

# **Statistics Calculator**

**By David S. Walonick, Ph.D.**

**Copyright © 1996-2010, StatPac Inc.**

# Contents

<b>Introduction</b>	<b>3</b>
Main menu.....	3
User interface.....	4
Formulas.....	4
Basic concepts.....	4
Problem recognition and definition.....	4
Creating the research design.....	5
Methods of research.....	5
Sampling.....	6
Data collection.....	7
Reporting the results.....	8
Validity.....	8
Reliability.....	8
Systematic and random error.....	9
Formulating hypotheses from research questions.....	9
Type I and Type II errors.....	10
Types of data.....	10
Significance.....	13
One-tailed and two-tailed tests.....	14
Procedure for significance testing.....	14
Bonferroni's theorem.....	15
Central tendency.....	16
Variability.....	16
Standard error of the mean.....	17
Inferences with small sample sizes.....	17
Degrees of freedom.....	17
Choosing a significance test.....	18
<b>Distributions</b>	<b>19</b>
Distributions menu.....	19
Normal distribution.....	19
Probability of a $z$ value.....	19
Critical $z$ for a given probability.....	20
Probability of a defined range.....	20
T distribution.....	21
Probability of a $t$ value.....	21
Critical $t$ value for a given probability.....	21
Probability of a defined range.....	22
F distribution.....	22
Probability of a $F$ -ratio.....	22
Critical $F$ for a given probability.....	23
Chi-square distribution.....	23
Probability of a chi-square statistic.....	23
Critical chi-square for a given probability.....	23

<b>Counts</b>	<b>25</b>
Counts menu .....	25
Chi-square test .....	25
Fisher's exact test .....	26
Binomial test .....	27
Poisson distribution events test .....	27
<b>Percents</b>	<b>29</b>
Percents menu .....	29
Choosing the proper test .....	29
One sample t-test between percents .....	31
Two sample t-test between percents .....	31
Confidence intervals around a percent .....	33
<b>Means</b>	<b>35</b>
Means menu .....	35
Mean and standard deviation of a sample .....	35
Matched pairs t-test between means .....	36
Independent groups t-test between means .....	37
Confidence interval around a mean .....	38
Compare a sample mean to a population mean .....	39
Compare two standard deviations .....	40
Compare three or more means .....	40
<b>Correlation</b>	<b>43</b>
Correlation types .....	43
Regression .....	44
<b>Sampling</b>	<b>47</b>
Sampling menu .....	47
Sample size for percents .....	47
Sample size for means .....	48

---

# Introduction

---

## Main menu

**Statistics Calculator** is an easy-to-use program designed to perform a series of basic statistical procedures related to distributions and probabilities. Most of the procedures are called *inferential* because data from a sample is used to infer to a population.

The menu bar of **Statistic Calculator** contains eight choices. These represent the basic types of operations that can be performed by the software.

**Exit Distributions Counts Percents Means Correlation Sampling Help**

The **Exit** menu item is used to exit the software.

The **Distributions** menu item is the electronic equivalent of probability tables. Algorithms are included for the z, t, F, and chi-square distributions. This selection may be used to find probabilities and critical values for the four statistics.

The **Counts** menu item contains routines to analyze a contingency table of counts, compute Fisher's exact probability for two-by-two tables, use the binomial distribution to predict the probability of a specified outcome, and the poisson distribution to test the likelihood of observing a specified number of events.

The **Percents** menu item is used to compare two percents. Algorithms are included to compare proportions drawn from one or two samples. There is also a menu option to calculate confidence intervals around a percent.

The **Means** menu item is used to calculate a mean and standard deviation of a sample, compare two means to each other, calculate a confidence interval around a mean, compare a sample mean to a population mean, compare two standard deviations to each other, and compare three or more standard deviations.

The **Correlation** menu item is used to calculate correlation and simple linear regression statistics for paired data. Algorithms are included for ordinal and interval data.

The **Sampling** menu item is used to determine the required sample size for a study. The software can be used for problems involving percents and means.

The **Help** menu item is used to get this on-line help.

---

## User interface

**Statistics Calculator** has a "fill in the form" user interface. After selecting a particular type of significance test from the menu, a form will be displayed. Fill in the form and press the Calculate button to calculate the answer.

Pressing the Exit button on a form will close the form and erase the data.

On-line help is provided for all menu selections. The on-line help describes the application of each statistical procedure. Practical examples are also included.

For most significance tests, the result is automatically copied to the clipboard when you click the Calculate button, thereby allowing you to do a paste operation in your word processor. The result is usually in APA (American Psychological Association) format, so it may be pasted directly into a research paper. The procedure would be to first open your word processor. Then run **Statistics Calculator**. Select the test, fill in the form, and click the Calculate button. Activate your word processor by clicking on the Windows menu bar. Set the cursor where you want to report the statistic and select Edit, Paste (or press Ctrl V). The APA formatted result will be inserted into the text.

You can also copy the screen image to the clipboard by pressing the Alt and Print Screen keys together.

---

## Formulas

Formulas used in the **Statistics Calculator** may be found in nearly any statistics textbook. The textbook written by the author of **Statistics Calculator** is called "*Survival Statistics*". It contains all the formulas with worked examples. "*Survival Statistics*" is available for purchase at the StatPac Web site: <http://www.statpac.com/statistics-book>

Finite population correction is incorporated into all relevant formulas. In a typical research scenario the population is very large compared to the sample and the correction is unnecessary. Leave the population size blank to ignore the correction. However, if the sample size is more than 10% of the population size, the population size should be specified.

---

## Basic concepts

### Problem recognition and definition

We understand the world by asking questions and searching for answers. Our construction of reality depends on the nature of our inquiry.

All research begins with a question. Intellectual curiosity is often the foundation for scholarly inquiry. Some questions are not testable. The classic philosophical example is to ask, "How many angels can dance on the head of a pin?" While the question might elicit profound and thoughtful revelations, it clearly cannot be tested with an empirical experiment. Prior to Descartes, this is precisely the kind of question that would engage the minds of learned men. Their answers came from within. The scientific method precludes asking questions that cannot be empirically tested. If the angels cannot be observed or detected, the question is considered inappropriate for scholarly research.

Defining the goals and objectives of a research project is one of the most important steps in the research process. Do not underestimate the importance of this step. Clearly stated goals keep a research project focused. The process of goal definition usually begins by writing down the broad and general goals of the study. As the process continues, the goals become more clearly defined and the research issues are narrowed.

Exploratory research (e.g., literature reviews, talking to people, and focus groups) goes hand-in-hand with the goal clarification process. The literature review is especially important because it obviates the need to reinvent the wheel for every new research question. More importantly, it gives researchers the opportunity to build on each others work.

The research question itself can be stated as a hypothesis. A hypothesis is simply the investigator's belief about a problem. Typically, a researcher formulates an opinion during the literature review process. The process of reviewing other scholar's work often clarifies the theoretical issues associated with the research question. It also can help to elucidate the significance of the issues to the research community.

The hypothesis is converted into a null hypothesis in order to make it testable because the only way to test a hypothesis is to eliminate alternatives of the hypothesis. Statistical techniques will enable us to reject or fail to reject a null hypothesis, but they do not provide us with a way to accept a hypothesis. Therefore, all hypothesis testing is indirect.

## Creating the research design

Defining a research problem provides a format for further investigation. A well-defined problem points to a method of investigation. There is no one best method of research for all situations. Rather, there are a wide variety of techniques for the researcher to choose from. Often, the selection of a technique involves a series of trade-offs. For example, there is often a trade-off between cost and the quality of information obtained. Time constraints sometimes force a trade-off with the overall research design. Budget and time constraints must always be considered as part of the design process.

## Methods of research

There are three basic methods of research: 1) survey, 2) observation, and 3) experiment. Each method has its advantages and disadvantages.

The *survey* is the most common method of gathering information in the social sciences. It can be a face-to-face interview, telephone, or mail survey. A personal interview is one of the best methods obtaining personal, detailed, or in-depth information. It usually involves a lengthy questionnaire that the interviewer fills out while asking questions. It allows for extensive probing by the interviewer and gives respondents the ability to elaborate their answers. Telephone interviews are similar to face-to-face interviews. They are more efficient in terms of time and cost, however, they are limited in the amount of in-depth probing that can be accomplished, and the amount of time that can be allocated to the interview. A mail survey is generally the most cost effective interview method. The researcher can obtain opinions, but trying to meaningfully probe opinions is very difficult.

*Observation* research monitors respondents' actions without directly interacting with them. It has been used for many years by A.C. Nielsen to monitor television viewing habits. Psychologists often use one-way mirrors to study behavior. Anthropologists and social scientists often study societal and group behaviors by simply observing them. The fastest growing form of observation research has been made possible by

the bar code scanners at cash registers, where purchasing habits of consumers can now be automatically monitored and summarized.

In an *experiment*, the investigator changes one or more variables over the course of the research. When all other variables are held constant (except the one being manipulated), changes in the dependent variable can be explained by the change in the independent variable. It is usually very difficult to control all the variables in the environment. Therefore, experiments are generally restricted to laboratory models where the investigator has more control over all the variables.

## Sampling

It is incumbent on the researcher to clearly define the target population. There are no strict rules to follow, and the researcher must rely on logic and judgment. The population is defined in keeping with the objectives of the study.

Sometimes, the entire population will be sufficiently small, and the researcher can include the entire population in the study. This type of research is called a *census* study because data is gathered on every member of the population.

Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen *sample* can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

Sampling methods are classified as either *probability* or *nonprobability*. In probability samples, each member of the population has a *known non-zero* probability of being selected. Probability methods include random sampling, systematic sampling, and stratified sampling. In nonprobability sampling, members are selected from the population in some nonrandom manner. These include convenience sampling, judgment sampling, quota sampling, and snowball sampling. The advantage of probability sampling is that *sampling error* can be calculated. Sampling error is the degree to which a sample might differ from the population. When inferring to the population, results are reported plus or minus the sampling error. In nonprobability sampling, the degree to which the sample differs from the population remains unknown.

*Random sampling* is the purest form of probability sampling. Each member of the population has an equal and known chance of being selected. When there are very large populations, it is often difficult or impossible to identify every member of the population, so the pool of available subjects becomes biased.

*Systematic sampling* is often used instead of random sampling. It is also called an *Nth name selection* technique. After the required sample size has been calculated, every Nth record is selected from a list of population members. As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method. Its only advantage over the random sampling technique is simplicity. Systematic sampling is frequently used to select a specified number of records from a computer file.

*Stratified sampling* is commonly used probability method that is superior to random sampling because it reduces sampling error. A *stratum* is a subset of the population that share at least one common characteristic. The researcher first identifies the relevant strata and their actual representation in the population. Random sampling is then used to select subjects from each stratum until the number of subjects in that stratum is proportional to its frequency in the population. Stratified sampling is often used when one or more of the strata in the population have a low incidence relative to the other strata.

*Convenience sampling* is used in exploratory research where the researcher is interested in getting an inexpensive approximation of the truth. As the name implies, the sample is selected because they are convenient. This nonprobability method is often used during preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample.

*Judgment sampling* is a common nonprobability method. The researcher selects the sample based on judgment. This is usually an extension of convenience sampling. For example, a researcher may decide to draw the entire sample from one "representative" city, even though the population includes all cities. When using this method, the researcher must be confident that the chosen sample is truly representative of the entire population.

*Quota sampling* is the nonprobability equivalent of stratified sampling. Like stratified sampling, the researcher first identifies the strata and their proportions as they are represented in the population. Then convenience or judgment sampling is used to select the required number of subjects from each stratum. This differs from stratified sampling, where the strata are filled by random sampling.

*Snowball sampling* is a special nonprobability method used when the desired sample characteristic is rare. It may be extremely difficult or cost prohibitive to locate respondents in these situations. Snowball sampling relies on referrals from initial subjects to generate additional subjects. While this technique can dramatically lower search costs, it comes at the expense of introducing bias because the technique itself reduces the likelihood that the sample will represent a good cross section from the population.

## Data collection

There are very few hard and fast rules to define the task of data collection. Each research project uses a data collection technique appropriate to the particular research methodology. The two primary goals for both quantitative and qualitative studies are to maximize response and maximize accuracy.

When using an outside data collection service, researchers often *validate* the data collection process by contacting a percentage of the respondents to verify that they were actually interviewed. *Data editing* and *cleaning* involves the process of checking for inadvertent errors in the data. This usually entails using a computer to check for out-of-bounds data.

*Quantitative* studies employ deductive logic, where the researcher starts with a hypothesis, and then collects data to confirm or refute the hypothesis. *Qualitative* studies use inductive logic, where the researcher first designs a study and then develops a hypothesis or theory to explain the results of the analysis.

Quantitative analysis is generally fast and inexpensive. A wide assortment of statistical techniques are available to the researcher. Computer software is readily available to provide both basic and advanced multivariate analysis. The researcher simply follows the preplanned analysis process, without making subjective decisions about the data. For this reason, quantitative studies are usually easier to execute than qualitative studies.

Qualitative studies nearly always involve in-person interviews, and are therefore very labor intensive and costly. They rely heavily on a researcher's ability to exclude personal biases. The interpretation of qualitative data is often highly subjective, and different researchers can reach different conclusions from the same data. However, the goal of qualitative research is to develop a hypothesis--not to test one.

Qualitative studies have merit in that they provide broad, general theories that can be examined in future research.



## Reporting the results

The most important consideration in preparing any research report is the nature of the audience. The purpose is to communicate information, and therefore, the report should be prepared specifically for the readers of the report. Sometimes the format for the report will be defined for the researcher (e.g., a thesis or dissertation), while other times, the researcher will have complete latitude regarding the structure of the report. At a minimum, the report should contain an abstract, problem statement, methods section, results section, discussion of the results, and a list of references.

## Validity

*Validity* refers to the accuracy or truthfulness of a measurement. Are we measuring what we think we are? This is a simple concept, but in reality, it is extremely difficult to determine if a measure is valid.

Face validity is based solely on the judgment of the researcher. Each question is scrutinized and modified until the researcher is satisfied that it is an accurate measure of the desired construct. The determination of face validity is based on the subjective opinion of the researcher.

*Content validity* is similar to face validity in that it relies on the judgment of the researcher. However, where face validity only evaluates the individual items on an instrument, content validity goes further in that it attempts to determine if an instrument provides adequate coverage of a topic. Expert opinions, literature searches, and open-ended pretest questions help to establish content validity.

*Criterion-related validity* can be either predictive or concurrent. When a dependent/independent relationship has been established between two or more variables, criterion-related validity can be assessed. A mathematical model is developed to be able to predict the dependent variable from the independent variable(s). *Predictive validity* refers to the ability of an independent variable (or group of variables) to predict a future value of the dependent variable. *Concurrent validity* is concerned with the relationship between two or more variables at the same point in time.

*Construct validity* refers to the theoretical foundations underlying a particular scale or measurement. It looks at the underlying theories or constructs that explain a phenomena. This is also quite subjective and depends heavily on the understanding, opinions, and biases of the researcher.

## Reliability

*Reliability* is synonymous with repeatability. A measurement that yields consistent results over time is said to be reliable. When a measurement is prone to random error, it lacks reliability. The reliability of an instrument places an upper limit on its validity. A measurement that lacks reliability will necessarily be invalid. There are three basic methods to test reliability: test-retest, equivalent form, and internal consistency.

A *test-retest* measure of reliability can be obtained by administering the same instrument to the same group of people at two different points in time. The degree to which both administrations are in agreement is a measure of the reliability of the instrument. This technique for assessing reliability suffers two possible drawbacks. First, a person may have changed between the first and second measurement. Second, the initial administration of an instrument might in itself induce a person to answer differently on the second administration.

The second method of determining reliability is called the *equivalent-form* technique. The researcher creates two different instruments designed to measure identical constructs. The degree of correlation between the instruments is a measure of equivalent-form reliability. The difficulty in using this method is that it may be very difficult (and/or prohibitively expensive) to create a totally equivalent instrument.

The most popular methods of estimating reliability use measures of *internal consistency*. When an instrument includes a series of questions designed to examine the same construct, the questions can be arbitrarily split into two groups. The correlation between the two subsets of questions is called the *split-half* reliability. The problem is that this measure of reliability changes depending on how the questions are split. A better statistic, known as Cronbach's alpha, is based on the mean (absolute value) interitem correlation for all possible variable pairs. It provides a conservative estimate of reliability, and generally represents the lower bound to the reliability of a scale of items. For dichotomous nominal data, the KR-20 (Kuder-Richardson) is used instead of Cronbach's alpha.

## Systematic and random error

Most research is an attempt to understand and explain *variability*. When a measurement lacks variability, no statistical tests can be (or need be) performed. Variability refers to the dispersion of scores.

Ideally, when a researcher finds differences between respondents, they are due to true difference on the variable being measured. However, the combination of systematic and random errors can dilute the accuracy of a measurement. *Systematic error* is introduced through a constant bias in a measurement. It can usually be traced to a fault in the sampling procedure or in the design of a questionnaire. *Random error* does not occur in any consistent pattern, and it is not controllable by the researcher.

## Formulating hypotheses from research questions

There are basically two kinds of research questions: testable and non-testable. Neither is better than the other, and both have a place in applied research.

Examples of non-testable questions are:

*How do managers feel about the reorganization?*

*What do residents feel are the most important problems facing the community?*

Respondents' answers to these questions could be summarized in descriptive tables and the results might be extremely valuable to administrators and planners. Business and social science researchers often ask non-testable research questions. The shortcoming with these types of questions is that they do not provide objective cut-off points for decision-makers.

In order to overcome this problem, researchers often seek to answer one or more testable research questions. Nearly all testable research questions begin with one of the following two phrases:

*Is there a significant difference between ...?*

*Is there a significant relationship between ...?*

For example:

*Is there a significant relationship between the age of managers and their attitudes towards the reorganization?*

*Is there a significant difference between white and minority residents with respect to what they feel are the most important problems facing the community?*

A research hypothesis is a testable statement of opinion. It is created from the research question by replacing the words "Is there" with the words "There is", and also replacing the question mark with a period. The hypotheses for the two sample research questions would be:

*There is a significant relationship between the age of managers and their attitudes towards the reorganization.*

*There is a significant difference between white and minority residents with respect to what they feel are the most important problems facing the community.*

It is not possible to test a hypothesis directly. Instead, you must turn the hypothesis into a null hypothesis. The null hypothesis is created from the hypothesis by adding the words "no" or "not" to the statement. For example, the null hypotheses for the two examples would be:

*There is no significant relationship between the age of managers and their attitudes towards the reorganization.*

*There is no significant difference between white and minority residents with respect to what they feel are the most important problems facing the community.*

All statistical testing is done on the null hypothesis...never the hypothesis. The result of a statistical test will enable you to either 1) reject the null hypothesis, or 2) fail to reject the null hypothesis. Never use the words "accept the null hypothesis".

## **Type I and Type II errors**

There are two types of hypothesis testing errors. The first one is called a *Type I error*. This is a very serious error where you wrongly reject the null hypothesis. Suppose that the null hypothesis is: Daily administrations of drug ABC will not help patients. Also suppose that drug ABC is really a very bad drug, and it causes permanent brain damage to people over 60. In your research, you ask for volunteers, and all of the sample is under 60 years of age. The sample seems to improve and you reject the null hypothesis. There could be very serious consequences if you were to market this drug (based on your sample). Type I errors are often caused by sampling problems.

A *Type II error* is less serious, where you wrongly fail to reject the null hypothesis. Suppose that drug ABC really isn't harmful and does actually help many patients, but several of your volunteers develop severe and persistent psychosomatic symptoms. You would probably not market the drug because of the potential for long-lasting side effects. Usually, the consequences of a Type II error will be less serious than a Type I error.

## **Types of data**

One of the most important concepts in statistical testing is to understand the four basic types of data: nominal, ordinal, interval, and ratio. The kinds of statistical tests that can be performed depend upon the type of data you have. Different statistical tests are used for different types of data.

Nominal and ordinal data are *nonparametric* (non-continuous or categorical). Interval and ratio scales are called *parametric* (continuous). Some statistical tests are called parametric tests because they use parametric data. Others are called nonparametric tests because they use nonparametric data. All statistical tests are

designed to be used with a specific kind of data, and may only be performed when you have that kind of data.

### ***Nominal data***

Nominal data is characterized by non-ordered response categories.

#### **Examples of nominal data**

What is your sex?

Male  Female

What program are you in?

Administration/Management

Health Services

Education

Human Services

Do you have health insurance?

Yes  No  Don't know

What school did you attend?

Park Elementary

West Side

Other

What should be done with the program?

Close it down

Seek government funding

Hold a private fund raiser

Other

What state do you live in? \_\_\_\_\_

Note: This question is called an *open-ended* question because it calls for a verbatim response. Even though the categories (i.e., the states) are not listed, the question is still considered nominal because the data can be categorized after it is collected.

Which of the following meats have you eaten in the last week? (Check all that apply)

Hamburger  Pot roast  Liver

Hotdogs  Bacon  Steak

Porkchops  Sausage  Other

Note: This question is called a *multiple response* item because respondents can check more than one category. Multiple response simply means that a respondent can make more than one response to the same question. The data is still nominal because the responses are non-ordered categories.

What are the two most important issues facing our country today?

\_\_\_\_\_ and \_\_\_\_\_

Note: This question is an open-ended multiple response item because it calls for two verbatim responses. It is still considered nominal data because the issues could be categorized after the data is collected.

### **Ordinal data**

Ordinal data is characterized by non-ordered response categories.

#### **Examples of ordinal data**

What is your highest level of education?

- Grade school
- Some high school
- High school graduate
- Some college
- College graduate
- Advanced degree

How many beers have you drunk in the last week?

- None
- One to five
- Six to ten
- Over ten

How would you rate your progress?

- Excellent
- Good
- Fair
- Poor

What has the trend been in your business over the past year?

- Decreasing
- Stable
- Increasing

Please rate the quality of this lecture?

- Low
- Medium
- High

Use a circle to indicate your level of agreement or disagreement with the following statement: Abortion should be a decision between a woman and her doctor.

- |          |       |         |          |          |
|----------|-------|---------|----------|----------|
| Strongly |       |         |          | Strongly |
| Agree    | Agree | Neutral | Disagree | Disagree |
| 1        | 2     | 3       | 4        | 5        |

What is your annual family income?

- Under \$12,000
- \$12,000 to \$23,999
- \$24,000 to \$49,999
- \$50,000 to \$74,999
- \$75,000 or more

## ***Interval and ratio data***

Interval and ratio data are such that each numeric *interval* represents one unit of measurement. *Ratio* scales also have the property of an absolute "zero-point". *Interval* and *ratio-scaled* questions are preferable in research design because they offer the most versatility in the kinds of analyses that may be performed.

### **Examples of interval and ratio data**

What is your age? \_\_\_\_\_

How many children do you have? \_\_\_\_\_

What was your SAT score? \_\_\_\_\_

How many years of school have you completed? \_\_\_\_\_

What percent of your work time do you spend .... ? \_\_\_\_\_

How many collective bargaining sessions have you personally been involved in?  
\_\_\_\_\_

What is the average class size in your school? \_\_\_\_\_

What was your family income last year? \_\_\_\_\_

How many units have you completed? (Circle) 0 1 2 3

What was your GPA as an undergraduate student? \_\_\_\_\_

How many times have you been arrested? \_\_\_\_\_

## **Significance**

What does significance really mean?

Many researchers get very excited when they have discovered a "significant" finding, without really understanding what it means. When a statistic is significant, it simply means that you are very sure that the statistic is reliable. It doesn't mean the finding is important.

For example, suppose we give 1,000 people an IQ test, and we ask if there is a significant difference between male and female scores. The mean score for males is 98 and the mean score for females is 100. We use an independent groups t-test and find that the difference is significant at the .001 level. The big question is, "So what?". The difference between 98 and 100 on an IQ test is a very small difference...so small, in fact, that it's not even important.

Then why did the t-statistic come out significant? Because there was a large sample size. When you have a large sample size, very small differences will be detected as significant. This means that you are very sure that the difference is real (i.e., it didn't happen by fluke). It doesn't mean that the difference is large or important. If we had only given the IQ test to 10 people instead of 1,000, the two-point difference between males and females would not have been significant.

Significance is a statistical term that tells how sure you are that a difference or relationship exists. To say that a significant difference or relationship exists only tells half the story. We might be very sure that a relationship exists, but is it a strong, moderate, or weak relationship? After finding a significant relationship, it is important to evaluate its strength. Significant relationships can be strong or weak. Significant differences can be large or small. It just depends on your sample size.

## One-tailed and two-tailed tests

One important concept in significance testing is whether to use a one-tailed or two-tailed test of significance. The answer is that it depends on your hypothesis. When your research hypothesis states (or implies) the direction of the difference or relationship, then you use a one-tailed probability. For example, a one-tailed test would be used to test these null hypotheses: Females will not score significantly higher than males on an IQ test. Blue collar workers will not have significantly lower education than white collar workers. Superman is not significantly stronger than the average person. In each case, the null hypothesis (indirectly) predicts the direction of the expected difference. A two-tailed test would be used to test these null hypotheses: There will be no significant difference in IQ scores between males and females. There will be no significant difference between blue collar and white collar workers. There is no significant difference in strength between Superman and the average person.

**Statistics Calculator** always reports the two-tailed probability. A one-tailed probability is exactly half the value of a two-tailed probability. Thus, if you have a one-tailed research question, you should divide the probability given by the **Statistics Calculator** by two.

There is a raging controversy (for about the last hundred years) on whether or not it is ever appropriate to use a one-tailed test. The rationale is that if you already know the direction of the difference, why bother doing any statistical tests. The safest bet is to always state your hypotheses so that two-tailed tests are appropriate.

## Procedure for significance testing

Whenever we perform a significance test, it involves comparing a test value that we have calculated to some critical value for the statistic. It doesn't matter what type of statistic we are calculating (e.g., a t-statistic, a chi-square statistic, an F-statistic, etc.), the procedure to test for significance is the same.

1. Decide on the *critical alpha level* you will use (i.e., the error rate you are willing to accept).
2. Conduct the research.
3. Calculate the statistic.
4. Compare the statistic to a *critical value* obtained from a table or compare the probability of the statistic to the *critical alpha level*.

If your statistic is higher than the *critical value* from the table or the probability of the statistic is less than the critical alpha level:

Your finding is significant.

You reject the null hypothesis.

The probability is small that the difference or relationship happened by chance, and  $p$  is less than the critical alpha level ( $p < \alpha$ ).

If your statistic is lower than the *critical value* from the table or the probability of the statistic is higher than the critical alpha level:

Your finding is not significant.

You fail to reject the null hypothesis.

The probability is high that the difference or relationship happened by chance, and  $p$  is greater than the critical alpha level ( $p > \alpha$ ).

Modern computer software can calculate exact probabilities for most test statistics. When the **Statistics Calculator** (or other software) gives you an exact probability, simply compare it to your critical alpha level. If the exact probability is less than the critical alpha level, your finding is significant, and if the exact probability is greater than your critical alpha level, your finding is not significant. Using a table is not necessary when you have the exact probability for a statistic.

## Bonferroni's theorem

Bonferroni's theorem states that as one performs an increasing number of statistical tests, the likelihood of getting an erroneous significant finding (Type I error) also increases. Thus, as we perform more and more statistical tests, it becomes increasingly likely that we will falsely reject a null hypothesis (very bad).

For example, suppose our critical alpha level is .05. If we performed one statistical test, our chance of making a false statement is .05. If we were to perform 100 statistical tests, and we made a statement about the result of each test, we would expect five of them to be wrong (just by fluke). This is a rather undesirable situation for social scientist.

Bonferroni's theorem states that we need to adjust the critical alpha level in order to compensate for the fact that we're doing more than one test. To make the adjustment, take the desired critical alpha level (e.g., .05) and divide by the number of tests being performed, and use the result as the critical alpha level. For example, suppose we had a test with eight scales, and we plan to compare males and females on each of the scales using an independent groups t-test. We would use  $.05/8$  as the critical alpha level for all eight tests.

Bonferroni's theorem should be applied whenever you are conducting two or more tests that are of the same "type" and the same "family". The same "type" means the same kind of statistical test. For example, if you were going to do one t-test, one ANOVA, and one regression, you would not make the adjustment because the tests are all different. The same "family" is a more elusive concept, and there are no hard and fast rules. "Family" refers to a series of statistical tests all designed to test the same (or very closely related) theoretical constructs. The bottom line is that it's up to the individual researcher to decide what constitutes a "family".

Some things are more obvious than others, for example, if you were doing t-tests comparing males and females on a series of questionnaire items that are all part of the same scale, you would probably apply the adjustment, by dividing your critical alpha level by the number of items in the scale (i.e., the number of t-tests you performed on that scale). The probabilities of the tests would be called the *family error rates*. However, suppose you have a series of independent questions, each focusing on a different construct, and you want to compare males and females on how they answered each question. Here is where the whole idea of Bonferroni's adjustment becomes philosophical. If you claim that each t-test that you perform is a test of a unique "mini"-hypothesis, then you would not use the adjustment, because you have defined each question as a different "family". In this case, the probability would be called a *statement error rate*. Another researcher might call the entire questionnaire a "family", and she would divide the critical alpha by the total number of items on the questionnaire.



Why stop there? From a statistician's perspective, the situation becomes even more complex. Since they are personally in the "statistics business", what should they call a "family"? When a statistician does a t-test for a client, maybe she should be dividing the critical alpha by the total number of t-tests that she has done in her life, since that is a way of looking at her "family". Of course, this would result in a different adjustment for each statistician--an interesting dilemma.

In the real world, most researchers do not use Bonferroni's adjustment because they would rarely be able to reject a null hypothesis. They would be so concerned about the possibility of making a false statement, that they would overlook many differences and relationships that actually exist. The "prime directive" for social science research is to discover relationships. One could argue that it is better to risk making a few wrong statements, than to overlook relationships or differences that are clear or prominent, but don't meet critical alpha significance level after applying Bonferroni's adjustment.

## Central tendency

The best known measures of central tendency are the mean and median. The mean average is found by adding the values for all the cases and dividing by the number of cases. For example, to find the mean age of all your friends, add all their ages together and divide by the number of friends. The mean average can present a distorted picture of central tendency if the sample is skewed in any way.

For example, let's say five people take a test. Their scores are 10, 12, 14, 18, and 94. (The last person is a genius.) The mean would be the sums of the scores  $10+12+14+18+94$  divided by 5. In this example, a mean of 29.6 is not a good measure of how well people did on the test in general. When analyzing data, be careful of using only the mean average when the sample has a few very high or very low scores. These scores tend to skew the shape of the distribution and will distort the mean.

When you have sampled from the population, the mean of the sample is also your best estimate of the mean of the population. The actual mean of the population is unknown, but the mean of the sample is as good an estimate as we can get.

The median provides a measure of central tendency such that half the sample will be above it and half the sample will be below it. For skewed distributions this is a better measure of central tendency. In the previous example, 14 would be the median for the sample of five people. If there is no middle value (i.e., there are an even number of data points), the median is the value midway between the two middle values.

The distribution of many variables follows that of a bell-shaped curve. This is called a "normal distribution". One must assume that data is approximately normally distributed for many statistical analyses to be valid. When a distribution is normal, the mean and median will be equal to each other. If they are not equal, the distribution is distorted in some way.

## Variability

Variability is synonymous with diversity. The more diversity there is in a set of data, the greater the variability. One simple measure of diversity is the range (maximum value minus the minimum value). The range is generally not a good measure of variability because it can be severely affected by a single very low or high value in the data. A better method of describing the amount of variability is to talk about the dispersion of scores away from the mean.

The variance and standard deviation are useful statistics that measure the dispersion of scores around the mean. The standard deviation is simply the square root of the variance. Both statistics measure the amount of diversity in the data. The higher the statistics, the greater the diversity. On the average, 68 percent of all the scores in a sample will be within plus or minus one standard deviation of the mean and 95 percent of all scores will be within two standard deviations of the mean.

There are two formulas for the variance and standard deviation of a sample. One set of formulas calculates the exact variance and standard deviation of the sample. The statistics are called *biased*, because they are biased to the sample. They are the exact variance and standard deviation of the sample, but they tend to underestimate the variance and standard deviation of the population.

Generally, we are more concerned with describing the population rather than the sample. Our intent is to use the sample to describe the population. The *unbiased estimates* should be used when sampling from the population and inferring back to the population. They provide the best estimate of the variance and standard deviation of the population.

## Standard error of the mean

The standard error of the mean is used to estimate the range within which we would expect the mean to fall in repeated samples taken from the population (i.e., confidence intervals). The standard error of the mean is an estimate of the standard deviation of those repeated samples.

The formula for the standard error of the mean provides an accurate estimate when the sample is very small compared to the size of the population. In marketing research, this is usually the case since the populations are quite large. However, when the sample size represents a substantial portion of the population, the formula becomes inaccurate and must be corrected. The finite population correction factor is used to correct the estimate of the standard error when the sample is more than ten percent of the population.

## Inferences with small sample sizes

When the sample size is small (less than 30), the  $z$  value for the area under the normal curve is not accurate. Instead of a  $z$  value, we can use a  $t$  value to derive the area under the curve. In fact, many researchers always use the  $t$  value instead of the  $z$  value. The reason is that the  $t$  values are more accurate for small sample sizes, and they are nearly identical to the  $z$  values for large sample sizes. Unlike the  $z$  value, the values for  $t$  depend upon the number of cases in the sample. Depending on the sample size, the  $t$  value will change.

## Degrees of freedom

Degrees of freedom literally refers to the number of data values that are free to vary.

For example, suppose I tell you that the mean of a sample is 10, and there are a total of three values in the sample. It turns out that if I tell you any two of the values, you will always be able to figure out the third value. If two of the values are 8 and 12, you can calculate that the third value is 10 using simple algebra.

$$(x + 8 + 12) / 3 = 10 \quad x = 10$$

In other words, if you know the mean, and all but one value, you can figure out the missing value. All the values except one are free to vary. One value is set once the others are known. Thus, degrees of freedom is equal to  $n-1$ .

## Choosing a significance test

The choice of which significance test you use depends upon the type of data you have collected. Use the following reference to help select an appropriate statistical test.

To explore the relationship between two categorical variables:

Chi-square

To explore the relationship between two ordinal scaled variables:

Chi-square

Spearman's rank-order correlation

To compare two percentages:

one-sample t-test between percents

two-sample t-test between percents

To compare two mean averages:

matched pairs t-test between means

independent groups t-test between means

To compare three or more variances:

one-way analysis of variance

To explore the relationship between two interval or ratio scaled variables:

Pearson's product-moment correlation

To explore the relationship between two interval or ratio scaled variables using one variable to predict the other:

simple regression

# Distributions

---

## Distributions menu

The **Distributions** menu selection is used to calculate critical values and probabilities for various distributions. The most common distributions are the  $z$  (normal) distribution,  $t$  distribution,  $F$  distribution, and the *chi-square* distribution. Within the last 20 years, computers have made it easy to calculate exact probabilities for the various statistics. Prior to that, researchers made extensive use of books containing probability tables.

The **Distributions** menu has four selections:

**Normal distribution**

**T distribution**

**F distribution**

**Chi-square distribution**

---

## Normal distribution

The normal distribution is the most well-known distribution and is often referred to as the  $z$  distribution or the bell shaped curve. It is used when the sample size is greater than 30. When the sample size is less than 30, the  $t$  distribution is used instead of the normal distribution.

The menu offers three choices: 1) probability of a  $z$  value, 2) critical  $z$  for a given probability, and 3) probability of a defined range.

### Probability of a $z$ value

When you have a  $z$  (standardized) value for a variable, you can determine the probability of that value. The software is the electronic equivalent of a normal distribution probability table. When you enter a  $z$  value, the area under the normal curve will be calculated. The area not under the curve is referred to as the *rejection region*. It is also called a *two-tailed* probability because both tails of the distribution are excluded. The **Statistics Calculator** reports the two-tailed probability for the  $z$  value. A one-tailed probability is used when your research question is concerned with only half of the distribution. Its value is exactly half the two-tailed probability.

### Example

z-value: 1.96

-----  
Two-tailed probability = .0500

## **Critical z for a given probability**

This menu selection is used to determine the critical z value for a given probability.

### Example

A large company designed a pre-employment survey to be administered to perspective employees. Baseline data was established by administering the survey to all current employees. They now want to use the instrument to identify job applicants who have very high or very low scores. Management has decided they want to identify people who score in the upper and lower 3% when compared to the norm. How many standard deviations away from the mean is required to define the upper and lower 3% of the scores?

The total area of rejection is 6%. This includes 3% who scored very high and 3% who scored very low. Thus, the two-tailed probability is .06. The z value required to reject 6% of the area under the curve is 1.881. Thus, new applicants who score higher or lower than 1.881 standard deviations away from the mean are the people to be identified.

Two tailed probability: .06

-----  
z-value = 1.881

## **Probability of a defined range**

Knowing the mean and standard deviation of a sample allows you to establish the area under the curve for any given range. This menu selection will calculate the probability that the mean of a new sample would fall between two specified values (i.e., between the limits of a defined range).

### Example

A manufacturer may find that the emission level from a device is 25.9 units with a standard deviation of 2.7. The law limits the maximum emission level to 28.0 units. The manufacturer may want to know what percent of the new devices coming off the assembly line will need to be rejected because they exceed the legal limit.

Sample mean = 25.9

Unbiased standard deviation = 2.7

Lower limit of the range = 0

Upper limit of the range = 28.0

-----  
Probability of a value falling within the range = .7817

Probability of a value falling outside the range = .2183

The area under the curve is the sum of the area defined by the lower limit plus the area defined by the upper limit.

The area under the normal curve is the probability that additional samples would fall between the lower and upper limits. In this case, the area above the upper limit is the rejection area (21.83% of the product would be rejected).

---

## T distribution

Mathematicians used to think that all distributions followed the bell shaped curve. In the early 1900's, an Irish chemist named Gosset, discovered that distributions were much flatter than the bell shaped curve when working with small sample sizes. In fact, the smaller the sample, the flatter the distribution. The  $t$  distribution is used instead of the normal distribution when the sample size is small. As the sample size approaches thirty, the  $t$  distribution approximates the normal distribution. Thus, the  $t$  distribution is generally used instead of the  $z$  distribution, because it is correct for both large and small sample sizes, where the  $z$  distribution is only correct for large samples.

The menu offers three choices: 1) probability of a  $t$  value, 2) critical  $t$  value for a given probability, and 3) probability of a defined range.

### Probability of a $t$ value

If you have a  $t$  value and the degrees of freedom associated with the value, you can use this program to calculate the two-tailed probability of  $t$ . It is the equivalent of computerized table of  $t$  values.

#### Example

t-value: 2.228

df: 10

-----  
Two-tailed probability = .050

### Critical $t$ value for a given probability

This program is the opposite of the previous program. It is used if you want to know what critical  $t$  value is required to achieve a given probability.

#### Example

Two-tailed probability: .050

Degrees of freedom: 10

-----  
t-value = 2.228

## Probability of a defined range

Knowing the mean and standard deviation of a sample allows you to establish the area under the curve for any given range. You can use this program to calculate the probability that the mean of a new sample would fall between two values.

### Example

A company did a survey of 20 people who used its product. The mean average age of the sample was 22.4 years and the unbiased standard deviation was 3.1 years. The company now wants to advertise in a magazine that has a primary readership of people who are between 18 and 24, so they need to know what percent of its potential customers are between 18 and 24 years of age?

Sample mean: 22.4

Unbiased standard deviation: 3.1

Sample size = 20

Lower limit of the range = 18

Upper limit of the range = 24

-----  
Probability of a value falling within the range = .608

Probability of a value falling outside the range = .392

Because of the small sample size, the *t* distribution is used instead of the *z* distribution. The area under the curve represents the proportion of customers in the population expected to be between 18 and 24 years of age. In this example, we would predict that 60.8% of its customers would be expected to be between 18 and 24 years of age, and 39.2% would be outside of the range. The company decided not to advertise.

---

## F distribution

The *F*-ratio is used to compare variances of two or more samples or populations. Since it is a ratio (i.e., a fraction), there are degrees of freedom for the numerator and denominator. This menu selection may be used to calculate the probability of an *F*-ratio or to determine the critical value of *F* for a given probability. These menu selections are the computer equivalent of an *F* table.

### Probability of a *F*-ratio

If you have a *F*-ratio and the degrees of freedom associated with the numerator and denominator, you can use this program to calculate the probability.

#### Example

*F*-ratio: 2.774

Numerator degrees of freedom: 20

Denominator degrees of freedom: 10

-----  
Two-tailed probability = .0500

### **Critical $F$ for a given probability**

If you know the critical alpha level and the degrees of freedom associated with the numerator and denominator, you can use this program to calculate the  $F$ -ratio.

#### **Example**

Two-tailed probability = .0500

Numerator degrees of freedom: 20

Denominator degrees of freedom: 10

-----  
F-ratio: 2.774

---

## **Chi-square distribution**

The chi-square statistic is used to compare the observed frequencies in a table to the expected frequencies. This menu selection may be use to calculate the probability of a chi-square statistic or to determine the critical value of chi-square for a given probability. This menu selection is the computer equivalent of an chi-square table.

### **Probability of a chi-square statistic**

If you have a chi-square value and the degrees of freedom associated with the value, you can use this program to calculate the probability of the chi-square statistic. It is the equivalent of computerized table of chi-square values.

#### **Example**

Chi-square value: 18.307

Degrees of freedom: 10

-----  
Probability = .050

### **Critical chi-square for a given probability**

If you have the critical alpha level and the degrees of freedom, you can use this program to calculate the probability of the chi-square statistic. It is the equivalent of computerized table of chi-square values.

#### **Example**

Probability = .0500

Degrees of freedom: 10

-----  
Chi-square value: 18.307





# Counts

---

## Counts menu

The **Counts** menu selection has four tests that can be performed for simple frequency data. The chi-square test is used to analyze a contingency table consisting of rows and columns to determine if the observed cell frequencies differ significantly from the expected frequencies. Fisher's exact test is similar to the chi-square test except it is used only for tables with exactly two rows and two columns. The binomial test is used to calculate the probability of two mutually exclusive outcomes. The poisson distribution events test is used to describe the number of events that will occur in a specific period of time.

The **Counts** menu has four selections:

**Chi-square test**

**Fisher's exact test**

**Binomial test**

**Poisson distribution events test**

---

## Chi-square test

The chi-square is one of the most popular statistics because it is easy to calculate and interpret. There are two kinds of chi-square tests. The first is called a one-way analysis, and the second is called a two-way analysis. The purpose of both is to determine whether the observed frequencies (counts) markedly differ from the frequencies that we would expect by chance.

The observed cell frequencies are organized in rows and columns like a spreadsheet. This table of observed cell frequencies is called a *contingency table*, and the chi-square test is part of a *contingency table analysis*.

The chi-square statistic is the sum of the contributions from each of the individual cells. Every cell in a table contributes something to the overall chi-square statistic. If a given cell differs markedly from the expected frequency, then the contribution of that cell to the overall chi-square is large. If a cell is close to the expected frequency for that cell, then the contribution of that cell to the overall chi-square is low. A large chi-square statistic indicates that somewhere in the table, the observed frequencies differ markedly from the expected frequencies. It does not tell which cell (or cells) are causing the high chi-square...only that they are there. When a chi-square is high, you must visually examine the table to determine which cell(s) are responsible.

When there are exactly two rows and two columns, the chi-square statistic becomes inaccurate, and Yate's correction for continuity is usually applied. In the **Statistics Calculator**, Yate's correction will automatically be applied to two-by-two tables if any cell has an expected frequency less than 5 or the total N is less than 50. The result will be shown as the "Corrected Chi-Square". Regardless of whether the Yate's correction checkbox is checked or not, the correction will only be applied to two-by-two tables.

If there is only one column or one row (a one-way chi-square test), the degrees of freedom is the number of cells minus one. For a two way chi-square, the degrees of freedom is the number of rows minus one times the number of columns minus one.

Using the chi-square statistic and its associated degrees of freedom, the software reports the probability that the differences between the observed and expected frequencies occurred by chance. Generally, a probability of .05 or less is considered to be a significant difference.

A standard spreadsheet interface is used to enter the counts for each cell. After you've finished entering the data, the program will print the chi-square, degrees of freedom and probability of chance.

Use caution when interpreting the chi-square statistic if any of the cell frequencies are less than five. Also, use caution when the total for all cells is less than 50.

### Example

A drug manufacturing company conducted a survey of customers. The research question is: Is there a significant relationship between packaging preference (size of the bottle purchased) and economic status? There were four packaging sizes: small, medium, large, and jumbo. Economic status was: lower, middle, and upper. The following data was collected.

	lower	middle	upper
small	24	22	18
medium	23	28	19
large	18	27	29
jumbo	16	21	33

-----  
Chi-square statistic = 9.743

Degrees of freedom = 6

Probability of chance = .1359

---

## Fisher's exact test

The chi-square statistic becomes inaccurate when used to analyze contingency tables that contain exactly two rows and two columns, and that contain less than 50 cases. Fisher's exact probability is not plagued by inaccuracies due to small N's. Therefore, it should be used for two-by-two contingency tables that contain fewer than 50 cases. The probability will be similar to the chi-square test with Yate's correction.

### Example

Here are the results of a recent public opinion poll broken down by gender. What is the exact probability that the difference between the observed and expected frequencies occurred by chance?

	Male	Female
Favor	2	3
Opposed	1	33

-----  
Fisher's exact probability = .0372

---

## Binomial test

The binomial distribution is used for calculating the probability of dichotomous outcomes in which the two choices are mutually exclusive. The program requires that you enter the number of trials, probability of the desired outcome on each trial, and the number of times the desired outcome was observed.

### Example

If we were to flip a coin one hundred times, and it came up heads seventy times, what is the probability of this happening?

Number of trials: 100

Probability of success on each trial (0-1): .5

Number of successes: 70

-----  
Probability of 70 or more successes < .0001

---

## Poisson distribution events test

The poisson distribution, like the binomial distribution, is used to determine the probability of an observed frequency. It is used to describe the number of events that will occur in a specific period of time or in a specific area or volume. You need to enter the observed and expected frequencies.

### Example

Previous research on a particular assembly line has shown that they have an average daily defect rate of 39 products. Thus, the expected number of defective products expected on any day is 39. The day after implementing a new quality control program, they found only 25 defects. What is the probability of seeing 25 or fewer defects on any day?

Observed frequency: 25

Expected frequency: 39

-----  
Probability of 25 or fewer events = .0226



# Percents

---

## Percents menu

Percents are understood by nearly everyone, and therefore, they are the most popular statistics cited in research. Researchers are often interested in comparing two percentages to determine whether there is a significant difference between them.

The **Percents** menu has three selections:

**One sample t-test between percents**

**Two sample t-test between percents**

**Confidence interval around a percent**

### Choosing the proper test

There are two kinds of t-tests between percents. Which test you use depends upon whether you're comparing percentages from one or two samples.

Every percentage can be expressed as a fraction. By looking at the denominator of the fraction we can determine whether to use a one-sample or two-sample t-test between percents. If the denominators used to calculate the two percentages represent the same people, we use a one-sample t-test between percents to compare the two percents. If the denominators represent different people, we use the two-sample t-test between percents.

For example suppose you did a survey of 200 people. Your survey asked,

*Were you satisfied with the program?*

*Yes*    *No*    *Don't know*

Of the 200 people, 80 said *yes*, 100 said *no*, and 20 didn't know. You could summarize the responses as:

Yes                     $80/200 = .4 = 40\%$

No                      $100/200 = .5 = 50\%$

Don't know          $20/200 = .1 = 10\%$

Is there a significant difference between the percent saying *yes* (40%) and the percent saying *no* (50%)? Obviously, there is a difference; but how sure are we that the difference didn't just happen by chance? In other words, how reliable is the difference?

Notice that the denominator used to calculate the percent of *yes* responses (200) represents the same people as the denominator used to calculate the percent of *no*

responses (200). Therefore, we use a one-sample t-test between proportions. The key is that the denominators represent the same people (not that they are the same number).

After you completed your survey, another group of researchers tried to replicate your study. They also used a sample size of 200, and asked the identical question. Of the 200 people in their survey, 60 said yes, 100 said no, and 40 didn't know. They summarized their results as:

Yes	$60/200 = .3 = 30\%$
No	$100/200 = .5 = 50\%$
Don't know	$40/200 = .2 = 20\%$

Is there a significant difference between the percent who said yes in your survey (40%) and the percent that said yes in their survey (30%)? For your survey the percent that said yes was calculated as  $80/200$ , and in their survey it was  $60/200$ . To compare the yes responses between the two surveys, we would use a two-sample t-test between percents. Even though both denominators were 200, they do not represent the same 200 people.

### **Examples that would use a one-sample t-test**

Which proposal would you vote for?

Proposal A     Proposal B

Which product do you like better?

Name Brand     Brand X

Which candidate would you vote for?

Johnson     Smith     Anderson

When there are more than two choices, you can do the t-test between any two of them. In this example, there are three possible combinations: Johnson/Smith, Johnson/Anderson, and Smith/Anderson. Thus, you could actually perform three separate t-tests...one for each pair of candidates. If this was your analysis plan, you would also use Bonferroni's theorem to adjust the critical alpha level because the plan involved multiple tests of the same type and family.

### **Examples that would use a two-sample t-test**

A previous study found that 39% of the public believed in gun control. Your study found the 34% believed in gun control. Are the beliefs of your sample different than those of the previous study?

The results of a magazine readership study showed that 17% of the women and 11% of the men recalled seeing your ad in the last issue. Is there a significant difference between men and women?

In a brand awareness study, 25% of the respondents from the Western region had heard of your product. However, only 18% of the respondents from the Eastern region had heard of your product. Is there a significant difference in product awareness between the Eastern and Western regions?

---

## One sample t-test between percents

This test can be performed to determine whether respondents are more likely to prefer one alternative or another.

### Example

The research question is: Is there a significant difference between the percent of people who say they would vote for candidate A and the percent of people who say they will vote for candidate B? The null hypothesis is: There is no significant difference between the percent of people who say they will vote for candidate A or candidate B. The results of the survey were:

Plan to vote for candidate A = 35.5%

Plan to vote for candidate B = 22.4%

Sample size = 107

The sum of the two percents does not have to be equal to 100 (there may be candidates C and D, and people that have no opinion). Use a one-sample t-test because both percentages came from a single sample.

Use a two-tailed probability because the null hypothesis does not state the direction of the difference. If the hypothesis is that one particular choice has a greater percentage, use a one-tailed test (divide the two-tailed probability by two).

Enter the first percent: 35.5

Enter the second percent: 22.4

Enter the sample size: 107

-----  
t-value = 1.808

Degrees of freedom = 106

Two-tailed probability = .074

You might make a statement in a report like this: A one-sample t-test between proportions was performed to determine whether there was a significant difference between the percent choosing candidate A and candidate B. The t-statistic was not significant at the .05 critical alpha level,  $t(106)=1.808$ ,  $p=.073$ . Therefore, we fail to reject the null hypothesis and conclude that the difference was not significant.

---

## Two sample t-test between percents

This test can be used to compare percentages drawn from two independent samples. It can also be used to compare two subgroups from a single sample.

### Example



After conducting a survey of customers, you want to compare the attributes of men and women. Even though all respondents were part of the same survey, the men and women are treated as two samples. The percent of men with a particular attribute is calculated using the total number of men as the denominator for the fraction. And the percent of women with the attribute is calculated using the total number of women as the denominator. Since the denominators for the two fractions represent different people, a two-sample t-test between percents is appropriate.

The research question is: Is there a significant difference between the proportion of men having the attribute and the proportion of women having the attribute? The null hypothesis is: There is no significant difference between the proportion of men having the attribute and the proportion of women having the attribute. The results of the survey were:

86 men were surveyed and 22 of them (25.6%) had the attribute.

49 women were surveyed and 19 of them (38.8%) had the attribute.

Enter the first percent: 25.6

Enter the sample size for the first percent: 86

Enter the second percent: 38.8

Enter the sample size for the second percent: 49

-----  
t-value = 1.603

Degrees of freedom = 133

Two-tailed probability = .111

You might make a statement in a report like this: A two-sample t-test between proportions was performed to determine whether there was a significant difference between men and women with respect to the percent who had the attribute. The t-statistic was not significant at the .05 critical alpha level,  $t(133)=1.603$ ,  $p=.111$ . Therefore, we fail to reject the null hypothesis and conclude that the difference between men and women was not significant.

### **Another example**

Suppose interviews were conducted at two different shopping centers. This two sample t-test between percents could be used to determine if the responses from the two shopping centers were different.

The research question is: Is there a significant difference between shopping centers A and B with respect to the percent that say they would buy product X? The null hypothesis is: There is no significant difference between shopping centers A and B with respect to the percent of people that say they would buy product X. A two-tailed probability will be used because the hypothesis does not state the direction of the difference. The results of the survey were:

89 people were interviewed at shopping center A and 57 of them (64.0%) said they would buy product X.

92 people were interviewed at shopping center B and 51 of them (55.4%) said they would buy product X.

Enter the first percent: 64.0

Enter the sample size for the first percent: 89

Enter the second percent: 55.4

Enter the sample size for the second percent: 92

-----

t-value = 1.179

Degrees of freedom = 179

Two-tailed probability = .240

You might write a paragraph in a report like this: A two-sample t-test between proportions was performed to determine whether there was a significant difference between the two shopping centers with respect to the percent who said they would buy product X. The t-statistic was not significant at the .05 critical alpha level,  $t(179)=1.179$ ,  $p=.240$ . Therefore, we fail to reject the null hypothesis and conclude that the difference in responses between the two shopping centers was not significant.

---

## Confidence intervals around a percent

Confidence intervals are used to determine how much latitude there is in the range of a percent if we were to take repeated samples from the population.

### Example

In a study of 150 customers, you find that 60 percent have a college degree. Your best estimate of the percent who have a college degree in the population of customers is also 60 percent. However, since it is just an estimate, we establish confidence intervals around the estimate as a way of showing how reliable the estimate is.

Confidence intervals can be established for any error rate you are willing to accept. If, for example, you choose the 95% confidence interval, you would expect that in five percent of the samples drawn from the population, the percent who had a college degree would fall outside of the interval.

What are the 95% confidence intervals around this percent? In the following example, note that no value is entered for the population size. When the population is very large compared to the sample size (as in most research), it is not necessary to enter a population size. If, however, the sample represents more than ten percent of the population, the formulas incorporate a finite population correction adjustment. Thus, you only need to enter the population size when the sample size exceeds ten percent of the population size.

Enter the percent: 60

Enter the sample size: 150

Enter the population size: (left blank)

Enter the desired confidence interval (%): 95

-----

Standard error of the proportion = .040

Degrees of freedom = 149

95% confidence interval = 60.0%  $\pm$  7.9%

Confidence interval range = 52.1% to 67.9%

Therefore, our best estimate of the population proportion with 5% error is 60%  $\pm$  7.9%. Stated differently, if we predict that the proportion in the population who have a college degree is between 52.1% and 67.9%, our prediction would be wrong for 5% of the samples that we draw from the population.

# Means

---

## Means menu

Researchers usually use the results from a sample to make inferential statements about the population. When the data is interval or ratio scaled, it usually described in terms of central tendency and variability. Means and standard deviations are usually reported in all research.

The **Means** menu has seven selections:

**Mean and standard deviation of a sample**

**Matched pairs t-test between means**

**Independent groups t-test between means**

**Confidence intervals around a mean**

**Compare a sample mean to a population mean**

**Compare two standard deviations**

**Compare three or more means**

---

## Mean and standard deviation of a sample

This menu selection will let you enter data for a variable and calculate the mean, unbiased standard deviation, standard error of the mean, and median. Data is entered using a standard spreadsheet interface. Finite population correction is incorporated into the calculation of the standard error of the mean, so the population size should be specified whenever the sample size is greater than ten percent of the population size.

### Example

A sample of ten was randomly chosen from a large population. The ten scores were:

20 22 54 32 41 43 47 51 45 35

-----  
Mean = 39.0

Unbiased standard deviation = 11.6

Standard error of the mean = 3.7

Median = 42.0

---

## Matched pairs t-test between means

The matched pairs t-test is used in situations where two measurements are taken for each respondent. It is often used in experiments where there are before-treatment and after-treatment measurements. The t-test is used to determine if there is a reliable difference between the mean of the before-treatment and the mean of the after treatment measurements.

Pretreatment                      Posttreatment

Johnny ----- Johnny

Martha ----- Martha

Jenny ----- Jenny

Sometimes, in very sophisticated (i.e., expensive) experiments, two groups of subjects are individually matched on one or more demographic characteristics. One group is exposed to a treatment (experimental group) and the other is not (control group).

Experimental                      Control

Johnny ----- Fred

Martha ----- Sharon

Jenny ----- Linda

The t-test works with small or large N's because it automatically takes into account the number of cases in calculating the probability level. The magnitude of the t-statistic depends on the number of cases (subjects). The t-statistic in conjunction with the degrees of freedom are used to calculate the probability that the difference between the means happened by chance. If the probability is less than the critical alpha level, then we say that a significant difference exists between the two means.

### Example

A example of a matched-pairs t-test might look like this:

Pretest	Posttest
8	31
13	37
22	45
25	28
29	50
31	37
35	49
38	25
42	36
52	69

-----  
Var.1: Mean = 29.5 Unbiased SD = 13.2

Var. 2: Mean = 40.7 Unbiased SD = 13.0

t-statistic = 2.69

Degrees of freedom = 9

Two-tailed probability = .025

You might make a statement in a report like this: The mean pretest score was 29.5 and the mean posttest score was 40.7. A matched-pairs t-test was performed to determine if the difference was significant. The t-statistic was significant at the .05 critical alpha level,  $t(9)=2.69$ ,  $p=.025$ . Therefore, we reject the null hypothesis and conclude that posttest scores were significantly higher than pretest scores.

---

## Independent groups t-test between means

This menu selection is used to determine if there is a difference between two means taken from different samples. If you know the mean, standard deviation and size of both samples, this program may be used to determine if there is a reliable difference between the means.

One measurement is taken for each respondent. Two groups are formed by splitting the data based on some other variable. The groups may contain a different number of cases. There is not a one-to-one correspondence between the groups.

<u>Score</u>	<u>Sex</u>		<u>Males</u>	<u>Females</u>
25	M		25	27
27	F	-----becomes---->	19	17
17	F			21
19	M			
21	F			

Sometimes the two groups are formed because the data was collected from two different sources.

<u>School A Scores</u>	<u>School B Scores</u>
525	427
492	535
582	600
554	
520	

There are actually two different formulas to calculate the  $t$ -statistic for independent groups. The  $t$ -statistics calculated by both formulas will be similar but not identical.

Which formula you choose depends on whether the variances of the two groups are equal or unequal. In actual practice, most researchers assume that the variances are unequal because it is the most conservative approach and is least likely to produce a Type I error. Thus, the formula used in **Statistics Calculator** assumes unequal variances.

### Example

Two new product formulas were developed and tested. A twenty-point scale was used to measure the level of product approval. Six subjects tested the first formula. They gave it a mean rating of 12.3 with a standard deviation of 1.4. Nine subjects tested the second formula, and they gave it a mean rating of 14.0 with a standard deviation of 1.7. The question we might ask is whether the observed difference between the two formulas is reliable.

Mean of the first group: 12.3

Unbiased standard deviation of the first group: 1.4

Sample size of the first group: 6

-----  
Mean of the second group: 14.0

Unbiased standard deviation of the second group: 1.7

Sample size of the second group: 9

-----  
t value = 2.03

Degrees of freedom = 13

Two-tailed probability = .064

You might make a statement in a report like this: An independent groups t-test was performed to compare the mean ratings between the two formulas. The t-statistic was not significant at the .05 critical alpha level,  $t(13)=2.03$ ,  $p=.064$ . Therefore, we fail to reject the null hypothesis and conclude that there was no significant difference between the ratings for the two formulas.

---

## Confidence interval around a mean

You can calculate confidence intervals around a mean if you know the sample size and standard deviation.

The standard error of the mean is estimated from the standard deviation and the sample size. It is used to establish the confidence interval (the range within which we would expect the mean to fall in repeated samples taken from the population). The standard error of the mean is an estimate of the standard deviation of those repeated samples.

The formula for the standard error of the mean provides an accurate estimate when the sample size is very small compared to the size of the population. In marketing research, this is usually the case since the populations are quite large. Thus, in most situations the population size may be left blank because the population is very large compared to the sample. However, when the sample is more than ten percent of the

population, the population size should be specified so that the finite population correction factor can be used to adjust the estimate of the standard error of the mean.

**Example**

Suppose that an organization has 5,000 members. Prior to their membership renewal drive, 75 members were randomly selected and surveyed to find out their priorities for the coming year. The mean average age of the sample was 53.1 and the unbiased standard deviation was 4.2 years. What is the 90% confidence interval around the mean? Note that the population size can be left blank because the sample size of 75 is less than ten percent of the population size.

Mean: 53.1

Unbiased standard deviation: 4.2

Sample size: 75

Population size: (left blank -or- 5000)

Desired confidence interval (%): 90

-----  
Standard error of the mean = .485

Degrees of freedom = 74

90% confidence interval = 53.1 ± .8

Confidence interval range = 52.3 - 53.9

---

## Compare a sample mean to a population mean

Occasionally, the mean of the population is known (perhaps from a previous census). After drawing a sample from the population, it might be helpful to compare the mean of your sample to the mean of the population. If the means are not significantly different from each other, you could make a strong argument that your sample provides an adequate representation of the population. If, however, the mean of your sample is significantly different than the population, something may have gone wrong during the sampling process.

**Example**

After selecting a random sample of 18 people from a very large population, you want to determine if the average age of the sample is representative of the average age of the population. From previous research, you know that the mean age of the population is 32.0. For your sample, the mean age was 28.0 and the unbiased standard deviation was 3.2. Is the mean age of your sample significantly different from the mean age in the population?

Sample mean = 28

Unbiased standard deviation = 3.2

Sample size = 18

Population size = (left blank)

Mean of the population = 32



-----  
Standard error of the mean = .754

t value = 5.303

Degrees of freedom = 17

Two-tailed probability = .0001

The two-tailed probability of the t-statistic is very small. Thus, we would conclude that the mean age of our sample is significantly less than the mean age of the population. This could be a serious problem because it suggests that some kind of age bias was inadvertently introduced into the sampling process. It would be prudent for the researcher to investigate the problem further.

---

## Compare two standard deviations

The F-ratio is used to compare variances. In its simplest form, it is the variance of one group divided by the variance of another group. When used in this way, the larger variance (by convention) is the numerator and the smaller is the denominator. Since the groups might have a different sample sizes, the numerator and the denominator have their own degrees of freedom.

### Example

Two samples were taken from the population. One sample had 25 subjects and the standard deviation 4.5 on some key variable. The other sample had 12 subjects and had a standard deviation of 6.4 on the same key variable. Is there a significant difference between the variances of the two samples?

First standard deviation: 4.5

First sample size: 25

Second standard deviation: 6.4

Second sample size: 12

-----  
F-ratio = 2.023

Degrees of freedom = 11 and 24

Probability that the difference was due to chance = .072

---

## Compare three or more means

Analysis of variance (ANOVA) is used when testing for differences between three or more means.

In an ANOVA, the F-ratio is used to compare the variance between the groups to the variance within the groups. For example, suppose we have two groups of data. In the best of all possible worlds, all the people in group one would have very similar scores. That is, the group is cohesive, and there would be very little variability in scores within the group. All the people in group two would also have similar scores (although different than group one). Again, there is very little variability within the group. Both groups have very little variability within their group, however, there

might be substantial variability between the groups. The ratio of the between groups variability (numerator) to the within groups variability (denominator) is the F-ratio. The larger the F-ratio, the more certain we are that there is a difference between the groups.

If the probability of the F-ratio is less than or equal to your critical alpha level, it means that there is a significant difference between at least two of groups. The F-ratio does not tell which group(s) are different from the others...just that there is a difference.

After finding a significant F-ratio, we do "post-hoc" (after the fact) tests on the factor to examine the differences between levels. There are a wide variety of post-hoc tests, but one of the most common is to do a series of special t-tests between all the combinations of levels for that factor. **Statistics Calculator** uses lsd (least significant difference) t-tests for all post-hoc comparisons. For the post-hoc tests, use the same critical alpha level that you used to test for the significance of the F-ratio.

### Example

A company has offices in four cities with sales representatives in each office. At each location, the average number of sales per salesperson was calculated. The company wants to know if there are significant differences between the four offices with respect to the average number of sales per sales representative.

Group	Mean	SD	N
1	3.29	1.38	7
2	4.90	1.45	10
3	7.50	1.38	6
4	6.00	1.60	8

Source	df	SS	MS	F	p
Factor	3	62.8	20.9	9.78	.0002
Error	27	57.8	2.13		
Total	30	120.6			

### Post-hoc t-tests

Group	Group	t-value	df	p
1	2	2.23	15	.0412
1	3	5.17	11	.0003
1	4	3.58	13	.0034
2	3	3.44	14	.0040
2	4	1.59	16	.1325
3	4	1.09	12	.0019



# Correlation

---

## Correlation types

Correlation is a measure of association between two variables. The variables are not designated as dependent or independent. The two most popular correlation coefficients are: Spearman's correlation coefficient rho and Pearson's product-moment correlation coefficient.

When calculating a correlation coefficient for ordinal data, select Spearman's technique. For interval or ratio-type data, use Pearson's technique.

The value of a correlation coefficient can vary from minus one to plus one. A minus one indicates a perfect negative correlation, while a plus one indicates a perfect positive correlation. A correlation of zero means there is no relationship between the two variables. When there is a negative correlation between two variables, as the value of one variable increases, the value of the other variable decreases, and vice versa. In other words, for a negative correlation, the variables work opposite each other. When there is a positive correlation between two variables, as the value of one variable increases, the value of the other variable also increases. The variables move together.

The standard error of a correlation coefficient is used to determine the confidence intervals around a true correlation of zero. If your correlation coefficient falls outside of this range, then it is significantly different than zero. The standard error can be calculated for interval or ratio-type data (i.e., only for Pearson's product-moment correlation).

The significance (probability) of the correlation coefficient is determined from the t-statistic. The probability of the t-statistic indicates whether the observed correlation coefficient occurred by chance if the true correlation is zero. In other words, it asks if the correlation is significantly different than zero. When the t-statistic is calculated for Spearman's rank-difference correlation coefficient, there must be at least 30 cases before the t-distribution can be used to determine the probability. If there are fewer than 30 cases, you must refer to a table to find the probability of the correlation coefficient.

### Example

A company wanted to know if there is a significant relationship between the total number of salespeople and the total number of sales. They collect data for five months.

Var. 1    Var. 2

207 6907  
180 5991  
220 6810  
205 6553  
190 6190

-----  
Correlation coefficient = .921

Standard error of the coefficient = .068

t-test for the significance of the coefficient = 4.100

Degrees of freedom = 3

Two-tailed probability = .0263

### **Another Example**

Respondents to a survey were asked to judge the quality of a product on a four-point Likert scale (excellent, good, fair, poor). They were also asked to judge the reputation of the company that made the product on a three-point scale (good, fair, poor). Is there a significant relationship between respondents perceptions of the company and their perceptions of quality of the product?

Since both variables are ordinal, Spearman's method is chosen. The first variable is the rating for the quality the product. Responses are coded as 4=excellent, 3=good, 2=fair, and 1=poor. The second variable is the perceived reputation of the company and is coded 3=good, 2=fair, and 1=poor.

Var. 1	Var. 2
4	3
2	2
1	2
3	3
4	3
1	1
2	1

-----  
Correlation coefficient rho = .830

t-test for the significance of the coefficient = 3.332

Number of data pairs = 7

Probability must be determined from a table because of the small sample size.

---

## **Regression**

Simple regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can

be used to predict the dependent variable when the independent variable is known. Regression goes beyond correlation by adding prediction capabilities.

People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children's diet results in higher energy levels. The ease of waking up in the morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

For example, a medical researcher might want to use body weight (independent variable) to predict the most appropriate dose for a new drug (dependent variable). The purpose of running the regression is to find a formula that fits the relationship between the two variables. Then you can use that formula to predict values for the dependent variable when only the independent variable is known. A doctor could prescribe the proper dose based on a person's body weight.

The regression line (known as the *least squares line*) is a plot of the expected value of the dependent variable for all values of the independent variable. Technically, it is the line that "minimizes the squared residuals". The regression line is the one that best fits the data on a scatterplot.

Using the regression equation, the dependent variable may be predicted from the independent variable. The slope of the regression line (b) is defined as the rise divided by the run. The y intercept (a) is the point on the y axis where the regression line would intercept the y axis. The slope and y intercept are incorporated into the regression equation. The intercept is usually called the constant, and the slope is referred to as the coefficient. Since the regression model is usually not a perfect predictor, there is also an error term in the equation.

In the regression equation, y is always the dependent variable and x is always the independent variable. Here are three equivalent ways to mathematically describe a linear regression model.

$$y = \text{intercept} + (\text{slope} \cdot x) + \text{error}$$

$$y = \text{constant} + (\text{coefficient} \cdot x) + \text{error}$$

$$y = a + bx + e$$

The significance of the slope of the regression line is determined from the t-statistic. It is the probability that the observed correlation coefficient occurred by chance if the true correlation is zero. Some researchers prefer to report the F-ratio instead of the t-statistic. The F-ratio is equal to the t-statistic squared.

The t-statistic for the significance of the slope is essentially a test to determine if the regression model (equation) is usable. If the slope is significantly different than zero, then we can use the regression model to predict the dependent variable for any value of the independent variable.

On the other hand, take an example where the slope is zero. It has no prediction ability because for every value of the independent variable, the prediction for the dependent variable would be the same. Knowing the value of the independent variable would not improve our ability to predict the dependent variable. Thus, if the slope is not significantly different than zero, don't use the model to make predictions.

The coefficient of determination (r-squared) is the square of the correlation coefficient. Its value may vary from zero to one. It has the advantage over the correlation coefficient in that it may be interpreted directly as the proportion of variance in the dependent variable that can be accounted for by the regression equation. For example, an r-squared value of .49 means that 49% of the variance in

the dependent variable can be explained by the regression equation. The other 51% is unexplained.

The standard error of the estimate for regression measures the amount of variability in the points around the regression line. It is the standard deviation of the data points as they are distributed around the regression line. The standard error of the estimate can be used to develop confidence intervals around a prediction.

### **Example**

A company wants to know if there is a significant relationship between its advertising expenditures and its sales volume. The independent variable is advertising budget and the dependent variable is sales volume. A lag time of one month will be used because sales are expected to lag behind actual advertising expenditures. Data was collected for a six month period. All figures are in thousands of dollars. Is there a significant relationship between advertising budget and sales volume?

IV	DV
4.2	27.1
6.1	30.4
3.9	25.0
5.7	29.7
7.3	40.1
5.9	28.8

-----  
Model:  $y = 10.079 + (3.700 \cdot x) + \text{error}$

Standard error of the estimate = 2.568

t-test for the significance of the slope = 4.095

Degrees of freedom = 4

Two-tailed probability = .0149

r-squared = .807

You might make a statement in a report like this: A simple linear regression was performed on six months of data to determine if there was a significant relationship between advertising expenditures and sales volume. The t-statistic for the slope was significant at the .05 critical alpha level,  $t(4)=4.10$ ,  $p=.015$ . Thus, we reject the null hypothesis and conclude that there was a positive significant relationship between advertising expenditures and sales volume. Furthermore, 80.7% of the variability in sales volume could be explained by advertising expenditures.

# Sampling

---

## Sampling menu

The formula to determine sample size depends upon whether the intended comparisons involve means or percents.

The **Sampling** menu has two selections:

Sample size for percents

Sample size for means

---

## Sample size for percents

This menu selection is used to determine the required size of a sample for research questions involving percents.

Four questions must be answered to determine the sample size:

1. Best estimate of the population size: You do not need to know the exact size of the population. Simply make your best estimate. An inaccurate population size will not seriously affect the formula computations. If the population is very large, this item may be left blank.
2. Best estimate of the rate in the population (%): Make your best estimate of what the actual percent of the survey characteristic is. This is based on the null hypothesis. For example, if the null hypothesis is "blondes don't have more fun", then what is your best estimate of the percent of blondes that do have more fun? If you simply do not know, then enter 50 (for fifty percent).
3. Maximum acceptable difference (%): This is the maximum percent difference that you are willing to accept between the true population rate and the sample rate. Typically, in social science research, you would be willing to accept a difference of 5 percent. That is, if your survey finds that 25 percent of the sample has a certain characteristic, the actual rate in the population may be between 20 and 30 percent.
4. Desired confidence level (%): How confident must you be that the true population rate falls within the acceptable difference (specified in the previous question)? This is the same as the confidence that you want to have in your findings. If you want 95 percent confidence (typical for social science research), you should enter 95. This means that if you took a hundred samples from the population, five of those samples



would have a rate that exceeded the difference you specified in the previous question.

**Example**

A publishing wants to know what percent of the population might be interested in a new magazine on making the most of your retirement. Secondary data (that is several years old) indicates that 22% of the population is retired. They are willing to accept an error rate of 5% and they want to be 95% certain that their finding does not differ from the true rate by more than 5%. What is the required sample size?

Best estimate of the population size: (left blank)

Best estimate of the rate in the population (%): 22

Maximum acceptable difference (%): 5

Desired confidence level (%): 95

-----  
Required sample size = 263

---

## Sample size for means

This menu selection is used to determine the required size of a sample for research questions involving means.

Three questions must be answered to determine the sample size:

1. Standard deviation of the population: It is rare that a researcher knows the exact standard deviation of the population. Typically, the standard deviation of the population is estimated a) from the results of a previous survey, b) from a pilot study, c) from secondary data, or d) or the judgment of the researcher.
2. Maximum acceptable difference: This is the maximum amount of error that you are willing to accept. That is, it is the maximum difference that the sample mean can deviate from the true population mean before you call the difference significant.
3. Desired confidence level (%): The confidence level is your level of certainty that the sample mean does not differ from the true population mean by more than the maximum acceptable difference. Typically, social science research uses a 95% confidence level.

**Example**

A fast food company wants to determine the average number of times that fast food users visit fast food restaurants per week. They have decided that their estimate needs to be accurate within plus or minus one-tenth of a visit, and they want to be 95% sure that their estimate does differ from true number of visits by more than one-tenth of a visit. Previous research has shown that the standard deviation is .7 visits. What is the required sample size?

Population standard deviation: .7

Maximum acceptable difference: .1

Desired confidence interval (%): 95

---

Required sample size = 188