22S:30/105, Statistical Methods and Computing
Spring 2013, 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 for each):

   (a) systolic blood pressure in women ages 65-74

      i. Binary
      ii. Nominal
      iii. Ordinal
      iv. Discrete quantitative
      v. Continuous quantitative

   (b) the number of auto thefts in Wyoming in each year from 1901-2000

      i. Binary
      ii. Nominal
      iii. Ordinal
      iv. Discrete quantitative
      v. Continuous quantitative

   (c) the ratings of high school choirs in a statewide competition (superior, excellent, good, fair, poor)

      i. Binary
      ii. Nominal
      iii. Ordinal
      iv. Discrete quantitative
      v. Continuous quantitative

2. For each of the following variables, which shape would you expect its distribution to have? (circle one answer for each)

   (a) lengths of oak leaves

      i. roughly symmetric
      ii. right skewed
      iii. left skewed

   (b) the amount of money spent on clothing in 2012 by each woman in Iowa City

      i. roughly symmetric
      ii. right skewed
      iii. left skewed

3. The respiratory rate in healthy dogs follows a normal density with mean 22 breaths per minute and standard deviation 6 breaths per minute.

   The body temperature in healthy dogs follows a normal density with mean 101.5 degrees Fahrenheit and standard deviation 0.5 degrees F.

   If my dog Donny has a respiratory rate of 12 breaths per minute and a body temperature of 100 degrees F, is he more unusual with regard to respiratory rate or temperature? Justify your answer with appropriate numeric calculations.

4. A dataset on predictors of low birthweights in infants is included in the textbook Hosmer and Lemeshow (2000) *Applied Logistic Regression: Second Edition.* These data are copyrighted by John Wiley & Sons Inc. The observations in the dataset are mother-ifant pairs. Three of the variables are:

```
bwt          --   birthweight of the infant in grams
age          --   age of the mother in years at time of giving birth
smoke        --   1 if the mother was a smoker; 0 otherwise
```

   Refer to the attached SAS output in answering the following questions about this dataset.

   (a) Based on the scatterplot of `bwt` versus `age`, what is your best guess of the correlation between these two variables? (circle one)

      i. -0.67
      ii. -0.33
      iii. 0.00
      iv. 0.33
      v. 0.67

   (b) If `bwt` were measured in pounds instead of grams, would the sample correlation coefficient r between `bwt` and `age` change? Explain briefly.

   (c) On the scatterplot, circle any points that would be influential if we fit a linear regression model to these data.

   (d) If `bwt` were measured in pounds instead of grams, would the value of the regression slope b change? Explain briefly.

(e) What is the lowest birthweight in the group in which mothers were smokers (`smoke` = 1)? Give a numeric answer and explain how you got it.

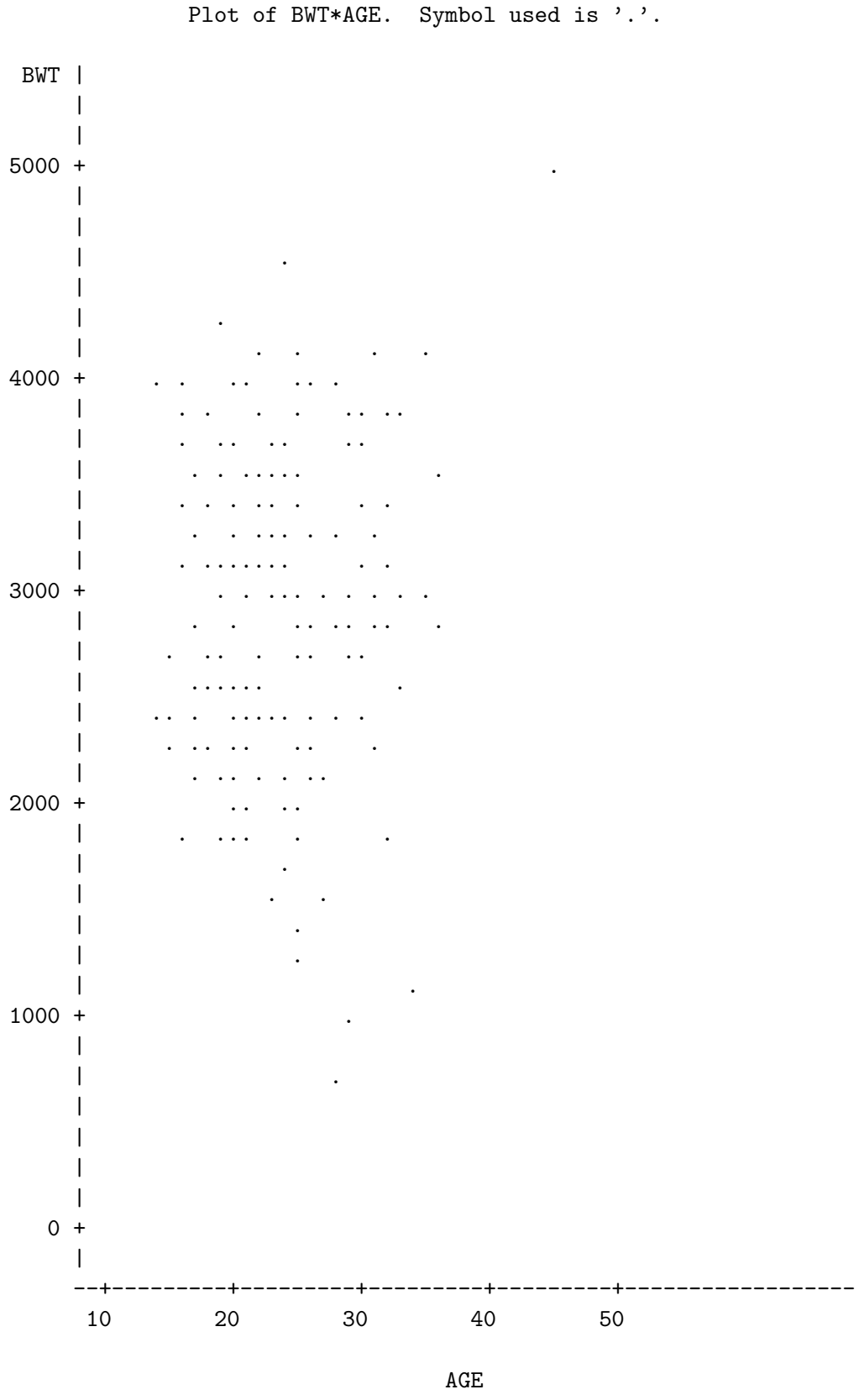
(f) The distribution of birthweights in the group in which mothers were nonsmokers (`smoke` = 0) is (circle one):

i. roughly symmetric
ii. skewed right
iii. skewed left
iv. multimodal
v. no information is given in the SAS output to answer this

(g) Would the mean and standard deviation be good numeric summaries to report for the birthweight variable `bwt` in each of the groups defined by `smoke` = 0 and `smoke` = 1? (Yes or no; briefly explain your answer.)

(h) In this dataset, are birthweights generally higher for nonsmoking mothers or smoking mothers? Justify your answer referring to SAS output.

(i) Does the difference in birthweights between smoking and nonsmoking mothers mean that there is a correlation between mother's smoking status and infant's birthweight? (Yes or no; briefly explain your answer.)


Plot of BWT*AGE. Symbol used is '.'.
BWT
5000
4000
3000
2000
1000
0
10
20
30
40
50
AGE

```
------------------------------ SMOKE=0 --------------------------------
            Stem Leaf                     #             Boxplot
              48 9                        1                |
              46                                           |
              44 9                        1                |
              42                                           |
              40 00551577                 8                |
              38 0366814478              10                |
              36 011355037779            12             +-----+
              34 0266789479              10             |     |
              32 002233377227            12             |     |
              30 6668990088              10             *--+--*
              28 144468822288            12             |     |
              26 24423558                 8             |     |
              24 04450025                 8             +-----+
              22 4480258                  7                |
              20 66809                    5                |
              18 9037                     4                |
              16 03                       2                |
              14 799                      3                |
              12 3                        1                |
              10 2                        1                |
                 ----+----+----+----+
             Multiply Stem.Leaf by 10**+2

------------------------------ SMOKE=1 --------------------------------
            Stem Leaf                     #             Boxplot
              42 4                        1                |
              40                                           |
              38 684                      3                |
              36 344556                   6                |
              34 347                      3                |
              32 0602237                  7             +-----+
              30 03488935                 8             |     |
              28 21255889                 8             |     |
              26 066778                   6             *--+--*
              24 11127770069             11             |     |
              22 12005788                 8             +-----+
              20 88239                    5                |
              18 28334                    5                |
              16 9                        1                |
              14                                           |
              12                                           |
              10 4                        1                |
               8
               6 1                        1                0
                 ----+----+----+----+
             Multiply Stem.Leaf by 10**+2
```

The UNIVARIATE Procedure
Variable: BWT

Schematic Plots