Session 1  “What 30 Data Can Tell About Thousands”


General question to be investigated:
How well does a random sample of 30 data represent a large data set of several thousand?

Overview:
A. Create a large set of data.
   1. Use the computer to generate the 5000 data.
   2. Construct the histogram of the data and calculate several numerical statistics.
   3. Describe the data set using ideas of shape, center, and spread.
B. Draw a random sample of 30.
   4. Construct the histogram and find the same numerical statistics for the sample.
   5. Describe the sample of data; shape, center, and spread.
C. Repeat Part B five (5) times.
   6. Describe what appears to be the “typical” sample.
   7. Compare the sample to the sampled data set.
D. Conclusions
   8. Write a paragraph answering the question, How well does a random sample of 30 data represent a data set of several thousand?
E. Repeat parts A - D using a different large data set.

Start Minitab

PART A - A LARGE DATA SET

1. Generate 5000 data from the theoretical normal distribution with mean 100, standard deviation 20.

   Calc > Random Data > Normal ...
   Enter: Generate: 5000
   Store in: C1
   Mean: 100
   Standard deviation: 20 > OK

2a. Construct a histogram of the 5000 data

   Graph > Histogram ...
   Enter: Graph variables: C1
   Select: Options ...
   Select: Type of Histogram: Frequency
   Type of Interval: Cutpoint
   Definition of Intervals:
   Midpoint/Cutpoint positions: 20:180/5 > OK > OK

   Print the histogram: Ctrl P > OK

2b. Inspect the histogram. Does the data seem to be normally distributed? (Y/N) Write a paragraph of 25+ words stating specific evidence seen in the histogram to support your answer.
2c. Find several numerical statistics describing the 5000 data and compare the normal distribution curve to the histogram.

Stat > Basic Statistics > Descriptive Statistics ...
Enter: Variable: C1
Select: Graphs ...
Select: Histogram of data with normal curve > OK > OK

Print the histogram: Ctrl P > OK
[After printing graphs, close graph windows using Window > Close All Graphs ]

3. Does the normal curve seem to follow the distribution of the 5000 data? (Y/N) Write a paragraph (50+ words) stating specific reasons why the numerical statistics and the normal curve demonstrate that the 5000 data have a normal distribution.

PART B - A SAMPLE

Let us now consider the 5000 data in C1 to be an unknown data set. Further suppose that because of some circumstances beyond our control, we can only access a sample of 30 data to be used to describe the entire set of 5000 values.

4a. Randomly select a sample of 30 data from the 5000 data in C1.

Calc > Random Data > Sample from Columns ...
Enter: Sample: 30 rows from columns: C1
Store sample in: C2
Select: Sample with replacement > OK

4b. Find numerical statistics and construct a histogram of the 30 sample data.

Stat > Basic Statistics > Descriptive Statistics ...
Enter: Variable: C2
Select: Graphs ...
Select: Histogram of data with normal curve > OK > OK

Print the histogram: Ctrl P > OK
[After printing graphs, close graph windows using Window > Close All Graphs ]

5. Describe the sample of data; shape, center and spread. Write a paragraph (50+ words) stating specific reasons why the numerical statistics, the histogram, and the normal curve demonstrate that the 30 data have a normal distribution.
PART C - ADDITIONAL SAMPLES

6a. If you could randomly select another sample of 30 data, describe the distribution that you would anticipate with regards to shape, center and spread. (25+ words)

6b. Repeat the instructions for 4a and 4b using copy and paste techniques.

Scroll up the Session window until you find the following block of commands:

MTB > Sample 30 c1 c2.
MTB > Describe C2;
SUBC> GNHist.

Highlight these commands
Copy [Ctrl C]
Go to MTB > prompt [Ctrl End]
Paste [Ctrl V]
Press: Enter

Print the histogram: Ctrl P > OK

[After printing graphs, close graph windows using Window > Close All Graphs ]

6c. Repeat (6b) three more times by using: Ctrl End; Ctrl V; Enter

6d. Using highlighting, deleting, copy and paste techniques, organize the output from 6b & c into a table for easier comparison. Leave out SE Mean.

6e. Based on the results from #4 and # 6 (a) through (d), describe the “typical” sample of 30 data; shape, center, and spread. (50+ words)

7a. Using highlighting, copy and paste techniques, organize the output from 2 (c) into a table for easier comparison with the table formed in 6(d). [Print the Session window and attach to finished assignment as an appendix.]

7b. Compare the sample described in # 6(e) to the data set in Part A. Be specific about the characteristics of shape, center and spread. (50+ words)
PART D - CONCLUSION

8. Write a paragraph answering the question, How well does a random sample of 30 data represent a data set of several thousand? (50+ words) Be convincing.

PART E - A DIFFERENT DATA SET

9. Repeat Parts A through D using a different data set. Generate 5000 data by combining 3000 from the theoretical normal distribution with mean 100, standard deviation 20 with 2000 from a normal distribution with mean 160, standard deviation 10.

   Calc > Random Data > Normal ...
   Enter: Generate: 3000
   Store in: C1
   Mean: 100
   Standard deviation: 20 > OK
   
   Calc > Random Data > Normal ...
   Enter: Generate: 2000
   Store in: C2
   Mean: 160
   Standard deviation: 10 > OK

To combine the two set, we must “Stack” them.

   Manip > Stack/Unstack > Stack ...
   Enter: Stack the following columns: C1 C2
   Store the stacked data in: C3 > OK

   Graph > Histogram ...
   Enter: Graph variables: C3
   Select: Options ...
   Select: Type of Histogram: Frequency
   Type of Interval: Cutpoint
   Definition of Intervals:
   Midpoint/Cutpoint positions: 20:200/5 > OK > OK

To select the random samples and describe the sample, use the same block of commands as used in 6b, except the column numbers must be changed.

   MTB > Sample 30 c3 c4.
   MTB > Describe C4;
   SUBC> GNHist.

PART F - A Class Project

10. Repeat Parts A through D using different data set or different sample sizes. For example; assign three different distributions each to 1/3 of class, or assign three different sample sizes, perhaps 10, 20 and 50. These projects could become individual or group written or oral reports.

Part G - HELP

11. What is the Tr Mean? Use Minitab’s Help window to find its definition. Describe how the Tr Mean is calculated.
Session 2 “Looking Beyond the Formula for Information”

General objective of this lesson:
Learning to use our own knowledge to look beyond the formulas for additional help to explain and understand the patterns displayed by the scatter diagram.

Overview:
A. Describing people using six variables
   1. One person
   2. People of all ages
   3. All college students
   4. College students by gender
B. Describing the commute to college using three variables

Start Minitab [double click the MTB icon]

PART A - VARIABLES THAT MEASURE PEOPLE
1. Let’s start by looking at several variables, and the corresponding data values for ourselves. Write down the value of each variable as it applies to you:
   Gender _____, Age _____, Height _____, Weight _____, Shoe size _____, Arm span _____

2a. Thinking about all the people of all ages you know, write down what you expect for a minimum value and a maximum value for each variable when a large set of data is studied:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2b. For the set People of all ages, do the same people tend to have both the shorter heights and smallest weights? (minimum values)

2c. For the set People of all ages, do the same people tend to have both the tallest height and largest weight? (maximum values)

2d. For the set People of all ages, does it make sense that shorter people tend to weigh less and taller people tend to weigh more? Explain.

2e. Estimate the correlation coefficient for height and weight of all people of all ages.
   Circle one: -0.9 -0.7 -0.3 0.0 0.3 0.7 0.9
2f. (Circle one) Would the slope of the line of best fit be positive, negative or near zero?

2g. On the graph paper provided to the right, plot several points to demonstrate the pattern you expect to see if the heights and weights for several hundred people of all ages were collected and the scatter diagram drawn. Label both axes, starting with minimum values and ending with maximum values for both variables.

2h. Mark the location of your own data on the scatter diagram with an X.

3a. Thinking about all the people you know that are college students, write down what you expect for a minimum value and a maximum value for each variable when a large set of data is studied:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3b. For the set college students, do the same people tend to have both the shorter heights and smallest weights? (minimum)

3c. For the set college students, do the same people tend to have both the tallest height and largest weight? (maximum)

3d. For the set college students, does it make sense that shorter people tend to weigh less and taller people tend to weigh more? Explain.

3e. Is the overall pattern, or tendency, for the relationship between height and weight of college students going to be {stronger, about the same, not as strong} as the relationship between height and weight for people of all ages? (circle one) Explain.

3f. Estimate the correlation coefficient for height and weight of college students.

Circle one: -0.9 -0.7 -0.3 0.0 0.3 0.7 0.9

3g. (Circle one) Would the slope of the line of best fit be positive, negative or near zero?
3h. On the graph paper provided to the right, plot several points to demonstrate the pattern you expect to see if the heights and weights for several hundred college students were collected and the scatter diagram drawn. Label both axes, starting with minimum values and ending with maximum values for both variables.

3i. Mark the location of your own data on the scatter diagram with an X. Has your location within the pattern of data changed from the answer to question (2h)? Explain.

3j. Have Minitab construct the scatter diagram of height (x) vs. weight (y) for the sample of students, M:ClasDta.mtw.
   1. To retrieve the set of data ClasDta.mtb
      a. Click: File > Open Worksheet ... [An Open Worksheet dialog window will appear.]
      b. Select: Drive M:
         Directory Courses / MTH / Conferen
         Type Minitab (*.mtw)
         Filename Clasdta.mtw > OK
      c. View the data.
   2. To construct the scatter diagram
      Graph > Plot ...
      Enter: Graph variables: Y: C5  X: C4
      Select: Annotations > Title ...
      Enter: College Students > OK
      Select: Annotations > Footnote ...
      Enter: Your Name > OK > OK
      To print the graph: With the graph window active, Ctrl P > Enter

3k. Locate yourself on this scatter diagram. Is your location as expected (3i)? Explain.

3l. Locate the point representing a height of 55. This point seems to be separated from the pattern displayed by the other 100 pairs of data. If it is an authentic data point, describe the person. If this point represents an error, offer at least one possible explanation.

3m. Does the scatter diagram of the 101 students look like your expectations as described in (3h)? Describe and explain differences.
3n. Inspect the scatter diagram on the right and describe the college student represented by the dots above line A with regards to height and weight.

3o. Inspect the scatter diagram on the right and describe the college students represented by the dots below line B with regards to height and weight.

4a. Thinking about all the people you know that are college students, but separated by gender. Write down what you expect for a minimum value and a maximum value when a large set of data is studied for each variable for each gender:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female coll. students</th>
<th>Male college students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4b. For each gender of college students, do the same people tend to have both the shorter heights and smallest weights? (minimum)

4c. For each gender of college students, do the same people tend to have both the tallest height and largest weight? (maximum)

4d. For each gender of college students, does it make sense that shorter people tend to weigh less and taller people tend to weigh more? Explain what is different about each gender population compared to the previous populations of people of all ages and all college students.

4e. Something about the data has changed. What about the data changes when all college students are separated into two groups by gender?
4f. Is the overall pattern, or tendency, for the relationship between height and weight for each gender of college students going to be [stronger, about the same, not as strong] as the relationship between height and weight for all college students? (circle one) Explain.

4g. Estimate the correlation coefficient for height and weight of each gender.

Males Circle one: -0.9 -0.7 -0.3 0.0 0.3 0.7 0.9

Females Circle one: -0.9 -0.7 -0.3 0.0 0.3 0.7 0.9

4h. (Circle one) Would the slope of the line of best fit be positive, negative or near zero?

4i. On the graph paper provided to the right, plot several points to demonstrate the pattern you expect to see if the heights and weights for several hundred college students of each gender were collected and the scatter diagram drawn. Label both axes, starting with minimum values and ending with maximum values.

4j. Mark the location of your own data on the scatter diagram with an X. Has your location within the pattern of data changed from the answer to question (3i)? Explain.

[Use pencil for males, pen for females.]

4k. Have Minitab construct the scatter diagram of height (x) vs. weight (y) for the sample of students, M:ClasDta.mtw.

To construct the scatter diagram showing both genders

Graph > Plot ...
Enter: Graph variables: Y: C5  X: C4
Select: Data display: For each: Group
  Group variable: Gender
Select: Annotations > Title ...
Enter: College Students > OK
Select: Annotations > Footnote ...
Enter: Your Name > OK > OK

To print the graph: With the graph window active, Ctrl P > Enter

4l. Locate yourself on this scatter diagram. Is your location as expected (4j)? Explain.

4m. Does the scatter diagram of the 101 students look like your expectations as described in (4i)? Describe and explain differences.
Let’s look at several variables that measure a person’s body.

5a. Have Minitab calculate the correlation coefficients for all possible pairs of the above variables using the sample of students, M:ClasDta.mtw.

![Stat > Basic Statistics > Correlation...](image)

Enter: Variables: C3 - C7 > OK

5b. Enter the calculated values on the table below.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Height</th>
<th>Weight</th>
<th>ShoeSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Weight</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>ShoeSize</td>
<td></td>
<td></td>
<td></td>
<td>---</td>
</tr>
<tr>
<td>ArmSpan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5c. Why is it unnecessary to list a correlation coefficient in to cells marked with “---”?  

5d. What would happen if the variables were entered into the correlation command in reverse order?  

6a. Which pair of variables had the strongest correlation coefficient for this set of data? Does that surprise you? Explain.

6b. Which one of the variables is not as strongly correlated with the others? Give a possible explanation for this.

6c. List the variables which when paired have a correlation coefficient greater than 0.50. Can you give a possible reason as to why these variables all correlate with each other?

**PART B - COMMUTING TO COLLEGE or The “Highway Battle of the Sexes”**

7a. Being a college with no dormitories, every student is a commuter, therefore everybody lives with this problem. Let’s look at it statistically. How far do you commute to college each day?  

One-way commute distance _____, One-way commute time _____, Gender ______

The sample data was collected from 101 students, and we will look at the data after a few questions, but first there is always this “gender rivalry.”
7b. The males think they are better drivers (meaning faster) than their female counterpart. If this were in fact the case, draw an ellipse around the points on the graph to the right what would represent these “better, faster” male drivers. Explain why you drew the ellipse where you did.

7c. The females recant that “there is no difference in the speed or anything else about the commuting distance and time between the two genders.” If this were in fact the case, draw an ellipse around the points on the graph to the right where the data for these "no different commuters” would be. Explain why you drew the ellipse where you did.

7d. The graph at the right shows a “60 miles per hour” line; that’s “one mile per minute” or \( y = x \). Describe what it means for a point on this graph to be below the line.

On the line.

Above the line. (Be careful.)

7e. Describe at least two different scenarios that could possibly explain the existence of the points on this graph that are below the “60 mph” line. State which scenario you tend to think has a better chance of being true and why.

7f. Have Minitab construct the scatter diagram of the 101 students data.

Graph > Plot...
Enter: Graph variables: Y: C9   X: C8
Select: Data display: For each: Graph > OK [Be sure to Select Graph]
7g. Notice on the graph for (7f) that there is one point at the top separate from the pattern, and there are 3 more points on the left side that are located at Time = 30 and seem to be above the pattern, also. Describe at least one possible explanation for these data being separated like this.

7h. Notice that the dots on part of the graph seem quite dense compared to the rest. For what values of x and y does the pattern seem more dense? Describe at least one possible explanation for this two-part pattern. [Could gender of driver be involved?]

7i. It is time to see how gender relates to these data; have Minitab draw the scatterplot showing both male and female on the same graph.

   Graph > Plot ...
   Enter: Graph variables: Y: C9   X: C8
   Select: Data display: For each: Group
           Group variable: Gender
   Select: Annotations > Title ...
           Enter: College Students > OK
   Select: Annotations > Footnote ...
           Enter: Your Name > OK > OK

7j. Inspect the graph. Describe at least two characteristics of the pattern that are visible now that the two genders are plotted using different symbols.

7k. Which gender commutes no more than 20 miles? __________
Which gender has a commute distance that span the complete domain of the data? __________
Which gender commutes for no more than 30 minutes? __________
Which gender has a commute time that span the complete range of the data? __________

7l. Why do males not commute the longer distance?
Why is it that some females do commute the longer distance?

7m. Describe an explanation for the 4 points identified in (7g) that is compatible with answer to (7l).
Session 3  “How Important Is the Assumption?”


General question to be investigated:
How important is the assumption “the sampled population is normally distributed” to the use of the Student’s t-distribution for statistical procedures about one mean?

Overview:
A. Generate a random sample of 1000 values from the probability distribution, N(100, 50).
   1. Construct the histogram of the 1000 data.
   2. Visually inspect the distribution to verify “normally distributed data.”
B. Draw a random sample of 10 from the theoretical distribution, N(100, 50).
   3. Calculate the mean and standard deviation for the sample.
   4. Calculate the t-statistic $t^* = (\bar{x} - 100)/\left(s/\sqrt{10}\right)$
C. Draw 100 random samples of 10 each from the theoretical distribution, N(100, 50).
   5. Calculate the mean and standard deviation for each of the 100 samples.
   6. Calculate the t-statistic $t^* = (\bar{x} - 100)/\left(s/\sqrt{10}\right)$ for each of the 100 samples.
D. Compare the empirical results to t-distribution with 9 degrees of freedom.
E. Repeat parts A - D using non-normal probability distributions.
F. Conclusions

Start Minitab

PART A - THE DATA BEING GENERATED

Let’s start by checking the data produced by the random number generator, “Does the data appear to be normally distributed?”

1a. Generate 1000 data from the normal distribution with mean 100, standard deviation 50 and place them in column C1.

   Calc > Random Data > Normal ...
   Enter: Generate: 1000
   Store in: ‘data’
   Mean: 100
   Standard deviation: 50 > OK

1b. Construct a histogram of the 1000 data

   Graph > Histogram ...
   Enter: Graph variables: C1 data
   Select: Options ...
   Select: Type of Histogram: Frequency
   Type of Interval: Cutpoint
   Definition of Intervals:
   Midpoint/Cutpoint positions: -100:300/25 > OK

   Select: Annotation > Title ...
   Enter: 1000 Data > OK
   Select: Annotation > Footnote ...
   Enter: “Your Name” > OK > OK

   Print the histogram: Ctrl P > OK

2. Inspect the histogram. Does the data seem to be normally distributed? (Y/N) Write a paragraph of 25+ words stating specific evidence seen in the histogram to support your answer.
PART B - ONE SAMPLE

3a. Randomly select one sample of 10 data from the normal distribution with mean 100, standard deviation 50 and place it in row 1 of columns C2 through C11.

Calc > Random Data > Normal ...
Enter: Generate: 1
Store in: C2 - C11
Mean: 100
Standard deviation: 50  > OK

Inspect your sample data.

3b. Calculate the mean and standard deviation for the sample in row 1 of C2 - C11.

Calc > Row Statistics ...
Select: Statistic: Mean
Enter: Input variables: C2 - C11  [Double click C2; type -; double click C11]
Enter: Store results in: ‘x-bar’  > OK  [Use of single quotes names the next available column.]

Calc > Row Statistics ...
Select: Statistic: Standard deviation
Enter: Input variables: C2 - C11
Enter: Store results in: ‘stdev’  > OK

View the sample mean and standard deviation; they're in row 1 of C12 and C13.

4a. Find the calculated \( t^* = \left( \bar{x} - 100 \right) / \left( s / \sqrt{10} \right) \), for sample in row 1, C2 - C11 and put in C14.

Calc > Calculator ...
Enter: Store results in: ‘t*’
Enter: Expression: \( \left( \bar{x} - 100 \right) / \left( s / (10^{**.5}) \right) \)  > OK
[To enter the expression, click in sequence: the expression window; (); dbl. click C12; -; 1; 0; 0; click right of ); / (divide); (); dbl. click C13; / (divide); (); 1; 0; **; . ; 5 It should now look like expression to above.]

4b. View the sample mean, standard deviation and \( t^* \); they’re in row 1 of C12, C13 and C14. Write a paragraph of 25+ words describing what has occurred in 3a, 3b and 4a.
PART C - ONE HUNDRED SAMPLES

5a. Randomly select 100 samples of 10 data each from the normal distribution with mean 100, standard deviation 50 and place them in rows 1 through 100 of columns C2 through C11.

```
Calc  >  Random Data  >  Normal ...
Enter: Generate:  100  [Bold indicates the values to be entered.]
Store in:  C2 - C11  [Not in bold indicates the value was previously entered, but check to be sure it is correct.]
Mean:  100
Standard deviation:  50  >  OK

Inspect the 100 samples.
```

Note: The information on a dialog window remains unchanged during a session, until you change it. Thus repeat uses are quick and easy.

5b. Find the mean and standard deviation for each of the 100 samples, rows 1-100 of C2 - C11.

```
Calc  >  Row Statistics ...
Select:  Statistics:  Mean
Enter:  Input variables:  C2 - C11  [Not in bold indicates the value was previously entered, but check to be sure it is correct.]
Enter:  Store results in:  ‘x-bar’  >  OK

Calc  >  Row Statistics ...
Select:  Statistics:  Standard deviation
Enter:  Input variables:  C2 - C11
Enter:  Store results in:  ‘stdev’  >  OK
```

View the means and standard deviations for the 100 samples; they’re in C12 and C13.

6a. Find the calculated t*,

\[ t^* = \frac{\bar{x} - 100}{s/\sqrt{10}} \]

for each of the 100 samples and put in C14.

```
Calc  >  Calculator ...
Enter:  Store results in:  ‘t*’
Enter:  Expression:  (‘x-bar’ - 100) / (stdev / (10**.5))  >  OK
```

6b. Inspect the 100 sample means, standard deviations and t* values; they’re in C12, C13 and C14.

Write a paragraph of 25+ words describing what has occurred in 5a, b and 6a.
PART D - COMPARISON OF RESULTS TO THE STUDENT’S t-DISTRIBUTION, df = 9

7a. Use the histogram command to express the cumulative relative frequency distribution of the 100 observed \( t^* \) values.

```
Graph > Histogram ...
Enter: Graph variables: \( t^* \)
Select: Options ...
Select: Type of Histogram: Cumulative Frequency
Type of Interval: Cutpoint
Definition of Intervals: Midpoint/Cutpoint positions: -5:5/1 > OK
Select: Annotations
Select: Data Labels ...
Select: Show data labels > OK
Select: Annotation > Title ...
Enter: 100 Calculated t Values > OK > OK
Print the graph: Ctrl P > OK
```

7b. Inspect the graph. Complete the listing of the corresponding cumulative relative frequency distribution (below) and write 25+ words describing the meaning of this cum. rel. freq. dist.

<table>
<thead>
<tr>
<th>( t^* )</th>
<th>Cum. Rel. Freq.</th>
<th>Cum. Prob. (for 8b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8a. Construct the cumulative probability distribution for Student’s \( t \) with df = 9.

1) List the possible \( t \)-values
```
Calc > Make Patterned Data > Simple Set of Numbers ...
Enter: Store patterned data in: ‘t’
From first value: -5
To last value: 5
In steps of: 1
List each value: 1 times
List the whole sequence: 1 times > OK
```

2) Calculate the cumulative probabilities for Student’s \( t \) with df = 9.
```
Calc > Probability Distributions > T ...
Select: Cumulative probability
Enter: Degrees of freedom: 9
Input column: C15
Optional storage: ‘CumProb’ > OK
```
8b. Copy the cumulative probability distribution from (8b) onto table in (7b). Write 25+ words describing the meanings of the information in C15 and C16.

9. Compare the cumulative relative frequency distribution (7b) and the cumulative probability distribution (8b). Write a short paragraph about the similarity between the two distributions.

10a. Repeat the instructions for 5a, 5b, 6a and 7a to observe another set of 100 samples.

Scroll up the Session window until you find the following block of commands:

```
MTB > Random 100 c2-c11;
SUBC>   Normal 100.0 50.0.  [Normal with mean = 100, stdev = 50]
MTB > RMean C2-C11 'x-bar'.
MTB > RStDev C2-C11 'stdev'.
MTB > Let 't*' = ('x-bar' - 100) / (stdev / (10**.5))
MTB > Histogram 't*';
SUBC>   Cumulative;
SUBC>   Frequency;
SUBC>   CutPoint -5:5/1;
SUBC>   Bar;
SUBC>   Title "100 Calculated t Values";
SUBC>   Footnote "your name";
SUBC>   Symbol;
SUBC>   Type 0;
SUBC>   Label;
SUBC>   Offset 0.0 0.0125;
SUBC>   Placement 0 1.
```

[If you have unwanted lines in between, highlight and delete these unwanted lines.]

Click, drag, and release to highlight this “block” of commands
Use **Ctrl C** to Copy
Use **Ctrl End** to go to MTB > prompt
Use **Ctrl V** to Paste
Press **Enter**

Compare the results shown on the graph to the distribution in C15 and C16.

10b. Repeat three more times using only: **Ctrl End; Ctrl V; Enter** each time.

10c. Write a short paragraph answering, “If the sampled population is normally distributed, does the sampling distribution for the calculated t seem to have a Student’s t-distribution?”.
PART E - USING NON-NORMAL DISTRIBUTIONS

11. Repeat the instructions for 1a, 1b, 5a, 5b, 6a and 7a to observe a set of 1000 data and 100 samples from a uniform probability distribution with low of 0 and upper of 200 (rectangular).

a. Scroll up the Session window until you find the following block of commands. Then replace the underlined information with the corresponding information in the box.

```
MTB > Random 1000 c1;
SUBC> Normal 100 50.
MTB > Histogram C1;
SUBC> CutPoint -100:300/25; 0:200/25
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
```

Copy and Paste at the active (Red) MTB > prompt. Press Enter.

View 1000 values. Do they seem to have a uniform distribution from 0 to 200? Describe the distribution.

b. Scroll up the Session window until you find the following block of commands. Then replace the underlined information with the corresponding information in the box.

```
MTB > Random 100 c2-c11;
SUBC> Normal 100 50.
MTB > RMean C2-C11 c12.
MTB > RStDev C2-C11 c13.
MTB > Let c14 = ('x-bar'-100) / (stdev / ((10)**.5))
MTB > Histogram c14;
SUBC> Cumulative;
SUBC> Frequency;
SUBC> CutPoint -5:5/1;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
```

Copy and Paste at the active (Red) MTB > prompt. Press Enter.

```
-5
-4
-3
-2
-1
 0
 1
 2
 3
 4
 5
```

```
c. Inspect the graph. Complete the listing of the corresponding cumulative relative frequency distribution (below). List the cumulative probability distribution for Student’s t with df = 9. Compare the two distributions. (25+ words)
```

```
-5
-4
-3
-2
-1
 0
 1
 2
 3
 4
 5
```
12. Repeat the instructions for 1a, 1b, 5a, 5b, 6a and 7a to observe a set of 1000 data and 100 samples from a skewed (exponential with mean 100) probability distributions.

   a. Scroll up the Session window until you find the following block of commands. Then replace the underlined information with the corresponding information in the box.

```
MTB > Random 1000 c1;
SUBC> Normal 100 50.
MTB > Histogram C1;
SUBC> CutPoint -100:300/25;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
```

Copy and Paste at the active (Red) MTB > prompt. Press Enter.

View 1000 values. Do they seem to have a skewed distribution from 0 to 700? Describe the distribution.

   b. Scroll up the Session window until you find the following block of commands. Then replace the underlined information with the corresponding information in the box.

```
MTB > Random 100 c2-c11;
SUBC> Normal 100 50.
MTB > RMean C2-C11 c12.
MTB > RStDev C2-C11 c13.
MTB > Let c14 = ('x-bar'-100) / (stdev / ((10)**.5))
MTB > Histogram c14;
SUBC> Cumulative;
SUBC> Frequency;
SUBC> CutPoint -5:5/1;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
```

Copy and Paste at the active (Red) MTB > prompt. Press Enter.

c. Inspect the graph. Complete the listing of the corresponding cumulative relative frequency distribution (below). List the cumulative probability distribution for Student’s t with df = 9.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compare the two distributions. (25+ words)
PART F - CONCLUSIONS

13. Write a paragraph answering the question, How important is the assumption “the sampled population is normally distributed?” to the use of the Student’s t-distribution for statistical procedures about one mean? (50+ words) Be convincing.
Lesson 3 - Attachment  “How Important Is the Assumption?”


The MINITAB routines

NORMAL (100, 50)

1000 values

MTB > Random 1000 c1;
SUBC>   Normal 100 50.
MTB > Histogram C1;
SUBC>   CutPoint -100:300/25;
SUBC>   Bar;
SUBC>   Symbol;
SUBC>   Type 0;
SUBC>   Label;
SUBC>   Offset 0.0 0.0125;
SUBC>   Placement 0 1.

1 sample

MTB > Random 1 c2-c11;
SUBC>   Normal 100 50.
MTB > Name c12 = 'x-bar'
MTB > RMean C2-C11 'x-bar'.
MTB > Name c13 = 'stdev'
MTB > RStDev C2-C11 'stdev'.
MTB > Name C14 = 't*
MTB > Let 't*' = ('x-bar' - 100) / (stdev / (10**.5))

100 samples

MTB > Random 100 c2-c11;
SUBC>   Normal 100 50.
MTB > RMean C2-C11 c12.
MTB > RStDev C2-C11 c13.
MTB > Let c14 = (x-bar'-100) / (stdev / (10)**.5))
MTB > Histogram c14;
    Cumulative;
SUBC>   Frequency;
SUBC>   CutPoint -5:5/1;
SUBC>   Bar;
SUBC>   Symbol;
SUBC>   Type 0;
SUBC>   Label;
SUBC>   Offset 0.0 0.0125;
SUBC>   Placement 0 1.
MTB > Set c15
DATA>   1 (-5 : 5 / 1)1
DATA>   End.
MTB > CDF c15 c16;
SUBC>   T 9.
UNIFORM (0, 200)

MTB > Random 1000 cl;
SUBC> Uniform 0 200.
MTB > Histogram C1;
SUBC> CutPoint 0:200/25;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
MTB > Random 100 c2-c11;
SUBC> Uniform 0 200.
MTB > RMean C2-C11 c12.
MTB > RStDev C2-C11 c13.
MTB > Let c14 = ('x-bar'-100) / (stdev / ((10)**.5))
MTB > Histogram c14;
SUBC> Cumulative;
SUBC> Frequency;
SUBC> CutPoint -5:5/1;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
MTB > Set c15
DATA> 1( -5 : 5 / 1 )1
DATA> End.
MTB > CDF c15 c16;
SUBC> T 9.
========================================

SKEWED - EXPONENTIAL(100)

MTB > Random 1000 c1;
SUBC> Exponential 100.
MTB > Histogram C1;
SUBC> CutPoint 0:700/50;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
MTB > Random 100 c2-c11;
SUBC> Expo 100.
MTB > RMean C2-C11 c12.
MTB > RStDev C2-C11 c13.
MTB > Let c14 = ('x-bar'-100) / (stdev / ((10)**.5))
MTB > Histogram c14;
SUBC> Cumulative;
SUBC> Frequency;
SUBC> CutPoint -5:5/1;
SUBC> Bar;
SUBC> Symbol;
SUBC> Type 0;
SUBC> Label;
SUBC> Offset 0.0 0.0125;
SUBC> Placement 0 1.
MTB > Set c15
DATA>  1( -5 : 5 / 1 )1
DATA>  End.
MTB >  CDF c15 c16;
SUBC>  T 9.