STAT425 Homework #4
Fall 2015 Nov 04, 2015

Homework 4

Due: Nov 11, 2015

Problem 1. The `aatemp` data come from the U.S. Historical Climatology network. They are the annual mean temperatures (in degrees F) in Ann Arbor, Michigan going back about 150 years.

a) (6pt) Suppose someone claims that the temperature was constant until 1930 and then began a linear trend, that is,

$$y_i = \begin{cases} \beta_0 + e_i, & \text{if } x_i \le 1930; \\ \beta_0 + \beta_1(x_i - 1930) + e_i, & \text{if } x_i > 1930. \end{cases}$$

where e_i denotes the normal error term. Explain how to set up a design matrix $X_{n\times 2}$ to fit such a regression model. Report R^2 and the estimated coefficients.

(b) (6pt) Produce a scatter plot the data with x-coordinate being `year` and y-coordinate being `temp`, along with the line from (a) and the SLR line.

(c) (6pt) Which model is better, the SLR model or the model suggested in (a)? To answer this question, you need to explain why the SLR is nested in the model suggested in (a), so you can use F-test. Report the F-stat, its distribution under the null, the corresponding p-value, and your conclusion.

Problem 2. Continue with the `aatemp` data.

(a) (6pt) Fit a polynomial model with degree 10 and use *backward* elimination to reduce the degree of the model. Whats the degree you decide to use? Use the this model to provide a prediction for the temperature in 2015, along with the corresponding 95% CI and PI.

(b) (6pt) Repeat (a) using *forward* approach to select the polynomial degree. Whats the degree you decide to use? Use the this model to provide a prediction for the temperature in 2015, along with the corresponding 95% CI and PI.

(c) (6pt) Fit the data using cubic splines. Use 10-fold CV to select the number of knots. Produce a plot of CV error versus the number of knots. How many knots you decide to use. Use the this model to provide a prediction for the temperature in 2015, along with the corresponding 95% CI and PI.

(d) (6pt) Produce a scatter plot, along with the three curves from (a), (b), and (c). Of course, if the polynomial you select in (a) agrees with the one you select in (b), then you only need to draw two curves. Remember to draw those curves using different line types and color, and add captions.

Problem 3. (2+8+8=18pt) Analyze the `chickwts` data from the package `faraway`.

(a) Provide a graphical display of the data: stripchart on top of the boxplot.

(b) Determine whether there are any differences in weights of chickens between the feed types. Explain your model assumption under H_0 and H_a. What's the test statistic? What's its distribution under H_0? What's your decision?

(c) Perform necessary model diagnostics, such as QQ-plot, residuals versus fitted plot, Levene's test for constant variance, and your comments.

Problem 4. (6+12+8=26pt) An experiment was conducted to compare the lifetime of three different brands of batteries. The data are presented in the table below:

Brand	Mean	SD	Group Size
A	75	9	5
B	57	13	5
C	48	15	5

where

$$\text{Mean}_i = \bar{y}_{i\cdot}, \qquad \text{SD}_i = \sqrt{\frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2}.$$

(a) Do the data suggest that there are differences between brands? State your null and alternative hypotheses, calculate the corresponding test statistic, and give the corresponding p-value.

(b) Consider the following contrasts:

$$\begin{aligned} \gamma_1 &: A \textit{ versus } B \\ \gamma_2 &: B \textit{ versus } C \\ \gamma_3 &: A \textit{ versus } \text{average of } B \text{ and } C \end{aligned}$$

Calculate the LS estimate for γ_j $(j = 1, 2, 3)$, as well as their standard deviation.

(c) Find the simultaneous 95% confidence intervals for γ_j $(j = 1, 2, 3)$ using Bonferroni and Scheffés methods.