22S:30/105, Statistical Methods and Computing
Spring 2012, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

Name: Solutions

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

1. A study of computer-assisted learning examined the learning of "Blissymbols" by children. Blissymbols are pictographs that are sometimes used to help learning-impaired children communicate. The researcher designed two computer lessions that taught the same content using the same examples. One lesson required the children to interact with the material, while in the other, the children controlled only the pace of the lesson. Call these two styles "Active" and "Passive." After the lesson, the computer presented a quiz that asked the children to identify 56 Blissymbols. The dataset contains the numbers of correct identifications by 24 children in the Active group and 24 children in the Passive group. The variables are:

```
group -- "A" for Active, "P" for Passive
score -- number of correct identifications
```

Refer to the attached SAS output to answer the following questions:

(a) What is the range of the score variable in the "A" group? (numeric answer; show your work)

44 - 15 = 29

(b) The range is (circle one):

i. a robust measure of center
ii. a sensitive measure of center
iii. a robust measure of spread
iv. a sensitive measure of spread
v. none of the above

(c) Is the range a good numeric summary to use with the score variable in the "A" group? Briefly state why or why not.

No. The range is sensitive to outliers. The high outlier in the A group will inflate the range.

(d) Which group – "A" or "P" – has the higher 3rd quartile? How could you tell from the SAS output?

A. The top of its box in the boxplot is higher than that for group B.

(e) Was this study an experiment or an observational study? Briefly explain.

Experiment. The researcher controlled which kind of lesson each child received.

2. Forced expiratory volume in 1 second (FEV1) is a commonly-used measure of lung function. It can be adjusted for body size by dividing by height squared. $\frac{FEV_1}{ht^2}$ has a roughly linear relationship with age. A journal article reports the following estimated regression line for adult men:

$$\widehat{\frac{FEV_1}{ht^2}} = 1.67 - 0.0131\ age$$

where age is measured in years.

(a) Which is the response variable (circle one):

i. age
ii. $\frac{FEV_1}{ht^2}$

(b) Harry and Fred are adult men. Harry is 5 years older than Fred. Would you expect Harry's $\frac{FEV_1}{ht^2}$ to be higher or lower than Fred's? By how much? (show your work for the numeric part of the answer)

I would expect Harry to be lower (since slope is negative) by about $5(.0131) = .0655$ units

(c) The correlation between $\frac{FEV_1}{ht^2}$ and age is (circle one):

i. positive
ii. negative
iii. exactly 0
iv. no way to tell from the information given

same sign as regression slope

(d) The article also discusses the fact that $\frac{FEV_1}{ht^2}$ on average is lower in women than in men. Would it be correct therefore to say that there is a correlation between $\frac{FEV_1}{ht^2}$ and gender? Why or why not?

No. Correlation is the strength of the linear association between 2 quantitative variables. Gender is binary.

3. A biology professor wishes to investigate whether students' scores on midterm exams are affected by whether they write the exams in pen or pencil. Before each midterm exam, she randomly assigns half of the students to use pencils on the exam and the other half of the students to use pen. She records the exam score for each student, along with whether the exam was written in pen or pencil.

(a) The factor(s) in this study is or are (circle one):

i. the individual students
ii. writing implement
iii. pen or pencil
iv. the exam scores
v. none of the above

(b) The treatments(s) in this study is or are (circle one):

i. the individual students

ii. writing implement
iii. pen or pencil
iv. the exam scores
v. none of the above

(c) The experimental unit(s) in this study is or are (circle one):
i. the individual students
ii. writing implement
iii. pen or pencil
iv. the exam scores
v. none of the above

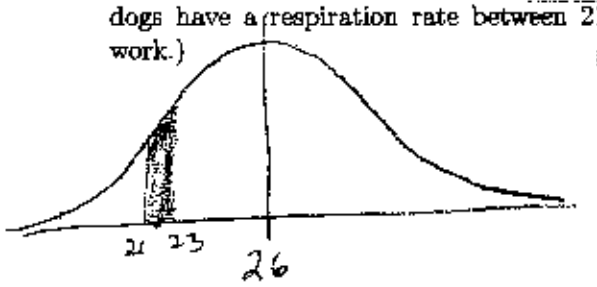
4. What is the data type of each of the following variables (circle one for each):

(a) responses to a survey question that asks: "Do you believe that the U.S. economy is (a) getting better, (b) staying about the same, or (c) getting worse?"
i. Binary
ii. Nominal
iii. Ordinal
iv. Discrete quantitative
v. Continuous quantitative

(b) responses to a survey question that asks: "Did you vote in the 2010 elections? (yes or no)"
i. Binary
ii. Nominal
iii. Ordinal
iv. Discrete quantitative
v. Continuous quantitative

(c) the body weights of adult male lions
i. Binary
ii. Nominal
iii. Ordinal
iv. Discrete quantitative
v. Continuous quantitative

5. The respiratory rate in healthy dogs follows a normal density with mean 26 breaths per minute and standard deviation 4 breaths per minute. What proportion of healthy dogs have a respiration rate between 21 and 23 breaths per minute? (Show your work.)



$Pr\left(Z < \frac{21-26}{4}\right) =$

$Pr(Z < -1.25) = .1056$

$Pr\left(Z < \frac{23-26}{4}\right) =$

$Pr(Z < -0.75) = 0.2266$

$.2266 - .1056 = \boxed{0.121}$

6. Draw a stem and leaf plot of the following values:

121 121 125 130 132 143 146 148 148 151 159 171

17 1

16

15 1 9

14 3 6 8 8

13 0 2

12 1 5

multiply stem. leaf by 10