PRACTICE PROBLEMS for MIDTERM 1, 2006
22S:30/105, Statistical Methods and Computing
Spring 2005, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

There are 50 total points on this midterm. Point values for each question are shown in parentheses. I will grade on a curve.

Name: _____ Course no. (30 or 105) _____

1. (6) The faculty in the department of Linguistics at the UI are:

```
Jill Beckman
Maureen Burke
Rob Chametzky
William Davies
Alice Davison
Elena Gavruseva
Marc Light
Rosemary Plapp
Catherine Ringen
Jerzy Rubach
Roumyana Slabakova
Bob Wachal
```

```
62964 88145 83083 69453 46109 59505 69680 00900 19687 12633
```

Use the list of random digits above to choose a simple random sample of three of these people to serve on a committee. Write enough on the list of names, and make markings on the list of random digits, so that I can tell what procedure you used. Write the names of the three people you selected here:

_____   _____   _____

2. The dataset for this problem includes distances and cheapest airline fares to certain destinations for passengers flying out of Baltimore, MD (as of 1/8/1995). The variables are:

```
dest:  destination
dist:  distance from Baltimore (in miles)
fare:  fare (in dollars)
```

(a) (4) Refer to the stem and leaf plot for the dist variable (below). What is the median of the distribution of values?

```
Stem Leaf                   #
 14 0                       1
 12 2                       1
 10 0                       1
  8 5                       1
  6 149                     3
  4 18                      2
  2 17                      2
  0 9                       1
    ----+----+----+----+
  Multiply Stem.Leaf by 10**+2
```

(b) Refer to the SAS output attached to the end of the exam to answer the following questions.

i. (1) Which is the response variable, dist or fare?

ii. (2) What proportion of the variability in the response variable is explained by the explanatory variable? (numeric answer)

iii. (2) Suppose that city A is 100 miles farther from Baltimore than city B. How much higher or lower a fare would the regression model predict for city A than for city B?

iv. (2) Two points are plotted as circles on the scatterplot. If these points were removed, would the sample correlation coefficient r be more likely to get larger or get smaller? Briefly justify your answer. vskip 1.0 in

v. (2) One point on the scatterplot is plotted as an "x." Is this point likely to be influential? Briefly justify your answer.

3. (5) In 2002, the scores of the 1.3 million students who took the Scholastic Aptitude Test (SAT) could be described by a Normal distribution with mean 1020 and standard deviation 207. What proportion of students scored between 1000 and 1400?

4. (2) At a political gathering, there are 25 people over age 50 and 15 people under age 50. You choose at random 5 of those over age 50 and separately choose at random 3 of those under age 50 to interview about attitudes toward Social Security reform. The sample of 8 people that you obtain is a (circle one):

   (a) convenience sample
   (b) judgment sample
   (c) simple random sample
   (d) stratified random sample
   (e) none of the above

5. Researchers wish to investigate the relationship between the number of hours of sleep that people living in the Iowa City/Coralville area get each night and the number of colds they get each year. The researchers randomly select 100 numbers from the Iowa City/Coralville telephone directory. They call these numbers and ask the following questions of the person who answers the telephone:

   a. How old are you?
   b. How many hours did you sleep last night?
   c. How would you describe your general health? (excellent, good, fair, poor)
   d. How many colds did you have in the past year?

   (a) (2) This research study is (circle one):
       i. an experiment
       ii. an observational study

iii. none of the above

   (b) (2) The population of interest is (circle one):
       i. all people living in the Iowa City/Coralville area
       ii. all people listed in the Iowa City/Coralville phone directory
       iii. the people living in the homes whose telephone numbers are selected
       iv. the people who actually answer the questions
       v. none of the above

   (c) (2) What data type is the variable "number of colds in the past year" from question d? (circle one)
       i. binary
       ii. nominal
       iii. ordinal
       iv. discrete quantitative
       v. continuous quantitative
       vi. none of the above

   (d) (2) What data type is the variable "general health" from question c? (circle one)
       i. binary
       ii. nominal
       iii. ordinal
       iv. discrete quantitative
       v. continuous quantitative
       vi. none of the above

   (e) (2) Which plot or plots from the list below could be used to summarize the distribution of the variable "general health" from question c? (circle as many as are correct)
       i. bar graph
       ii. boxplot
       iii. histogram
       iv. line plot
       v. pie chart
       vi. scatterplot
       vii. stem and leaf plot
       viii. none of the above

6. Consider the variable age in undergraduate students at the university of Iowa.

   (a) (1) The distribution of ages of undergraduate students is most likely to be (circle one)
       i. skewed to the left
       ii. skewed to the right
       iii. approximately symmetric

   (b) (2) Briefly justify your answer to the preceding question.

(c) (1) From the list below, circle the best choice of numeric summary of this variable.

    i. frequency table

    ii. mean and standard deviation

    iii. 5-number summary

(d) (2) Briefly justify your answer to the preceding question.

```
                        The REG Procedure
                          Model: MODEL1
                     Dependent Variable: fare

       Number of Observations Read          12
       Number of Observations Used          12


                      Analysis of Variance

                             Sum of         Mean
Source                DF    Squares       Square  F Value  Pr > F

Model                  1      24735        24735    17.59  0.0018
Error                 10      14060   1406.02467
Corrected Total       11      38795


        Root MSE              37.49700    R-Square     0.6376
        Dependent Mean       166.50000    Adj R-Sq     0.6013
        Coeff Var             22.52072


                      Parameter Estimates

                   Parameter      Standard
Variable      DF    Estimate         Error    t Value    Pr > |t|

Intercept      1    82.57767      22.74905       3.63      0.0046
dist           1     0.11776       0.02808       4.19      0.0018
```

```
                          Model: MODEL1
                     Dependent Variable: fare


         ---+---------+---------+---------+---------+---------+---------+---
     300 +                                                                 +    +
         |                                                                 |    |
         |                                              .                  |    |
         |                                                            x    |    |
     250 +                                                                 +    +
         |                                                                 |    |
   fare  |                                                                 |    |
         |                                                                 |    |
         |                                                                 |    |
     200 +                                         .                       +    +
         |                                       .                         |    |
         |                                  .                              |    |
         |                          .                                      |    |
         |                   .                                             |    |
     150 +                                                                 +    +
         |           .       .                                             |    |
         |                                                                 |    |
         |                                                                 |    |
         |                                                                 |    |
     100 +           .               o                                     +    +
         |                   o                                             |    |
         |                                                                 |    |
         |                                                                 |    |
         |                                                                 |    |
      50 +                                                                 +    +
         ---+---------+---------+---------+---------+---------+---------+---
            0        250       500       750      1000      1250      1500


                                      dist
```

7