

**What does our data REALLY mean?**

**Part 1 – Box Plots**

1. Go to the following website:

[www.shodor.org/interactive/discussions/Outliers/](http://www.shodor.org/interactive/discussions/Outliers/)

2. Read through and study the discussion about outliers.

3. What is an outlier?

4. In the example of the second grade students and their teacher, whose height was the outlier?

5. What is the mean height of the second grade class when the teacher is included?

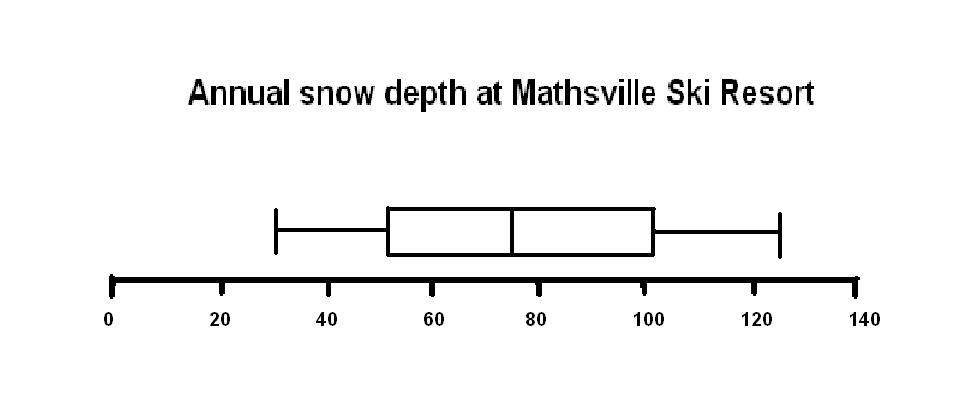
6. What is the mean height of the second grade class when the teacher is NOT included?

7. In summary, outliers can affect the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

8. Go to the following website. Read and study that page carefully.

[www.shodor.org/interactive/discussions/BoxPlot](http://www.shodor.org/interactive/discussions/BoxPlot)

9. In the box plot shown below, label the following parts:

Lower extreme (minimum)

Lower Quartile (Q1)

Median

Upper quartile (Q3)

Upper extreme (maximum)

10. Now go to the following website:

[www.shodor.org/interactive/activities/BoxPlot/](http://www.shodor.org/interactive/activities/BoxPlot/)

11. Under the pull-down menu labeled “Select a data set”, select “Body Fat Percentages in 2006”.

12. Approximately what was the median percentage body fat for everyone in the data set? (make sure “Graph all data” is selected under the graph)

13. What is the interquartile (within the box) range for the overall percentage body fat?

14. Select “Graph by Category” under the graph. Do men or women seem to have higher overall body fat?

15. How do the middle 50% of the data for body fat compare?

16. Based upon this data, is there a REAL difference in body fat of males compared to females? Or is the difference we see due to merely chance?

17. How did you determine your answer to #16?

**Part 2 Testing for Statistical Significance**

A certain researcher developed the inspired hypothesis that people are taller when they are wearing shoes than when they are not wearing shoes. To test this hypothesis, he took a random sample of 15 adults, measuring the height of each individual subject first with shoes on, and then again with shoes off.

18. Calculate the mean height with shoes on and with shoes off. What is the difference in the means? Can the researcher draw any conclusions?

|  |  |  |
| --- | --- | --- |
| Subject | Height with shoes on (in) | Height with shoes off (in) |
| 1 | 64.8 | 63.5 |
| 2 | 70.5 | 68.8 |
| 3 | 69.3 | 67.6 |
| 4 | 55.5 | 54.1 |
| 5 | 61.4 | 59.9 |
| 6 | 69.7 | 68.6 |
| 7 | 68.8 | 66.7 |
| 8 | 64.6 | 63 |
| 9 | 63.8 | 61.8 |
| 10 | 61.9 | 59.4 |
| 11 | 69.4 | 68.4 |
| 12 | 63 | 61.1 |
| 13 | 75.5 | 73.9 |
| 14 | 69.4 | 68.2 |
| 15 | 59.1 | 58.1 |

19. Create a box plot for the following data on a sheet of

graph paper. Be sure to put a title at the top. Label all

the parts you labeled on #9 above.

19. Based on the information in the box plot, is there a

REAL difference in the heights or is it due to chance?

20. We can use mathematics to know the answer to #19

for sure. We will perform a test for **statistical significance** called a **t-test**.

21. We will use Excel to help us find what is known as

The ***p-value*** (if you take AP Stats, you will learn how to

do the math that Excel will do for us to get a better understanding of what the t-test really does.) The p-value

is called the “**confidence level**”, or how sure you are

that your data is good.

By the way – who should someday take AP Stats????

ANSWER: anybody planning to go into science, medicine, engineering, business, law………basically anybody going into ANYTHING!!!

Back to p-values……

22. Many statisticians figure that **if we have a p-value of less than 0.05, then we can be 95% confident that our results are statistically significant.** How to run a t-test with Excel:

a. Copy and paste the above data into Excel.

 b. Click on any open cell.

c. Click on the “formula builder” that looks like

d. Double-click TTEST

e. To input into “Array 1”, select data set Height with shoes on

f. To input into “Array 2”, select data set height with shoes off

g. Enter 2 tails and Type 2

h. Hit return/enter on the keyboard and Excel will display the p-value.

**If p is less than 0.05, then the difference in height with shoes on and off**

**is statistically significant.**

In his investigation, the researcher used 15 subjects. What happens to the p-value if he uses more subjects? Try again using the following data.

23. What is the p-value now?

24. Based on your answer to #23, what is the moral of the story? In other words, what “rule of thumb” should we remember to follow when collecting and analyzing data?

|  |  |  |
| --- | --- | --- |
| Subject | Height with shoes on | Height with shoes off |
| 1 | 64.8 | 63.5 |
| 2 | 70.5 | 68.8 |
| 3 | 69.3 | 67.6 |
| 4 | 55.5 | 54.1 |
| 5 | 61.4 | 59.9 |
| 6 | 69.7 | 68.6 |
| 7 | 68.8 | 66.7 |
| 8 | 64.6 | 63 |
| 9 | 63.8 | 61.8 |
| 10 | 61.9 | 59.4 |
| 11 | 69.4 | 68.4 |
| 12 | 63 | 61.1 |
| 13 | 75.5 | 73.9 |
| 14 | 69.4 | 68.2 |
| 15 | 59.1 | 58.1 |
| 16 | 62.4 | 60.1 |
| 17 | 70.1 | 68.3 |
| 18 | 69.7 | 67.1 |
| 19 | 63.5 | 62.1 |
| 20 | 70.3 | 68 |
| 21 | 64.9 | 62.5 |
| 22 | 65.9 | 63.1 |
| 23 | 69.9 | 67.1 |
| 24 | 59.7 | 58.1 |
| 25 | 67.7 | 67 |
| 26 | 69.4 | 68.1 |
| 27 | 70.1 | 69.1 |
| 28 | 59.9 | 58.2 |
| 29 | 62.4 | 58.8 |
| 30 | 70.3 | 68.9 |
| 31 | 63.3 | 61.1 |
| 32 | 65.5 | 63.1 |
| 33 | 70.4 | 65.8 |
| 34 | 61.3 | 60.1 |
| 35 | 68.3 | 66.9 |
| 36 | 64.5 | 62.3 |
| 37 | 67.2 | 64.3 |
| 38 | 69 | 66.7 |
| 39 | 68.8 | 66.8 |
| 40 | 75.3 | 71.5 |