Test 1 Key

1. In an experiment, researchers observe how a response variable behaves when they manipulate one or more factors. However, in an observational study, the researchers do not manipulate any factors. Instead, they observe characteristics of a subset of the members of one or more existing populations.

2. In an experiment, the explanatory variable is the one that researchers manipulate in order to observe changes in the response variable. An extraneous variable is any other variable which is thought to affect the response variable, but is not of interest in the study.

3. a) Bias is the tendency for a sample to differ from the corresponding population in some systematic way.

   b) Selection bias occurs when some part of the population is systematically excluded from the sample. However, non-response bias occurs when responses are not actually obtained from the individuals who were selected for the sample.

4. a) 45 students (.15 x 300)

   b) More students preferred Coke than any other drink, and colas (Coke and Pepsi) account for more than half of the preferred drinks.

   c) Relative frequency bar chart:
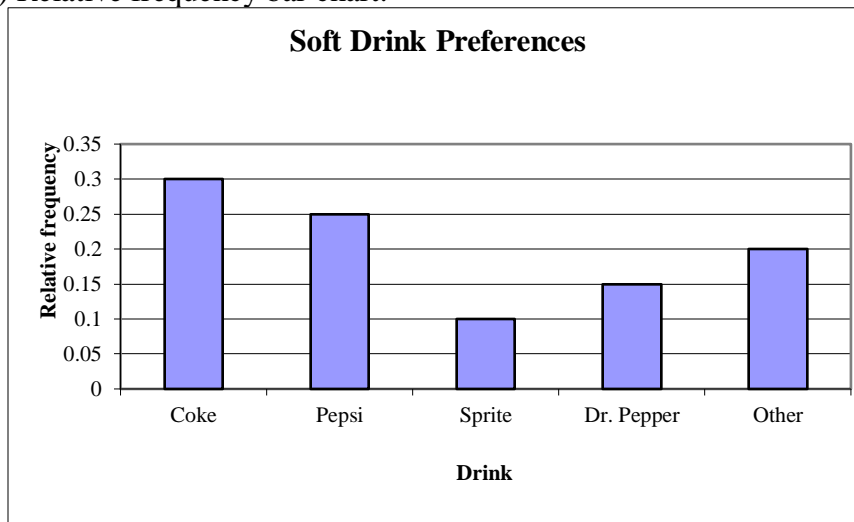
5. a)   comparative bar chart:
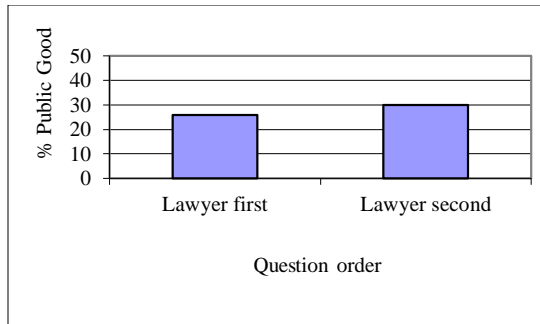
b) There is only a small difference that is shown in this data, from 26% to 30%. The order of the questions has not made much of a difference in the responses to the question in this case.

6.a) .15% or .0015

b) 84th percentile

c) Since a height of 67" is one standard deviation below the mean, we would expect only 16% of observations to be below 67". Since 25% of the observations should be less than the first quartile, the first quartile must be 68".
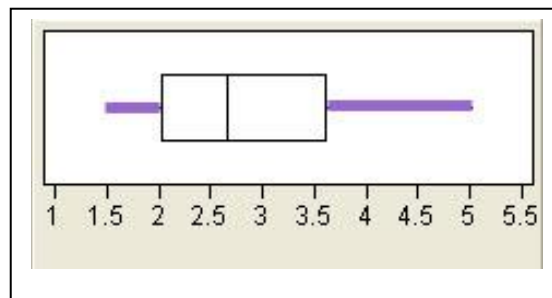
7.a) mean = 2.842
standard deviation = 1.002
median = 2.67
IQR = 3.61-2.03 = 1.58

b) 5-number summary: 1.45, 2.03, 2.67, 3.61, 5.08

c) This distribution is slightly skewed to the right. The distance from the minimum to the median is less than the distance from the median to the maximum.

8.a) 20% of n = 73 is 14.6, so we are looking for the 15th value in the ordered list, which is 8".

b) 7.854" and 15.844"

c) 3.859" and 19.839"

9. (a) Individual pairs of students would be randomly assigned to "trade papers" or "not trade papers" treatment groups. The non-trading students' work would be graded by the teacher each day and given back the next day.

(b) Each class would be considered a "block." Within each block both treatments would be randomly assigned as indicated in part (a).

(c) The results might be generalizable to other classes, but w/o doing the experiment in those classes there is no evidence suggest one could generalize. Statistics and history seem like they might be different enough that, although they are both classes with homework, the subject matter might be learned differently and the instant checking of the quizzes might be less or more of a help in one class or the other.

10. a) The proportion of households with 4 people has decreased slightly, from 15.5% of households to 14.6%.

b) The largest households, with 5 or more members decreased sharply, from 20.9% to 10.4% during the 30 years. This is a decrease of more than 50%. The other households either grew in size or only decreased slightly.

11. a) The two treatments will be: 1. The subject is recorded and 2. The subject is not recorded.

b) I would use blocking (pairing) to control the experience level of players. I would pair the two most experienced together as one block, the next two most experienced as the next block, and so on. Then the two members of each block would be randomly split into the two treatment groups. This way each treatment group should be roughly the equivalent with regard to experience level.

c) No, the results of this study should not be generalized to all male tennis players for at least two reasons. One, competitive tennis players are presumably more used to playing in front of crowds and would be less bothered by video recording than the typical player. Two, volunteers are not generally representative of any larger population.
Note: either reason should be sufficient to receive credit.

12. a) Because $\sum (x - \bar{x})$ will always be 0.

b) Because it is in the same units as the original observations.

c) Because on average, dividing by n-1 tends to give a value closer to σ^2 than dividing by n.

d) If a distribution is symmetric, then either the mean or median can be used as a measure of center. However, if the distribution is not symmetric, then the median should be used since it is resistant.