

The Statistics You Need to Know

When given a set of data, you can be asked to calculate and/or interpret several statistics:

- Mean, Mode, Median
- Range
- Box-and-Whisker Plot, Stem-and-Leaf Plot
- Standard Deviation (σ), Variance (σ^2)
- Mean Absolute Deviation (MAD)
- Z-Scores

You should begin by entering the data into a list on the calculator.

To clear a list:

STAT

EDIT

highlight the list to be cleared

CLEAR

ENTER or DOWN ARROW

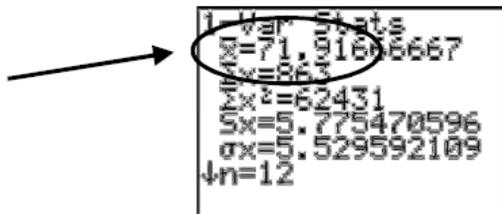
Example: A class of students opened bags of M&Ms and counted how many candies each bag had. Here are their totals:

{67, 73, 81, 62, 79, 72, 68, 80, 71, 73, 67, 70}

The **Mean** of a set of data is the average.

- ✓ In notation, it is represented by the Greek letter mu (μ).
- ✓ On the calculator, it is represented by x-bar (\bar{x}).

Solution:



STAT

1: EDIT

Enter the date

2nd QUIT

STAT → CALC

1: 1-Var Stats...

2nd L₁

Enter

The class had an average (mean) of 71.92 M&Ms per bag.

The **Mode** of a set of data is the value that occurs most often.

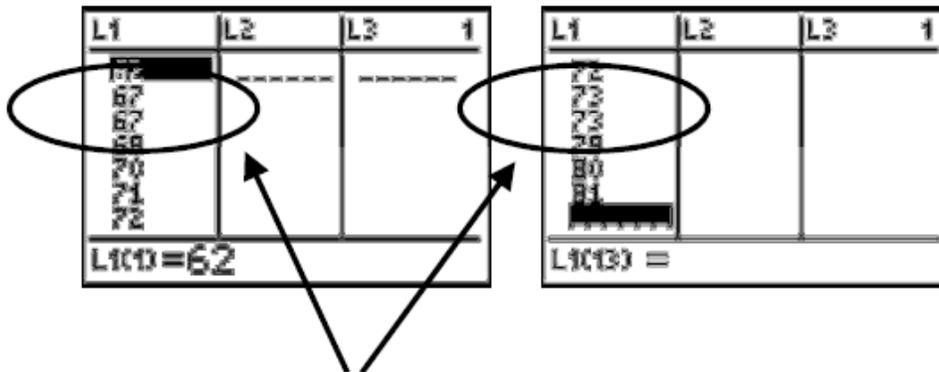
- ✓ The calculator will not directly display this for you, but you can arrange the list in order (lowest to highest) and then scroll through to see what value(s) are repeated.



STAT

2: SortA(
2nd L₁
Enter
(The screen will display done.)

- ✓ Now go back in and look at L1 – it's been arranged in order for you! Scroll through to see what the mode is.



STAT

1: EDIT
Look at Data
2nd QUIT

Solution: The class has two modes – 67 and 73.

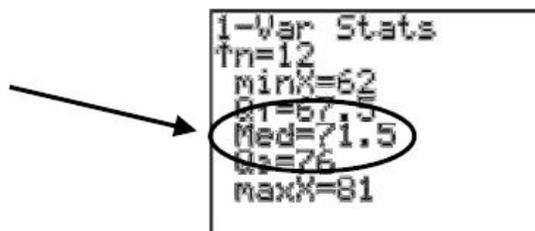
Note:

- A set of data may have one mode, more than one mode, or no mode at all (if none of the values are repeated).

The **Median** of a set of data is the value that is located in the middle of the list once it has been arranged from least to greatest.

- You may have the calculator arrange the list of data for you (like you just did in finding the mode) and choose the one in the middle, or...
- You may return to 1-Var Stats and scroll down to see the Med value.

Solution: The class has a median of 71.5 M&Ms.



STAT → CALC

1: 1-Var Stats...
2nd L₁
Enter

Notes:

Be sure to arrange the values first!

- If there are an even number of data values (as there are in this example), then find the mean of the "two" middle values.
- If there are an odd number of data values, there will be exactly one value perfectly in the middle.

The **Range** of a set of data is the difference between the largest and smallest values.

- There is no commonly-accepted notation for the range.
- The calculator will not compute the range directly; you will need to subtract the values yourself.

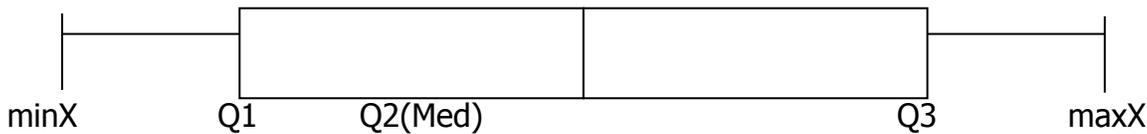
Solution: The class had a range of 19 M&Ms.

The **Box-and-Whisker Plot** is one common way to display the data.

- ✓ This display evenly breaks the list of data into four equal sections, called Quartiles.
- ✓ The first quartile (Q1) is located 25% of the way through the data.
- ✓ The second quartile (also called the Median) is located 50% of the way through the data.
- ✓ The third quartile (Q3) is located 75% of the way through the data.
- ✓ The other values that are needed to construct the display are the Min and Max.
- ✓ All five of these values are found under 1-Var Stats, and the calculator will draw the display for you.

```
1-Var Stats
n=12
minX=62
Q1=67.5
Med=71.5
Q3=76
maxX=81
```

```
STAT → CALC
1: 1-Var Stats...
2nd L1
Enter
```



The **Stem-and-Leaf Plot** is another common way to display the data.

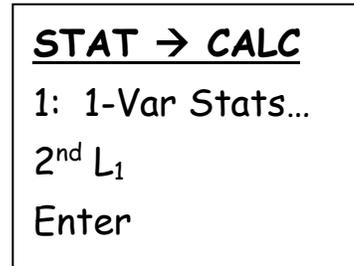
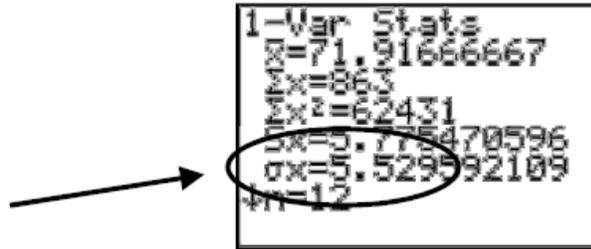
- ✓ This display resembles a bar graph that is rotated on its side.
- ✓ The calculator will not construct this display for you.
- ✓ You need the data arranged in order (SortA has already done this.)
- ✓ The "leafs" are the last single digit in each value.
- ✓ The "stems" are all the digits that precede the last digit.
- ✓ The display needs a Key to explain what the data is representing.

6	2778
7	012339
8	01

Solution: Key: 6 2 = 62 M&Ms.

The **Standard Deviation** (σ) of a set of data is a measurement of how spread out the data are.

- ✓ In notation, it is represented by the Greek letter sigma (σ).
- ✓ On the calculator, it is represented by σx .



Solution: The class had a standard deviation of 5.53.

Notes:

- Standard deviations are a measurement, and therefore must be positive values.
- The smaller the standard deviation, the closer the data are to each other.
- The larger the standard deviation, the farther away the data are from each other.
- A standard deviation is most useful when you compare it against another one.
- If a second class of students had a SD of 6.04 M&Ms, the number of candies per bag in this second class would be more spread out than the original example.

The **Variance** of a set of data is the standard deviation squared.

In notation, it is represented by σ^2 .

- ✓ The calculator will not directly calculate variance. You will need to calculate the σ and then square it yourself.

Solution: The class had a variance of 30.58.

Notes:

- Just like σ 's, variances are a measurement, and therefore must be positive values.
- The smaller the variance, the closer the data are to each other.
- The larger the variance, the farther away the data are from each other.
- Just like σ 's, a variance is most useful when you compare it against another one.

Why do we need two measures of spread?

- ❖ Since the variance is the square of the σ , the value will be much larger. The variance is helpful when you want to emphasize that some of the data values are "weighted" more heavily than others.
- ❖ Data values farther away from the average (either above it or below it) are "weighted" more when computing the variance.
- ❖ The class average was 71.92 M&Ms per bag. The student who only had 62 M&Ms will "influence" the variance more than the student who had 73 M&Ms.

The **Z-Score** (standard score) is the position of one specific data value with respect to the rest of the data values.

- In notation, it is represented by Z.
- The calculator will not compute z-scores directly. They can be calculated using the mean and σ

of the data set with the formula:
$$Z = \frac{X - \bar{X}}{\sigma}$$

Example: To convert the student who had 81 M&Ms in his/her bag into a z-score, find the difference between 81 and the class mean. Then, divide by the σ .

Solution: This student's bag of M&Ms has a z-score of 1.64.

Notes:

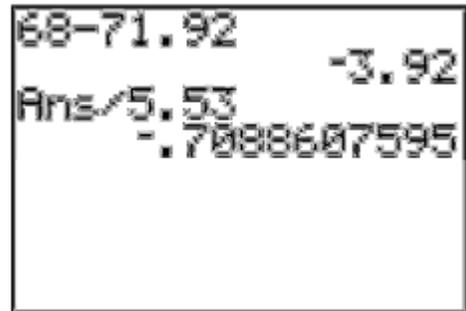
z-scores with positive values indicate the data was above the mean.

z-scores with negative values indicate the data was below the mean.

z-scores are helpful when you compare one against another one – the bag with 68

M&Ms had a z-score of -0.71. This new bag of M&Ms has a negative z-score,

so it has *fewer* candies than the class average. But it's not *as far away* from the class average as the bag of 81 candies is.



Finally...

- ✓ If you have a z-score, the mean, and the SD of a data set, you can work backwards to

calculate the "original" data value using the formula:
$$X = Z\sigma + \bar{X}$$

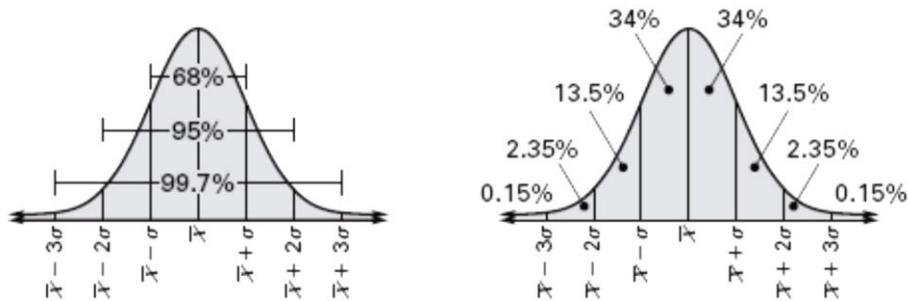
Example: If a bag of M&Ms has a z-score of 2.79...

Solution: ...there would have been 87.35 M&Ms in the bag.

Video link: <http://youtu.be/dend1sIvli4>

Normal curve Notes

You may have heard of “bell-shaped curves”. Statisticians call them Normal models. Normal models are appropriate for distributions whose shapes are unimodal (one high point) and roughly symmetric. Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean. It turns out that in a Normal Model, the “**68-95-99.7” Rule (Empirical Rule)** is true.



Properties of Normal Curve

1. It is **bell** shaped and it is symmetrical about a vertical line, $x = \text{mean}$.
2. The **mean, median** and **mode** are the same and are located at the highest point.
3. **50%** of data values are greater than the **mean**, **50%** of the data values are below the **mean**.
4. It approaches but never touches or crosses the **x**- axis.
5. Approximately 68% of the data values fall between $\mu - 1\sigma$ and $\mu + 1\sigma$.
6. Approximately 95% of the data values fall between $\mu - 2\sigma$ and $\mu + 2\sigma$.
7. Approximately 99.7% of the data values fall between $\mu - 3\sigma$ and $\mu + 3\sigma$.
8. The area under the curve is 1.
9. Z-scores convert all normal distributions to a **standard normal curve** with mean = **0** and standard deviation = **1**.

Empirical Rule

The Empirical Rule gives us approximate percentages. However, we can use the calculator to find the exact percentages...

TI Inspire: In calc>menu>6:Statistics>5:Distributions>2:Normalcdf(lb,ub, μ, σ)

In calc>menu>6:Statistics>5:Distributions>3:Inverse Normal(area to left, μ, σ)

TI-84: In calc>2nd>VARS (DISTR)>2:normacdf(lb,ub, μ, σ)

In calc>2nd> VARS (DISTR)>3:inv Norm(area to left, μ, σ)

Example: A machine fills 12 ounce Potato Chip bags. It places chips in the bags. Not all bags weigh exactly 12 ounces. The weight of the chips placed is normally distributed with a mean of 12.4 ounces and with a standard deviation of 0.2 ounces.

The company has asked you to determine the following probabilities to aid in consumer relations concerning the weight of the bags purchased.

- a. If you purchase a bag filled by this dispenser what is the likelihood it has less than 12 ounces? *In calc>menu>6:Statistics>5:Distribut2:Normalcdf(lb,ub, μ , σ)*

```
normalcdf(-100,12,12.4,.2)
.022750062
```

- b. If you purchase a bag filled by this dispenser what is the likelihood it has more than 12 ounces?

```
normalcdf(12,100,12.4,.2)
.977249938
```

- c. Describe the content of the bottom 5% of bags.

In calc>menu>6:Statistics>5:Distributions>3:Inverse Normal(area to left, μ , σ)

```
invNorm(.05,12.4,.2)
12.07102927
```

- f. Describe the content of the top 2.5% of bags

```
invNorm(1-.025,12.4,.2)
12.7919928
```

Video Link: <http://youtu.be/2jYwxGCCYGO>

Use Combinations

Goal • Use combinations

VOCABULARY **Combination:** A selection of r objects from a group of n objects where the order is not important

COMBINATIONS OF n OBJECTS TAKEN r AT A TIME

The number of combinations of r objects taken from a group of n distinct objects is denoted by ${}_nC_r$

$${}_nC_r = \frac{n!}{(n-r)! \cdot r!}$$

Example 1

Find combinations

Books You are picking 7 books from a stack of 32. If the order of the books you choose is not important, how many different 7 book groups are possible?

The number of ways to choose 7 books from 32 is:

$$\begin{aligned} {}_{32}C_7 &= \frac{32!}{25! \cdot 7!} \\ &= \frac{32 \cdot 31 \cdot 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25!}{25! \cdot 7!} \\ &= 3,365,856 \end{aligned}$$

Example 2

Decide to multiply or add combinations

Movie Rentals The local movie rental store is having a special on new releases. The new releases consist of 12 comedies, 8 action, 7 drama, 5 suspense, and 9 family movies.

- You want exactly 2 comedies and 3 family movies. How many different movie combinations can you rent?
- You can afford at most 2 movies. How many movie combinations can you rent?

Solution

- You can choose 2 of the 12 comedies and 3 of the 9 family movies. So, the number of possible sets of movies is:

$${}_{12}C_2 \cdot {}_9C_3 = \frac{12!}{10! \cdot 2!} \cdot \frac{9!}{6! \cdot 3!} = 66 \cdot 84 = 5544$$

b. You can rent 0,1, or 2 movies. Because there are 41 movies to choose from, the number of possible sets of movies is:

$${}_{41}C_0 + {}_{41}C_1 + {}_{41}C_2 = \underline{1 + 41 + 820} = \underline{862}$$

Example 3

Solve a multi-step problem

Reading A popular magazine has 11 articles. You want to read at least 2 of the articles. How many different combinations of articles can you read?

Solution

For each of the 11 articles, you can choose to read or not read the article, so there are 2^{11} total combinations. If you read at least 2 articles, you do not read only a total of 0 or 1 articles. So, the number of ways you can read at least 2 articles is:

$$\underline{2^{11}} - ({}_{11}C_0 + {}_{11}C_1) = \underline{2048 - (1 + 11)} = \underline{2036}.$$

Notes: Counting Principle and Permutations

Permutation is an ordering of n objects

Factorial: represented by the symbol !, n factorial is defined as: $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$

Fundamental counting principle:

Two events If one event can occur in m ways and another event can occur in n ways, then the number of ways that both events can occur is $m \cdot n$.

Three or More events: $m \cdot n \cdot p$

Example: Use the fundamental counting principle

Pizza You are buying a pizza. You have a choice of 3 crusts, 4 cheeses, 5 meat toppings and 8 vegetable toppings. How many different pizzas with one crust, one cheese, one meat, and one vegetable can you choose?

Solution: Number of pizza = $3 \cdot 4 \cdot 5 \cdot 8 = 480$

Permutations of n Objects taken r at a time:

The number of permutation of r objects taken from a group of n distinct objects is denoted by ${}_n P_r$

$${}_n P_r = \frac{n!}{(n-r)!}$$

Example 2: You have 6 homework assignments to complete over the weekend. However, you only have time to complete 4 of them on Saturday. In how many orders can you complete 4 of the assignments?

Solution: Find the number of permutation of 6 objects taken 4 at a time

$${}_6 P_4 = \frac{6!}{(6-4)!} = \frac{720}{2} = 360 \quad \text{You can complete the 4 assignments in 360 different orders.}$$

Permutations with repetition

The number of distinguishable permutations of n objects where one object is repeated s_1 times, another is repeated s_2 times, and so on, is:

$$\frac{n!}{s_1! \cdot s_2! \cdot \dots \cdot s_k!}$$

Example: Find the number of distinguishable permutations of the letters in (a) EVEN and (b) CALIFORNIA

Solution:

a. EVEN had 4 letters of which E is repeated 2 times. So, the number of distinguishable permutations is

$$\frac{4!}{2!} = \frac{24}{2} = 12$$

b. CALIFORNIA has 10 letters of which A and I are each repeated 2 times. So, the number of distinguishable

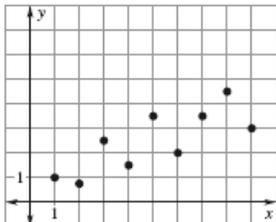
$$\text{permutations is } \frac{10!}{2!2!} = \frac{3,628,800}{2 \cdot 2} = 907,200$$

Statistics: Regression Notes

Goal: Given data, determine the appropriate model. Use the model to make predictions.

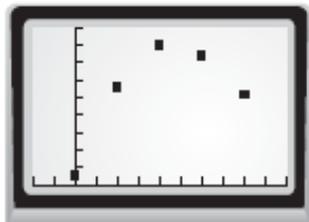
The models are...

Linear



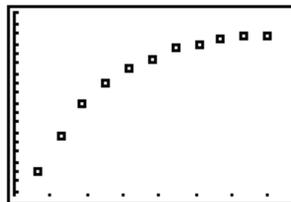
Calc: Linear Reg
Form: $y = ax + b$

Quadratic



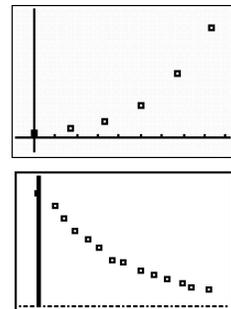
Calc: Quadreg
Form: $y = ax^2 + bx + c$

Logarithmic



Calc: LnReg
Form: $y = a + b \ln x$

Exponential



Calc: ExpReg
Form: $y = a \cdot b^x$

r^2 : is a measure of the proportion of variability explained by, or due to the regression in a sample of data. The closer r^2 is to 100%, the better the fit of the data. Use to determine which model is the best fit

On the calculator: You must enter the data into the lists and look at the scatterplot before running the appropriate regression model.

Links for calculator help:

<http://www.dummies.com/how-to/content/regression-modeling-on-the-ti84-plus.html>

<http://www.dummies.com/how-to/content/perform-regressions-from-within-tinspire-calculato.html>

General steps for regression:

1. Enter data into lists on the calculator. Look at the scatterplot to help determine the model that best fits the data.
2. Using the calculator, find the equation for the model. Enter the equation into $y =$ and graph.
3. When making predictions, if you are given the independent value (x) and asked to predict the dependent value (y), then substitute the number into x in the equation and evaluate to find y . If you are given the dependent value (y) and asked to find the independent value (x), then set the equation equal to the value of y and solve for x . You can also do this on the calculator by graphing the regression equation and $y = \text{value of } y$ and finding the intersection.