

HOMework #9

Read:

- Chapters 11 in 'Regression Analysis by Example'.

R Assignment:

Solve the following questions using R. Hand in your R code and output file for Questions 1-3, as well as answers to all questions.

1. The salary of the highest paid professional baseball players has increased substantially over the years. The salaries for some notable players can be found in

www.stat.columbia.edu/~martin/W2024/Data/Salary.txt

- (a) Fit a simple linear regression using year as the explanatory variable, and salary as the response variable.
 - (b) Construct a scatter plot of year and salary with the regression line overlayed. Also make a residual plot and a QQ-plot. Comment on whether the assumptions of the regression appear to be valid.
 - (c) Fit a simple linear regression using year as the explanatory variable, and the logarithm of salary as the response variable.
 - (d) Construct a scatter plot of year against the logarithm of salary with the regression line overlayed. Also make a residual plot and a QQ-plot. Comment on whether the assumptions of the regression appear to be valid.
 - (e) Use the model to predict the salary of the highest paid player in 2003.
2. A company that provides transportation services uses a telemarketing division to help sell its services. The manager is interested in the time spent on the phone by telemarketers in the division. Data on the number of months of employment and the number of calls placed per day is recorded for 20 employees. The data can be found on the web page:

www.stat.columbia.edu/~martin/W2024/Data/telemarketing.txt

- (a) Fit a simple linear regression model using months of employment as the explanatory variable and the number of calls as the response variable.
- (b) Make a scatter plot of the explanatory and response variables with the regression line overlayed. Comment on the fit of the model.
- (c) Make a residual plot. Comment on whether the assumptions of the regression model appear to be valid.
- (d) Fit a second order polynomial regression model.
- (e) Test whether the term associated with the second order term is significantly different from 0.
- (f) Make the appropriate residual plots. Comment on whether the assumptions of the regression model appear to be valid.

3. Researchers measured the global mean temperatures (in degrees °C) of the earth's surface every five years between 1950 and 2005. The data is listed below where the variable year is re-coded as follows: 1950 is coded as 1, 1955 as 2, and so on.

Year	1	2	3	4	5	6	7	8	9	10	11	12
Temp	13.8	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4	14.8

- (a) Fit a simple linear regression model to the data using temperature as a response variable and year as an explanatory variable.
- (b) Make a scatter plot of the explanatory and response variables with the regression line overlayed. Comment on the fit of the model.
- (c) Make all relevant diagnostic plots. Comment on whether the assumptions of the regression model appear to be valid.
- (d) Fit a second order polynomial regression model.
- (e) Overlay the new regression line on the scatter plot in (b).
- (f) Make all relevant diagnostic plots. Comment on whether the assumptions of the regression model appear to be valid.
- (g) Use the model in (d) to predict the global mean temperature for the year 2010.