Semester 2	Statistical Tests (Advanced)	2011

Tutorial 10

1. A study was constructed to determine whether a linear relationship exists between the age (x) and the diameter (y in inches) at the 1.5 metre level of chestnut trees. Ten randomly selected chestnut trees were examined and the data were recorded as follows:

Age (x)	Diameter (y) at the 1.5 metre level
4	0.8
5	0.8
8	1.0
8	2.3
10	3.2
12	4.9
13	3.7
14	4.5
17	5.6
19	6.5

$$\sum_i x_i^2 = 1428, \quad \sum_i y_i^2 = 149.37, \quad \sum_i x_i y_i = 454.2.$$

   (a) Fit a regression line to the data.

   (b) Use an F-test to test whether or not a linear relationship exists between the age and the diameter at the 1.5 metre level of chestnut trees.

   (c) Find an estimate of the average diameter for a tree of age 30 and give a 95% confidence interval for the average diameter.

   (d) Find an estimate of the diameter of a chestnut tree at age 20 and give a 95% prediction interval for the diameter.

2. A behavioural biologist believes that performance of a laboratory rat on an intelligence test depends, to a large extent, on the amount of protein in the rat's daily diet. To check out the theory, he accumulated the following data after working with 10 rats.

Rat	1	2	3	4	5	6	7	8	9	10
No. of units of protein daily(x)	7	8	9	10	10	11	12	15	18	20
Score on a standard test(y)	20	19	23	25	23	25	26	24	28	26

$$\sum_i x_i^2 = 1608, \quad \sum_i y_i^2 = 5781, \quad \sum_i x_i y_i = 2950.$$

   (a) Fit a regression line to the data.

   (b) Decide whether or not a linear relationship between the amount of protein and the score on the standard test, as described by the data, is significant using a t-test.

   (c) Calculate the sample correlation coefficient r and comment on the result.

   (d) Calculate a 95% confidence interval for the average score of a rat with an average daily intake of 16 units of protein.

   (e) Which protein amount x_0 will result in the narrowest 95% confidence interval?

3. As an additional theoretical excercise do problem 7 on p. 593 of Rice's book.

Computer Exercises 10

1. Consider the data frame `fuel.frame`, which has information on makes of cars taken from the April 1990 issue of Consumer Reports. This exercise continues the discussion in week 9 about the relationship between weight and fuel recorded in `fuel.frame`.

   (a) Create two vectors x and y whose elements correspond to `Weight` (in pounds) and `Fuel` (in gallons per 100 miles) in `fuel.frame`.

   (b) Calculate S_{xx}, S_{yy} and S_{xy}, and store your results as `Sxx`, `Syy` and `Sxy` respectively.

   (c) Test "manually" whether the weight has an influence on the fuel using an F-test.

   (d) Check your calculations by displaying the regression ANOVA table using `summary`.

   (e) Find the regression line for Fuel on Weight using `lsfit`, and then find an estimate of average fuel in gallons per 100 miles corresponding to cars with weight 2500 pounds. Provide a 90 % confidence interval for the average.

   (f) Obtain a 90 % prediction interval for the prediction of the fuel in gallons per 100 miles for a car with weight 2500 pounds.

   (g) Calculate the sample correlation coefficient r between weight and fuel, and comment on the result.

   (h) Test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$ using the built-in command `cor.test`.