Statistics 7 NAME:___________________________________

Sample Midterm 1 Solutions

Student ID#:________________________________

One page of notes allowed. *You should have 8 problems on 3 pages.* Each part of each problem is 5 points.

1. Eleven students were asked to measure their pulses for 30 seconds and multiply by two to get their one-minute pulse rates. The results were: 32, 60, 62, 66, 70, 72, 74, 74, 78, 80, 84.
   a. Create a five-number summary for these pulse rates.

	72	
62		78
32		84

   b. The pulse rate of 32 is an outlier. The student forgot to multiply by two. Which one of the three reasons for outliers discussed in class and in the book applies here?

   A mistake was made while taking a measurement or entering it into the computer.

   c. Correct the outlier and redo the five-number summary.

	72	
64		78
60		84

   d. For each of the following statistics, specify whether it would increase, decrease or remain the same after the outlier is corrected.

   Mean: ____*Increase*________ Median: _______*Remain the same*____

   Standard deviation: ____*Decrease*_______ Interquartile range: ____*Decrease*______

   Standardized score for a height equal to the mean of the current data: ___*Remain the same*___

2. A study done by the Center for Academic Integrity at Rutgers University asked 2116 students at 21 colleges and universities a series of questions. Some of the schools had an "honor code" and others did not. Of the students at schools with an honor code, 7% reported having plagiarized a paper via the Internet, while at schools with no honor code, 13% did so. (*Sacramento Bee*, Feb 29, 2000, D1.)
   a. Was this a randomized experiment or an observational study?

   It was an observational study. (Schools cannot be randomly assigned to have an honor code.)

   b. What are the explanatory and response variables for this study?

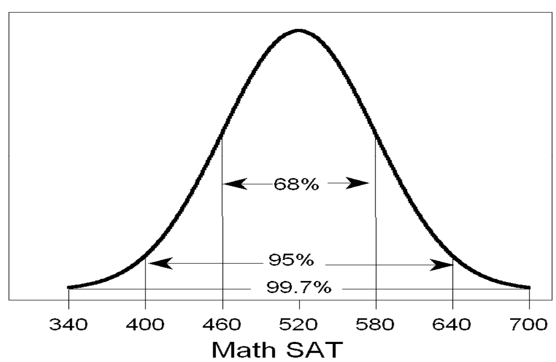
   Explanatory variable is whether the respondent's school had an honor code or not.
   Response variable is whether the student had ever plagiarized a paper via the Internet or not.

   c. Give an example of a possible confounding variable and explain why it fits the criteria for a confounding variable.
   Answers will vary. But you must explain that your confounding variable (1)Affects the response (plagiarism) and (2)Is related to the explanatory variable of whether the school has an honor code or not. An example in this case is how strict the admissions policy is for each school. Schools that accept better students may be less likely to have dishonest students, and may be more likely to have an honor code.

3. Over many years, the average rainfall during the month of November in San Francisco, California, is 2.62 inches. The standard deviation is 2.79 inches. Based on this information, explain how you can tell that the distribution of rainfall values cannot be bell-shaped.

   For bell-shaped data, we know from the Empirical Rule that a substantial proportion of the data should fall more than one standard deviation below the mean. But in this case, one standard deviation below the mean is −0.17, which is already an impossible rainfall value. Rainfall in November could be 0, but cannot be negative. So, the data must be skewed to the right with some high values.

4. Math SAT scores for students admitted to a university are bell-shaped with a mean of 520 and a standard deviation of 60.
   a. Draw a picture of these SAT scores, indicating the cutoff points for the middle 68%, 95% and 99.7% of the scores.



   b. A student had a math SAT score of 490. Find the standardized score for this student *and* draw where her score would fall on your picture in part (a).
   $z = \frac{490-520}{60} = \frac{-30}{60} = -0.5$. *It should be drawn halfway between 460 and 520 on the picture above.*

5. The table below shows the opinions of 908 respondents in the General Social Survey to the question "Do you believe there is life after death?" The purpose of examining the data is to see if there is a gender difference in how people would respond to this question.

	Yes	No	Total
Male	282	109	391
Female	408	109	517
Total	690	218	908

   a. Write the null and alternative hypotheses for this study.
   <u>Null hypothesis</u>: *There is no relationship in the population between gender and belief in life after death.*
   <u>Alternative</u>: *There is a relationship in the population between gender and belief in life after death.*

   b. Find the expected count for the number of males who believe there is life after death.
   $$\frac{(391)(690)}{908} = 297.13$$

   c. The chi-square statistic for this situation is 5.63 and the p-value is 0.018. State the appropriate conclusion in the context of this situation.
   Reject the null hypothesis and conclude that there is a gender difference in the population regarding belief in life after death. You could also say that there is a statistically significant relationship between gender and belief in life after death.

6. In the previous question, there were 517 females in the sample and 408 of them said they believe there is life after death.
   a. What is the margin of error for this survey (for females only)?
   $$\frac{1}{\sqrt{517}} = .044 \textit{ or } 4.4\%$$
   b. Compute a 95% confidence interval for the percent of females in the population who believe there is life after death.

The sample proportion is $\frac{408}{517} = .789$ *or 78.9%. So the interval is 78.9% ± 4.4% or 74.5% to 83.3%*

7. In a study of acupuncture for treating pain, 100 volunteers were recruited. Half were randomly assigned to receive acupuncture and the other half to receive a sham acupuncture treatment. The patients were followed for 6 months and the treating physician measured their degree of pain relief. The patients did not know which treatment they actually received, but the treating physicians were aware of who was getting acupuncture and who wasn't. For the following list of terms, *circle* the ones that apply to this study and *cross out* the ones that do not apply to this study:

   The ones that apply are in bold:
   Randomized experiment, Observational study, Blocking, **Single blind**, Double blind

8. The percent of high school graduates who took the SAT exam in 1998 varied widely from state-to-state (including Washington DC), with a high of 83% in Washington DC and a low of 4% in Mississippi and Utah. The regression equation relating the average 1998 verbal SAT score in the states to the percent of graduates who took the SAT in the state is:

$$\text{Verbal Average} = 573 - 1.08\ (\text{PercentTook})$$

   a. Is the association between the percent who took the exam and the average verbal SAT score positive or negative? Explain why, logically, that would be expected.

We know the association is negative because the slope is negative. This makes sense, because the good students will generally take the SAT exam no matter what. When the state has a higher percentage of graduates taking the exam, it means that the weaker students are taking it as well. Therefore, the higher the percentage taking it, the lower the average is likely to be.

   b. In California 47% of high school graduates took the exam. What is the *predicted* average verbal SAT score for a state in which 47% of high school graduates took the exam?

      Predicted verbal average = 573 − 1.08 (47) = 522.24.

   c. The average verbal SAT score for California was 497. What is the *residual* for California?

      Residual = actual – predicted = 497 – 522.24 = −25.24

   d. Does the *intercept* of 573 have a logical interpretation in this situation? Explain.

   No. It would be the average score if no one took the exam. But if no one took it, there can be no average score.