Methods of Data Analysis II
Homework 1

Name: **OSU-ID:**

Instructions

- The homework is due on Friday, Apr. 10th (in class). Late homeworks will not be accepted under any circumstances.

- You may work individually or in groups of 2 people. If you work in groups, you must include the names and OSU-ID numbers of all the students in the group and submit only one assignment per group.

- You must provide complete answers in order to receive full credit. The homework is worth 25 points.

- You must clearly indicate the problem that you are working.

- Homeworks must be stapled when submitted (including this sheet as a cover page). Do not use folders, paper clips or any other objects to keep the pages together.

1. Data on corn yield and rainfall in six U.S. corn-producing states (Iowa, Nebraska, Illinois, Indiana, Missouri and Ohio), was recorded for each year from 1890 to 1927. The data set (available in the textbook's library) consists of 38 observations on the following 3 variables:

   Year: year of observation (1890 - 1927)

   Yield: average corn yield for the six states (in $bu/acre$)

   Rainfall: average rainfall in the six states (in $in/year$)

   (a) Construct a scatter-plot of yield vs. rainfall. Does the plot suggest a linear relationship between these variables? Explain.

   (b) Regardless your answer from part (a), fit a multiple linear regression model to explain the mean response of *corn yield* in terms of *rain* and $rain^2$. Write down the model equation and report the values of the estimated coefficients with the corresponding P-values.

   (c) Construct 95% confidence intervals for all the parameters in your model.

   (d) Construct a scatter-plot of the residuals vs year. Does the model seem adequate?

   (e) Re-fit the model, but now include the variable *year*. That is, fit a multiple linear regression model to explain the mean response of *corn yield* in terms of $rain$, $rain^2$ and *year*. Report the estimated coefficients and the corresponding P-values.

   (f) Without doing any further analysis. Do you think it would be reasonable to include an interaction between *rain* and *year* in the model? Explain.

   Note: The data set can be read in R using the following commands

```
> library('Sleuth3')
> head(ex0915) # To see the data structure
> ex0915 # To read the full data set
```