Practice 5 – Statistics Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Measuring Variability for Skewed Distributions (IQR)

Lesson Summary

* Non-symmetrical data distributions are referred to as skewed.
* Left-skewed or skewed to the left means the data spreads out longer (like a tail) on the left side.
* Right-skewed or skewed to the right means the data spreads out longer (like a tail) on the right side.
* The center of a skewed data distribution is described by the median.
* Variability of a skewed data distribution is described by the interquartile range (IQR).
* The IQR describes variability by specifying the length of the interval that contains the middle $50\%$ of the data values.
* Outliers in a data set are defined as those values more than $1.5$(IQR) from the nearest quartile. Outliers are usually identified by an “\*” or a “•” in a box plot.

Consider the following scenario. Transportation officials collect data on flight delays (the number of minutes a flight takes off after its scheduled time). Consider the dot plot of the delay times in minutes for $60$ BigAir flights during December 2012:



1. How many flights left more than $60$ minutes late?
2. Why is this data distribution considered skewed?
3. Is the tail of this data distribution to the right or to the left? How would you describe several of the delay times in the tail?
4. Draw a box plot over the dot plot of the flights for December.
5. What is the interquartile range, or IQR, of this data set?
6. The mean of the $60$ flight delays is approximately $42$ minutes. Do you think that $42$ minutes is typical of the number of minutes a BigAir flight was delayed? Why or why not?
7. Based on the December data, write a brief description of the Big Air flight distribution for December.
8. Calculate the percentage of flights with delays of more than $1$ hour. Were there many flight delays of more than $1$ hour?
9. Big Air later indicated that there was a flight delay that was not included in the data. The flight not reported was delayed for $48$ hours. If you had included that flight delay in the box plot, how would you have represented it? Explain your answer.
10. Consider a dot plot and the box plot of the delay times in minutes for $60$ BigAir flights during January 2013.

How is the January flight delay distribution different from the one summarizing the December flight delays? In terms of flight delays in January, did Big Air improve, stay the same, or do worse compared to December? Explain your answer.

Compare the medians and the IQRs for the two sets of data.

