Statistics and Risk Management

Confidence Intervals

Video URL:
jukebox.esc13.net/untdeveloper/Videos/Confidence%20Intervals.mov

Vocabulary List:

**Frequency Distribution**: the number of times a value of a variable occurs in a data set.

**Normal Distribution**: occurs when the frequency of a random variable falls in a symmetrical bell-curved shaped distribution, with the data points concentrated around the mean.

**Skewed Distribution**: occurs when one tail of the distribution curve is longer than the other; asymmetrical distribution.

**Kurtosis**: used to describe the curve height to width of a bell type distribution curve.

**Outliers**: data collected that would be considered extreme values compared to the rest of the data collected.

**Z-score**: a devised scale with intervals coinciding with the standard deviations points that identifies a point location along a normal distribution

**Critical Value**: the level that when exceeded, finds a result that is significantly different.

**Confidence Interval**: the level on which we feel confident that we can accept.

**Sample**: a subset of a population

**Random Sample**: a sample in which each unit of the sample is randomly drawn from a given population, and each unit has an equal chance of being chosen.

**Sample size**: the number of units in a population, or \( n \).
Resources:

**Normal Distribution**
Tutorial of Histograms, Scatter Plots, Central Tendency, Standard Deviation, and Confidence Intervals. Previous lessons of Correlation and future lessons on Z Scores available as well.
http://www.gla.ac.uk/sums/users/lhornibrook/Sensor_Comparisons/normdist1.html

**Basic Statistics Home Page**
Mean, median, mode, Standard Deviation, Normal Curve, Comparing Means, the t-Test, Statistical Significance, the Null Hypothesis, Correlation and Causation are some of the relevant terms.
http://www.fgse.nova.edu/edl/secure/stats/index.htm

Through the experiment provided on this site teachers can have students collect data in the field and analyze it to discover the importance and power statistical sampling has in searching for information. The site provides a through description of each part of the activity and sample charts that the teacher can duplicate for the students to use in recording and analyzing their data.
http://slincs.coe.utk.edu/gtelab/learning_activities/33botg.html

**Measures of Shape: Skewness and Kurtosis**

This site defines the processes of both skewness and kurtosis, breaks down the concepts behind both of them, gives examples of each, and explains when they are important to determining the measures of shapes.
http://www.tc3.edu/instruct/sbrown/stat/shape.htm
Confidence Intervals Practice Test

Name:_____________________

TRUE and FALSE:

1. The larger the sample the more likely it will be a normal distribution.
   A. True            B. False
2. The z Score is based upon Standard Deviation divisions.
   A. True            B. False
3. When you collect sample data, it should not be assumed that the data is in the form of a normal distribution.
   A. True            B. False
4. Symmetric and Unimodal data is often referred to as “Skew Shaped.”
   A. True            B. False
5. If an obtained value is more extreme than the critical value, you reject.
   A. True            B. False
6. All data scores fit the normal distribution.
   A. True            B. False

MATCHING:

A. Confidence Interval
B. Normal Distribution
C. Z-Score
D. Critical Value
E. Skewed

7. ________ A uni-modal symmetrical bell style sample distribution.
8. ________ A devised scale with intervals coinciding with the standard deviations points that identifies a point location along a normal distribution.
9. ________ The point in which we reject.
10. ________ The level on which we feel confident that we can accept.
11. ________ Those who consider probability to be the relative frequency “in the long run” of outcomes.

MULTIPLE CHOICE:

12. A “___________” Normal Distribution has a Mean of 0 and a Standard Deviation of 1.0.
    A. Skewed Left            B. Skewed Right
    C. Standard                D. Horizontal
13. The _________ is used to describe the curve height to width of a bell type distribution.
    A. Z-score             B. Outliers
    C. Critical Value        D. Kurtosis
14. To express a confidence interval you need ______ pieces of information.
    A. 1            B. 2
    C. 3            D. 4
Confidence Intervals Practice Test

Name: _______________________

15. If the bulk of the data is to the left and the right tail is longer, we say the data is:
   A. Negatively Skewed  
   B. Positively Skewed  
   C. Bell Shaped  
   D. Normal Distribution

16. If the bulk of the data is to the right and the left tail is longer, we say the data is:
   A. Negatively Skewed  
   B. Positively Skewed  
   C. Bell Shaped  
   D. Normal Distribution

17. The larger the sample, the more likely it will be a _____ distribution.
   A. Skewed left  
   B. Skewed Right  
   C. Normal  
   D. Controlled

18. The _____ gives you a standardized way to measure the distance left or right of the mean value.
   A. Kurtosis  
   B. Outliers  
   C. Critical Value  
   D. Z-Score

19. __________ replace the trimmed scores with the closest normal scores.
   A. Winsorized Sample  
   B. Trimmed Sample  
   C. Kurtosis  
   D. Outliers

20. A _____ distribution is a distribution with 2 normal distributions.
   A. Unimodal  
   B. Bimodal  
   C. Trimodal  
   D. None of the Above
Confidence Intervals Practice Test KEY

1. A
2. A
3. B
4. B
5. A
6. B
7. B
8. C
9. D
10. A
11. E
12. C
13. D
14. C
15. B
16. A
17. C
18. D
19. A
20. B
Student Assignment

5.1a

Name:_____________________

Using a search engine find three examples (images) of various charts exhibiting various distributions. Explain what the chart is showing. Identify its characteristics: Normal, Standardized, Kurtosis, Uni or Bi modal, Skewed, etc.

Find a chart that you believe demonstrates outliers. Be prepared to explain the chart.
Student Assignment

5.2a Confidence Intervals Z-Score

Name:_____________________

Assume the above chart is a normal standardized distribution of 100 scores on an exam you took. The exam was worth 200 points. The mean score for this exam was 125 with a standard deviation of 20.

Your Z-Score was +2.0. What was your point score out of 200 points?

Answer: ________________

Out of the 100 students estimate how many student scored below you?

Answer: ________________

What is the percentage of the question you answered correctly?

Answer: ________________

Approximately, what is the highest exam score earned?

Answer: ____________
The above tolerance inspection chart exhibits a day’s worth of production. 1000 units were produced. The center line represents the number of inspections getting a perfect measurement. To the left and right of the center line represents the number of inspections with less than a perfect measurement. However all inspections in the blue area still passed. How many inspections failed?

Answer: _________________________

Be prepared to defend your answer.
A fast food chain fills millions a cup of soda each year. The 22 ounce cup must be filled to 21 ounce plus or minus 1 ounce with a confidence interval of 95%. The fountain is automated to first fill the glass ¾ full and then 5 second later finish topping it off to the specifications. If an employee has to spent time putting in more soda, customer service will slow and labor costs will increase. Feeling that the machine was not working correctly the store manager ran a test sample. Using a marked test cup, he filled it 10 times and recorded the number of ounces for each fill.

He calculated a Mean of 20.5 ounces and a Standard Deviation of .75 ounces. Using the Z-score of +- 2.0 as the confidence interval points, was the test acceptable? Answer: ___________

If the SD was .5 ounces would the test be Acceptable? Answer: ___________
Explore Activity:
*Requires the “Confidence Intervals NBA Height” spreadsheet in Common Documents

What is a 95% Confidence Interval? – A confidence interval is a tool for estimating some characteristic of a population when it is not practical (or possible) to observe and measure every member of the population. For the 2006-2007 NBA season, there were 439 players. Suppose you wanted to know the average height of an NBA player from that season, but you didn’t have time to look up every height, let alone calculate the mean height. So you decide to take a random sample and use the sample mean to estimate the actual average height of all NBA players.

You are going to randomly select the heights of 10 players and compute a 95% confidence interval to estimate the mean height of an NBA player from the 2006-2007 season.

If you are working at a computer from the electronic list of player heights, you can randomly select a player’s height by entering the following into an empty cell: =RANDBETWEEN(1,439). Look up the height of the selected player, and repeat for a total of 10 heights. If you have more advanced experienced with formulas in your spreadsheet software, you can write a formula to produce all 10 (without repeats, of course!).

If you are working from a paper copy of the player heights, you might have access to a graphing calculator that can generate random numbers. For a graphing calculator to Math Menu and right-arrow to the PrB sub-Menu. Then select “randInt(“ and type randInt(1,439) to select one player height. Repeat until you have 10 unique heights.
Calculate the mean height in your sample. Then (unrealistically) assume you know the population of all heights has a standard deviation of about 3.7 inches. Compute a 95% confidence interval from your sample.

You can plot your interval on a graph as a horizontal line segment from the lower bound to the upper bound of your interval. If your classmates add their intervals to the plot (in the manner shown below) you can see how your sample estimate compares to other estimates.

The actual average height is about 79.2 inches. How many of your classmates’ confidence intervals actually captured the true mean height? A 95% Confidence Interval is determined with a procedure that is likely to correctly capture the truth in about 95% of samples. So, in a class of 20 students, you would expect 19 of the 20 intervals to have included the true mean height of 79.2 inches (and one interval will have missed due to bad luck in sampling). Of course, your class might have been “lucky” and all intervals captured the true height. Or, possibly, fewer than 95% of intervals may have captured the truth due to “bad luck” in sampling. That’s the nature of a confidence interval. We don’t know whether our one interval actually captured the truth. But we are using a method with a high probability (95%) of success.
Video Links – Check out the relevant links at Khan Academy for more information and examples:

http://www.khanacademy.org/#statistics