Minitab Exercise 6 – Confidence Interval Concepts (Rev 1/15)

In this session you will

- Better understand the concepts of confidence intervals
- Better understand the impact of sample size
- Generate random data of a specified distribution

Step 1. Impact of Sample size on the confidence interval

We want to see what the impact of the number of observations is on our estimation of a population mean with a given standard deviation. We will first **generate several samples of different sizes from the same population with true mean 100 and standard deviation of 5.**

Label columns C1-C5 as follows: n10 n41 n61 n144

Choose **Calc>Random Data>Normal**
Generate **10** rows of data
Store in column: **N10**
Mean: **100**
Standard Deviation: **5**

Generate 41 rows for n41, 61 rows for n61, 144 rows for n144 using the same mean and standard deviation

We now want to know, using the different sample sizes, in what interval are we 95% sure that the actual mean of the population falls.

To do this graphically

Choose **Graph>Interval Plot>Multiple Y's-Simple OK**
Graph variable: *select columns – n10 n41 n61 n144*
OK

Hopefully, you can see the interval line decreasing as the sample size increases. This relates directly back to the central limit theorem.

Click on each of the lines to find the mean, and 95% confidence interval limits. Fill in the table below and calculate the range in the interval (Upper CI limit minus Lower CI limit).

Sample size	Mean	Lower CI Limit	Upper CI limit	Range of CI
10				
41				
61				
144				

What was the true mean of the population from which these samples were taken? _____

Did the confidence interval for prediction of the true mean based on the sample increase or decrease as the sample size increases?

_____

Step 2. How often does the true mean fall into the confidence interval?

A 95 % confidence interval for the sample means that 95% of the time the true mean will fall into the interval. That would mean that on average, one out of every 20 times the confidence interval is run, the true mean will not be in the interval.

If we take a sample of 40 observations from the population, and calculate a 95% confidence interval around each sample mean*:

What is the probability of all the intervals containing the true mean?

What is the probability of exactly 1 out of 40 not containing the true mean?

What is the probability of exactly 2 out of 40 not containing the true mean?

What is the probability of exactly 3 out of 40 not containing the true mean?

What is the probability of exactly 4 out of 40 not containing the true mean?

More than 4?

**Show manual calculations. Hopefully you recognize this as a binomial problem – two outcomes of success (define this as not containing the true mean, p=.05) or failure (does contain the true mean) with a constant sample size of n.*

You can check your calculations using the binomial probability mass function use Minitab

Calc>Probability Distributions>Binomial
Probability – *check this one to get the exact probability*
Number of trials: 40
Event probability: .05
Input constant: *enter the number of "successes" you are looking for*.

Step 3. Testing the probabilities

Now let's generate some samples and see how many have the true mean in it. We are going to simulate taking samples of size 5 from a normal population.

Use an empty column and generate 200 rows of data, normal distribution, mean of 100 and standard deviation of 5.

In the next column we want to make patterned data so that we will be able to make our column into samples of size 5.

Calc>Make Patterned Data>Simple Set of Numbers
Store data in: *enter the open column number to the right of the 200 rows*
From first value: **1**
To last value: **40**
In steps of: **1**
Number of times to list each value: **5**
Number of times to list the sequence: **1**
OK

Now we have 40 samples of size 5 identified. But we need to put them into their own columns to create the graph.

Data>Unstack Column
Unstack the data in: *select the column with the 200 data values*
Using Subscripts in: *select the column with the sample numbers*.
Select After last column in use.
OK

You should now have 40 new columns each with 5 values in it.

Now let's run the interval plots for them all

Choose **Graph>Interval Plot>Multiple Y's-Simple OK**
Graph variable: *select all 40 columns – remember, you can click on the top one and drag the mouse down to the last one – be sure you get all 40*
OK

Examine your interval plots. How many do not contain the true mean (100)? _____

What was the probability of getting the number that you did? _____

.