# Week Five: Normal Distributions and Sampling Distributions

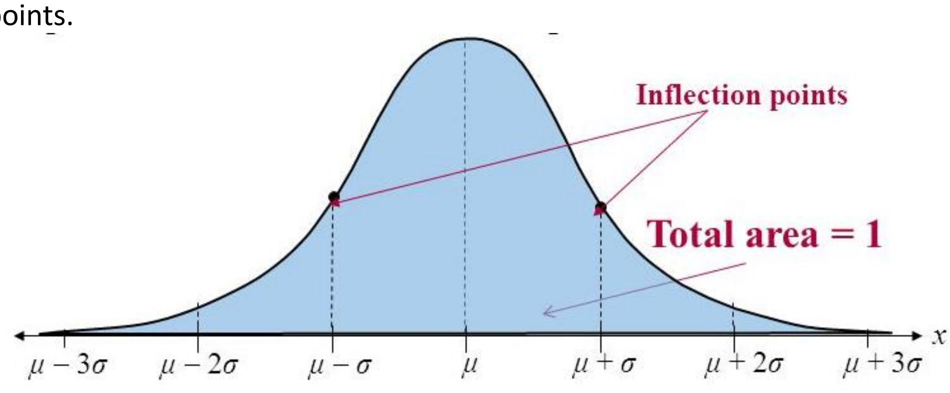
## Week Five Goals

- The Standard Normal Distribution
- Examples Finding a Probability with StatHelper and Minitab Examples
- Examples of Finding Percentiles
- Examples Finding a Value given the Probability with StatHelper and Minitab Examples
- The Normal Distribution and Examples
- Describe the Sampling distribution of $\bar{x}$ (shape, center, spread)
- Examples with Sampling distribution of $\bar{x}$
- The Central Limit Theorem
- The Normal Probability Plot

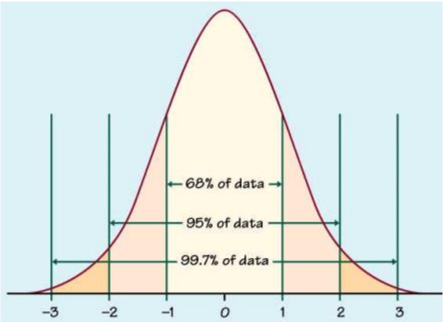
# The Standard Normal Distribution

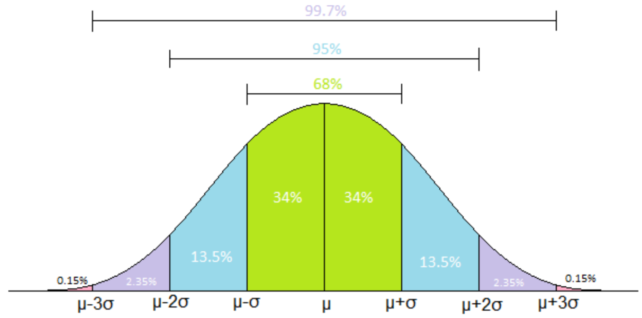
Let's begin by learning about the properties of a normal distribution.

- The mean, median and mode are all equal.
- The normal curve is bell-shaped and symmetric about the mean
- The total area under the curve is equal to one.
- The normal curve approaches, but never touches, the x-axis as it extends farther and farther away from the mean.
- In the center of curve, within the first standard deviation, the graph curves downward (concave down). The graph then curves upward (concave up). The points where concavity changes are called the inflection points.



Next, let's recall The Empirical Rule and the intervals that you learned a few weeks prior.




The Empirical Rule intervals and percentages are only *approximations*. It is helpful to know these percentages so that, under the assumption of normality, we can easily find the middle 68% (or roughly two-thirds) of your data or the middle 95% of your data. Statisticians are able to calculate the probabilities under this normal curve at absolutely any point, not just one, two or three standard deviations from the mean.

The Standard Normal Table (Z-table) was developed in the late 1700's and is still widely used today. Knowing that a z-score of ZERO is at the center of the normal curve, we are able to calculate the probability of being less than or greater than any Z-value (rounded to the nearest hundredth). Below is a snapshot of a Z-table.

For example: Assuming Z ~ N(0, 1), if we wanted the probability of getting a Z-score that is less than 1.54, we would locate 1.54 on the table. Assuming Z ~ N(0, 1), P( Z < 1.54) = **.93822**.

If we wanted the probability of getting a Z-score that is **at least 1.54** (or greater than 1.54) we would subtract from 1. **1 - .93822 = 0.06178.**

**STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.**

| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .50000 | .50399 | .50798 | .51197 | .51595 | .51994 | .52392 | .52790 | .53188 | .53586 |
| 0.1 | .53983 | .54380 | .54776 | .55172 | .55567 | .55962 | .56356 | .56749 | .57142 | .57535 |
| 0.2 | .57926 | .58317 | .58706 | .59095 | .59483 | .59871 | .60257 | .60642 | .61026 | .61409 |
| 0.3 | .61791 | .62172 | .62552 | .62930 | .63307 | .63683 | .64058 | .64431 | .64803 | .65173 |
| 0.4 | .65542 | .65910 | .66276 | .66640 | .67003 | .67364 | .67724 | .68082 | .68439 | .68793 |
| 0.5 | .69146 | .69497 | .69847 | .70194 | .70540 | .70884 | .71226 | .71566 | .71904 | .72240 |
| 0.6 | .72575 | .72907 | .73237 | .73565 | .73891 | .74215 | .74537 | .74857 | .75175 | .75490 |
| 0.7 | .75804 | .76115 | .76424 | .76730 | .77035 | .77337 | .77637 | .77935 | .78230 | .78524 |
| 0.8 | .78814 | .79103 | .79389 | .79673 | .79955 | .80234 | .80511 | .80785 | .81057 | .81327 |
| 0.9 | .81594 | .81859 | .82121 | .82381 | .82639 | .82894 | .83147 | .83398 | .83646 | .83891 |
| 1.0 | .84134 | .84375 | .84614 | .84849 | .85083 | .85314 | .85543 | .85769 | .85993 | .86214 |
| 1.1 | .86433 | .86650 | .86864 | .87076 | .87286 | .87493 | .87698 | .87900 | .88100 | .88298 |
| 1.2 | .88493 | .88686 | .88877 | .89065 | .89251 | .89435 | .89617 | .89796 | .89973 | .90147 |
| 1.3 | .90320 | .90490 | .90658 | .90824 | .90988 | .91149 | .91309 | .91466 | .91621 | .91774 |
| 1.4 | .91924 | .92073 | .92220 | .92364 | .92507 | .92647 | .92785 | .92922 | .93056 | .93189 |
| 1.5 | .93319 | .93448 | .93574 | .93699 | .93822 | .93943 | .94062 | .94179 | .94295 | .94408 |
| 1.6 | .94520 | .94630 | .94738 | .94845 | .94950 | .95053 | .95154 | .95254 | .95352 | .95449 |

Here is a link to the Standard Normal Table. I am sharing it only to provide background information.
**We will use technology to access the Z-tables**. StatHelper, Minitab and the TI calculators are programmed with this table.

**<u>IMPORTANT INFORMATION</u>**

The probability **<u>less than</u>** a Z-score results in the same value as **<u>less than or equal to</u>**.
The probability that a z-score is **AT MOST** a value is the same as **<u>less than or equal to</u>**.

The probability **<u>greater than</u>** a Z-score results in the same value as **<u>greater than or equal to</u>**.
The probability that a z-score is **AT LEAST** a value is the same as **<u>greater than or equal to</u>**.

Probabilities smaller than 0.05 are considered RARE events and
we say the event is UNLIKELY to occur; we would not expect this event to occur if there is less than a 5% chance of it happening.

Probabilities greater than 0.05 are considered LIKELY events and
we would **<u>expect</u>** the event to occur.

## Examples 1 - 3

Here are three examples of finding probabilities using the standard normal probability distribution. I work through these examples in my videos.

1. Assuming $Z \sim N(0, 1)$, $P( Z < 1.54) =$ **.938**

2. Assuming $Z \sim N(0, 1)$, $P( Z \geq 1.54) =$ **.062**

3. Assuming $Z \sim N(0, 1)$, $P( -.94 < Z < 1.77) =$ **.788**

## Video Explaining the Normal Distribution and Examples with Z-Tables from Prof. Coffey

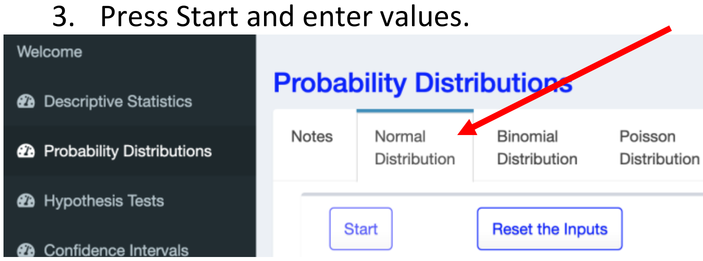
Here is a video introduction to the standard normal table (Z-table): https://youtu.be/IB_97fqjXms

Here is the same video with interpreting: https://youtu.be/fhRWzLRNvsM

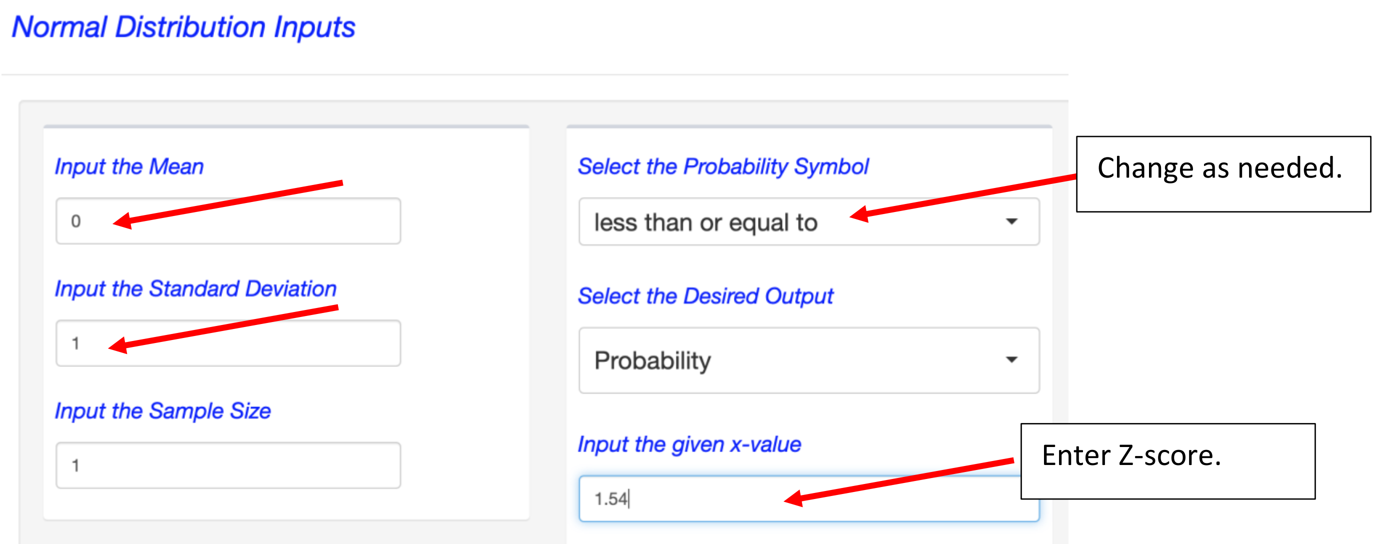
The videos below show these examples completed with StatHelper and Minitab for Mac. Now that you know what the table is, please use technology to solve all remaining problems.

# Examples Finding a Probability with StatHelper

1. Select the feature: **Probability Distributions** on the left.
2. Select the Normal Distribution tab.
3. Press Start and enter values.




If we are working with the standard normal curve, we must assume a mean = 0 and standard deviation = 1. These values must be entered.




Video Showing Examples 1 - 3 using StatHelper from Prof. Coffey

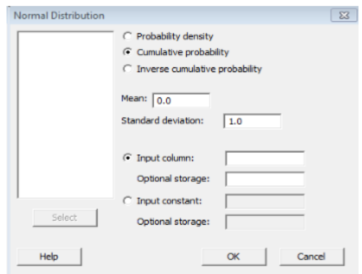
Here is a video with StatHelper examples: [https://youtu.be/IvHuKHVGOQQ](https://youtu.be/IvHuKHVGOQQ)

# Examples Finding a Probability with Minitab

NOTE: The keystrokes depend on whether you are using Minitab 19 for Windows or Minitab 19 for Mac.

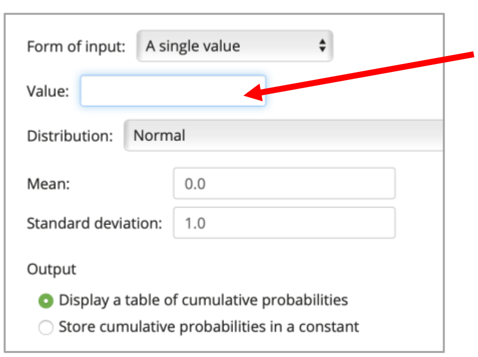
## Minitab 19 for Windows

1. Choose **Calc** > **Probability Distributions** > **Normal**.
2. Select **Cumulative probability**.
3. In **Mean**, enter the mean. ← this should be 0 (default)
4. In **Standard deviation**, enter the standard deviation. ← this should be 1 (default)
5. In **Input value**, enter the Z-score.
6. Click **OK**.




## Minitab 19 for Mac

1. Choose **Calc** > **Probability Distributions** > **Cumulative probability**
2. Form of Input: A single value
3. Value: Enter the Z-score here
4. Distribution: Normal (default)
5. Mean: 0 (default)
6. Standard deviation: 1 (default)
7. Output: Display a table of cumulative probabilities




## Video showing Examples 1 -3 using Minitab for Mac from Prof. Coffey

Here is a video with Minitab for Mac examples: [https://youtu.be/cXSabyvAygg](https://youtu.be/cXSabyvAygg)

# Examples of Finding Percentiles

If we know the probability of getting a value and want THE VALUE itself, it is the same as finding the percentile.

If a question asks: find the z-score such that 20% of z-scores are less than or equal to it then this is the same as asking for the **20th percentile**.

If a question asks: find the z-score such that 80% of z-scores are greater than the value then this is the same as the **20$^{th}$ percentile**.

For example: Assuming Z ~ N(0, 1), find the z-score that represents the 63rd percentile (rounded to the nearest hundredth).
To answer this with the Z-table, we would first find .63 in the table. We are not able to find it exactly, but we are quite close with .62930. The Z-score that corresponds with this probability is the answer.
**ANSWER = 0.33**

**STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.**

| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0.0** | .50000 | .50399 | .50798 | .51197 | .51595 | .51994 | .52392 | .52790 | .53188 | .53586 |
| **0.1** | .53983 | .54380 | .54776 | .55172 | .55567 | .55962 | .56356 | .56749 | .57142 | .57535 |
| **0.2** | .57926 | .58317 | .58706 | .59095 | .59483 | .59871 | .60257 | .60642 | .61026 | .61409 |
| **0.3** | .61791 | .62172 | .62552 | .62930 | .63307 | .63683 | .64058 | .64431 | .64803 | .65173 |
| **0.4** | .65542 | .65910 | .66276 | .66640 | .67003 | .67364 | .67724 | .68082 | .68439 | .68793 |
| **0.5** | .69146 | .69497 | .69847 | .70194 | .70540 | .70884 | .71226 | .71566 | .71904 | .72240 |
| **0.6** | .72575 | .72907 | .73237 | .73565 | .73891 | .74215 | .74537 | .74857 | .75175 | .75490 |
| **0.7** | .75804 | .76115 | .76424 | .76730 | .77035 | .77337 | .77637 | .77935 | .78230 | .78524 |
| **0.8** | .78814 | .79103 | .79389 | .79673 | .79955 | .80234 | .80511 | .80785 | .81057 | .81327 |
| **0.9** | .81594 | .81859 | .82121 | .82381 | .82639 | .82894 | .83147 | .83398 | .83646 | .83891 |
| **1.0** | .84134 | .84375 | .84614 | .84849 | .85083 | .85314 | .85543 | .85769 | .85993 | .86214 |
| **1.1** | .86433 | .86650 | .86864 | .87076 | .87286 | .87493 | .87698 | .87900 | .88100 | .88298 |
| **1.2** | .88493 | .88686 | .88877 | .89065 | .89251 | .89435 | .89617 | .89796 | .89973 | .90147 |
| **1.3** | .90320 | .90490 | .90658 | .90824 | .90988 | .91149 | .91309 | .91466 | .91621 | .91774 |
| **1.4** | .91924 | .92073 | .92220 | .92364 | .92507 | .92647 | .92785 | .92922 | .93056 | .93189 |
| **1.5** | .93319 | .93448 | .93574 | .93699 | .93822 | .93943 | .94062 | .94179 | .94295 | .94408 |
| **1.6** | .94520 | .94630 | .94738 | .94845 | .94950 | .95053 | .95154 | .95254 | .95352 | .95449 |

Here are examples of finding Z-scores using the standard normal probability distribution and knowing the probability. The videos below show these examples completed with StatHelper and Minitab for Mac.

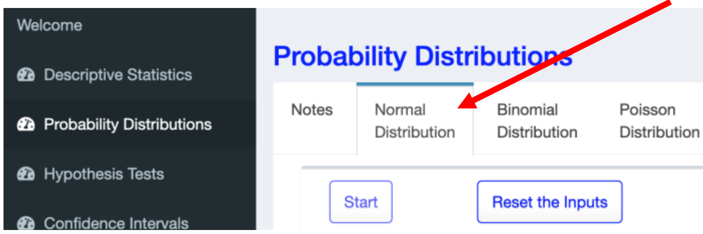
## Examples 4 - 5

4. Assuming Z ~ N(0, 1), find the z-score that represents the 63rd percentile (rounded to the nearest hundredth). **ANSWER = 0.33**

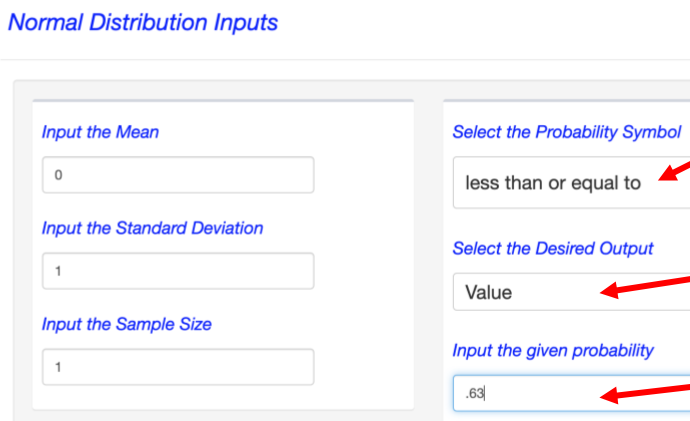
5. Assuming Z ~ N(0, 1), find the z-score such that 13% of z-scores are greater than it (rounded to the nearest hundredth). **ANSWER = 1.13**

# Examples Finding a Value given the Probability with StatHelper

1. Select the feature: **Probability Distributions** on the left.
2. Select the Normal Distribution tab.
3. Press Start and enter values.




If we are working with the standard normal curve, we must assume a mean = 0 and standard deviation = 1. These values must be entered.




Finding a percentile: choose less than or equal to.

Finding a value such that X% is greater: choose greater than or equal to.

We are finding a **VALUE**.

Type in the probability (i.e. decimal; a $0 \leq P \leq 1$)

## Video showing the Inverse Cumulative Examples using StatHelper from Prof. Coffey

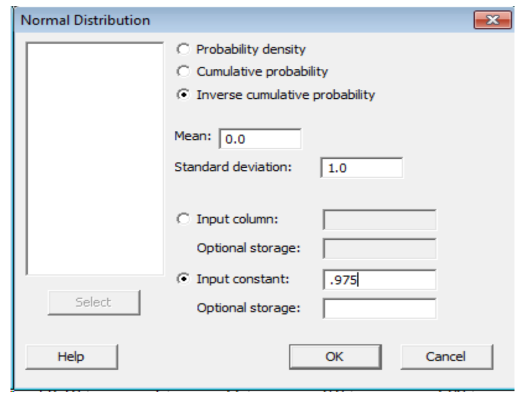
Here is a video with StatHelper examples: [https://youtu.be/wVf_tOreaoQ](https://youtu.be/wVf_tOreaoQ)

# Examples Finding a Value given the Probability with Minitab

NOTE: The keystrokes depend on whether you are using Minitab 19 for Windows or Minitab 19 for Mac.

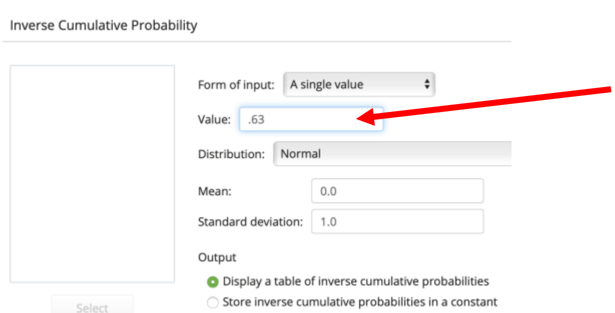
## Minitab 19 for Windows

1. Choose **Calc** > **Probability Distributions** > **Normal**.
2. Select **Inverse cumulative probability**.
3. In **Mean**, enter the mean. ← this should be 0 (default)
4. In **Standard deviation**, enter the standard deviation. ← this should be 1 (default)
5. In **Input value**, enter the probability (i.e. decimal; a 0 ≤ P ≤ 1)
6. Click **OK**.




## Minitab 19 for Mac

1. Choose **Calc** > **Probability Distributions** > **Inverse Cumulative probability**
2. Form of Input: A single value
3. Value: Enter the probability here (i.e. decimal; a 0 ≤ P ≤ 1)
4. Distribution: Normal (default)
5. Mean: 0 (default)
6. Standard deviation: 1 (default)
7. Output: Display a table of cumulative probabilities




## Video showing the Inverse Cumulative Examples using Minitab for Mac from Prof. Coffey

Here is a video with Minitab for Mac examples: https://youtu.be/mB3zRz4ACGE

# The Normal Distribution and Examples

The normal distribution uses the standard normal distribution table by first standardizing the values into z-scores. If we are told that a population can be assumed normal and we know the mean and standard deviation, we can answer probability questions and percentile questions without having to standardize the values. When you hear 'standard normal distribution' it means that you are working with Z-scores. When you hear 'normal distribution' it means that you are working in the units your variable is measured.

We will assume the variable we are studying, X, is a random variable and if it follows a NORMAL distribution with a known center (typical) value and spread, then we can say X ~ N(center, spread). If you were asked to find a probability by hand and with the table, then you would first have to convert the value of X into a Z-score, and then follow the process taught in the previous section on the Standard Normal Distribution.

**There is no need to use the table anymore. We have technology.**
**We will not do this by hand; we will use the cumulative and inverse cumulative features that you just learned in StatHelper, Minitab or your technology of choice.**

**Assuming X ~ N(center, spread), we can find the probability of getting**

- less than (or equal to) a certain value of X (or at most a certain value)
- greater (or equal to) a certain value of X (or at least a certain value)
- between two distinct values of X.

Since we can use technology to answer such questions will not need to convert to a Z-score. The technology can calculate the z-score for X behind the scenes and find the corresponding probability or value.

**Assuming X ~ N(center, spread), we can find a certain value of x, when given a probability or percentile by using the inverse cumulative probability distribution.**

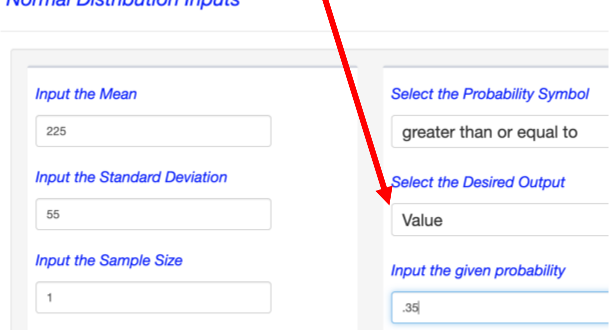
## Example 6: Finding a Value, Given a Probability with Technology

**People's monthly electric bills in S-town are normally distributed with a mean of $225 and a standard deviation of $55. Those folks spend a lot of time online. Name the monthly electric bill cost, in dollars, such that 35% of bills are greater. Round to the nearest dollar.** **ANSWER: $246**

If you are given the probability, you chose INVERSE Cumulative to find the value.

StatHelper

If you are given the probability, the desired output is the value.

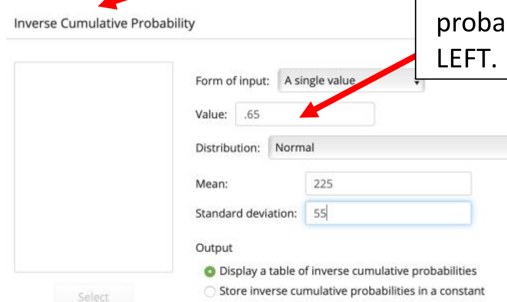



$x = 246.45$

**Now the value that 35 percent of the data fall below is 246.45**

Minitab 19 for Mac

You must provide the probability to the LEFT.




Inverse Cumulative Distribution Function

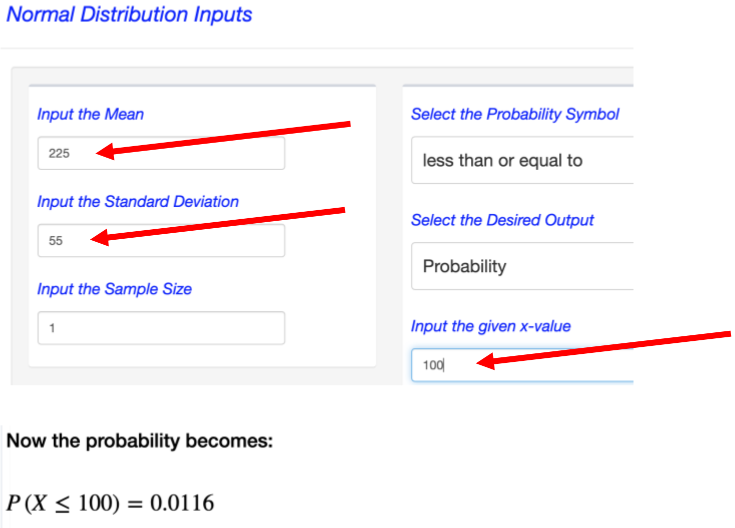
Normal with mean = 225 and standard deviation = 55

| P( X ≤ x ) | x |
|---|---|
| 0.65 | 246.193 |

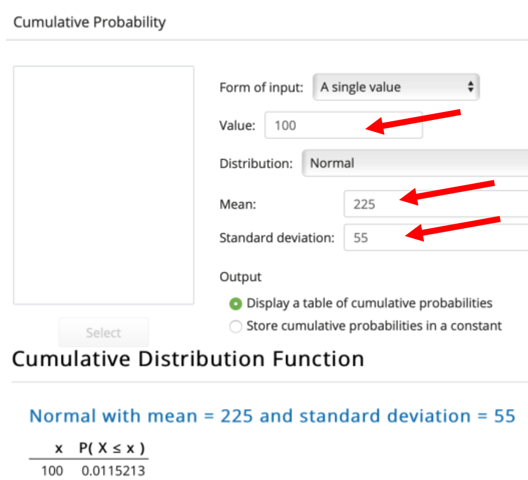
## Example 7: Finding a Probability with Technology

People's monthly electric bills in S-town are normally distributed with a mean of $225 and a standard deviation of $55. Those folks spend a lot of time online. What is the probability that a customer has a bill that is $100 or less? Round to three decimal places. **ANSWER: 0.012**

StatHelper




Minitab 19 for Mac




## Video explaining the Normal Distribution with Examples 6 & 7 from Prof. Coffey

Here is a video on the Normal Distribution: https://youtu.be/Ula6TnAvtnQ

Here is the same video with interpreting: https://youtu.be/ZAtcvAnGj2o

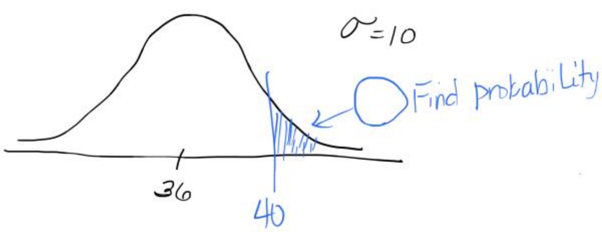
## Example 8 adapted from e-book homework section #75

The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 36 and a standard deviation of 10. Suppose that one individual is randomly chosen. Let *X* = percent of fat calories.

A. Write using this notation: *X* ~ _____(_____,_____)

ANSWER: X ~ N(36, 10)

B. Find the probability that the percent of fat calories a person consumes is more than 40. Graph the situation. Shade in the area to be determined.

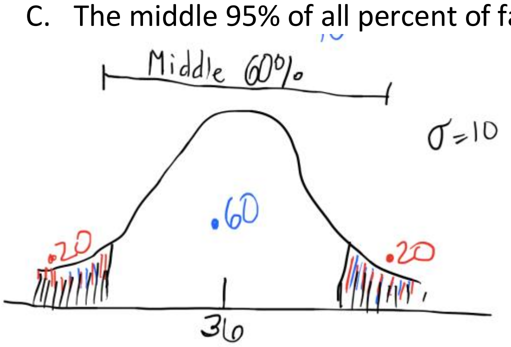



ANSWER: .4168

StatHelper output:

$$P(X \geq 40) = 1 - P(X \leq 40)$$
$$P(X \geq 40) = 1 - 0.5832$$
$$P(X \geq 40) = 0.4168$$

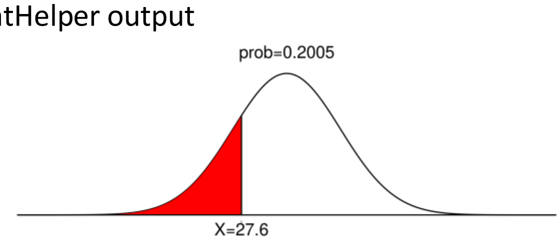
C. The middle 95% of all percent of fat calories fall between which two values?

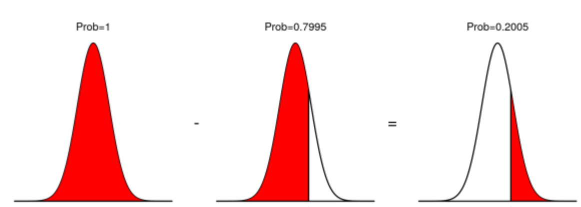



ANSWER: The middle 60% leaves 40% to be split up into the two tails of the graph. That results in 20% in each tail (since the curves are symmetric). Notice that the three regions add up to 1 (100%). The middle 60% falls between 27.6 and 44.4

Find the value such that 20% is less than it: 27.6 Find the value such that 20% is greater than it: 44.4

StatHelper output



# The Central Limit Theorem

Here's a simple example of the theory: when you roll a single die, your odds of getting any number (1, 2,3 ,4 ,5, or 6) are the same $= {}^{1}/_{6}$. The mean for any roll is (1 + 2 + 3 + 4 + 5 + 6) / 6 = 3.5. The results from a one-die roll are shown in the first figure below: it looks like a uniform distribution. However, as the sample size is increased (two dice, three dice…), the mean of the sampling distribution of the means looks more and more like a **normal** distribution. That is what the central limit theorem predicts.

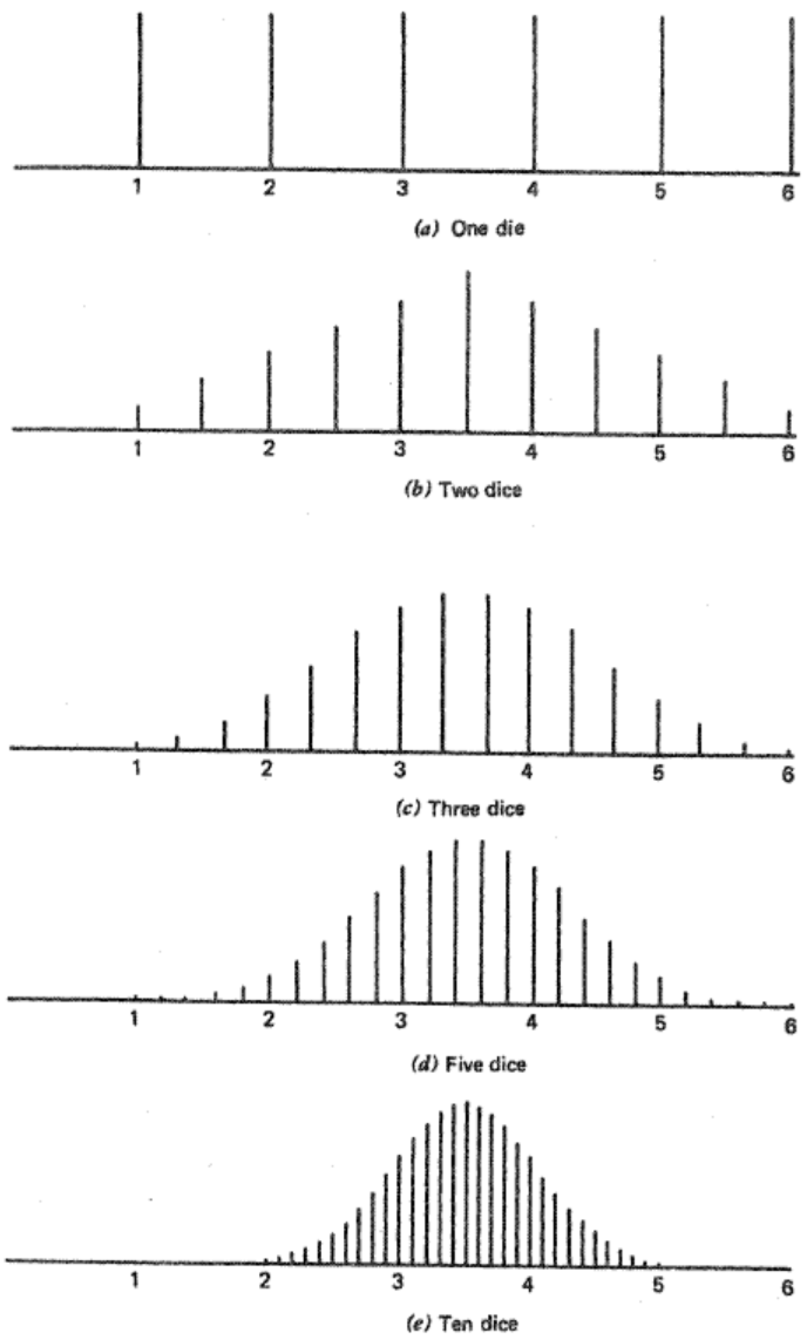



*Image: U of Michigan.*

**The Central Limit Theorem - CLT**

If simple random samples of size n are selected from a distribution with mean μ and standard deviation σ, then the distribution of sample means approaches a normal distribution, as sample size increases, no matter the shape of the parent population's distribution. The greater the sample size, the better is the approximation.

To continue, given a sufficiently large sample size from a population, the mean of all samples from that same population will be approximately equal to the mean of the population. Furthermore, all of the samples will follow an approximate normal distribution pattern, with the standard deviation of the sample means being approximately equal to the standard deviation of the population divided by the square root of the sample size.

According to the central limit theorem, the mean of a sample of data will be closer to the mean of the overall population in question as the sample size increases, notwithstanding the actual distribution of the data, and whether it is normal or non-normal. As a general rule, sample sizes equal to or greater than 30 are considered sufficient for the central limit theorem to hold, meaning the distribution of the sample means is fairly normally distributed. Learn more: http://www.investopedia.com/terms/c/central_limit_theorem.asp

In Other Words…

Regardless of the shape of the underlying population, the sampling distribution of sample means can be assumed normal for a 'large enough' sample size. We will consider n $\geq$ 30 to be 'large enough'.

The sampling distributions center will be the population mean.

The spread of the sampling distribution will be $\frac{\sigma}{\sqrt{n}}$.

In my video, I use the following website to demonstrate CLT and sampling distributions of x-bar: http://onlinestatbook.com/stat_sim/sampling_dist/

## Video explaining the Central Limit Theorem (CLT) from Prof. Coffey

Here is a video on the CLT: https://youtu.be/ckkogy4rUi8

Here is the same video with interpreting: https://youtu.be/DkfHDkJexWs

# Describe the Sampling Distribution of X-bar

When asked to describe the sampling distribution of the sample means, $\bar{x}$, I want you to address the shape, the center and the spread of the sampling distribution.

**Shape: You can assume the sampling distribution is normal if you satisfy ONE of the following:**

- ***Is the population distribution normal?*** If yes, then you can state that the sampling distribution of the sample means is normal (regardless of the sample size). In other words, if the population is normal (bell-shaped), then when I sample from it, the shape of that distribution is expected to also be normal.
  or
- ***Is the sample size large (n $\geq$ 30)?*** If yes, then state that since the sample size is large enough (n $\geq$ 30), we can assume that the sampling distribution of the sample means is approximately normal (regardless of the underlying shape of the population distribution).

Side note: Later in the course, we will encounter situations when the sample size is too small and we have raw data. I will introduce you to the technique of building a Normal Probability Plot (NPP) at the end of this lesson. This technique is used when n < 30. It is not needed in this part of the lesson since we will not work with raw data here.

**Center: The mean of the sample means is the population mean. In symbols we write:** $\boldsymbol{\mu_{\bar{x}} = \mu}$

**Spread: The standard deviation of the sample means is the population standard deviation divided by the square root of the sample size. This is also called the standard error of the mean (SE). In symbols we write:** $\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

## Video explaining Sampling Distributions from Prof. Coffey

Here is a video describing Sampling Distributions: https://youtu.be/xGuIA23nhn8

Here is the same video with interpreting: https://youtu.be/reo8jAiZfEs

Once we have established the shape to be assumed normal, determined the center and calculated the standard error (the spread), we can use the standard normal tables— that are programmed into the technology we have been using—to answer probability questions.

1. Determine shape, center and spread and [optional] rewrite the problem X ~ N(mean, standard error)
2. If asked to find a probability, use the CUMMULATIVE distribution function.
3. If asked to find a sample mean, given a percentile/probability, use the INVERSE CUMMULATE distribution function.

NOTE: StatHelper calculates the standard error for you and shows you the background work as long as you enter in the mean, standard deviation, sample size (n) and the value and direction (<, >, or between).

NOTE: Minitab 19 requires you to calculate the standard error and for you to enter this value in where is asks for standard deviation. When you work with sampling distributions, the standard deviation of the sample means is used IN PLACE OF THE standard deviation.

# Examples with Sampling Distributions of $\bar{x}$

## Example 9: StatHelper Sampling Distribution Example

The population of RIT student commuters has a mean commute time of 17 minutes and a standard deviation of 6 minutes and a skewed right distribution. A random sample of n = 57 commuters is taken.

**A. Describe the sampling distribution of the sample means.**

1. Describe the shape by answering the question: what is the shape of the sampling distribution of the sample means, $\bar{x}$ ?

ANSWER: Since the sample size n = 57 is greater than 30, the CLT allows us to assume a **normal** sampling distribution.

2. Describe the center by answering the question: what is the mean of the sampling distribution of the sample means, $\bar{x}$ ?

ANSWER: The mean of the sampling distribution should be the same as the mean of the population distribution. In other words, the mean of the sampling distribution is 17 minutes.

3. Describe the spread by answering the question: what is the standard deviation of the sampling distribution of the sample means, $\bar{x}$ ? In other words, what is the standard error of $\bar{x}$ ? Round to four decimal places.

ANSWER: The standard error of $\bar{x}$ is $\frac{6}{\sqrt{57}} = .7947$

Confirmed with StatHelper output

$\frac{6}{\sqrt{57}}$ 0.7947

**B. Using the population and sampling information above, what is the probability of a sample mean commute time between 17. 5 and 19 minutes? Round to four decimal places.**

ANSWER: .2585

StatHelper output

$$P(17.5 \le X \le 19) = P(X \le 19) - P(X \le 17.5)$$
$$P(17.5 \le X \le 19) = 0.9941 - 0.7357$$
$$P(17.5 \le X \le 19) = 0.2585$$

**C. Using the population and sampling information above, what is the sample mean commute time that represents the 70th percentile? Interpret this value in the context of the problem. Round to the nearest tenth of a minute.**

ANSWER: 70% of sample mean commute times will be less than 17.4 minutes.

StatHelper output

$x = 17.413$

Now the value that 70 percent of the data fall below is 17.413

## Video discussing Example 9 using StatHelper from Prof. Coffey

Here is a video discussing the example above with StatHelper: https://youtu.be/86lFg3TeOJQ

Here is the same video with interpreting: https://youtu.be/CJkPcnPG5kc

## Example 10: Minitab Sampling Distribution Example

[This example was taken and modified from Example 7.2 of your e-text.]

**The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a mean of two hours and a standard deviation of 0.5 hours. A sample of size *n* = 50 is drawn randomly from the population.**

**A. Describe the sampling distribution of the sample means.**

1. Describe the shape:

ANSWER: Since the sample size n = 50 is greater than 30, the CLT allows us to assume a **normal** sampling distribution.

2. Describe the center:

ANSWER: The mean of the sampling distribution should be the same as the mean of the population distribution. In other words, the mean of the sampling distribution is 2 hours.

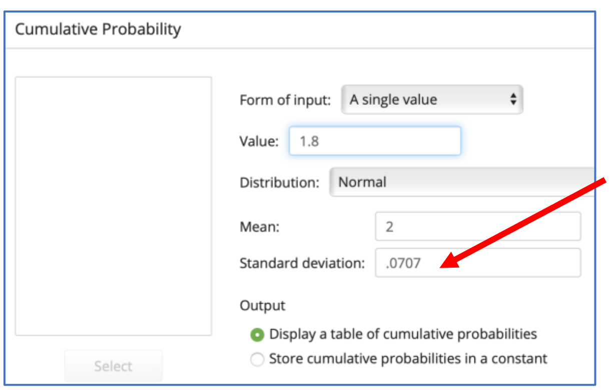
3. Describe the spread (round to four decimal places):

ANSWER: The standard error of $\bar{x}$ is $\frac{0.5}{\sqrt{50}}$ =.0707

**B. Find the probability that the sample mean is between 1.8 hours and 2.3 hours.**

Answer: We need to find $P(1.8 \le \bar{x} \le 2.3)$. Find P(X $\le$ 1.8) = 0.0023357, Find P (X $\le$ 2.3) = .999989; Subtract: .99989 - .0023357 = .9975

Minitab 19 for Mac (repeat for X = 1.8 and X = 2.3) Minitab 19 for Windows

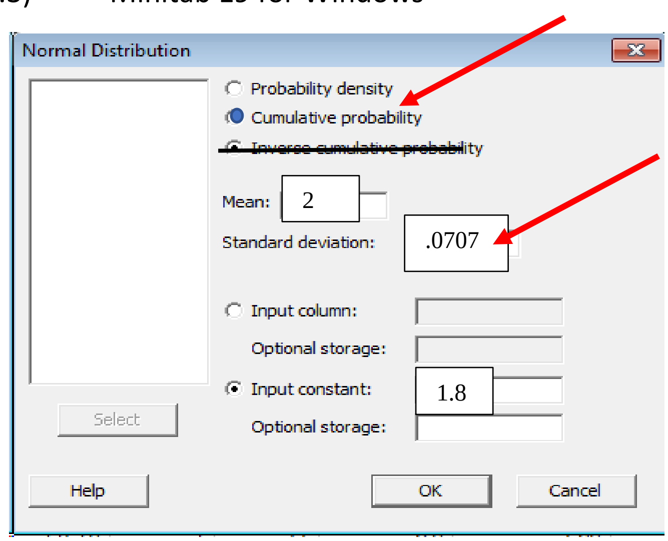



Normal with mean = 2 and standard deviation = 0.0707

| x | P( X ≤ x ) |
|---|---|
| 1.8 | 0.0023357 |

Normal with mean = 2 and standard deviation = 0.0707

| x | P( X ≤ x ) |
|---|---|
| 2.3 | 0.999989 |




**C. Find the probability that the sample mean is at most 1.5 hours.**

ANSWER: We are still assuming a normal distribution, center = 2, standard error = .0707

NOTE: At most means less than or equal to.

We will find P(x $\le$ 1.5) = 0.0000 ← [we call this VERY SMALL in statistics. It is not exactly zero; it was rounded]

It is unlikely to have a sample mean length of a soccer match (for the over 40 group) that is at most 1.5 hours.

Video discussing Example 10 above using Minitab 19 for Mac from Prof. Coffey

Here is a video discussing the example above with Minitab 19: https://youtu.be/lqwX6n0jKJY

Here is the same video with interpreting: https://youtu.be/6BSn5u7JnOo

# The Normal Probability Plot

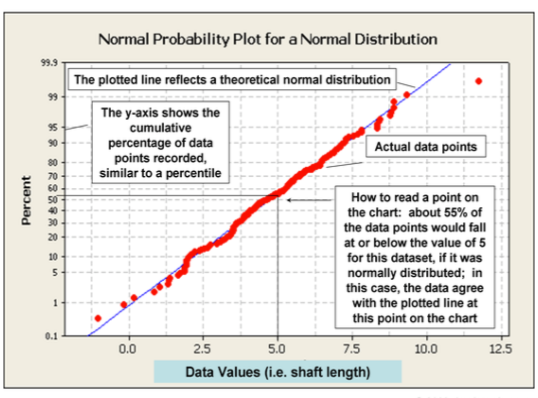
Later in this course, we will be required to determine whether or not our sample data can be assumed to follow a normal distribution. If the sample size is large enough, n > 30, we will use the CLT and assume a normal distribution. If the sample size is considered small, n < 30, and we have the raw sample data, then we will construct a Normal Probability Plot (NPP). If the NPP is somewhat linear (Anderson Darling P-value > 0.05), then we can assume a normal distribution.

The Normal Probability Plot (NPP) is a graphical technique for assessing whether or not sample data come from an approximately normally distributed population. We generally build such a plot when the sample size is considered too small, n < 30.

The normal probability plot (NPP) is a graph that plots observed data versus normal scores (the expected z-score of the data value, assuming that the distribution of the random variable is normal). If this graph is ***approximately linear***, then we can assume the sample data come from a normally distributed population.

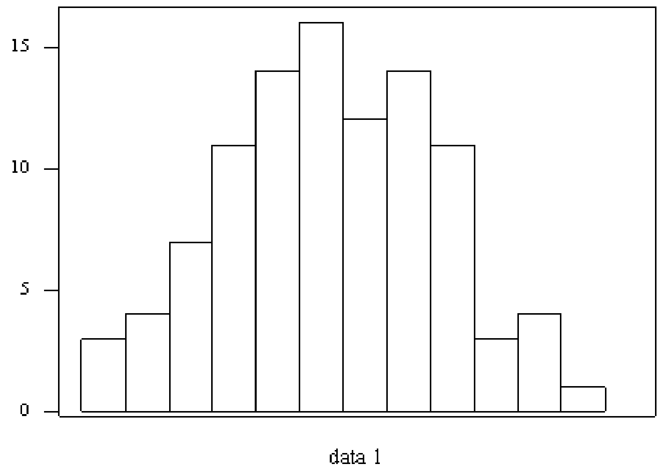
The basic premise is that the plot compares the data with what would be expected of data that is perfectly normally distributed. The 'original data' and 'the idealized normally distributed data' are then compared. If the two generally agree, that means the data agree with what would be expected from a normal distribution. The normal probability plot is then *linear*.

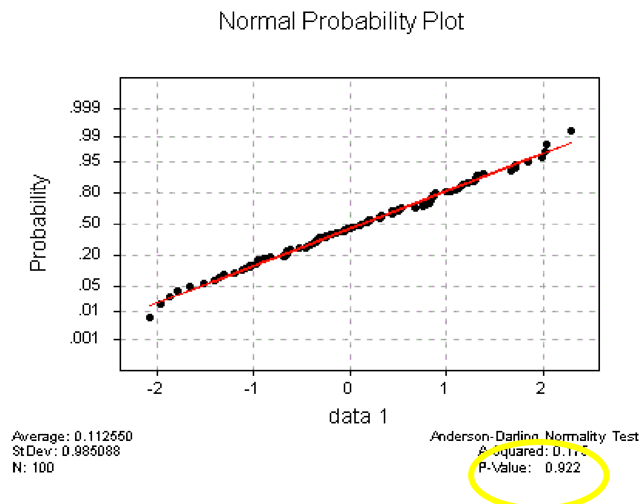
Otherwise, the plot will not be linear. Of course, no plot will be exactly linear, because data is subject to randomness in its collection. ***We look for a general pattern of linearity and a P-value > 0.05.***
http://www.oswego.edu/~srp/stats/normal_prb_plot.htm



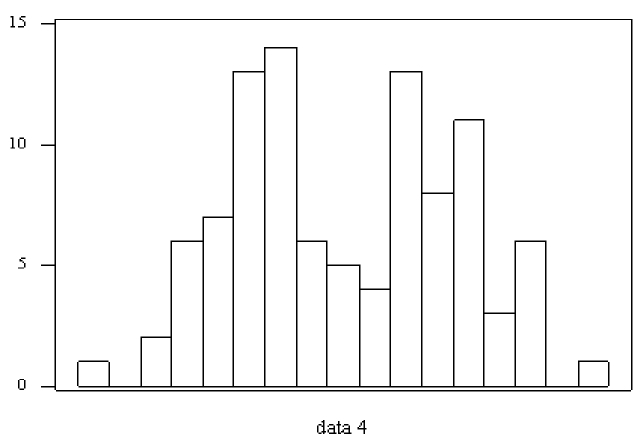
Understanding the relationship between a histogram and a normal probability plot: [http://www.oswego.edu/~srp/stats/normal_prb_plot.htm](http://www.oswego.edu/~srp/stats/normal_prb_plot.htm)

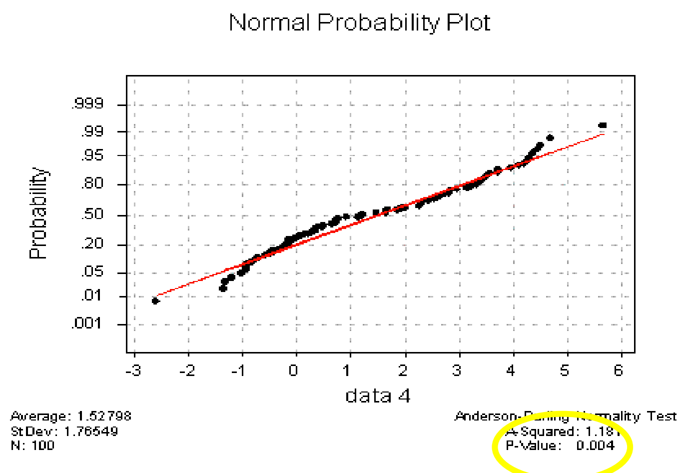
Notice that the histogram below is bell-shaped. The corresponding NPP for this data is basically a straight line. That is the idea: Normal data → NPP that is linear. Note the P-value is high.




Notice that the histogram below is bi-modal. The corresponding NPP for this data is not very straight line. That is the idea: Non-normal data → NPP that is NOT linear. Note the P-value is very small.



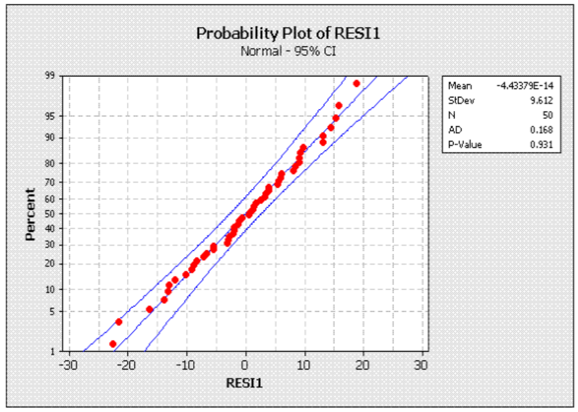

At this point in our study of statistics, we will rely on the possible linear relationship of the normal probability plot (NPP) to determine normality and you may choose to look at the P-value if you questioning the linearity of the plot. It is too early to teach you why this is true, but here is a fact: If the P-value of a normality test is greater than 0.05, we will assume that the sample data come from a normal distribution.

**In summary: if the P-value > 0.05, then we will assume the population follows a normal distribution. If the P-value ≤ 0.05 we will not be able to assume the population follows a normal distribution and we should not go on to use the standard normal tables programmed into our technology to answer questions.**

***NOTE: I am aware that I have not explained WHY the P-value needs to be greater than 0.05. This will occur later in the course.***

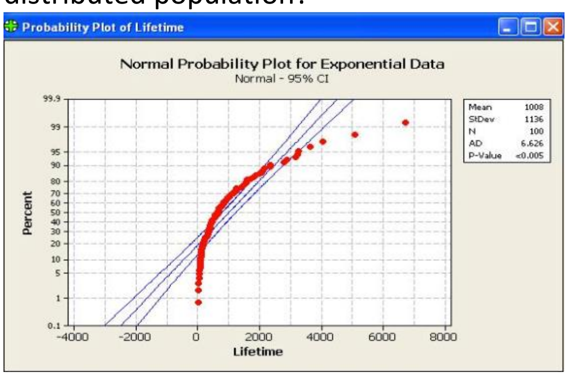
# Interpreting a Normal Probability Plot

Example 1 Given the Normal Probability Plot (NPP) below, can we assume the data come from a normally distributed population?




ANSWER: Since the P-value = .931, which is greater than .05, we can assume the population follows a normal distribution.

Example 2 Given the Normal Probability Plot (NPP) below, can we can assume the data come from a normally distributed population?




ANSWER: NO! Since the P-value <.005 (which is less than .05), we CANNOT assume the population follows a normal distribution.

***NOTE: I am aware that I have not explained WHY the P-value needs to be greater than 0.05. This will occur later in the course.***

## Video discussing Normal Probability Plots from Prof. Coffey

Here is a video discussing NPP: https://youtu.be/sf11jvkPbfI

Here is the same video with interpreting: https://youtu.be/qrA2Zu3n8uE

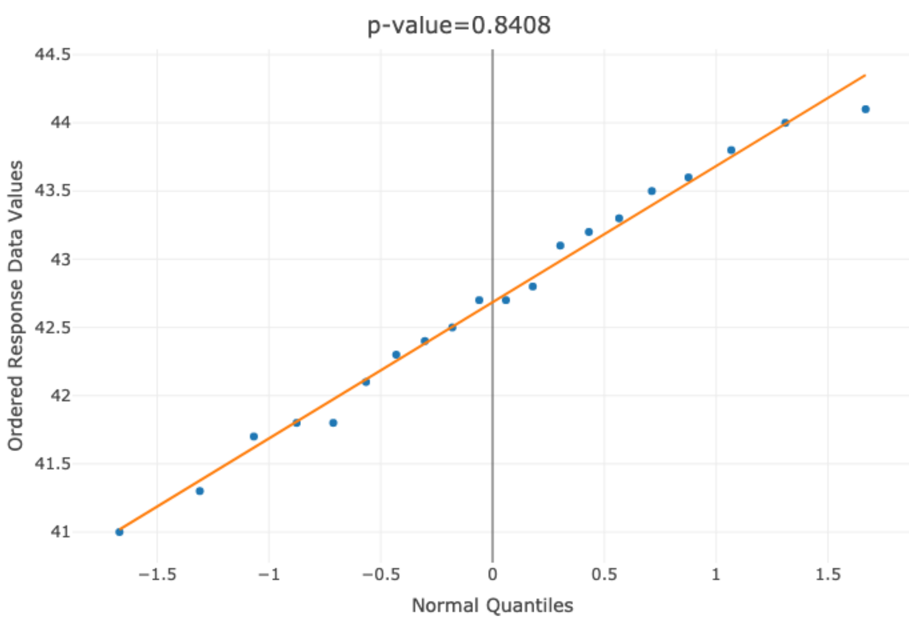
# StatHelper Instructions for Constructing a Normal Probability Plot (NPP)

1. Choose Descriptive Statistics and Start.
2. Browse and select the data; Click Run.
3. The NPP will be on the Normal Probability Plot tab. The P-value and conclusions are provided.


## Example using StatHelper to construct a Normal Probability Plot (NPP)

A sample of 20 packing times will be studied. Before we run any analysis, construct a normal probability plot and determine if the population distribution can be assumed normal. The data is stored in the file: **5_Week 5 STAT 145 Data.xlsx** under the Sheet: **Packing Times**.

ANSWER: I made the NPP using StatHelper. Since the P-value = .8408, which is greater than .05, we can assume the population follows a normal distribution.



# Minitab 19 Instructions for Constructing a Normal Probability Plot (NPP)

1. Graph→ Probability Plot→ Single (if you have one column of data)
2. Bring in the data and produce the graph by clicking OK.
3. If the P-value > 0.05, we will assume a normal distribution.

## Example using Minitab to construct a Normal Probability Plot (NPP)

A sample of 28 students were sampled and the number of hours they work per week will be studied. Before we run any analysis, construct a normal probability plot and determine if the population distribution can be assumed normal. The data is stored in the file: **5_Week 5 STAT 145 Data.xlsx** under the Sheet: **Work Hours**.

ANSWER: I made the NPP using Minitab 19. Since the P-value is smaller than 0.005 (which is less than .05), we CANNOT assume the population follows a normal distribution.