Statistics with the

TI-83 & TI-84 Calculator

Copyright © 2015 Martin Sauer

Table of Contents

Welcome!

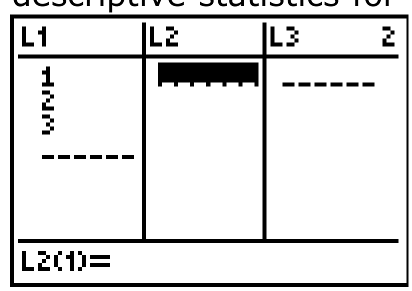
Welcome to College Statistics with the TI-83/84 Series Calculator. This guide is intended as a tutorial for statistics students who use a TI-83 or TI-84 calculator for their statistics class. The TI-83 works the same as the TI-84, keystroke for keystroke. The only major differences are that the TI-84 has more programming memory and operates a little faster than the TI-83.

The goal of this guide is to provide the reader with a quick and easy way to get up to speed on the most commonly used college statistics commands, in the shortest amount of time possible. For the more advanced features, advanced tutorials can be purchased on Amazon.com.

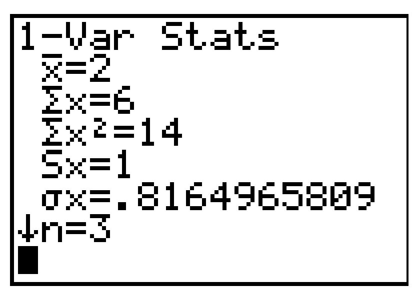
Descriptive Statistics

Measure Of Central Tendency refers to values that describe the central (middle) characteristics of data. The three main measures of central tendency are the mean, median, and mode. The ***Mean*** (*also called the **Average***) is the sum of data entries, divided by the number of data entries. The Median is the middle point of a data set, when the data is arranged in ascending order (from low to high). The ***Mode*** is the data entry or data entries which occur with the highest frequency (the most number of times).

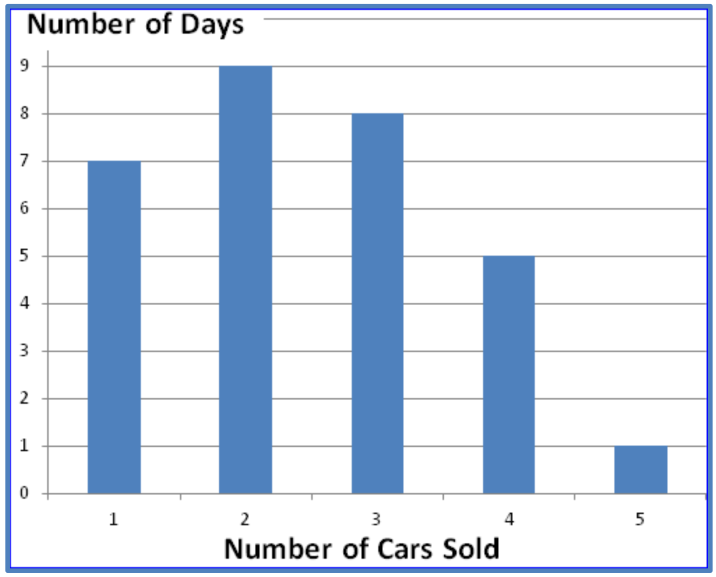
Often the first step to statistical analysis with the TI-83/84 is to put your data into a list. Let's start by first calculating simple descriptive statistics for the numbers {1, 2, 3}. We'll start with this simple example because we already know that the average will be 2 and the median will also be 2 just by looking at the dataset. We can also see that the sum of the values is 6. To put your data into a list press **STAT > ENTER (Edit)**, and then enter your data into list **L1** as shown on the right.



To find the descriptive statistics press **STAT > CALC > 1-Var Stats > ENTER > 2nd > L1 > ENTER**. The descriptive statistics are shown on the right. We see that the mean is 2, the sum of the values is 6, and by scrolling down using the down arrow, the median is also 2.

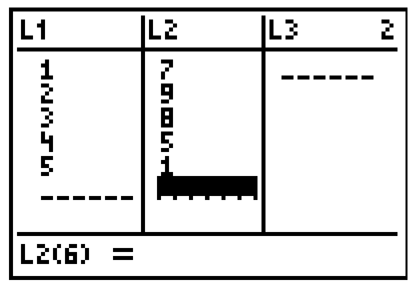


Our next example will be to generate descriptive statistics for the number of cars sold at an auto dealership during the past month. This example is more complicated than the last one because the data consists of x-axis and y-axis data. The variable (what we are actually measuring) is on the x-axis.

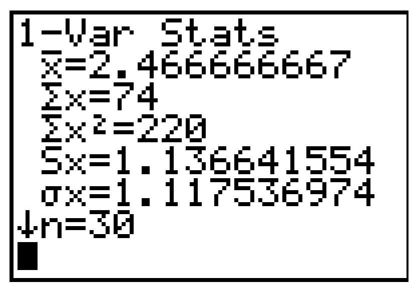


The data is presented in the histogram on the right and shows the variable **Number of Cars Sold**, and the Frequency of occurrence, **Number of Days**. For example, in the histogram we see that three cars were sold on eight days during the past month.

To calculate descriptive statistics for the cars sold, start by inputting the data into a list and then use the *Descriptive Statistics* function on your calculator. Enter the data as shown on the right by selecting **STAT > ENTER** (Edit) and then enter the values.



To find the mean, median, standard deviation, and other descriptive statistics press **STAT > CALC > 1-Var Stats > ENTER > 2nd > L1 > , > 2nd > L2 > ENTER**. The descriptive statistics are shown on the right. To see more stats press the down arrow on the right side of your TI-83/84. In this example the mean is 2.467 and the total number of cars sold was 74.



As you scroll down the screen notice that your TI-83/84 also gives the output for the first quartile (Q1), the second quartile (Med), and the third quartile (Q3), as well as the maximum value and other information about the data.

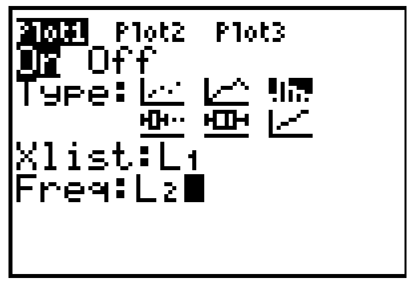
You will also notice that there are two values for the standard deviation, *Sx* and *σx*. When calculating the standard deviation it is important to distinguish the difference between a population and a sample. The standard deviation formula will be different depending on where the data comes from (a sample, or a population). If the data comes from a population we use the formula with the N in the denominator and if it comes from a sample, the formula with n-1 is used (below). A good way to remember this is that the sample formula has less precision. The n-1 in the denominator forces the sample standard deviation to always be a little higher. So when you see the two outputs the TI-83/84 gives for standard deviation (Sx and σx), the higher number will always indicate your TI-83/84 is treating this as a sample.

Population Standard Deviation $$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

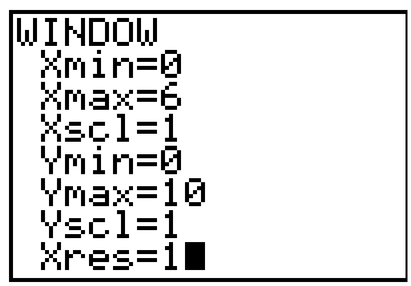
Sample Standard Deviation $$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}}$$

Graph the Data

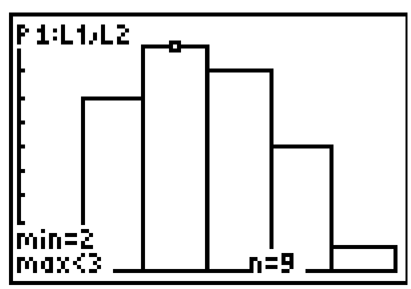
To now graph this data, press **2nd > STAT PLOT > ENTER > On > ENTER > HISTOGRAM** (shown only as a picture) **> Xlist: L1 > Freq: L2 > GRAPH**.



You must now set the appropriate graph parameters in order to view the whole picture. Press **WINDOW**. The WINDOW function allows you to set the parameters – or range – for your histogram graph. Notice that **Ymax = 10** is just above the highest value on the histogram above (which is 9). The same idea follows for **Xmax**.



You are now ready to plot your graph by pressing **GRAPH**. To view each individual data point press the **TRACE** button and then move the cursor to the left or right to view the data that corresponds to each individual bar. The cursor’s coordinates are listed at the bottom of the graph.



Principles of Counting

Before initiating many calculations, you first need to be on the Home screen. <u>To get out of most nested menus</u> – select **2ND > QUIT**.

Fundamental Counting Principle

The Multiplication Formula states that if there are a ways of doing one thing and b ways of doing another thing, then there are a x b ways of doing both. For example, If a menu consists of 4 choices of soup, 9 main dishes and 5 desserts, there are (4)(9)(5) = 180 different meals which can be selected for dinner. This formula is quite easy to calculate with the TI-83/84 : **4 > X > 9 > X > 5 > ENTER**. There are 180 different meals which can be selected for dinner.

Permutations

A permutation is an ordered arrangement of objects. The number of permutations for n objects is *n!*, where *n! = n (n - 1) (n -2)*…..3 * 2 * 1. The permutation formula is:

$$ {}_nP_r = \frac{n!}{(n-r)!} \qquad {}_{10}P_3 = \frac{10!}{(10-3)!} = 720 $$

For example, suppose 10 people apply for the positions of President, Vice President, and Commissioner. How many different arrangements of these three positions (President, Vice President, Commissioner) are possible from the pool of 10 candidates? Press **10 > MATH > PRB > 2 (nPr) > 3 > ENTER**. There are 720 different ways these three positions can be allocated.

Combinations

A Combination is the number of ways to choose ***r*** objects from a group of ***n*** objects without regard to order in the group ***r***. In other words, how many groups of size ***r*** are possible from an initial population of size ***n***? The formula is:

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} \qquad {}_{10}C_{3} = \frac{10!}{3!(10-3)!} = 120 $$

Using the same example as above, suppose 10 people apply for the positions of President, Vice President, and Commissioner. How many different groups of 3 people can be chosen, without regard to the position they are selected for? On the **Home** screen press **10 > MATH > PRB > 3 (nCr) > 3 > ENTER**. There are 120 different groups of three possible.

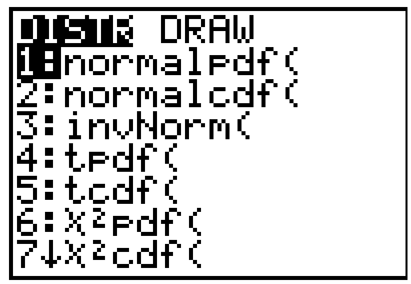
In the above example the order/position of the candidates doesn't matter. So Nguyen, Jones, Martinez for the positions of President, Vice President, and Commissioner, is the same as Martinez, Jones, and Nguyen for the same slots. All we care about is the three individuals elected when using a combination.

Probability Distributions

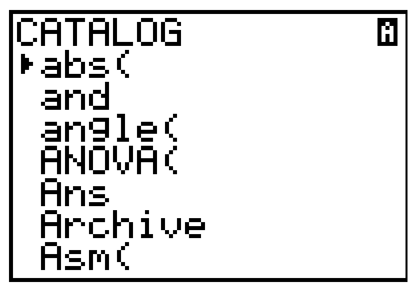
A Probability Distribution is the listing of all possible outcomes of an experiment and their corresponding probabilities. It gives the entire range of values that can occur, along with the probability of each outcome occurring. Distributions are either discrete or continuous.

A *Random Variable* is an experimental outcome that generates exactly one numerical value. For example, the number of couches sold during one day at a furniture dealership has the random variable X = Number of couches sold. X represents one particular outcome for an experiment (a day). Random variables can be either discrete, or continuous. A discrete random variable is a countable number, such as {0,1,2,3, …}. A continuous random variable has an uncountable number of possible outcomes. For example, there are an infinite number of different values between the numbers 3 and 4.

To find the various distributions that your calculator has to offer, select **2nd > DISTR**. Depending on how new your calculator is, you may have more menu options than the image shown on the right.



If you ever can't find a calculator function, they are all available in alphabetical order by selecting the **CATALOG** function, which can be found on the Zero (0) button by selecting **2ND > CATALOG**. Simply scroll down to find every function your calculator has to offer.



Discrete Distributions

Binomial Distribution

A binomial experiment satisfies the following conditions:

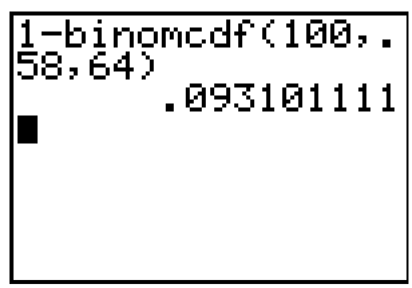
1. There are a fixed number of independent trials n.
2. There are only 2 possible outcomes for each trial, Success (S) or Failure (F).
3. The probabilities do not change for each trial.

If you didn't have a calculator to calculate probabilities automatically, you would need to use the binomial formula on the right to calculate each individual probability. From the formula you can see that this is a tedious task which isn't practical to calculate manually. Also, binomial tables are useless if the probabilities can't be looked up in a table (for example, 58% or 15.5%).

$$f(x) = \frac{n!}{x!(n-x)!} p^x (q)^{n-x}$$

In this example suppose that 58% of households in the US own a computer. If you randomly select 100 households, what is the probability that exactly 65 households will own a computer? You enter the appropriate parameters into your TI-83/84: **2nd > DISTR > binompdf(100, .58, 65) > ENTER** and get the answer 0.0299. So the probability that exactly 65 households out of 100 will own a computer is only about 3%.

Suppose you want to know the probability that <u>*at least*</u> 65 households will own a computer? This is a binomial distribution with n = 100, p = .58, and X ≥ 65. Use the **Binomial Cdf** function since there are many outcomes (65 + 66 + 67 + … + 100). To accomplish this task we will use a modified form of the compliment rule: P(X ≥ 65) = 1 – P(X ≤ 64). On your calculator select **1 – binomcdf(100, .58, 64).** The output shows that the probability that at least 65 will own a computer is approximately 9.3%.



The syntax for both the binompdf is **binompdf(n, p, x).**
The syntax for both the binomcdf is **binomcdf(n, p, x).**

Poisson Distribution

The *Poisson Distribution* is another discrete distribution. A Poisson distribution satisfies the following conditions:

- The experiment consists of counting the number of times, X, an event occurs in the interval. The interval can be **time, area, distance,** or **volume**.
- The probability of the event occurring is the same for each interval.
- The events are independent, regardless of what might occur in another interval.

Suppose the average number of airplane crashes annually throughout the world is 8. What is the probability that exactly 4 airplanes will crash this year? To answer this question enter the following formula into the TI-83/84: **2nd > DISTR > poissonpdf(8, 4) > ENTER** to get the answer 0.0573.

If you were interested in the probability of more than 4 airplanes crashing, you would use the cumulative case similar to what was done in the binomial distribution: Enter the formula **1 – poissoncdf(8,4).** The answer is 0.9004. There is a 90% likelihood that there will be more than 4 airplane crashes this year.

Continuous Distributions

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)/2\sigma^2}$$

The normal distribution is the most important continuous probability distribution in all of statistics. Many natural phenomena conform to a normal distribution. For example, the height of people and trees is normally distributed. The normal distribution is defined by two parameters, the standard deviation and the mean.

Sometimes you may not know the upper or lower bounds even though you know it is endless. In those cases where you have no idea what the upper and/or lower bounds should be, you can approximate ± ∞ (± infinity) by using ± E99. (Use the **EE** key to the left of the 4 key.) To enter a negative value into your calculator use the **(-)** button on the **ANS** key (under the **3** key).

Suppose the average lifespan of a certain Redwood Tree is 215 years with a standard deviation of 25 years. If you randomly select a tree just after it has died, what is the probability that the tree was less than 175 years old? As before, we enter the equation such that the following is on your screen: **normalcdf(-1E99,175,215,25)**, and get an answer of 0.0548. [On your calculator "-1E99" is entered as follows: **(-) > 1 > 2nd > EE > 99**.]

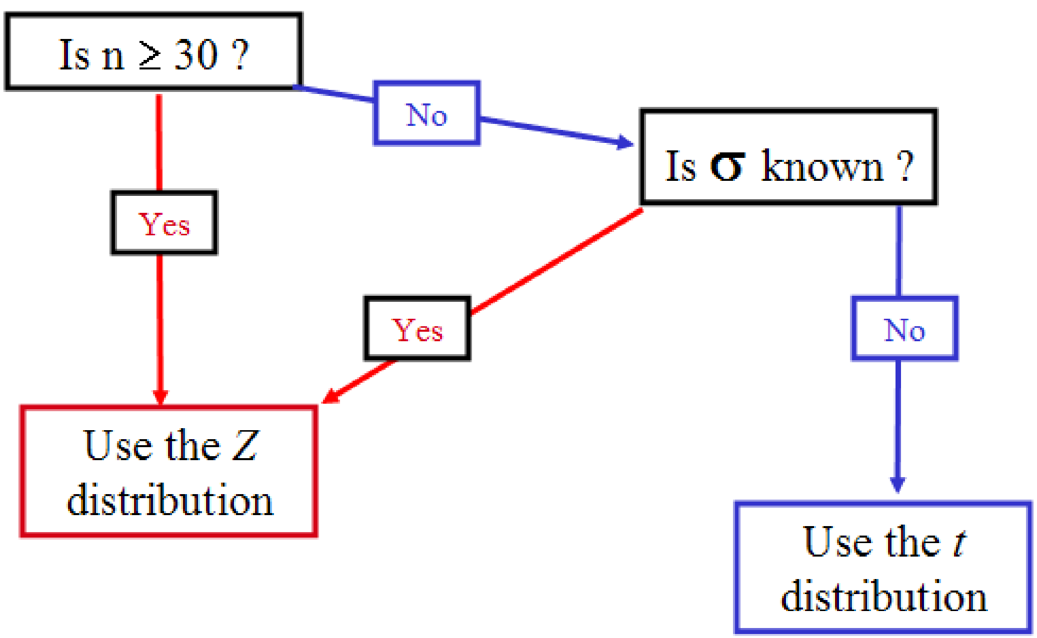
But suppose we desire to know the probability that the tree was between 190 and 240 years old? Using the cumulative functions again we enter: **normalcdf(190, 240, 215, 25)**, and get an answer of 0.6827.

The syntax is **normalcdf(L, U, μ, σ)**, where L is the lower bound and U is the upper bound.

Confidence Intervals

A confidence interval allows us to make meaningful estimates of true population means, standard deviations, and proportions. We specify a confidence interval of values on a number line, and then state the degree to which we are confident that the interval contains the actual population parameter. An interval estimate is an interval, or range of values, that estimates the population parameter. The wider the interval, the more confident we are that the true parameter lies within the interval. The level of confidence is the probability that the interval contains the population parameter.

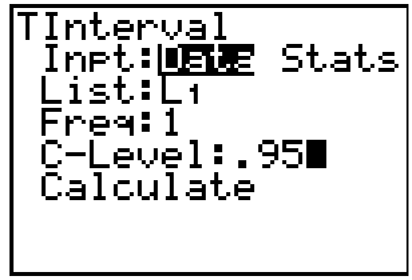
Generating confidence intervals is very simple. For normally distributed data where σ is known, use the **ZInterval.** For cases where σ is unknown, use the **TInterval**. There is a little bit more to it. Some textbooks use Z automatically if the sample size n ≥ 30. The following graphic shows this more clearly.



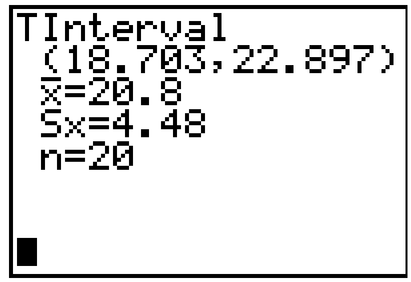
We must still assume that the population is approximately normally distributed.

Suppose you want a 95% confidence interval for the average fuel mileage of 20 foreign made sports cars. Since n < 30, calculate a t confidence interval

(**TInterval**). Enter the following data into list **L1: 26, 22, 23, 12, 19, 25, 23, 21, 25, 10, 17, 26, 23, 24, 20, 14, 21, 23, 20, 22**. Since the data is in a list, we choose **STAT > TESTS > TInterval > Data** and enter the rest of the information as show in the screen capture. Select **Calculate**. The confidence interval is centered around a mean of 20.8, and the standard deviation of the data is 4.48. It may be written as: (18.7 < μ < 22.9).



If there is no data but you know that the sample average is 20.8 and the sample standard deviation is 4.48, simply enter that data directly into the t-test function as such: **STAT > TESTS > TInterval > Stats** and enter the mean and standard deviation data. The results are the same. There is a 95% probability that the confidence interval we just created contains the true population parameter μ.



Hypothesis Testing

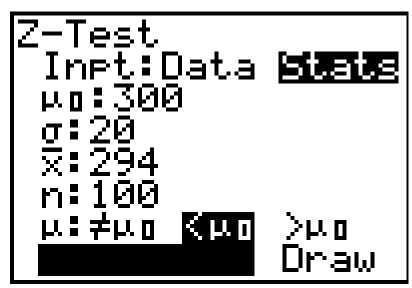
Hypothesis testing is a process that uses sample statistics to test a statement about the value of a population parameter. The statistic that is derived from the sample is compared to the parameter in the null hypothesis, and is called the *Test Statistic*. If the test statistic is very far away from the parameter in the null hypothesis, we then reject the null hypothesis and accept the alternative hypothesis.

We decide if the test statistic is too far away from the parameter in the null hypothesis by calculating the ***P-Value***. The p-value is the probability of obtaining a sample statistic with a value as extreme – or even more extreme – than the sample data if the value in the null hypothesis were true.

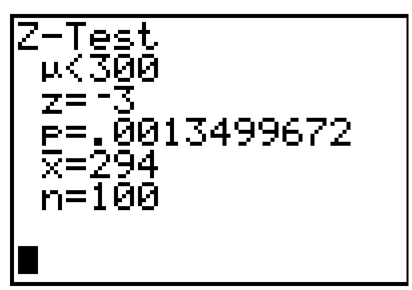
Like confidence intervals, hypothesis testing is also a snap. Suppose a corporation comes out with a claim that the mean salary for all of its senior consultants is 300 dollars per hour with a standard deviation of 20 dollars per hour. You are interested in working for the company but believe that salaries are lower. So you decide to test the validity of this statement by testing the salary claim at the 1% level of significance. 100 employees are randomly sampled. The mean salary is found to be only 294 dollars per hour. The null and alternative hypothesis are as follows:

$$H_0: \mu = 300$$
$$H_a: \mu < 300$$

This is a left tailed test since the alternative hypothesis is $\mu < 300$. To solve this problem easily, select **STAT > TESTS > Z-Test** and enter the data as shown on the screen. Select **Calculate** at the bottom and **ENTER**.



The output shows the p-value is 0.0013. There is significant evidence to reject the null hypothesis and conclude that the corporations claim is not accurate. You conclude that the average salary is less than $300 per hour.



The F Test

The F distribution is used to test whether two samples are from populations having similar variances/standard deviations. The null and alternative hypothesis are as follows:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

There is a "family" of *F* Distributions. Each member of the family is determined by two parameters: the numerator degrees of freedom and the denominator degrees of freedom.

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$

As an example, suppose a manager of a taxi service wants to examine the time it takes to drive to SeaTac airport from South-Western Tacoma using two different routes, one using Highway 99 and the other using I-5. Since tight scheduling is a necessary part of the business, it's important that the variation in the time of each route is kept to a minimum. From past experience it seems that the Hwy 99 route has much more variation and the manager wants to prove this to his boss. This test is conducted at the 5% level of significance.

The following sample data of minutes per trip on each route was collected and entered into **L4** and **L5**:

Hwy 99	I-5
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

$\bar{x}_{99} = 58.3$ $\quad \bar{x}_{I-5} = 59.0$

$s_{99} = 9.0$ $\quad s_{I-5} = 4.38$

$n_{99} = 7$ $\quad n_{I-5} = 8$

It appears that the taxi will be late much less frequently using this route.

$H_0: \sigma_{99}^2 \le \sigma_{I-5}^2$

$H_1: \sigma_{99}^2 > \sigma_{I-5}^2$

$\alpha = .05$

To test the hypothesis on your TI-83/84 press **STAT > TESTS > 2-SampFTest** and enter the Stats data above. Remember that this is a right tailed test. The output shows a p-value of 0.0405, so the null hypothesis is rejected and you conclude that the Hwy 99 route does indeed have more variation. The p-value is less than the 5% level of significance, so the conclusion is that there is more variation in the HWY 99 route. In addition, note that the average time for both routes is nearly identical, it's just that the HWY 99 route is more unpredictable. The taxi service will be late more often and may miss their pickups. They will also be early more often and be waiting around with nothing to do.

ANOVA (Analysis of Variance) – 1-Factor

Another use of the F distribution is a technique where we compare 3 or more population means to determine whether they are not all the same (at least one mean is different from another). Sample means are compared through their variances.

H_0: $\mu_A = \mu_B = \mu_C$
H_1: The means are not all the same.

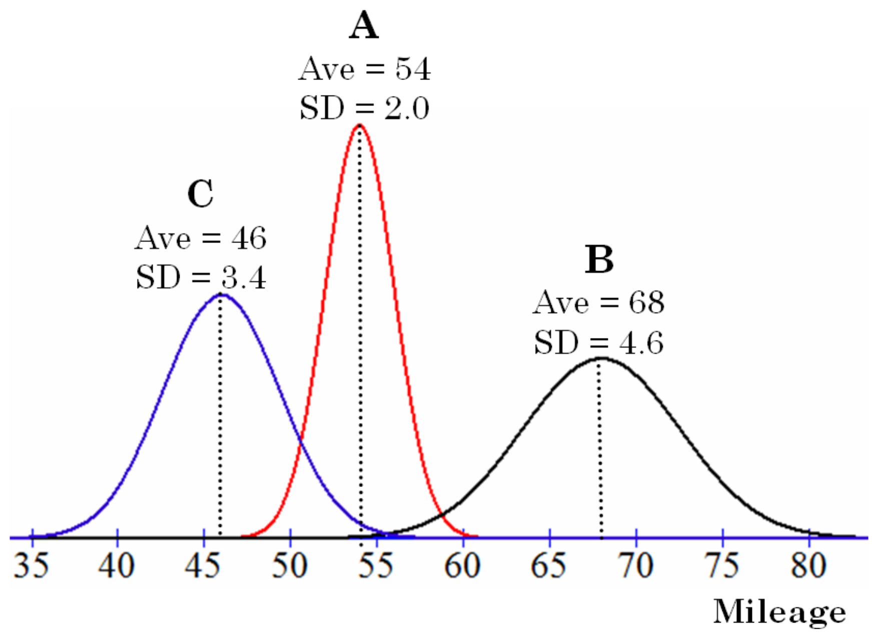
ANOVA requires the following conditions be met:

- The samples are randomly selected and are independent.
- The populations have equal standard deviations.
- The sampled populations follow the normal distribution.

The data below shows the number of miles driven (in thousands) of three different brands of tires. The test was stopped when each tire was too worn to meet highway safety standards, after which the tire was recycled. Is there a difference in the mileage among the three brands? This is referred to as one-factor ANOVA because we are investigating just one factor, or effect. The variation between groups represents the variation due to the effect, which in this case is the effect that each brand has on mileage, where Mileage is the dependent variable.

	Tire Style A	**Tire Style B**	**Tire Style C**
	55	66	47
	52	74	51
	53	62	44
	57	67	46
	53	71	42
Ave	**54**	**68**	**46**
SD	**2.0**	**4.6**	**3.4**

In the graphs below it appears that the means are not all the same – particularly when comparing the distribution of brand C with brand B. So we might expect our ANOVA analysis to show that the means of the three data sets are not all the same.



The next step is to enter the data for each into three separate lists, L1, L2, and L3. Next, press **STAT > TESTS > ANOVA > ANOVA(L1, L2, L3) > ENTER**. The results show a p-value of approximately 0.0000, and we conclude that the means are not all the same. Reject the null hypothesis. There is a difference in tire life among the three brands.

Chi-Square Test with Contingency Tables

The purpose of a Chi-Square Test is to compare an observed distribution to an expected distribution. Chi-Square techniques are appropriate for use with data in the form of frequencies (counts). The Chi-Square test should only be used if all of the observations are independent of each other. As a result, Chi-Square should not be used in repeated measures designs or for pairs of scores (paired; dependent).

A contingency table is used to investigate whether two traits or characteristics are related. Each observation is classified according to two criteria:

The *degrees of freedom* is equal to:
df = (# rows - 1)(# columns - 1)

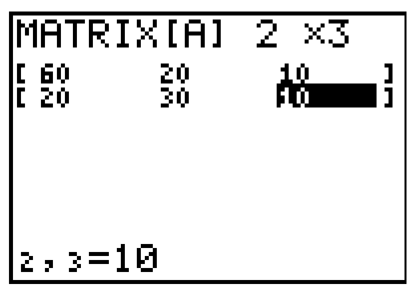
The *expected frequency* is computed as:
Expected Frequency = (row total)(column total)/Grand Total

Let's look at an example:

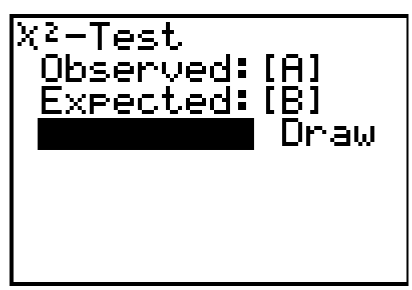
Is there a relationship between the type of HMO health plan a consumer selects and the types of complaints received by the HMO? A sample of 150 calls received at a call center were classified by Health Plan and type of Complaint. At the .05 level of significance, can we conclude that Health Plan and Complaint are related?

	Complaint			
Health Plan	Wait Time	Price	Other	**Totals**
Standard	60	20	10	**90**
Premium	20	30	10	**60**
Totals	**80**	**50**	**20**	**150**

To solve this problem we need to now put our data not in a list, but in a matrix. Press **2nd > MATRIX > EDIT > ENTER** to edit MATRIX [A]. Create a 2 X 3 matrix and enter the data as shown.



Now Press **STAT > Tests > X^2 Test > Calculate** as shown.



The results show a p-value of approximately 0.0002, and Ho is rejected. We conclude that health plan and type of complaint are indeed related.