## Homework 11 - Solutions

1) Sincich, 12.2 (don't interpret R^2-adj on part f)

a. $\hat{\beta}_0 = 506.346,\ \hat{\beta}_1 = -941.900,\ \hat{\beta}_2 = -429.060$

b. $\hat{y} = 506.346 - 941.900x_1 - 429.060x_2$

c. SSE = 151,016, MSE = 8883, $s = 94.251$

We expect about 95% of the $y$-values to fall within $\pm 2s$ or $\pm 2(94.251)$ or $\pm 188.502$ units of the fitted regression equation.

d. $H_0$: $\beta_1 = 0$
$H_a$: $\beta_1 \neq 0$

The test statistic is $t = \dfrac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \dfrac{-941.900}{275.08} = -3.42$

The rejection region requires $\alpha/2 = .05/2 = .025$ in each tail of the $t$ distribution with df $= n - (k + 1) = 20 - (2 + 1) = 17$. From Table V, Appendix B, $t_{.025} = 2.110$. The rejection region is $t < -2.110$ or $t > 2.110$.

Since the observed value of the test statistic falls in the rejection region ($t = -3.42 < -2.110$), $H_0$ is rejected. There is sufficient evidence to indicate $\beta_1 \neq 0$ at $\alpha = .05$.

e. For confidence coefficient .95, $\alpha = .05$ and $\alpha/2 = .025$. From Table V, Appendix B, with df $= n - (k + 1) = 20 - (2 + 1) = 17$, $t_{.025} = 2.110$. The 95% confidence interval is:

$$\hat{\beta}_2 \pm t_{.025}\, s_{\hat{\beta}_2} \Rightarrow -429.060 \pm 2.110(379.83) \Rightarrow -429.060 \pm 801.441$$
$$\Rightarrow (-1230.501, 372.381)$$

f. $R^2 = R\text{-Sq} = 45.9\%$. 45.9% of the total sample variation of the y values is explained by the model containing $x_1$ and $x_2$.

$R^2_a = R\text{-Sq(adj)} = 39.6\%$. 39.6% of the total sample variation of the $y$ values is explained by the model containing $x_1$ and $x_2$, adjusted for the sample size and the number of parameters in the model.

g. To determine if at least one of the independent variables is significant in prediction $y$, we test:

$H_0$: $\beta_1 = \beta_2 = 0$
$H_a$: At least one $\beta_i \neq 0$

From the printout, the test statistic is $F = 7.22$

Since no $\alpha$ level was given, we will choose $\alpha = .05$. The rejection region requires $\alpha = .05$ in the upper tail of the $F$-distribution with $\nu_1 = k = 2$ and $\nu_2 = n - (k + 1) = 20 - (2 + 1) = 17$. From Table VIII, Appendix B, $F_{.05} = 3.59$. The rejection region is $F > 3.59$.

Since the observed value of the test statistic falls in the rejection region ($F = 7.22 > 3.59$), $H_0$ is rejected. There is sufficient evidence to indicate at least one of the variables, $x_1$ or $x_2$, is significant in predicting $y$ at $\alpha = .05$.

h. The observed significance level of the test is $p$-value = 0.005. Since the $p$-value is so small, we will reject $H_0$ for most reasonable values of $\alpha$. There is sufficient evidence to indicate at least one of the variables, $x_1$ or $x_2$, is significant in predicting $y$ at $\alpha$ greater than 0.005.

2) Sincich, 12.72

a. Let $x_1 = \begin{cases} 1 \text{ if race is black} \\ 0 \text{ otherwise} \end{cases}$

Let $x_2 = \begin{cases} 1 \text{ if availability is high} \\ 0 \text{ otherwise} \end{cases}$

Let $x_3 = \begin{cases} 1 \text{ if position is quarterback} \\ 0 \text{ otherwise} \end{cases}$

Let $x_4 = \begin{cases} 1 \text{ if position is running back} \\ 0 \text{ otherwise} \end{cases}$

Let $x_5 = \begin{cases} 1 \text{ if position is wide receiver} \\ 0 \text{ otherwise} \end{cases}$

Let $x_6 = \begin{cases} 1 \text{ if position is tight end} \\ 0 \text{ otherwise} \end{cases}$

Let $x_7 = \begin{cases} 1 \text{ if position is defensive lineman} \\ 0 \text{ otherwise} \end{cases}$

Let $x_8 = \begin{cases} 1 \text{ if position is linebacker} \\ 0 \text{ otherwise} \end{cases}$

Let $x_9 = \begin{cases} 1 \text{ if position is defensive back} \\ 0 \text{ otherwise} \end{cases}$

b. The model is: $E(y) = \beta_0 + \beta_1 x_1$
$\beta_0$ = mean price for race black
$\beta_1$ = difference in mean price between races white and black

c. The model is: $E(y) = \beta_0 + \beta_2 x_2$
$\beta_0$ = mean price for card availability low
$\beta_2$ = difference in mean price between card availabilities high and low

d. The model is: $E(y) = \beta_0 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9$
$\beta_0$ = mean price for position offensive lineman
$\beta_3$ = difference in mean price between player positions quarterback and offensive lineman
$\beta_4$ = difference in mean price between player positions running back and offensive lineman
$\beta_5$ = difference in mean price between player positions wide receiver and offensive lineman
$\beta_6$ = difference in mean price between player positions tight end and offensive lineman
$\beta_7$ = difference in mean price between player positions defensive lineman and offensive lineman
$\beta_8$ = difference in mean price between player positions linebacker and offensive lineman
$\beta_9$ = difference in mean price between player positions defensive back and offensive lineman

3) Sincich, 12.74

a. The model would be: $E(y) = \beta_0 + \beta_1 x$

b. $\beta_0$ = mean relative optimism for analysts who worked for sell-side firms
$\beta_1$ = difference in mean relative optimism for analysts who worked for buy-side and sell-side firms

c. Yes.

d. Yes. If the buy-side analysts are less optimistic, then their estimates will be smaller than the sell-side estimates. Thus, the estimate of $\beta_1$ will be negative.

4) Sincich, 12.78 parts (a)-(d)

a. Let $x_1 = \begin{cases} 1 \text{ if study group complete solution} \\ 0 \text{ otherwise} \end{cases}$ Let $x_2 = \begin{cases} 1 \text{ if study group check figures} \\ 0 \text{ otherwise} \end{cases}$

A possible model would be: $E(y) = \beta_o + \beta_1 x_1 + \beta_2 x_2$

b. The difference between the mean knowledge gains of students in the "completed solution" and "no help groups" would be $\beta_1$.

c. Using MINITAB, the results are:

**Regression Analysis: IMPROVE versus X1, X2**

```
The regression equation is
IMPROVE = 2.43 - 0.483 X1 + 0.287 X2


Predictor      Coef  SE Coef      T      P
Constant     2.4333   0.4941   4.92  0.000
X1          -0.4833   0.7813  -0.62  0.538
X2           0.2867   0.7329   0.39  0.697


S = 2.70636   R-Sq = 1.2%   R-Sq(adj) = 0.0%


Analysis of Variance

Source          DF       SS     MS     F      P
Regression       2    6.643  3.322  0.45  0.637
Residual Error  72  527.357  7.324
Total           74  534.000


Source  DF  Seq SS
X1       1   5.523
X2       1   1.121
```

The least squares prediction equation is: $\hat{y} = 2.4333 - .4833x_1 + .2867x_2$

d. To determine if the model is useful, we test:

$H_o$: $\beta_1 = \beta_2 = 0$
$H_a$: At least one $\beta_i \neq 0$

From the printout, the test statistic is $F = .45$ and the $p$-value is $p = .637$. Since the p-value is not less than $\alpha$ ($p = .637 \not< .05$), $H_0$ is not rejected. There is insufficient to indicate that the model was useful at $\alpha = .05$..

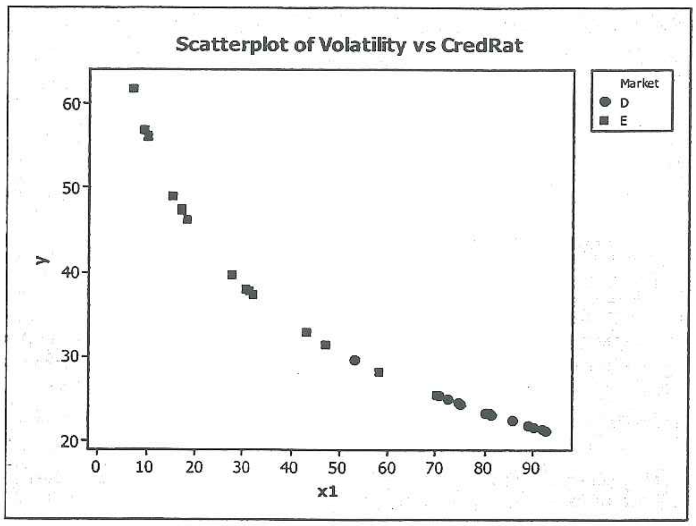
5) Sincich, 12.96

a. Let $x_2 = \begin{cases} 1 & \text{if Developing} \\ 0 & \text{otherwise} \end{cases}$

The model would be:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

b. Using MINITAB, the plot of the data is:




From the plot, it appears that the model is appropriate. The two lines appear to have different slopes.

c. Using MINITAB, the output is:

**Regression Analysis: y versus x1, x2, x1x2**

```
The regression equation is
y = 58.8 - 0.557 x1 - 18.7 x2 + 0.354 x1x2


Predictor       Coef  SE Coef       T      P
Constant      58.786    1.217   48.30  0.000
x1          -0.55743  0.03669  -15.19  0.000
x2           -18.718    5.572   -3.36  0.002
x1x2         0.35368  0.07615    4.64  0.000


S = 2.66123   R-Sq = 96.1%   R-Sq(adj) = 95.7%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       3  4596.5  1532.2  216.34  0.000
Residual Error  26   184.1     7.1
Total           29  4780.6


Source  DF  Seq SS
x1       1  4388.0
x2       1    55.7
x1x2     1   152.8
```

The fitted regression model is:

$$\hat{y} = 58.786 - .557x_1 - 18.718x_2 + .354x_1x_2$$

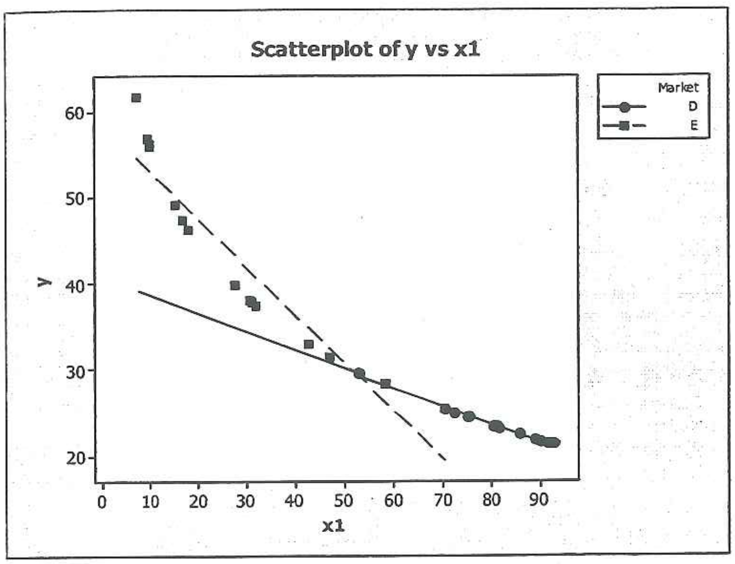
For the emerging countries, $x_2 = 0$. The fitted model is:

$$\hat{y} = 58.786 - .557x_1 - 18.718(0) + .354x_1(0) = 58.786 - .557x_1$$

For the developed countries, $x_2 = 1$. The fitted model is:

$$\hat{y} = 58.786 - .557x_1 - 18.718(1) + .354x_1(1) = 40.068 - .203x_1$$

d. The plot of the fitted lines is:




e. To determine if the slope of the linear relationship between volatility and credit rating depends on market type, we test:

$H_0$: $\beta_3 = 0$
$H_a$: $\beta_3 \neq 0$

The test statistic is $t = 4.64$.

The $p$-value is 0.000. Since the $p$-value is less than $\alpha = .01$, $H_0$ is rejected. There is sufficient evidence to indicate that the slope of the linear relationship between volatility and credit rating depends on market type at $\alpha = .01$.

6) Gesell.CSV

a.We would expect the score and age to be negatively correlated.

b. Yes, somehow

c. score = 110 - 1.13 * age

d.

H0: beta1 = 0

Ha: beta1 ~= 0

p-value = 0.002

p-value < 0.01 => H0 is rejected. There is sufficient evidence that Score is related to Age, at alpha = 0.01. The p-value does not establish directionality since this is a two-sided test.

e. R^2 = 0.41 => 41% of the variations in Score are explained by Age.

f. The point (Age, Score) = (42, 57) has the highest leverage.

Leverage = .6516, Cook'sD = .6781; leverage large, and Cook’s distance is moderate.

p-value increased to .149, R^2 decreased to 11.2%

Removing the leverage point shows there is no longer evidence of a linear relationship between Score and Age.

h. Since the point had a strong influence on the fit, we are justified in treating it separately.