Data & Statistics Day #2: KEY

Below is the number of days it took for soybean seeds to germinate.

5 7 7 8 8 8 9 9 9 9 10 11 12 12 12 12 14 15 17 21 22 22 22 23 25 27 27 28 28 28 30 30 30

1. Find the 5 number summary for this data set:

Min: **5** Q1: **9**  Median: **14**  Q3: **26** Max: **30**

What is the range? 30 – 5 = **25** What is the inner-quartile range(IQR)? 26 – 9 = **17**

****

Make a box plot. Decide whether or not there are any outliers. Justify your reasoning.

**9 – 17\*1.5 = -16.5**

**26 + 17\*1.5 = 51.5**

**No outliers, as all of the data falls within the Tukey fences for outliers.**

 **Soybean Germination**



 Number of Days

Make a histogram – bar width 10 where the bars include (1-10,11-20 etc)

  **Soybean Germination**



 Number of Days

2. Given this data set, is 15 an outlier? Justify your solution.

3 3 4 4 5 5 5 5 5 5 5 6 6 6 6 7 7 8 8 8 8 8 9 9 10 15



Outlier

Figure 1. Boxplot of Data from Question 2.

**IQR = 8 – 5 = 3**

**5 – 4.5 = 0.5**

**8 + 4.5 = 12. 5**

**As shown in figure one, when the Tukey fences for outliers are placed on either side of the box plot, the number 15 remains outside of this range, therefore 15 is an outlier. Thank you John Tukey. ☺**

3. Given these graphs, decide whether or not there is a significant difference between Group A and Group B. Justify your answer using the “rule of thumb” for box plots. Discuss which boxplots appear skewed and which appear symmetric.

![[image]]()

On the first graph, what percent of SAT scores were between 400 and 550, for group B? Between 550 and 650?

**On the first graph it is likely that there is a difference between Group A and B as the boxes overlap, but the medians do not. Group A appears to be greater than group B. One conjecture is that Group A represents the girls at MMSTC while Group B represents the boys. 😂**

 **Group A appears to be skewed to the left and Group B appears to be roughly symmetric.**

**25% of the scores were between 400 and 550 for Group B.**

**25% of the scores were also between 550 and 600 for Group B.**

![[image]]()

On the second graph, where in the median for Group B? How many scores in the interquartile range?

**On the second graph there DOES appear to be a difference between Group A and B as the boxes do not overlap. Since this couldn’t have happened by chance alone, one conjecture could be that A represents MMSTC juniors while B represents MMSTC freshmen.**

**The median for Group B is in one of two places; either at the beginning of the box (Q1) or the end of the box (Q3). Its location cannot be determined.**

**50% of the scores are located in the interquartile range. A number cannot be given as the total number of scores is not listed for either group.**