The second project (worth 50 points, 10% of your final grade) has three objectives:

1. Demonstrating you can use the JMP Stepwise platform.
2. Building a model that is “best”
   - The model is statistically useful at the 5% level.
   - Each variable in the model is statistically significant at the 5% level, given the other variables in the model.
   - Among those models that meet the two criteria above, the one with the highest $R^2$.
3. Performing regression diagnostics including identifying outliers, high leverage and influential values.

The data set for this project consists of body measurements on a random sample of 80 men. A brief description of the data set along with response and explanatory variables is given on the next page. This can be an individual or group (maximum 4 people) project. The result of the project is a thorough, but concise, professional quality technical report of no more than 6 typed pages. The report is to be handed in no later than 5 pm on Friday, May 2, 2014. You can turn in your report sooner if you wish. Your report should contain:

1. Demonstrating you can use the JMP Stepwise platform.
   - Provide a description of the “best” and second “best” single variable models. Give the variable, value of $R^2$, RMSE, and the test that the variable is statistically significant for each of the models.
   - Provide a description of the final models fit by JMP using Forward, Backward and Mixed (Forward) selection procedures (using the P-value Threshold Stopping rule and the default settings for the Prob to Enter and Prob to Leave). Give the complete prediction equation. Give the value of $R^2$, RMSE, $C_p$ and indicate whether each variable in the model is statistically significant or not by reporting the parameter estimate, appropriate test statistic and P-value.

2. Building a model that is “best.”
   - Provide a complete description of your final fitted model with supporting evidence that it is the “best” model. This evidence should include a description of the process you used to find the “best” model and how that process assures you that there is not a better model to be found.

3. Performing regression diagnostics including identifying outliers, high leverage and influential values.
   - Provide a complete evaluation of the results for your “best” model. This evaluation should include appropriate plots and tests for outliers, high leverage points and influential points. You should also comment on whether the conditions for regression analysis are satisfied.
Simply attaching a ream of computer printout in the appendix and expecting the instructor to find what is important is not acceptable. Points will be deducted for including extraneous JMP output that does not address the questions asked. The main body of the report should include only the end products of any statistical calculations. It is appropriate to include plots and important summary values within the body of the report that you feel support statements that you make.

You will have time to work on the project during lab on April 8, 15, 22 and 29.

May 2 – The final reports are due. These are to be typed (or word processed) on plain white paper and should not exceed 6 pages (12 pt font or larger) in length. If you have more than 6 pages, I will stop reading at the end of page 6 and evaluate the project based on the first 6 pages.

A grading rubric will be posted on the course web page.

**Percentage body fat as it relates to body measurements.**

Percentage body fat is determined from a person’s density. The density is obtained from the displacement of water in a large tub. This is a time consuming and inconvenient activity. It would be more convenient if the percentage body fat could be predicted from simple body measurements, ones that can be obtained with a scale and a tape measure. Data are obtained on a random sample of 80 men. A list of the variables appears below and the data are available as **BodyFat_S14.jmp** on the course homepage.

- **BodyFat**: Percentage Body Fat
- **Age**: Age in years
- **Weight**: Weight (kg)
- **Height**: Height (m)
- **BMI**: Weight(kg)/Height(m)^2
- **Neck**: Neck circumference (cm)
- **Chest**: Chest circumference (cm)
- **Abdomen**: Abdomen circumference (cm) at the umbilicus
- **Hip**: Hip circumference (cm)
- **BAI**: \[
\frac{\text{Hip} \text{ (cm)}}{\text{Height} \text{ (m)}} \sqrt[3]{\text{Height} \text{ (m)}} - 18
\]
- **Thigh**: Thigh circumference (cm)
- **Knee**: Knee circumference (cm)
- **Ankle**: Ankle circumference (cm)
- **Biceps**: Extended biceps circumference (cm)
- **Forearm**: Forearm circumference (cm)