50 pts

22S:30/105, Statistical Methods and Computing
Spring 2014, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

Name: __Solutions_____

Course no. (30, 105, or 197) _____

1. What type of plot would be appropriate to represent each of the following variables? (Circle all that apply for each.)

    (a) body temperatures in women ages 65-74

    5
        i. bar graph
        ii. boxplot
        iii. histogram
        iv. pie chart
        v. stem-and-leaf plot

    (b) the ratings of restaurants in a local newspaper (superior, excellent, good, fair, poor)

    5
        i. bar graph
        ii. boxplot
        iii. histogram
        iv. pie chart
        v. stem-and-leaf plot

2. The respiratory rate in healthy dogs follows a normal density with mean 22 breaths per minute and standard deviation 4 breaths per minute. What proportion of healthy dogs have a repiratory rate between 17 and 27 breaths per minute? Show your work.

5

$\sigma = 4$

17   22   27
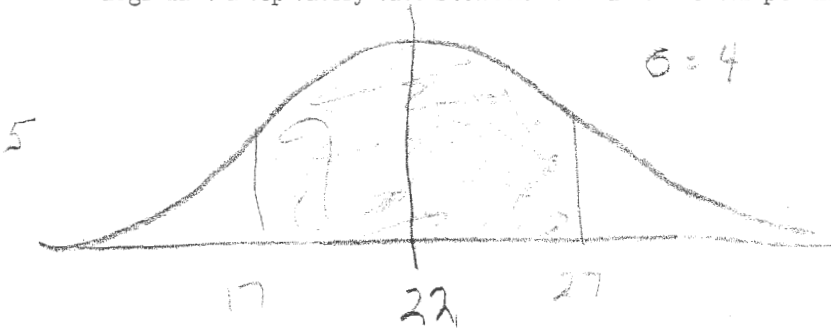
$$\frac{27-22}{4} = 1.25$$

$$\frac{17-22}{4} = -1.25 \qquad \text{area to left} = .1056$$

$$1.00 - 2(.1056) = .79$$

3. The Federal Highway Administration collects data on many variables related to motor fuel consumption. This question involves data reported for each state and the District of Columbia for the year 2001. The two variables are:

- mpc – the estimated miles driven per capita
- tax – the state gasoline tax rate in cents per gallon

We wish to examine the relationship between these two variables and to see whether tax can be used to explain or predict mpc

Refer to the attached SAS output in answering the following questions.

(a) What was the highest value of miles per capita (mpc)? (Numeric answer; tell which part of the SAS output you used to get it.)

3

$17.5 \times 10^3 = 17,500$ miles
from stem and leaf plot

(b) Circle all the true statements about the distribution of mpc.

5

  i. The distribution is skewed to the left.
  ii. The distribution is skewed to the right.
  iii. The distribution is roughly symmetric.
  iv. There are high outliers.
  v. There are low outliers.
  vi. None of the above.

(c) Give the five-number summary of tax. (Numeric answer; tell which part of the SAS output you used to get it.

5

$Q_3 = 23.5$
Min = 7.5   Max = 29.0
Q1 = 18
Median = 20   From Quantiles in proc univariate

(d) Based on the scatterplot, which of the following numbers is closest to the correlation coefficient, r? (Circle one).

2

  i. -0.8
  ii. -0.2
  iii. 0.2
  iv. 0.8

1

(e) In the scatterplot, which variable is the response variable, mpc or tax?

(f) Write the equation of the regression line relating mpc to tax.

$\widehat{mpc} = 12096 - 81.76 \, tax$

2½ formula *3*

(g) Explain what the numeric value of the slope means in terms of tax and miles per capita.

For each 1-cent increase in tax we would expect on average an

3

81.76-mile decrease in miles per capita.

2½ if imply causation

2½

(h) What is the predicted value of mpc for a state in which tax is 15 cents per gallon? (Numeric answer; show your work.)

3

$$12096 - 81.76(15) = 10870$$

(i) Was extrapolation involved in the prediction in the previous question? Explain briefly.

2

No. 15 cents is within the range of values of tax in the dataset.

(j) For the state of Iowa, tax = 20 cents and mpc = 10258.4 miles. What is the residual for the state of Iowa? Numeric answer; show your work.

$$12096 - 81.76(20) =$$

4

$$10258.4 - 10461 = -202$$

(k) Note that the units for the mpc variable are miles and for the tax variable are cents. What are the units for each of the following quantities?

4

  i. standard deviation of tax   cents
  ii. slope for the regression of mpc on tax   miles/cent
  iii. correlation between mpc and tax   no units
  iv. median of mpc   miles

(13)