CROWDMARK

UNIVERSITY OF WATERLOO

Please print in pen:

Waterloo Student ID Number:

WatIAM/Quest Login Userid:

Examination
Final
Spring 2019
STAT 331

Times: Friday 2019-08-09 at 09:00 to 11:30

Duration: 2 hours 30 minutes (150 minutes)

Exam ID: 4097543

Sections: STAT 331 LEC 001

Instructors: Nathaniel Stevens

SOLUTIONS

Special Materials

Candidates may bring only the listed aids.

- Calculator - Pink Tie
- Study Notes - Double-Sided 8.5x11

Instructions:

- This test consists of 20 pages including this cover page.
- Page 17 contains information relevant to Question 3.
- Page 18 contains tables of quantiles from the $t_{(10)}$, $t_{(45)}$, $F_{(1,45)}$, $F_{(2,45)}$, $F_{(3,45)}$ and $F_{(4,45)}$ distributions.
- Pages 19 and 20 contain additional space for rough work. If you use these pages for work that you would like to have marked, you must clearly indicate this.
- For your convenience you may remove pages 17-20.
- All numeric answers should be rounded to four decimal places (unless the answer is exact to fewer than four decimal places).
- Incorrect answers may receive partial credit if your work is shown. An incorrect answer with no work shown will receive 0 points.

Question	Points
Q1	20 /20
Q2	25 /25
Q3	32 /32
Q4	5 /5
Q5	11 /11
Q6	7 /7
Total	100 /100

Signature:

- Please identify yourself by signing here: ____________

CROWDMARK

Question 1 [20 points]

(a) [1] Suppose that you wish to use forward selection to choose among $q = 7$ explanatory variables for inclusion in a model. In the *worst-case scenario* how many separate models would you have to fit when applying this algorithm?

i. 28
ii. 29
iii. 21
iv. 22

(b) [1] The regression model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$ is a *linear* regression.

i. True
ii. False

(c) [1] In the context of a simple linear regression relating y to x, the point $(\bar{x}, \bar{y})$ lies on the line-of-best fit.

i. True
ii. False

(d) [1] A 99% prediction interval for y_0 is narrower than the 99% confidence interval for μ_0.

i. True
ii. False

(e) [1] In the context of a linear regression model relating to y to x_1, x_2,…, x_p the null hypothesis corresponding to the test of *overall significance* in the linear regression is

$$H_0: \beta_0 = \beta_1 = \cdots = \beta_p = 0$$

i. True
ii. False

(f) [1] Suppose that the explanatory variables x_1, x_2 and x_3 are together in a model. The variance inflation factor for x_1 is $VIF_1 = 2$. What percentage of the variation of x_1 is explained by x_2 and x_3?

i. 10%
ii. 20%
iii. 50%
iv. 80%

(g) [1] Suppose that you wish to identify observations that have a large influence on your analysis. Which of the following metrics is best-suited for this purpose?

i. Cook's D-Statistics
ii. Leverages
iii. Studentized Residuals

(h) [1] Suppose that a model with $p = 77$ explanatory variables and $n = 100$ observations is fit and the resulting R^2 value is 0.9. What is R^2_{adj}? Write your answer in the space below.

0.55

(i) [1] The addition of an explanatory variable to a linear regression model always decreases the error sum of squares.

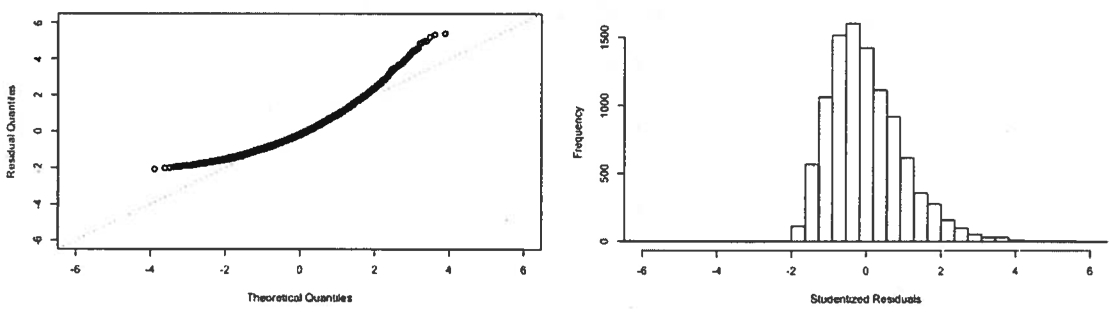
i. True
ii. False

(j) [1] *Leave-one-out* cross validation is equivalent to:

i. 1-fold cross validation
ii. n-fold cross validation
iii. $(n-1)$-fold cross validation

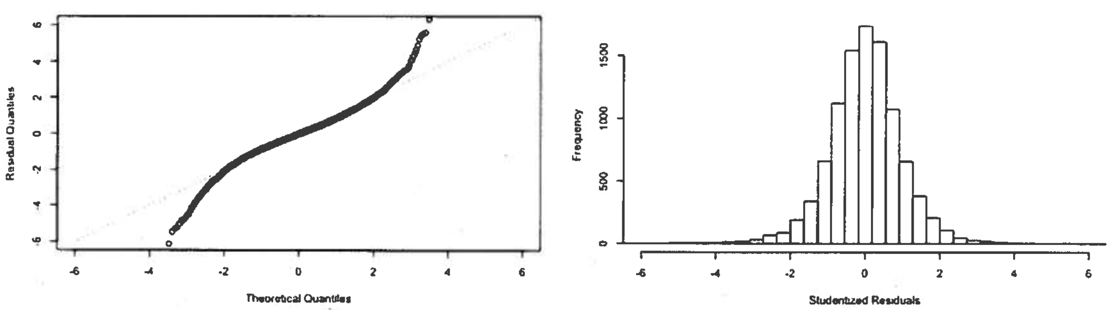
(k) [4] Four datasets gave rise to the following histograms and QQ-plots. Match each QQ-plot with the histogram that is most likely based on the same data. To identify your matching, beside each histogram indicate which QQ-plot – either i., ii., iii., or iv. – matches it.

i.



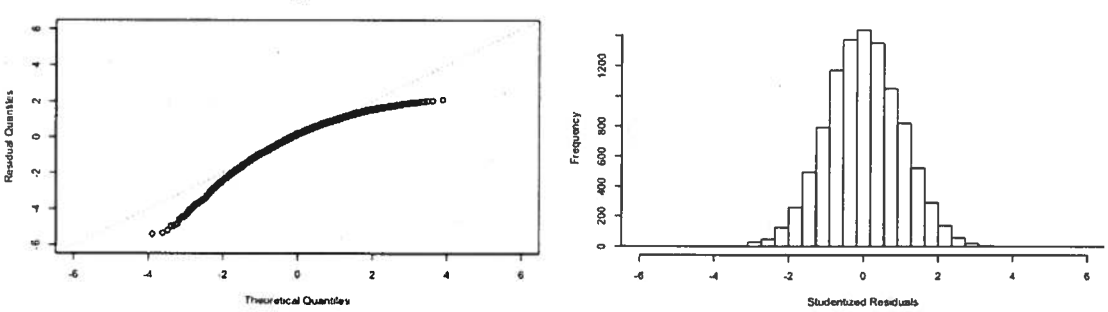
(i)

ii.



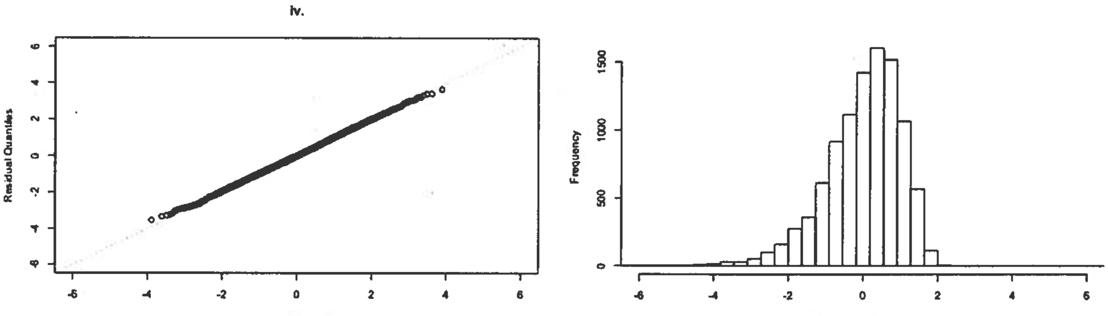
(ii)

iii.



(iv)

iv.



(iii)

(l) [1] Suppose that we fit the following model: $\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$. Suppose also that the estimate of β_1 is 0.4. By what factor do we expect y to increase if x_1 is increased by 3 units (and x_2 and x_3 are held fixed)? Write your answer in the space below.

$e^{3(0.4)} = 3.3201$

(m) [1] In *every* linear regression model, the average residual $\bar{e}$ is zero.

i. True
ii. False

Please initial:

The context of questions (n)-(q) is the following model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$$

for $i = 1,2,\dots,50$.

(n) [1] The null t-distribution associated with a test of $\mathrm{H_0}: \beta_4 = 0$ vs. $\mathrm{H_A}: \beta_4 \neq 0$ has how many degrees of freedom? Write your answer in the space below.

44

(o) [1] The rejection region associated with the test in (n) is given by:

$$\{t | t \geq 2.015 \text{ or } t \leq -2.015\}$$

If $\hat{\beta}_4 = 1.07$ and $\mathrm{SE}[\hat{\beta}_4] = 0.06$, is the hypothesis rejected? In the space below, state YES or NO.

YES

(p) [1] The null F-distribution associated with a test of $\mathrm{H_0}: \beta_3 = \beta_4 = \beta_5 = 0$ has how many degrees of freedom? Write your answer in the space below.

3, 44

(q) [1] The least squares estimate of σ^2 is $\hat{\sigma}^2 = 5$ and the sample variance of the response observations is 20. What is the value of R^2? Write your answer in the space below.

$R^2 = 1 - \frac{220}{980} = 0.7755$

Question 2 [25 points]

Consider the following regression equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

for $i = 1,2,\dots,13$.

(a) [3] This model may be equivalently written in matrix form as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

Define each of $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{\beta}$ in this case (i.e., when $n = 13$ and $p = 2$).

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{13} \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{13,1} & x_{13,2} \end{bmatrix} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

(b) [9] Suppose that the following summaries are available:

$$(X^TX)^{-1} = \begin{bmatrix} 2 & -5 & -3 \\ -5 & 1 & -1 \\ -3 & -1 & 5 \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 52 \\ 10 \\ 3 \end{bmatrix} \quad \text{and} \quad SSE = 30$$

i. [1] Write down the equation for the fitted model.

$$\hat{\mu} = 52 + 10x_1 + 3x_2$$

ii. [1] Compute $\hat{\sigma}^2$.

$$\hat{\sigma}^2 = \frac{SSE}{n-p-1} = \frac{30}{13-2-1} = \frac{30}{10} = 3$$

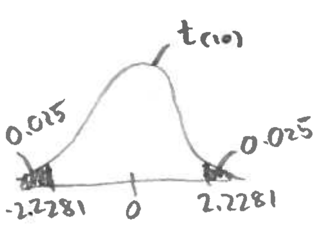
iii. [1] Compute the estimated variance of $\hat{\beta}_1$, $\mathrm{Var}[\hat{\beta}_1]$.

$$\mathrm{Var}[\hat{\beta}_1] = \hat{\sigma}^2 (X^TX)^{-1}_{22} = 3 \times 1 = 3$$

iv. [3] Test the following hypothesis at a 5% level of significance. [N.B. State the value of the test statistic and draw your conclusion by referring to the relevant quantiles found on page 18].

$$H_0: \beta_1 = 0 \text{ vs. } H_A: \beta_1 \neq 0$$

$$t = \frac{\hat{\beta}_1}{SE[\hat{\beta}_1]} = \frac{10}{\sqrt{3}} = 5.7735$$



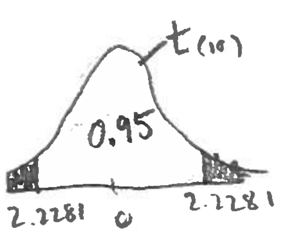
Since $P(t_{(10)} \geq 2.2281) = 0.025$, this implies that p-value $= 2P(t_{(10)} \geq 5.7735) < 0.05$ and so we reject H_0.

v. [3] Calculate a 95% confidence interval for β_2. [N.B. It will be useful for you to refer to the relevant quantiles found on page 18].

$$\hat{\beta}_2 \pm t_{(10,\, 0.975)} \times SE[\hat{\beta}_2]$$

$$= 3 \pm 2.2281 \times \sqrt{3 \times 5}$$

$$= (-5.8343,\ 11.8343)$$



(c) [11] Interest lies in doing inference for the response variable for specific values of x_1 and x_2.

i. [3] Consider the row vector $\boldsymbol{x_0} = [1 \quad x_{01} \quad x_{02}]$. Given $\widehat{\boldsymbol{\beta}} \sim \text{MVN}(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$, derive the distribution of $\hat{\mu}_0 = \boldsymbol{x_0}\widehat{\boldsymbol{\beta}}$. For full points you must name the distribution and derive general expressions for the expected value and variance of $\hat{\mu}_0 = \boldsymbol{x_0}\widehat{\boldsymbol{\beta}}$.

- Distribution name: Normal
- $E[\hat{\mu}_0] =$ $E[\vec{x}_0 \hat{\vec{\beta}}] = \vec{x}_0 E[\hat{\vec{\beta}}] = \vec{x}_0 \vec{\beta}$
- $\text{Var}[\hat{\mu}_0] =$ $\text{Var}[\vec{x}_0 \hat{\vec{\beta}}] = \vec{x}_0 \text{Var}[\hat{\vec{\beta}}] \vec{x}_0^{\,T}$

$$= \sigma^2 \vec{x}_0 (X^TX)^{-1} \vec{x}_0^{\,T}$$

ii. [2] By substituting the estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$ and $x_1 = 1$ and $x_2 = 3$ into the equations above, calculate estimates of the $E[\hat{\mu}_0]$ and $\text{Var}[\hat{\mu}_0]$ when $x_1 = 1$ and $x_2 = 3$ (i.e., when $\boldsymbol{x_0} = [1 \quad 1 \quad 3]$).

$$\hat{E}[\hat{\mu}_0] = \vec{x}_0 \hat{\vec{\beta}} = [1 \quad 1 \quad 3] \begin{bmatrix} 52 \\ 10 \\ 3 \end{bmatrix} = 71$$

$$\hat{\text{Var}}[\hat{\mu}_0] = \hat{\sigma}^2 [1 \quad 1 \quad 3] \begin{bmatrix} 2 & -5 & -3 \\ -5 & 1 & -1 \\ -3 & -1 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

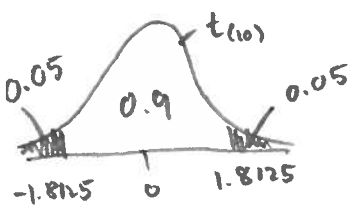
$$= 3 [1 \quad 1 \quad 3] \begin{bmatrix} -12 \\ -7 \\ 11 \end{bmatrix}$$

$$= 3(14)$$

$$= 42$$

iii. [3] Calculate a 90% *confidence* interval for the expected response observation when $x_1 = 1$ and $x_2 = 3$. [N.B. It will be useful for you to refer to the relevant quantiles found on page 18].

$$\hat{\mu}_0 \pm t_{(10,\,0.9)} \times SE[\hat{\mu}_0]$$



$$= 71 \pm 1.8125 \times \sqrt{42}$$

$$= (59.2537,\ 82.7463)$$

iv. [3] Calculate a 90% *prediction* interval for the predicted response observation y_0 when $x_1 = 1$ and $x_2 = 3$. [N.B. It will be useful for you to refer to the relevant quantiles found on page 18].

$$\hat{y}_0 \pm t_{(10,\,0.9)} \times SE[\hat{y}_0]$$

$$= 71 \pm 1.8125 \times \sqrt{\underset{\hat{\sigma}^2}{3} + 42}$$

$$= (58.8414,\ 83.1586)$$

v. [2] Briefly explain why the widths of these intervals differs.

The prediction interval is wider because it accounts for two types of uncertainty: estimation uncertainty and prediction uncertainty. Whereas the confidence interval only accounts for estimation uncertainty.

CROWDMARK

Question 3 [32 points]
There is a great deal of interest in determining the traits of a high school that predict the success of its students. In the US, student success beyond high school is often linked to high SAT (scholastic aptitude test) scores. A recent study aimed to identify attributes of a high school that may be used to predict the SAT scores of its students. In this study the following information was collected for each high school in a sample of $n = 50$ American high schools:

- `SAT`: average SAT (scholastic aptitude test) score (y)
- `expend`: average expenditures per student (in \$1000s) (x_1)
- `ratio`: average student-teacher ratio (x_2)
- `salary`: average annual salary of teachers (in \$1000s) (x_3)
- `takers`: percentage of eligible students that actually took the SAT test (x_4)

The linear regression model that relates `SAT` to the other variables may be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Partial R output for this model is shown on page 17. You may refer to this in the questions that follow.

(a) [2] It is usually believed that students thrive in smaller classes.

- [1] Interpret $\hat{\beta}_2$ in the context of the problem.

For a unit increase in the student-teacher ratio (and holding all else fixed) we expect the average SAT score at the school to decrease by 3.6242 points.

- [1] Does the value of $\hat{\beta}_2$ suggest that students perform better when class sizes are smaller? Circle one.

YES NO

(b) [6] It is also typically believed that students perform better in schools that pay their teachers higher salaries.

- [1] Interpret $\hat{\beta}_3$ in the context of the problem

For every additional \$1000 a teacher is paid, (all else equal) we expect the average SAT score at the school to increase by 1.6379 points.

- [1] Does the value of $\hat{\beta}_3$ suggest that students perform better when their teachers are paid more money? Circle one.

YES NO

- [4] Using an appropriate hypothesis test, at a 5% level of significance, test whether higher salaries are associated with a *significantly higher* expected SAT score. Draw your conclusions in the context of the problem.
[N.B. Clearly state the hypothesis you are testing; state the value of the test statistic; draw your conclusion by referring to the relevant quantiles found on page 18].

Alternatively, you could test $H_0: \beta_3 \geq 0$ vs. $H_A: \beta_3 < 0$. The only thing that changes is that the reference value is -1.6794 and you would not reject H_0 because $6.8618 > -1.6794$

$H_0: \beta_3 \leq 0$ vs. $H_A: \beta_3 > 0$

$$t = \frac{\hat{\beta}_3}{SE[\hat{\beta}_3]} = \frac{1.6379}{0.2387} = 6.8618$$

Since $P(t_{(45)} \geq 1.6794) = 0.05$ we know that p-value $= P(t_{(45)} \geq 6.8618) < 0.05$. Thus we reject H_0 and conclude that increase in expected SAT scores associated with higher salaries is statistically significant.

(c) [1] What percentage of the variability in `SAT` scores is *not* explained by `expend`, `ratio`, `salary` and `takers`?

$$1 - R^2 = 1 - 0.8246 = 0.1754 \quad \therefore 17.54\%$$

(d) [3] Complete the following ANOVA table:

Source	*df*	SS	MS	*F*
Regression	4	226215.19	56553.80	52.8891
Error	45	48118.05	1069.29	
Total	49	274333.24		

SPACE LEFT FOR ROUGH WORK:

$$\hat{\sigma}^2 = 32.7^2 = 1069.29 = \frac{SSE}{45} \Rightarrow SSE = 45 \times 1069.29 = 48118.05$$

$$R^2 = 0.8246 = 1 - \frac{48118.05}{SST} \Rightarrow SST = \frac{48118.05}{0.1754} = 274333.24$$

Alternatively: $SSE = 48123.90$ from Table 1 $\Rightarrow MSE = 1069.42$

And $SST = \frac{48123.90}{0.1754} = 274366.59 \Rightarrow SSR = 226242.69 \Rightarrow MSR = 56560.67$

$\Rightarrow F = 52.8891$

(e) [2] The F-ratio from the ANOVA table in part (d) tests the following hypothesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \text{ vs. } H_A: \beta_j \neq 0 \text{ for some } j = 1,2,3,4$$

At a 1% level of significance, do you reject the null hypothesis? State YES or NO and provide a brief justification.
[N.B. Draw your conclusion by referring to the relevant quantiles found on page 18]

YES

Since $P(F_{(4,45)} \geq 3.7674) = 0.01$ then

p-value $= P(F_{(4,45)} \geq 52.8891) < 0.01$ and

so p-value $< \alpha$ which means we reject H_0.

(f) [3] To evaluate the extent of multicollinearity, four separate regressions were fit between the explanatory variables. These models and their R^2 values are summarized in the table below. In the space provided, state the value of the variance inflation factor for each explanatory variable.

Response	Explanatory Variables	R^2	*VIF*
`expend`	`ratio, salary, takers`	0.894	9.4340
`ratio`	`expend, salary, takers`	0.589	2.4331
`salary`	`expend, ratio, takers`	0.892	9.2593
`takers`	`expend, ratio, salary`	0.430	1.7544

Does multicollinearity appear to be a problem for this data? Circle one.

YES NO

SPACE LEFT FOR ROUGH WORK:

$VIF_1 = \frac{1}{1-0.894}$

$VIF_2 = \frac{1}{1-0.589}$

$VIF_3 = \frac{1}{1-0.892}$

$VIF_4 = \frac{1}{1-0.430}$

(g) [5] Interest lies in finding a subset of the explanatory variables (and hence a reduced model) that adequately accounts for variation in SAT scores. The information provided in Table 1 on page 17 gives the model summary of *all possible regressions*, which in this case corresponds to $2^4 = 16$ different models. Note: all models contain an intercept. Using the information in this table, answer the following questions.

i. [2] Perform *backward elimination* using the AIC as a basis for eliminating variables from the model. Specifically, indicate the order in which variables exit the model and state the final model.

Step	Variable eliminated
1	x_1

STOP!

Final Model: $y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$

ii. [2] Perform *forward selection* using the AIC as a basis for adding variables into the model. Specifically, indicate the order in which variables enter the model and state the final model.

Step	Variable Added
1	x_4
2	x_1

STOP!

Final Model: $y = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \varepsilon$

iii. [1] The best overall model among *all possible regressions* is the one with the smallest AIC. Did either of these stepwise selection techniques choose the best overall model? Circle one.

YES NO

(h) [4] The optimal model according to *all possible regressions* and the AIC metric is the reduced model

$$y = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \varepsilon$$

This model arises if the following null hypothesis is true:

$$H_0: \beta_2 = \beta_3 = 0 \tag{1}$$

i. [1] This null hypothesis may be equivalently stated as

$$H_0: A\boldsymbol{\beta} = \mathbf{0} \tag{2}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$. Define the matrix A that makes hypotheses (1) and (2) equivalent.

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Alternatively, if $SSE_{Full} = 48118.05$ was used, then $t = 0.6556$

ii. [3] Using the additional sum of squares principle, formally test this hypothesis at a 5% level of significance.
[N.B. State the value of the test statistic and draw your conclusion by referring to the relevant quantiles found on page 18].

$$t = \frac{(SSE_{Red} - SSE_{Full})/\ell}{SSE_{Full}/(n-p-1)} = \frac{(49520.06 - 48123.90)/2}{48123.90/45}$$

$$= 0.6528$$

Since $P(F_{(2,45)} \geq 3.2043) = 0.05$ then

p-value $= P(F_{(2,45)} \geq 0.6528) > 0.05$ and so we do not reject H_0.

(i) [6] The six plots in Figure 1 (on page 17) were constructed using the reduced model from part (h). For each of the questions below answer YES or NO and provide a one-sentence justification. In each case, be sure to indicate which plot(s) you used to make your decision.

Accept "NO" if they're referred to plot [B] and say the residuals depend on the fitted values

i. Does the independence assumption appear to be met?

YES - the residuals vs index (in plot [A]) show a random scattering of points

ii. Does the constant variance assumption appear to be met?

NO - the residuals vs fitted values (in plot [B]) show a quadratic trend.

iii. Does the normality assumption appear to be met?

YES - aside from one outlier the points on the QQ-plot generally lie on the line of equality (plot [C])

iv. Do there appear to be any outliers in the y-dimension?

YES - Plots [A]-[D] indicate an observation with studentized residual < -3.

v. Do there appear to be any outliers in the x-dimension?

YES - the plot of the leverages ([F]) show 3 observations with leverages $> 2\bar{h}$.

vi. Do there appear to be any highly influential observations?

NO - the plot of Cook's D values ([E]) shows no observation with a substantially larger influence than the others.

CROWDMARK

Question 4 [5 points]
In class we showed that when SD[y] is proportional to μ (where $\mu = E[y]$) that the log-transformation is a variance stabilizing transformation. We also showed that when Var[y] is proportional to μ that the square-root transformation is variance stabilizing. By the same methods, find the transformation $g(\mu)$ that stabilizes the variance when SD[y] $\propto \mu^2$.

$$Var[y] = [h(\mu)]^2\sigma^2$$

where here $h(\mu) = \mu^2$. We want to choose $g(\cdot)$ such that

$$Var[g(y)] \simeq [g'(\mu)]^2[h(\mu)]^2\sigma^2$$

is constant.

Thus we want to choose $g(\cdot)$ such that

$$g'(\mu) = \frac{1}{h(\mu)} = \frac{1}{\mu^2}$$

The transformation $g(\cdot)$ that satisfies this is:

$$g(\mu) = -\frac{1}{\mu}$$

the negative reciprical transformation.

* note that the negative sign could be viewed to get absorbed into the proportionality constant so it's okay if they simply state the reciprical transformation.

CROWDMARK

Question 5 [11 points]
The *Gauss-Markov Theorem* states that in the linear regression model

$$y = X\beta + \varepsilon$$

the least squares estimator of β is the *best linear unbiased estimator* (BLUE) because among all possible linear and unbiased estimators of β, $\widehat{\beta} = (X^TX)^{-1}X^Ty$ has the smallest variance. In this question, you will prove this fact by considering another linear transformation $\widehat{\beta}^* = My$ where M is any $(p+1) \times n$ matrix of fixed numbers. The matrix $D = M - (X^TX)^{-1}X^T$ summarizes the difference between the two transformations.

(a) [5] Show that if $DX = O$ (where O is the $(p+1) \times (p+1)$ matrix of zeros), then
$$\mathrm{E}[\widehat{\beta}^*] = \beta$$

$$E[\hat{\vec{\beta}}^*] = E[M\vec{y}] = M E[\vec{y}]$$
$$= [D + (X^TX)^{-1}X^T] X\vec{\beta}$$
$$= DX\vec{\beta} + \vec{\beta}$$
$$= \vec{\beta} \quad \text{if } DX = 0$$

(b) [5] Show that if $\widehat{\beta}^*$ is unbiased, then
$$\mathrm{Var}[\widehat{\beta}^*] = \mathrm{Var}[\widehat{\beta}] + \sigma^2 DD^T$$

$$Var[\hat{\vec{\beta}}^*] = Var[M\vec{y}] = M\, Var[\vec{y}]\, M^T$$
$$= [D + (X^TX)^{-1}X^T]\, \sigma^2 I\, [D + (X^TX)^{-1}X^T]^T$$
$$= \sigma^2 [D + (X^TX)^{-1}X^T][D^T + X(X^TX)^{-1}]$$
$$= \sigma^2 [DD^T + \overset{0}{\cancel{DX}}(X^TX)^{-1} + (X^TX)^{-1}\overset{0}{\cancel{X^TD^T}}$$
$$+ (X^TX)^{-1}X^TX(X^TX)^{-1}]$$
$$= \sigma^2 [DD^T + (X^TX)^{-1}]$$
$$= \sigma^2 DD^T + Var[\hat{\beta}]$$

(c) [1] Why does the result in part (b) imply that $\mathrm{Var}[\widehat{\beta}] \le \mathrm{Var}[\widehat{\beta}^*]$?

Because DD^T is a positive semidefinite matrix.

Question 6 [7 points]

(a) [5] Suppose that we wish to model the relationship between a response variable y and p explanatory variables: $x_1, x_2, \dots, x_p$ and we observe the data in two batches: we observe n_1 observations initially, and then another n_2 observations at a later point in time.
For time point $k = 1,2$, let

- $\boldsymbol{y}_k$ be the $n_k \times 1$ response vector containing the response observations
- X_k be the $n_k \times (p+1)$ matrix containing the explanatory variable observations
- $\widehat{\boldsymbol{\beta}}_k = (X_k^T X_k)^{-1} X_k^T \boldsymbol{y}_k$ be the least squares estimate of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$

Show that the least squares estimate of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ based on all of the data can be written as

$$\widehat{\boldsymbol{\beta}} = (X_1^T X_1 + X_2^T X_2)^{-1} (X_1^T X_1 \widehat{\boldsymbol{\beta}}_1 + X_2^T X_2 \widehat{\boldsymbol{\beta}}_2)$$

Hint: the full X matrix and $\boldsymbol{y}$ vector (based on all $n = n_1 + n_2$ observations) can be obtained by stacking the respective time-specific matrices and vectors on top of each other.

$$X = \begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} \vec{y}_1 \\ \hline \vec{y}_2 \end{bmatrix}$$

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \vec{y}$$

$$= \left(\begin{bmatrix} X_1^T & \vdots & X_2^T \end{bmatrix} \begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1^T & \vdots & X_2^T \end{bmatrix} \begin{bmatrix} \vec{y}_1 \\ \hline \vec{y}_2 \end{bmatrix}$$

$$= (X_1^T X_1 + X_2^T X_2)^{-1} (X_1^T \vec{y}_1 + X_2^T \vec{y}_2)$$

$$= (X_1^T X_1 + X_2^T X_2)^{-1} (X_1^T X_1 \hat{\beta}_1 + X_2^T X_2 \hat{\beta}_2)$$

since $\hat{\vec{\beta}}_k = (X_k^T X_k)^{-1} X_k^T \vec{y}_k$ and so

$(X_k^T X_k) \hat{\vec{\beta}}_k = X_k^T \vec{y}_k$

(b) [2] Now suppose that you've collected data in $K > 2$ batches, and with each batch you estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ by $\widehat{\boldsymbol{\beta}}_k = (X_k^T X_k)^{-1} X_k^T \boldsymbol{y}_k$, $k = 1, 2, \dots, K$. Write an expression for the least squares estimate of $\boldsymbol{\beta}$ based on all $n = n_1 + n_2 + \dots + n_K$ observations and in terms of $\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \dots, \widehat{\boldsymbol{\beta}}_K$.

$$\hat{\vec{\beta}} = \left(\sum_{k=1}^{K} X_k^T X_k \right)^{-1} \left(\sum_{k=1}^{K} X_k^T X_k \hat{\vec{\beta}}_k \right)$$

This page is left blank.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1045.9715    52.8698   ?????    ?????
expend         4.4626    10.5465   ?????    ?????
ratio         -3.6242     3.2154   ?????    ?????
salary         1.6379     0.2387   ?????    ?????
takers        -2.9045     0.2313   ?????    ?????
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.7 on ?? degrees of freedom
Multiple R-squared:  0.8246,     Adjusted R-squared:  0.809
F-statistic: ????? on ?? and ?? DF,  p-value: ?????
```

Partial R Output for the full model in Question 3

Variables in Model	AIC	SSE
None (intercept only)	576.39	274307.68
x_1	570.57	234585.62
x_2	578.06	272496.65
x_3	567.64	221229.86
x_4	501.07	58433.15
x_1 x_2	572.33	233442.95
x_1 x_3	569.64	221225.01
x_1 x_4	494.80	49520.06
x_2 x_3	569.23	219441.23
x_2 x_4	500.14	55096.98
x_3 x_4	498.51	53338.38
x_1 x_2 x_3	570.63	216811.94
x_1 x_2 x_4	495.89	48627.32
x_1 x_3 x_4	496.76	49482.54
x_2 x_3 x_4	495.57	48315.37
x_1 x_2 x_3 x_4	497.37	48123.90

Table 1: AIC and SSE for various models in Question 3(g)

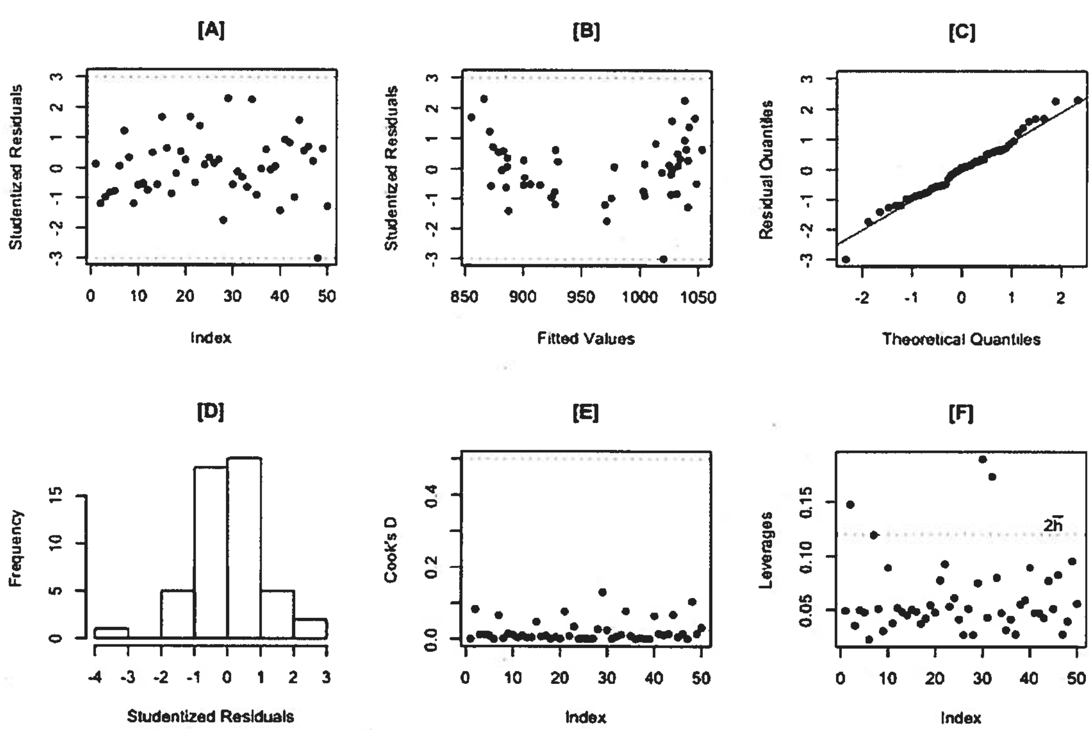


Figure 1: Diagnostic plots for the reduced model in Question 3(h)

For the indicated value of p, the following tables provide x^* where $P(X \geq x^*) = p$

$X \sim t_{(10)}$

p	x^*
0.005	3.1693
0.01	2.7638
0.025	2.2281
0.05	1.8125
0.1	1.3722

$X \sim t_{(45)}$

p	x^*
0.005	2.6896
0.01	2.4121
0.025	2.0141
0.05	1.6794
0.1	1.3006

$X \sim F_{(1,45)}$

p	x^*
0.005	8.7148
0.01	7.2339
0.025	5.3773
0.05	4.0566
0.1	2.8205

$X \sim F_{(2,45)}$

p	x^*
0.005	5.9741
0.01	5.1103
0.025	4.0085
0.05	3.2043
0.1	2.4245

$X \sim F_{(3,45)}$

p	x^*
0.005	4.8918
0.01	4.2492
0.025	3.4224
0.05	2.8115
0.1	2.2097

$X \sim F_{(4,45)}$

p	x^*
0.005	4.2941
0.01	3.7674
0.025	3.0860
0.05	2.5787
0.1	2.0742

LEFT BLANK FOR ROUGH WORK

CROWDMARK

LEFT BLANK FOR ROUGH WORK