The University of Sydney
STAT2012 Statistical Tests

Semester 2	Computer Practice Week 11	2015

Useful R commands

- Regression model:

  If $Y_i = \alpha + \beta x_i + \epsilon_i,\ i = 1, \ldots, n$ where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, the ANOVA table for the test

  $$H_0 :\ \beta = 0 \quad \text{vs} \quad H_1 :\ \beta \neq 0$$

  of the significance of the regression model can be obtained in R by

  ```
  summary(aov(y~x))
  ```

  where `x` and `y` are `n` $\times$ `1` vectors of observations.

- Correlation:

  If (X_i, Y_i) follow a bivariate normal distribution, the test

  $$H_0 :\ \rho = 0 \quad \text{vs} \quad H_1 :\ \rho \neq 0,$$

  for the significance of the linear relationship between X and Y can be conducted in R by

  ```
  cor.test(x,y,alt="two.sided")
  ```

Important points

- You will test the significance of a regression model.
- You will construct confidence intervals for intercept, slope, mean prediction and point prediction in regression analysis.
- You will calculate correlation coefficient and coefficient of determination for a regression model.

Practice Problems

1. The data `fuel` contain information on makes of cars taken from the April 1990 issue of Consumer Reports. Open the data set `fuel`.

   ```
   fuel=read.csv("http://www.maths.usyd.edu.au/u/UG/IM/STAT2012/r/fuel.csv")
   attach(fuel)
   ```

(a) Create two vectors x and y which correspond to `Weight` and `Fuel` respectively. Test for the significance of the regression model of `Fuel` on `Weight`. State the null and alternative hypotheses, the test statistic, the p-value and the result of the test.

(b) Check the test statistic and p-value in (a).

(c) Provide 95% confidence intervals for the slope and the y-intercept estimates. From the confidence intervals, comment on whether the slope estimate is significantly different from zero. Compare this result with the result in (a).

(d) Predict the `Fuel` consumption for a car which weighs 3000 lb and provide a 95% prediction interval for the predicted `Fuel` consumption.

(e) Test for the significance of the linear relationship between the `Weight` and `Fuel`. State the null and alternative hypotheses, the test statistic, the p-value and the result of the test. State the correlation coefficient and calculate the coefficient of determination from the test. Explain the proportion of variation in `Fuel` explained by `Weight`. Verify that the square of the test statistic equals to the test statistic in (a).