

## Quiz 1 - solutions

1. (50pts) Jaw width,  $y$ , and body length,  $x$ , were measured on a random sample of  $n = 41$  sharks. Summary statistics are given below:

Variable	mean	std deviation	Corrected sum of squares
$y$ : Jaw width (in <i>in</i> )	$\bar{y} = 14.7$	$s_y = 3$	$S_{yy} = 360$
$x$ : Body length (in <i>ft</i> )	$\bar{x} = 15$	$s_x = 2$	$S_{xx} = 160$

The least squares regression line is  $\hat{y} = 1.2 + 0.9x$

- (a) (8pts) Is the sample correlation coefficient  $r_{yx}$  positive or negative?

positive, since  $\hat{\beta}_1 > 0$

- (b) (8pts) Report the sample correlation coefficient,  $r_{yx}$ .

Since  $\hat{\beta}_1 = r_{yx} \frac{s_y}{s_x}$ ,  $r_{yx} = 0.9 \frac{2}{3} = 0.6$

- (c) (8pts) Estimate mean jaw width among sharks with body length  $x = 15$  *ft*.

$$\hat{\mu}(x = 15) = 1.2 + 0.9(15) = 14.7$$

- (d) (8pts) Estimate the difference between the mean in part (c) and mean jaw width among sharks with body length  $x = 20$  *ft*.

$$5\hat{\beta}_1 = 4.5$$

- (e) (8pts) Report and interpret (very briefly) the error mean square,  $MS[E]$ .

$$MS[E] = \frac{SS[E]}{n - 2} = \frac{(1 - r^2)SS[Total]}{39} = 5.9$$

This is  $\text{Var}(Y|X = x)$ .

- (f) (10pts) Report standard errors for the estimates in parts (c) and (d).

$$\begin{aligned} SE(\hat{\mu}(x = 15)) &= \sqrt{\frac{MS[E]}{n}} \\ &= 0.38 \\ SE(5\hat{\beta}_1) &= 5\sqrt{\frac{MS[E]}{S_{xx}}} \\ &= 0.96 \end{aligned}$$

2. (50pts) The weight ( $y$ , in  $g$ ), age ( $x_1$ , in  $yr$ s) and length ( $x_2$ , in  $mm$ ) are measured on a random sample of  $n = 17$  whitefish. Consider several regression models for describing the association between weight and the predictors, *age* and *length*:

- Model 0:  $\mu(x_1, x_2) = \beta_0$
- Model 1:  $\mu(\text{age}, \text{length}) = \beta_0 + \beta_1 \text{ age}$
- Model 2:  $\mu(\text{age}, \text{length}) = \beta_0 + \beta_2 \text{ length}$
- Model 3:  $\mu(\text{age}, \text{length}) = \beta_0 + \beta_1 \text{ age} + \beta_2 \text{ length}$

(a) (20pts) Using the output at the end of the exam, report appropriate  $F$ -ratios and corresponding degrees of freedom for each of the model comparisons below. (Some may be found directly in the output, some require computation.)

i. Model 0 versus Model 1

$$\begin{aligned} R(\beta_1|\beta_0) &= 59033 \\ F &= \frac{59033}{(305695 - 59033)/(n - 2)} \\ &= 3.59(df = 1, 15) \end{aligned}$$

ii. Model 0 versus Model 2

$$\begin{aligned} R(\beta_2|\beta_0) &= R(\beta_1, \beta_2|\beta_0) - R(\beta_1|\beta_2, \beta_0) \\ &= 181364 - 135 \\ &= 181229 \\ F &= \frac{181229}{(305695 - 181229)/(n - 2)} \\ &= 21.84(df = 1, 15) \end{aligned}$$

iii. Model 0 versus Model 3 from output,  $F = 10.2$ ,  $df = 2, 14$

iv. Model 1 versus Model 3 from output,  $F = 13.8$ ,  $df = 1, 14$

v. Model 2 versus Model 3 from output,  $F = 0.02$ ,  $df = 1, 14$

(b) (15pts) Which model is best?

i. Explain very briefly.

Model 2 is best. No evidence anywhere that weight depends on age.

ii. Use the full model to estimate the mean weight among whitefish aged 10 yrs with length 380 mm

$$\hat{\mu}(x_1 = 10, x_2 = 380) = -510.9 - 1.11(10) + 3.06 * (380) = 640.8$$

iii. Report the squared correlation between the observed weights and those predicted by the model you chose as best. The squared correlation using model 2 is just  $r^2$  from the model, or

$$r^2 = \frac{SS[Model]}{SS[Tot]} = \frac{R(\beta_2|\beta_0)}{SS[Tot]} = \frac{181229}{305695} = .59$$

(c) (15pts) Below, write what should go in the blanks in the output labelled AAA, BBB and CCC: Here, use the fact that the test for the partial slope for age is equivalent to a comparison of models 1 and 3 so that the  $p$ -values should agree ( $p = 0.9038$ ) and same for the  $t$  statistic:  $t = -\sqrt{F} = -\sqrt{0.015} = -0.12$  which implies  $SE = \hat{\beta}_1/t = -1.11/-0.12 = 9.2$  (If output value of  $F = 0.02$  is used, there is some rounding error)

/\*\*\*\*\* BEGIN OUTPUT \*\*\*\*\*/

The SAS System  
The GLM Procedure

1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	181364.0227	90682.0114	10.21	0.0018
Error	14	124331.0361	8880.7883		
Corrected Total	16	305695.0588			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
age	1	59033.9029	59033.9029	6.65	0.0219
length	1	122330.1198	122330.1198	13.77	0.0023

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age	1	134.5398	134.5398	0.02	0.9038
length	1	122330.1198	122330.1198	13.77	0.0023

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	-510.8786593	286.0797763	-1.79	0.0958
age	-1.1126251	AAA_____	BBB_____	CCC_____
length	3.0632745	0.8253630	3.71	0.0023