Review 1 Solutions

1a) $A = \bar{Y} - B\bar{X}$

Using sum of square of residuals should be minimized to obtain the best fitted model :-

$$RSS = S(A,B) = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y})^2 = \sum_{i=1}^{n} Y_i - A - BX_i)^2$$

Partial derivative with respect to A, set to zero for roots

$$\frac{\partial S(A,B)}{\partial (A)} = \sum (-1)(2)(Y_i - A - BX_i) = 0$$

$$\sum_{i=1}^{n} (Y_i - A - BX_i) = 0$$

$$= \sum_{i=1}^{n} Y_i - nA - B\sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} Y_i - B\sum_{i=1}^{n} X_i = nA$$

Divide by n

$$\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{B}{n}\sum_{i=1}^{n} X_i = A$$

$$\bar{Y} - B\bar{X} = A \quad \text{PROVED}$$

b) $\sum Y_i = \sum \hat{Y}_i$

As a consequence of

$\sum_{i=1}^{\hat{}} E_i = 0 \quad \sum_{i=1}^{\hat{}} (Y_i - \hat{Y}) = 0$

$\Rightarrow \sum Y_i = \sum \hat{Y}_i$ PROVED

c) $RegSS = B^2 S_{xx}$

$RegSS = \sum_{i=1}^{\hat{}} (\hat{Y}_i - \bar{Y})^2$

$= \sum_{i=1}^{\hat{}} \left((A + BX_i) - (A + B\bar{X}) \right)^2$

$= \sum_{i=1}^{\hat{}} \left(B(X_i - \bar{X}) \right)^2$

$RegSS = \sum_{i=1}^{\hat{}} B^2 (X_i - \bar{X})^2$

$RegSS = B^2 S_{xx}$ PROVED

CS STAT

2a)

	X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$\hat{Y}_i$	$Y_i - \hat{Y}_i$
	1	1	1	1	1	.6	.4
	2	1	4	1	2	+1.3	-.3
	3	2	9	4	6	2.0	0
	4	2	16	4	8	2.7	-.7
	5	4	25	16	20	3.4	.6
Σ	15	10	55	26	37		

$$B = \frac{\sum_{i=1} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1} X_i^2 - n\bar{X}^2} = \frac{37 - 5(3)(2)}{55 - 5(3)^2}$$

$$B = .7$$

$$A = \bar{Y} - B\bar{X} = 2 - (.7)(3) = -.10$$

Fitted Equation $\hat{Y} = -.10 + .7X$

b) $RSS = \sum (Y_i - \hat{Y}_i)^2 = 1.1$

c) A and B do not make sense in this case study

d) $H_0: \beta = 0 \quad H_1: \beta \neq 0$

$$RMS = \frac{RSS}{n-2} = \frac{1.1}{3} = \frac{1.1}{3} = .3666$$

$$t = \frac{B}{\sqrt{\frac{RMS}{S_{xx}}}} = \frac{.7}{\sqrt{\frac{.3666}{10}}} = 3.356$$

t_c with $\alpha = .025$, dof $= 3 = 3.18$

$H_0 : \beta_1 = 0$ REJECTED

e) ANOVA TABLE

Sources of Variation	Sum of Squares	Degree of Freedom	Mean Square	F
Explained Regression	Reg SS $= B^2 S_{xx}$ $(.7)^2 10 = 4.9$	1	RegMS $= \frac{\text{Reg SS}}{1}$ $= 4.9$	$F = \frac{\text{RegMS}}{\text{RMS}}$ $= \frac{4.9}{.367}$ $= 13.6$
UnExplained Residual	RSS 1.1	3	RMS $= \frac{\text{RSS}}{3}$ $= 1.1/3 = .367$	
Total	6.0			

$F_{.05,1,3} = 10.13$ F critical

$F_t > F_{critical}$ ∴ Reject H_0

g) This corraborates our result in d)

h) $t^2 = (3.18)^2 = 10.1124$, $F = 10.13$

These are approximately equal

i) $R^2 = \frac{RegSS}{TSS} = \frac{4.9}{6.0} = .81\bar{6} \approx .82$

R^2 informs us that 82% of variation in # number of STAT courses can be attributed to number of CS courses taken

j) Confidence Interval for β

looking at t table: $t_{\alpha/2, n-2} = t_{.025, 3} = 3.182$

$$B - t_{\alpha/2, n-2}\sqrt{\frac{RMS}{S_{xx}}} \leq \beta \leq B + t_{\alpha/2, n-2}\sqrt{\frac{RMS}{S_{xx}}}$$

$$.7 - (3.182)\sqrt{\frac{.367}{10}} \leq \beta \leq .7 + (3.182)\sqrt{\frac{.367}{10}}$$

$$Pr\{.1 \leq \beta_1 \leq 1.3\} = .95.$$

3a) Hypothesis ANOVA

$H_0: \beta_1 = \beta_2 = 0$ H_1: At least one of the slope is not zero.

P-value for F test $= 4.249e^{-09}$

This is significant.

Rejecting H_0. At least one of the slopes are non zero.

At least one of the explanatory variable poverty, minority is significantly effecting crime

b) Hypothesis t test for crime

$H_0: \beta_1 = 0$ $H_1: \beta_1 \neq 0$

P-value for poverty $= .072$

This value is non significant

So we cannot Reject H_0,

H_1 is rejected.

c) Hypothesis t test for minority

$H_0: \beta_2 = 0 \quad H_1: \beta_2 \neq 0$

p-value for minority $= 6.86e^{-08}$

This is significant so we reject H_0 and conclude that minority effects crime.

d) Look at the Residuals, the min and max value is far apart so data are not very closely aligned to the fitted line

Look at the Residual Standard Error $= 18.62$ which shows that the variability of residuals is high since residual standard error is standard deviation of residuals

e) Since poverty p-value is not significant we will only minority whose p-value is significant.

44.05% of the variation in creme can be attributed to minority

f) we would use adjusted R^2 to ensure eliminating inflation of R^2.

g) $\hat{Y} = 58.6701 + 1.1949X$

$\hat{Y}$ ↓ CRIME

X ↓ MINORITY