Math 121 **Test #1** **Fall 2007**

(100 pts) $(35, 53, 69, 81, 90)$, $\overline{x} = 67.0$.

1. (18 pts) $(2, 8.75, 9.5, 14.25, 18)$, $\overline{x} = 10.7$.

   (a) (3 pts) We would like α and β both to be small, so a good choice of rule would be: "Reject H_0 if the selected value is less than or equal to 66." You could also use 68. Mostly I checked to see that your rule made sense in the situation. That is, you rejected H_0 for smaller observed values. Some people said "Reject H_0 if the value is less than or equal to 66 and accept H_0 if it is greater than or equal to 70." We should never state both parts. If we fail to meet the rejection criterion, then we automatically accept. However, a further problem with that rule is that it overlooks the case of 68.

   (b) (3 pts) The direction of extreme to the left (towards the smaller values). That is the direction in which the values become less likely under H_0 and more likely under H_1.

   (c) (3 pts) If you used 66 in part (a), then $\alpha = 2/20 = 0.10$ (a value of 66 or less under H_0). If you used 68, then $\alpha = 5/20 = 0.25$.

   (d) (3 pts) If you used 66 in part (a), then $\beta = 4/20 = 0.20$ (a value greater than 66 under H_1). If you used 68, then $\beta = 1/20 = 0.05$.

   (e) (3 pts) Use the direction of extreme, just as in the case of α. The p-value of 72 is $13/20 = 0.65$ because there are 13 tokens $(0 + 0 + 2 + 3 + 3 + 5)$ that are less than or equal to 72.

   (f) (3 pts) You made a Type II error, to accept the null hypothesis when you should have rejected it.

2. (22 pts) $(6, 16.5, 18, 19.25, 22)$, $\overline{x} = 17.4$.

   (a) (4 pts) The null hypothesis should state the neutral position, that the "treatment" has no effect. Therefore, appropriate hypotheses are

   H_0: The dogs kept on a leash for eight hours are no meaner than those kept on a leash for no longer than one hour.

   H_1: The dogs kept on a leash for eight hours are meaner than those kept on a leash for no longer than one hour.

   (b) (6 pts)

      i. Selection bias is possible because the dogs were obtained from a dog pound. Thus, dogs in the dog pound are more likely to be selected than are dogs who are not in the dog pound. There is no need to explain the consequences of the bias. Selection bias simply says that some members of the population are more likely than others to be selected.

ii. Response bias is not possible. The dogs have no way of knowing what response was desired.

iii. Nonresponse bias is not possible. The dogs were not permitted to opt out of the study.

iv. Experimenter bias is possible because the researcher made the observations himself, knowing which group the dogs came from. Even if he tries to be fair, he could still be biased.

(c) (4 pts) Enter 250 → `rand` to set the seed. Then enter `randInt(1,100)` to get the first random integer. It should be 40. Press `ENTER` four more times to get 98, 34, 31, and 68.

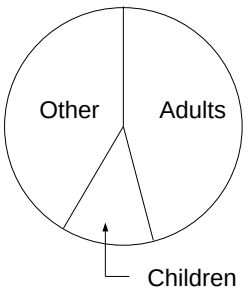
(d) (2 pts) The explanatory variable is the length of time that the dog was kept on a leash.

(e) (2 pts) The response variable was the level of meanness of the dog.

(f) (4 pts) This is an experimental study because the researcher manipulated the explanatory variable. He intentionally kept some dogs on a leash for 8 hours and others for only 1 hour.

3. (12 pts) $(3, 5.75, 9.5, 12, 12)$, $\overline{x} = 8.9$.

(a) (5 pts) The number of observations in the "Other" category is $147 - 67 - 20 = 60$. The three proportions are $67/146 = 0.46$, $60/146 = 0.41$, and $20/147 = 0.14$. The first two are close to equal and the last is just a bit more than $1/8$. If you compute the angles, the last angle should be $49°$, which is just a bit more than a $45°$ angle. Your pie chart should exhibit these features.



(b) i. (3 pts) The value of k should be $496/20 = 24.8$, or 25. I would accept 24.

ii. (4 pts) Enter 75 → `rand` to set the seed. Then enter `randInt(1,25)` (or `randInt(1,24)`). You should get 23. That is the starting point. Then repeatedly add 25 (or 24) to this number to get the sequence 23, 48, 73, 98, 123 (or 23, 47, 71, 95, 119).

4. (20 pts) $(6, 9.5, 12, 15, 18)$, $\overline{x} = 12.0$.

   (a) (2 pts) The precentage would be 77% of 90%, or $(0.90) \times (0.77) = 0.693$, or 69.3%.

   (b) (4 pts) The most likely type of bias is nonresponse bias because the majority of people who were mailed a survey did not respond. It is conceivable that there are other types of bias, but no other type of bias is indicated in the problem.

   (c) (5 pts) The two variables of interest are *whether* the person has been trained in domestic-violence prevention, and *whether* the person thinks that he needs such training. Both of these can be observed for each person in the sample. Quantities such as *the number of* people who have been trained in domestic-violence prevention would not be variables because you cannot observe their values for each member of the sample.

   (d) (3 pts) The population of interest is all health-care providers in Virginia.

   (e) (3 pts) The sample is the 2061 health-care providers who responded to the survey. I also accepted the 5581 who received a survey. However, although they were asked to be in the sample, they opted not to be.

   (f) (3 pts) The 77% figure is a statistic.

5. (8 pts) $(3, 6, 6.5, 7.25, 8)$, $\overline{x} = 6.5$.

   (a) The country is a qualitative variable. It is not a number.

   (b) The per capita GDP is a continuous quantitative variable. It is a number, and because it is computed as gross domestic product (the total of all goods and services) divided by the number of people, it is continuous, just like averages.

   (c) The population is a quantitative discrete variable. It is a number, but only a whole number.

   (d) The low estimate of total firearms is quantitative. It is probably discrete, but I allowed that it is conceivable that it is continuous. It would depend on how it is calculated and we don't know that.

6. (14 pts) $(3, 7, 10, 11.5, 14)$, $\overline{x} = 9.5$.

   (a) (4 pts) Based on the last sentence, the hypotheses would be

   H_0: The national uninsured rate in 2006 was the same as the national uninsured rate in 2005.

   H_1: The national uninsured rate in 2006 was higher than the national uninsured rate in 2005.

   In the alternative hypothesis, you could have said "not the same as." If you based your hypotheses on the first sentence in the article, then you would have said "Virginia" instead of "national."

(b) (3 pts) Because the increase was statistically significant, we know that they accepted the alternative hypothesis. (They rejected the null hypothesis.)

(c) (3 pts) Since they rejected the null hypothesis, we know that the p-value must have been less than or equal to 0.05. So choose any number you like that is less than or equal to 0.05. Do not choose 0.05 itself because then I cannot tell whether you were thinking "less than or equal to" or "greater than or equal to."

(d) (4 pts) Because they collected comprehensive data from three cities, they were using a cluster sample. Each city is presumed to be like the whole country in miniature. Had it been a stratified sample, they would had to have taken a random sample from all regions, or strata, of the country and in that case the strata should be chosen to be homogeneous groups.

7. (6 pts) $(0, 1.5, 4, 5.25, 6)$, $\overline{x} = 3.4$.

(a) They can control for selection bias by assigning individuals to the treatment group *at random*.

(b) They can control for response bias by using a placebo and a single-blind experiment where the subjects do not know whether they are receiving the treatment or the placebo.

(c) They can control for experimenter bias by making the experiment double-blind. That is, the people who make the observations do not know who is in the treatment group and who is in the control group.