Problem Set 5 - Solutions

Econ-310, Spring 2004

10.10 a.

x_i	y_i	x_i^2			x_iy_i		
7	2	7^2	=	49	7(2)	=	14
4	4	4^2	=	16	4(4)	=	16
6	2	6^2	=	36	6(2)	=	12
2	5	2^2	=	4	2(5)	=	10
1	7	1^2	=	1	1(7)	=	7
1	6	1^2	=	1	1(6)	=	6
3	5	3^2	=	9	3(5)	=	15

$$\sum x_i = 24 \sum y_i = 31$$
$$\sum x_i^2 = 116 \sum x_iy_i = 80$$

b. $SS_{xy} = \sum x_iy_i - \frac{(\sum x_i)(\sum y_i)}{n} = 80 - 106.2857143 = -26.2857143$

c. $SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 116 - 82.28571429 = 33.71428571$

d. $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-26.2857143}{33.71428571} = -.7797$

e. $\overline{x} = \frac{(\sum x_i)}{n} = \frac{24}{7} = 3.42857$ $\overline{y} = \frac{(\sum y_i)}{n} = \frac{31}{7} = 4.42857$

f. $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 7.102$

g. The least square line is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 7.102 - .7797x$.

10.28 a. From the printout, SSE = 926.46437, s_2 = MSE = 77.20536, and s = Standard Error = 8.78666.

b. We would expect that most of the observations will fall within $2s$ or 2(8.78666) or 17.573 games of their predicted values.

10.46 a. Some preliminary calculations are:

$$\sum x = 0 \quad \sum x^2 = 10 \quad \sum xy = 20$$
$$\sum y = 12 \quad \sum y^2 = 70$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 20 - \frac{(0)(12)}{5} = 20$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 10 - \frac{(0)^2}{5} = 10$$

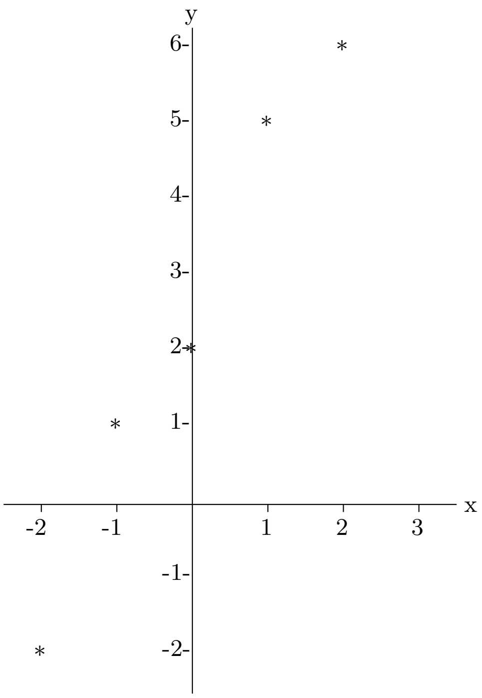
$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 70 - \frac{(12)^2}{5} = 41.2$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{20}{\sqrt{10(41.2)}} = .9853$$

$$r^2 = .9853^2 = .9709$$

Since r = .9853, there is a very strong positive linear relationship between x and y.

Since $r^2 = .9709$, 97.09% of the total sample variability around $\overline{y}$ is explained by the linear relationship between x and y.



b. Some preliminary calculations are:

$$\sum x = 0 \quad \sum x^2 = 10 \quad \sum xy = -15$$
$$\sum y = 16 \quad \sum y^2 = 74$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = -14 - \frac{(0)(16)}{5} = -15$$

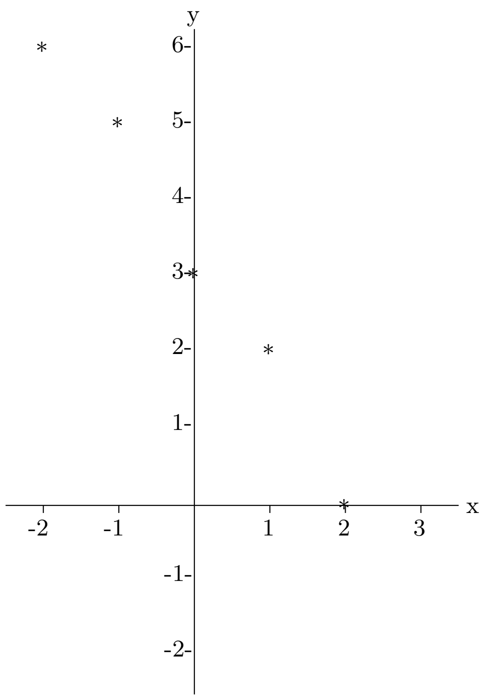
$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 10 - \frac{(0)^2}{5} = 10$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 74 - \frac{(16)^2}{5} = 22.8$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-15}{\sqrt{10(22.8)}} = -.9934$$

$$r^2 = .9934^2 = .9868$$

Since r = .9934, there is a very strong negative linear relationship between x and y.



Since $r^2 = .9868$, 98.68% of the total sample variability around $\bar{y}$ is explained by the linear relationship between x and y.

c. Some preliminary calculations are:

$$\sum x = 18 \quad \sum x^2 = 52 \quad \sum xy = 36$$
$$\sum y = 14 \quad \sum y^2 = 32$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 36 - \frac{(18)(14)}{7} = 0$$

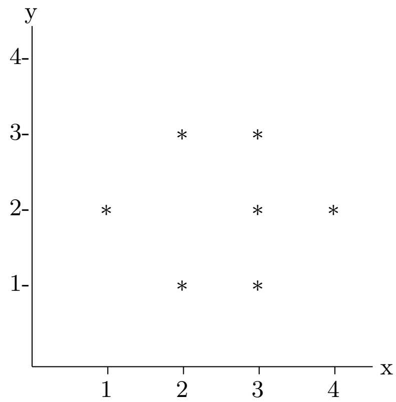
$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 52 - \frac{(18)^2}{7} = 5.71428571$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 32 - \frac{(14)^2}{7} = 4$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{0}{\sqrt{5.71428571(4)}} = 0$$

$$r^2 = 0^2 = 0$$

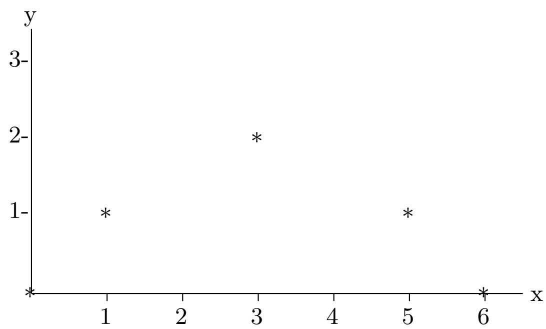
Since $r = 0$, this implies that x and y are not related.
Since $r^2 = 0$, 0% of the total sample variability around $\overline{y}$ is explained by the linear relationship between x and y.



d. Some preliminary calculations are:

$$\sum x = 15 \quad \sum x^2 = 71 \quad \sum xy = 12$$
$$\sum y = 4 \quad \sum y^2 = 6$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 12 - \frac{(15)(4)}{5} = 0$$



$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 71 - \frac{(15)^2}{5} = 26$$

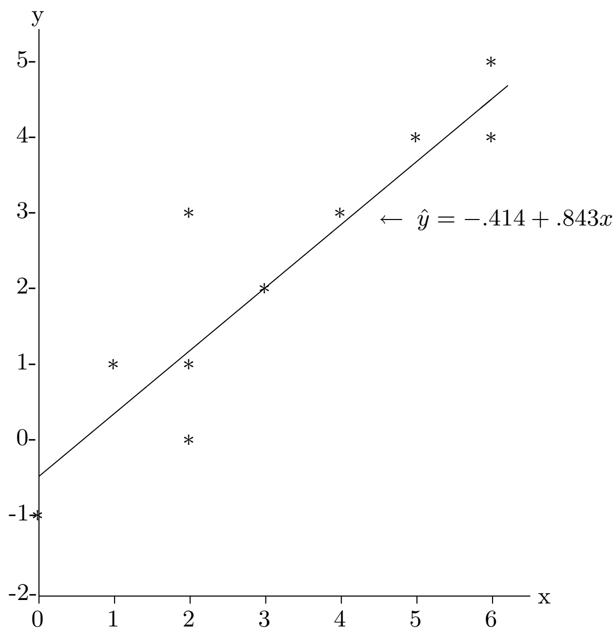
$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 74 - \frac{(16)^2}{5} = 2.8$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{0}{\sqrt{26(2.8)}} = 0$$

$$r^2 = 0^2 = 0$$

Since $r = 0$, this implies that x and y are not related.
Since $r^2 = 0$, 0% of the total sample variability around $\overline{y}$ is explained by the linear relationship between x and y.

10.58 a,b. The scattergram is:



c. $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 33.6 - .84318766(32.8) = 5.94344473$

$$s^2 = \frac{SSE}{n-2} = \frac{5.94344473}{10-2} = .742930591$$

$$s = \sqrt{.742930591} = .8619$$

$$\overline{x} = \frac{31}{10} = 3.1$$

The form of the confidence interval is $\hat{y} \pm t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{(x_p - \overline{x})^2}{SS_{xx}}}$

For $x_p = 6, \hat{y} = -.414 + .843(6) = 4.644$

For confidence coefficient .95, $\alpha = .05$ and $\alpha/2 = .025$. From Table VI, Appendix B, with df=n-2=10-2=8, $t_{.025} = 2.306$. The confidence interval is:

$$4.644 \pm 2.306(.8619)\sqrt{\frac{1}{10} + \frac{(6-3.1)^2}{38.9}} \Rightarrow 4.644 \pm 1.118 \Rightarrow (3.526, 5.762)$$

d. For $x_p = 3.2, \hat{y} = -.414 + .843(3.2) = 2.284$

The confidence interval is:

$$2.284 \pm 2.306(.8619)\sqrt{\frac{1}{10} + \frac{(3.2-3.1)^2}{38.9}} \Rightarrow 2.284 \pm .629 \Rightarrow (1.655, 2.913)$$

For $x_p = 0, \hat{y} = -.414 + .843(0) =$ -.414

The confidence interval is:

$$2.284 \pm 2.306(.8619)\sqrt{\frac{1}{10} + \frac{(0-3.1)^2}{38.9}} \Rightarrow 2.284 \pm 1.717 \Rightarrow (-1.585, .757)$$

W a. Generally, the coefficient is interpreted as the change in the left-hand side (dependent) variable for a given unit change in the right-hand side (independent) variable. For the units that these variables are measured in, this coefficient is the percentage point change in the GDP growth rate in the developing countries for a one percentage point change in the GDP growth rate in G-7 countries. When the GDP growth rate in G-7 countries rises by one percentage point, the GDP growth rate in the developing countries increases by 0.275 percentage points.

b. The test in this question is a 2-sided t-test for the null hypothesis that the true value of the parameter is 0, i.e.:

$H_0 : \ \beta_1 = 0$

$H_a : \ \beta_1 \neq 0$

Reject the null if $| \, t \, | > t_{\alpha/2}$. For 5% significance, $t_{\alpha/2} = 2.074$ for 22 d.f.

Calculating $t = (0.274541 - 0)/0.212348 \simeq 1.2929$

Since $| \, 1.2929 \, | = 1.2929 < 2.074 \Rightarrow$we do not reject the null hypothesis.

c. The entry 'Sum Squared Resid' is the same as what the book calls 'Sum of Squared Errors (SSE)'.

$s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{32.88168}{22}} = 1.2225$

which is the same as the entry under 'S.E. of regression'.

d. Same as c). This question aimed at checking whether you understand the regression output and the concepts in the textbook.

e. $R^2 = 1 - \frac{SSE}{SS_{yy}}$. While SS_{yy} is not reported in the regression output, notice that the standard deviation of the dependent variable y is reported. In fact $SS_{yy} = (SD^2) \times (n-1) = (1.240266)^2 \times 23 = 35.38$

Hence $R^2 = 1 - \frac{SSE}{SS_{yy}} = 1 - \frac{32.88168}{35.38} = 0.070614$ which is exactly equal to the reported R^2 in the output.

f. This coefficient is the percentage point change in the GDP growth rate of developing countries for a one percentage increase in the real interest rate, *holding all other independent variables constant*. When the real interest rate rises by one percentage point, and *all other variables remain constant*, the GDP growth rate in developing countries *rises* by 0.728 percentage points.

g. $s = \sqrt{\frac{SSE}{n-(k+1)}} = \sqrt{\frac{14.46213}{24-(3+1)}} = \sqrt{\frac{14.46213}{20}} = 0.85036$
which is the same as the entry under 'S.E. of regression'.

h. Note that a 95% confidence interval implies $\alpha = 0.05$, and hence $\alpha/2 = 0.025$. We have 20 d.f. in this problem. Therefore $t_{\alpha/2} = 2.086$.

The confidence interval for β_2 is given by:

$$\widehat{\beta}_2 \pm t_{\alpha/2} \times s_{\widehat{\beta}_2} = -0.314391 \pm 2.086 \times 0.102988$$
$$= [-0.52922, -0.09956]$$