

Chapter 6: Continuous Probability Distributions

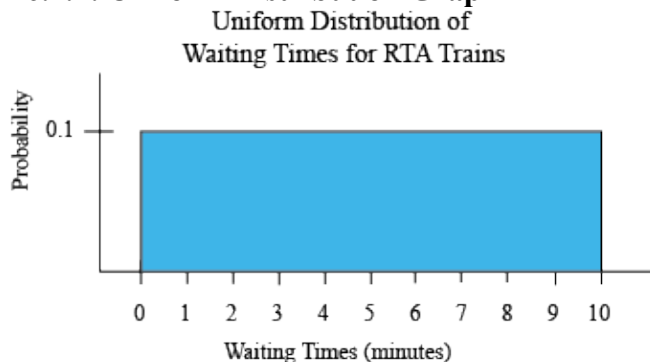
Chapter 5 dealt with probability distributions arising from discrete random variables. Mostly that chapter focused on the binomial experiment. There are many other experiments from discrete random variables that exist but are not covered in this book.

Chapter 6 deals with probability distributions that arise from continuous random variables. The focus of this chapter is a distribution known as the normal distribution, though realize that there are many other distributions that exist. A few others are examined in future chapters.

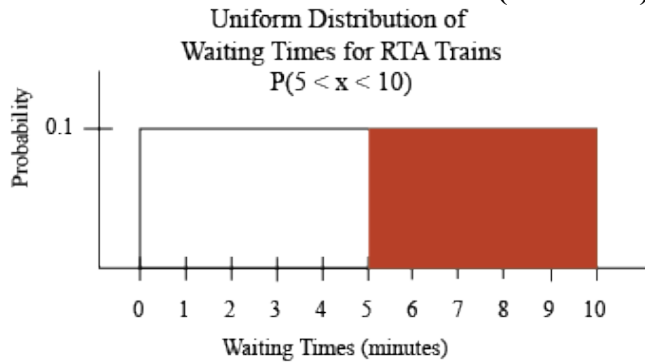
Section 6.1: Uniform Distribution

If you have a situation where the probability is always the same, then this is known as a uniform distribution. An example would be waiting for a commuter train. The commuter trains on the Blue and Green Lines for the Regional Transit Authority (RTA) in Cleveland, OH, have a waiting time during peak hours of ten minutes ("2012 annual report," 2012). If you are waiting for a train, you have anywhere from zero minutes to ten minutes to wait. Your probability of having to wait any number of minutes in that interval is the same. This is a uniform distribution. The graph of this distribution is in figure #6.1.1.

Figure #6.1.1: Uniform Distribution Graph



Suppose you want to know the probability that you will have to wait between five and ten minutes for the next train. You can look at the probability graphically such as in figure #6.1.2.

Figure #6.1.2: Uniform Distribution with $P(5 < x < 10)$ 

How would you find this probability? Calculus says that the probability is the area under the curve. Notice that the shape of the shaded area is a rectangle, and the area of a rectangle is length times width. The length is $10 - 5 = 5$ and the width is 0.1. The probability is $P(5 < x < 10) = 0.1 * 5 = 0.5$, where x is the waiting time during peak hours.

Example #6.1.1: Finding Probabilities in a Uniform Distribution

The commuter trains on the Blue and Green Lines for the Regional Transit Authority (RTA) in Cleveland, OH, have a waiting time during peak rush hour periods of ten minutes ("2012 annual report," 2012).

a.) State the random variable.

Solution:

x = waiting time during peak hours

b.) Find the probability that you have to wait between four and six minutes for a train.

Solution:

$$P(4 < x < 6) = (6 - 4) * 0.1 = 0.2$$

c.) Find the probability that you have to wait between three and seven minutes for a train.

Solution:

$$P(3 < x < 7) = (7 - 3) * 0.1 = 0.4$$

d.) Find the probability that you have to wait between zero and ten minutes for a train.

Solution:

$$P(0 < x < 10) = (10 - 0) * 0.1 = 1.0$$

- e.) Find the probability of waiting exactly five minutes.

Solution:

Since this would be just one line, and the width of the line is 0, then the
 $P(x = 5) = 0 * 0.1 = 0$

Notice that in example #6.1.1d, the probability is equal to one. This is because the probability that was computed is the area under the entire curve. Just like in discrete probability distributions, where the total probability was one, the probability of the entire curve is one. This is the reason that the height of the curve is 0.1. In general, the height of a uniform distribution that ranges between a and b , is $\frac{1}{b-a}$.

Section 6.1: Homework

- 1.) The commuter trains on the Blue and Green Lines for the Regional Transit Authority (RTA) in Cleveland, OH, have a waiting time during peak rush hour periods of ten minutes ("2012 annual report," 2012).
 - a.) State the random variable.
 - b.) Find the probability of waiting between two and five minutes.
 - c.) Find the probability of waiting between seven and ten minutes.
 - d.) Find the probability of waiting eight minutes exactly.

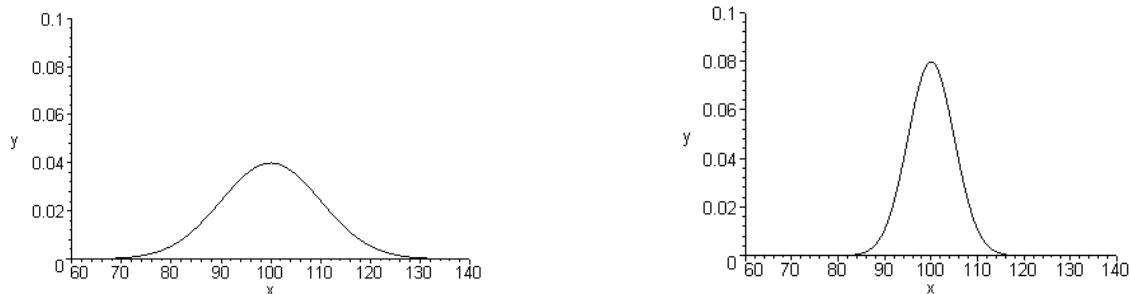
- 2.) The commuter trains on the Red Line for the Regional Transit Authority (RTA) in Cleveland, OH, have a waiting time during peak rush hour periods of eight minutes ("2012 annual report," 2012).
 - a.) State the random variable.
 - b.) Find the height of this uniform distribution.
 - c.) Find the probability of waiting between four and five minutes.
 - d.) Find the probability of waiting between three and eight minutes.
 - e.) Find the probability of waiting five minutes exactly.

Section 6.2: Graphs of the Normal Distribution

Many real life problems produce a histogram that is a symmetric, unimodal, and bell-shaped continuous probability distribution. For example: height, blood pressure, and cholesterol level. However, not every bell shaped curve is a normal curve. In a normal curve, there is a specific relationship between its “height” and its “width.”

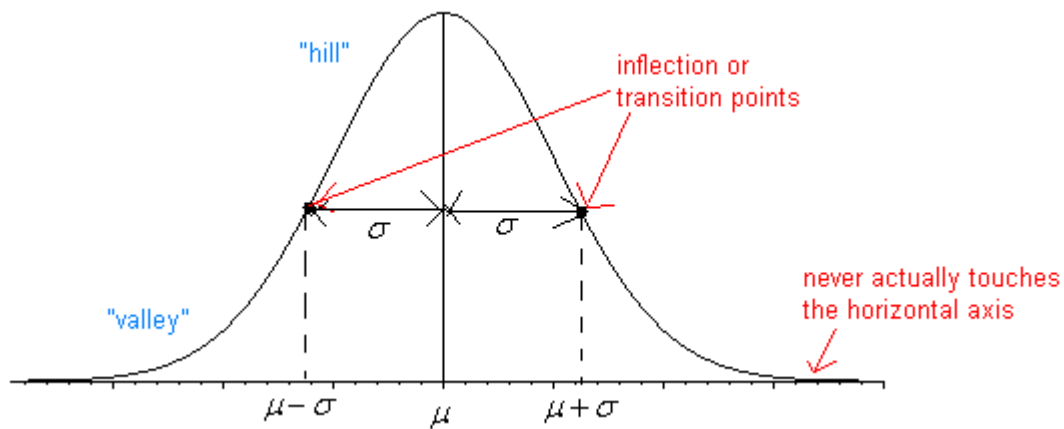
Normal curves can be tall and skinny or they can be short and fat. They are all symmetric, unimodal, and centered at μ , the population mean. Figure #6.2.1 shows two different normal curves drawn on the same scale. Both have $\mu = 100$ but the one on the left has a standard deviation of 10 and the one on the right has a standard deviation of 5. Notice that the larger standard deviation makes the graph wider (more spread out) and shorter.

Figure #6.2.1: Different Normal Distribution Graphs



Every normal curve has common features. These are detailed in figure #6.2.2.

Figure #6.2.2: Typical Graph of a Normal Curve



- The center, or the highest point, is at the population mean, μ .
- The transition points (inflection points) are the places where the curve changes from a “hill” to a “valley”. The distance from the mean to the transition point is one standard deviation, σ .
- The area under the whole curve is exactly 1. Therefore, the area under the half below or above the mean is 0.5.

The equation that creates this curve is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

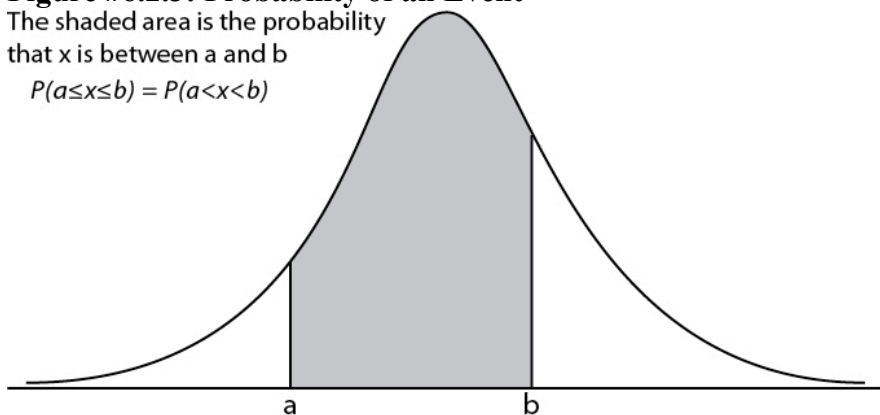
Just as in a discrete probability distribution, the object is to find the probability of an event occurring. However, unlike in a discrete probability distribution where the event can be a single value, in a continuous probability distribution the event must be a range. You are interested in finding the probability of x occurring in the range between a and b , or $P(a \leq x \leq b) = P(a < x < b)$. Calculus tells us that to find this you find the area under the curve above the interval from a to b .

$P(a \leq x \leq b) = P(a < x < b)$ is the area under the curve above the interval from a to b .

Figure #6.2.3: Probability of an Event

The shaded area is the probability that x is between a and b

$$P(a \leq x \leq b) = P(a < x < b)$$



Before looking at the process for finding the probabilities under the normal curve, it is somewhat useful to look at the **Empirical Rule** that gives approximate values for these areas. The Empirical Rule is just an approximation and it will only be used in this section to give you an idea of what the size of the probabilities is for different shadings. A more precise method for finding probabilities for the normal curve will be demonstrated in the next section. Please do not use the empirical rule except for real rough estimates.

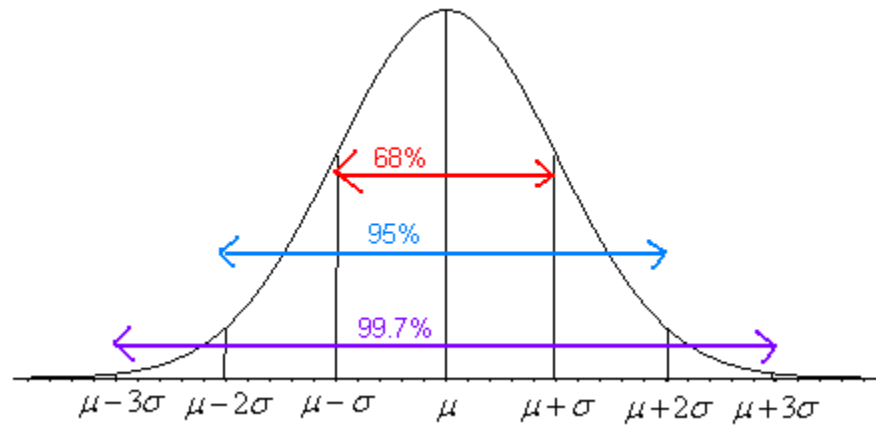
The Empirical Rule for any normal distribution:

Approximately 68% of the data is within one standard deviation of the mean.

Approximately 95% of the data is within two standard deviations of the mean.

Approximately 99.7% of the data is within three standard deviations of the mean.

Figure #6.2.4: Empirical Rule



Be careful, there is still some area left over in each end. Remember, the maximum a probability can be is 100%, so if you calculate $100\% - 99.7\% = 0.3\%$ you will see that for both ends together there is 0.3% of the curve. Because of symmetry, you can divide this equally between both ends and find that there is 0.15% in each tail beyond the $\mu \pm 3\sigma$.

Section 6.3: Finding Probabilities for the Normal Distribution

The Empirical Rule is just an approximation and only works for certain values. What if you want to find the probability for x values that are not integer multiples of the standard deviation? The probability is the area under the curve. To find areas under the curve, you need calculus. Before technology, you needed to convert every x value to a standardized number, called the z -score or z -value or simply just z . The z -score is a measure of how many standard deviations an x value is from the mean. To convert from a normally distributed x value to a z -score, you use the following formula.

z -score

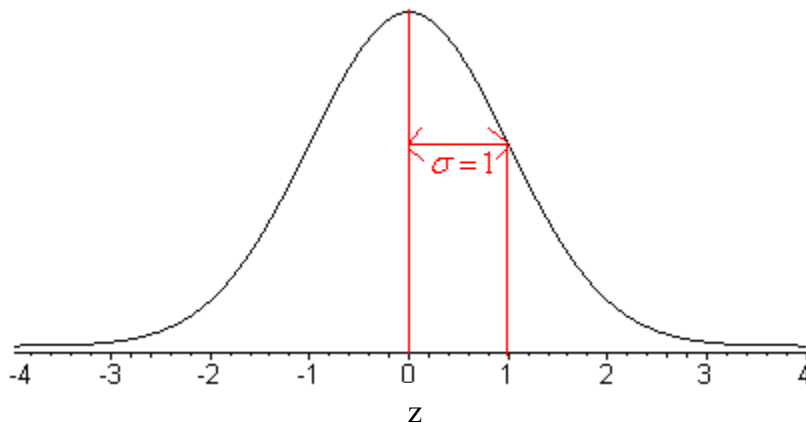
$$z = \frac{x - \mu}{\sigma}$$

where μ = mean of the population of the x value and σ = standard deviation for the population of the x value

The z -score is normally distributed, with a mean of 0 and a standard deviation of 1. It is known as the standard normal curve. Once you have the z -score, you can look up the z -score in the standard normal distribution table.

The **standard normal distribution, z** , has a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$.

Figure #6.3.1: Standard Normal Curve



Luckily, these days technology can find probabilities for you without converting to the z -score and looking the probabilities up in a table. There are many programs available that will calculate the probability for a normal curve including Excel and the TI-83/84. There are also online sites available. The following examples show how to do the calculation on the TI-83/84 and with R. The command on the TI-83/84 is in the DISTR menu and is `normalcdf`(. You then type in the lower limit, upper limit, mean, standard deviation in that order and including the commas. The command on R to find the area to the left is `pnorm`(z -value or x -value, mean, standard deviation).

Example #6.3.1: General Normal Distribution

The length of a human pregnancy is normally distributed with a mean of 272 days with a standard deviation of 9 days (Bhat & Kushtagi, 2006).

a.) State the random variable.

Solution:

x = length of a human pregnancy

b.) Find the probability of a pregnancy lasting more than 280 days.

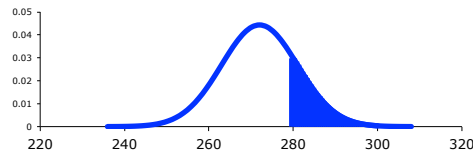
Solution:

First translate the statement into a mathematical statement.

$$P(x > 280)$$

Now, draw a picture. Remember the center of this normal curve is 272.

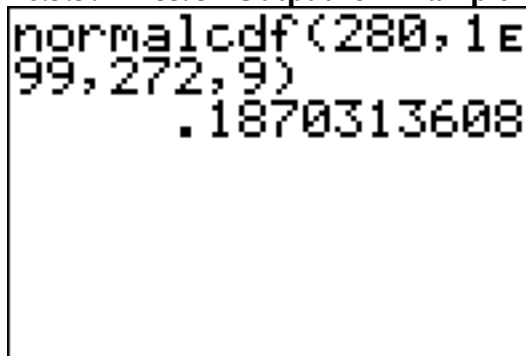
Figure #6.3.2: Normal Distribution Graph for Example #6.3.1b



To find the probability on the TI-83/84, looking at the picture you realize the lower limit is 280. The upper limit is infinity. The calculator doesn't have infinity on it, so you need to put in a really big number. Some people like to put in 1000, but if you are working with numbers that are bigger than 1000, then you would have to remember to change the upper limit. The safest number to use is 1×10^{99} , which you put in the calculator as 1E99 (where E is the EE button on the calculator). The command looks like:

`normalcdf(280,1E99,272,9)`

Figure #6.3.3: TI-83/84 Output for Example #6.3.1b



To find the probability on R, R always gives the probability to the left of the value. The total area under the curve is 1, so if you want the area to the right, then you find the area to the left and subtract from 1. The command looks like:

$$1 - \text{pnorm}(280, 272, 9)$$

$$\text{Thus, } P(x > 280) \approx 0.187$$

Thus 18.7% of all pregnancies last more than 280 days. This is not unusual since the probability is greater than 5%.

c.) Find the probability of a pregnancy lasting less than 250 days.

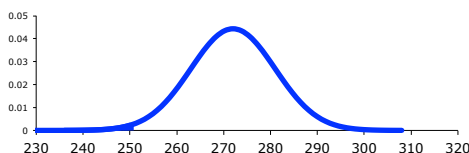
Solution:

First translate the statement into a mathematical statement.

$$P(x < 250)$$

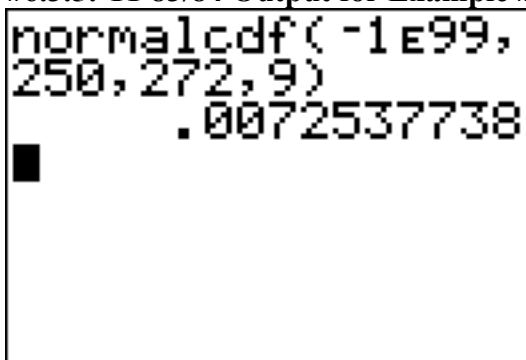
Now, draw a picture. Remember the center of this normal curve is 272.

Figure #6.3.4: Normal Distribution Graph for Example #6.3.1c



To find the probability on the TI-83/84, looking at the picture, though it is hard to see in this case, the lower limit is negative infinity. Again, the calculator doesn't have this on it, put in a really small number, such as $-1 \times 10^{99} = -1E99$ on the calculator.

Figure #6.3.5: TI-83/84 Output for Example #6.3.1c



$$P(x < 250) = \text{normalcdf}(-1E99, 250, 272, 9) = 0.0073$$

To find the probability on R, R always gives the probability to the left of the value. Looking at the figure, you can see the area you want is to the left. The command looks like:

$$P(x < 250) = \text{pnorm}(250, 272, 9) = 0.0073$$

Thus 0.73% of all pregnancies last less than 250 days. This is unusual since the probability is less than 5%.

d.) Find the probability that a pregnancy lasts between 265 and 280 days.

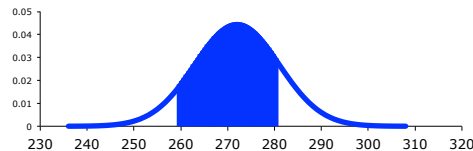
Solution:

First translate the statement into a mathematical statement.

$$P(265 < x < 280)$$

Now, draw a picture. Remember the center of this normal curve is 272.

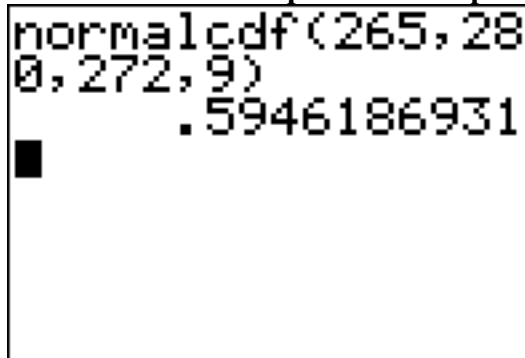
Figure #6.3.6: Normal Distribution Graph for Example #6.3.1d



In this case, the lower limit is 265 and the upper limit is 280.

Using the calculator

Figure #6.3.7: TI-83/84 Output for Example #6.3.1d



$$P(265 < x < 280) = \text{normalcdf}(265, 280, 272, 9) = 0.595$$

To use R, you have to remember that R gives you the area to the left. So

$P(x < 280) = \text{pnorm}(280, 272, 9)$ is the area to the left of 280 and

$P(x < 265) = \text{pnorm}(265, 272, 9)$ is the area to the left of 265. So the area is

between the two would be the bigger one minus the smaller one. So,

$$P(265 < x < 280) = \text{pnorm}(280, 272, 9) - \text{pnorm}(265, 272, 9) = 0.595$$

Thus 59.5% of all pregnancies last between 265 and 280 days.

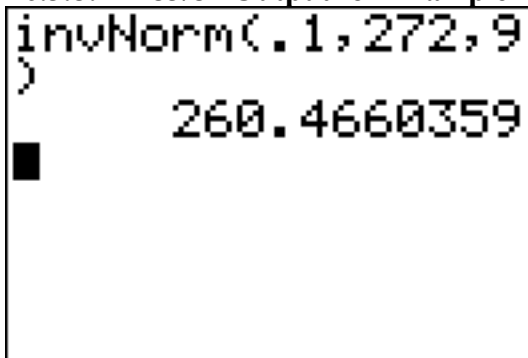
e.) Find the length of pregnancy that 10% of all pregnancies last less than.

Solution:

This problem is asking you to find an x value from a probability. You want to find the x value that has 10% of the length of pregnancies to the left of it. On the TI-83/84, the command is in the DISTR menu and is called `invNorm(`. The `invNorm(` command needs the area to the left. In this case, that is the area you are

given. For the command on the calculator, once you have `invNorm(` on the main screen you type in the probability to the left, mean, standard deviation, in that order with the commas.

Figure #6.3.8: TI-83/84 Output for Example #6.3.1e



On R, the command is `qnorm(area to the left, mean, standard deviation)`. For this example that would be `qnorm(0.1, 272, 9)`

Thus 10% of all pregnancies last less than approximately 260 days.

- f.) Suppose you meet a woman who says that she was pregnant for less than 250 days. Would this be unusual and what might you think?

Solution:

From part (c) you found the probability that a pregnancy lasts less than 250 days is 0.73%. Since this is less than 5%, it is very unusual. You would think that either the woman had a premature baby, or that she may be wrong about when she actually became pregnant.

Example #6.3.2: General Normal Distribution

The mean mathematics SAT score in 2012 was 514 with a standard deviation of 117 ("Total group profile," 2012). Assume the mathematics SAT score is normally distributed.

- a.) State the random variable.

Solution:

x = mathematics SAT score

- b.) Find the probability that a person has a mathematics SAT score over 700.

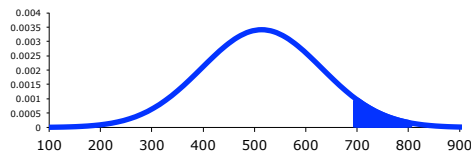
Solution:

First translate the statement into a mathematical statement.

$$P(x > 700)$$

Now, draw a picture. Remember the center of this normal curve is 514.

Figure #6.3.9: Normal Distribution Graph for Example #6.3.2b



On TI-83/84: $P(x > 700) = \text{normalcdf}(700, 1E99, 514, 117) \approx 0.056$

On R: $P(x > 700) = 1 - \text{pnorm}(700, 514, 117) \approx 0.056$

There is a 5.6% chance that a person scored above a 700 on the mathematics SAT test. This is not unusual.

- c.) Find the probability that a person has a mathematics SAT score of less than 400.

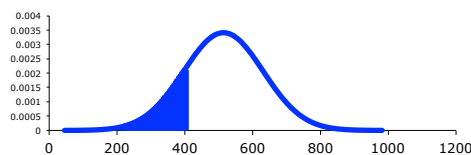
Solution:

First translate the statement into a mathematical statement.

$$P(x < 400)$$

Now, draw a picture. Remember the center of this normal curve is 514.

Figure #6.3.10: Normal Distribution Graph for Example #6.3.2c



On TI-83/84: $P(x < 400) = \text{normalcdf}(-1E99, 400, 514, 117) \approx 0.165$

On R: $P(x < 400) = \text{pnorm}(400, 514, 117) \approx 0.165$

So, there is a 16.5% chance that a person scores less than a 400 on the mathematics part of the SAT.

- d.) Find the probability that a person has a mathematics SAT score between a 500 and a 650.

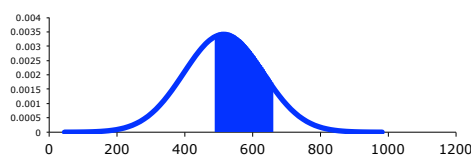
Solution:

First translate the statement into a mathematical statement.

$$P(500 < x < 650)$$

Now, draw a picture. Remember the center of this normal curve is 514.

Figure #6.3.11: Normal Distribution Graph for Example #6.3.2d



On TI-83/84: $P(500 < x < 650) = \text{normalcdf}(500, 650, 514, 117) \approx 0.425$

On R:

$$P(500 < x < 650) = \text{pnorm}(650, 514, 117) - \text{pnorm}(500, 514, 117) \approx 0.425$$

So, there is a 42.5% chance that a person has a mathematical SAT score between 500 and 650.

e.) Find the mathematics SAT score that represents the top 1% of all scores.

Solution:

This problem is asking you to find an x value from a probability. You want to find the x value that has 1% of the mathematics SAT scores to the right of it. Remember, the calculator and R always need the area to the left, you need to find the area to the left by $1 - 0.01 = 0.99$.

On TI-83/84: $\text{invNorm}(.99, 514, 117) \approx 786$

On R: $\text{qnorm}(.99, 514, 117) \approx 786$

So, 1% of all people who took the SAT scored over about 786 points on the mathematics SAT.

Section 6.3: Homework

- 1.) Find each of the probabilities, where z is a z -score from the standard normal distribution with mean of $\mu = 0$ and standard deviation $\sigma = 1$. Make sure you draw a picture for each problem.
 - a.) $P(z < 2.36)$
 - b.) $P(z > 0.67)$
 - c.) $P(0 < z < 2.11)$
 - d.) $P(-2.78 < z < 1.97)$
- 2.) Find the z -score corresponding to the given area. Remember, z is distributed as the standard normal distribution with mean of $\mu = 0$ and standard deviation $\sigma = 1$.
 - a.) The area to the left of z is 15%.
 - b.) The area to the right of z is 65%.
 - c.) The area to the left of z is 10%.
 - d.) The area to the right of z is 5%.
 - e.) The area between $-z$ and z is 95%. (Hint draw a picture and figure out the area to the left of the $-z$.)
 - f.) The area between $-z$ and z is 99%.
- 3.) If a random variable that is normally distributed has a mean of 25 and a standard deviation of 3, convert the given value to a z -score.
 - a.) $x = 23$
 - b.) $x = 33$
 - c.) $x = 19$
 - d.) $x = 45$

- 4.) According to the WHO MONICA Project the mean blood pressure for people in China is 128 mmHg with a standard deviation of 23 mmHg (Kuulasmaa, Hense & Tolonen, 1998). Assume that blood pressure is normally distributed.
- State the random variable.
 - Find the probability that a person in China has blood pressure of 135 mmHg or more.
 - Find the probability that a person in China has blood pressure of 141 mmHg or less.
 - Find the probability that a person in China has blood pressure between 120 and 125 mmHg.
 - Is it unusual for a person in China to have a blood pressure of 135 mmHg? Why or why not?
 - What blood pressure do 90% of all people in China have less than?
- 5.) The size of fish is very important to commercial fishing. A study conducted in 2012 found the length of Atlantic cod caught in nets in Karlskrona to have a mean of 49.9 cm and a standard deviation of 3.74 cm (Ovegard, Berndt & Lunneryd, 2012). Assume the length of fish is normally distributed.
- State the random variable.
 - Find the probability that an Atlantic cod has a length less than 52 cm.
 - Find the probability that an Atlantic cod has a length of more than 74 cm.
 - Find the probability that an Atlantic cod has a length between 40.5 and 57.5 cm.
 - If you found an Atlantic cod to have a length of more than 74 cm, what could you conclude?
 - What length are 15% of all Atlantic cod longer than?
- 6.) The mean cholesterol levels of women age 45-59 in Ghana, Nigeria, and Seychelles is 5.1 mmol/l and the standard deviation is 1.0 mmol/l (Lawes, Hoorn, Law & Rodgers, 2004). Assume that cholesterol levels are normally distributed.
- State the random variable.
 - Find the probability that a woman age 45-59 in Ghana, Nigeria, or Seychelles has a cholesterol level above 6.2 mmol/l (considered a high level).
 - Find the probability that a woman age 45-59 in Ghana, Nigeria, or Seychelles has a cholesterol level below 5.2 mmol/l (considered a normal level).
 - Find the probability that a woman age 45-59 in Ghana, Nigeria, or Seychelles has a cholesterol level between 5.2 and 6.2 mmol/l (considered borderline high).
 - If you found a woman age 45-59 in Ghana, Nigeria, or Seychelles having a cholesterol level above 6.2 mmol/l, what could you conclude?
 - What value do 5% of all woman ages 45-59 in Ghana, Nigeria, or Seychelles have a cholesterol level less than?

- 7.) In the United States, males between the ages of 40 and 49 eat on average 103.1 g of fat every day with a standard deviation of 4.32 g ("What we eat," 2012). Assume that the amount of fat a person eats is normally distributed.
- a.) State the random variable.
 - b.) Find the probability that a man age 40-49 in the U.S. eats more than 110 g of fat every day.
 - c.) Find the probability that a man age 40-49 in the U.S. eats less than 93 g of fat every day.
 - d.) Find the probability that a man age 40-49 in the U.S. eats less than 65 g of fat every day.
 - e.) If you found a man age 40-49 in the U.S. who says he eats less than 65 g of fat every day, would you believe him? Why or why not?
 - f.) What daily fat level do 5% of all men age 40-49 in the U.S. eat more than?
- 8.) A dishwasher has a mean life of 12 years with an estimated standard deviation of 1.25 years ("Appliance life expectancy," 2013). Assume the life of a dishwasher is normally distributed.
- a.) State the random variable.
 - b.) Find the probability that a dishwasher will last more than 15 years.
 - c.) Find the probability that a dishwasher will last less than 6 years.
 - d.) Find the probability that a dishwasher will last between 8 and 10 years.
 - e.) If you found a dishwasher that lasted less than 6 years, would you think that you have a problem with the manufacturing process? Why or why not?
 - f.) A manufacturer of dishwashers only wants to replace free of charge 5% of all dishwashers. How long should the manufacturer make the warranty period?
- 9.) The mean starting salary for nurses is \$67,694 nationally ("Staff nurse -," 2013). The standard deviation is approximately \$10,333. Assume that the starting salary is normally distributed.
- a.) State the random variable.
 - b.) Find the probability that a starting nurse will make more than \$80,000.
 - c.) Find the probability that a starting nurse will make less than \$60,000.
 - d.) Find the probability that a starting nurse will make between \$55,000 and \$72,000.
 - e.) If a nurse made less than \$50,000, would you think the nurse was under paid? Why or why not?
 - f.) What salary do 30% of all nurses make more than?

- 10.) The mean yearly rainfall in Sydney, Australia, is about 137 mm and the standard deviation is about 69 mm ("Annual maximums of," 2013). Assume rainfall is normally distributed.
- a.) State the random variable.
 - b.) Find the probability that the yearly rainfall is less than 100 mm.
 - c.) Find the probability that the yearly rainfall is more than 240 mm.
 - d.) Find the probability that the yearly rainfall is between 140 and 250 mm.
 - e.) If a year has a rainfall less than 100mm, does that mean it is an unusually dry year? Why or why not?
 - f.) What rainfall amount are 90% of all yearly rainfalls more than?

Section 6.4: Assessing Normality

The distributions you have seen up to this point have been assumed to be normally distributed, but how do you determine if it is normally distributed. One way is to take a sample and look at the sample to determine if it appears normal. If the sample looks normal, then most likely the population is also. Here are some guidelines that are use to help make that determination.

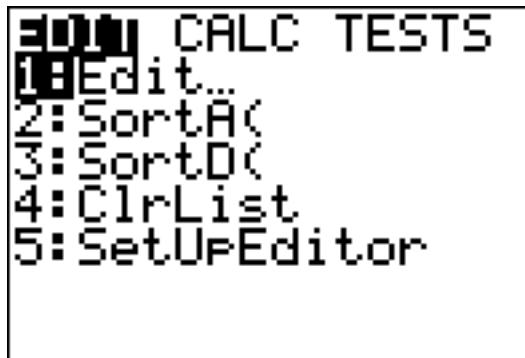
1. **Histogram:** Make a histogram. For a normal distribution, the histogram should be roughly bell-shaped. For small samples, this is not very accurate, and another method is needed. A distribution may not look normally distributed from the histogram, but it still may be normally distributed.
2. **Outliers:** For a normal distribution, there should not be more than one outlier. One way to check for outliers is to use a modified box plot. Outliers are values that are shown as dots outside of the rest of the values. If you don't have a modified box plot, outliers are those data values that are:
 - Above Q_3 , the third quartile, by an amount greater than 1.5 times the interquartile range (IQR)
 - Below Q_1 , the first quartile, by an amount greater than 1.5 times the interquartile range (IQR)

Note: if there is one outlier, that outlier could have a dramatic effect on the results especially if it is an extreme outlier. However, there are times where a distribution has more than one outlier, but it is still normally distributed. The guideline of only one outlier is just a guideline.
3. **Normal quantile plot (or normal probability plot):** This plot is provided through statistical software on a computer or graphing calculator. If the points lie close to a line, the data comes from a distribution that is approximately normal. If the points do not lie close to a line or they show a pattern that is not a line, the data are likely to come from a distribution that is not normally distributed.

To create a histogram on the TI-83/84:

1. Go into the STAT menu, and then Chose 1:Edit

Figure #6.4.1: STAT Menu on TI-83/84



2. Type your data values into L1.

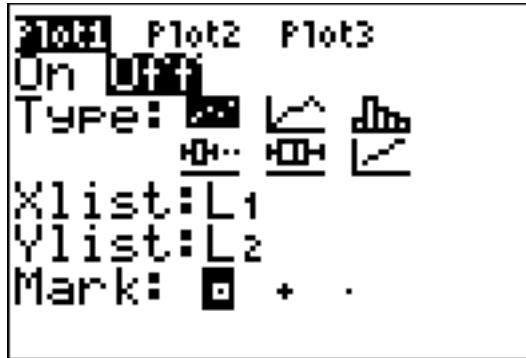
- Now click STAT PLOT (2nd Y=).

Figure #6.4.2: STAT PLOT Menu on TI-83/84



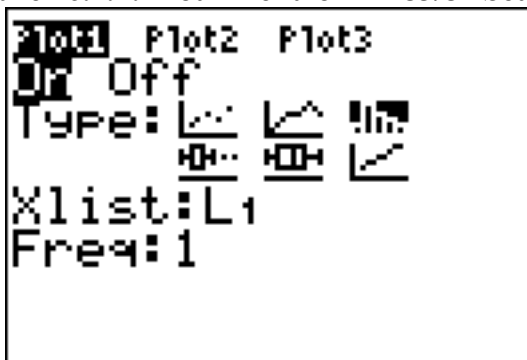
- Use 1:Plot1. Press ENTER.

Figure #6.4.3: Plot1 Menu on TI-83/84



- You will see a new window. The first thing you want to do is turn the plot on. At this point you should be on On, just press ENTER. It will make On dark.
- Now arrow down to Type: and arrow right to the graph that looks like a histogram (3rd one from the left in the top row).
- Now arrow down to Xlist. Make sure this says L1. If it doesn't, then put L1 there (2nd number 1). Freq: should be a 1.

Figure #6.4.4: Plot1 Menu on TI-83/84 Setup for Histogram

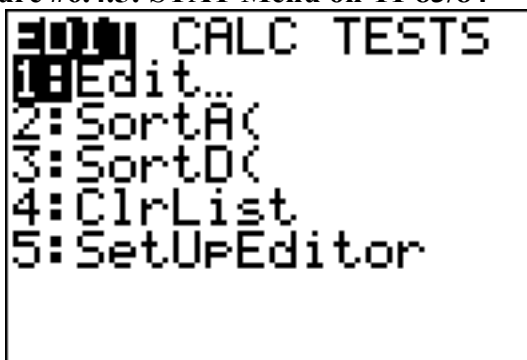


8. Now you need to set up the correct window to graph on. Click on WINDOW. You need to set up the settings for the x variable. Xmin should be your smallest data value. Xmax should just be a value sufficiently above your highest data value, but not too high. Xscl is your class width that you calculated. Ymin should be 0 and Ymax should be above what you think the highest frequency is going to be. You can always change this if you need to. Yscl is just how often you would like to see a tick mark on the y -axis.
9. Now press GRAPH. You will see a histogram.

To find the IQR and create a box plot on the TI-83/84:

1. Go into the STAT menu, and then Chose 1:Edit

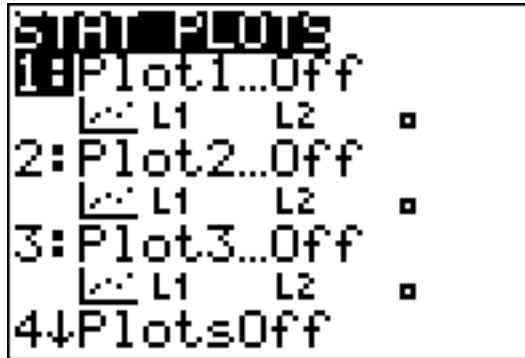
Figure #6.4.5: STAT Menu on TI-83/84



2. Type your data values into L1. If L1 has data in it, arrow up to the name L1, click CLEAR and then press ENTER. The column will now be cleared and you can type the data in.
3. Go into the STAT menu, move over to CALC and choose 1-Var Stats. Press ENTER, then type L1 (2nd 1) and then ENTER. This will give you the summary statistics. If you press the down arrow, you will see the five-number summary.

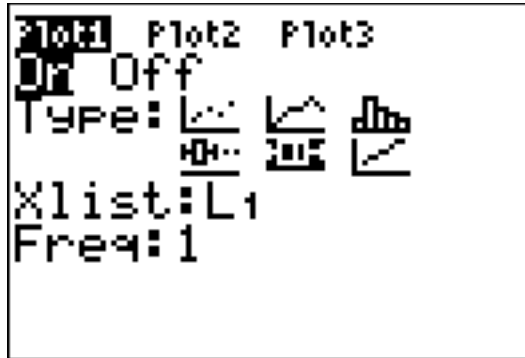
4. To draw the box plot press 2nd STAT PLOT.

Figure #6.4.6: STAT PLOT Menu on TI-83/84



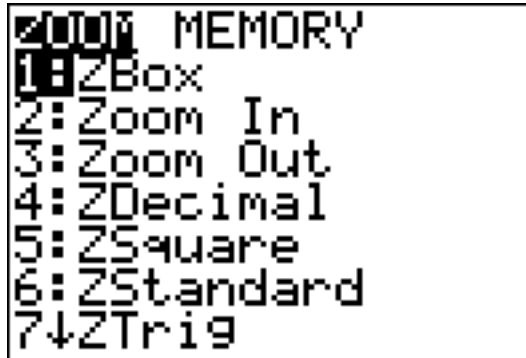
5. Use Plot1. Press ENTER

Figure #6.4.7: Plot1 Menu on TI-83/84 Setup for Box Plot



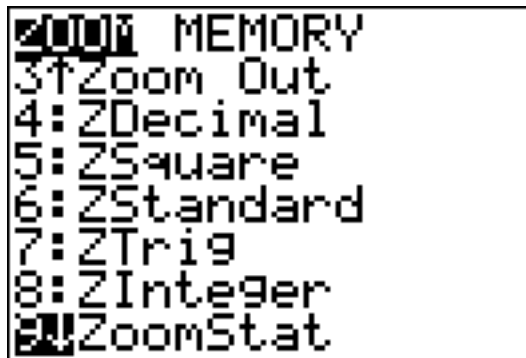
6. Put the cursor on On and press Enter to turn the plot on. Use the down arrow and the right arrow to highlight the boxplot in the middle of the second row of types then press ENTER. Set Data List to L1 (it might already say that) and leave Freq as 1.
7. Now tell the calculator the set up for the units on the x-axis so you can see the whole plot. The calculator will do it automatically if you press ZOOM, which is in the middle of the top row.

Figure #6.4.8: ZOOM Menu on TI-83/84



Then use the down arrow to get to 9:ZoomStat and press ENTER. The box plot will be drawn.

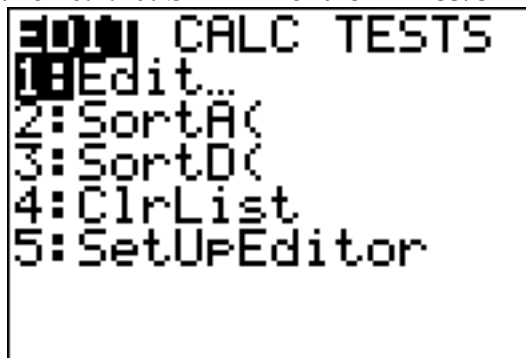
Figure #6.4.9: ZOOM Menu on TI-83/84 with ZoomStat



To create a normal quantile plot on the TI-83/84:

1. Go into the STAT menu, and then Chose 1:Edit

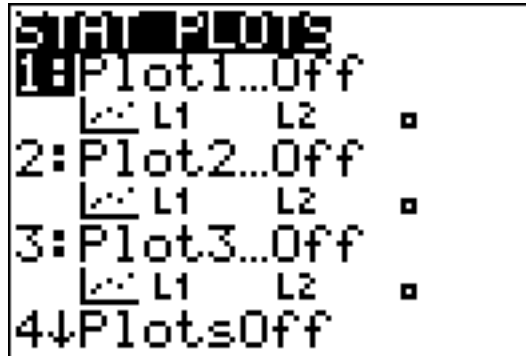
Figure #6.4.10: STAT Menu on TI-83/84



2. Type your data values into L1. If L1 has data in it, arrow up to the name L1, click CLEAR and then press ENTER. The column will now be cleared and you can type the data in.

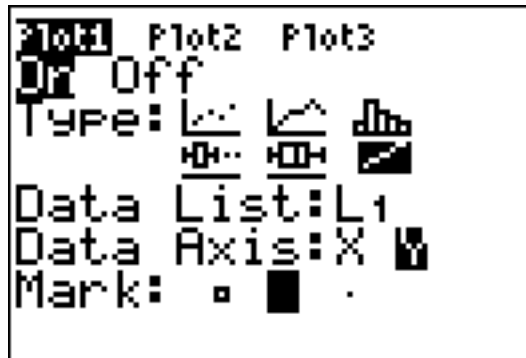
- Now click STAT PLOT (2nd Y=). You have three stat plots to choose from.

Figure #6.4.11: STAT PLOT Menu on TI-83/84



- Use 1:Plot1. Press ENTER.
- Put the cursor on the word On and press ENTER. This turns on the plot. Arrow down to Type: and use the right arrow to move over to the last graph (it looks like an increasing linear graph). Set Data List to L1 (it might already say that) and set Data Axis to Y. The Mark is up to you.

Figure #6.4.12: Plot1 Menu on TI-83/84 Setup for Normal Quantile Plot



- Now you need to set up the correct window on which to graph. Click on WINDOW. You need to set up the settings for the x variable. Xmin should be -4 . Xmax should be 4 . Xscl should be 1 . Ymin and Ymax are based on your data, the Ymin should be below your lowest data value and Ymax should be above your highest data value. Yscl is just how often you would like to see a tick mark on the y -axis.
- Now press GRAPH. You will see the normal quantile plot.

To create a histogram on R:

Put the variable in using `variable<-c(type in the data with commas between values)` using a name for the variable that makes sense for the problem. The command for histogram is `hist(variable)`. You can then copy the histogram into a

word processing program. There are options that you can put in for title, and axis labels. See section 2.2 for the commands for those.

To create a modified boxplot on R:

Put the variable in using `variable<-c`(type in the data with commas between values) using a name for the variable that makes sense for the problem. The command for box plot is `boxplot(variable)`. You can then copy the box plot into a word processing program. There are options that you can put in for title, horizontal orientation, and axis labels. See section 3.3 for the commands for those.

To create a normal quantile plot on R:

Put the variable in using `variable<-c`(type in the data with commas between values) using a name for the variable that makes sense for the problem. The command for normal quantile plot is `qqnorm(variable)`. You can then copy the normal quantile plot into a word processing program.

Realize that your random variable may be normally distributed, even if the sample fails the three tests. However, if the histogram definitely doesn't look symmetric and bell shaped, there are outliers that are very extreme, and the normal probability plot doesn't look linear, then you can be fairly confident that the data set does not come from a population that is normally distributed.

Example #6.4.1: Is It Normal?

In Kiama, NSW, Australia, there is a blowhole. The data in table #6.4.1 are times in seconds between eruptions ("Kiama blowhole eruptions," 2013). Do the data come from a population that is normally distributed?

Table #6.4.1: Time (in Seconds) Between Kiama Blowhole Eruptions

83	51	87	60	28	95	8	27
15	10	18	16	29	54	91	8
17	55	10	35	47	77	36	17
21	36	18	40	10	7	34	27
28	56	8	25	68	146	89	18
73	69	9	37	10	82	29	8
60	61	61	18	169	25	8	26
11	83	11	42	17	14	9	12

- a.) State the random variable

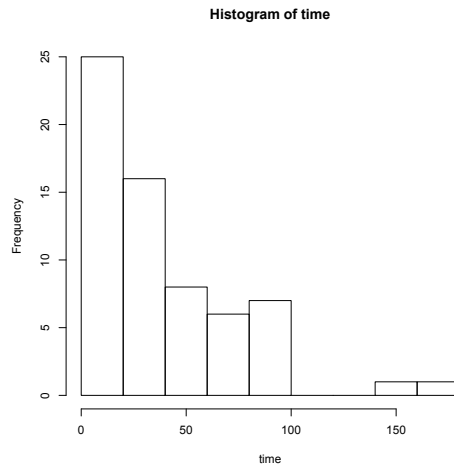
Solution:

x = time in seconds between eruptions of Kiama Blowhole

- b.) Draw a histogram.

Solution:

The histogram produced is in figure #6.4.13.

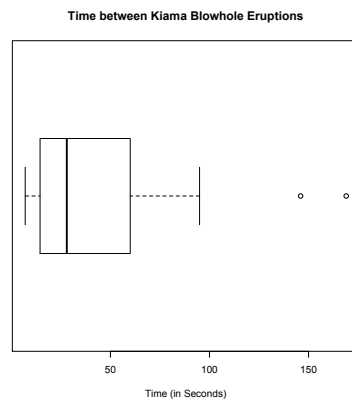
Figure #6.4.13: Histogram for Kiama Blowhole

This looks skewed right and not symmetric.

c.) Find the number of outliers.

Solution:

The box plot is in figure #6.4.14.

Figure #6.4.14: Modified Box Plot from TI-83/84 for Kiama Blowhole

There are two outliers.

Instead using:

$$IQR = Q3 - Q1 = 60 - 14.5 = 45.5 \text{ seconds}$$

$$1.5 * IQR = 1.5 * 45.5 = 68.25 \text{ seconds}$$

$$Q1 - 1.5 * IQR = 14.5 - 68.25 = -53.75 \text{ seconds}$$

$$Q3 + 1.5 * IQR = 60 + 68.25 = 128.25 \text{ seconds}$$

Outliers are any numbers greater than 128.25 seconds and less than -53.75 seconds. Since all the numbers are measurements of time, then no

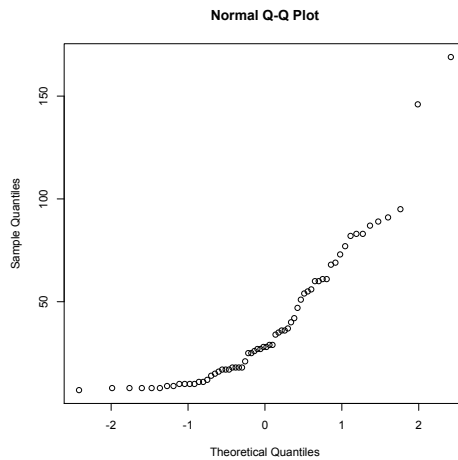
data values are less than 0 or -53.75 seconds for that matter. There are two numbers that are larger than 128.25 seconds, so there are two outliers. Two outliers are not real indications that the sample does not come from a normal distribution, but the fact that both are well above 128.25 seconds is an indication of an issue.

d.) Draw the normal quantile plot.

Solution:

The normal quantile plot is in figure #6.4.15.

Figure #6.4.15: Normal Probability Plot



This graph looks more like an exponential growth than linear.

e.) Do the data come from a population that is normally distributed?

Solution:

Considering the histogram is skewed right, there are two extreme outliers, and the normal probability plot does not look linear, then the conclusion is that this sample is not from a population that is normally distributed.

Example #6.4.2: Is It Normal?

One way to measure intelligence is with an IQ score. Table #6.4.2 contains 50 IQ scores. Determine if the sample comes from a population that is normally distributed.

Table #6.4.2: IQ Scores

78	92	96	100	67	105	109	75	127	111
93	114	82	100	125	67	94	74	81	98
102	108	81	96	103	91	90	96	86	92
84	92	90	103	115	93	85	116	87	106
85	88	106	104	102	98	116	107	102	89

a.) State the random variable

Solution:

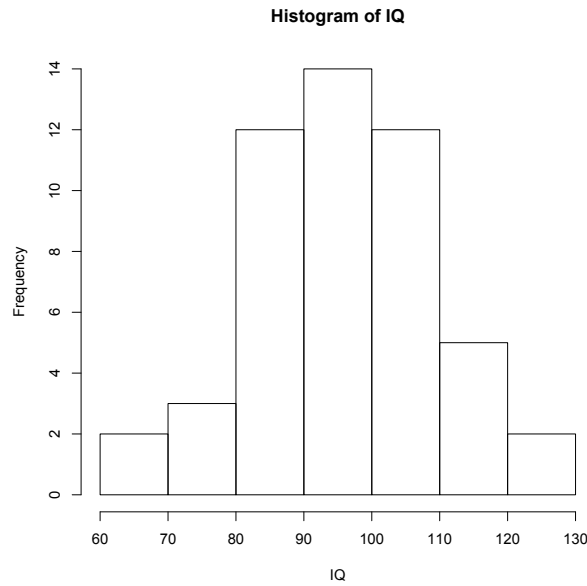
$x = \text{IQ score}$

b.) Draw a histogram.

Solution:

The histogram is in figure #6.4.16.

Figure #6.4.16: Histogram for IQ Score



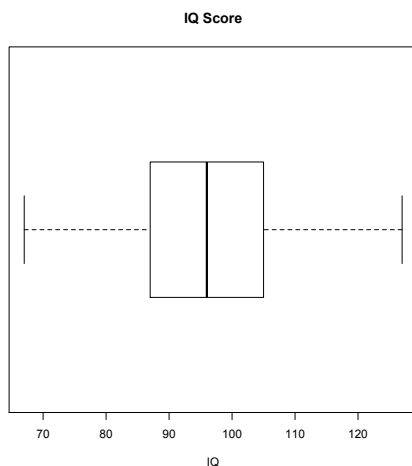
This looks somewhat symmetric, though it could be thought of as slightly skewed right.

c.) Find the number of outliers.

Solution:

The modified box plot is in figure #6.4.17.

Figure #6.4.17: Output from TI-83/84 for IQ Score



There are no outliers.

Or using

$$IQR = Q3 - Q1 = 105 - 87 = 18$$

$$1.5 * IQR = 1.5 * 18 = 27$$

$$Q1 - 1.5IQR = 87 - 27 = 60$$

$$Q3 + 1.5IQR = 105 + 27 = 132$$

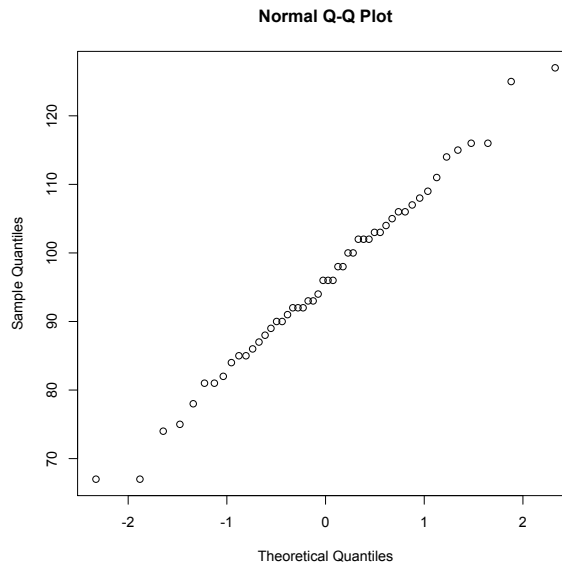
Outliers are any numbers greater than 132 and less than 60. Since the maximum number is 127 and the minimum is 67, there are no outliers.

d.) Draw the normal quantile plot.

Solution:

The normal quantile plot is in figure #6.4.18.

Figure #6.4.18: Normal Quantile Plot



This graph looks fairly linear.

e.) Do the data come from a population that is normally distributed?

Solution:

Considering the histogram is somewhat symmetric, there are no outliers, and the normal probability plot looks linear, then the conclusion is that this sample is from a population that is normally distributed.

Section 6.4: Homework

- 1.) Cholesterol data was collected on patients four days after having a heart attack. The data is in table #6.4.3. Determine if the data is from a population that is normally distributed.

Table #6.4.3: Cholesterol Data Collected Four Days After a Heart Attack

218	234	214	116	200	276	146
182	238	288	190	236	244	258
240	294	220	200	220	186	352
202	218	248	278	248	270	242

- 2.) The size of fish is very important to commercial fishing. A study conducted in 2012 collected the lengths of Atlantic cod caught in nets in Karlskrona (Ovegard, Berndt & Lunneryd, 2012). Data based on information from the study is in table #6.4.4. Determine if the data is from a population that is normally distributed.

Table #6.4.4: Atlantic Cod Lengths

48	50	50	55	53	50	49	52
61	48	45	47	53	46	50	48
42	44	50	60	54	48	50	49
53	48	52	56	46	46	47	48
48	49	52	47	51	48	45	47

- 3.) The WHO MONICA Project collected blood pressure data for people in China (Kuulasmaa, Hense & Tolonen, 1998). Data based on information from the study is in table #6.4.5. Determine if the data is from a population that is normally distributed.

Table #6.4.5: Blood Pressure Values for People in China

114	141	154	137	131	132	133	156	119
138	86	122	112	114	177	128	137	140
171	129	127	104	97	135	107	136	118
92	182	150	142	97	140	106	76	115
119	125	162	80	138	124	132	143	119

- 4.) Annual rainfalls for Sydney, Australia are given in table #6.4.6. ("Annual maximums of," 2013). Can you assume rainfall is normally distributed?

Table #6.4.6: Annual Rainfall in Sydney, Australia

146.8	383	90.9	178.1	267.5	95.5	156.5	180
90.9	139.7	200.2	171.7	187.2	184.9	70.1	58
84.1	55.6	133.1	271.8	135.9	71.9	99.4	110.6
47.5	97.8	122.7	58.4	154.4	173.7	118.8	88
84.6	171.5	254.3	185.9	137.2	138.9	96.2	85
45.2	74.7	264.9	113.8	133.4	68.1	156.4	

Section 6.5: Sampling Distribution and the Central Limit Theorem

You now have most of the skills to start statistical inference, but you need one more concept.

First, it would be helpful to state what statistical inference is in more accurate terms.

Statistical Inference: to make accurate decisions about parameters from statistics

When it says “accurate decision,” you want to be able to measure how accurate. You measure how accurate using probability. In both binomial and normal distributions, you needed to know that the random variable followed either distribution. You need to know how the statistic is distributed and then you can find probabilities. In other words, you need to know the shape of the sample mean or whatever statistic you want to make a decision about.

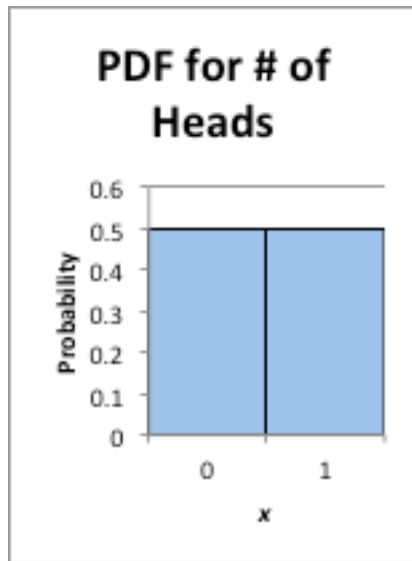
How is the statistic distributed? This is answered with a sampling distribution.

Sampling Distribution: how a sample statistic is distributed when repeated trials of size n are taken.

Example #6.5.1: Sampling Distribution

Suppose you throw a penny and count how often a head comes up. The random variable is x = number of heads. The probability distribution (pdf) of this random variable is presented in figure #6.5.1.

Figure #6.5.1: Distribution of Random Variable



Repeat this experiment 10 times, which means $n = 10$. Here is the data set: {1, 1, 1, 1, 0, 0, 0, 0, 0, 0}. The mean of this sample is 0.4. Now take another sample. Here is that data set: {1, 1, 1, 0, 1, 0, 1, 1, 0, 0}. The mean of this sample is 0.6. Another sample looks like: {0, 1, 0, 1, 1, 1, 1, 1, 0, 1}. The mean of this sample is 0.7. Repeat this 40 times. You could get these means:

Table #6.5.1: Sample Means When $n = 10$

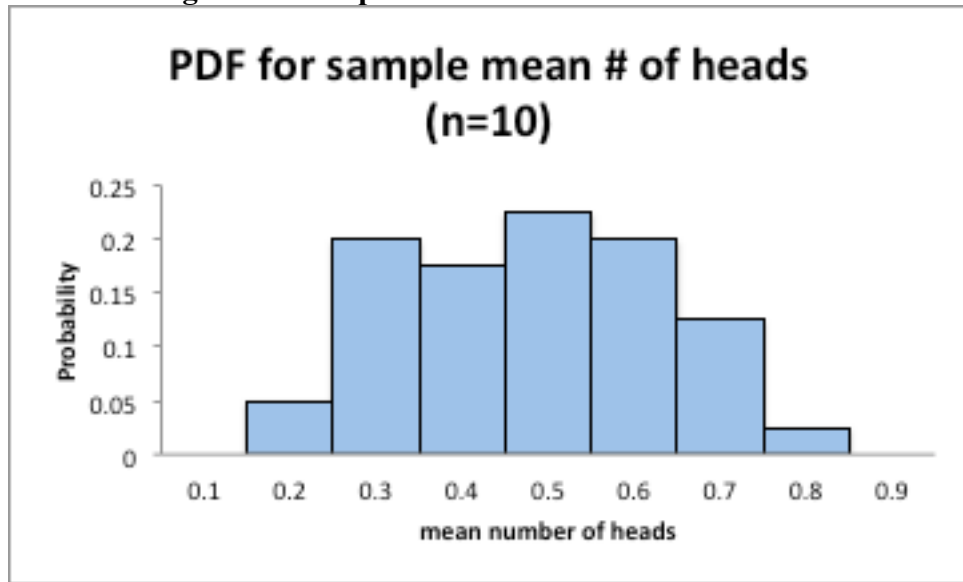
0.4	0.6	0.7	0.3	0.3	0.2	0.5	0.5	0.5	0.5
0.4	0.4	0.5	0.7	0.7	0.6	0.4	0.4	0.4	0.6
0.7	0.7	0.3	0.5	0.6	0.3	0.3	0.8	0.3	0.6
0.4	0.3	0.5	0.6	0.5	0.6	0.3	0.5	0.6	0.2

Table #6.5.2 contains the distribution of these sample means (just count how many of each number there are and then divide by 40 to obtain the relative frequency).

Table #6.5.2: Distribution of Sample Means When $n = 10$

Sample Mean	Probability
0.1	0
0.2	0.05
0.3	0.2
0.4	0.175
0.5	0.225
0.6	0.2
0.7	0.125
0.8	0.025
0.9	0

Figure #6.5.2 contains the histogram of these sample means.

Figure #6.5.2: Histogram of Sample Means When $n = 10$ 

This distribution (represented graphically by the histogram) is a sampling distribution. That is all a sampling distribution is. It is a distribution created from statistics.

Notice the histogram does not look anything like the histogram of the original random variable. It also doesn't look anything like a normal distribution, which is the only one you really know how to find probabilities. Granted you have the binomial, but the normal is better.

What does this distribution look like if instead of repeating the experiment 10 times you repeat it 20 times instead?

Table #6.5.3 contains 40 means when the experiment of flipping the coin is repeated 20 times.

Table #6.5.3: Sample Means When $n = 20$

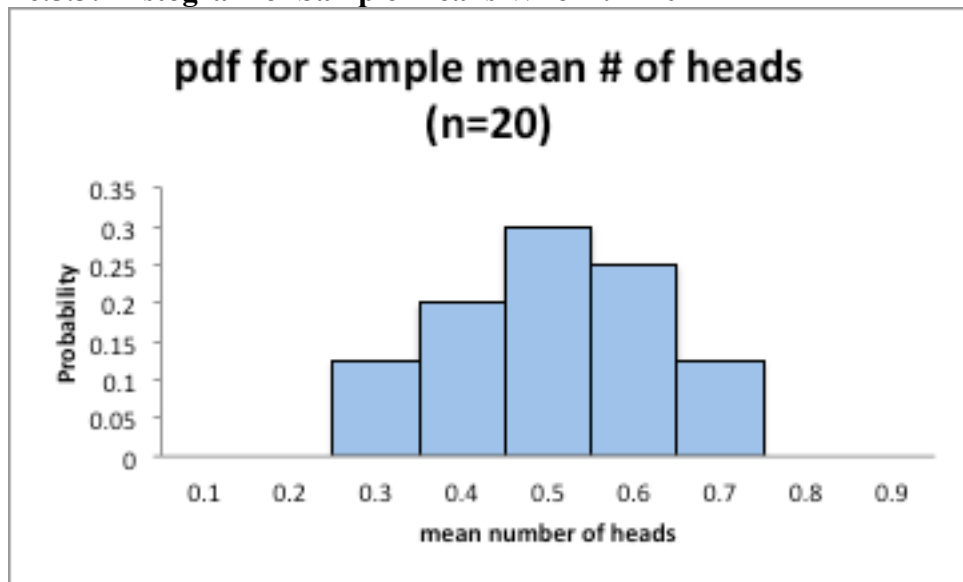
0.5	0.45	0.7	0.55	0.65	0.6	0.4	0.35	0.45	0.6
0.5	0.5	0.65	0.5	0.5	0.35	0.55	0.4	0.65	0.3
0.4	0.5	0.45	0.45	0.65	0.7	0.6	0.5	0.7	0.7
0.7	0.45	0.35	0.6	0.65	0.55	0.35	0.4	0.55	0.6

Table #6.5.3 contains the sampling distribution of the sample means.

Table #6.5.3: Distribution of Sample Means When $n = 20$

Mean	Probability
0.1	0
0.2	0
0.3	0.125
0.4	0.2
0.5	0.3
0.6	0.25
0.7	0.125
0.8	0
0.9	0

This histogram of the sampling distribution is displayed in figure #6.5.3.

Figure #6.5.3: Histogram of Sample Means When $n = 20$ 

Notice this histogram of the sample mean looks approximately symmetrical and could almost be called normal. What if you keep increasing n ? What will the sampling distribution of the sample mean look like? In other words, what does the sampling distribution of $\bar{x}$ look like as n gets even larger?

This depends on how the original distribution is distributed. In Example #6.5.1, the random variable was uniform looking. But as n increased to 20, the distribution of the mean looked approximately normal. What if the original distribution was normal? How big would n have to be? Before that question is answered, another concept is needed.

Suppose you have a random variable that has a population mean, μ , and a population standard deviation, σ . If a sample of size n is taken, then the sample mean, $\bar{x}$ has a mean $\mu_{\bar{x}} = \mu$ and standard deviation of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. The standard deviation of $\bar{x}$ is lower because by taking the mean you are averaging out the extreme values, which makes the distribution of the original random variable spread out.

You now know the center and the variability of $\bar{x}$. You also want to know the shape of the distribution of $\bar{x}$. You hope it is normal, since you know how to find probabilities using the normal curve. The following theorem tells you the requirement to have $\bar{x}$ normally distributed.

Theorem #6.5.1: Central Limit Theorem.

Suppose a random variable is from any distribution. If a sample of size n is taken, then the sample mean, $\bar{x}$, becomes normally distributed as n increases.

What this says is that no matter what x looks like, $\bar{x}$ would look normal if n is large enough. Now, what size of n is large enough? That depends on how x is distributed in the first place. If the original random variable is normally distributed, then n just needs to be 2 or more data points. If the original random variable is somewhat mound shaped and symmetrical, then n needs to be greater than or equal to 30. Sometimes the sample size can be smaller, but this is a good rule of thumb. The sample size may have to be much larger if the original random variable is really skewed one way or another.

Now that you know when the sample mean will look like a normal distribution, then you can find the probability related to the sample mean. Remember that the mean of the sample mean is just the mean of the original data ($\mu_{\bar{x}} = \mu$), but the standard deviation of the sample mean, $\sigma_{\bar{x}}$, also known as the standard error of the mean, is actually $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Make sure you use this in all calculations. If you are using the z-score, the formula when working with $\bar{x}$ is $z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$. If you are using the TI-83/84 calculator, then the

input would be `normalcdf(lower limit, upper limit,  $\mu$ ,  $\sigma/\sqrt{n}$ )`. If you are using R, then the input would be `pnorm( $\bar{x}$ ,  $\mu$ ,  $\sigma/\sqrt{n}$ )` for find the area to the left of $\bar{x}$. Remember to subtract `pnorm( $\bar{x}$ ,  $\mu$ ,  $\sigma/\sqrt{n}$ )` from 1 if you want the area to the right of $\bar{x}$.

Example #6.5.1: Finding Probabilities for Sample Means

The birth weight of boy babies of European descent who were delivered at 40 weeks is normally distributed with a mean of 3687.6 g with a standard deviation of 410.5 g (Janssen, Thiessen, Klein, Whitfield, MacNab & Cullis-Kuhl, 2007). Suppose there were nine European descent boy babies born on a given day and the mean birth weight is calculated.

- a.) State the random variable.

Solution:

x = birth weight of boy babies (Note: the random variable is something you measure, and it is not the mean birth weight. Mean birth weight is calculated.)

- b.) What is the mean of the sample mean?

Solution:

$$\mu_{\bar{x}} = \mu = 3687.6 \text{ g}$$

- c.) What is the standard deviation of the sample mean?

Solution:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{410.5}{\sqrt{9}} = \frac{410.5}{3} \approx 136.8 \text{ g}$$

- d.) What distribution is the sample mean distributed as?

Solution:

Since the original random variable is distributed normally, then the sample mean is distributed normally.

- e.) Find the probability that the mean weight of the nine boy babies born was less than 3500.4 g.

Solution:

You are looking for the $P(\bar{x} < 3500.4)$. You use the `normalcdf` command on the calculator. Remember to use the standard deviation you found in part c. However to reduce rounding error, type the division into the command. On the TI-83/84 you would have

$$P(\bar{x} < 3500.4) = \text{normalcdf}(-1E99, 3500.4, 3687.6, 410.5 \div \sqrt{9}) \approx 0.086$$

On R you would have

$$P(\bar{x} < 3500.4) = \text{pnorm}(3500.4, 3687.6, 410.5 / \text{sqrt}(9)) \approx 0.086$$

There is an 8.6% chance that the mean birth weight of the nine boy babies born would be less than 3500.4 g. Since this is more than 5%, this is not unusual.

- f.) Find the probability that the mean weight of the nine babies born was less than 3452.5 g.

Solution:

You are looking for the $P(\bar{x} < 3452.5)$.

On TI-83/84:

$$P(\bar{x} < 3452.5) = \text{normalcdf}(-1E99, 3452.5, 3687.6, 410.5 \div \sqrt{9}) \approx 0.043$$

On R:

$$P(\bar{x} < 3452.5) = \text{pnorm}(3452.5, 3687.6, 410.5 \div \sqrt{9}) \approx 0.043$$

There is a 4.3% chance that the mean birth weight of the nine boy babies born would be less than 3452.5 g. Since this is less than 5%, this would be an unusual event. If it actually happened, then you may think there is something unusual about this sample. Maybe some of the nine babies were born as multiples, which brings the mean weight down, or some or all of the babies were not of European descent (in fact the mean weight of South Asian boy babies is 3452.5 g), or some were born before 40 weeks, or the babies were born at high altitudes.

Example #6.5.2: Finding Probabilities for Sample Means

The age that American females first have intercourse is on average 17.4 years, with a standard deviation of approximately 2 years ("The Kinsey institute," 2013). This random variable is not normally distributed, though it is somewhat mound shaped.

a.) State the random variable.

Solution:

x = age that American females first have intercourse

b.) Suppose a sample of 35 American females is taken. Find the probability that the mean age that these 35 females first had intercourse is more than 21 years.

Solution:

Even though the original random variable is not normally distributed, the sample size is over 30, by the central limit theorem the sample mean will be normally distributed. The mean of the sample mean is $\mu_{\bar{x}} = \mu = 17.4$ years. The standard

deviation of the sample mean is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{35}} \approx 0.33806$. You have all the

information you need to use the normal command on your technology. Without the central limit theorem, you couldn't use the normal command, and you would not be able to answer this question.

On the TI-83/84:

$$P(\bar{x} > 21) = \text{normalcdf}(21, 1E99, 17.4, 2 \div \sqrt{35}) \approx 9.0 \times 10^{-27}$$

On R:

$$P(\bar{x} > 21) = 1 - \text{pnorm}(21, 17.4, 2 / \text{sqrt}(35)) \approx 9.0 \times 10^{-27}$$

The probability of a sample mean of 35 women being more than 21 years when they had their first intercourse is very small. This is extremely unlikely to happen. If it does, it may make you wonder about the sample. Could the population mean have increased from the 17.4 years that was stated in the article? Could the sample not have been random, and instead have been a group of women

who had similar beliefs about intercourse? These questions, and more, are ones that you would want to ask as a researcher

Section 6.5: Homework

- 1.) A random variable is not normally distributed, but it is mound shaped. It has a mean of 14 and a standard deviation of 3.
 - a.) If you take a sample of size 10, can you say what the shape of the sampling distribution for the sample mean is? Why?
 - b.) For a sample of size 10, state the mean of the sample mean and the standard deviation of the sample mean.
 - c.) If you take a sample of size 35, can you say what the shape of the distribution of the sample mean is? Why?
 - d.) For a sample of size 35, state the mean of the sample mean and the standard deviation of the sample mean.

- 2.) A random variable is normally distributed. It has a mean of 245 and a standard deviation of 21.
 - a.) If you take a sample of size 10, can you say what the shape of the distribution for the sample mean is? Why?
 - b.) For a sample of size 10, state the mean of the sample mean and the standard deviation of the sample mean.
 - c.) For a sample of size 10, find the probability that the sample mean is more than 241.
 - d.) If you take a sample of size 35, can you say what the shape of the distribution of the sample mean is? Why?
 - e.) For a sample of size 35, state the mean of the sample mean and the standard deviation of the sample mean.
 - f.) For a sample of size 35, find the probability that the sample mean is more than 241.
 - g.) Compare your answers in part d and f. Why is one smaller than the other?

- 3.) The mean starting salary for nurses is \$67,694 nationally ("Staff nurse -," 2013). The standard deviation is approximately \$10,333. The starting salary is not normally distributed but it is mound shaped. A sample of 42 starting salaries for nurses is taken.
 - a.) State the random variable.
 - b.) What is the mean of the sample mean?
 - c.) What is the standard deviation of the sample mean?
 - d.) What is the shape of the sampling distribution of the sample mean? Why?
 - e.) Find the probability that the sample mean is more than \$75,000.
 - f.) Find the probability that the sample mean is less than \$60,000.
 - g.) If you did find a sample mean of more than \$75,000 would you find that unusual? What could you conclude?

- 4.) According to the WHO MONICA Project the mean blood pressure for people in China is 128 mmHg with a standard deviation of 23 mmHg (Kuulasmaa, Hense & Tolonen, 1998). Blood pressure is normally distributed.
- a.) State the random variable.
 - b.) Suppose a sample of size 15 is taken. State the shape of the distribution of the sample mean.
 - c.) Suppose a sample of size 15 is taken. State the mean of the sample mean.
 - d.) Suppose a sample of size 15 is taken. State the standard deviation of the sample mean.
 - e.) Suppose a sample of size 15 is taken. Find the probability that the sample mean blood pressure is more than 135 mmHg.
 - f.) Would it be unusual to find a sample mean of 15 people in China of more than 135 mmHg? Why or why not?
 - g.) If you did find a sample mean for 15 people in China to be more than 135 mmHg, what might you conclude?
- 5.) The size of fish is very important to commercial fishing. A study conducted in 2012 found the length of Atlantic cod caught in nets in Karlskrona to have a mean of 49.9 cm and a standard deviation of 3.74 cm (Ovegard, Berndt & Lunneryd, 2012). The length of fish is normally distributed. A sample of 15 fish is taken.
- a.) State the random variable.
 - b.) Find the mean of the sample mean.
 - c.) Find the standard deviation of the sample mean
 - d.) What is the shape of the distribution of the sample mean? Why?
 - e.) Find the probability that the sample mean length of the Atlantic cod is less than 52 cm.
 - f.) Find the probability that the sample mean length of the Atlantic cod is more than 74 cm.
 - g.) If you found sample mean length for Atlantic cod to be more than 74 cm, what could you conclude?
- 6.) The mean cholesterol levels of women age 45-59 in Ghana, Nigeria, and Seychelles is 5.1 mmol/l and the standard deviation is 1.0 mmol/l (Lawes, Hoorn, Law & Rodgers, 2004). Assume that cholesterol levels are normally distributed.
- a.) State the random variable.
 - b.) Find the probability that a woman age 45-59 in Ghana has a cholesterol level above 6.2 mmol/l (considered a high level).
 - c.) Suppose doctors decide to test the woman's cholesterol level again and average the two values. Find the probability that this woman's mean cholesterol level for the two tests is above 6.2 mmol/l.
 - d.) Suppose doctors being very conservative decide to test the woman's cholesterol level a third time and average the three values. Find the probability that this woman's mean cholesterol level for the three tests is above 6.2 mmol/l.
 - e.) If the sample mean cholesterol level for this woman after three tests is above 6.2 mmol/l, what could you conclude?

- 7.) In the United States, males between the ages of 40 and 49 eat on average 103.1 g of fat every day with a standard deviation of 4.32 g ("What we eat," 2012). The amount of fat a person eats is not normally distributed but it is relatively mound shaped.
- a.) State the random variable.
 - b.) Find the probability that a sample mean amount of daily fat intake for 35 men age 40-59 in the U.S. is more than 100 g.
 - c.) Find the probability that a sample mean amount of daily fat intake for 35 men age 40-59 in the U.S. is less than 93 g.
 - d.) If you found a sample mean amount of daily fat intake for 35 men age 40-59 in the U.S. less than 93 g, what would you conclude?
- 8.) A dishwasher has a mean life of 12 years with an estimated standard deviation of 1.25 years ("Appliance life expectancy," 2013). The life of a dishwasher is normally distributed. Suppose you are a manufacturer and you take a sample of 10 dishwashers that you made.
- a.) State the random variable.
 - b.) Find the mean of the sample mean.
 - c.) Find the standard deviation of the sample mean.
 - d.) What is the shape of the sampling distribution of the sample mean? Why?
 - e.) Find the probability that the sample mean of the dishwashers is less than 6 years.
 - f.) If you found the sample mean life of the 10 dishwashers to be less than 6 years, would you think that you have a problem with the manufacturing process? Why or why not?

Data Sources:

Annual maximums of daily rainfall in Sydney. (2013, September 25). Retrieved from <http://www.statsci.org/data/oz/sydrain.html>

Appliance life expectancy. (2013, November 8). Retrieved from <http://www.mrappliance.com/expert/life-guide/>

Bhat, R., & Kushtagi, P. (2006). A re-look at the duration of human pregnancy. *Singapore Med J.*, 47(12), 1044-8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17139400>

College Board, SAT. (2012). *Total group profile report*. Retrieved from website: <http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf>

Greater Cleveland Regional Transit Authority, (2012). *2012 annual report*. Retrieved from website: <http://www.riderta.com/annual/2012>

Janssen, P. A., Thiessen, P., Klein, M. C., Whitfield, M. F., MacNab, Y. C., & Cullis-Kuhl, S. C. (2007). Standards for the measurement of birth weight, length and head circumference at term in neonates of european, chinese and south asian ancestry. *Open Medicine*, 1(2), e74-e88. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2802014/>

Kiama blowhole eruptions. (2013, September 25). Retrieved from <http://www.statsci.org/data/oz/kiama.html>

Kuulasmaa, K., Hense, H., & Tolonen, H. World Health Organization (WHO), WHO Monica Project. (1998). *Quality assessment of data on blood pressure in the who monica project* (ISSN 2242-1246). Retrieved from WHO MONICA Project e-publications website: <http://www.thl.fi/publications/monica/bp/bpqa.htm>

Lawes, C., Hoorn, S., Law, M., & Rodgers, A. (2004). High cholesterol. In M. Ezzati, A. Lopez, A. Rodgers & C. Murray (Eds.), *Comparative Quantification of Health Risks* (1 ed., Vol. 1, pp. 391-496). Retrieved from <http://www.who.int/publications/cra/chapters/volume1/0391-0496.pdf>

Ovegard, M., Berndt, K., & Lunneryd, S. (2012). Condition indices of atlantic cod (*gadus morhua*) biased by capturing method. *ICES Journal of Marine Science*, doi: 10.1093/icesjms/fss145

Staff nurse - RN salary. (2013, November 08). Retrieved from <http://www1.salary.com/Staff-Nurse-RN-salary.html>

The Kinsey institute - sexuality information links. (2013, November 08). Retrieved from <http://www.iub.edu/~kinsey/resources/FAQ.html>

US Department of Agriculture, Agricultural Research Service. (2012). *What we eat in America*. Retrieved from website:

<http://www.ars.usda.gov/Services/docs.htm?docid=18349>

