Basic Statistics

with the

TI-89 Calculator

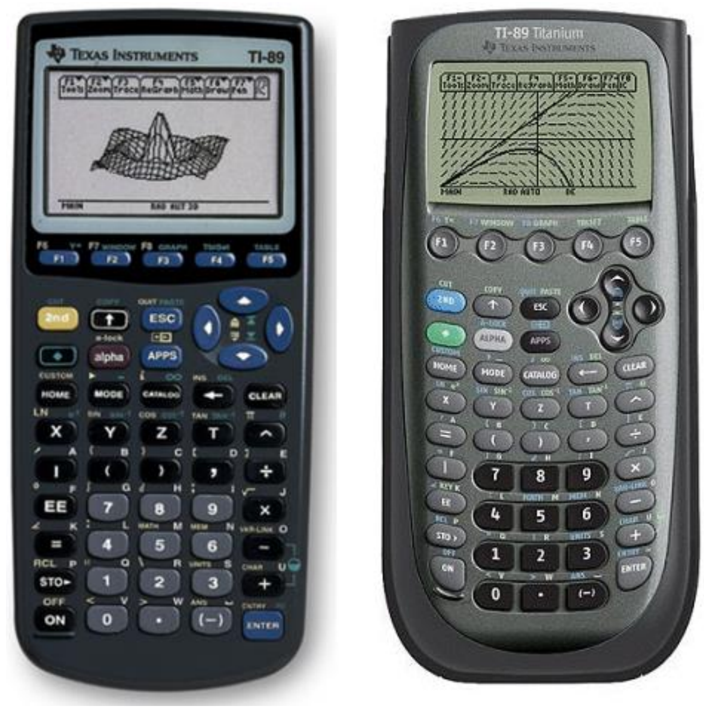
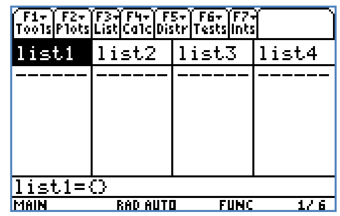


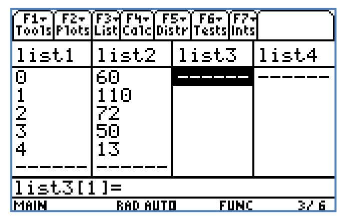
Table of Contents

Welcome!

Welcome to Statistics with the TI-89 calculator. This guide is intended as a tutorial for students of statistics who use a Texas Instruments TI-89 calculator for their statistics class. 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 statistics commands, in the shortest possible amount of time possible. This tutorial is designed to be used for all TI-89 models. For the more advanced features available on the TI-89 Titanium, advanced tutorials can be purchased on Amazon.com.



Often the first step to statistical analysis with the TI-89 is to put your data into a list. To find the main default lists select **Stats/List Editor** from the home screen until your calculator looks like the screen capture on the right.

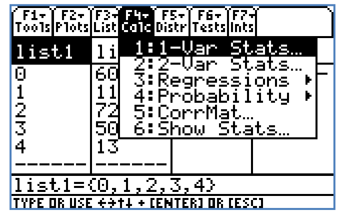


Now scroll down and put your data into **list1**. In this example we will analyze last year's sales of couches by a large furniture store. The business was open a total of 305 days over the past year. Enter the data into each list as shown. We observe that no sales took place on 60 days and 4 couches were sold 13 times (days) over the past year. We are now ready to calculate basic descriptive statistics.

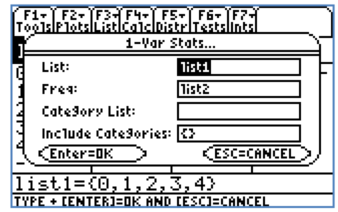
Descriptive Statistics

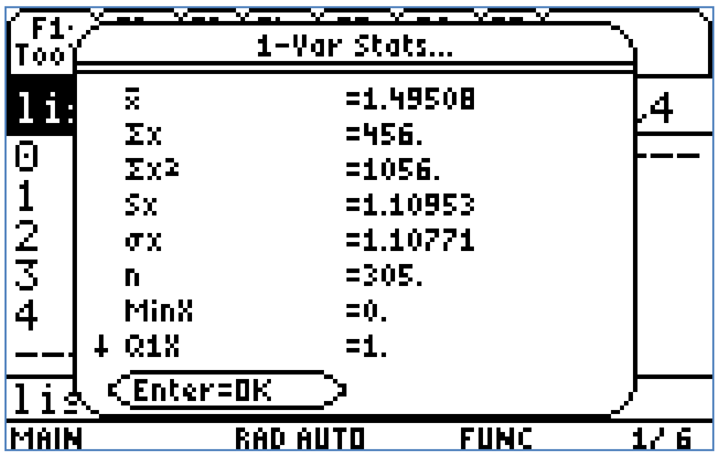
Measures Of Central Tendency refers to values that describe the central (middle) characteristics of data. The three main measures of central tendency are the mean, median, and mode. The ***Mean (****also called the* ***Average)*** is the sum of data entries, divided by the total number of data entries.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

From the dataset created above we can easily see that the mode is 1 couch. But what about the other descriptive statistics such as mean, median, standard deviation, quartiles, etc? To find these measures we will calculate descriptive statistics using **list1** as the variable (what we are measuring, which refers to numbers of couches), and list2, which is the frequency of occurrence for each observation. So while the lists are still showing on your screen, select **F4-Calc > 1:1-Var Stats** as shown on the right.



Now select **list1** for the variable and **list2** for the frequency as shown in the next screen capture on the right. To get these lists into your menu select **2ND > VAR-LINK** and choose the appropriate lists. Now just press ENTER=OK and the descriptive statistics appear. If you scroll down more descriptive information is given.





The screen capture above shows that the average of this dataset is 1.49508 couches sold per given day. There were a total of 456 couches sold. The standard deviations are also shown as S_x (if the data came from a sample), or σ_x (if the data consists of the entire population). The total number of observations is revealed by n = 305 days of economic activity, and the median is 1 sale. Notice also that your TI-89 also gives the output for the first quartile (Q1), the second quartile (Med), and the third quartile (Q3).

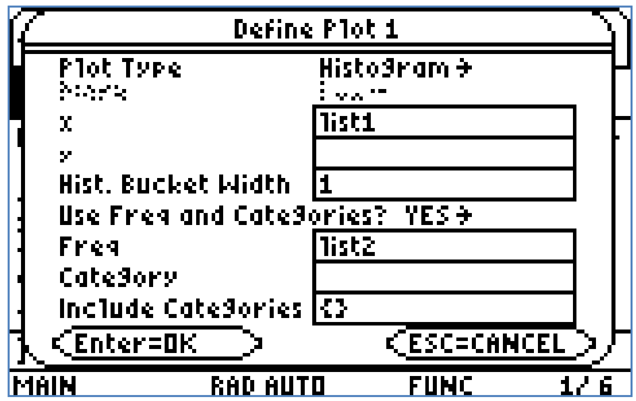
When calculating standard deviations it is important to distinguish the difference between a population and a sample. For example, if we wish to know the standard deviation, the standard deviation formula will be different depending on where the data comes from (a sample, or a population). If the data comes from a population we use the formula with the N in the

Population Standard Deviation $\sigma = \sqrt{\dfrac{\Sigma(x-\mu)^2}{N}}$ *Sample Standard Deviation* $s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}}$

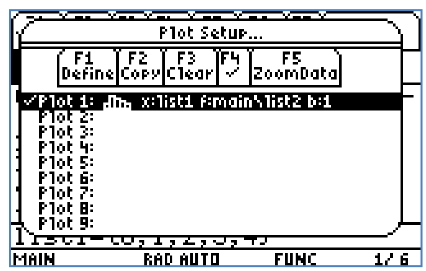
denominator and if it comes from a sample, the formula with n-1 is used. A good way to remember this is that the sample formula has less precision. Then n-1 in the denominator forces the sample standard deviation to always be higher. So when you see the two outputs the TI-89 gives for standard deviation (S_x and σ_x), the higher number will always indicate your TI-89 is treating this as a sample.

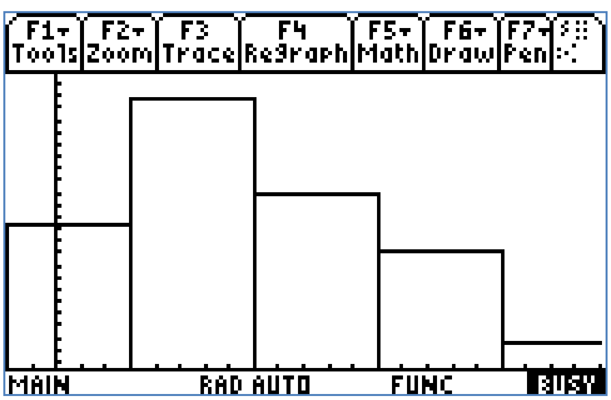
Graph the Data

To graph this data select **F2-Plots > 1:Plot Setup > F1 Define > Histogram** and select the appropriate lists and **Use Freq Categories** as shown. Now select **Enter=OK.**

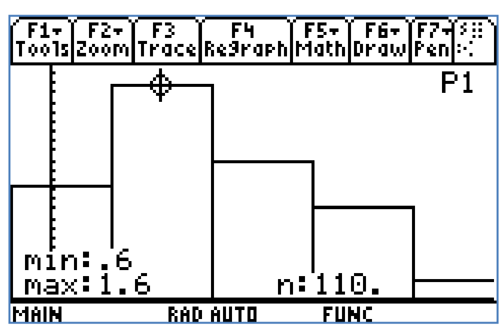


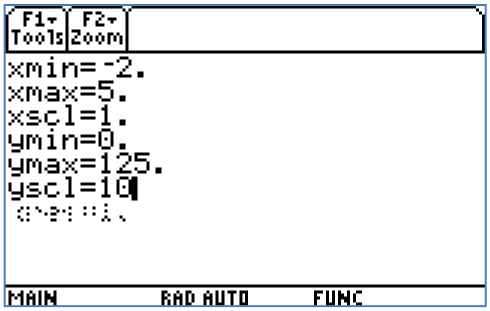
Make sure the **Plot 1** indicator is checked. If it is not, then select **F4** to check it. Now select **Enter=OK** and then **F5 ZoomData**. Your histogram is shown below.



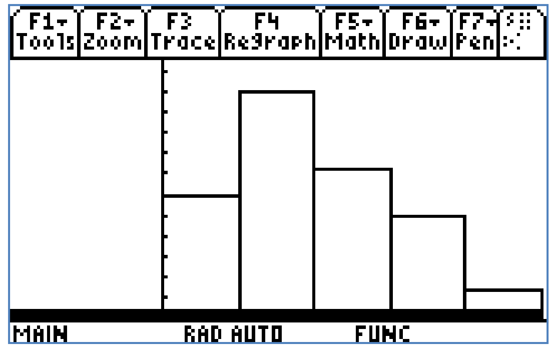


To confirm the correct histogram frequencies select **F3 Trace**.





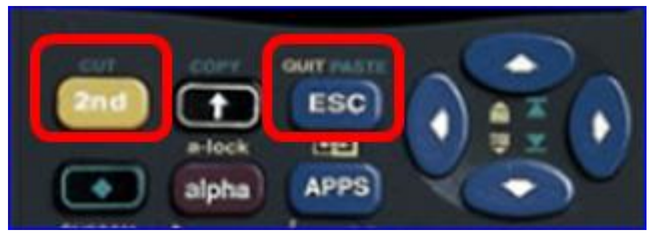
To change the window setting in order to see more of the graph, select the **◇ > WINDOW (F2)** key and change the settings as desired. Your new graph makes viewing easier.



Principles of Counting

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 dinner menu consists of 5 appetizers, 8 main dishes, and 6 desserts, then there are (5)(8)(6) = 240 different meals which can be selected for dinner. This formula is quite easy to calculate with the TI-89 by just multiplying the numbers, but you need to go to the **Home** screen first. To do so - and to get out of most nested menus - select **2ND > ESC**.



Permutations

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

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 a pool of 10 candidates? Assume any candidate can take on any position.

On the **Home** screen press **2ND > MATH > Probability > nPr > ENTER > (10,3) > ENTER.** There are 720 different ways these positions can be allocated. (NOTE: the comma key is above the 9 button). The formula is:

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

Combinations

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

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

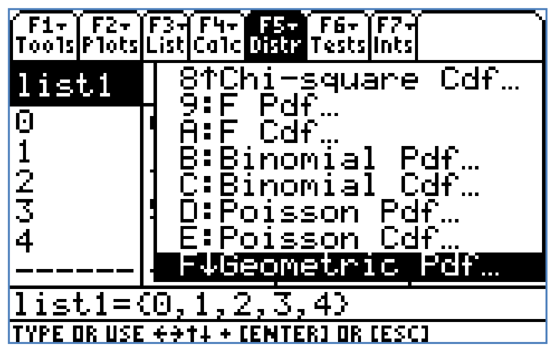
Using the same example as above, suppose 10 people apply for the positions of President, Vice President, and Commissioner. How many different groups of 3 people can be chosen, without regard to the position they are selected for? On the **Home** screen press **2ND > MATH > Probability > nCr > ENTER > (10,3) > ENTER**. There are 120 different groups of three possible. (The order/position of the candidates doesn't matter.)

Probability Distributions

A Probability Distribution is the listing of all possible outcomes of an experiment and the corresponding probability. 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 (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 the TI-89 has to offer select **Stats/List Editor** from the home screen and then **F5 Distr**.



Discrete Distributions

Binomial Distribution

A binomial experiment is an experiment that satisfies the following conditions:

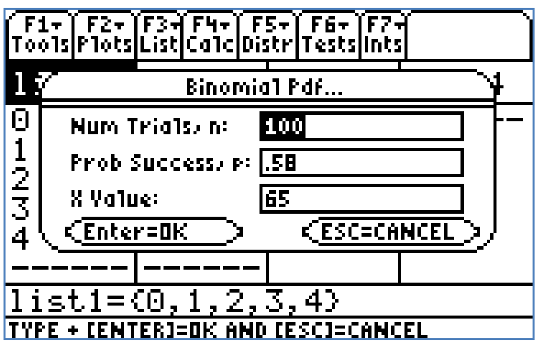
1. There are a fixed number of independent trials n.
2. There are only 2 possible outcomes for each trial, Success (S) or Failure (F).
3. The probabilities do not change for each trial.

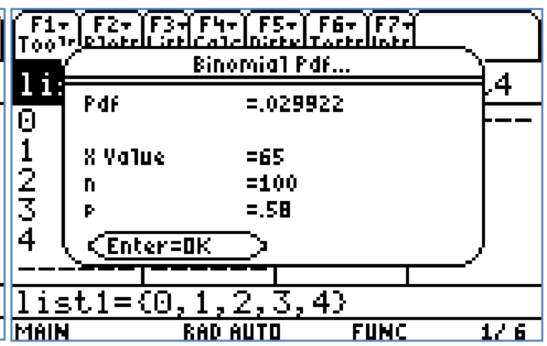
If we did not have a calculator to calculate probabilities for us automatically, we would need to use the binomial formula to calculate each individual probability. From the binomial formula shown on the right, you can see that this is a tedious task that isn't practical to calculate manually. Also,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

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

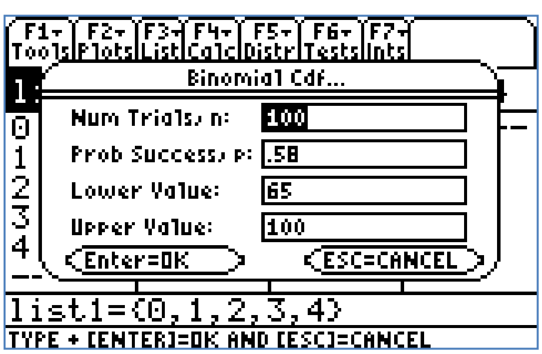
As an example, suppose that 58% of all households in the US own a computer. If you randomly select 100 households, what is the probability that exactly 65 households will own a computer? This is a binomial distribution with n = 100, p = .58, and X = 65.

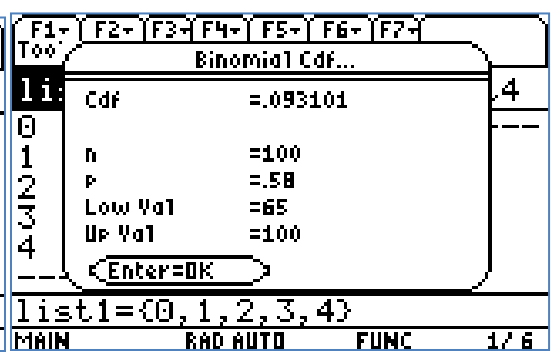
To find the various distributions that the TI-89 has to offer select **Stats/List Editor** from the home screen and then **F5 Distr**. Now scroll down to **Binomial Pdf** and enter the data as shown below. The answer is 0.0299. So the probability that exactly 65 households out of 100 will own a computer is only about 3%.





Suppose you want to know the probability that _at least_ 65 will own a computer? This is a binomial distribution with n = 100, p = .58, and X ≥ 65. We use the **Binomial Cdf** function since we are accumulating the outcomes (65 + 66 + 67 + … + 100).



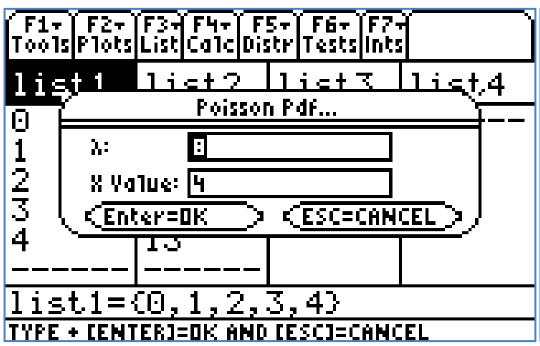


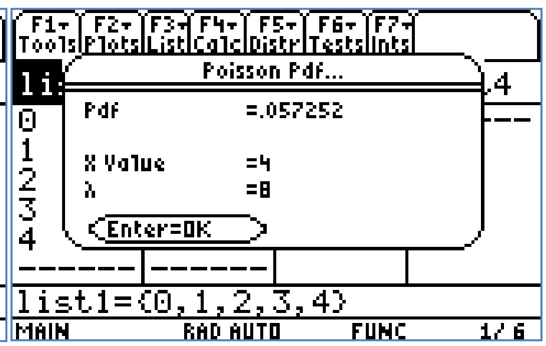
Poisson Distribution

The Poisson Distribution is another discrete distribution. A Poisson Distribution is a discrete probability distribution of a random variable that satisfies the following conditions:

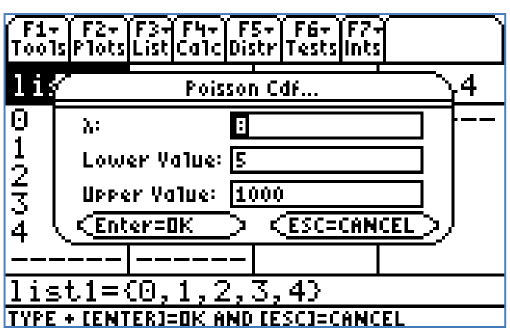
- The experiment consists of counting the number of times, X, an event occurs in the interval. The interval can be **time, area, distance,** or **volume**.
- The probability of the event occurring is the same for each interval.
- The events are independent, regardless of what might occur in another interval.

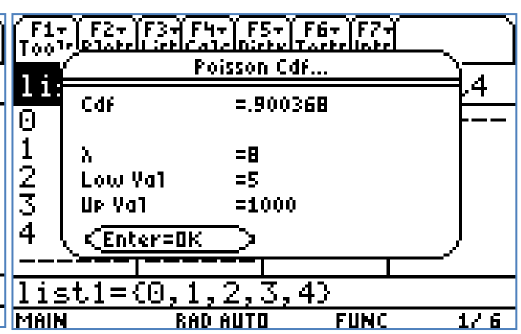
Suppose the mean number of annual major airplane crashes throughout the world is 8. What is the probability that exactly 4 airplanes will crash this year? From the **Stats/List Editor** select **F5 Distr > Poisson Pdf.** The average is **λ (lambda) = 8** and the number of occurrences **X = 4**. The probability of this occurring is 0.0573, or about 5.73%.





If we were interested in the probability of more than 4 airliners crashing, then you will use the cumulative case similar to what was done in the binomial distribution: Enter **F5 Distr > Poisson Cdf** as shown. But since the upper value has no limit, we will use 1000, which is reasonable since the average is 8. The answer is 0.9004.





Continuous Distributions

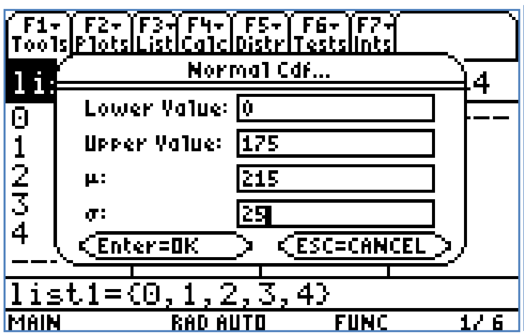
Normal Distribution

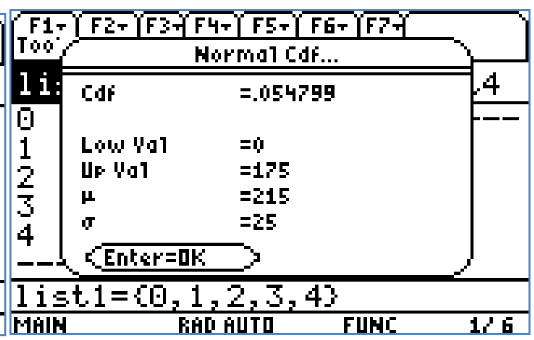
The normal distribution is the most important continuous probability distribution in all of statistics. Many natural phenomena conform to a normal distribution. For example, the height of people and trees is normally distributed. The normal distribution is defined by two parameters, the standard deviation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

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

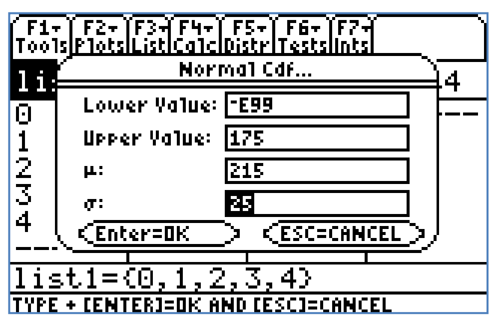
Suppose the average lifespan of a certain Redwood Tree is 215 years with a standard deviation of 25 years. If you randomly select a tree just after it has died, what is the probability that the tree is was less than 175 years old?

Enter **F5 Distr > Normal Cdf** as shown. The lower value is zero years. The probability of this occurring is about 5.5%.





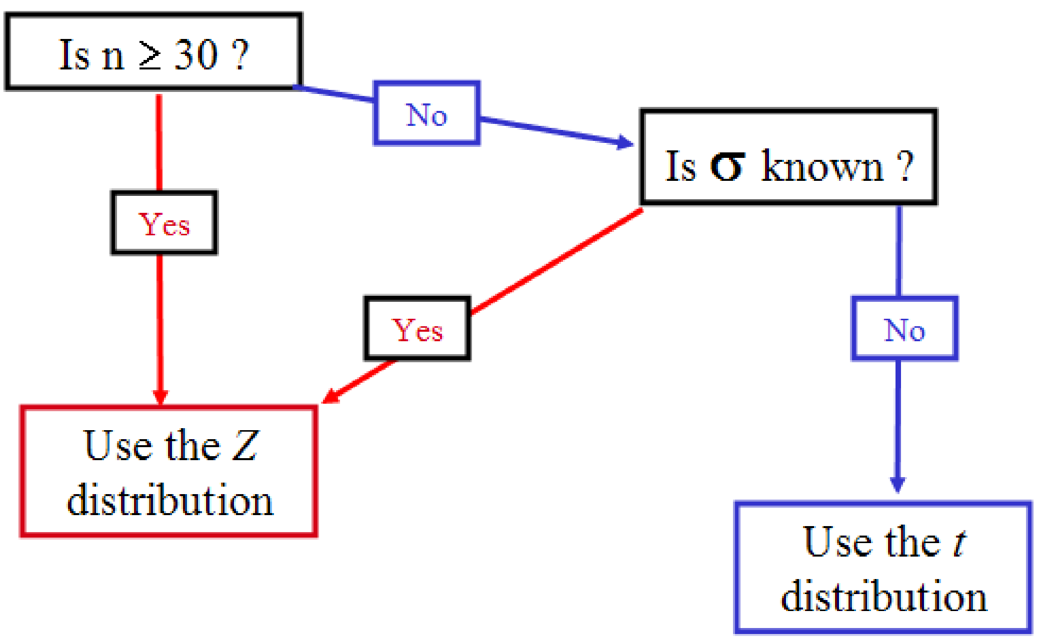
NOTE: Sometimes you may not know the upper or lower bounds even though you know it is endless. In those cases where you have no idea what the upper and/or lower bounds should be, you can approximate ± ∞ by using ± E99. (Use the **EE** key to the left of the 4 key.) To enter a negative value into your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



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 used to estimate 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 estimate contains the population parameter.

Generating confidence intervals with the TI-89 is very simple. For normally distributed data where we know σ, we use the **ZInterval**, and for cases where we don't know σ we use the **TInterval**. There is a little bit more to it. Some textbooks use Z automatically if the sample size $n \geq 30$. The following graphic that shows this more clearly.



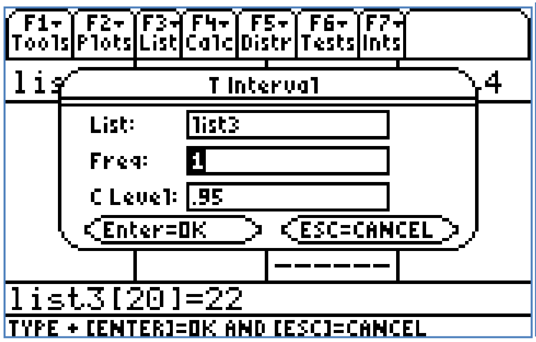
We must still assume that the population is approximately normally distributed.

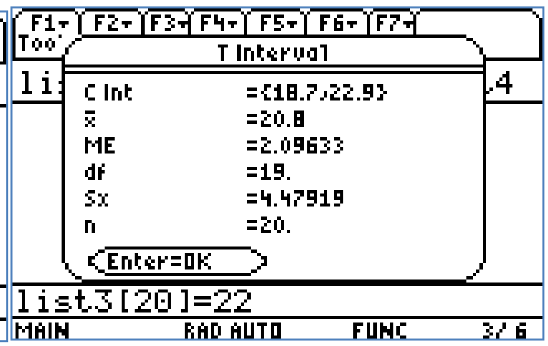
T Interval Example

Suppose we are want a 95% confidence interval for the mean fuel mileage of 20 foreign made sports cars. Since n < 30 and we don't know σ, we will calculate a T confidence interval (TInterval). We will also assume the fuel mileage for all foreign made sports cars is normally distributed.

Enter the following data into **list3: 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 **2ND > F7 [F2] > TInterval > Data** and then **list3** (using **2ND > VAR-LINK**). The **Freq:** is **1** since every item in list3 occurs just one time.

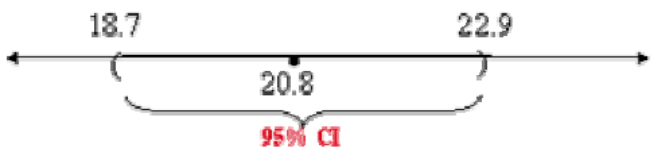




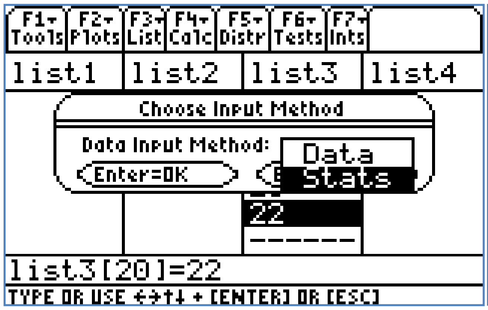
The confidence interval is centered around a mean of 20.8, and the standard deviation of the data is 4.48. It may be written as such:

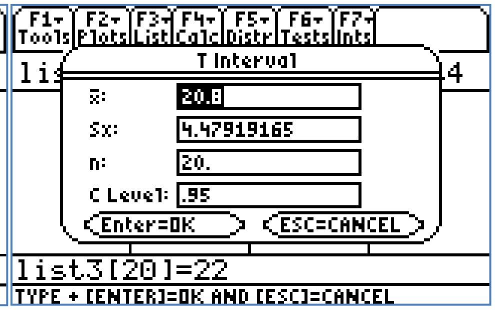
$$(18.7 < \mu < 22.9)$$

or



If we do not have any data but know that the sample average is 20.8 and the sample standard deviation is 4.48, we can simple enter that data directly into the TInterval function using **Stats**. 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.

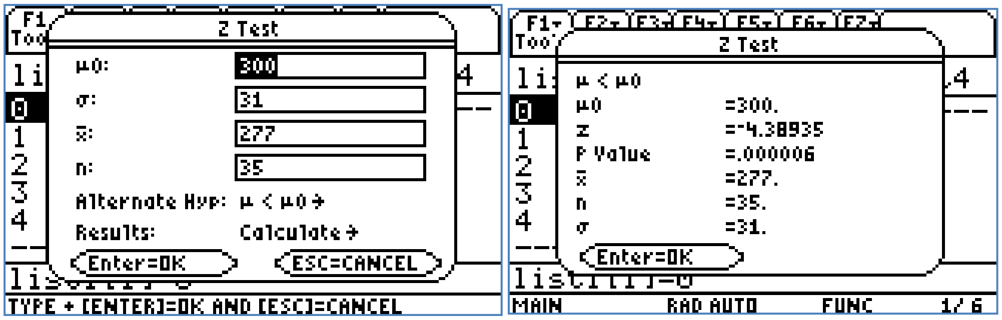
We decide if the test statistic is too far away from the parameter in the null hypothesis by calculating the ***P-Value***. The p-value is the probability of obtaining a sample statistic with a value as extreme - or even more extreme - than the sample data if the value in the null hypothesis were true.

Like confidence intervals, hypothesis testing is also a snap with the TI-89. 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 31 dollars per hour.

You are interested in working for this company and decide to test the validity of this statement and decide to test this claim at the 1% level of significance. So you randomly sample 35 employees and find that the mean salary is only 277 dollars per hour. The null and alternative hypothesis are as follows:

$$H_0: \mu = 300$$
$$H_1: \mu < 300$$

This is a left tailed test since the alternative hypothesis is $\mu < 300$. To solve this problem easily with the TI-89 select **Stats/List Editor** from the home screen and **2ND > F6 Tests > Z-Test > Stats**. You select **Stats** because you will be entering the statistic, parameters, and sample size directly from information already known, rather than having the TI-89 calculate them from a list like we did with the confidence interval example.



The output shows the p-value is 0.000006, which is far less than our 1% level of significance. Therefore, we 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.

In other words, if it's really true that the average salary is $300 per hour, the probability we would get a sample mean of $277 or less is only .000006. So only 6 in 1 million times will we get a sample this extreme or more extreme if the mean actually was $300. Not likely at all!

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.

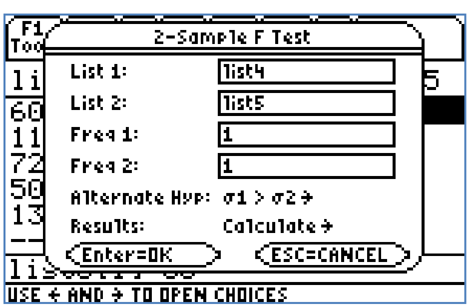
$$\boxed{F = \frac{s_1^2}{s_2^2}}$$

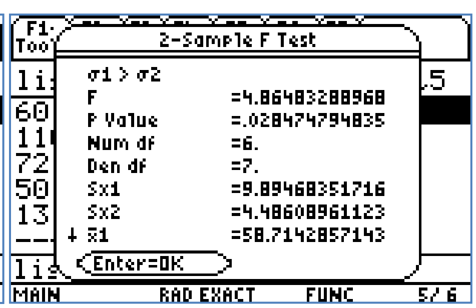
For example, suppose a manager of a taxi service wants to examine the time it takes to drive to SeaTac airport from South-Western Tacoma using two different routes, one using highway 99 and the other using I-5. Since tight scheduling is a necessary part of the business, it's important that the variation 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. You decide to conduct this test at the 5% level of significance.

The following sample data of minutes per trip on each route was collected and entered into **list4** and **list5**:

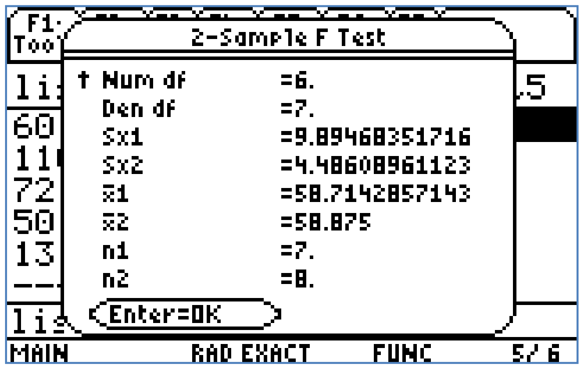
HWY 99	I-5
list4	**list5**
51	60
44	60
71	59
68	52
56	63
55	55
66	66
	56

Select **Stats/List Editor** from the home screen and **2ND > F6 Tests > 2-SampFTest > Data**. You select **Data** because you will have the TI-89 calculate the information from a list like we did before with the confidence interval example.





The p-value is less than the 5% level of significance, So we conclude that there is definitely more variation in the HWY 99 route. In addition, you will 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 could be equal. We actually compare sample means through their variances.

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

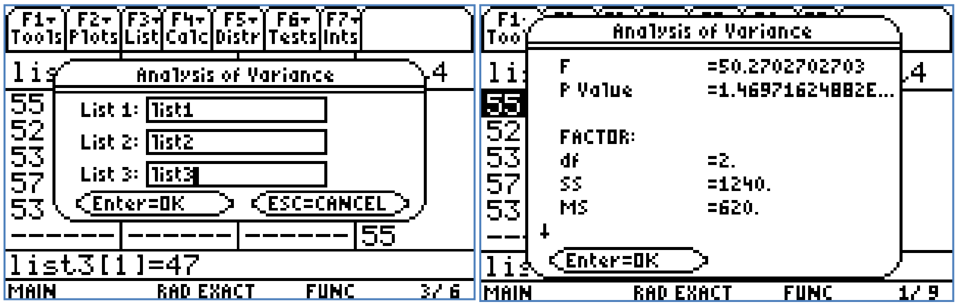
<u>ANOVA requires the following conditions be met:</u>

- 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 on 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**

Put the data into **list1**, **list2**, and **list3**. Select **2ND > F6 Tests > ANOVA > Data**. The output for ANOVA is a weak point on the TI-89, depending on how you have it setup using MODE. In this example the p-value floats out of sight. It should read 1.4697162E-6, which is about .0000015. This is very significant and we conclude that the means are not all the same. The graphics also don't give you the means of all three lists, which is important since that's what ANOVA does - it test differences in means through their variances.



Since the TI-89 output can be cumbersome and squished together, I have added a graphic. By visually inspecting the graph below it appears that the means are not all the same – particularly when comparing the distribution of brand C with brand B.