Alanna Hazlett
2/14/24

M03 HW

2.) Suppose that for $n = 6$ students, we want to predict their scores on the second quiz using scores from the first quiz. The estimated regression line is $\hat{y} = 20 + 0.8x$.

a.) For each individual observation calculate its predicted score on the second quiz $\hat{y}_i$ and the residual e_i

	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6
X_i	70	75	80	80	85	90
y_i	75	82	80	86	90	91
$\hat{y}_i$	76	80	84	84	88	92
e_i	-1	2	-4	2	2	-1

$e_i = y_i - \hat{y}_i$

Col 1: $\hat{y}_i = 20 + 0.8(70) = 76$
$e_i = 75 - 76 = -1$

Col 2: $\hat{y}_i = 20 + 0.8(75) = 80$
$e_i = 82 - 80 = 2$

Col 3: $\hat{y}_i = 20 + 0.8(80) = 84$
$e_i = 80 - 84 = -4$

Col 4: $\hat{y}_i = 20 + 0.8(80) = 84$
$e_i = 86 - 84 = 2$

Col 5: $\hat{y}_i = 20 + 0.8(85) = 88$
$e_i = 90 - 88 = 2$

Col 6: $\hat{y}_i = 20 + 0.8(90) = 92$
$e_i = 91 - 92 = -1$

b.) Complete ANOVA table for this dataset.

	DF	SS	MS	F-Stat	p-value
Regression	1	190	190	25.3	0.0099
Residual	4	30	7.5	***	***
Total	5	220	***	***	***

$\bar{y} = (75 + 82 + 80 + 86 + 90 + 91) \div 6 = 84$

$$SS_R = \sum_{i=1}^{6} (\hat{y}_i - \bar{y})^2 = (75-84)^2 + (82-84)^2 + (80-84)^2 + (86-84)^2 + (90-84)^2 + (91-84)^2$$
$$= 81 + 4 + 16 + 4 + 36 + 49$$
$$\boxed{SS_R = 190}$$

$$SS_{res} = \sum_{i=1}^{6} (y_i - \hat{y}_i)^2 = (75-76)^2 + (82-80)^2 + (80-84)^2 + (86-84)^2 + (90-88)^2 + (91-92)^2$$
$$= 1 + 4 + 16 + 4 + 4 + 1$$
$$\boxed{SS_{res} = 30}$$

$$SS_T = SS_R + SS_{res}$$
$$= 190 + 30$$
$$\boxed{SS_T = 220}$$

$MS_R = SS_R / df_R \quad = 190/1 \quad = \boxed{190 = MS_R}$

$MS_{res} = SS_{res} / df_{res} \quad = 30/4 \quad = \boxed{7.5 = MS_{res}}$

$F \text{ statistic} = MS_R / MS_{res} \quad = 190/7.5 = \boxed{25.\overline{3} = F \text{ statistic}}$

c.) Calculate the sample estimate of the variance σ^2 for the regression model.

$s^2 = MS_{res} = \boxed{7.5 = s^2}$

d.) What is value of R^2?

$R^2 = \frac{SS_R}{SS_T} = \frac{190}{220} = \boxed{0.8636 \doteq R^2}$

This is a strong/good fit to our model.

e.) Carry out ANOVA F test. What is an appropriate conclusion?

$H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0$

$F \text{ statistic} = 25.\overline{3}$ critical value $= qf(1-0.05, 1, 4)$
$= 7.708647$

F statistic > critical value,
so we reject the null hypothesis. Our data supports the alternative hypothesis of the slope being different from zero. This indicates that there is a linear association between our variables.

- alternatively our p-value $= (1 - pf(25.333, 1, 4)) \times 2$
$= 0.01463$

p-value < 0.05, so we reject the null hypothesis. Our data supports the alternative hypothesis.

Alanna Hazlett
2/15/24

MD3 HW Continued

3.) $SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ $\quad e_i = y_i - \hat{y}_i \quad y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$

$\hat{\beta}_0$ = intercept $\quad SS_{res} = \sum_{i=1}^{n} (e_i)^2 \quad e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

$\hat{\beta}_1$ = slope $\quad SS_{res} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$$\frac{\partial SS_{res}}{\partial \hat{\beta}_0} = \frac{\partial}{\partial \hat{\beta}_0}\left[\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right]$$

$$= \sum_{i=1}^{n} \left[\frac{\partial}{\partial \hat{\beta}_0} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right]$$

$$= \sum_{i=1}^{n} 2 \cdot -1 (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$\frac{\partial SS_{res}}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$0 = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

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

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$\frac{\partial SS_{res}}{\partial \hat{\beta}_1} = \frac{\partial}{\partial \hat{\beta}_1}\left[\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right] = \sum_{i=1}^{n} \left[\frac{\partial}{\partial \hat{\beta}_1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right]$$

$$= \sum_{i=1}^{n} 2 \cdot -1 \cdot x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$0 = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \qquad 0 = \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$0 = \sum_{i=1}^{n} (y_i x_i - \hat{\beta}_0 x_i - \hat{\beta}_1 (x_i^2))$$

$$= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \hat{\beta}_0 x_i - \sum_{i=1}^{n} \hat{\beta}_1 (x_i^2)$$

$$= \sum_{i=1}^{n} y_i x_i - \left(\hat{\beta}_0 \sum_{i=1}^{n} x_i\right) - \left(\hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2\right)$$

$$\hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2 = \sum_{i=1}^{n} y_i x_i - \left(\hat{\beta}_0 \sum_{i=1}^{n} x_i\right)$$

$$\hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2 = \sum_{i=1}^{n} y_i x_i - \left[(\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i\right]$$

$$\hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2 = \left(\sum_{i=1}^{n} y_i x_i\right) - \bar{y}\bar{x} + \hat{\beta}_1 (\bar{x}^2)$$

$$\hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2 - \hat{\beta}_1 (\bar{x}^2) = \left(\sum_{i=1}^{n} y_i x_i\right) - \bar{y}\bar{x}$$

$$\hat{\beta}_1 \left[\left(\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2\right] = \left(\sum_{i=1}^{n} y_i x_i\right) - \bar{y}\bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$