THE UNIVERSITY OF SYDNEY

STAT2012 STATISTICAL TESTS

Semester 2	**Tutorial Week 10**	2015

Summary of week 9

- Regression: For the simple linear regression model

$$Y_i = \alpha + \beta\, x_i + \epsilon_i, \quad \text{where} \quad \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2),$$

$$\text{Slope:} \qquad \hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2},$$

$$\text{y-intercept:} \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}, \quad \hat{\alpha}$$

Tutorial Questions

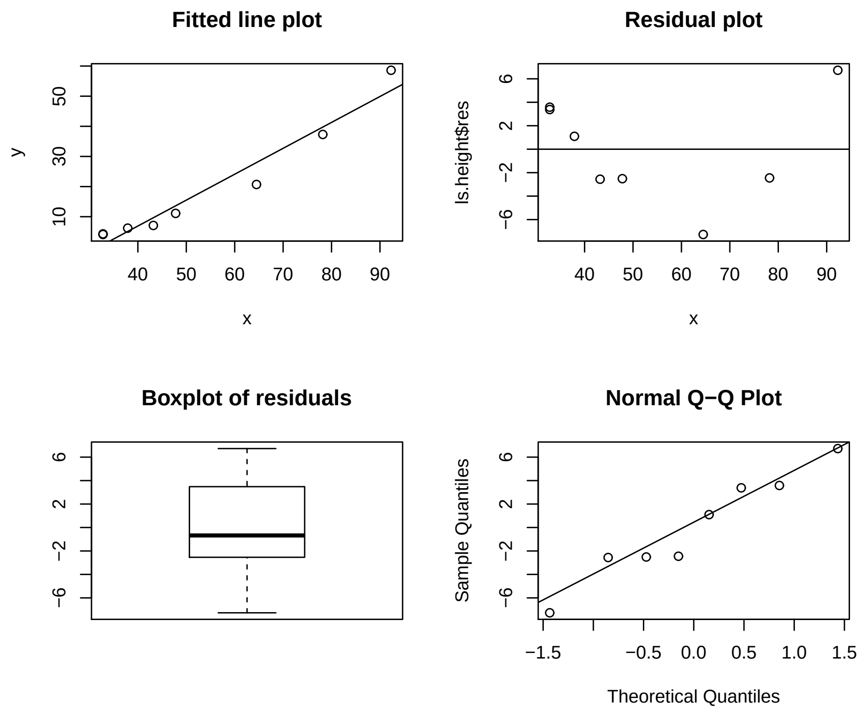
1. An experiment is conducted to study the relationship between the velocity X in km/hr of a vehicle at the time of braking and the stopping distance Y in m required. The following is the data:

Distance Y	4.1	4.3	6.2	7.1	11.1	20.7	37.3	58.6
Velocity X	32.8	32.8	37.9	43.2	47.8	64.5	78.2	92.3

The summary statistics are

$$\sum_{i=1}^{8} x_i = 429.5, \ \sum_{i=1}^{8} y_i = 149.4, \ \sum_{i=1}^{8} x_i^2 = 26533.95, \ \sum_{i=1}^{8} y_i^2 = 5501.1, \ \sum_{i=1}^{8} x_i y_i = 11008.59.$$

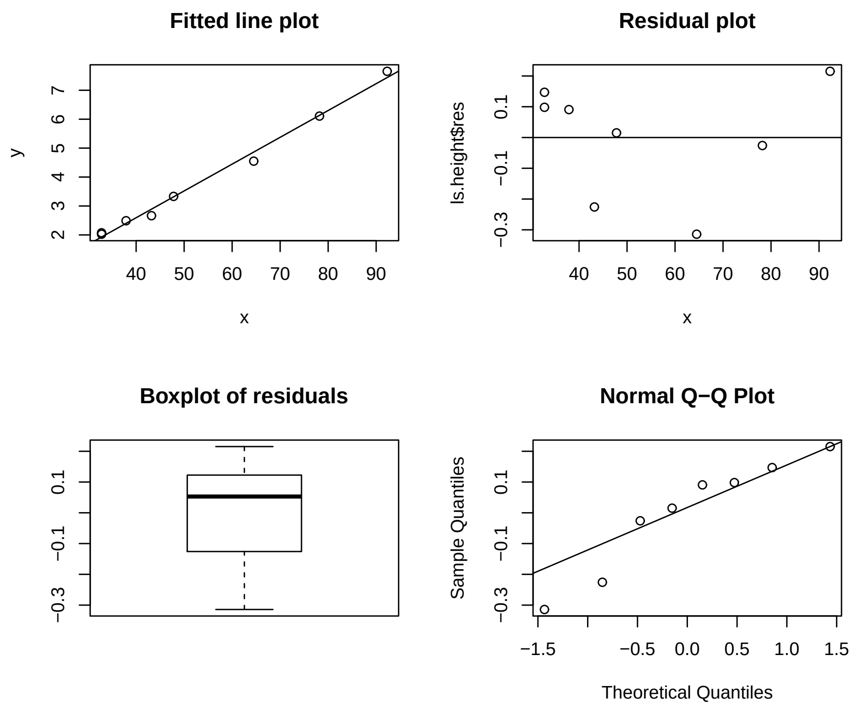
(a) Fit a least squares regression line to the data. Plot the data and the regression line. Does a straight line relationship between X and Y seem to be appropriate?



(b) Transformation is employed to improve the fitting. Complete another row of the data $\sqrt{Y}$ and fit a revised least squares regression line of $\sqrt{Y}$ against X. Plot the data. Does a straight line relationship between X and $\sqrt{Y}$ seem to be appropriate?

The revised summary statistics are

$$\sum_{i=1}^{8} y_i = 30.89688, \ \sum_{i=1}^{8} y_i^2 = 149.4, \ \sum_{i=1}^{8} x_i y_i = 1980.781.$$



(c) Predict the stopping distance when the velocity at the time of braking is 50km/hr.

2. A statistician investigates the relationship between the amount of precipitation (in inches x) and the number of automobile accidents (y). He gathered data for 10 randomly selected days as shown below:

Day i	Number of accidents y	Amount of precipitation (in inches) x
1	5	0.05
2	6	0.12
3	2	0.05
4	4	0.08
5	8	0.10
6	14	0.35
7	7	0.15
8	13	0.30
9	7	0.10
10	10	0.20

The summary statistics are

$$\sum_{i=1}^{10} x_i = 1.5,\ \sum_{i=1}^{10} y_i = 76,\ \sum_{i=1}^{10} x_i^2 = 0.3208,\ \sum_{i=1}^{10} y_i^2 = 708,\ \sum_{i=1}^{10} x_i y_i = 14.74.$$

(a) Determine the least squares regression line.

(b) Calculate the standard error estimate for the residuals.

(c) Calculate the 95% confidence interval for α and β. Draw your conclusion about the significance of β.

Extra Practice Problems

1. By late 1971, all cigarette packs in the United States had to be labeled with the words, "Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous To Your Health." The case against smoking rested heavily on statistical, rather than laboratory, evidence. Extensive surveys of smokers and nonsmokers had revealed the former to have a much higher risks of dying from a variety of causes, most notably lung cancer and heart disease. Other types of studies, some designed with a much broader focus, painted much the same picture. Typical are the data below: 21 countries were the subjects. Recorded for each country was x, its annual cigarette consumption and y (per adult per year), its mortality rate due to coronary heart disease (per 100,000 ages 35-64) in the year of 1962.

Country	1	2	3	4	5	6	7	8	9	10	11
x	3900	3350	3220	3220	2790	2780	2770	2290	2160	1890	1810
y	256.9	211.6	238.1	211.8	194.1	124.5	187.3	110.5	233.1	150.3	124.7
Country	12	13	14	15	16	17	18	19	20	21	
x	1800	1770	1700	1680	1510	1500	1410	1270	1200	1090	
y	141.2	82.1	118.1	71.9	114.3	95.2	136.3	126.9	59.7	42.6	

The summary statistics are:

$$\sum_{i=1}^{21} x_i = 45,110,\ \sum_{i=1}^{21} y_i = 3031.2,\ \sum_{i=1}^{21} x_i^2 = 109,957,100,\ \sum_{i=1}^{21} y_i^2 = 513,248.16,\ \sum_{i=1}^{21} x_i y_i = 7,340,085.$$

(a) Plot the data. Does a straight line relationship between X and Y seem to be appropriate?

(b) Find the estimated regression line $\widehat{Y} = \widehat{\alpha} + \widehat{\beta} X$. Draw the estimated regression line on the scatter plot in (a).

(c) Construct a 95% confidence interval for α.