

22S:30/105, Statistical Methods and Computing  
Spring 2008, Instructor: Cowles  
Midterm 1

Show your work on any problems that involve calculations.

Name: Solutions  
Course no. (30, 105, or 191) \_\_\_\_\_

1. A variable that can be measured on individual earthworms is their length in centimeters. This variable is (circle one):

- 2
- (a) nominal
  - (b) ordinal
  - (c) binary
  - (d) discrete quantitative
  - (e) continuous quantitative

Numeric summaries appropriate to describe this variable in a large sample of earthworms are (circle all that apply, and explain very briefly why you did or did not include each one):

2

(a) mean

(b) median

(c) mode

(d) frequency table

OK if distribution is roughly symmetric  
OK for both skewed & symmetric distributions  
} Not very useful for continuous data  
in a large sample with lots of different values.

2. One variable that describes medical doctors is their area of specialization. It can take on values such as pediatrics, cardiology, geriatrics, etc. This variable is (circle one):

2

(a) nominal

(b) ordinal

(c) binary

(d) discrete quantitative

(e) continuous quantitative

No ordering; just categories.

Numeric summaries appropriate to describe this variable as evaluated in a group of medical doctors are (circle all that apply):

2

(a) mean

(b) median

(c) mode

(d) frequency table

3. The National Oceanic and Atmospheric Administration reports the following annual numbers of deaths from tornadoes in the U.S. from 1990 through 2000:

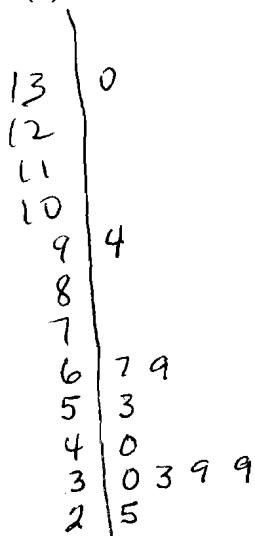
6 3 4 3 8 2 7 16 9 5  
53 39 39 33 69 30 25 67 130 94 40

(a) Calculate the median and the first and third quartiles of these values by hand. Show your work.

25 30  $Q_1$  33  $M$  39 39 40 53 67  $Q_3$  94 130

eleven values; median will be middle (6<sup>th</sup>)  
 $Q_1$  and  $Q_3$  are medians of values to left and to right respectively of median.

(b) Make a stem-and-leaf plot of these values.



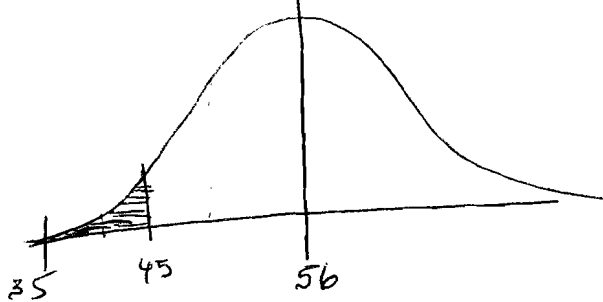
(c) The distribution of these values is (circle one):

- i. symmetric
- ii. right skewed
- iii. left skewed
- iv. impossible to tell from information provided

(d) The best numeric summary of these data would be (circle one):

- i. mode and range
- ii. mean and standard deviation
- iii. standard deviation and interquartile range
- iv. 5-number summary

4. A physical fitness test was given to all the students at a large high school. The distribution of the scores was approximately Normal, with mean 56 and standard deviation 7. What proportion of students scored between 35 and 45? Draw a picture representing this proportion, and show your work to produce a numeric answer.



$$z_1 = \frac{35-56}{7} = -3; \quad P(Z \leq -3) = .00135$$

$$z_2 = \frac{45-56}{7} = -1.57; \quad P(Z \leq -1.57) = .05804$$

$$P(-3 \leq Z \leq -1.57) = .05804 - .00135 = .0567$$

5. A student group wishes to survey restaurants in Iowa City and Coralville about the importance of students as customers. From telephone book listings, the committee chooses 20 restaurants at random. Of these, 9 return the questionnaire mailed by the committee. The population for this study is (circle one):

(a) all restaurants in Iowa City and Coralville

(b) the 20 restaurants chosen from the phone book

(c) the 9 restaurants that returned the survey

(d) none of the above

6. The "pretty" plot on the first page of the attached SAS output shows the U.S. population (in millions) for the years 1800 - 2000.

(a) What kind of plot is this? (just the name)

*line plot or time plot*

(b) Use the plot to determine the approximate population of the U.S. in 1900. (numeric answer)

*around 70 million*

7. The remaining pages of SAS output are based on a dataset of 130 healthy adults. The three variables are:

**gender** (coded as 1 = male and 2 = female)

**temp** (body temperature in degrees Fahrenheit)

**heart** (heart rate in beats per minute)

1 (a) According to the side-by-side boxplots, do the males in this sample have any outliers in temperature? (yes no)

1 (b) In the way the scatterplot and the regression analysis were done, which variable is the response variable? (temp heart)

(c) Does the output indicate a strong linear relationship between temp and heart? Justify your answer using both graphical and numeric evidence.

2 *No. The scatterplot shows that the points do not lie close to any straight line, and  $R^2$  is only 0.0643 (very far from 1).*

(d) What is the predicted value of heart for a person with temp of 98.6 degrees?

2 
$$-166.3 + 2.44(98.6) = 74.28$$