

## MATH 204 - EXERCISES 3- SOLUTIONS

1. (a) The derivative can be obtained using the chain rule; the derivative of the term

$$(y_i - \beta_0 - \beta_1 x_i)^2$$

with respect to  $\beta_0$  is

$$2 \times (y_i - \beta_0 - \beta_1 x_i) \times (-1) = -2(y_i - \beta_0 - \beta_1 x_i)$$

whereas the derivative with respect to  $\beta_1$  is

$$2 \times (y_i - \beta_0 - \beta_1 x_i) \times (-x_i) = -2(y_i - \beta_0 - \beta_1 x_i)x_i$$

so summing over  $i = 1, \dots, n$  and equating to zero yields

$$\text{For } \beta_0 : \sum_{i=1}^n -(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\text{For } \beta_1 : \sum_{i=1}^n -(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

after dividing both sides of the equation by 2. Note that we can also remove the negative sign, yielding the simultaneous equations

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (1)$$

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)x_i = 0 \quad (2)$$

These are two simultaneous equations in two variables, so they can be solved using standard methods to give  $\hat{\beta}_0$  and  $\hat{\beta}_1$ . From (1), by summing terms, we obtain

$$\sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i = 0$$

or

$$S_y - n\hat{\beta}_0 - \hat{\beta}_1 S_x = 0 \quad (3)$$

where

$$S_x = \sum_{i=1}^n x_i \quad S_y = \sum_{i=1}^n y_i$$

and from (2), using similar arguments

$$S_{xy} - \hat{\beta}_0 S_x - \hat{\beta}_1 S_{xx} = 0 \quad (4)$$

where

$$S_{xy} = \sum_{i=1}^n x_i y_i \quad S_{xx} = \sum_{i=1}^n x_i^2.$$

From (3), we obtain

$$\hat{\beta}_0 = \frac{1}{n} (S_y - \hat{\beta}_1 S_x) = (\bar{y} - \hat{\beta}_1 \bar{x}) \quad (5)$$

and thus from (4), substituting in this expression for  $\hat{\beta}_0$ , we obtain

$$S_{xy} - (\bar{y} - \hat{\beta}_1 \bar{x})S_x - \hat{\beta}_1 S_{xx} = 0 \quad \therefore \quad S_{xy} - \bar{y}S_x + \hat{\beta}_1(\bar{x}S_x - S_{xx}) = 0$$

and hence

$$\hat{\beta}_1 = \frac{S_{xy} - \bar{y}S_x}{S_{xx} - \bar{x}S_x}$$

But

$$S_{xy} - \bar{y}S_x = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}.$$

Now

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \bar{x} - \sum_{i=1}^n x_i \bar{y} + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{aligned}$$

Therefore

$$S_{xy} - \bar{y}S_x = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = SS_{xy}.$$

Similarly

$$S_{xx} - \bar{x}S_x = \sum_{i=1}^n (x_i - \bar{x})^2 = SS_{xx}$$

Thus

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

and by back substitution into (5) we obtain the final expression for  $\hat{\beta}_0$ .

(b) From (1), we have by construction that

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

But by definition

$$\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

and thus

$$\sum_{i=1}^n \hat{e}_i = 0.$$

2. For the Longley data, here is the SPSS analysis:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.991 <sup>a</sup>	.982	.981	13.703405

a. Predictors: (Constant), Population (millions)

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-1275.210	59.826		-21.315	.000	-1403.523	-1146.897
	Population (millions)	14.162	.509	.991	27.842	.000	13.071	15.253

a. Dependent Variable: GNP (billions of dollars)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	145561.3	1	145561.339	775.156	.000 <sup>a</sup>
	Residual	2628.966	14	187.783		
	Total	148190.3	15			

a. Predictors: (Constant), Population (millions)

b. Dependent Variable: GNP (billions of dollars)

Thus

- the estimates (standard errors) for  $\hat{\beta}_0$  and  $\hat{\beta}_1$  are

$$-1275.210(59.826) \quad 14.161(0.509)$$

respectively.

- the 95 % confidence intervals for both parameters exclude zero, so both parameters are significantly different from zero (at the  $\alpha = 0.05$  significance level)
- the correlation between  $x$  and  $y$  is 0.991.
- the  $R^2$  statistic is 0.982
- the ANOVA-F test yields an  $F$  statistic equal to 775.156, and a  $p$ -value of less than 0.001.

Thus there is a significant linear relationship between the  $x$  and  $y$  variables.