1. [25 pts] On the first exam you looked at the alcohol content of beer (\%). Consider an indicator variable, Type, that equals 0 if the beer is a light beer and 1 if the beer is a regular beer. Refer to the JMP output “Alcohol Content of Beer.”
   a) [4] Describe the distribution of % Alcohol. Be sure to comment on shape, center and spread.
   b) [5] Compute a 95% prediction interval for the % Alcohol of the next beer selected at random from the population.
   c) [2] Give the equation for predicting % Alcohol from Type.
d) [2] How much of the variation in % Alcohol is explained by the linear relationship with Type?


g) [5] Is there a statistically significant difference between light beers and regular beers in terms of mean % Alcohol? Support your answer with an appropriate statistical analysis.
2. [40 pts] Data on one year old GM vehicles from the 2005 model year are given in the November 2008 issue of *Journal of Statistics Education*. A random sample of 100 vehicles (Buicks, Cadillacs, Chevrolets, Pontiacs, Satrums and SAABs) was selected. Using these data we wish to predict the Price ($1000) of the vehicles. The explanatory variables are listed below. Refer to the JMP output entitled “Analysis of the Price of Used GM Vehicles”.

- Mileage – number of miles the vehicle has been driven
- Cylinder – the number of cylinders in the engine
- Liter – the size of the engine in liters
- Doors – the number of doors
- Cruise = 1 if vehicle has cruise control, = 0 otherwise
- Sound = 1 if vehicle has upgraded sound system, = 0 otherwise
- Leather = 1 if vehicle has leather interior, = 0 otherwise
- Sedan = 1 if vehicle is a sedan, = 0 otherwise
- Buick = 1 if vehicle is a Buick, = 0 otherwise
- Cadillac = 1 if vehicle is a Cadillac, = 0 otherwise
- Chevrolet = 1 if vehicle is a Chevrolet, = 0 otherwise
- Pontiac = 1 if vehicle is a Pontiac, = 0 otherwise
- Saturn = 1 if vehicle is a Saturn, = 0 otherwise

Note: If a vehicle is not a Buick nor a Cadillac nor a Chevrolet nor a Pontiac nor a Saturn it is a SAAB.

a) [7] Consider the Forward selection process with the 13 explanatory variables listed above.
   i. [4] What will be the first variable entered into the model? How do you know this?
   ii. [3] What will be the value of $R^2$ once this first variable is entered?

b) [5] At the end of the Forward procedure eleven variables have been entered.
   i. [2] What is the value of $R^2$ for this eleven variable model?

c) [7] Consider the Backward selection process with the 13 explanatory variables listed above.
   i. [4] What will be the first variable removed from the model? How do you know this?

ii. [3] What will be the value of $R^2$ once this first variable is removed?

d) [5] At the end of the Backward procedure four variables have been removed.
   i. [2] What is the value of $R^2$ for the resulting nine variable model?

e) [8] After six steps of the Mixed procedure, six variables have been entered.
   i. [2] What are the six variables that have been entered?

   ii. [6] What will happen at the next step of the Mixed procedure? Be sure to support your answer by referring to the JMP output.


   i. [3] What is the value of $C_p$ for this nine variable model?

   ii. [2] What is the estimate of the error standard deviation for this nine variable model?

   iii. [3] Use this “best” model to predict the Price of a 2005 Buick sedan which has a 6 cylinder 3.8-liter engine with 20,000 miles and an upgraded sound system.
3. [25] The regression diagnostics for the “best” model that contains the nine variables: Mileage, Cylinder, Liter, Sound, Buick, Chevrolet, Pontiac, Saturn, and Sedan are computed. Refer to the JMP output entitled “Regression Diagnostics for the “Best” Model”.
   
   i. [3] According to the distribution of residuals, what values are potential outliers? Give the residual values and the make and model of the vehicles.

   ii. [6] Are there any vehicles with statistically significant standardized residuals? If so, what vehicles are they and what are the values of their standardized residuals? Be sure to support your answer statistically.

   iii. [2] What value of leverage would be considered high?

   iv. [3] Are there any vehicles with high leverage? If so, what vehicles are they and what are the values of their leverage?

   v. [5] Compute the value of the F statistic for the vehicle with the highest leverage. Is this a statistically significant F statistic? Be sure to support your answer statistically.

vii. [3] Are there any vehicles with statistically significant Studentized residuals? If so, what vehicles are they and what are their Studentized residuals?

4. [35] A company collected data on a random sample of 25 employees, 10 women and 15 men. The data included annual salary ($1000), number of years employed at the company and the gender of the employee. The number of years employed goes from 0 (a new employee) to 25. Refer to the JMP output entitled “Salary Study”.
   a) [3] Describe the general relationship between the number of years employed at the company and salary.

   b) [2] Give the equation for predicting salary given only the number of years employed at the company.
c) [4] Comment on the plot of residuals for the simple linear regression of salary on number of years. Be sure to indicate what this plot tells you about predicting salaries for men and women.

d) [3] Give the equation for predicting salary given number of years with the company, an indicator variable for gender and an interaction term gender*years? The indicator variable gender = 0 if the employee is a man, gender = 1 if the employee is a woman.


f) [4] Give an interpretation of the estimated slope coefficient for years for the equation in d).

g) [4] Give an interpretation of the estimated slope coefficient for gender for the equation in d).
h) [5] Does the average yearly change in salary for women differ significantly from the average yearly change in salary for men? Support your answer with an appropriate test of hypothesis.

i) [5] Construct a 95% confidence interval for the difference in mean starting salaries for men and women. Use $t^* = 2.080$. Give an interpretation of the confidence interval.

j) [2] After 17 years what are the predicted salaries for a man and a woman at the company?

You can pick up your graded final exam after noon on Friday, December 19 or at the beginning of next semester at 1407 Wilson Hall.