Practice Exam 2

Statistics 0800 Fall 2014 Dr. Nancy Pfenning

This is a closed book exam worth 150 points. You are allowed to bring a calculator, normal table, and a two-sided sheet of notes. If you want to receive partial credit for wrong answers, show your work. Don't spend too much time on any one problem.

1. (15 pts.) For each of the following situations, circle the name of the best picture to display the data:

   (a) students' gender and self-rated driving (on a scale of 1 to 10)
   (i) histogram (ii) piechart (iii) scatterplot (iv) multiple boxplots

   (b) students' favorite color
   (i) histogram (ii) piechart (iii) scatterplot (iv) multiple boxplots

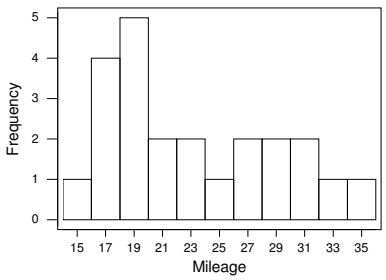
   (c) students' score on first homework and score on first exam
   (i) stemplot (ii) bar graph (iii) scatterplot (iv) multiple boxplots

2. (15 pts.) Mileages (in miles per gallon) were recorded for 23 models of car. A histogram of the data is shown below.

   (a) The shape of the distribution is
   (i) skewed left/low outliers (ii) fairly symmetric (iii) skewed right/high outliers

   (b) The median mileage is 22 mpg. Considering the shape of the distribution, which of these is your best guess for the mean?
   (i) 12 (ii) 21 (iii) 22 (iv) 23 (v) 32

   (c) Which of these is your best guess for standard deviation, keeping in mind that it measures the typical distance of values from their mean?
   (i) 0.06 (ii) 0.6 (iii) 6 (iv) 60



3. (5 pts.) Suppose Pitt students are asked to report the number of credits they're taking and the number of hours they spend in classes. The correlation should be close to
(a) -1 (b) -0.5 (c) 0 (d) $+0.5$ (e) $+1$

4. (5 pts.) For which of these situations is linear regression appropriate?

   (a) Predicting height from age based on ages from 10 to 30.

   (b) Predicting living situation (on or off campus) from age, based on students' ages from 18 to 22.

   (c) Linear regression is appropriate for both of the above.

   (d) Linear regression is appropriate for neither of the above.

5. (10 pts.) In 1980, when the Consumer Price Index was 82, PAT bus fares were 0.60.

   (a) If the amount were to keep pace with inflation, how much should bus fares have been in 2013, considering that the CPI was 230?_____

   (b) In fact, PAT bus fare in 2013 was \$2.50. Which of the following is true?

      i. A fare of \$2.50 was roughly keeping pace with inflation.

      ii. A fare of \$2.50 was too high, even accounting for inflation.

      iii. A fare of \$2.50 was a bargain when inflation is taken into account.

   (c) When fares were regressed on date from 1980 to 2002, the resulting regression line equation was $y = -92 + 0.0468x$. What fare would the regression line predict for 2013? _____

   (d) Why might your answer from part (c) do a poor job in predicting the fare for 2013? (Answer with a single word.) _____

6. (20 pts.) Suppose battery lifetimes are normally distributed with mean 12 hours and standard deviation 2 hours.

   (a) What proportion of batteries last less than 7 hours?_____

   (b) What proportion of batteries last more than 14.8 hours?_____

   (c) The worst 1% last less than how many hours? _____

   (d) The best 10% last more than how many hours? _____

   (e) How do the mean and median of battery lifetimes compare? (i) mean less than median (ii) mean equals median (iii) mean greater than median

   (f) Use the 68-95-99.7 Rule to sketch the distribution of battery lifetimes.

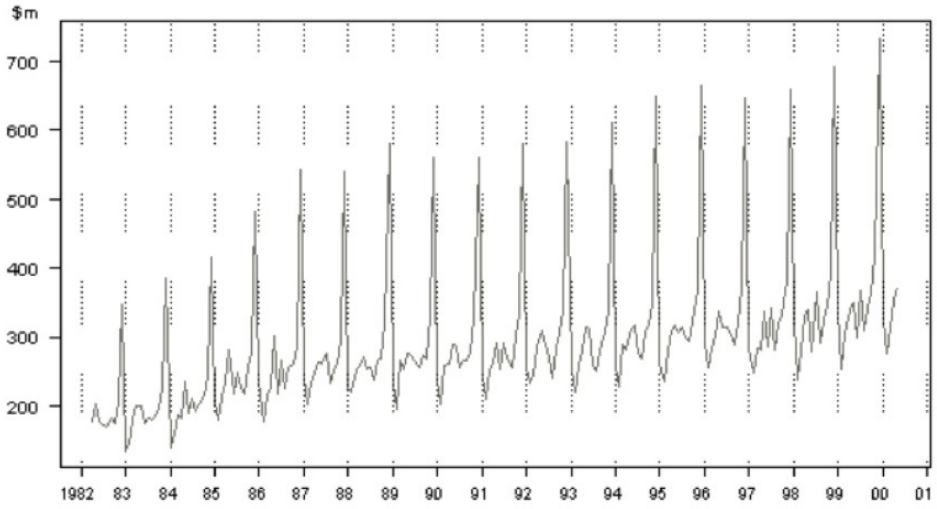
7. (15 pts.) Monthly sales were recorded for Australian retail department stores over about twenty years' time.

   (a) The overall change from 1982 to 2001 is best described as
   (i) a long-term downward trend (ii) a long-term upward trend
   (iii) seasonal components (iv) an irregular cycle

   (b) Each year sales increase dramatically during the holidays in December, showing
   (i) a long-term downward trend (ii) a long-term upward trend
   (iii) seasonal components (iv) an irregular cycle

   (c) Suppose we used the data for the entire period, from 1982 to 2001, to produce a regression line to predict retail sales based on month. The predictions would be

      i. quite accurate, as long as the calculations are done correctly
      ii. too high for holiday time and too low for the rest of the year
      iii. too low for holiday time and too high for the rest of the year


$m
700
600
500
400
300
200
1982
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
00
01

7. (30 pts.) Researchers have found a clear association between TV viewing time (explanatory variable) x and excess weight (response variable) y in children.

   For each of the possible explanations (a) through (g), circle the "reason for a relationship" that applies; each number 1 through 7 should be used exactly once. [The first has been done for you.]

   (a) There happened to be more overweight kids in the study who watched a lot of TV.
   1 2 3 4 5 6 **(7)**

   (b) Having a slower metabolism causes some children to prefer sedentary activities like watching TV, and also causes them to gain weight easily.
   1 2 3 4 5 6 7

   (c) Watching too much TV is what makes children obese.
   1 2 3 4 5 6 7

   (d) Many modern lifestyle habits, including TV, are responsible for childhood obesity.
   1 2 3 4 5 6 7

   (e) Heavy kids choose to watch a lot of TV because other activities don't suit them.
   1 2 3 4 5 6 7

   (f) Kids watch TV now more than ever, and obesity is on the rise.
   1 2 3 4 5 6 7

   (g) Low-income kids tend to watch more TV, and they also tend towards obesity.
   1 2 3 4 5 6 7

1. explanatory variable is direct cause of response variable
2. response variable is causing change in explanatory variable
3. explanatory variable is contributing but not sole cause of response variable
4. confounding variables may exist
5. both variables result from a common cause
6. both variables are changing over time
7. the association may be nothing more than coincidence

8. (30 pts.) A two-way table is constructed from survey data to see if gender is related to whether or not a student eats breakfast:

	Breakfast	No Breakfast	Total
Females	89	151	240
Males	61	59	120
Total	150	210	360

(a) Which group has a higher proportion eating breakfast?
(i) females (ii) males (iii) both the same

(b) Compute a table of counts expected if eating breakfast were independent of gender.

Expected	Breakfast	No Breakfast
Females		
Males		

(c) Compare the observed and expected numbers by computing each $\frac{(observed-expected)^2}{expected}$ rounded to the nearest tenth (one decimal place).

Compare	Breakfast	No Breakfast
Females		
Males		

(d) Calculate the chi-square statistic (round to the nearest tenth).

(e) The difference between observed and expected counts is
(i) relatively small (ii) relatively large (iii) borderline

(f) Which one of these claims is supported by your conclusions in (e)?

i. Proportions eating breakfast are bound to differ somewhat in a sample; the difference that we see in this sample could easily have come about by chance, and there really is no difference for males and females in general.
ii. Women are more likely than men to skip breakfast, for whatever reason.
iii. Women skip breakfast more than men because they are dieting.
iv. Women are more health-conscious than men, and are less likely to miss breakfast because they know it's an important meal.