Math 121 Test #1 Answers Spring 2009

$(31, 62, 80, 92, 98)$, $\overline{x} = 75.0$.

1. (24 pts) $(0, 12, 16, 20, 24)$, $\overline{x} = 14.7$.

   (a) (3 pts) The direction extreme is both left and right. That is because both graphs are symmetric.

   (b) (4 pts) The most common decision rule was to reject H_0 if the observed value is ≤ 20 or ≥ 50. An equally good rule would be to reject H_0 if the observed value is ≤ 10 or ≥ 60.

   (c) (8 pts) Using the first rule in part (b), the rejection region includes 10, 20, 50, and 60, so $\alpha = \frac{6}{18}$. The acceptance region includes 30 and 40, so $\beta = \frac{4}{18}$. By the second rule, $\alpha = \frac{2}{18}$ and $\beta = \frac{10}{18}$.

   (d) (5 pts) The p-value of 20 is $\frac{6}{18}$. It is calculated in a manner similar to α.

   (e) (2 pts) A Type I error would be to conclude that you were holding Bag B when you were really holding Bag A. Therefore, if you conclude that the bag you are holding is Bag A, then it is not possible that you made a Type I error

   (f) (2 pts) By similar reasoning, if the bag you are holding really is Bag A, then it is possible that you will make a Type I error. You will make a Type I error if you conclude that you are holding Bag B.

2. (14 pts) $(4, 7, 10, 14, 14)$, $\overline{x} = 10.0$.

   (a) (6 pts) The null hypothesis should state the "neutral" hypothesis. The hypotheses should be

   H_0: White men (born around 1980), on the average, are just as tall as black men (born around 1980).

   H_1: White men (born around 1980), on the average, are not just as tall as black men (born around 1980).

   Instead of "not just as tall as," you could say "taller than."

   (b) (4 pts) Results with large p-values (larger than α) are not statistically significant. Therefore, if $\alpha = 0.05$ and the results were not statistically significant, then the p-value must be greater than 0.05. For example, p-value $= 0.10$.

   (c) (4 pts) Because the design was to take 100 individuals from each group, this would be a *stratified* sample.

3. (16 pts) $(8, 12, 16, 16, 16)$, $\overline{x} = 13.9$.

   (a) (4 pts) The explanatory variable is the number of hours of sleep the individual gets, on average.

   (b) (4 pts) The response variable is whether the person got sick from the cold virus.

   (c) (4 pts) This is an interesting situation. It could be viewed as either an observational study or as an experiment. In an experiment, it is the explanatory variables that are manipulated. The explanatory variable in this situation was the number of hours of sleep. There is no indication that that variable was manipulated.

   On the other hand, the researchers did spray cold viruses up the subjects' noses. That is manipulation, except that they did the same to every subject. There was only one level to that variable.

   I accepted either answer provided the explanation was sensible.

   (d) (4 pts) A confounding variable would be the subject's state of health upon entering the study. Another would be the strength of his immune system. Age is another, as older people are often less healthy. One could think of others.

   However, factors that cause a person to lose sleep would not be confounding variables. It does not matter *why* the person slept less than 7 hours. Every person who slept less than 7 hours had a reason, but the reasons had no impact on whether the person caught a cold.

4. (16 pts) $(0, 10, 12, 15, 16)$, $\overline{x} = 11.5$.

   (a) (4 pts) Enter `30→rand` to set the seed. Then enter `randInt(1,50)` and press `ENTER` five times. Or enter `randInt(1,50,5)` and get the same 5 numbers.

   (b) (4 pts) This randomizing is designed to avoid *selection* bias. It guarantees that each individual has the same chance to be in the group that sleeps at most 7 hours. It also eliminates one type of *experimenter* bias. If the experimenters are trying to show that more sleep prevents colds, then they may have a tendency to place the healthier individuals in the group that sleeps 8 hours or more.

   (c) (4 pts) The most obvious source of experimenter (given that one form of it was eliminated by the randomized design) is that the researchers may have a tendency to report the health more favorably of the group that sleeps at least 8 hours.

   Various forms of fraud do not count as experimenter bias. Of course, the whole experiment could be rigged or the researchers could lie about the results, but that is a different matter altogether. There is no indication that any sort of dishonesty was going on.

(d) (4 pts) The best way to avoid experimenter bias is to have the experimenters who make the observations not know who is in which group. This would be called double-blind except that it is hard to blind the subjects as to which group they are in. They can easily tell whether they are getting 7 hours of sleep or 8 hours of sleep.

5. (10 pts) $(1, 8, 9, 10, 10)$, $\overline{x} = 8.0$.

(a) (2 pts) The number of heart-attack patients that a hospital received in 2008 is *quantitative discrete* (only whole numbers).

(b) (2 pts) The city or county in which one resides is *qualitative* (not a number).

(c) (2 pts) The height of a person is *quantitative continuous*. It could be any value within a reasonable range.

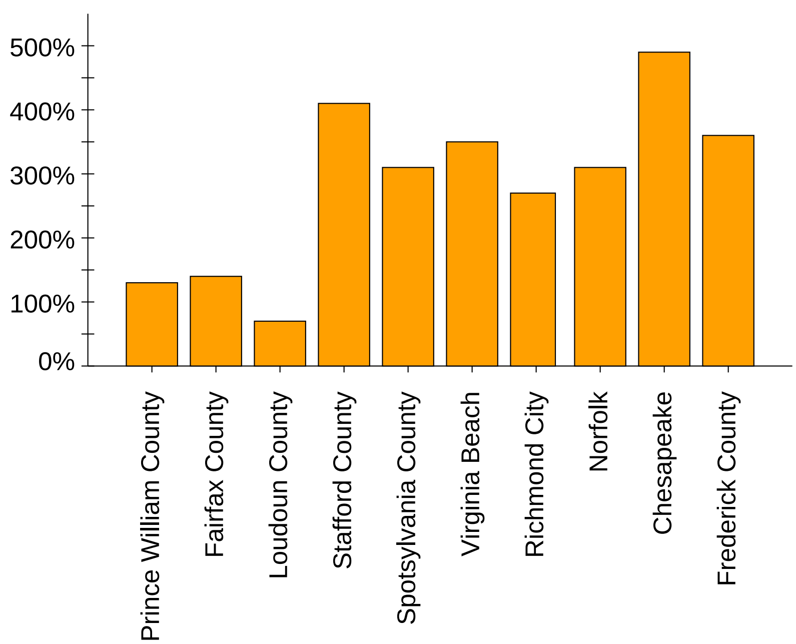
(d) (2 pts) Whether a person catches a cold during a specified 5-day period is *qualitative*. The values are "yes" and "no."

(e) (2 pts) The amount of sleep a person gets per night, on average, is *quantitative continuous*. It could be any value within a reasonable range.

6. (10 pts) $(4, 9, 10, 10, 10)$, $\overline{x} = 9.0$.

(a) (4 pts) The best type of graph is a bar graph. I also accepted histogram, but bar graph is better.

(b) (6 pts) The bar graph:



7. (10 pts) $(0, 6, 8, 10, 10)$, $\overline{x} = 7.9$.

(a) (6 pts) The stem-and-leaf display:

Stem	Leaf
0	4 8 2 5 8 7 6
1	6 5
2	1 1 3
3	1 0 0
4	3
5	8
6	
7	
8	3
9	1
10	
11	
12	9

(b) (4 pts) The shape of the distribution is unimodal and skewed right.