Name: _____

Lab Problems 1-4 (worth 15 pts.)

Statistics 1000
Dr. Nancy Pfenning

1. Some physical characteristics, like shoe size, naturally follow a normal shape for specific age and gender groups. Use software to access the student survey data and complete the following tasks.

   (a) Find the mean _____and standard deviation _____for shoe sizes of all surveyed students.

   (b) Separate the data by gender and find the mean _____and standard deviation _____for shoe sizes of surveyed females.

   (c) Find the mean _____and standard deviation _____for shoe sizes of surveyed males.

   (d) The combined mean is (i) halfway between the mean for females and males (ii) closer to that for females (iii) closer to that for males. Explain how this result is affected by the relative numbers of females and males surveyed.

   (e) The combined standard deviation is (i) less than the individual standard deviations (ii) in between them (iii) more than the individual standard deviations. Explain how this occurs.

   (f) Report medians for all surveyed students _____, for females only _____, and for males only _____. For which group is the mean noticeably higher than the median? (i) combined (ii) females (iii) males

2. On-campus students may get *more* sleep than off-campus students because they don't have to get up earlier to commute; or perhaps they get *less* sleep because of late-night noise in the dormitories; or perhaps they get about the *same* amount of sleep as off-campus students. Which of these do you expect is the case? _____

   (a) Use software to access the student survey data, and report the mean hours of sleep for on-campus _____and off-campus students _____.

   (b) Which sample mean is higher? (i) on-campus (ii) off-campus

   (c) Does the difference appear to be (i) major or (ii) minor?

   (d) If we want to use mean hours slept by samples of on- and off-campus students to decide if there is evidence that all on-campus students sleep longer than all off-campus students, we are mainly concerned with (i) data production (ii) displaying and summarizing data (iii) probability (iv) statistical inference

3. Students taking introductory statistics courses at Pitt were asked to report whether or not they smoked, and whether they were non-vegetarian, a sometimes-vegetarian, or vegetarian.

   (a) First, we'll treat smoking as the explanatory variable. Suppose a professor anticipates that smokers tend not to be health-conscious and that very health-conscious students tend to be vegetarians. Would the professor expect to see a (i) higher or (ii) lower percentage of vegetarians (and sometimes-vegetarians) among the smokers, compared to the non-smokers?

   (b) Use software to access the student survey data. Combine percentages for vegetarians and sometimes-vegetarians, and report these separately for the smokers _____and for the non-smokers _____. Which group has a higher percentage of vegetarians? (i) smokers (ii) non-smokers

   (c) Next, we'll treat vegetarianism as the explanatory variable. If the professor anticipates that students' degree of health-consciousness could play a role in their values for both variables, would she expect to see a (i) higher or (ii) lower percentage of smokers among the vegetarians (and sometimes-vegetarians), compared to the non-vegetarians?

   (d) The combined percentage smoking for vegetarians and sometimes-vegetarians is 25%, and the percentage smoking for non-vegetarians is 18%. Is this consistent with what we would anticipate if health-consciousness plays a significant role in these two variables' values? _____

   (e) Often, the variable whose values occur first chronologically is taken to be the explanatory variable. Explain why, according to this criterion, there is no clear choice of roles for explanatory and response in this situation.

   (f) In fact, the differences in the percentages that you reported in part (b) could easily have come about by chance if we were sampling at random from a larger population where smoking and vegetarianism were not related. Can we conclude that these variables are related for the larger population? _____

4. A survey was taken of students at a large university in the fall of 2003; among other things, students were asked to report their weight and their age.

   (a) Use software to access the student survey data and produce a scatterplot of weights versus ages.

   (b) Report the correlation r _____. The relationship is (i) strong and negative (ii) weak and negative (iii) weak and positive (iv) strong and positive.

   (c) It is possible that the relationship seen between weight and age is really due to gender. If male students tended to be older than female students, the correlation between weight and age be (i) positive (ii) negative

   (d) Separate the weights and ages for males and females, and produce correlations separately for males _____and females _____.

   (e) The relationship seems to be a bit stronger for one group: (i) males (ii) females

   (f) The regression equation for females is approximately $Weight = 112 + 1(Age)$. This suggests that for every additional year, a female student will weigh how many more pounds? _____

   (g) The regression equation for males is approximately $Weight = 92 + 3.9(Age)$. Predict the weight of a 20-year-old male. _____

   (h) The standard deviation of male weights is 33 pounds. If we use the regression equation to predict a male's weight, given his age, the typical prediction error will be (i) less than 33 (ii) equal to 33 (iii) greater than 33