Mgmt 230: Introductory Statistics
In-class Exercise: Predicting Wins in Major League Baseball

Work in groups of up to four people to answer the following questions. The questions concern predicting the number of wins a major league baseball team will have based on performance measures such as ERA, runs scored, hits allowed, walks allowed, saves, errors, and also what league the team plays in (American League vs. National League). Download the Excel sheet `BB2005.xls` on the class website.

By signing below, you agree that the following work represents the efforts of everyone in the group, and you are willing to accept as your own grade for the group project the grade earned from representation of your group's work. Every member must agree to these terms to earn a non-zero grade for this assignment.

Signature Group Member 1	Print Name	Date
Signature Group Member 2	Print Name	Date
Signature Group Member 3	Print Name	Date
Signature Group Member 4	Print Name	Date

1. Run a regression that predicts number of wins based on all the other variables given. Print your output and attach it to this sheet. Write down the equation for the regression line.

2. Test the hypothesis that none of the explanatory variables actually help explain the number of wins a team has.

3. What percentage of the variability in wins is explained by your seven variables? Explain your answer.

4. Which league does your regression predict has a greater likelihood of winning? Is this result statistically significant. Test the appropriate hypothesis.

5. What does your regression predict is the impact of a one run increase in ERA? Provide a point estimate for your regression coefficient and also construct and interpret a 95% confidence interval. Does this result make sense? Ask someone what an ERA is if you do not know.

6. Is the effect of a one run increase in ERA statistically significantly less than zero? Test the appropriate hypothesis.

7. What performance measures do you statistical significance as important predictors for the number of wins a team has? For which performance measures do you find there is no statistical significance in predicting the number of wins a team has?

8. Re-run a standardized regression. What do your point estimates for the coefficients predict for which explanatory variables have the greatest influence on the number of wins? Attach your output for the standardized regression.