STAT 2509 B – Test #1

SOLUTION

//90

Q.1.:

[1] (a) the response variable, y, is price of a tire (1)

[1] (b) the explanatory variable, x, is limited tread warranty (1)

[6] (c) SLR model: $y = \beta_0 + \beta_1 x + \varepsilon$ (1), $n = 12$

assumptions: (i) x's are observed without error (1)

(ii) y's (or ε's) are independently distributed with mean $E(y) = \beta_0 + \beta_1 x$ (1) (or $E(\varepsilon) = 0$)

(iii) variance of y's (or ε's) is constant for all x's (1)

(iv) $y \sim N(E(y), \sigma^2)$ (or $\varepsilon \sim N(0, \sigma^2)$) for any value of x (1)

[5] (d) $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} =$ (1)

$= \frac{68\,050 - \frac{(780)(1036)}{12}}{51\,450 - \frac{(780)^2}{12}} = \frac{710}{750} = 0.946667$

(1) $\doteq 0.9467$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\sum y_i}{n} - \hat{\beta}_1 \left(\frac{\sum x_i}{n}\right) =$ (1)

$= \frac{1036}{12} - (0.9467)\left(\frac{780}{12}\right) = 86.33333 - (0.9467)65 =$

$= 24.79783 \doteq 24.8$ (1)

∴ the least squares fitted regr. line is given by

$\hat{y} = 24.8 + 0.9467x$ (1)

(e) [6] $s^2 = \frac{SSE}{n-2}$ (1), where $SSE = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$ (1) =

$$= \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right] - \frac{\left[\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}\right]^2}{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]} =$$

$n = 12$

$$= \left[91\,944 - \frac{(1036)^2}{12}\right] - \frac{(710)^2}{750} = 2502.667 - 672.1333 =$$ (1) ↙ given (1)

$$= 1830.533$$ (1)

$$\therefore s^2 = \frac{SSE}{n-2} = \frac{1830.533}{10} = \underline{183.0533}$$ (1)

$$\therefore s = \sqrt{s^2} = 13.52972$$

(f) [6] $H_0: \beta_1 = 0$, $H_a: \beta_1 \neq 0$ } (1) $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

-test-stat: $t = \frac{\hat{\beta}_1}{s/\sqrt{S_{xx}}}$ (1) $= \frac{0.9467}{13.52972/\sqrt{750}} =$

$$= \underline{1.916259}$$ (1)

-R.R. - we reject H_0 if $t < -t_{\alpha/2;\,n-2} = -t_{0.025;10} = -2.228$ } (1)

or $t > t_{\alpha/2;\,n-2} = t_{0.025;10} = 2.228$

(1) - since $t = 1.916259 \not> 2.228$, we do not reject H_0 (1/2) and conclude that at 5% level of sign. there is not enough evidence to say that a lin. relationship between linsulved tread warranty and price of a tire exists. (ie. no significant relationship between x & y). (1/2)

g) 95% C.I. for β_1 : $1-\alpha = 0.95$, $\alpha = 0.05$, $\alpha/2 = 0.025$

[4]

$$\beta_1 \in \left(\hat{\beta}_1 \pm t_{\alpha/2;n-2} \times S/\sqrt{S_{xx}}\right) = (1)$$

$$= \left(0.9467 \pm t_{0.025;10} \times \frac{13.52972}{\sqrt{750}}\right) = (1)$$

$$= (0.9467 \pm 2.228 \times (0.494036)) = (1)$$

$$= (0.9467 \pm 1.100711) = (-0.154011, 2.047411) =$$

$$= (-0.154, 2.047)$$

ie. We are 95% confident that in repeated (1) sampling the true value of the popuer slope β_1 would lie in the interval (-0.154, 2.047)

(h) [23]

$TSS = S_{yy} = 2502.667$ (1) (given)

$SSE = 1830.533$ (1) (calculated in part (e))

$SSR = TSS - SSE = 672.134$ (1)

$MSR = \frac{SSR}{1} = 672.134$ (1)

$MSE = \frac{SSE}{n-2} = \frac{1830.533}{10} = 183.0533$ (1)

$F = \frac{MSR}{MSE} = 3.671791$ (1)

ANOVA table:

Source of variation	d.f.	SS	MS	F
Regression	1 (1)	672.134 (1)	672.134 (1)	3.671791 (1)
Error	10 (1)	1830.533 (1)	183.0533 (1)	
Total	11 (1)	2502.667		

$H_0: \beta_1 = 0$
$H_{a1}: \beta_1 \neq 0$ } (1) $\alpha = 0.05$

- test-stat.: $F = \frac{MSR}{MSE} = 3.671791$ (1)

- R.R. - we reject H_0 if $F > F_\alpha(1; n-2) = F_{0.05}(1, 10) = 4.96$ (1)

- since $F = 3.671791 \not> 4.96$ (1), we do not reject H_0 and conclude that at 5% level of signif. there is not enough evidence to say that a lin. rel. between X & Y exists. (1)

(i) [5] $r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{710}{\sqrt{(750)(2502.667)}} = 0,5189234 \doteq 51,89\%$ (1)

ie. the limited tread warranty and price of a tire are positively correlated (related) with the strength of their relationship approx. 51.89%. (1/2)

$r^2 = \frac{SSR}{TSS} = 0.268567 \doteq 26.86\%$ (1)

ie. approx. 26.86% of the total variation in the data is explained by the regr. line (and 73.14% is due to error). I.e. model is not good. (1/2) (1)

(j) [5] 95% C.I. for $E(y)$ when $x_p = 73$

$\Rightarrow \hat{y} = 24.8 + 0.9467(73) = 93.9091$ (1)

and

$1 - \alpha = 0.95$
$\alpha = 0.05$
$\alpha/2 = 0.025$

$\Rightarrow E(y) \in \left(\hat{y} \pm t_{\alpha/2;n-2} \times s\sqrt{\frac{1}{n} + \frac{(x_p-\bar{x})^2}{S_{xx}}}\right) =$ (1)

$= \left(93.9091 \pm t_{0.025;10} \times (13.52972)\sqrt{\frac{1}{12} + \frac{(73-65)^2}{750}}\right) =$ (1)

$= (93.9091 \pm 2.228 \times (5.556527)) =$

$= (93.9091 \pm 12.37994) = (81.52916, 106.289) =$

$\doteq (81.523, 106.289)$ (1)

ie. We are 95% confident that an average price of a tire when a limited tread warranty is 73 (thousand miles) falls in interval (81.523, 106.289). (1)

(k) 95% P.I. for y when $x_p = 73$

[5] $\Rightarrow \hat{y} = 93.9091$ (1)

$y \in \left(\hat{y} \pm t_{\alpha/2;n-2} \times s\sqrt{1 + \frac{1}{n} + \frac{(x_p-\bar{x})^2}{S_{xx}}}\right) =$ (1)

$= \left(93.9091 \pm t_{0.025;10} \times (13.52972)\sqrt{1 + \frac{1}{12} + \frac{(73-65)^2}{750}}\right) =$ (1)

$= (93.9091 \pm 2.228 \times (14.62629)) =$

$= (93.9091 \pm 32.58737) = (61.32173, 126.4965)$

$\doteq (61.322, 126.496)$ (1)

ie. We are 95% confident that the price of a tire with limited tread warranty of 73 thousand miles will fall in interval (61.322, 126.496). (1)

$SSE = SSPE + SSLF$ (1)

Q.2:

(a) [5] $SSE = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 1830.533$ (1) (1)

(calculated in 1e)

$SSPE = 823.5$ (given)

$\therefore SSLF = SSE - SSPE = 1007.033$ (1) (1)

(b) [6] $y = \beta_0 + \beta_1 x + \epsilon$, $\alpha = 0.05$

H_0: model is appropriate

H_a: model is not appropriate (1)

test-stat. $F = \frac{MSLF}{MSPE} = \frac{SSLF/(n-2-\sum_i^l (n_i-1))}{SSPE/\sum_i^l (n_i-1)} =$ (1)

$= \frac{1007.033/(10-5)}{823.5/5} = \frac{201.4067}{164.7} = 1.22287$ (1)

R.R. - we reject H_0 if $F > F_2(n-2-\sum_i^l (n_i-1), \sum_i^l (n_i-1)) =$

$= F_{0.05}(5,5) = 5.05$ (1)

- since $F = 1.22287 \not> 5.05$ (1), we do not reject H_0 and conclude that at 5% level of signif. there is not enough evidence to say that the lin. model is not appropriate. (1/2) (1/2)

Q.3: (a) [3] - violation of the assumption of constant variance (1), since the residuals are increasing with x's (2)

(b) [3] • violation of the assumption of errors being normally distributed (1), since the histogram of errors is not bell-shaped (1), nor is it symmetric (it is negatively skewed) (1).

(c) [3] • violation of linearity (or independence) (1), since curve-linear pattern (2)

(d) [3] • no violations (1), since residuals are randomly scattered around their mean (ie. no pattern) (2).