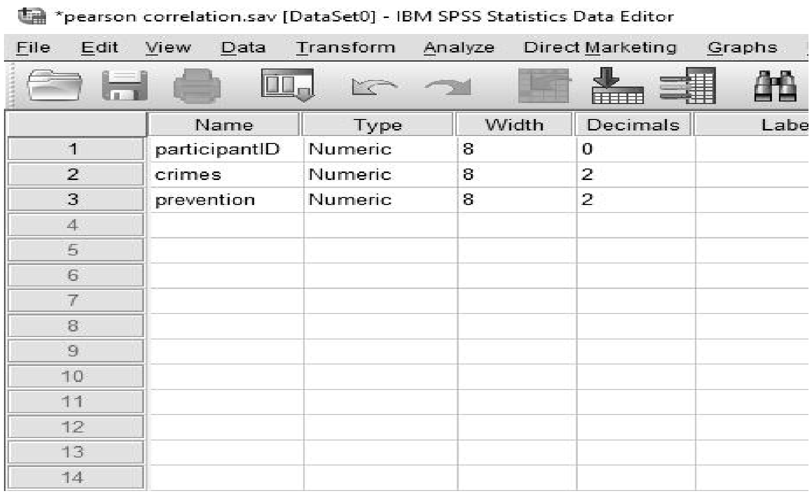
This means that when we look at a scatterplot of scores there should be see a straight line (roughly), not a curve.

For Example, a researcher records the annual number of serious crimes and the amount spent on crime prevention. The data are given in table. Compute the Pearson correlation to measure the degree of relationship between the two variables. Is the correlation statistically significant? Use a two-tailed test with alpha .01.

Number of crimes	Amount spent for prevention
3	6
4	7
6	3
7	4
8	11
9	12
11	8
12	9
13	16
14	17
16	13
17	14

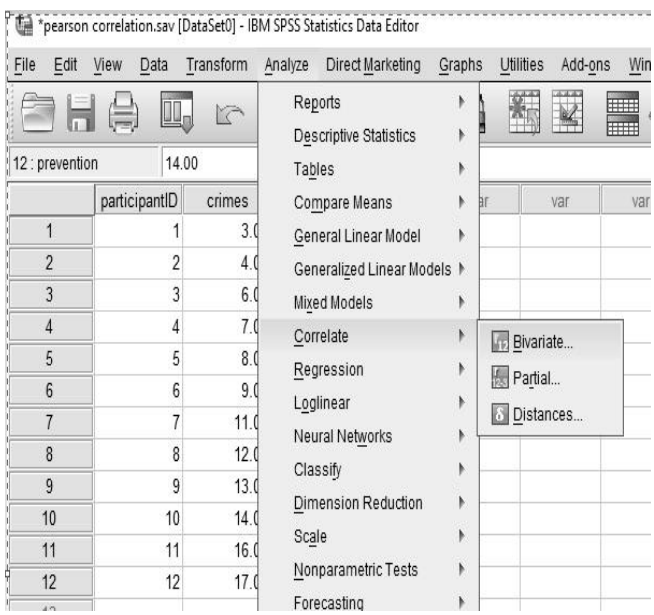
Data Entry

First of all, we need to create two variables on scale measurement namely crime (independent variable) and prevention (dependent variable) in variable view. The data are entered into two columns in the data editor, one for the *X* values (variable crimes) and one for the *Y* values (Variable prevention), with the two scores for each individual in the same row.

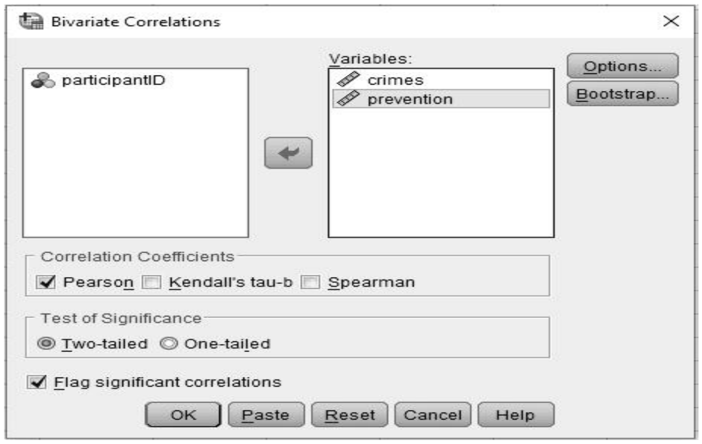


Running Analysis

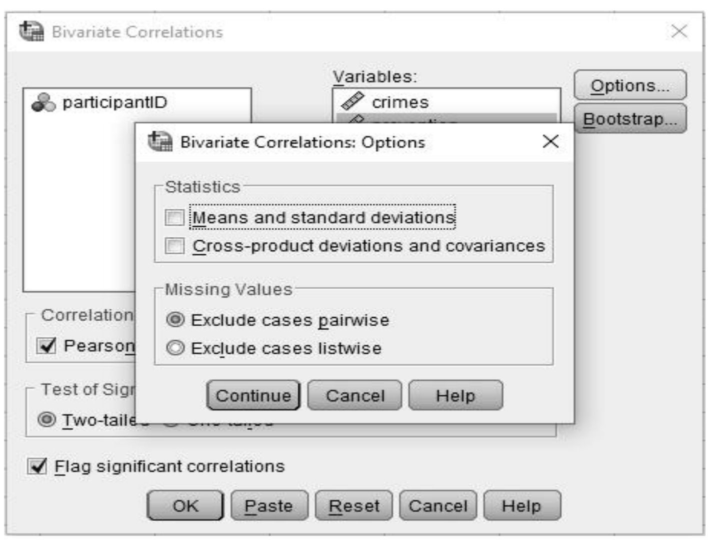
- From the menu at the top of the screen, click on **Analyze**, then select **Correlate**, then **Bivariate**.



- Select two variables and move them into the box marked **Variables** (e.g., crimes and prevention).
- In the **Correlation Coefficients** section, the **Pearson** box is the default option.



- Click on the **Options** button. For **Missing Values**, click on the **Exclude cases pairwise** box.
- Under **Options**, we can also obtain means and standard deviations if we wish.
- Click on **Continue** and then on **OK**



Interpreting Results

SPSS provides a table giving the correlation coefficients between each pair of variables listed, the significance level and the number of cases. In the example given here, the Pearson correlation coefficient (r=.77, p<.01) indicating a significant positive correlation between number of crimes and the amount spent on prevention. The more number of crimes people commit the more time is spent to prevent these crimes.

Correlations

		crimes	prevention
crimes	Pearson Correlation	1	.765**
	Sig. (2-tailed)		.004
	N	12	12
prevention	Pearson Correlation	.765**	1
	Sig. (2-tailed)	.004	
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

The size of the value of the correlation coefficient ranges from –1 to 1. This value indicates the strength of the relationship between your two variables. Cohen (1988) suggests the following guidelines;

Small r=.10 to .29

Medium r=.30 to .49

Large r=.50 to 1.0

The relationship between number of crimes and amount spent on prevention was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions. There was a strong significant positive correlation between the two variables, $r = .77$, $n = 12$ $p < .01$, with more number of crimes associated with more time spent on prevention

Lesson 36

TYPES OF CORRELATION-I

Partial Correlation

Sometimes a researcher may suspect that the relationship between two variables is being distorted by the influence of a third variable. A statistical technique, known as *partial correlation,* allows a researcher to measure the relationship between two variables while eliminating the influence of a third variable.

For example, a study showed a positive correlation between number or churches and number of crimes in towns and cities. However, it was unlikely that there was a direct relationship between churches and crime. Instead, both variables were influenced by population. If population were controlled, there probably would be no real correlation between churches and crime.

In a situation with three variables, *X*, *Y*, and *Z*, it is possible to compute three individual Pearson correlations:

1. *rXY* measuring the correlation between *X* and *Y*
2. *rXZ* measuring the correlation between *X* and *Z*
3. *rYZ* measuring the correlation between *Y* and *Z*

The statistical significance of a partial correlation is determined using the same procedure as is used to evaluate a regular Pearson correlation. For a partial correlation, however, we must use $df=n-3$ instead of the $n-2$ value that is used for the Pearson correlation. A significant correlation means that it is very unlikely ($p<\alpha$) that the sample correlation would occur without a corresponding relationship in the population.

The Spearman Correlation

When the Pearson correlation formula is used with **data from an ordinal scale (ranks),** the result is called the *Spearman correlation*. The Spearman correlation is used in two situations. It is identified by the symbol ***rS*** to differentiate it from the Pearson correlation. The Spearman correlation is used to measure the relationship between *X* and *Y* when both variables are measured on ordinal scales. Additionally, it can be used as a valuable alternative to the Pearson correlation, even when the original raw scores are on an interval or a ratio scale.

As we have noted, the Pearson correlation measures the degree of *linear relationship* between two variables—that is, how well the data points fit on a straight line. However, a researcher often expects the data to show a consistently one-directional relationship but not necessarily a linear relationship. If the Pearson correlation were computed for these data, it would not produce a correlation of 1.00 because the data do not fit perfectly on a straight line. In a situation like this, the Spearman correlation can be used to measure the consistency of the relationship, independent of its form.

The Spearman correlation measures consistency, rather than form. When two variables are consistently related, their ranks are linearly related. For example, a perfectly consistent positive relationship means that every time the *X* variable increases, the *Y* variable also increases. When there is a consistently one-directional relationship between two variables, the relationship is said to be *monotonic*. Thus, the Spearman correlation measures the degree of monotonic relationship between two variables.

The Point-Biserial Correlation

The point-Biserial correlation is used to measure the relationship between two variables in situations in which one variable consists of regular, numerical scores, But the second variable has only two values. A variable with only two values is called a *dichotomous variable* or a *binomial variable*. **For example,** Male versus female, College graduate versus not a college graduate, Older than 30 years versus younger than 30 years.

To compute the point-biserial correlation, the dichotomous variable is first converted to numerical values by assigning a value of zero (0) to one category and a value of one (1) to the other category. Then the regular Pearson correlation formula is used with the converted data. Squaring the value of the point-biserial correlation produces r^2 which is exactly the value of r^2 we obtain measuring effect size of independent sample t-test. As with most correlations, the *strength* of the relationship is best described by the value of r^2. r^2 is the coefficient of determination, which measures how much of the variability in one variable is predicted or determined by the association with the second variable.

The Phi-Coefficient

When both variables (*X* and *Y*) measured for each individual are dichotomous, the correlation between the two variables is called the *phi-coefficient.* It is represented by the symbol (ϕ). A two-step procedure is followed:

- Convert each of the dichotomous variables to numerical values by assigning a 0 to one category and a 1 to the other category for each of the variables.

- Use the regular Pearson formula with the converted scores.
- The phi-coefficient can be used to assess the relationship between two dichotomous variables.
- The more common statistical procedure is a chi-square statistic.

Lesson 37

INTRODUCTION TO REGRESSION

LINEAR EQUATIONS

In general, a *linear relationship* between two variables *X* and *Y* can be expressed by the Equation

$$Y = bX + a \quad (16.1)$$

where *a* and *b* are fixed constants.

For example, a local video store charges a membership fee of $5 per month, which allows you to rent videos and games for $2 each. With this information, the total cost for 1 month can be computed using a *linear equation* that describes the relationship between the total cost (*Y*) and the number of videos and games rented (*X*).

$$Y = 2X + 5$$

In the general linear equation, the value of *b* is called the *slope*. The slope determines how much the *Y* variable changes when *X* is increased by 1 point. For the video store example, the slope is *b* =2 and indicates that your total cost increases by $2 for each video you rent. The value of *a* in the general equation is called the Y-*intercept* because it determines the value of *Y* when *X* = 0. (On a graph, the *a* value identifies the point where the line intercepts the *Y*-axis.) For the video store example, *a* = 5; there is a $5 membership charge even if you never rent a video.

Because a straight line can be extremely useful for describing a relationship between two variables, a statistical technique has been developed that provides a standardized method for determining the best-fitting straight line for any set of data. The statistical procedure is *regression,* and the resulting straight line is called the *regression line.*

The goal for regression is to find the best-fitting straight line for a set of data. To accomplish this goal, however, it is first necessary to define precisely what is meant by "best fit." For any particular set of data, it is possible to draw lots of different straight lines that all appear to pass through the center of the data points. Each of these lines can be defined by a linear equation of the form

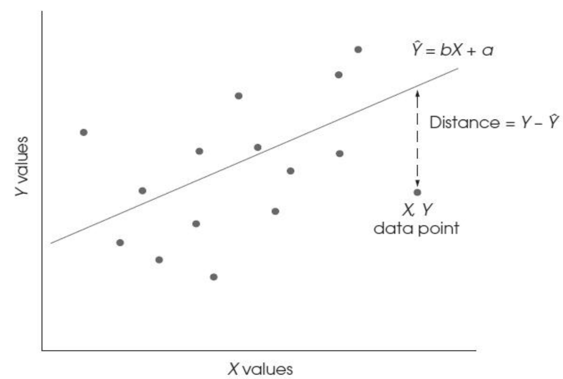
$$\boldsymbol{Y = bX + a}$$

Where *b* and *a* are constants that determine the slope and *Y*-intercept of the line, respectively. Each individual line has its own unique values for *b* and *a*. The problem is to find the specific line that provides the best fit to the actual data points.

The Least-Squares Solution

To determine how well a line fits the data points, the first step is to define mathematically the distance between the line and each data point. For every *X* value in the data, the linear equation determines a *Y* value on the line. This value is the predicted *Y* and is called $\hat{Y}$ ("*Y* hat"). The distance between this predicted value and the actual *Y* value in the data is determined by

$$\textbf{distance} = Y - \hat{Y}$$



Because some of these distances are positive and some are negative, the next step is to square each distance to obtain a uniformly positive measure of error. Finally, to determine the total error between the line and the data, we add the squared errors for all of the data points. The result is a measure of overall squared error between the line and the data:

$$\text{total squared error} = \Sigma(Y - \hat{Y})^2$$

Now we can define the *best-fitting* line as the one that has the smallest total squared error. For obvious reasons, the resulting line is commonly called the *least-squared-error solution*. In symbols, we are looking for a linear equation of the form

$$\hat{Y} = bX + a$$

For each value of X in the data, this equation determines the point on the line ($\hat{Y}$) *that gives the best prediction of* Y. The problem is to find the specific values for a and b that make this the best-fitting line.

The calculations that are needed to find this equation require calculus and some sophisticated algebra, so we do not present the details of the solution. The results, however, are relatively straightforward, and the solutions for b and a are as follows:

$$b = \frac{SP}{SS_X} \qquad (16.2)$$

where SP is the sum of products and SS_X is the sum of squares for the X scores.

A commonly used alternative formula for the slope is based on the standard deviations for X and Y. The alternative formula is

$$b = r\frac{s_Y}{s_X} \qquad (16.3)$$

where s_Y is the standard deviation for the Y scores, s_X is the standard deviation for the X scores, and r is the Pearson correlation for X and Y. The value of the constant a in the equation is determined by

$$a = M_Y - bM_X \qquad (16.4)$$

Note that these formulas determine the linear equation that provides the best prediction of *Y* values. This equation is called the *regression equation for* Y. The **regression equation for *Y*** is the linear equation

$$\hat{Y} = bX + a$$

For example, the scores in the following table are used to demonstrate the calculation and use of the regression equation for predicting *Y*.

X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)^2$	$(Y - M_X)^2$	$(X - M_X)(Y - M_Y)$
2	3	−2	−5	4	25	10
6	11	2	3	4	9	6
0	6	−4	−2	16	4	8
4	6	0	−2	0	4	0
7	12	3	4	9	16	12
5	7	1	−1	1	1	−1
5	10	1	2	1	4	2
3	9	−1	1	1	1	−1
				$SS_X = 36$	$SS_Y = 64$	$SP = 36$

For these data, $\Sigma X = 32$, so $M_X = 4$. Also, $\Sigma Y = 64$, so $M_Y = 8$. These values have been used to compute the deviation scores for each X and Y value. The final three columns show the squared deviations for X and for Y, and the products of the deviation scores.

Our goal is to find the values for b and a in the regression equation. Using Equations 16.2 and 16.4, the solutions for b and a are

$$b = \frac{SP}{SS_X} = \frac{36}{36} = 1.00$$

$$a = M_Y - bM_X = 8 - 1(4) = 4.00$$

The resulting equation is

$$\hat{Y} = X + 4$$

In general, there is some error between the predicted *Y* values and the actual data. Although the amount of error varies from point to point, on average the errors are directly related to the magnitude of the correlation. With a correlation near 1.00 (or 1.00), the data points generally are clustered close to the line and the error is small. As the correlation gets nearer to zero, the points move away from the line and the magnitude of the error increases.

The Standard Error of Estimate

The **standard error of estimate** gives a measure of the standard distance between the predicted *Y* values on the regression line and the actual *Y* values in the data. Conceptually, the standard error of estimate is very much like a standard deviation: Both provide a measure of standard distance. Also, the calculation of the standard error of estimate is very similar to the calculation of standard deviation. To calculate the standard error of estimate, we first find the sum of squared deviations (*SS*). Each deviation measures the distance between the actual *Y* value (from the data) and the predicted *Y* value (from the regression line). This sum of squares is commonly called *SS*residual because it is based on the remaining distance between the actual *Y* scores and the predicted values.

$$SS_{\text{residual}} = \Sigma(Y - \hat{Y})^2 \quad (16.8)$$

The obtained *SS* value is then divided by its degrees of freedom to obtain a measure of variance. This procedure should be very familiar:

$$\text{Variance} = \frac{SS}{df}$$

The degrees of freedom for the standard error of estimate are $df = n - 2$. The reason for having $n - 2$ degrees of freedom, rather than the customary $n - 1$, is that we now are measuring deviations from a line rather than deviations from a mean. To find the equation for the regression line, you must know the means for both the *X* and the *Y* scores. Specifying these two means places two restrictions on the variability of the data, with the result that the scores have only $n - 2$ degrees of freedom. (*Note*: the $df = n - 2$ for SS_{residual} is the same $df = n - 2$ that we encountered when testing the significance of the Pearson correlation on page 529.)

The final step in the calculation of the standard error of estimate is to take the square root of the variance to obtain a measure of standard distance. The final equation is

$$\text{standard error of estimate} = \sqrt{\frac{SS_{\text{residual}}}{df}} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - 2}} \quad (16.9)$$

Assumptions of Regression

- Data should be linear.
- For any two observations the residual terms should be uncorrelated (i.e., independent). This eventuality is sometimes described as a lack of autocorrelation.
- The assumption of autocorrelation can be tested with the Durbin–Watson test, which tests for serial correlations between errors. Specifically, it tests whether adjacent residuals are correlated. The test statistic varies between 0 and 4, with a value of 2 meaning that the residuals are uncorrelated.
- At each level of the predictor variable(s), the variance of the residual terms should be constant. This assumption means that the residuals at each level of the predictor(s) should have the same variance (homoscedasticity); when the variances are very unequal there is said to be heteroscedasticity.
- Normally distributed errors: It can be helpful if the residuals in the model are random, normally distributed variables with a mean of 0.

- This assumption means that the differences between the predicted and observed data are most frequently zero or very close to zero, and that differences much greater than zero happen only occasionally.
- All predictor variables must be quantitative (continuous) or categorical (with two categories).
- The outcome variable must be quantitative (continuous).
- No perfect multicollinearity: If the model has more than one predictor then there should be no perfect linear relationship between two or more of the predictors. So, the predictor variables should not correlate too highly.
- To identify multicollinearity we need to scan the correlation matrix for predictor variables that correlate very highly (values of r above 0.80 or 0.90).
- SPSS Statistics also compute the variance inflation factor (VIF), which indicates whether a predictor has a strong linear relationship with the other predictor(s).
- SPSS also gives the tolerance statistic, which is the reciprocal of VIF, (1/VIF).

Some general guidelines for interpreting the VIF include:

1. If the largest VIF is greater than 10 (or the tolerance is below 0.1) then this indicates a serious problem.
2. If the average VIF is substantially greater than 1 then the regression may be biased
3. Tolerance below 0.2 indicates a potential problem

Lesson 38

SIMPLE LINEAR REGRESSION

Multiple regression is a family of techniques that can be used to explore the relationship between one continuous dependent variable and a number of independent variables or predictors. It allows a more sophisticated exploration of the interrelationship among a set of variables. Multiple regression can be used to address a variety of research questions. It can tell you how well a set of variables is able to predict a particular outcome. For example, you may be interested in exploring how well a set of subscales on an intelligence test is able to predict performance on a specific task. Multiple regression will provide you with information about the model as a whole (all subscales) and the relative contribution of each of the variables that make up the model (individual subscales). As an extension of this, multiple regression will allow you to test whether adding a variable (e.g. motivation) contributes to the predictive ability of the model, over and above those variables already included in the model. Multiple regression can also be used to statistically control for an additional variable (or variables) when exploring the predictive ability of the model. Some of the main types of research questions that multiple regression can be used to address are:

- how well a *set of variables* is able to predict a particular outcome
- *which variable* in a set of variables is the best predictor of an outcome
- Whether a particular predictor variable is still able to predict an outcome when the effects of another variable are *controlled for* (e.g. socially desirable responding).

Major Types of Multiple Regression

1. **Multiple linear regression** also called standard or simultaneous regression: All the independent (or predictor) variables are entered into the model simultaneously. Each independent variable is evaluated in terms of its predictive power, over and above that offered by all the other independent variables.
2. **Hierarchical multiple regression** (also called sequential regression): Variables or sets of variables are entered in steps (or blocks), with each independent variable being assessed in terms of what it adds to the prediction of the dependent variable after the previous variables have been controlled for.

3. **Stepwise multiple regression:** The researcher provides a list of independent variables and then allows the program to select which variables it will enter and in which order they go into the equation, based on a set of statistical criteria. There are three different versions of this approach: **forward selection, backward deletion and stepwise regression**.

The **forward selection method** is often used to provide an initial screening of the variables when a large group of variables exists. SPSS selects the variable that has the highest R-Squared. At each step, it selects the predictor variable that increases R-Squared the most. Finally, it stops adding variables when none of the remaining variables are significant.

The **backward deletion procedure** begins with a model that includes all the independent variables. It deletes one variable at a time by determining whether the least significant variable currently in the model can be removed because its p-value is less than the user-specified or default value.

Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, all variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If a non-significant variable is found, it is removed from the model.

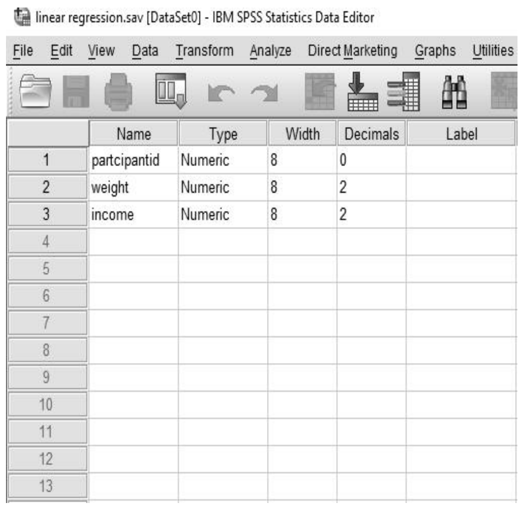
Simple Linear Regression Through SPSS-I

Linear regression is the next step up after correlation. It is used when we want to predict the value of one variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable). The variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable). For simple linear regression dependent variable should be measured at the continuous level (i.e., it is either interval or ratio variable). The independent variable could be categorical (having only two categories) or continuous If the dependent variable is categorical, that is having two levels, the appropriate procedure would be a logistic regression (non-parametric test) instead of linear regression.

For example, for the following set of data for income (DV) and weight (IV). To simplify the weight variable, the women are classified into five categories that measure actual weight relative to height, from 1=thinnest to 5=heaviest. find the linear regression. Does the regression account for a significant portion of the variance? Use alpha level .05 to evaluate the *F*-ratio.

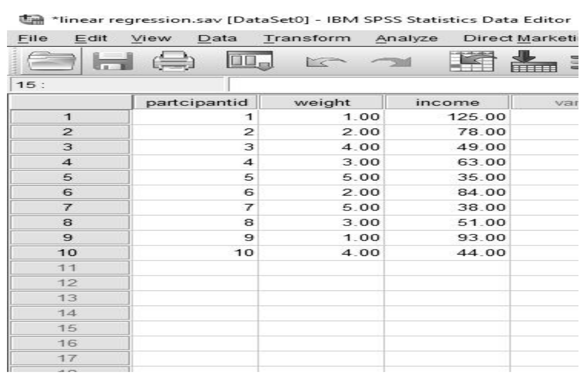
Data Entry

We need to create two variables namely weight (IV) and income (DV), in the SPSS data editor. For predictor variable (weight), enter the values in one column and for outcome variable (income), enter the values in a second column of the SPSS data editor as shown.



	Name	Type	Width	Decimals	Label
1	partcipantid	Numeric	8	0	
2	weight	Numeric	8	2	
3	income	Numeric	8	2	

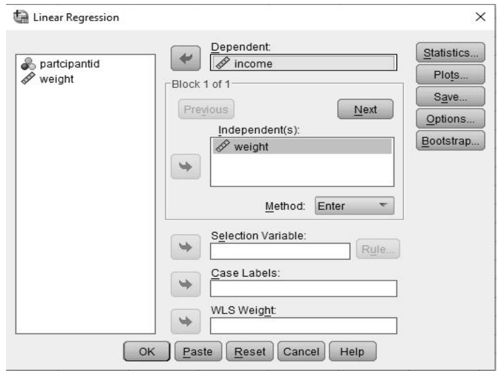
For predictor variable (weight), enter the values in one column and for outcome variable (income), enter the values in a second column of the SPSS data editor as shown.



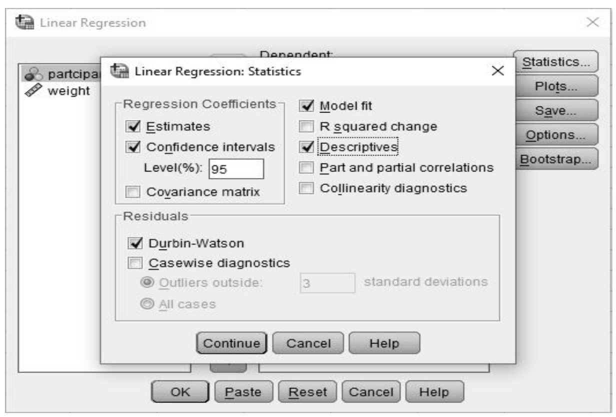
	partcipantid	weight	income
1	1	1.00	125.00
2	2	2.00	78.00
3	3	4.00	49.00
4	4	3.00	63.00
5	5	5.00	35.00
6	6	2.00	84.00
7	7	5.00	38.00
8	8	3.00	51.00
9	9	1.00	93.00
10	10	4.00	44.00

Running Analysis

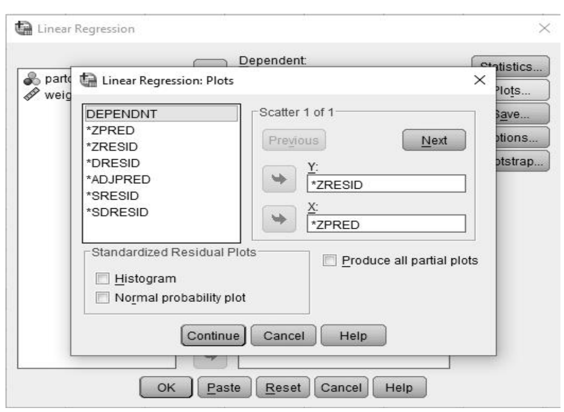
- Click Analyze on the tool bar, select Regression, and click on Linear.
- In the left-hand box, highlight the column label for the outcome variable, then click the arrow to move the column label into the Dependent Variable box.
- For the predictor variable, highlight the column label for the weight and click the arrow to move it into the Independent Variable(s) box.



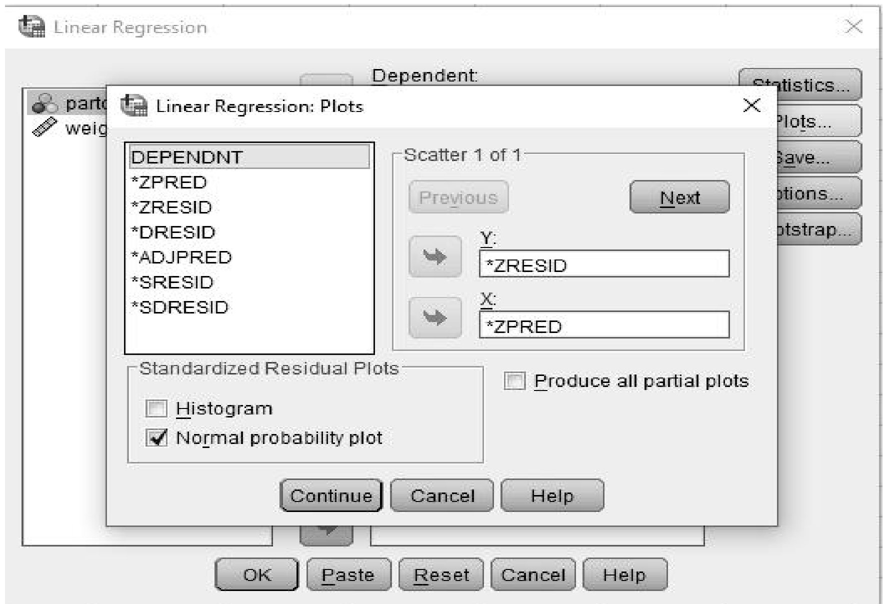
- From the Statistics section, check the Estimates, Confidence intervals, Descriptives and model fit boxes.
- To assess the assumption of independence of observations (no autocorrelation) check the Durbin-Watson box.
- Click Continue.



- In Plots section, move the ZRESID (standardized residuals) and ZPRED (standardized predicted values) in Y and X boxes respectively.
- The scatter plot will assess the homoscedasticity in the data.

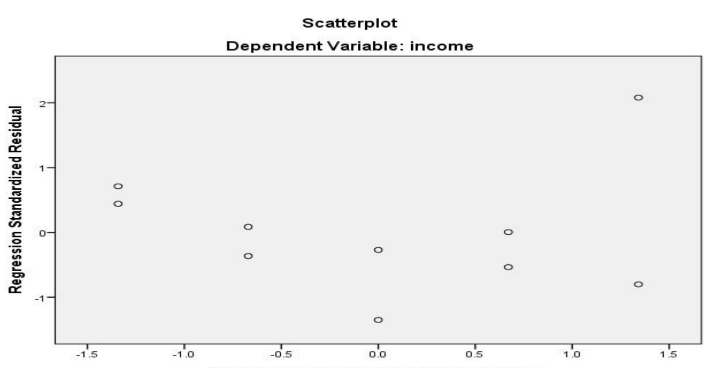


- In the section headed Standardized Residual Plots, tick the Normal probability plot option to assess the normality of the data.
- Click on Continue and **OK**.

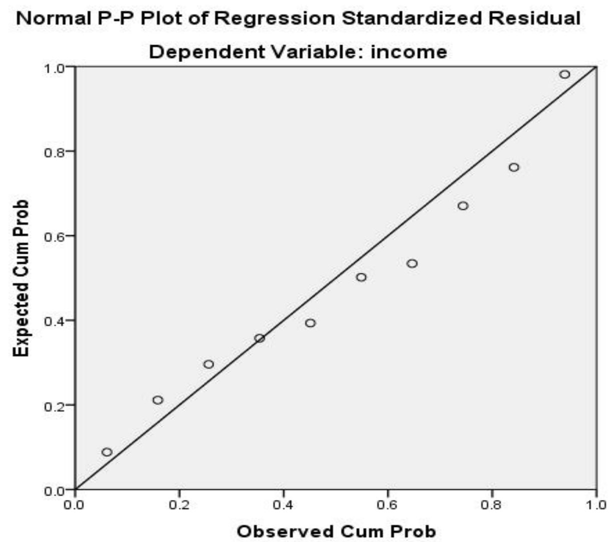


Interpreting Results

In the Scatterplot of the standardized residuals we are hoping that the residuals will be roughly rectangularly distributed, with most of the scores concentrated in the center (along the 0 point). There should not be curvilinear pattern in the data. Deviations from a centralized rectangle suggest some violation of the assumptions.



In the Normal P-P Plot, we check whether the data points lie in a reasonably straight diagonal line from bottom left to top right. This would suggest no major deviations from normality.



Now we need to look at the Model Summary box. This table provides R and R square values. The R value represents the correlation and is (R=.931) which indicates a high correlation. The R square value tells us how much of the variance in the DV (income) is explained by the IV (weight).

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.931[a]	.867	.851	11.10152	1.663

a. Predictors: (Constant), weight

b. Dependent Variable: income

The table also gives the Durbin-Watson value which is 1.6 and is under acceptable range.
This indicates that there is no autocorrelation and the assumption is satisfied.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.931[a]	.867	.851	11.10152	1.663

a. Predictors: (Constant), weight

b. Dependent Variable: income

The next table is the ANOVA table, which reports how well the regression equation fits the data (i.e., predicts the dependent variable). This tests the null hypothesis that R in the population equals 0. The model in this example reaches statistical significance (F=52.29, p<.001).

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6444.050	1	6444.050	52.287	.000[b]
	Residual	985.950	8	123.244		
	Total	7430.000	9			

a. Dependent Variable: income

b. Predictors: (Constant), weight

The Coefficients table provides the necessary information to predict income from weight. The weight is significantly negatively predicting income (B=-17.95, p<.001). This means that for every 1-unit increase in the weight, the income will decrease by 17.95 units.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	119.850	8.233		14.557	.000
	weight	-17.950	2.482	-.931	-7.231	.000

a. Dependent Variable: income

Simple linear regression was run to assess whether weight predicts income in women. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. The total variance explained by the model as a whole was 86.7%, F (1, 8) = 52.28, p < .001. Weight was found to be significant negative predictor of income (B=-17.95, p<.001).

Lesson 39

MULTIPLE LINEAR REGRESSION

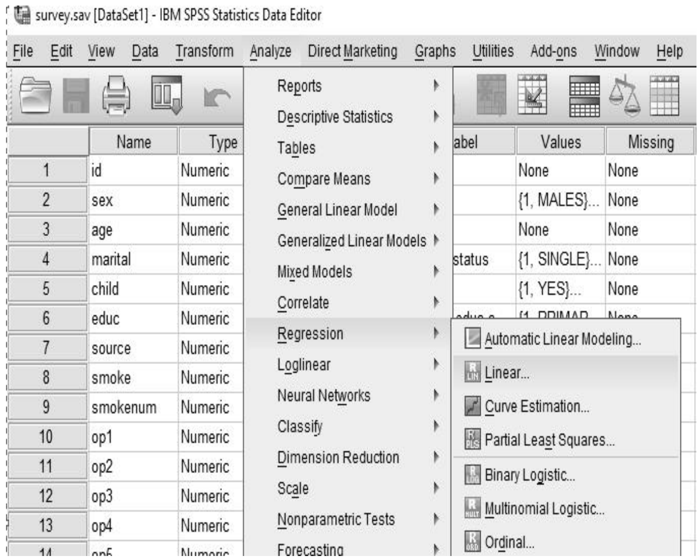
Multiple linear regression tells us how much of the variance in dependent variable can be explained by the independent variables. Tests allow us to determine the statistical significance of the results, in terms of both the model itself and the individual independent variables. Let's take an example of the impact of respondents' perceptions of control on their levels of perceived stress. In this example, we are interested in exploring how well the Mastery Scale and the PCOISS are able to predict scores on a measure of perceived stress. In the questionnaire, the Mastery Scale, measured the degree to which people feel they have control over the events in their lives; and the Perceived Control of Internal States Scale (PCOISS) measured the degree to which people feel they have control over their internal states (their emotions, thoughts and physical reactions).

Multiple Linear Regression in SPSS

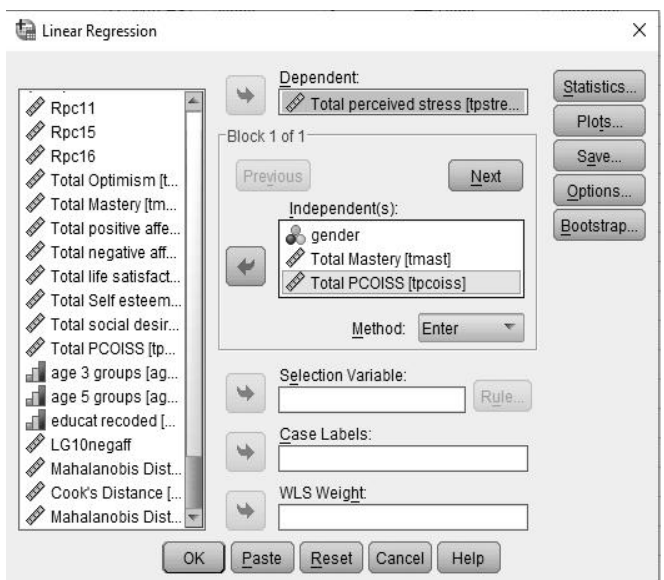
- To run the analysis we need,
- One continuous dependent variable (Total perceived stress)
- Two or more continuous independent variables (mastery, PCOISS)
- We can also use dichotomous independent variables having two categories, (e.g., gender: males=1, females=2).

Running Analysis

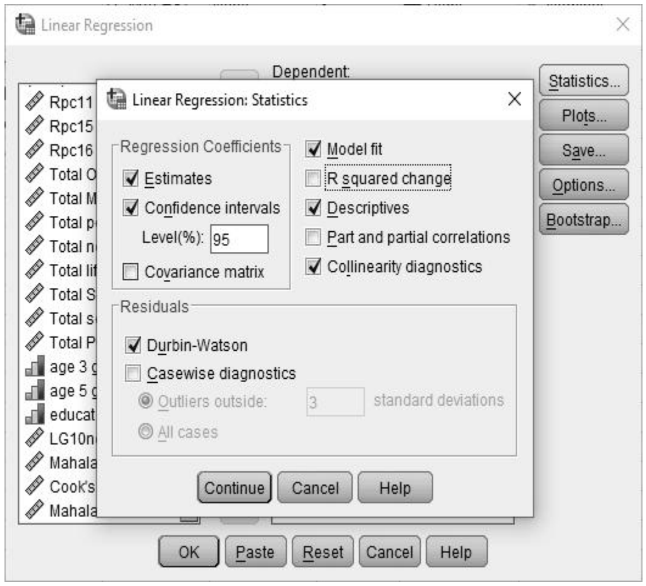
- From the menu at the top of the screen, click on **Analyze**, then select **Regression**, then **Linear**.
- Click on the continuous dependent variable (Total perceived stress: tpstress) and move it into the **Dependent** box.
- Click on the independent variables (Total Mastery: tmast; Total PCOISS: tpcoiss; Gender) and click on the arrow to move them into the **Independent(s)** box.



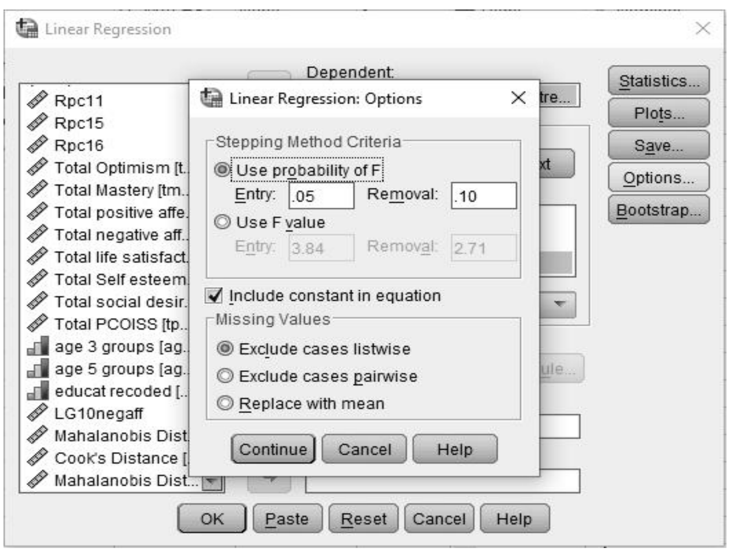
- Click on the continuous dependent variable (Total perceived stress: tpstress) and move it into the **Dependent** box.
- Click on the independent variables (Total Mastery: tmast; Total PCOISS: tpcoiss; Gender) and click on the arrow to move them into the **Independent(s)** box.



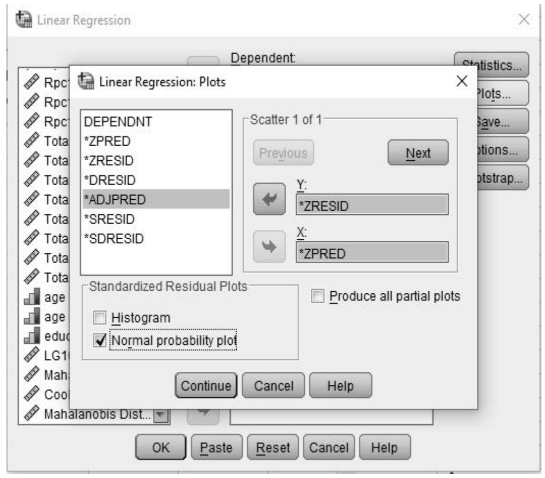
- Click on the **Statistics** button.
- Select the following: **Estimates**, **Confidence Intervals**, **Model fit**, **Descriptives**, **Collinearity diagnostics** and **Durbin-Watson**.



- Click on the **Options** button. In the **Missing Values** section, select **Exclude cases list wise**.
- Click on **Continue**.



- Click on the **Plots** button.
- Click on ***ZRESID** and the arrow button to move this into the **Y** box.
- Click on ***ZPRED** and the arrow button to move this into the **X** box.
- In the section headed **Standardized Residual Plots**, tick the **Normal probability plot** option.
- Click on **Continue** and **OK**.



Assumptions of Multiple Linear Regression

1. **Multicollinearity**

- The correlations between the variables in the model are provided in the table labelled **Correlations**.
- Check that if independent variables show at least some relationship with the dependent variable (above .3 preferably).

Correlations

		Total perceived stress	gender	Total Mastery	Total PCOISS
Pearson Correlation	Total perceived stress	1.000	.147	-.611	-.581
	gender	.147	1.000	-.115	-.102
	Total Mastery	-.611	-.115	1.000	.527
	Total PCOISS	-.581	-.102	.527	1.000
Sig. (1-tailed)	Total perceived stress	.	.001	.000	.000
	gender	.001	.	.009	.018
	Total Mastery	.000	.009	.	.000
	Total PCOISS	.000	.018	.000	.
N	Total perceived stress	426	426	426	426
	gender	426	426	426	426
	Total Mastery	426	426	426	426
	Total PCOISS	426	426	426	426

- The other value given is the **VIF** (Variance inflation factor), which is just the inverse of the Tolerance value (1 divided by Tolerance).
- VIF values above 10 would be a concern here, indicating multicollinearity.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	49.387	1.510		32.704	.000	46.419	52.356		
	gender	.740	.422	.063	1.755	.080	-.089	1.568	.985	1.016
	Total Mastery	-.613	.061	-.417	-9.973	.000	-.734	-.492	.718	1.392
	Total PCOISS	-.173	.020	-.354	-8.484	.000	-.213	-.133	.720	1.388

a. Dependent Variable: Total perceived stress

2. **Independence of Observations**
   - Th **Model summary** table gives the Durbin-Watson value which is 1.91 for this example.
   - This indicates that there is no autocorrelation and the assumption is satisfied.

Model Summary[b]

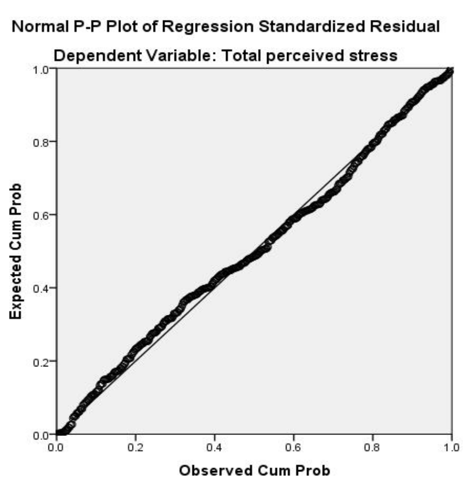
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.685[a]	.470	.466	4.267	1.906

a. Predictors: (Constant), Total PCOISS, gender, Total Mastery

b. Dependent Variable: Total perceived stress

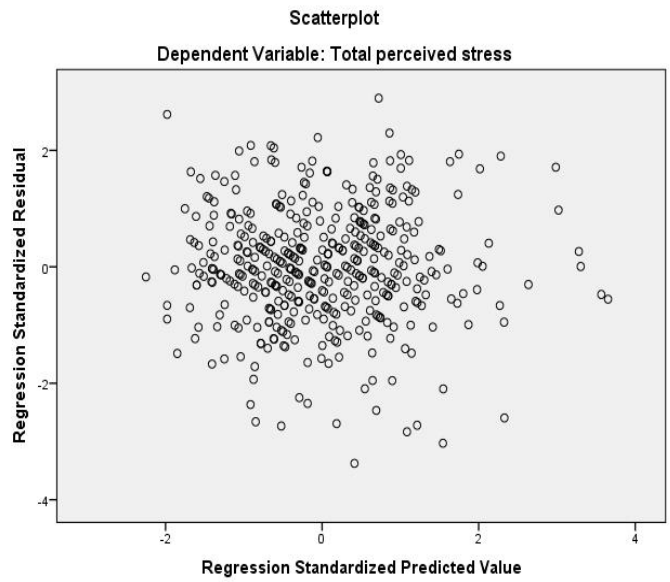
3. **Normality**

- In the **Normal P-P Plot**, the data points lie along the line from bottom left to top right.
- In the current example, the data suggest no deviations from normality, hence, the assumption is fulfilled.



4. **Linearity and Homoscedasticity**

- The **Scatterplot** indicates that the residuals are roughly distributed, with most of the scores concentrated in the center.
- the pattern of residuals in the scatter plot suggests that the assumptions of homoscedasticity and linearity are satisfied.



Interpreting the Results

- First of all, the **Model Summary** box gives the **R Square**.
- This tells you how much of the variance in the dependent variable (perceived stress) is explained by the model (which includes the variables of gender, Total Mastery and Total PCOISS).
- In this case, the value is .468.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.685[a]	.470	.466	4.267	1.906

a. Predictors: (Constant), Total PCOISS, gender, Total Mastery

b. Dependent Variable: Total perceived stress

- To assess the statistical significance of the result, it is necessary to look in the table labelled **ANOVA**.

- The model in this example gives the statistical significance (Sig. = .000 or p <.001).

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6807.615	3	2269.205	124.623	.000[b]
	Residual	7684.009	422	18.209		
	Total	14491.624	425			

a. Dependent Variable: Total perceived stress

b. Predictors: (Constant), Total PCOISS, gender, Total Mastery

- The next thing we want to know is which of the variables included in the model contributed to the prediction of the dependent variable.
- We find this information in the output box labelled **Coefficients**.
- Look in the column labelled **Beta** under **Standardized Coefficients**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	49.387	1.510		32.704	.000	46.419	52.356		
	gender	.740	.422	.063	1.755	.080	-.089	1.568	.985	1.016
	Total Mastery	-.613	.061	-.417	-9.973	.000	-.734	-.492	.718	1.392
	Total PCOISS	-.173	.020	-.354	-8.484	.000	-.213	-.133	.720	1.388

a. Dependent Variable: Total perceived stress

- In this case the largest beta coefficient is –.42 (ignoring the sign), which is for Total Mastery. This means that this variable makes the strongest unique contribution to explaining the dependent variable, when the variance explained by all other variables in the model is controlled for.
- In this case the largest beta coefficient is –.42 (ignoring the sign), which is for Total Mastery. This means that this variable makes the strongest unique contribution to explaining the dependent variable, when the variance explained by all other variables in the model is controlled for.

Lesson 40

NON-PARAMETRIC TEST

Often, researchers are confronted with experimental situations that do not conform to the requirements of parametric tests. When the assumptions of a test are violated, the test may lead to an erroneous interpretation of the data. In these situations, it may not be appropriate to use a parametric test. There are several hypothesis-testing techniques that provide alternatives to parametric tests. These alternatives are called nonparametric tests. Nonparametric tests sometimes are also called *assumption-free tests*. The main reasons to apply the nonparametric test include:

- The underlying data do not meet the assumptions about the population sample
- The population sample size is too small
- The analyzed data is ordinal or nominal
- Nonparametric statistics do not assume that data is drawn from a normal distribution, therefore, called distribution free tests.
- Nonparametric statistics includes nonparametric descriptive statistics, statistical models, inference, and statistical tests.
- The model structure of nonparametric models is not specified *a priori* but is instead determined from data.
- Nonparametric test usually involves qualitative data or Nominal / Ordinal data
- Most of the time even quantitative data is converted to nominal/ordinal data for use with non-parametric tests
- For example, numerical scores measuring self-esteem can be converted to categories like high, medium, and low self-esteem etc.

Chi-Square Test

The name of the test comes from the Greek letter (chi, pronounced "kye"), which is used to identify the test statistic. The most obvious difference between the chi- square tests and the other hypothesis tests we have considered (t and ANOVA) is the nature of the data. For chi- square, the data are in frequencies rather than numerical scores. The chi-square test can be defined as the frequency data from a sample to evaluate the relationship between two variables in the population. Each individual in the sample is classified on both of the two variables, creating a two-dimensional frequency-distribution matrix. The frequency distribution for the sample is then used to test hypotheses about the corresponding frequency distribution for the population. In this situation, each individual in the sample is measured or classified on two separate variables. In general terms, this chi-square test uses the proportions obtained for sample data to test hypotheses about the corresponding proportions in the population

Assumptions of Chi-Square Test

To use a chi-square test for goodness of fit or a test of independence, several conditions must be satisfied. For any statistical test, violation of assumptions and restrictions casts doubt on the results. Some important assumptions and restrictions for using chi-square tests are the following:

1. **Independence of Observations**

   This is not to be confused with the concept of independence between variables, as seen in the chi-square test for independence. One consequence of independent observations is that each observed frequency is generated by a different individual. A chi-square test would be inappropriate if a person could produce responses that can be classified in more than one category or contribute more than one frequency count to a single category.

2. **Size of Expected Frequencies**

   A chi-square test should not be performed when the expected frequency of any cell is less than 5. The chi-square statistic can be distorted when f_e is very small. Consider the chi-square computations for a single cell. Suppose that the cell has values of f_e 1 and *fo* 5. Note that there is a 4-point difference between the observed and expected frequencies.

Chi-square statistic involves two tests:

- The chi-square test for goodness of fit
- The chi-square test for independence

Lesson 41

CHI-SQUARE TEST FOR INDEPENDENCE-I

The chi-square statistic may also be used to test whether there is a relationship between two variables. In this situation, each individual in the sample is measured or classified on two separate variables. For example, a group of students could be classified in terms of personality (introvert, extrovert) and in terms of color preference (red, yellow, green, or blue). Usually, the data from this classification are presented in the form of a matrix, where the rows correspond to the categories of one variable and the columns correspond to the categories of the second variable. For example, in table below a hypothetical data for a sample of n 200 students who have been classified by personality and color preference. The number in each box, or cell, of the matrix indicates the frequency, or number of individuals in that particular group. In Table, for example, there are 10 students who were classified as introverted and who selected red as their preferred color. To obtain these data, the researcher first selects a random sample of n 200 students. Each student is then given a personality test and is asked to select a preferred color from among the four choices.

Note that the classification is based on the measurements for each student; the researcher does not assign students to categories. Also, note that the data consist of frequencies, not scores, from a sample. The goal is to use the frequencies from the sample to test a hypothesis about the population frequency distribution. Specifically, are these data sufficient to conclude that there is a significant relationship between personality and color preference in the population of students?

	Red	Yellow	Green	Blue	
Introvert	10	3	15	22	50
Extrovert	90	17	25	18	150
	100	20	40	40	$n = 200$

The null hypothesis for the chi-square test for independence states that the two variables being measured are independent; that is, for each individual, the value obtained for one variable is not related to (or influenced by) the value for the second variable. This general hypothesis can be expressed in two different conceptual forms, each viewing the data and the test from slightly different perspectives. The null hypothesis for the chi-square test for independence states that the two variables being measured are independent.

***H*₀:** For the general population of students, there is no relationship between color preference and personality.

The alternative hypothesis, *H*₁, states that there is a relationship between the two variables.

Calculating Chi-square test for independence

The chi-square test for independence uses the same basic logic that was used for the goodness-of-fit test. First, a sample is selected and each individual is classified or categorized. Because the test for independence considers two variables, every individual is classified on both variables, and the resulting frequency distribution is presented as a two-dimensional matrix. As before, the frequencies in the sample distribution are called observed frequencies and are identified by the symbol f_o.

The next step is to find the expected frequencies, or f_e values, for this chi-square test. As before, the expected frequencies define an ideal hypothetical distribution that is in perfect agreement with the null hypothesis. Once the expected frequencies are obtained, we compute a chi-square statistic to determine how well the data (observed frequencies) fit the null hypothesis (expected frequencies). Although you can use either version of the null hypothesis to find the expected frequencies, the logic of the process is much easier when you use H0 stated in terms of equal proportions. To find the expected frequencies, we first determine the overall distribution of color preferences and then apply this distribution to both categories of personality. For these data, 100 people selected red as their preferred color. Because the total sample consists of 200 people, the proportion selecting red is 100 out of 200, or 50%. The complete set of color preference proportions is as follows: 100 out of 200 50% preferred 20 out of 200 10% prefer yellow 40 out of 200 20% prefer green 40 out of 200 20% prefer blue. The row totals in the matrix define the two samples of personality types. For example, the matrix in Table shows a total of 50 introverts (the top row) and a sample of 150 extroverts (the bottom row). According to the null hypothesis, both personality groups should have the same proportions for color preferences.

Beginning with the sample of 50 introverts in the top row, we obtain expected frequencies of:

50% preferred: f_e 50% of 50 0.50(50) 25

10% prefer yellow: f_e 10% of 50 0.10(50) 5

20% prefer green: f_e 20% of 50 0.20(50) 10

20% prefer blue: f_e 20% of 50 0.20(50) 10

Using exactly the same proportions for the sample of n 150 extroverts in the bottom row, we obtain expected frequencies of:

50% preferred: f_e 50% of 150 0.50(50) 75

10% prefer yellow: f_e 10% of 150 0.10(50)

15 20% prefer green: f_e 20% of 150 0.20(50) 30

20% prefer blue: f_e 20% of 150 0.20(50) 30

A simple formula for determining expected frequencies

Although expected frequencies are derived directly from the null hypothesis and the sample characteristics, it is not necessary to go through extensive calculations to find f_e values. In fact, there is a simple formula that determines f_e for any cell in the frequency distribution matrix:

$$f_e = fc f_r / n$$

where fc is the frequency total for the column (column total), f_r is the frequency total for the row (row total), and n is the number of individuals in the entire sample. The chi-square test for independence uses exactly the same chi-square formula as the test for goodness of fit:

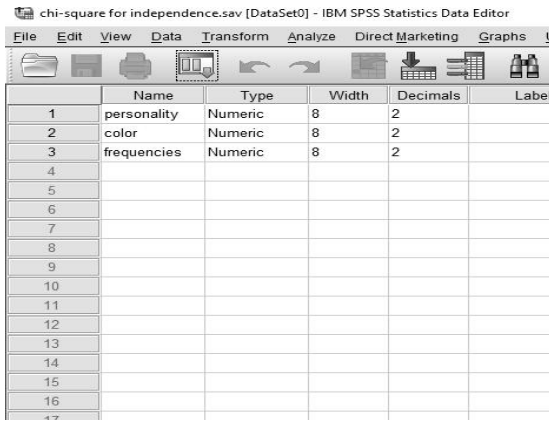
$$\chi^2 = \Sigma\ (f_o - f_e)^2 / f_e$$

Thus, the total number of f_e values that you can freely choose is (R – 1) (C – 1), and the degrees of freedom for the chi-square test of independence are given by the formula df (R – 1)(C – 1) (17.5) Also note that once you calculate the expected frequencies to fill the smaller matrix, the rest of the f_e values can be found by subtraction.

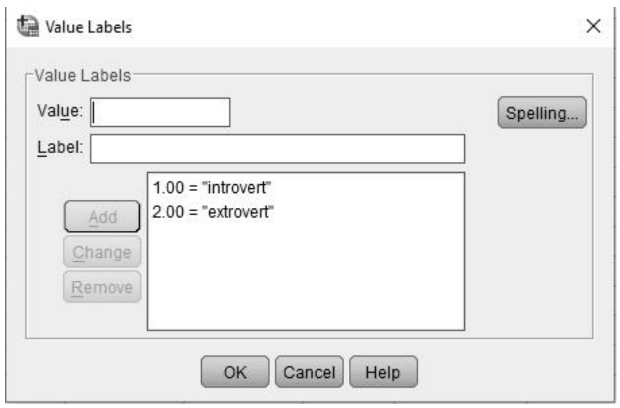
Chi-square test for independence in SPSS

Entering Data

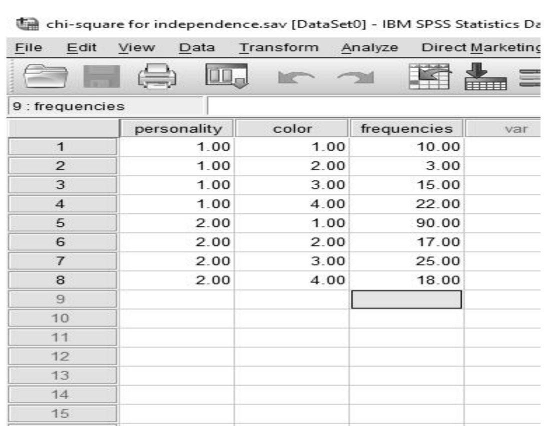
- In the SPSS data editor, create three variables, one for frequencies and two categorical variables (e.g., personality and color).



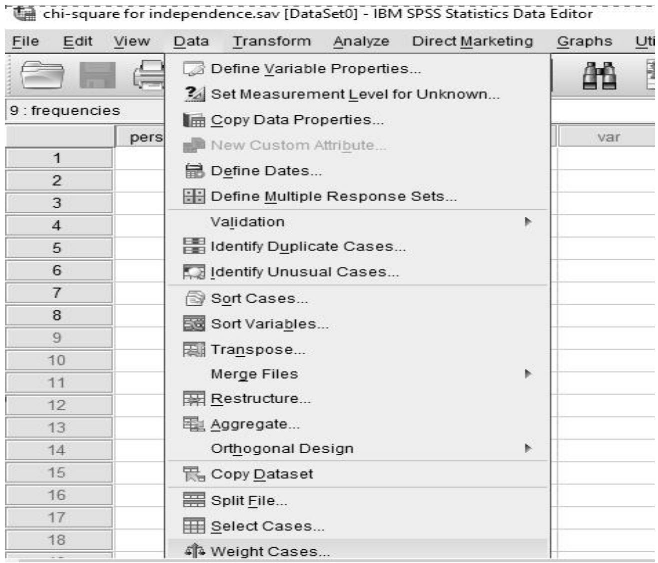
- In Value Labels section, assign codes to the levels of the variables.



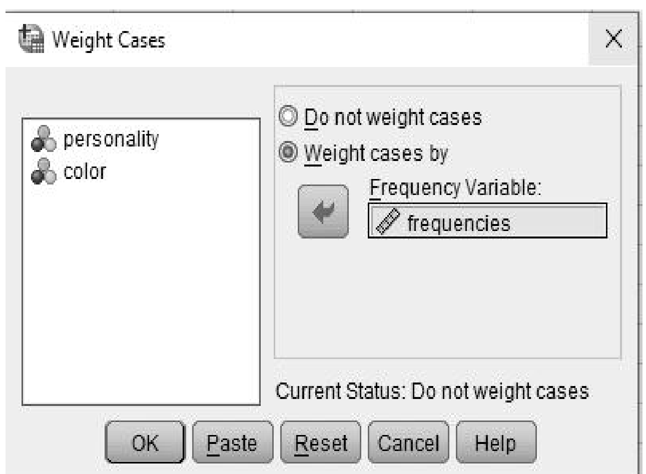
- In a first column, enter a code (e.g., 1 and 2) that identifies the row corresponding to each observed frequency.
- In the second column, enter a code (e.g., 1, 2, 3 and 4) that identifies the column corresponding to each observed frequency.
- Enter the complete set of observed frequencies in a column named frequencies.



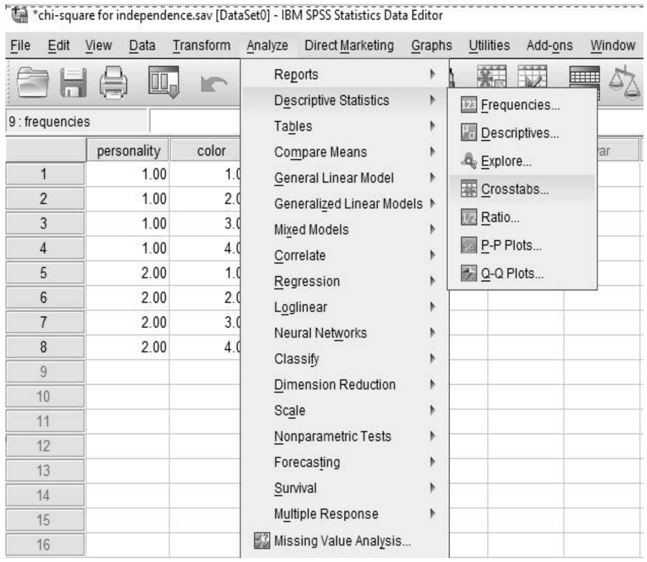
- We need to Weight our cases to tell the SPSS Statistics that we have summated the categories.
- Click Data in the top menu and go to Weight Cases.



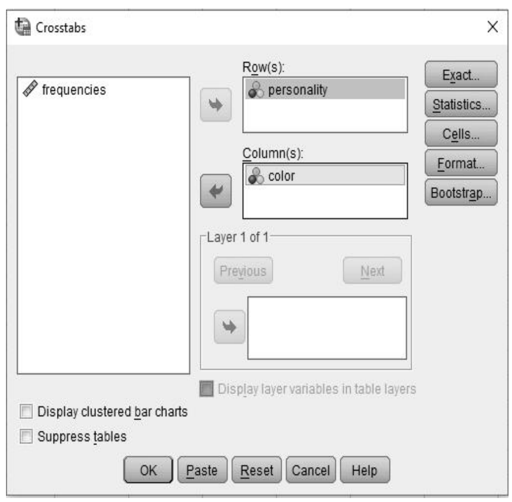
- Click on the circle Weight cases by and drag the frequencies and variable into the Frequency Variable box.
- Click Ok.



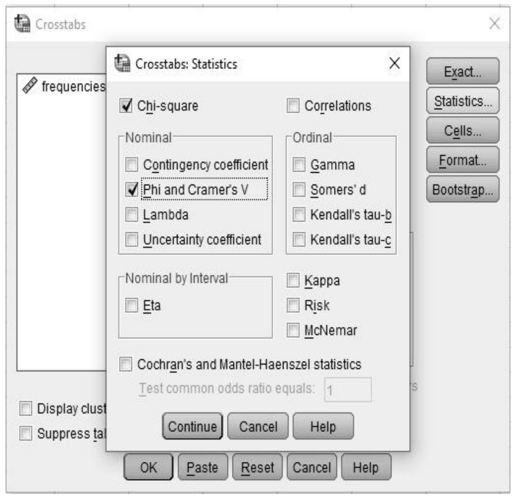
- Click Analyze on the tool bar at the top of the page, select Descriptive Statistics, and click on Crosstabs.



- Highlight the label for the column containing the rows (personality) and move it into the Rows box by clicking on the arrow.
- Highlight the label for the column containing the columns (color) and move it into the Columns box by clicking on the arrow.



- Click on the Statistics button. Tick Chi-square and Phi and Cramer's V. Click on Continue.
- Click OK.



- The first thing we should check is whether we have violated one of the assumptions of chi-square concerning the ‘minimum expected cell frequency’, which should be 5 or greater. The main value that we are interested in is the Pearson Chi-Square value, which is presented in the Chi-Square Tests.
- In the example the chi-square value is $\chi^2(3)=35.6$, with an associated significance level of $p<.001$, presented in the column labelled Asymptotic Significance (2-sided).
- This indicates that our result is significant and there appears to be an association between personality type and color preference.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	35.600[a]	3	.000
Likelihood Ratio	35.033	3	.000
Linear-by-Linear Association	34.476	1	.000
N of Valid Cases	200		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.00.

- For 2 by 2 tables the most commonly used effect size is the phi coefficient, which is a correlation coefficient and can range from 0 to 1, with higher values indicating a stronger association between the two variables.
- For tables larger than 2 by 2 the value to report is Cramer’s V, which takes into account the degrees of freedom.

- In this example the Cramer's V value (shown in the table Symmetric Measures) is .422 indicating a large effect size.
- For tables larger than 2 by 2 the value to report is Cramer's V, which takes into account the degrees of freedom.
- In this example the Cramer's V value (shown in the table Symmetric Measures) is .422 indicating a large effect size.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.422	.000
	Cramer's V	.422	.000
N of Valid Cases		200	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Lesson 42

CHI-SQUARE TEST FOR INDEPENDENCE-II

The Pearson correlation measures the degree of linear relationship between two variables when the data (X and Y values) consist of numerical scores from an interval or ratio scale of measurement. However, other correlations have been developed for nonlinear relationships and for other types of data. When the Pearson correlation formula is used with data from an ordinal scale (ranks), the result is called the Spearman correlation.

Pearson’s r and Spearman’s rho are very similar. They are both correlation coefficients, interpreted in the same way. Pearson’s r is used when our data meet the assumptions for a parametric test. Spearman’s rho (ρ) is used when the data do not conform to these assumptions. To summarize, the Spearman correlation measures the relationship between two variables when both are measured on ordinal scales (ranks). There are two general situations in which the Spearman correlation is used:

1. Spearman is used when the original data are ordinal; that is, when the X and Y values are ranks. In this case, you simply apply the Pearson correlation formula to the set of ranks.
2. Spearman is used when a researcher wants to measure the consistency of a relationship between X and Y, independent of the specific form of the relationship.

In this case, the original scores are first converted to ranks, then the Pearson correlation formula is used with the ranks. Because the Pearson formula measures the degree to which the ranks fit on a straight line, it also measures the degree of consistency in the relationship for the original scores. Incidentally, when there is a consistently one-directional relationship between two variables, the relationship is said to be monotonic. Thus, the Spearman correlation measures the degree of monotonic relationship between two variables. In either case, the Spearman correlation is identified by the symbol rs to differentiate it from the Pearson correlation. When we have small participant numbers and are uncertain as to whether we meet the assumptions for Pearson’s r, it is advised to use the Spearman’s rho.

Spearman’s rho transforms the original scores into ranks before performing further calculations. This test is less sensitive to bias due to the effect of outliers. This can be used to reduce the weight of outliers (large distances get treated as a one-rank difference). It does not require assumption of normality.

Look at the following data. Nine people were asked to rate the attractiveness of a target person, and then rate themselves (myattract) on a 10-point scale from 1 (awful) to 10 (wonderful). Small participant numbers, the nature of the data and the fact that many participants rated themselves as near-wonderful should make you suspicious about such data conforming to the assumptions for a parametric test.

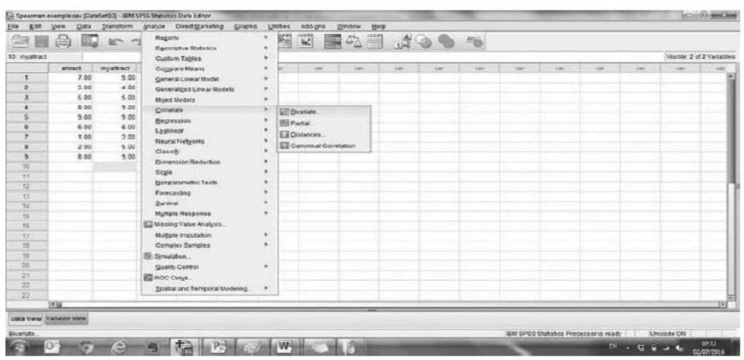
Hypotheses

The hypothesis in this example are given below:

- H_o: There is no correlation between the X (person attractiveness) and the Y (self-attractiveness).
- H_1: There is a correlation between the X and the Y.

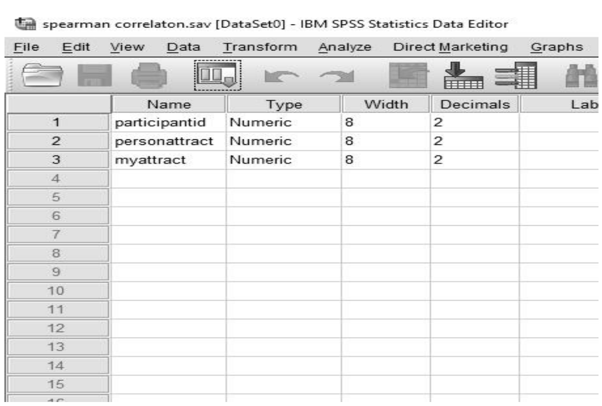
attract	myattract	attract	myattract
7.00	9.00	6.00	8.00
5.00	4.00	1.00	3.00
5.00	5.00	2.00	5.00
8.00	9.00	8.00	9.00
9.00	9.00		

Open your datafile. Choose *Analyze, Correlate, Bivariate*:

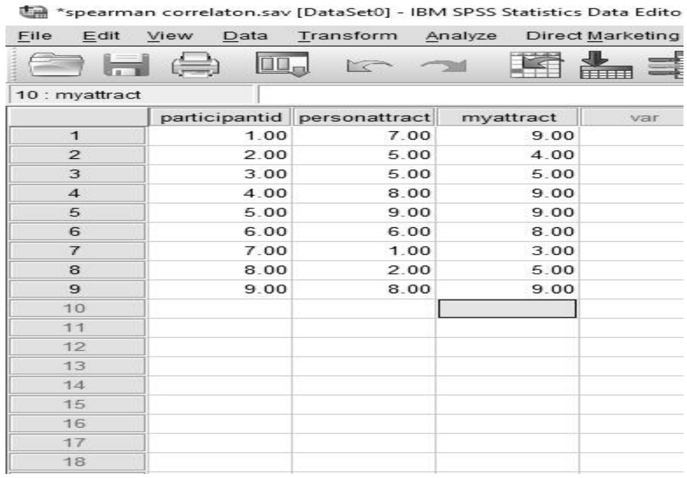


Entering Data

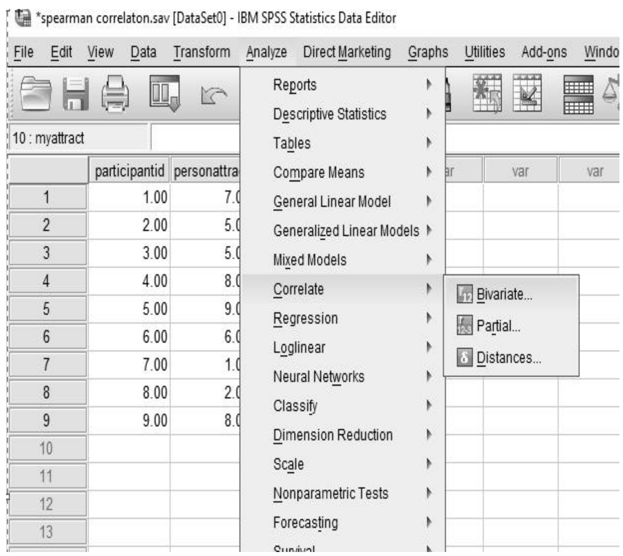
- Create two variables (i.e., personattract and myattract) in the SPSS data editor as shown.



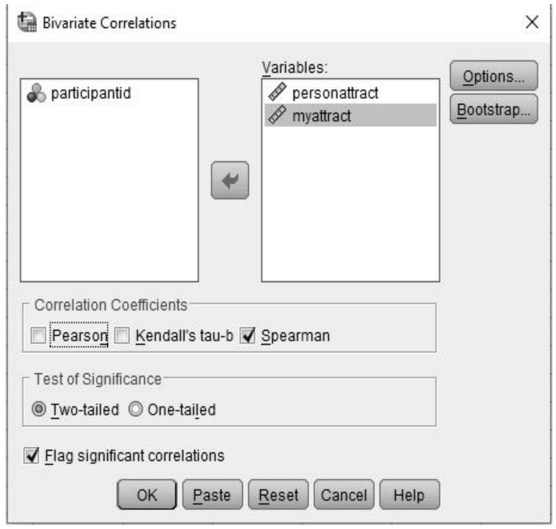
- Enter data for personal attract and my attract in two columns respectively.



- In the top menu, go to Analyze > Correlate > Bivariate.



- This brings us to the bivariate correlations dialogue box.
- We need to uncheck the default Pearson option and select Spearman.
- Move the variables of interest from the left-hand side to the variable list on the right.
- Click OK.



- As in the matrix from Pearson's r, we only need to focus on one half of the matrix.
- The results indicate that there is significant positive correlation between the ratings given to a target person and oneself (ρ=.92, p<.001).

Nonparametric Correlations

[DataSet0] C:\Users\Peaceful Tears\Desktop\spearman correlaton.sav

Correlations

			personattract	myattract
Spearman's rho	personattract	Correlation Coefficient	1.000	.921**
		Sig. (2-tailed)	.	.000
		N	9	9
	myattract	Correlation Coefficient	.921**	1.000
		Sig. (2-tailed)	.000	.
		N	9	9

**. Correlation is significant at the 0.01 level (2-tailed).

Reporting Results

The relationship between person attractiveness and self-attractiveness was investigated using Spearman correlation coefficient. There was a strong, positive correlation between the two variables, **ρ = .92, *n* = 9, *p* < .001.**

Lesson 43

MANN-WHITNEY U-TEST

The Mann-Whitney U Test is used to test for differences between two independent groups on a continuous measure. The Mann-Whitney test is designed to use the data from two separate samples to evaluate the difference between two treatments (or two populations). The calculations for this test require that the individual scores in the two samples be rank-ordered. For example, do males and females differ in terms of their self-esteem? This test is the non-parametric alternative to the t-test for independent samples. Instead of comparing means of the two groups, as in the case of the t-test, the Mann-Whitney U Test actually compares **medians**.

It converts the scores on the continuous variable to ranks across the two groups. It then evaluates whether the ranks for the two groups differ significantly. As the scores are converted to ranks, the actual distribution of the scores does not matter. This test is far simpler than the t-tests in that it does not involve calculations of means, standard deviations and standard errors.

The mathematics of the Mann-Whitney test are based on the following simple observation: A real difference between the two treatments should cause the scores in one sample to be generally larger than the scores in the other sample. If the two samples are combined and all the scores are ranked, then the larger ranks should be concentrated in one sample and the smaller ranks should be concentrated in the other sample. **For example, f**or the Mann Whitney U test, we will be exploring the impact of gender on participants' levels of self-esteem using Survey.sav SPSS data file. For this purpose, research questions are given below:

Research Question

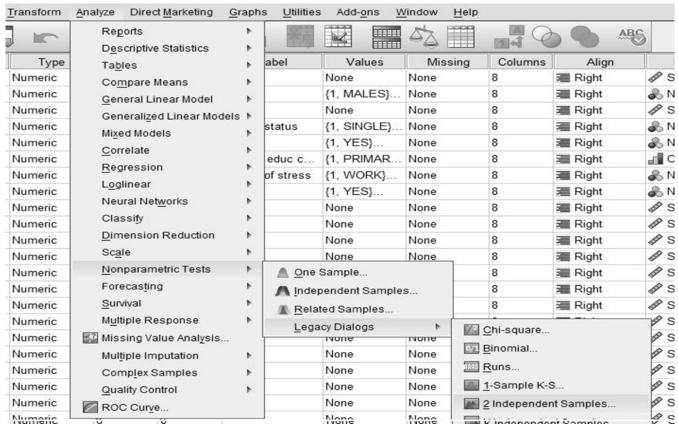
- Do males and females differ in terms of their levels of self-esteem?
- Do males have higher levels of self-esteem than females?

Mann Whitney U Test in SPSS

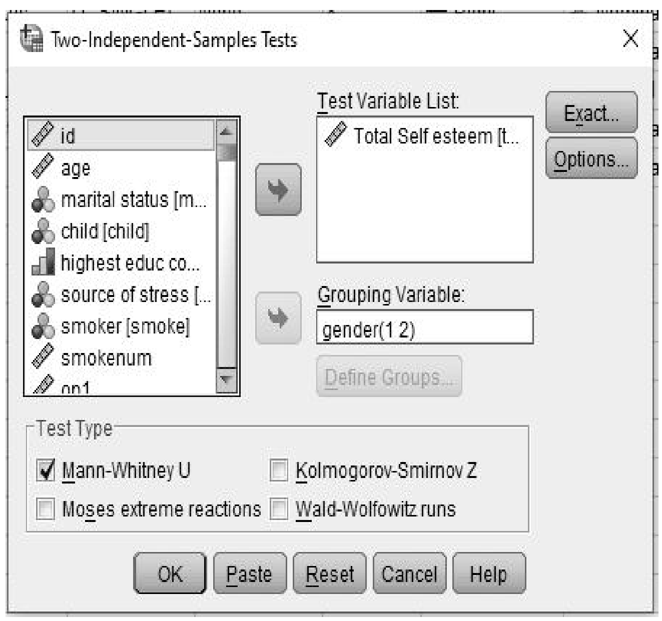
To run this test in SPSS we need two variables: One categorical variable with two groups (e.g., gender) and one continuous variable (e.g., total self-esteem).

Running Analysis

- Choose Analyze, Nonparametric Tests, Legacy Dialogs and 2 Independent Samples as shown.



- Move the DV to the Test variable list and IV to the grouping variable boxes.
- Click on define groups as we did with the t-test. Make sure to define them.
- Also check Mann-Whitney U option. Press OK.



- The first section of the output shows the mean ranks of both groups, and the sum of ranks.
- We can also check the number of participants in each condition.

Ranks

	gender	N	Mean Rank	Sum of Ranks
Total Self esteem	MALES	184	227.14	41794.00
	FEMALES	252	212.19	53472.00
	Total	436		

Test Statistics[a]

	Total Self esteem
Mann-Whitney U	21594.000
Wilcoxon W	53472.000
Z	-1.227
Asymp. Sig. (2-tailed)	.220

a. Grouping Variable: gender

- The next section shows the value of the test statistic, in this case the Mann–Whitney U, which is 21594.0.
- In this case the **Sig**. value is P=.220.
- This is not less than .05, therefore the result is not significant.
- In other words the distribution of self-esteem scores is same across both categories of gender.
- SPSS does not provide an effect size statistic for Mann-Whitney U test, however, we can calculate an approximate value of r using the z value (called **standardized test statistic**) which is available in the Test statistic table (z= -1.23).
- r = z / square root of N where N = total number of cases.
- In this example, z = –1.23 and N = 436; $-1.23/\sqrt{436}$
- **Therefore the r value is .06.**
- This would be considered a very small effect size using Cohen (1988) criteria of .1=small effect, .3=medium effect, .5=large effect.

Reporting Results

A Mann-Whitney U Test revealed no significant difference in the self-esteem levels of males (Mean Rank = 227.1, n =184) and females (Mean Rank = 212.9, n = 252), U = 21594, z = –1.23, p = .22, r = .06.

Lesson 44

WILCOXON SIGNED RANK TEST

Alternative to the paired t-test: the Wilcoxon

The Wilcoxon Signed Rank Test (also referred to as the Wilcoxon matched pairs signed ranks test) is designed for use with repeated measures; that is, when the participants are measured on two occasions, or under two different conditions. The Wilcoxon test is designed to evaluate the difference between two treatments, using the data from a repeated-measures experiment. Recall that a repeated-measures study involves only one sample, with each individual in the sample being measured twice. The difference between the two measurements for each individual is recorded as the score for that individual. The Wilcoxon test requires that the differences be rank-ordered from smallest to largest in terms of their absolute magnitude, without regard for sign or direction.

For the Wilcoxon test, there are two possibilities for tied scores:

1. A participant may have the same score in treatment 1 and in treatment 2, resulting in a difference score of zero.
2. Two (or more) participants may have identical difference scores (ignoring the sign of the difference).

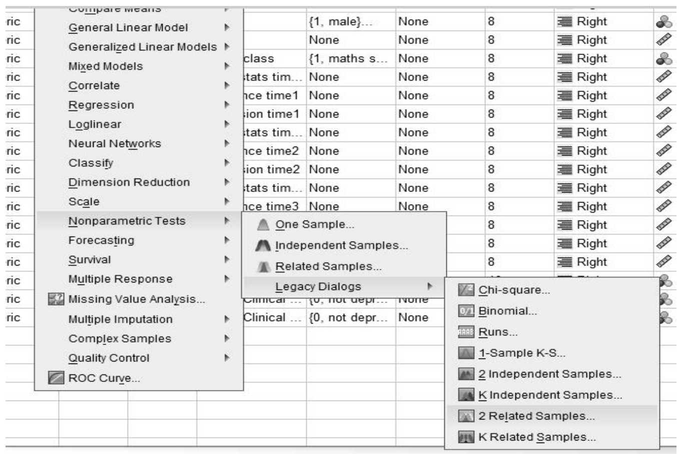
Imagine the psychologist is interested in the change in depression levels. She wants to compare the BDI scores on Sunday to those on Wednesday. Suppose the distributions of scores for drugs is not normal on one of the two days. So the psychologist would have to use a non-parametric test. The Wilcoxon signed-rank test is based on ranking the differences between scores in the two conditions we're comparing. Once these differences have been ranked, the sign of the difference (positive or negative) is assigned to the rank.

To illustrate the use of this test, we will be using data from the experim.sav file. In this example, in which we are studying "Is there a change in the scores on the Fear of Statistics Test from Time 1 to Time 2?" We will compare the scores on a Fear of Statistics Test administered before and after an intervention.

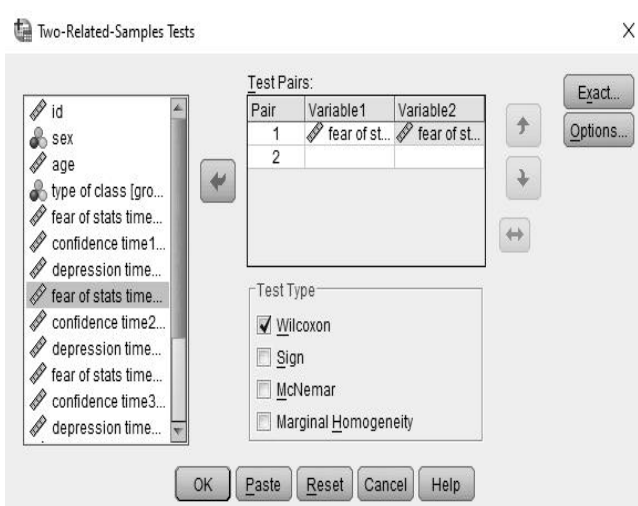
Running Analysis

- To run this test in SPSS, there is need of one group of participants measured on the same continuous scale or measured on two different occasions.

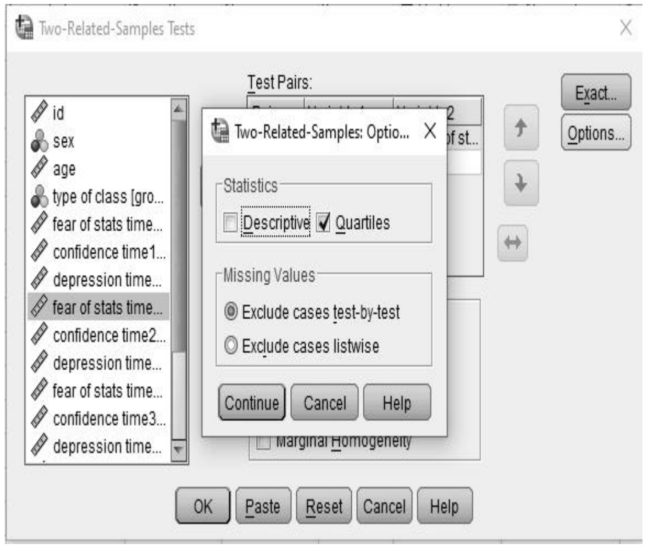
- The variables involved are scores at Time 1 or Condition 1, and scores at Time 2 or Condition 2.
- To follow along with this example, open the **experim.sav** data file.
- From the menu at the top of the screen, click on **Analyze**, then select **Nonparametric Tests**, then **Legacy Dialogs** and then **2 Related Samples**.



- Click on the variables that represent the scores at Time 1 and at Time 2 (e.g., fear of stats time1: fost1, fear of stats time2: fost2).
- Click on the arrow to move these into the Test Pairs box.
- Make sure that the Wilcoxon box is ticked in the Test Type section.



- Click on the Options button. Choose Quartiles (this will provide the median scores for each time point).
- Click on Continue and then on OK.



- The first table in the output gives descriptive statistics (Median values) for time 1 and time 2.

Descriptive Statistics

	N	Percentiles 25th	50th (Median)	75th
fear of stats time1	30	37.00	40.00	44.00
fear of stats time2	30	34.50	38.00	40.00

Interpreting Results

- The two things of interest are the Z value and the associated significance levels, presented as Asymp. Sig. (2-tailed).
- If the significance level is equal to or less than .05. We can conclude that the difference between the two scores is statistically significant.

Test Statistics[a]

	fear of stats time2 - fear of stats time1
Z	-4.180[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

- The effect size for this test can be calculated using the same procedure as described for the Mann-Whitney U Test; that is, by dividing the z value by the square root of N.
- In this situation, N = the number of observations over the two time points, not the number of cases.
- For this calculation we can ignore any negative sign out the front of the z value.
- In this example, Z = 4.18, N = 60 (cases × 2); therefore r = .54, indicating a large effect size using Cohen (1988) criteria of .1 = small effect, .3 = medium effect, .5 = large effect.

Test Statistics[a]

	fear of stats time2 - fear of stats time1
Z	-4.180[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Reporting Results

A Wilcoxon Signed Rank Test revealed a statistically significant reduction in fear of statistics following participation in the training program, z=-4.18, p <.001, with a large effect size (r =.54). The median score on the Fear of Statistics Scale decreased from pre-program (Md = 40) to post-program (Md = 38).

Lesson 45

KRUSKAL WALLIS & FRIEDMAN TEST

The Kruskal Wallis Test: An Alternative to the Independent- Measure ANOVA

The Kruskal-Wallis Test (sometimes referred to as the Kruskal-Wallis H Test) is the non-parametric alternative to a one-way between-groups analysis of variance. It allows you to compare the scores on some continuous variable for *three or more groups*. It is similar in nature to the Mann-Whitney U Test, but it allows you to compare more than just two groups. Scores are converted to ranks and the mean rank for each group is compared. This is a '*between groups*' analysis, so different people must be in each of the different groups.

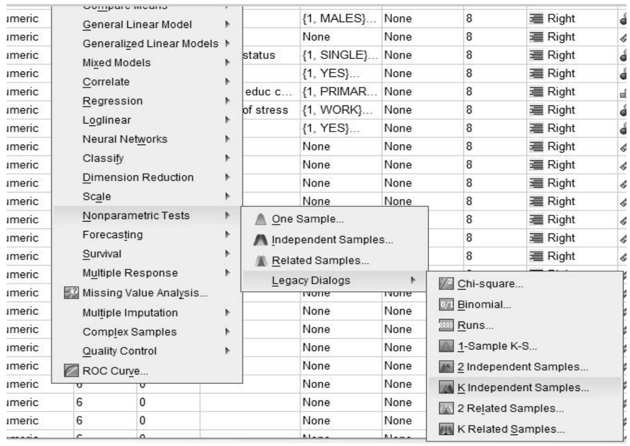
To begin with, scores are ordered from lowest to highest, ignoring the group to which the score belongs. The lowest score is assigned a rank of 1, the next highest a rank of 2 and so on. Once ranked, the scores are collected back into their groups and their ranks are added within each group. The Kruskal–Wallis test tells us that, overall, groups come from different populations. However, it doesn't tell us which groups differ. The simplest way to break down the overall effect is to compare all pairs of groups (known as pairwise comparisons). Let's take an example to run this Kruskal Wallis Test in SPSS. In this example, we will explore levels of optimism across three age groups. The research question is**:** Is there a difference in optimism levels across three age levels?

We need two variables:

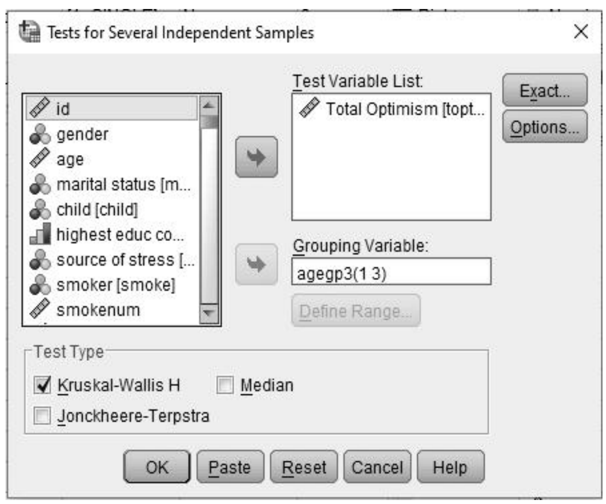
- One categorical independent variable with three or more categories (e.g. age 3: 18–29, 30–44, 45+)
- One continuous dependent variable (e.g. total optimism).

Running Analysis

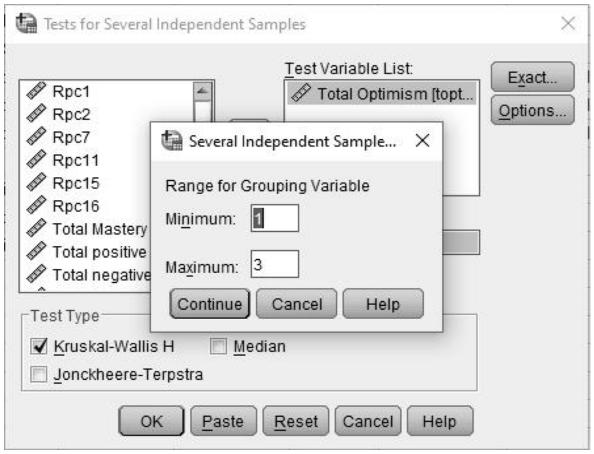
- To follow along with this example, open the survey.sav data file.
- From the menu at the top of the screen click on Analyze, select Nonparametric Tests, Legacy Dialogs and K Independent samples.



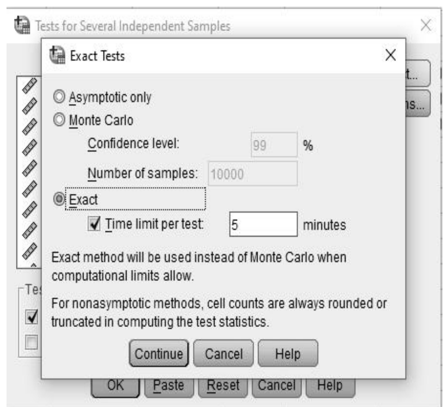
- Select the continuous dependent variable (e.g., Total optimism) and move it into the **Test Variable List** box.
- Click on the categorical independent variable (e.g., age 3 groups; agegp3) and move it into the **Grouping Variable** box.



- Then press the **Define Range** button. This enables us to define the range of the groups.
- Here, our groups are coded 1, 2, and 3, so the range is 1-3.
- Enter the range into the dialogue box and press **Continue**.



- Select the **Exact button** and check the **Exact** option.
- Click **Continue.**



- Ensure that the **Kruskal–Wallis H** box is checked.
- For Descriptives, press the **Options** button and select descriptive option.
- Press **OK**.