22S:30/105, Statistical Methods and Computing
Spring 2014, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

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

1. What type of plot would be appropriate to represent each of the following variables? (Circle **all** that apply for each.)

   (a) body temperatures in women ages 65-74

      i. bar graph
      ii. boxplot
      iii. histogram
      iv. pie chart
      v. stem-and-leaf plot

   (b) the ratings of restaurants in a local newspaper (superior, excellent, good, fair, poor)

      i. bar graph
      ii. boxplot
      iii. histogram
      iv. pie chart
      v. stem-and-leaf plot

2. The respiratory rate in healthy dogs follows a normal density with mean 22 breaths per minute and standard deviation 4 breaths per minute. What proportion of healthy dogs have a repiratory rate between 17 and 27 breaths per minute? Show your work.

3. The Federal Highway Administration collects data on many variables related to motor fuel consumption. This question involves data reported for each state and the District of Columbia for the year 2001. The two variables are:

   - `mpc` – the estimated miles driven per capita
   - `tax` – the state gasoline tax rate in cents per gallon

   We wish to examine the relationship between these two variables and to see whether `tax` can be used to explain or predict `mpc`

   Refer to the attached SAS output in answering the following questions.

   (a) What was the highest value of miles per capita (`mpc`)? (Numeric answer; tell which part of the SAS output you used to get it.)

   (b) Circle all the true statements about the distribution of `mpc`.

   i. The distribution is skewed to the left.
   ii. The distribution is skewed to the right.
   iii. The distribution is roughly symmetric.
   iv. There are high outliers.
   v. There are low outliers.
   vi. None of the above.

   (c) Give the five-number summary of `tax`. (Numeric answer; tell which part of the SAS output you used to get it.

   (d) Based on the scatterplot, which of the following numbers is closest to the correlation coefficient, r? (Circle one).

   i. -0.8
   ii. -0.2
   iii. 0.2
   iv. 0.8

   (e) In the scatterplot, which variable is the response variable, `mpc` or `tax`?

   (f) Write the equation of the regression line relating `mpc` to `tax`.

   (g) Explain what the numeric value of the slope means in terms of tax and miles per capita.

(h) What is the predicted value of `mpc` for a state in which `tax` is 15 cents per gallon? (Numeric answer; show your work.)

(i) Was extrapolation involved in the prediction in the previous question? Explain briefly.

(j) For the state of Iowa, `tax` = 20 cents and `mpc` = 10258.4 miles. What is the residual for the state of Iowa? Numeric answer; show your work.

(k) Note that the units for the `mpc` variable are miles and for the `tax` variable are cents. What are the units for each of the following quantities?

i. standard deviation of `tax`

ii. slope for the regression of `mpc` on `tax`

iii. correlation between `mpc` and `tax`

iv. median of `mpc`

Selected Proc univariate output for `tax`

The UNIVARIATE Procedure
Variable: tax

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	31.67066	Pr > |t|	<.0001
Sign	M	25.5	Pr >= |M|	<.0001
Signed Rank	S	663	Pr >= |S|	<.0001

The UNIVARIATE Procedure
Variable: tax

Quantiles (Definition 5)

Quantile	Estimate
100% Max	29.0
99%	29.0
95%	27.0
90%	25.0
75% Q3	23.5
50% Median	20.0
25% Q1	18.0
10%	15.0
5%	10.5
1%	7.5
0% Min	7.5

Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
7.5	11	25.65	49
8.0	2	26.00	39
10.5	31	27.00	27
13.6	10	27.30	50
14.0	51	29.00	40

Plots for variable mpc

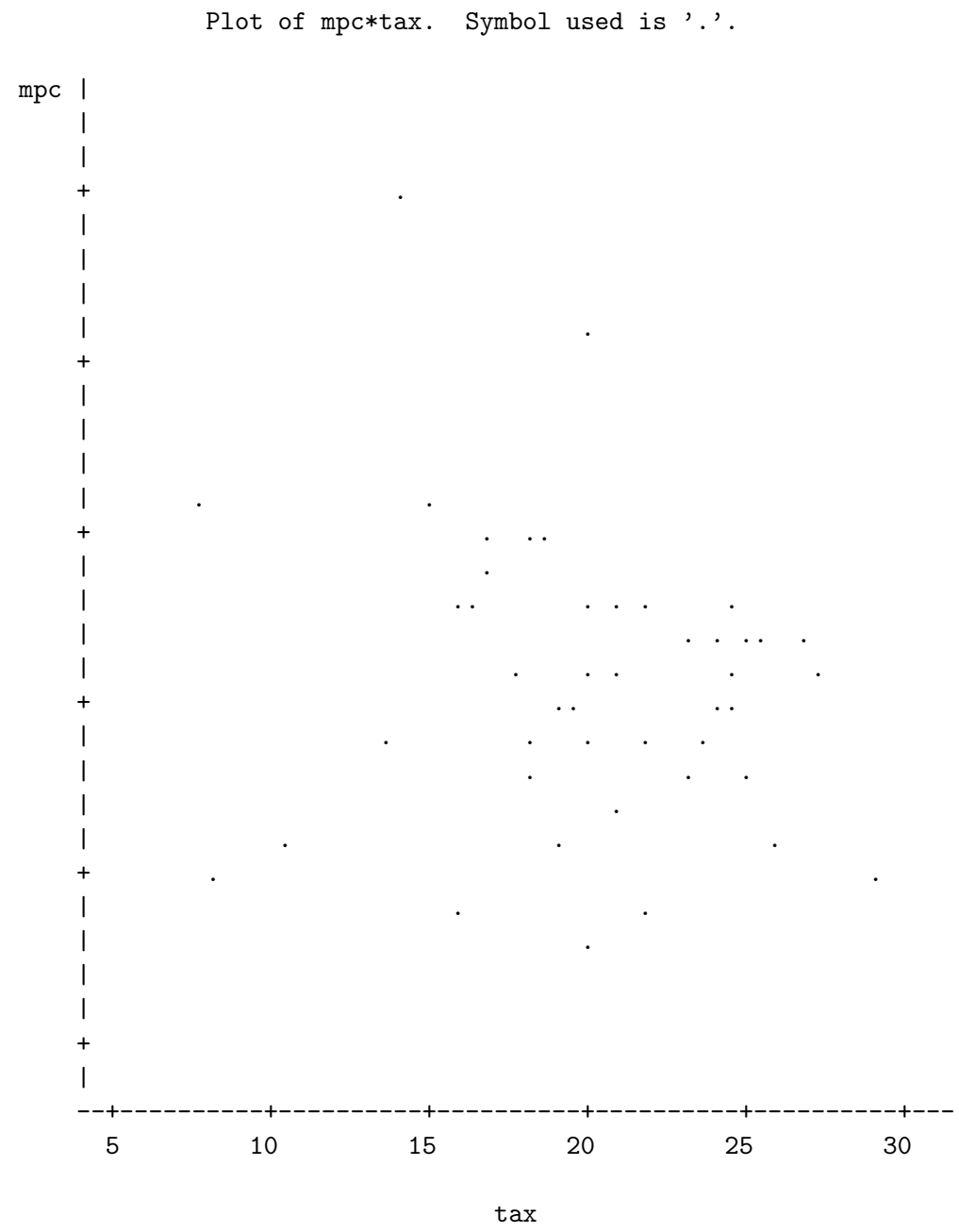
The UNIVARIATE Procedure
Variable: mpc

Stem	Leaf	#	Boxplot
17	5	1	*
17			
16			
16			
15	7	1	0
15			
14			
14			
13			
13	2	1	|
12	6679	4	|
12	03	2	|
11	56	2	|
11	02333333	8	+-----+
10	5677999	7	*-----*
10	0333	4	| + |
9	5577899	7	| |
9	0044	4	+-----+
8	9	1	|
8	1123	4	|
7	66	2	|
7	1	1	|
6	69	2	|

----+----+----+----+

Multiply Stem.Leaf by 10**+3

Scatterplot



The REG Procedure
Model: MODEL1
Dependent Variable: mpc

Number of Observations Read	51
Number of Observations Used	51

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6902740	6902740	1.68	0.2007
Error	49	201095182	4103983		
Corrected Total	50	207997922			

Root MSE	2025.82904	R-Square	0.0332
Dependent Mean	10448	Adj R-Sq	0.0135
Coeff Var	19.38911		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12096	1301.82684	9.29	<.0001
tax	1	-81.75552	63.03897	-1.30	0.2007