11.1 Given $n = 30$,

$\bar{x} = 126$, $\bar{y} = 0.04$

$s_x = 35$, $s_y = 0.01$ and $r = 0.86$

Compute the least squares estimates,

$$b_1 = r\left(\frac{s_y}{s_x}\right) = 0.86\left(\frac{0.01}{35}\right) = 0.000246$$

$$b_0 = \bar{y} - b_1\bar{x} = 0.04 - (0.000246)(126) = 0.009.$$

The ~~fittest~~ regression line has an equation

$$\boxed{y = 0.009 + 0.000246x}$$

The time it takes to transmit a 400kbyte file is predicted as

$$\hat{y}_n = 0.009 + 0.000246x = 0.009 + (0.00246)400$$

$$= \boxed{0.107 \text{ seconds}}$$

11.2 Here $n = 75$, $\bar{x} = 32.2$, $s_x^2 = 6.4$, $\bar{y} = 8.4$, $s_y^2 = 2.8$ and $s_{xy} = 3.6$.

(a) Compute the least squares Estimates

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{3.6}{6.4} = 0.5625$$

$$b_0 = \bar{y} - b_1\bar{x} = 8.4 - (0.5625)(32.2) = -9.7125$$

The Sample regression line is

$$\boxed{y = -9.7125 + 0.5625x}$$

(b) Compute the sums of squares,

$$SS_{TOT} = (n-1)s_y^2 = (75-1)(2.8) = 207.2$$

$$SS_{REG} = b_1^2 S_{xx} = b_1^2 s_x^2 (n-1) = (0.5625)^2 6.4(75-1) = 149.85$$

$$SS_{ERR} = SS_{TOT} - SS_{REG} = 57.35$$

Also, compute degrees of freedom,

$$df_{TOT} = n-1 = 74, \quad df_{REG} = 1 \text{ \& } df_{ERR} = n-2 = 73.$$

and the mean squares

$$MS_{REG} = \frac{SS_{REG}}{df_{REG}} = 149.85, \quad MS_{ERR} = \frac{SS_{ERR}}{df_{ERR}}$$

$$= \frac{57.35}{73} = 0.7856$$

finally, we compute the F-ratio

$$F = \frac{MS_{REG}}{MS_{ERR}} = 190.75$$

and from the table for $df_{REG} = 1$ and 73 d.f this is significant at the 0.1% level.

Complete the ANOVA table:

Source	Sum of squares	df	Mean squares	F
Model	149.85	1	149.5	190.75
Error	57.35	73	0.7856	
Total	207.2	74		

predictor X can explain

$$R^2 = SS_{REG} / SS_{TOT} = \boxed{0.7232 \text{ or } 72.32\%}$$

of the total variation.

(c) The 99% CI for β_1 is

$$b_1 \pm t_{\alpha/2} \frac{s}{\sqrt{S_{xx}}} = b_1 \pm t_{0.005} \frac{\sqrt{MS_{ERR}}}{\sqrt{(n-1)S_x^2}}$$

$$= 0.5625 \pm \frac{(2.648)\sqrt{0.7856}}{\sqrt{74(6.4)}}$$

$$= 0.5625 \pm 0.1078 \text{ or } [0.4547, 0.6703]$$

where $t_{0.05}$ is obtained from the table with 73 d.f

This interval does not contain 0 and therefore the slope is significant at 1% level of significance.

11.5 (a)

$$X = \begin{pmatrix} 1 & 3 & 0 \\ 1 & 4 & 0 \\ 1 & 5 & 1 \\ 1 & 6 & 0 \\ 1 & 7 & 1 \\ 1 & 8 & 1 \\ 1 & 9 & 1 \\ 1 & 10 & 1 \\ 1 & 11 & 0 \\ 1 & 12 & 1 \\ 1 & 13 & 1 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 17 \\ 23 \\ 31 \\ 29 \\ 33 \\ 39 \\ 39 \\ 40 \\ 41 \\ 44 \\ 47 \end{pmatrix}$$

The 3rd column of the design matrix X is the dummy variable Z that equals 1 if the company reports profit and 0 otherwise. In the 2nd column of X, we again defined

$X = \text{year} - 2000,$

to simplify the calculations,

we then compute

$$X^T X = \begin{pmatrix} 11 & 88 & 7 \\ 88 & 814 & 64 \\ 7 & 64 & 7 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.6742 & -0.0707 & -0.0278 \\ -0.0707 & 0.0118 & -0.0370 \\ -0.0278 & -0.0370 & 0.5093 \end{pmatrix}$$

and

$$b = (X^T X)^{-1} X^T y = \begin{pmatrix} 13.3586 \\ 2.3569 \\ 4.0926 \end{pmatrix}$$

$$\boxed{b_0 = 13.3586, \; b_1 = 2.3569, \; b_2 = 4.0926}$$

(b) The estimated regression equation is

$$\hat{y} = 13.3586 + 2.3569x + 4.0296z$$

for a company reporting profit $(z_n = 1)$ in 2015 (that is $x_* = 15$), we predict the investment amount as

$$\hat{y}_* = 13.3586 + 2.3569(15) + 4.0926(1)$$

$$= \boxed{52,8047 \text{ thousand dollars}}$$

(c) The slope β_2 is the change in the response variable when the dummy variable z changes from 0 to 1. Thus, if the company reports a loss during year 2007 instead of a gain, it's expected amount reduces by β_2. Our prediction will decrease by $b_2 = 4.0926$ thousand dollars.

(a) The total sum of squares

$$SS_{TOT} = S_{yy} = \Sigma (Y_i - \bar{Y})^2 = 841.64$$

The Error sum of squares can be computed, say, by filling the table.

Y_i	$\hat{Y}_i$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
17	20.4293	-3.4293	11.7600
23	22.7862	0.2138	0.0457
31	29.2357	1.7643	3.1128
29	27.5000	1.5000	2.2500
33	33.9495	-0.9495	0.9015
39	36.3064	2.6936	7.2555
39	41.0202	0.3367	0.1134
40	39.2845	-1.0262	1.0468
41	38.6633	1.7155	2.9429
44	45.7340	-1.7340	3.0068
47	48.0909	-1.0909	1.1901

Alternatively, one can multiply matrices

$$SS_{ERR} = (y - \hat{y})^T (y - \hat{y}) = (y - Xb)^T (y - Xb)$$

Then

$$SS_{ERR} = \sum_i (Y_i - \hat{Y}_i)^2 = 33.62$$

and $SS_{REG} = SS_{TOT} - SS_{ERR} = 841.64 - 33.62$

$$= 808.02$$

Complete the ANOVA table:

Source	Sum of squares	df	Mean squares	F
Model	808.02	2	404.01	96.14
Error	33.62	8	4.20	
Total	841.64	10		

Comparing $F_{obs} = 96.14$ against table with 2 and 8 df, we find that the model is significant at the 0.1% level (in fact, it's p-value is less than 0.0001)

(e) $SS_{ERR}(\text{Reduced}) = (MS_{ERR})(df_{ERR})$

$= (9)(7.39) = 66.51 \; (9\,d.f)$

This Error Sum of Squares is obtained from the reduced model where investment is predicted based on the year only.

$SS_{ERR}(\text{Full}) = 33.62 \; (8\,d.f)$

Hence, the new variable explains additional

$$\frac{SS_{EX}}{SS_{TOT}} \cdot 100\% = \frac{6.51 - 33.62}{841.64} \cdot 100\%$$

$= 3.9\%$ of the total variation

Significance of the new dummy variable (reporting profit) in addition to the time trend is tested by the partial F-statistic

$$F_{obs} = \frac{SS_{EX} / df_{EX}}{MS_{ERR}(Full)} = \frac{(SS_{ERR}(Reduced) - SS_{ERR}(Full)/(9-8))}{MS_{ERR}(Full)}$$

$$= \frac{6.51 - 33.62}{4.20} = \underline{7.83}$$

from the table with 1 and 8 d.f.. addition of the new variable is significant at the 2.5% but not at the 1% level. The p-value of this test is between 0.01 and 0.025.

<u>11.9</u>

(a) $$R^2_{adj} = 1 - \frac{SS_{ERR} / df_{ERR}}{SS_{TOT} / df_{TOT}}$$

$$\Rightarrow 1 - \frac{617.1/5}{1452/6} = 0.49$$

(b) $$MS_{REG} = 834.9/1 = 834.9$$

$$MS_{ERR} = 617.1/5 = 123.42$$

$$F\text{-ratio} = \frac{MS_{REG}}{MS_{ERR}} = \frac{834.9}{123.42} = 6.79$$

<u>Special problem :</u>

$$E(LF(y)) = x^3 \sin^2\theta\, \beta^2 + y^2 \cos^3\theta\, c^5 + z^5 \sin\theta \cos\theta$$

$$\frac{\partial E}{\partial x} = 3x^2 \sin^2\theta\, \beta^2$$

$$\frac{\partial E}{\partial y} = 2y \cos^3\theta \, c^5$$

$$\frac{\partial E}{\partial z} = 5 z^4 \sin\theta \cos\theta$$

$$\frac{\partial E}{\partial \theta} = x^3 \cdot 2\sin\theta\cos\theta B^2 + y^2 3\cos^2\theta$$

$$= (-\sin\theta) c^5 + z^5 \cos^2\theta + z^5(-\sin^2\theta)$$

$$= 2x^3 \sin\theta\cos\theta B^2 + 3y^2 \sin\theta \cos^2\theta c^5 - z^5(\cos 2\theta)$$

$$= 2x^3 \sin\theta \cos\theta B^2 - 3y^2 \sin\theta \cos^2\theta c^5 - z^5 \cos 2\theta$$

$$\sigma^2_{F(x)} = 3x^2 \sin^2\theta B^2 \sigma_x^2 + 2y \cos^3\theta \, c^5 \sigma_y^2 + 5 \cdot z^4 \sin\theta\cos\theta \, \sigma_z^2$$
$$+ (x^3 \sin^2\theta B_z^2 - 3y^2 \sin\theta \cos^2\theta c^5 + z^5 \cos^2\theta) \sigma_\theta^2$$