Homework #11 Solution
ENCE 302 – FALL 2001
Due W, 12/5

Problem 1:
Use least-squares regression to fit a straight line to the following data:

x	4	6	8	10	14	16	20	22	24	28	28	34	36	38
y	30	22	22	28	14	22	16	8	20	8	14	14	0	4

Along with the slope and the intercept, compute the standard deviation, the standard error of estimate, and the correlation coefficient. Plot the data and the regression line (in the same plot), and assess the prediction of the regression model.

***** SOLUTION *****

i	*X*	*Y*	*X*²	*Y*²	*XY*
1	4	30	16	900	120
2	6	22	36	484	132
3	8	22	64	484	176
4	10	28	100	784	280
5	14	14	196	196	196
6	16	22	256	484	352
7	20	16	400	256	320
8	22	8	484	64	176
9	24	20	576	400	480
10	28	8	784	64	224
11	28	14	784	196	392
12	34	14	1156	196	476
13	36	0	1296	0	0
14	38	4	1444	16	152
Σ	288	222	7592	4524	3476

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{3476 - \frac{1}{14}(288)(222)}{(7592) - \frac{(288)^2}{14}} = -0.6542$$

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{b_1}{n}\sum_{i=1}^{n} x_i = \frac{1}{14}(222) - \frac{-0.6542}{14}(288) = 29.315$$

Therefore, the regression line is

$$\hat{Y} = 29.315 - 0.6542X$$

The standard deviation is computed as follows:

$$S_Y = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{m} y_i\right)^2\right]} = \sqrt{\frac{1}{14-1}\left[4524 - \frac{1}{14}(222)^2\right]} = 8.7869$$

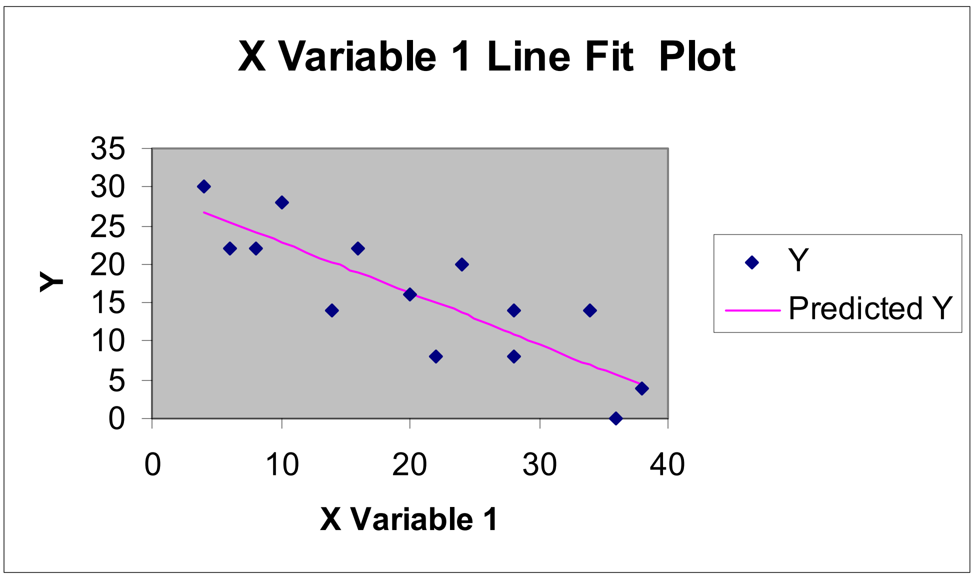
The correlation coefficient is

$$R = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2}} = \frac{3476 - \frac{288(222)}{14}}{\sqrt{\left(7592 - \frac{288^2}{14}\right)\left(4524 - \frac{222^2}{14}\right)}} = -0.8432$$

The standard error of estimate is obtained as

$$S_e = \sqrt{\left(\frac{n-1}{n-p-1}\right) S_Y^2 \left(1 - R^2\right)} = \sqrt{\left(\frac{14-1}{14-1-1}\right)(8.7869)^2\left(1 - (-0.8432)^2\right)} = 4.9167$$

<u>Plot of the Data:</u>



<u>Assessment of the model:</u>
The value of the coefficient of determination $R^2 = 0.711$ suggests that about 71% of the original uncertainty has been explained by the model.
The value of $S_e = 4.92$ is less that $S_Y = 8.79$, almost by half. So the regression line has improved the prediction somewhat. The ideal value for $S_e \approx 0$.

<u>Problem 2:</u>

Textbook: 6-3

	X_i	Y_i	$\hat{Y}$	TV $(y_i - \bar{y})^2$	EV $(\hat{y}_i - \bar{y})^2$	UV $(y_i - \hat{y}_i)^2$	EV +UV
	2	1	4.7735	16	0.051302	14.2393	14.2906045
	5	3	5.0564	4	0.003181	4.228781	4.23196192
	7	5	5.245	0	0.060025	0.060025	0.12005
	6	7	5.1507	4	0.02271	3.419913	3.44262098
	2	9	4.7735	16	0.051302	17.8633	17.9146045
$\sum$	22	25	24.9991	40	0.188521	39.81132	40.0

Therefore,

TV = EV +UV = $\sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 = 0.1888521 + 39.81132 = 40$

Problem 3:

Textbook: 6-14

(a)

$n = 12$, $R = 0.582$, $\alpha = 5\%$

$\rho = 0$

H_0: $\rho = 0$

H_A: $\rho \neq 0$

$$t = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}} = \frac{0.582}{\sqrt{\frac{1-0.582}{12-2}}} = 2.263$$

Since ($t > t_{0.025,\,10} = 2.228$), reject the null hypothesis H_0.

(b)

$\rho = -0.1$

H_0: $\rho = -0.1$

H_A: $\rho \neq -0.1$

$$R' = 0.5\ln\left(\frac{1+R}{1-R}\right) = 0.5\ln\left(\frac{1+0.582}{1-0.582}\right) = 0.6655$$

$$\bar{R}' = 0.5\ln\left(\frac{1+\rho_0}{1-\rho_0}\right) = 0.5\ln\left(\frac{1-0.1}{1+0.1}\right) = -0.1003$$

$$s_{R'} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = \frac{1}{3}$$

$$Z = \frac{R' - \bar{R}'}{s_{R'}} = \frac{0.6655 - (-0.1003)}{1/3} = 2.2975$$

Since ($Z > t_{0.975} = 1.96$), reject the null hypothesis H_0.

Problem 4:

Textbook: 6-32

x	y	xy	x^2	y^2	
2	1	2	4	1	
1	2	2	1	4	
3	4	12	9	16	
2	5	10	4	25	
8	12	26	18	46	Summation

a)

$$\text{Slope} = b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2} = 1$$

$$\text{Intercept} = b_0 = \bar{Y} - b_1 \bar{X} = \frac{\sum_{i=1}^{n} Y_i}{n} - b_1 \frac{\sum_{i=1}^{n} X_i}{n} = 1$$

Hence,

$$\hat{Y} = b_0 + b_1 X = 1 + X$$

b)

$$\text{Slope} = b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2} = 0.2$$

$$\text{Intercept} = b_0 = \bar{X} - b_1 \bar{Y} = \frac{\sum_{i=1}^{n} X_i}{n} - b_1 \frac{\sum_{i=1}^{n} Y_i}{n} = 1.4$$

Hence,

$$\hat{X} = b_0 + b_1 Y = 1.4 + 0.2Y$$

c)

$$R_a = R_b = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2} \cdot \sqrt{\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2}} = 0.44721$$

d) The transformed regression equation of part (a) is

$$\hat{X} = -1 + Y$$

with regression coefficients b_1 = 1 and b_0 = -1 as opposed to b_1 = 1 and b_0 = 1 computed in part (b).

(d) When using regression, it is necessary to specify which variable is the criterion and which is the predictor. The distinction is necessary with regression because a regression equation is not transformable, unless the correlation coefficient equals 1.0. That is, if *Y* is the criterion variable when the equation is calibrated, the equation cannot be rearranged algebraically to get an equation for predicting *X*.

Problem 5:

Textbook: 6-41

$$\hat{Y} = aX^b$$

Let

$\hat{Z} = \ln \hat{Y}, \quad c = \ln a, \quad$ and $W = \ln X$

Then

$$\ln \hat{Y} = \ln(aX^b) = \ln a + b \ln X$$

or

$$\hat{Z} = c + bW$$

$$F = \min \sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n} (\hat{Z}_i - Z_i)^2 = \min \sum_{i=1}^{n} (c + bW_i - Z_i)^2$$

The resulting derivatives are

$$\frac{\partial F}{\partial c} = 2\sum_{i=1}^{n} (c + bW_i - Z_i)\,(1) = 0$$

$$\frac{\partial F}{\partial b} = 2\sum_{i=1}^{n} (c + bW_i - Z_i)\,(W_i) = 0$$

Rearranging above two equations, the following set of normal equations can be obtained:

$$\begin{array}{lcccl} cn & + & b\sum_{i=1}^{n} W_i & = & \sum_{i=1}^{n} Z_i \\ c\sum_{i=1}^{n} W_i & + & b\sum_{i=1}^{n} W_i^2 & = & \sum_{i=1}^{n} W_i Z_i \end{array}$$

The coefficient *a* can be determined using

$$a = e^c$$

X	Y	$W(\ln X)$	$Z(\ln Y)$	$W\,Z$	W^2
1	1	0	0	0	0
2	1	0.69315	0	0	0.48045
3	2	1.09861	0.69315	0.7615	1.20695
4	4	1.38629	1.38629	1.92181	1.92181
5	4	1.60944	1.38629	2.23115	2.59029
15	12	4.78749	3.46574	4.91447	6.1995 Summation

$$\begin{array}{lcccl} 5c & + & 4.78749b & = & 3.46574 \\ 4.78749c & + & 6.1995b & = & 4.91447 \end{array} \quad \Rightarrow \quad \begin{bmatrix} c \\ b \end{bmatrix} = \begin{bmatrix} -0.252821 \\ 0.987958 \end{bmatrix}$$

From which

$a = e^{-0.252821} = 0.776607$