Practice First Midterm Exam

Statistics 200 Spring 2008 (Pfenning)

This is a closed book exam worth 150 points. You are allowed to use a calculator and a two-sided sheet of notes. There are 8 problems, with point values as shown. If you want to receive partial credit for wrong answers, show your work. Don't spend too much time on any one problem.

1. (25 pts.) An article entitled **Prevent Migraines Naturally** states that "Patients who took 100 milligrams of the supplement Coenzyme Q10 (CoQ10) three times a day had up to 50 percent fewer migraines and less nausea after three months, finds a new Swiss study. Their headaches were also shorter and not as severe. Researchers believe that CoQ10 prevents migraines by boosting energy production in cells." Identify each of the following as well as you can.

   (a) explanatory variable _____

   (b) explanatory variable type: (i) quantitative (ii) categorical (iii) not clear

   (c) response variable _____

   (d) What comparison is being made?

      i. patients who do or do not take CoQ10
      ii. patients before and after taking CoQ10

   (e) Which one of the following additional pieces of information would be **most** helpful in deciding whether CoQ10 is really beneficial for migraine sufferers?

      i. Was there a control group taking a placebo?
      ii. Were the patients randomly chosen to participate in the study?
      iii. How did researchers define a migraine?
      iv. Were different dosages of CoQ10 tested?
      v. How realistic was the setting?

2. (20 pts.) Researchers are interested in how pesticides affect children's health. Here are designs for two actual studies:

   Design A: The EPA proposes to pay each family in their study $970, some children's clothing and a camcorder in exchange for their participation for two years, during which they would use various household pesticides and have their children tested for absorption rates.

   Design B: A study will recruit and enroll participants in 96 locations over the next four years and track them from conception through their children's 21st year. The 2.7 billion dollar study will monitor the children's health, along with pesticide exposures in the air, food, and water.

   (a) Which design is an observational study? (i) A (ii) B (iii) both (iv) neither

   (b) Which design is more vulnerable to confounding variables?
   (i) A (ii) B (iii) both the same (iv) neither

(c) What is the most worrisome flaw in Design A?

(d) Which of the above in this problem are closed questions?
(a), (b), or (c) (Circle any that are closed.)

3. (20 pts.) Weights of male mallard ducks are normally distributed with mean 800 grams, standard deviation 100 grams.

   (a) According to the 68-95-99.7 Rule, 95% of male mallards weigh between _____and _____grams.

   (b) Almost none of the mallards weighed less than how many grams?

   (c) What is the z-score for a male mallard weighing 740 grams? _____

   (d) A male mallard weighing 740 grams could best be described as
   (i) very light (ii) a bit light (iii) a bit heavy (iv) very heavy

4. (15 pts.) Breadths (in centimeters) of a sample of male Egyptian skulls from 4000 B.C. and 150 A.D. are listed below and displayed with a side-by-side boxplot.

4000 B.C.	131	119	138	125	129	126	131	132	126	128	128	131
150 A.D.	136	130	126	126	139	141	137	138	133	131	134	129

   (a) During which time did skulls tend to be broader?
   (i) 4000 B.C. (ii) 150 A.D. (iii) both about the same

   (b) During which time did breadth vary the most for the middle half of values? (i) 4000 B.C. (ii) 150 A.D. (iii) both about the same

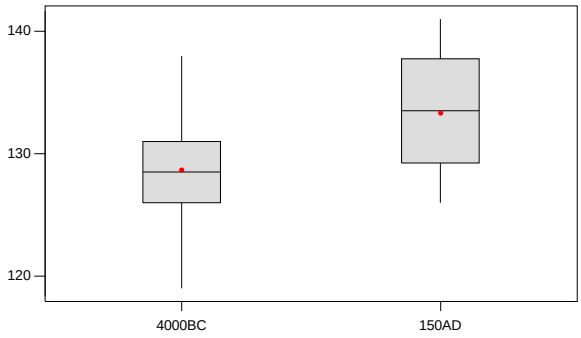
   (c) Which of these is your best guess for standard deviation of skull breadths in 150 A.D.? (i) 5 cm (ii) 20 cm (iii) 50 cm

   (d) We could characterize shapes as being
   (i) both skewed left
   (ii) both skewed right
   (iii) 4000 B.C. skewed left, 150 A.D. skewed right
   (iii) 4000 B.C. skewed right, 150 A.D. skewed left
   (v) both symmetric

   (e) Sample mean skull breadth in 4000 B.C. would be denoted (i) $\bar{x}_1$ (ii) s_1 (iii) μ_1 (iv) σ_1 (v) $\hat{p}_1$ (vi) p_1 (vii) n_1

Boxplots of 4000BC and 150AD

(means are indicated by solid circles)



5. (30 pts.) Women taking a certain combination of blood pressure drugs (diuretics combined with calcium channel blockers) were suspected of having a higher risk of fatal heart attacks than those taking other blood pressure drugs. This table represents counts observed in recent years.

	Fatal Heart Attack	No Fatal Heart Attack	Total
Combination Drugs	29	1171	1200
Other Drugs	21	1779	1800
Total	50	2950	3000

(a) Find the probability of a fatal heart attack for those taking the combination drugs.

(b) Find the overall probability of a fatal heart attack._____

(c) For those 3000 women, fatal heart attacks were (i) more likely if combination drugs were taken (ii) less likely if combination drugs were taken (iii) neither (combination drugs made no difference)

(d) If fatal heart attacks and type of drug treatment were independent of each other, how many of the 1800 women who took other drugs would have suffered a fatal heart attack?

(e) Before concluding that the combination drugs are responsible for higher rates of heart attacks, which one of the following is most important?

i. Make sure the researchers were blind.

ii. Make sure there wasn't a tendency for women who were more at risk of fatal heart attacks to be prescribed the combination drugs.

iii. Make sure the women studied were all the same ages.

iv. Make sure the women studied were a variety of ages.

(f) Question (e) concerns itself with (i) data production
(ii) displaying and summarizing (iii) probability (iv) statistical inference

6. (5 pts.) Suppose we are interested in finding out how age of fish fossils relates to the depth below sea floor at which they are found.

(a) What would be an appropriate display?
(i) bar graph (ii) histogram (iii)side-by-side boxplots (iv) scatterplot

(b) Which of these would provide the best summary?
(i) compare percentages (ii) compare means and standard deviations
(iii) compare Five Number Summaries (iv) report the correlation

7. (5 pts.) Suppose we are interested in finding out how common it is for students to take part in study-abroad programs.

(a) What would be an appropriate display?
(i) pie chart (ii) histogram (iii)side-by-side boxplots (iv) scatterplot

(b) Which of these would provide the best summary?
(i) report percentage (ii) report mean and standard deviation
(iii) report Five Number Summary (iv) report the correlation

8. (30 pts.) Size (in hundreds of square feet) and price (in thousands of dollars) were recorded for 9 custom-home resales in an exclusive subdivision in Phoenix, Arizona.

   (a) What is the explanatory variable?_____

   (b) The regression equation predicts that for each additional hundred square feet, price increases by _____thousand dollars.

   (c) Which of the following is the best guess for r?
   (i) -.83 (ii) -.43 (iii) -.13 (iv) .13 (v) .43 (vi) .83

   (d) Predict the price of a home with size 40 hundred square feet. _____

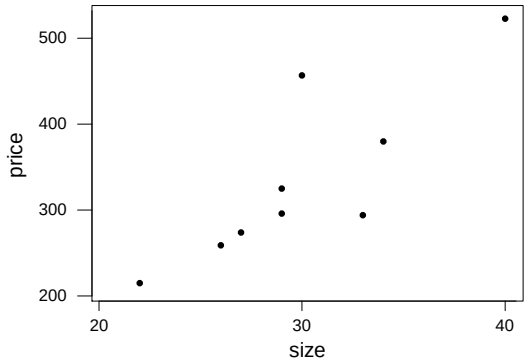
   (e) If a home with size 40 hundred square feet actually cost 523 thousand dollars, then the residual is _____(be sure to include sign + or -).

   (f) What is the typical size of prediction errors made by the regression line?

   (g) If the home with size 40 and cost 523 were omitted from the regression, the value of r would (i) increase (ii) decrease (iii) stay the same

   (h) According to the output, there was an outlier home which had size 30 and cost 457. If it were omitted from the regression, the value of r would (i) increase (ii) decrease (iii) stay the same

```
The regression equation is
price = - 141 + 15.9 size
Predictor        Coef     SE Coef          T        P
Constant       -140.9       123.3      -1.14    0.291
size           15.894       4.057       3.92    0.006
S = 59.62       R-Sq = 68.7%     R-Sq(adj) = 64.2%
Unusual Observations
Obs       size      price         Fit      SE Fit    Residual    St Resid
  7       30.0      457.0       335.9        19.9       121.1        2.15R
R denotes an observation with a large standardized residual
```


price
500
400
300
200
20
30
40
size