Statistics 160 Midterm #1 – October 7, 2008

This exam has 11 problems on 8 numbered pages and is worth a total of 60 points.

You have 75 minutes to complete this exam. Please read all instructions carefully, and check your answers. Show all work neatly and in order, and clearly indicate your final answers. Answers must be justified whenever possible in order to earn full credit. **Unless otherwise specified, no credit will be given for unsupported answers, even if your final answer is correct.** *Points will be deducted for incoherent, incorrect, and/or irrelevant statements.*

This exam is closed-book, except that one $8\frac{1}{2} \times 11$ double-sided page of handwritten notes is permitted as well as a calculator. No other aids are allowed. A copy of Table A will be provided.

You must answer all of the questions in the space provided. Note that blank space is NOT an indication of a question's difficulty.

Name: _____

Instructor: Michael Kozdron

| Problem	Score
1	
2	
3 and 4 and 5	
6 and 7	
8	
9	
10	
11	

TOTAL: _____

1. (*8 points*) The number of people living on American farms has declined steadily during the last century. Shown below are data on the farm population (millions of persons) for selected years between 1935 and 1980.

Year	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
Population	32.1	30.5	24.4	23.0	19.1	15.6	12.4	9.7	8.9	7.2

The summary statistics for Year and Population are as follows.

	mean	standard deviation
Year	1957.5	15.13825
Population	18.29	8.986712

The square of the correlation is $r^2 = 0.9770393$ which indicates that most of the observed variation in farm population is accounted for by the linear change over time. A scatterplot for this data is shown below.

American Farm Population, selected years (1935–1980)
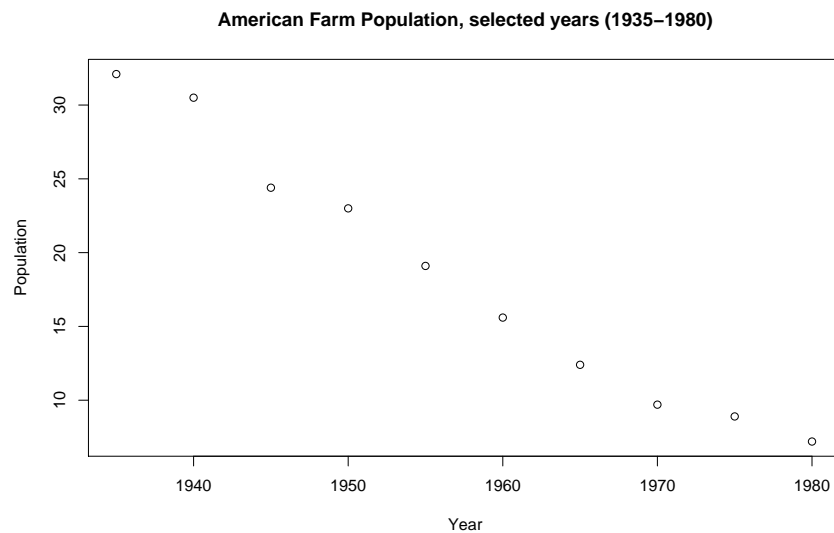
(a) Identify the explanatory variable and the response variable.

(b) Determine the value of the correlation r.

(continued)

1

(c) Determine the equation of the least-squares regression line of farm population on year.

(d) Use your regression model from **(b)** to predict the number of people living on farms in 2000. Is this result reasonable? Why?

For the following problem, circle the best answer.

2. (*2 points*) The ages of people in a class (to the nearest year) are as follows:

Age	18	19	20	21	22	23	24	25	32
Number of students	14	120	200	200	90	30	10	2	1

What is true abut the median age?

A It must be less than 20.

B It must be 20.

C It must be 20.5.

D It must be 21.

E It must be over 21.

For each of the following problems, circle the best answer.

3. (*2 points*) A researcher reports that, on average, participants in a study lost 4.1 kg after two months on their new diet. A friend of yours comments that she tried the diet for two months and lost no weight. Which of the following statements is then correct?

A Your friend must not have followed the diet correctly, since she did not lose weight.

B Because your friend did not lose weight, the report must not be correct.

C The report gives only the average. This does not imply that all participants in the study lost 4.1 kg or even that all lost weight. Your friend's experience does not necessarily contradict the study results.

D For the study to be correct, we must now add your friends results to those of the study and recompute the new average.

4. (*2 points*) The average salary of all female workers at a large plant is \$35 000. The average salary of all male workers at the plant is \$41 000. If there are more male workers than female workers at the plant, then the average salary at the plant must be

A exactly \$38 000.

B larger than \$38 000.

C smaller than \$38 000.

D not able to be determined from the information given.

5. (*2 points*) The median age of five people in a meeting is 30 years. One of the people, whose age is 50 years, leaves the room. The median age of the remaining four people in the room is

A exactly 30 years.

B less than 30 years.

C more than 30 years.

D not able to be determined from the information given.

6. (*6 points*) The business magazine *Forbes* reports that 4567 companies sold their first stock to the public between 1990 and 2000. The *mean* change in the stock prices of these companies since the first stock was issued was $+111\%$. The *median* change was -31%. Explain how this could happen. (*Hint*: Start with the fact that Cisco Systems stock went up 60 600%.)

7. (*6 points*) Joe's retirement plan invests in stocks through an "index fund" that follows the behaviour of the stock market as a whole, as measured by the S&P 500 stock index. Suppose that Joe now wants to buy a "mutual fund" that does not track the S&P 500 stock index closely. He reads that monthly returns from the Fidelity Technology Mutual Fund have correlation $r = 0.77$ with the S&P 500 stock index and that monthly returns from the Fidelity Real Estate Mutual Fund have correlation $r = 0.37$ with the S&P 500 stock index.

  (a) Which of these two mutual funds has the closer relationship to returns from the stock market as a whole?

  (b) Does the information given tell Joe anything about which fund has the higher returns?

8. (*8 points*) Observational studies had suggested that vitamin E reduces the risk of heart disease. Careful experiments, however, showed that vitamin E has no effect, at least for women. According to a commentary in the *Journal of the American Medical Association*:

> Thus, vitamin E enters the category of therapies that were promising in epidemiologic and observational studies but failed to deliver in adequately powered randomized controlled trials. As in other studies, the "healthy user" bias must be considered, i.e., the healthy lifestyle behaviours that characterize individuals who care enough about their health to take various supplements are actually responsible for the better health, but this is minimized with rigorous trial design.

A friend who knows no statistics asks you to explain this.

(a) What is the difference between observational studies and experiments?

(b) What is a "randomized controlled trial"?

(c) How does "healthy user bias" explain how people who take vitamin E supplements have better health in observational studies but not in controlled experiments?

9. (*8 points*) Data from Environment Canada indicate that the yearly snowfall in Regina is normally distributed with mean 100 cm and standard deviation 10 cm.

(a) Find the probability that next year's snowfall is between 90 cm and 120 cm.

(b) What must next year's snowfall be so that there is an 80% chance of receiving at least that much snowfall?

10. (*8 points*) A national restaurant chain conducts a simple random sample of some of its customers. The card these customers are given to fill out asks their opinions about their meal, the service, the cleanliness, and so on. One question asks the customer to rate the quality of the service as poor, below average, average, above average, or outstanding. The following data represent the results obtained for the Regina branch of this restaurant chain.

Poor	Below average	Average	Above average	Outstanding
7	14	33	67	19

(a) Describe **(i)** the population of interest, and **(ii)** the sample. **(iii)** What is the variable of interest? Is it categorical or quantitative?

(b) What proportion of customers rated the sampled restaurant as either above average or outstanding?

(c) Can the results derived from the given data be extended to *all* other branches in this restaurant chain? Explain why or why not.

11. (*8 points*) Suppose that you are interested in estimating the average amount of money that was spent on books for Fall 2008 classes by full-time undergraduate students at the University of Regina.

Explain how you would design a survey which uses regression to help you estimate this average amount of money spent on books.

Be sure to define the population of interest, how and who you would sample, as well as which explanatory variable x and response variable y you would measure for each sampled individual.