22S:30/105, Statistical Methods and Computing
Spring 2011, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

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

1. What is the data type of each of the following variables? Circle one choice for each.

    (a) types of electronic devices manufactured by a company (cell phone, netbook, tablet pc, etc.)
        i. Binary
        ii. Nominal
        iii. Ordinal
        iv. Discrete quantitative
        v. Continuous quantitative

    (b) the ratings awarded to high school cheerleading squads in a state competition (Superior, Excellent, Very Good, Fair)
        i. Binary
        ii. Nominal
        iii. Ordinal
        iv. Discrete quantitative
        v. Continuous quantitative

    (c) the number of dogs housed by the Iowa City animal shelter each year
        i. Binary
        ii. Nominal
        iii. Ordinal
        iv. Discrete quantitative
        v. Continuous quantitative

2. In lectures at the beginning of the semester, we used a dataset containing nutritional information on different kinds of cereals. Two of the variables in the dataset are:

    - potass: milligrams of potassium per serving
    - fiber: grams of fiber per serving

    Refer to the SAS output provided in answering the following questions about these data.

    (a) The distribution of the fiber variable is (circle one):
        i. right skewed
        ii. left skewed
        iii. roughly symmetric

    iv. no way to tell from output provided

    (b) Give the range of the fiber variable (numeric answer). Show your work, and tell what part of the SAS output you used to get it.

    (c) Would the mean and standard deviation provide a good numeric summary of the fiber variable? (yes/no) Why or why not?

    (d) Based on the scatterplot of fiber and potass, the value of the sample correlation coefficient r is likely to be closest to (circle one):
        - 2.0
        - 0.9
        - 0.5
        - 0.1
        - -0.5
        - -1.0

    (e) Use the regression equation to predict the number of milligrams of potassium in a serving of cereal with 1.5 grams of fiber. (numeric answer; show your work).

    (f) What proportion of the variability in the potass variable is explained by the fiber variable? (numeric answer based on SAS output)

    (g) Is there anything in the SAS output that makes you think that the correlation coefficient might not be the best measure of association between fiber and potassium? Explain briefly.

3. A television news show reports that scientists have found a strong correlation between whether people live in Canada or the U.S. and their degree of satisfaction with their children's schools. Briefly explain the misuse of a statistical term in this statement.

4. The respiratory rate in healthy dogs follows a normal density with mean 26 breaths per minute and standard deviation 4 breaths per minute. Find the number of breaths per minute such that only 5% of dogs have a respiratory rate below it. (numeric answer; show your work)

5. Marketing researchers wish to determine what proportion of U.S. fathers do the food shopping for their families. They randomly draw 2000 telephone numbers from telephone lists in 16 major cities. From calling these households, the researches are able to get 1216 fathers who answer the question "Do you do the food shopping for your family?"

   (a) The population of interest is (circle one):
       i. all U.S. fathers
       ii. the proportion of U.S. fathers who do the food shopping for their families
       iii. the telephone lists in 16 major cities
       iv. the 2000 households drawn from the telephone lists
       v. the 1216 fathers who responded to the survey
       vi. none of the above
   (b) The sampling frame is (circle one):
       i. all U.S. fathers
       ii. the proportion of U.S. fathers who do the food shopping for their families
       iii. the telephone lists in 16 major cities
       iv. the 2000 households drawn from the telephone lists
       v. the 1216 fathers who responded to the survey
       vi. none of the above

6. There are 17 tenured and tenure-track faculty in the Statistics department at UI, 12 men and 5 women. The department chair needs to pick a random sample of 3 faculty members to serve on a committee. He wants to have 2 men and 1 woman on the committee.

   (a) Simple random sampling is considered the "gold standard" sampling procedure. Should the department chair use it to choose his committee members? (yes/no) Why or why not?

(b) Use your random digit table beginning at line 136 to randomly select 2 male faculty and 1 female for the committee from the list below. Write enough below to show how you arrived at your answer.

```
Men
      Chan
      Dykstra
      Huang
      Lang
      Lenth
      Pansera
      Russo
      Shiu
      Shyamalkumar
      Tang
      Tierney
      Zimmerman

Women
      Cowles
      DeCook
      Ghosh
      Stramer
      Tan
```

(c) The kind of sampling that you used to draw the committee members is called (circle one)
    i. biased sampling
    ii. convenience sampling
    iii. judgment sampling
    iv. simple random sampling
    v. stratified random sampling
    vi. volunteer response sampling
    vii. none of the above

```
                        Fiber


Stem Leaf                      #        Boxplot
  14 0                         1           *
  13
  12
  11
  10 0                         1           *
   9 0                         1           0
   8
   7
   6 0                         1           |
   5 0000                      4           |
   4 0000                      4           |
   3 000000000000000          15        +-----+
   2 000000000057             12        *--+--*
   1 0000000000000000555      19        +-----+
   0 0000000000000000000      19           |
     ----+----+----+----+
```

```
                    Plot of potass*fiber.  Symbol used is '.'.

    potass |
           |
           |
       375 +
           |
           |
           |                                        .
           |                             .
       300 +
           |                                   .
           |
           |                        .
           |                     .
       225 +           .
           |                 .
           |                  .
           |
           |           . .
       150 +
           |       .  . .  .
           |         ..   . .
           |           .  ..
           |.        ..   .
        75 +      ..
           |.    .
           |.    .
           |.    .
           |.    .
           |.    .
         0 +
           |
           -+-----------+-----------+-----------+
            0           5          10          15

                              fiber
```
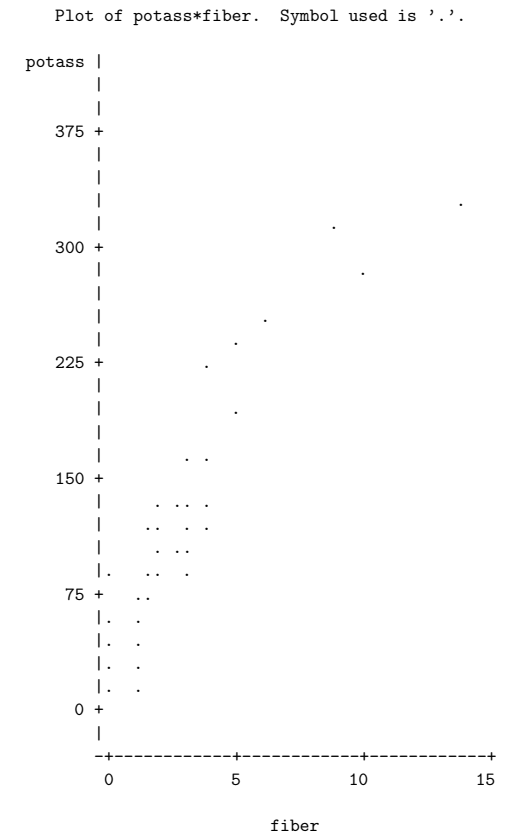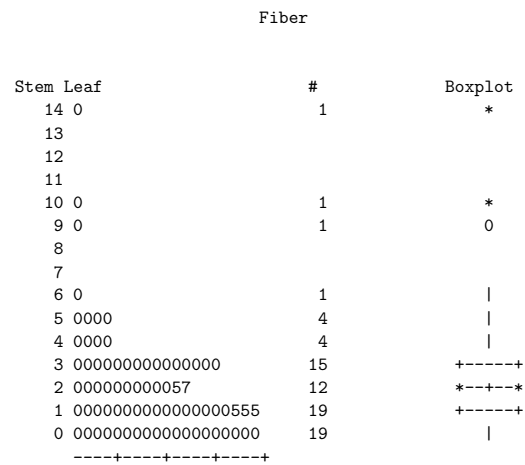
```
                          The REG Procedure
                           Model: MODEL1
                     Dependent Variable: potass


     Number of Observations Read                    77
     Number of Observations Used                    75
     Number of Observations with Missing Values      2


                       Analysis of Variance

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

     Model               1      304823      304823   358.65  <.0001
     Error              73       62043   849.91078
     Corrected Total    74      366867


          Root MSE              29.15323   R-Square    0.8309
          Dependent Mean        98.66667   Adj R-Sq    0.8286
          Coeff Var             29.54719

                       Parameter Estimates

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

     Intercept     1     40.48460       4.55748       8.88    <.0001
     fiber         1     26.65641       1.40755      18.94    <.0001
```