Methods of Data Analysis II
Homework 3

Name: **OSU-ID:**

Instructions

- The homework is due on Friday, May 1st (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. The file "density.csv" contains data from a growth experiment with blue-green algae. In the experiment, CO_2 is bubbled through the culture during 14 consecutive days. Two replicates were considered for this experiment, each one corresponding to a measurement (in log-scale) of the increased absorbance of light of a solution containing the algae, which is typically interpreted as a measure of algae density.

   (a) Construct a scatter-plot (including the regression line) for *log-density* vs *day*. Does the plots suggest a linear relationship between these variables? Explain.

   (b) Use the R-function `cor(x,y)` to obtain the correlation between the variables *log-density* and *day*. Comment on this value.

   (c) Fit 4 polynomial regression models for *log-density* vs *day* including: a linear term, a quadratic term, a cubic term and a quartic term. For each one of these models reports the values of R^2 and R^2_{adj}. Comment on these results.

   (d) What is the highest order polynomial that can be fit using these data?

   (e) Construct the residual plots for all the 4 models under considerations and produce a 2 by 2 display in which the top-left corner corresponds to the residual plot for the model with the linear term only, the top-right corresponds to the residual plot for the quadratic model, the bottom-left the cubic model and the bottom-right the quartic model.

   (f) Use the ESS test to compare the models: linear vs. quadratic, quadratic vs. cubic, cubic vs. quartic. Based on these results, which model seems more adequate? Does this result agree with your analysis in part c?

   Note: To read the data make sure you are working on the right directory. Then, you can use the following command:

```
> data <- read.csv("density.csv",header=TRUE)
```