Return to Cover Page

LESSON 13 - HYPOTHESIS TESTING; SMALL SAMPLE

We will use Problem 29 on page 405 as an example. First enter the data and name the variable (let's call it Servings). Maximize the session Window, clear it below the date/time stamp then type your name, Lesson 13, Example, and define the variable (X = number of 12-ounce servings of soda).

Since the number of data values is less than 30, it is necessary to check the data is reasonably normal and that the t-test would be valid. One way to check for normality is to create a stem-and-leaf display of the data. Click on Graph > Stem-and- Leaf and select the variable into the "Graph variable:" box. Type 1 in the box marked "Increment:" then click "OK". After deleting excess blank lines and other unnecessary output from the session Window, your display should look like the figure above. Notice 2|8 = 2.8 in the stem-and-leaf plot.

—— 7/10/2007 12:51:52 PM ————

Jeonghun Kim
Lesson 13
Example

X = 12-ounce servings of soda

Stem-and-Leaf Display: Servings

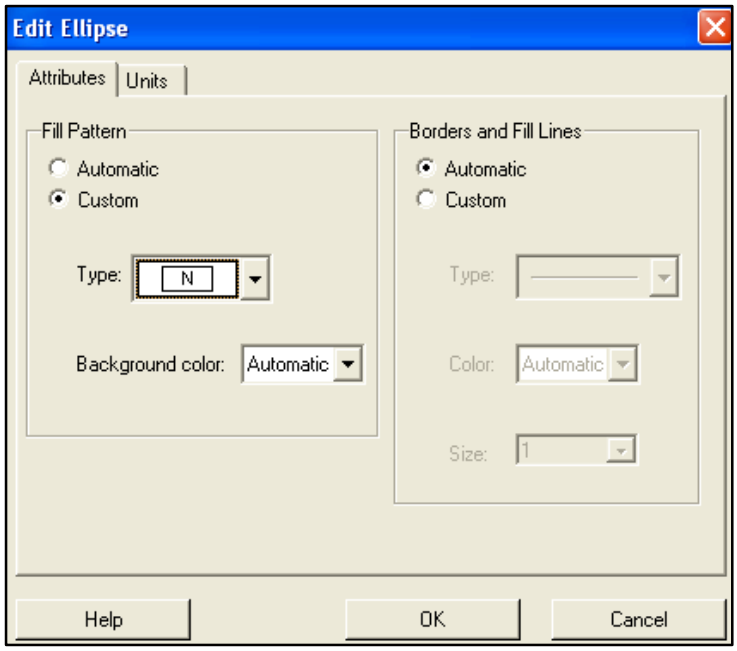
```
Stem-and-leaf of Servings  N  = 20
Leaf Unit = 0.10

 2    0  58
 7    1  22468
(12)  2  001111223458
 1    3  2
```

ANDERSON-DARLING NORMALITY TEST

A better way to test for normality is the "Anderson-Darling Normality Test". This is a test that has "The data is from a normal distribution" as its null hypothesis. We use the same level of significance on this test as we do on the t-test that we would like to run on the data in question. If we fail to reject the hypothesis of normality, we go ahead with the t-test; if the P-Value of the Anderson-Darling test is smaller than our choice of alpha, we reject normality and the t-test would be invalid. To get to the Anderson-Darling test click Stat > Basic Statistics > Graphical Summary. Select C1 Servings into the "Variables:" box and click "OK". There are several things we would like to fix. We will want everything black and white as in our previous graphs. To edit the bars in the histogram, you will need to right click in them, then click "Edit Bars." Do the same for the other areas that need to be fixed. Be sure to make the blue lines black. We will now need two new tools from the graphics tool bars, the ADD MENU and the ELLIPSE TOOL. These are shown in the figure below.

To put your name on the graph select either "Graph Region" or "Figure Region" from the "Item to edit menu" then select "Footnote" from the "ADD MENU", type your name in the box provided, then click OK. Your name will appear in the lower left corner of the graph. Notice that the P-Value for the Anderson-Darling test is .184 in this case. That is certainly larger than .05, (or any other alpha we would ever pick), so we will not reject normality and it is OK to use the t-test. To indicate that, choose "Subtitle" from the "ADD MENU" and type "t-test OK" in the box that is provided. The subtitle will appear right below the title, but use the mouse to drag it to the right and leave it right above the box with the Anderson-Darling test.

NOTE: If the P-Value had been less than our chosen value of alpha, we would have typed "t-test not OK" as the subtitle. We would then have had to abort the procedure since the Student's t-test may only be used if the sample data follows an approximately normal distribution.



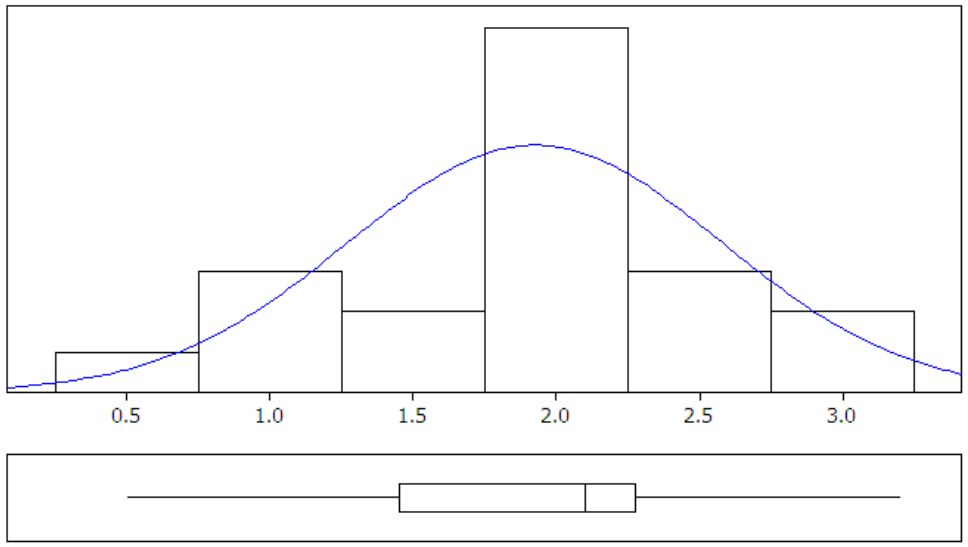
Now click on the "ELLIPSE" tool. The cursor will become a cross hair. Place it a little above and to the left of the P-Value for the Anderson-Darling test and drag it down and to the right to create an ellipse with the P-Value inside. When you release the mouse button, the ellipse will turn gray and cover the value. To fix this, click on the "Edit" tool and change the "Type:" under "Fill Pattern" to none. The dialog box should look like the figure to the right. Click "OK" and the gray will disappear from the ellipse and the P-Value will show through.

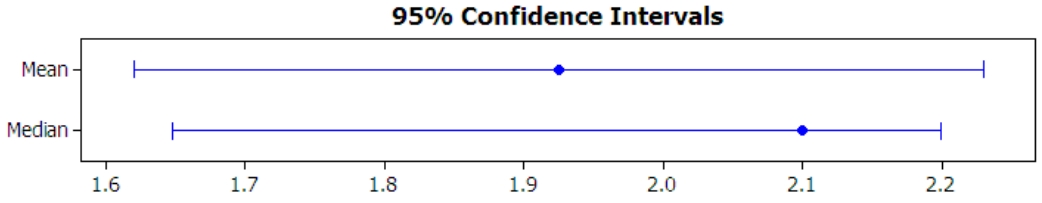
Your completed graphical summary should now look like the figure on the next page. This is the procedure we will use any time we need to verify normality of data. We will find that we will need it again in several future lessons.

If this were an assignment, we would now print the graph to turn in with the rest of the problem, but for this exercise you may now close it.

Summary for Servings

t-test OK





Anderson-Darling Normality Test	
A-Squared	0.50
P-Value	0.184
Mean	1.9250
StDev	0.6536
Variance	0.4272
Skewness	-0.440628
Kurtosis	0.335028
N	20
Minimum	0.5000
1st Quartile	1.4500
Median	2.1000
3rd Quartile	2.2750
Maximum	3.2000
95% Confidence Interval for Mean	
1.6191	2.2309
95% Confidence Interval for Median	
1.6470	2.2000
95% Confidence Interval for StDev	
0.4971	0.9547

Jeonghun Kim

ONE SAMPLE t - TEST

We will now use the t-test and the P-Value method to test the hypothesis that the mean is less than 3. Click on Stat > Basic Statistics > 1-Sample t and you see a dialog box that is almost the same as it is for the 1-Sample Z that we saw in Lesson 12, except that it is not necessary to enter a value for sigma since the t-test automatically uses the sample standard deviation. Select C1 Servings into the "Samples in columns:" box. As with the z-test, you must check the "Perform hypothesis test" box and type the value of μ from the null hypothesis into the "Hypothesized mean:" box (3 in this case) then click on "Options" and set the "Alternative:" box to "less than" since our alternative hypothesis is < in this case. Now click "OK" on the option dialog box and "OK" on the 1-Sample t dialog box. Using the P-Value approach we would normally stop here, but since we have been told to use $\alpha = .05$ we will type the decision and the conclusion based on a 5% level of significance. The results are shown in the figure below.

One-Sample T: Servings

```
Test of mu = 3 vs < 3


                                          95% Upper
Variable   N   Mean  StDev  SE Mean      Bound      T      P
Servings  20  1.925  0.654    0.146      2.178  -7.36  0.000
```

Decision: Since the P-Value = 0.000 < .05, we reject H0.
Conclusion: There is enough evidence at the 5% level of significance to support the claim that teenage males drink fewer than three 12-ounce servings of soda per day.

OUTLIER EXAMPLE

As a second example, let us consider the following 16 data values:

12 83 92 92 97 98 103 104
107 109 109 111 113 118 122 128

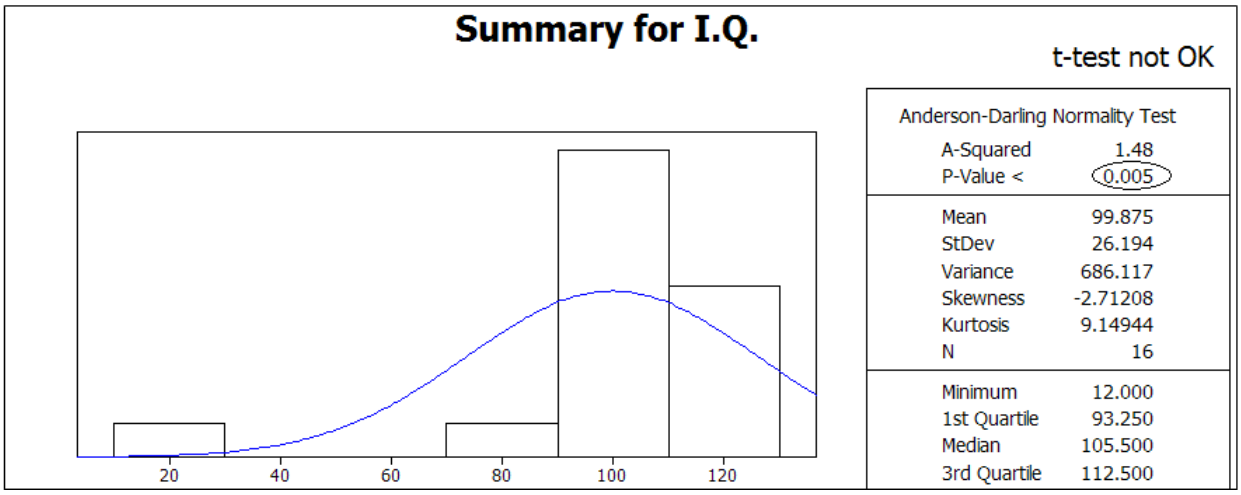
Since the example doesn't say what this data is, let us assume it is I.Q. scores of a random sample of students from Podunk High and we want to test the principal's claim that the mean is 100 using the P-Value method with a 5% level of significance. Start a new instance of Minitab, type your name and so on as usual, and enter the data. Now create a Stem-and-Leaf display. In this case we should type 10 into the "Increment" box since we want the stems to be 10's. It should look like the figure on the right.

Stem-and-Leaf Display: I.Q.

```
Stem-and-leaf of I.Q.  N  =
16
Leaf Unit = 1.0

 1   1   2
 1   2
 1   3
 1   4
 1   5
 1   6
 1   7
 2   8   3
 6   9   2278
(5)  10  34799
 5   11  138
 2   12  28
```

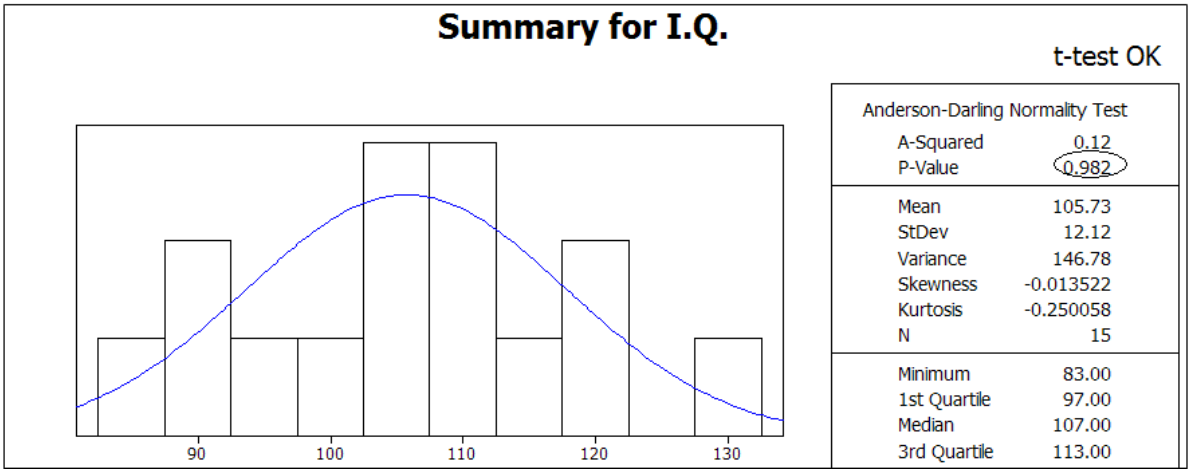
NOTE: Since the Stem-and-Leaf Display gives all of the information that is in a data display, it is not necessary to do both.

Except for the outlier at 12, the data looks pretty good. Let us see what the Anderson-Darling test tells us. Create a Graphical Summary for this data as we did above. The upper portion of the summary is shown below.



Notice that the outlier at 12 is also quite obvious here. Also notice that the Aderson-Darling test would reject the assumption of normality because of the small P-value, so the t-test would not be valid.

Since an I.Q. of 12 is very unlikely, we will assume that data point is a mistake and delete it, (to delete the outlier return to the Data Window, click in the cell with the 12, then click Edit > Delete cells) then try the Anderson-Darling test again.



We see that it is now OK to use the t-test so we shall do so. In this case we must make the "Test mean:" 100 then click on "Options" to set the "Alternative:" menu to "not equal". After typing in the decision and conclusion, the results are shown at the top of the next page.

One-Sample T: I.Q.

```
Test of mu = 100 vs not = 100

Variable   N     Mean   StDev  SE Mean         95% CI          T      P
I.Q.      15  105.733  12.115    3.128  (99.024, 112.443)  1.83  0.088
```

Decision: Since the P-Value = 0.088 > .05, fail to reject H0.
Conclusion: The mean I.Q. score of students at Podnuk High is not significantly different from 100 at the .05 level of significance. There is insufficient evidence to refute the principal's claim.

SUMMARY DATA

As was in the case with the z-test, we may have only summary data to work with. Consider Problem 25 on page 404. Start a new instance of Minitab, clear the text below the date/time stamp and enter your name and so on as usual. Define the variable as X = waste recycled (in pounds) by an adult per day. The problem tells us that we may assume the data to be normal, so we need not do anything with the Anderson-Darling test. Now click on Stat > Basic Statistics > 1-Sample t. Click on the "Summarized data" button then fill in the boxes with the information given in the problem: "Sample size:" is 12, "Mean:" is 1.46 and "Standard deviation:" is 0.28. The test mean is 1 and we must select "greater than" from the "Alternative:" menu in the "Options" box. The results, after typing in the decision and conclusion, are shown below.

Jeonghun Kim
Lesson 13
Example

X = waste recycled (in pounds) by an adult per day

One-Sample T

```
Test of mu = 1 vs > 1


                              95% Lower
 N    Mean   StDev  SE Mean      Bound     T      P
12  1.4600  0.2800   0.0808     1.3148  5.69  0.000
```

Decision: Since the P-Value = 0.000 < 0.05, reject H0.
Conclusion: The mean waste recycled by adults per day in U.S. is significantly greater than 1 at the 5% level of significance. There is sufficient evidence to support the environmentalist's claim.

MINITAB ASSIGNMENT 13

See instructions on page 8.

1. Do the following using the data from Problem 30 on page 405 as the data from Problem 29 was used above.
   (a) Display the data and create a stem-and-leaf display in the session window.

   (b) Create a graphical summary as was done in the lesson above, then print it.

   (c) Use the t-test and P-Value to test the company's claim at the 5% level of significance.

   (d) Type in your decision and conclusion.

2. Do the following using the data below. The table gives scores of 12 students in a statistics class.

7	43	60	65	74	77
80	84	85	89	91	95

   (a) Display the data and create a stem-and-leaf display in the session window.

   (b) Create a graphical summary and determine if it is eligible to test the hypothesis that the mean of this data is 80 at the 5% level of significance.

   (c) Now remove the outlier 7 in the data, then create a new graphical summary and test the hypothesis that the mean score is 80 at the 5% level of significance.

Return to Cover Page