STAT 2509 A
Assignment #3

SOLUTION // 40

1. [11 marks]

[10] (a) $SSE = SSPE + SSLF$

$$SSE = \text{[1/2]}\ S_{yy} - \frac{S_{xy}^2}{S_{xx}} = \left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}\right]^2}{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]} =$$

[1/2]

$$= \text{[1/2]} \left[10\,742 - \frac{(344)^2}{12}\right] - \frac{\left[969 - \frac{(30)(344)}{12}\right]^2}{\left[90 - \frac{(30)^2}{12}\right]} = 880.6666667 - \frac{(109)^2}{15} = \underline{\underline{88.6}}\ \text{[1/2]}$$

[1/2]

[1/2] $SSPE = \sum_i \sum_j \left(y_{ij} - \bar{y}_i\right)^2 = 8.666667 + 8 + 12.66667 + 50.66667 = \underline{\underline{80}}$ **[1/2]**

Since **[1/2]** $SSE = SSPE + SSLF \Rightarrow SSLF = SSE - SSPE = 88.6 - 80 = \underline{\underline{8.6}}$ **[1/2]**

H_0 : *model is appropriate*
H_a : *model is not appropriate* **[1]** $\alpha = 0.05$

test-statistics: **[1/2]** $F_{LF} = \frac{MSLF}{MSPE} = \frac{SSLF \big/ \left[(n-2) - \sum_i (n_i - 1)\right]}{SSPE \big/ \sum_i (n_i - 1)} = \frac{8.6/(10-8)}{80/8} =$ [1/2] [1/2]

$$= \frac{4.3}{10} = \underline{\underline{0.43}}\ \text{[1/2]}$$

R.R: **we reject** H_0 **if** $F_{LF} > F_{\alpha(n-2-\sum_i(n_i-1),\sum_i(n_i-1))} = F_{0.05(2,8)} =$ **4.46** **[1]**

[1/2]

Since F_{LF} **= 0.43** $\ngtr$ **4.46, we do not reject** H_0 **[1/2] and conclude that at 5% level of significance there is not enough evidence to say that a linear model is not appropriate (i.e. model is appropriate). [1/2]**

[1] (b)

Lack of Fit Tests

Dependent Variable: cars

[1]

Source	Sum of Squares	df	Mean Square	F	Sig.
Lack of Fit	8.600	2	4.300	.430	.665
Pure Error	80.000	8	10.000		

2. [29 marks]

[3] (a) **Model:** **[1/2]** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, $n = 20$

Assumptions: (i) X_1, X_2 **are observed without error** **[1/2]**

(ii) y**'s (or** ε**'s) are independently distributed** **[1/2]** **with mean** $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ **[1/2]** **(or** $E(\varepsilon) = 0$**)**

(iii) variance of y**'s (or** ε**'s) is constant,** σ^2 **[1/2]** **for all** X_1, X_2

(iv) $y \sim N(E(y), \sigma^2)$ **[1/2]** **for any value of** X_1, X_2 **(or** $\varepsilon \sim N(0, \sigma^2)$ **for any value of** X_1, X_2**).**

NOTE: Assumptions (ii) – (iv) can be summarized also as $y \overset{i.i.d.}{\sim} N(E(y), \sigma^2)$ **(or** $\varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2)$**)**

[3] (b)

Coefficients[a] **[1]**

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	17.349	11.060		1.569	.135
	midterm_grade	.537	.115	.677	4.661	<.001
	hmwk_grade	.257	.120	.312	2.144	.047

a. Dependent Variable: final_grades

$\hat{\beta}_0 = 17.349$ **[1/2]** , $\hat{\beta}_1 = 0.537$ **[1/2]** , $\hat{\beta}_2 = 0.257$ **[1/2]**

Hence the least squares line is given by:

$\hat{y} = 17.349 + 0.537x_1 + 0.257x_2$ **[1/2]**

[4.5] (c)

ANOVA[a] **[1]**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2054.683	2	1027.342	16.853	<.001[b]
	Residual	1036.317	17	60.960		
	Total	3091.000	19			

a. Dependent Variable: final_grades
b. Predictors: (Constant), hmwk_grade, midterm_grade

$H_0 : \beta_1 = \beta_2 = 0$
H_a : *at least one of the* β's $\neq 0$ } **[1]** $\alpha = 0.01$

test-statistics: $F = \frac{MSR}{MSE} =$ **16.853**

R.R: **we reject** H_0 **if p-value <** α **(or if** $F > F_{\alpha(k,n-(k+1))} = F_{0.01(2,17)} =$ **6.11) [1]**

Since p-value < 0.001 < 0.01 (or *F* = 16.853 > 6.11) [1/2] , we reject H_0 **[1/2] and conclude that at 1% level of significance there is an evidence to say that a linear relationship between the final grade, midterm grade and/or homework grade exists. [1/2]**

[6] (d)

Coefficients[a] **[1]**

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	17.349	11.060		1.569	.135
	midterm_grade	.537	.115	.677	4.661	<.001
	hmwk_grade	.257	.120	.312	2.144	.047

a. Dependent Variable: final_grades

1. **midterm grade (x_1):**

$$H_0: \beta_1 = 0, \quad H_a: \beta_1 \neq 0 \quad \text{[1]} \quad \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$$

test-statistics: **t =** 4.661

Since *p*-value < 0.001 < 0.01 (or t = 4.661 > $t_{0.005,17} = 2.898$) **[1/2], at** α **= 0.01, we reject** H_0**, [1/2] and hence we conclude that 'the midterm grade (x_1)' is useful in predicting 'the final grade'. [1/2]**

2. **homework grade (x_2):**

$$H_0: \beta_2 = 0, \quad H_a: \beta_2 \neq 0 \quad \text{[1]} \quad \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$$

test-statistics: **t =** 2.144

Since *p*-value = 0.047 > 0.01 (or t = 2.144 $\not>$ $t_{0.005,17} = 2.898$) **[1/2], at** α **= 0.01, we do not reject** H_0**, [1/2] and hence we conclude that we do not have enough evidence to say that 'the homework grade (x_2)' is useful in predicting 'the final grades'. [1/2]**

[5] (e)

[1/2] $r^2 = \frac{SSR}{TSS} = \frac{2\ 054.683}{3\ 091} =$ **0.66473** $\cong$ **66.5%** [1/2]

i.e. approximately 66.5% of the total variation in the data is explained by the regr. line (and 33.5% is due to error). [1]

[1/2] $r^2_{adj} = 1 - \frac{SSE/n-(k+1)}{TSS/n-1} = 1 - \frac{MSE}{TSS/n-1} = 1 - \frac{60.960}{3091/19} = 1 - 0.374713684 =$ **0.62528** $\cong$

= 62.53% [1/2]

Since r^2_{adj} **is approx. 62.53% (not very high) we can conclude that the full model is not very good. Probably some higher order terms are needed (or perhaps because** X_2 **is not needed). [1]**

Model Summary [1]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.815[a]	.665	.625	7.808

a. Predictors: (Constant), hmwk_grade, midterm_grade

[7.5] (f)

Reduced Model: $y = \beta_0 + \beta_1 x_1 + \varepsilon$

ANOVA[a] **[1]**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1774.441	1	1774.441	24.260	<.001[b]
	Residual	1316.559	18	73.142		
	Total	3091.000	19			

a. Dependent Variable: final_grades
b. Predictors: (Constant), midterm_grade

Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2054.683	2	1027.342	16.853	<.001[b]
	Residual	1036.317	17	60.960		
	Total	3091.000	19			

a. Dependent Variable: final_grades
b. Predictors: (Constant), hmwk_grade, midterm_grade

$H_0 : \beta_2 = 0$ $\alpha = 0.01$
$H_a : \beta_2 \neq 0$ **[1]**

- **full model**: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
- **reduced model**: $y = \beta_0 + \beta_1 x_1 + \varepsilon$

***SSR$_f$* = 2054.683 (d.f. = 2) ,** ***SSR$_r$* = 1774.441 (d.f. = 1)**
***SSE$_f$* = 1036.317 (d.f. = 17) ,** ***SSE$_r$* = 1316.559 (d.f. = 18)** **[1/2]** if correct SPSS values were used in the calculation of test statistics

test-statistics :

[1/2] $F_{part} = \dfrac{[SSR_f - SSR_r]/[df_{SSR_f} - df_{SSR_r}]}{SSE_f / df_{SSE_f}} = \dfrac{(2054.683 - 1774.441)/(2-1)}{1036.317/17} =$ [1/2] [1/2] for d.f.

$= \dfrac{280.242/1}{60.960} = \underline{\underline{4.59714}}$ **[1/2]**

or equivalently,

$$F_{drop} = \frac{\left[SSE_r - SSE_f\right] / \left[df_{SSE_r} - df_{SSE_f}\right]}{SSE_f / df_{SSE_f}} = \frac{(1316.559 - 1036.317)/(18-17)}{1036.317/17} =$$

$$= \frac{280.242/1}{60.960} = \underline{\underline{\mathbf{4.59714}}}$$

<u>R.R:</u> we reject H_0 if $F_{part} > F_{\alpha(1,17)} = F_{0.01(1,17)} = 8.40$ [1]

Since F_{part} = 4.597 $\not>$ 8.40 [1/2], <u>we do not reject</u> H_0 [1/2] and conclude that at 1% level of significance there is not enough evidence to say that the X_2 term (i.e. the homework grade) contributes to the model. [1/2]

This confirms the results in part d) as the individual t-tests confirm that x_2 is not useful in predicting 'the final grade'. [1/2]