Williams College
Department of Mathematics and Statistics

Stat 101
Midterm Exam I

Professors R. D. De Veaux
Due Thursday, October 16, 2008 in class

Your Name _____

There are 100 points on this exam. **Please show all your work and write your answers on this exam. No credit will be given for correct answers without correct justification.** You may use your book, anything on our Blackboard site, your notes, but not other sources on the web or any other assistance.

1. (20 points) Our survey before class started (Survey.jmp) asked several questions including:

   Sex (M or F)

   Politics: Rate yourself on the following political scale:

   1. Communist/Socialist
   2. Very Liberal
   3. Moderately Liberal
   4. Conservative Democrat
   5. Independent/Middle of Road
   6. Liberal Republican
   7. Moderatly Conservative
   8. Extremely Conservative
   9. Libertarian/Extreme Right
   10. Other (please specify)

   Height: (in inches)

   Describe briefly, using appropriate summaries and graphical displays, the differences (if any) between responses on *Politics* and *Height* between men and women.

2. (20 points) Alexandra and Scyrine are two forwards on the Watsamatta U. women's basketball team. They play one game a week and both have played in all 10 weeks of the regular season. Here's a list of their *points scored* in each week:

Week	Alexandra	Scyrine
1	2	17
2	12	15
3	5	14
4	2	16
5	18	18
6	28	15
7	13	17
8	24	15
9	25	15
10	30	17

Some summary statistics follow:

	Min	Q1	Med	Q3	Max
Alexandra	2	4.25	15.5	25.75	30
Scyrine	14	15	15.5	17	18

a. (15 points) Using whatever displays and summary statistics you think appropriate, discuss the differences and similarities between the two players. What are the advantages and disadvantages of each player?

Problem 2 continued

b. (5 points) The regular season is over and now the coach has to decide on *one* of the two players to take to the playoffs. What would you do in the coach's place? Explain briefly. Use graphical displays and data to back up your decision.

3. (60 points) In the summer of 2005 Tour De France, Lance Armstrong averaged 41.65 kilometers per hour (kph) for the entire course for the fastest average time in tour history. On the web site you'll find data (Tour.jmp) on the winning speed and other information for all the races between 1903 and 2007.

For the rest of this problem we will use *only the data from 1930 to 2007*. You can use the **Rows → Exclude/Unexclude** switch to exclude any rows.

a) (6 points) Describe the distribution of Speeds in this time period (1930-2007). Use whatever summary statistics and graphical displays you think are appropriate.

b) (6 points) Describe the relationship of *Speed* by *Year* (again only from 1930 to 2007) being careful to point out any unusual features in the plot.

Problem 3 continued

c) (12 points) Fit a regression to *Speed* by *Year*. What is the equation? Do the assumptions and conditions for regression appear to be met? (Include in this discussion any time periods for which the model seems to work especially well or poorly).

d) (3 points) What percentage of the total variation in Speed is accounted for by Year?

e) (4 points) What is the predicted speed for the year 2020?

f) (4 points) How much confidence do you have in the prediction you just made? Explain.

Problem 3 continued

g) (4 points) If the residual speed in the year 2008 turns out to be -1.02 kph, what was the winner's actual speed?

h) (4 points) Suppose that next July, the temperature is 105 degrees F and the winner of the Tour in 2009 averages a *speed* of only 30 kph. If we fit a new regression of *speed* on *year*, including this point, will the R^2 value increase or decrease? Explain briefly.

i) (2 points) From the regression output, find the SD of the residuals.

j) (3 points) Remembering that there are 71 residuals, about how many would we expect to be greater than 2.81 kph? What did you assume to answer this?

Problem 3 continued

k) (4 points) In 1958 Charly Gaul of Luxembourg averaged 36.9 kph, while in 1979 Bernard Hinault averaged 39.8 kph. Which of the two performances is more "remarkable"? Explain briefly.

l) (3 points) What was unusual about the 2007 race?

m) (3 points) If we standardized both variables, what would the regression of *Standardized Speed* on *Standardized Year* be?

n) (3 points) Suppose we reversed the regression and regressed *Standardized Year* on *Standardized Speed.* What would the regression equation be now?