STAT 270 Q&A for Mastering Ch 1 Ideas Jan 12, 2006

1. How do you decide whether a mean or a median is an more appropriate measure of central location of a distribution? Skewness & Symmetry.

2. Why is the interquartile range preferred to the range for reporting the spread of a data distribution? Is there a situation when this would not be an issue? IQR does not increase systematically with n, the number of observations.

3. When is the standard deviation a good measure of spread? For data distributions that are symmetric about their mean. Also, if the distribution is strongly multi-modal, one would obviously need a more elaborate description of the spread (not just one number).

4. What is "distributed" in a data distribution? Frequency.

5. How is the shape of a distribution described, and why would it matter to the summary of the distribution? Symmetry, skewness, "gappiness", outliers.

6. What is the difference between an absolute frequency distribution and a relative frequency distribution? rel freq=abs freq/n

7. What feature of a data set makes a histogram a suitable summary? If the order of observations has no relevant information – most time series certainly should not be summarized this way.

8. Provide an example of a distribution for which the spread is a more important summary measure than the location? Distribution of repeated measures on a weigh scale that can be calibrated. Distribution of errors in tests on watch mechanisms. Comparison of lumber sizes between two different sawmilling processes.

9. Describe the pros and cons of using a dotplot vs a boxplot to display a single distribution. Dotplot is more informative and easier to explain. What about for multiple distributions? The box plot is good for comparing distributions, since the box represents the quartiles, and the outliers are easily identified as such. Dotplots can be positioned for comparison, but the quartiles would have to be judged informally; also, box plots are a bit better for comparing more than 2 or 3 distributions.

10. When would a stem-and-leaf plot be preferred to a dotplot? When you are constructing it by hand, or when you are learning what a frequency distribution is.

11. If you had to choose between MINITAB and R, and you did not have any experience with either one, which would you choose, and why? MINITAB is the one most often used in first courses. R is more common in advanced courses. But for basic operations, there is not much to choose between them.