**West Coast Collaborative**

**Mathematics**

**Specialist Unit 4**

**Investigation 3: Sample Means**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Due: Thursday 30 August**

Complete Part 1 the preparation component. On completion of Part 1 there will be Part 2, the validation component on Tuesday 4 September.

**Extended investigation Part 1:** **Preparation activity**

Note: For .For .

The distribution formed when the following procedure is carried out is called the **sampling distribution of means**.

* Take a random sample of  independent observations from a population. If the population is finite, sampling should be with replacement to ensure that the observations are independent.
* Calculate the mean of these  sample values. This is known as the sample mean.
* Repeat the procedure until all possible samples of size  have been taken, calculating the sample mean of each one.
* Form a distribution of the sample means.

A sample value may be used to estimate an unknown population parameter by constructing an interval estimate, known as a **confidence interval**. This is an interval that has a specified probability of including the parameter. The probabilities most often used in confidence intervals are 90%, 95% and 99%. If the mean  of a particular population is unknown, then determining a 95% confidence interval for it would mean constructing the interval  such that .

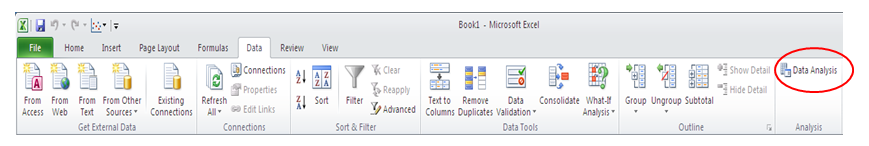
The interval constructed uses the value of the mean  of a random sample of size  taken from the population.

Before constructing a confidence interval for , the following questions need to be answered.

* Is the distribution of the population normal?
* Is the variance of the population known?
* Is the sample small or large? (Usually  is considered a large sample.)

When calculating confidence intervals it is often the case that the population standard deviation is not known. Provided that the sample size is large, the sample standard deviation  may be used as an unbiased estimate for .

Open an Excel workbook and look for **Data Analysis** on the **Data** tab.



If the **Data Analysis** command does not appear in the **Data** tab, then follow the instructions below:

1. Click the **File** tab, click **Options**, and then click the **Add-ins** category.

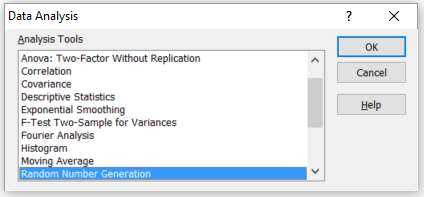
2. In the **Manage** box, select **Excel Add-ins** and then click **Go**.

3. In the **Add-ins available** box, select the **Analysis ToolPak** check box, and then click **OK**.

**Task One:**

Step 1: Open an Excel workbook and generate 100 samples each of size  from uniformly distributed numbers between 10 and 20, i.e. .

Method: Click on **Data Analysis** and then double click on **Random Number Generation**.



|  |  |
| --- | --- |
|  | Number of variables: sample size  Number of Random Numbers: number of samples  Distribution: select distribution  Parameter: enter parameters for distribution  Random Seed: leave blank  Output options  Output Range: insert address of first cell of generated values |

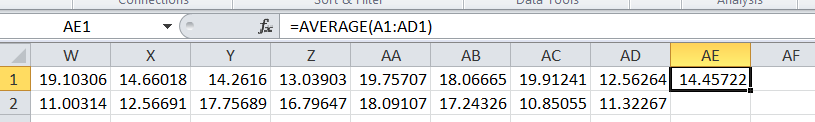
Click **OK**.

Each of the 100 rows contains 30 numbers between 10 and 20.

Step 2a: For each of the 100 samples, calculate the sample mean.

Method: Click in cell AE1 and type **=AVERAGE(A1:AD1)**

Click in cell AE1, grab the small square in the lower right hand corner of the cell and fill down to cell AE100.



Step 2b: For each of the 100 samples, calculate the sample standard deviation.

Method: Click in cell AF1 and type **=STDEV.S(A1:AD1)**

Click in cell AF1, grab the small square in the lower right hand corner of the cell and fill down to cell AF100.

Step 3a: Calculate the mean of the sample means, .

Method: Click in cell AE101 and type **=AVERAGE(AE1:AE100)**

Step 3b: Calculate the standard deviation of the sample means, .

Method: Click in cell AE102 and type **=STDEV.P(AE1:AE100)**

Step 3c: Record the sample size, the number of samples, the type of distribution (including the distribution parameters), the mean of the sample means and the standard deviation of the sample means.

Step 4: For each sample, construct the 95% confidence interval for the population mean . Since the samples were not taken from a Normal population, the variance of the population is known and the sample size is large, a 95% confidence interval is given by , where  is the sample mean,  is the population standard deviation and  is the sample size.

Method: Click in cell AG1 and type **=AE1-1.96\*(10/SQRT(12))/SQRT(30)**

Click in cell AG1, grab the small square in the lower right hand corner of the cell and fill down to cell AG100.

Click in cell AH1 and type **= AE1+1.96\*(10/SQRT(12))/SQRT(30)**

Click in cell AH1, grab the small square in the lower right hand corner of the cell and fill down to cell AH100.

Step 5a: For each sample, test whether or not the population mean lies within the 95% confidence interval.

Method: Click in cell AI1 and type **=IF(AND(15>AG1,15<AH1),1,0)**

Click in cell AI1, grab the small square in the lower right hand corner of the cell and fill down to cell AI100.

If the population mean lies within the 95% confidence interval the value 1 is returned, if not the value of 0 is returned.

Step 5b: Determine how many of the 95% confidence intervals contain the population mean.

Method: Click in cell AI101 and type **=SUM(AI1:AI100)**

Step 6: Arrange the values of the sample means that are currently in AE1 to AE100 into intervals

13 to 13.25, 13.25 to 13.5, …, 16.5 to 16.75, 16.75 to 17 and draw a histogram.

Method: Click in cell AM1 and type **13**

Click in cell AM2 and type **=AM1+0.25**

Click in cell AM2, grab the small square in the lower right hand corner of the cell and fill down to cell AM17. (Cell AM17 should show the value 17.)

Click on **Data Analysis** and then double click on **Histogram**.

|  |  |
| --- | --- |
|  | Input Range: AE1 to AE100  Bin Range: AM1 to AM17  Output options  New Worksheet Ply:  Check Chart Output |

Click **OK**.

Click on the word **Bin** and press **Delete**.

To remove the space between the bars, right click on a bar, select **Format Data Series** and change the **Gap Width** to 0%. Select **Border Color** to add a border.

Step 7a: For each sample, calculate .

Method: Click in cell AJ1 and type **=(AE1-15)/((10/SQRT(12))/SQRT(30))**

Click in cell AJ1, grab the small square in the lower right hand corner of the cell and fill down to cell AJ100.

Step 7b: Calculate the mean of these  values.

Method: Click in cell AJ101 and type **=AVERAGE(AJ1:AJ100)**

Step 7c: Calculate the standard deviation of these  values.

Method: Click in cell AJ102 and type **=STDEV.P(AJ1:AJ100)**

Step 8: Arrange the values  that are currently in AJ1 to AJ100 into intervals -3.5 to -3,

-3 to -2.5, -2.5 to -2, …, 2.5 to 3, 3 to 3.5 and draw a histogram.

Method: Click in cell AK1 and type **-3.5**

Click in cell AK2 and type **=AK1+0.5**

Click in cell AK2, grab the small square in the lower right hand corner of the cell and fill down to cell AK15. (Cell AK15 should show the value 3.5.)

Click on **Data Analysis** and then double click on **Histogram**.

|  |  |
| --- | --- |
|  | Input Range: AJ1 to AJ100  Bin Range: AK1 to AK15  Output options  New Worksheet Ply:  Check Chart Output |

Click **OK**.

Click on the word **Bin** and press **Delete**.

To remove the space between the bars, right click on a bar, select **Format Data Series** and change the **Gap Width** to 0%. Select **Border Color** to add a border.

**It is often the case that the population standard deviation  is not known. Provided that the sample size is large, the sample standard deviation  may be used as an unbiased estimate for . Steps 4, 5, 7 and 8 will be repeated using  instead of .**

Step 9: For each sample, construct the 95% confidence interval for the population mean  using , where  is the sample mean,  is the sample standard deviation and  is the sample size.

Method: Click in cell AO1 and type **=AE1-1.96\*AF1/SQRT(30)**

Click in cell AO1, grab the small square in the lower right hand corner of the cell and fill down to cell AO100.

Click in cell AP1 and type **= AE1+1.96\*AF1/SQRT(30)**

Click in cell AP1, grab the small square in the lower right hand corner of the cell and fill down to cell AP100.

Step 10a: For each sample, test whether or not the population mean lies within the 95% confidence interval.

Method: Click in cell AQ1 and type **=IF(AND(15>AO1,15<AP1),1,0)**

Click in cell AQ1, grab the small square in the lower right hand corner of the cell and fill down to cell AQ100.

If the population mean lies within the 95% confidence interval the value 1 is returned, if not the value of 0 is returned.

Step 10b: Determine how many of the 95% confidence intervals contain the population mean.

Method: Click in cell AQ101 and type **=SUM(AQ1:AQ100)**

Step 11a: For each sample, calculate .

Method: Click in cell AR1 and type **=(AE1-15)/(AF1/SQRT(30))**

Click in cell AR1, grab the small square in the lower right hand corner of the cell and fill down to cell AR100.

Step 11b: Calculate the mean of these  values.

Method: Click in cell AR101 and type **=AVERAGE(AR1:AR100)**

Step 11c: Calculate the standard deviation of these  values.

Method: Click in cell AR102 and type **=STDEV.P(AR1:AR100)**

Step 12: Arrange the values  that are currently in AR1 to AR100 into intervals

-3.5 to -3, -3 to -2.5, -2.5 to -2, …, 2.5 to 3, 3 to 3.5 and draw a histogram.

Click on **Data Analysis** and then double click on **Histogram**.

|  |  |
| --- | --- |
|  | Input Range: AR1 to AR100  Bin Range: AK1 to AK15  Output options  New Worksheet Ply:  Check Chart Output |

Click **OK**.

Click on the word **Bin** and press **Delete**.

To remove the space between the bars, right click on a bar, select **Format Data Series** and change the **Gap Width** to 0%. Select **Border Color** to add a border.

**Task One Questions**

1 In Step 2a, the mean  of each of the 100 samples of size  was calculated. Do each of your samples have the same sample mean or does this value vary between the samples?

2 (a) Calculate the population mean  for .

(b) In Step 3a, the mean of the 100 sample means was calculated. Compare your answer with the value of .

3 In Step 2b, the standard deviation  of each of the 100 samples of size  was calculated. Do each of your samples have the same sample standard deviation or does this value vary between the samples?

4 (a) Calculate the population standard deviation  for .

(b) In Step 3b, the standard deviation of the 100 sample means was calculated. Why is the standard deviation of the 100 sample means less than the population standard deviation ?

5 In Step 4, 95% confidence intervals for  were constructed using whilst in Step 5 each confidence interval was tested to determine whether or not  was within the interval. What percentage of the confidence intervals contained ?

6 In Step 6, a histogram was drawn using the values of the sample means. What do you notice about the shape of this distribution?

7 In Step 7a, the value of was calculated for each sample. In Steps 7b and 7c, the mean and standard deviation of the values of  were calculated. What do you notice about the mean and standard deviation of ?

8 In Step 8, a histogram was drawn using the values of . What do you notice about the shape of this distribution?

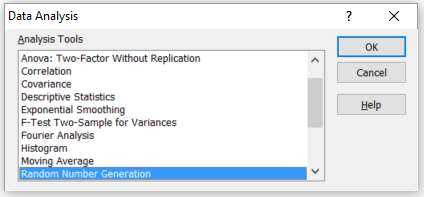
9 In Step 9, 95% confidence intervals for  were constructed using whilst in Step 10 each confidence interval was tested to determine whether or not  was within the interval. What percentage of the confidence intervals contained ?

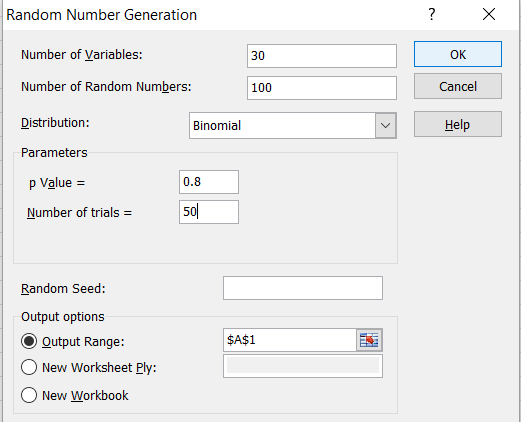
10 In Step 11a, the value of  was calculated for each sample. In Steps 11b and 11c, the mean and standard deviation of the values of  were calculated. What do you notice about the mean and standard deviation of ?

11 In Step 12, a histogram was drawn using the values of . What do you notice about the shape of this distribution

**Task Two:**

Step 1 Using and Excel workbook, generate 100 samples each of size 30 from the binomial distribution .





Step 2 For each of the 100 samples, calculate the sample mean and the sample standard deviation.

Step 3 Calculate the mean of the sample means, , and the standard deviation of the sample means, . Record the sample size, the number of samples, type of distribution, the mean of the sample means and the standard deviation of the sample means.

Step 4 For each sample, construct the 95% confidence interval for the population mean  using , where  is the sample mean,  is the population standard deviation and  is the sample size.

Step 5 For each sample, test whether or not the population mean lies within the 95% confidence interval . Determine how many of the 95% confident intervals contain the population mean.

Step 6 Arrange the sample means into appropriate intervals and draw a histogram.

Step 7 For each sample, calculate . Calculate the mean and the standard deviation of the  values.

Step 8 Arrange the values  into intervals -3.5 to -3, -3 to -2.5, -2.5 to -2, …, 2.5 to 3, 3 to 3.5 and draw a histogram.

**It is often the case that the population standard deviation  is not known. Provided that the sample size is large, the sample standard deviation  may be used as an unbiased estimate for . Steps 4, 5, 7 and 8 will be repeated using  instead of .**

Step 9: For each sample, construct the 95% confidence interval for the population mean  using , where  is the sample mean,  is the sample standard deviation and  is the sample size.

Step 10: For each sample, test whether or not the population mean lies within the 95% confidence interval . Determine how many of the 95% confident intervals contain the population mean.

Step 11: For each sample, calculate . Calculate the mean and the standard deviation of the  values.

Step 12: Arrange the values  that are currently in AR1 to AR100 into intervals

-3.5 to -3, -3 to -2.5, -2.5 to -2, …, 2.5 to 3, 3 to 3.5 and draw a histogram.

**Part Two Questions**

1 In Step 2, the mean  of each of the 100 samples of size  was calculated. Do each of your samples have the same sample mean or does this value vary between the samples?

2 (a) Calculate the population mean  for .

(b) In Step 3, the mean of the 100 sample means was calculated. Compare your answer with the value of .

3 In Step 2, the standard deviation  of each of the 100 samples of size  was calculated. Do each of your samples have the same sample standard deviation or does this value vary between the samples?

4 (a) Calculate the population standard deviation  for .

(b) In Step 3, the standard deviation of the 100 sample means was calculated. Why is the standard deviation of the 100 sample means less than the population standard deviation ?

5 In Step 4, 95% confidence intervals for  were constructed using whilst in Step 5, each confidence interval was tested to determine whether or not  was within the interval. What percentage of the confident intervals contained ?

6 In Step 6, a histogram was drawn using the values of the sample means. What do you notice about the shape of this distribution?

7 In Step 7, the value of was calculated for each sample and the mean and standard deviation of the values of  were calculated. What do you notice about the mean and standard deviation of ?

8 In Step 8, a histogram was drawn using the values of . What do you notice about the shape of this distribution?

9 In Step 9, 95% confidence intervals for  were constructed using whilst in Step 10 each confidence interval was tested to determine whether or not  was within the interval. What percentage of the confidence intervals contained ?

10 In Step 11, the value of  was calculated for each sample and the mean and standard deviation of the values of  were calculated. What do you notice about the mean and standard deviation of ?

11 In Step 12, a histogram was drawn using the values of . What do you notice about the shape of this distribution?