Stat 320, Spring 2006

Lab_assignment 2 February 3, 2006

Dr. Bensmail

(A) Simple Linear Regression Example

Let us suppose the following is a data set we want to regress to build a prediction equation for Y.

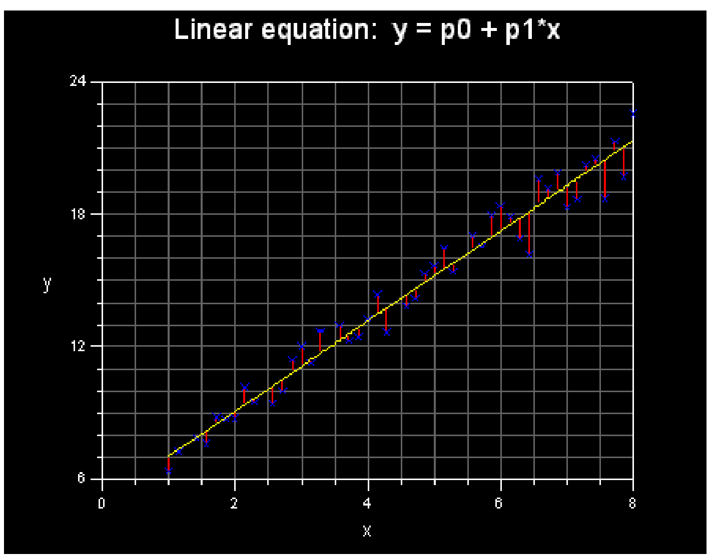
X	3	4	5	3	4	5	4	5	7	6	8	9	10	8	11
Y	20	18	21	16	20	19	21	20	24	22	26	27	30	28	32

1-Give the following summary statistics of $\mathbf{X}$ and $\mathbf{Y}$ (mean, min, max, std)

2-Fit a linear equation $\mathbf{Y} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{X}$ to this data using (a) the least square estimate, (b) JMP outputs. Compare the two results.

3- Provide a scatter plot of the data with the regression line.

4-A straight line can be fitted to any set of data. The formulas for the coefficients of the least squares fit are the same for a sample, a population, or any arbitrary batch of numbers. However, regression is usually used to let analysts generalize from the sample in hand to the population from which the sample was drawn. We suppose that a population regression equation is given by $\mathbf{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{X} + \mathbf{e}_i$ where $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are the population regression coefficients and $\mathbf{e}_i$ is a random error peculiar to the i-th observation.



Thus, each response is expressed as the sum of a value predicted from the corresponding $\mathbf{X}$, plus a random error. Thus the sample regression equation is an estimate of the population regression equation.

b_1 is the point estimate of $\boldsymbol{\beta}_1$. If $\boldsymbol{\beta}_1 = 0$, do you think it is necessary to use a regression model to check for a possible linear relationship between $\mathbf{X}$ and $\mathbf{Y}$?

5- Test if $\boldsymbol{\beta}_1 = 0$ versus $\beta_1 \neq 0$. Conclusion?

(B) Stock life insurance

A financial analyst is investigating the performance of a stock life insurance company in the second quarter of its fiscal period in 1991 .Because of a series of abnormal charges, second quarter earnings are atypical for this company .The analyst would like to determine what the "normal" earnings for this company would have been in the absence of these abnormal charges. To do this, the analysts elects a sample of size $\mathbf{n} = \mathbf{36}$ stock insurance companies and compares the relationship between their 1991 second quarter earnings per share to their 1990 earnings per share.

1. Graph **EPS91** versus **EPS90**?

2. Does the graph show a strong relationship between **EPS91**($\mathbf{Y}$) and **EPS90**($\mathbf{X}$) ?

3. Calculate $\bar{\mathbf{Y}}$, $\bar{\mathbf{X}}$, $\mathbf{S}_X$, $\mathbf{S}_Y$ and $\mathbf{r}$.

5. We may see an unusual performance of a particular company? Let's call this company **USF&G** corporations.

6. We wanted to provide the summary statistics without **USF&G**. Calculate the new $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$?

7. Calculate the new values of $\mathbf{S}_Y$ and $\mathbf{S}_X$ using the following relationship

$$(\mathbf{n}-\mathbf{1})\mathbf{S}_Y^2 = \sum_{i=1}^{n} \mathbf{Y}_i^2 - \mathbf{n}\bar{\mathbf{Y}}^2$$

8. Calculate the new value of r using the following relationship

$$\sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{Y}_i - \bar{\mathbf{Y}}) = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{Y}_i - \mathbf{n}\bar{\mathbf{X}}\bar{\mathbf{Y}}$$

8. What is the percentage change in r between the original data set analyzed and the data set with USF&G deleted?