BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

AMS 315, Data Analysis, Second Examination, November 14, 2019

Name: SOLUTION ID:

You may use an approved calculator in the examination provided that you permit the proctor to inspect them during the examination. You may also use one sheet of notes on paper that is the size of the paper on this examination (that is, standard paper size). The contents of this paper must be in your handwriting, and your name must be on the paper. The statistical tables from your text will be provided. You must turn in the tables and your notes with your examination. You may not give or accept any other assistance during the examination. The value of each question or part of question is stated with the question. The total value of the examination is 280 points, and the last problem is number 7. In the event of a fire alarm, gather your books and papers and leave the building. Find a quiet place to complete your examination. Please turn in the examination to my office (Mathematics Tower, Room 1-113) by 6:00 pm this evening. You are on your honor not to give or take assistance in this event.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

Common Information for Questions 1, 2, 3, and 4

A research team seeks to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y is a measure of the participant's anxiety at age 18 (where a larger number indicates greater anxiety). The variable x is a measure of the participant's body mass index (BMI) at age 15 (where a larger number indicates more BMI, a measure of relative weight). The variable w is a measure of the participant's anxiety at age 12 (where a larger number indicates greater anxiety). They observed values of y, x, and w for 617 participants. The mean and variance of Y (using $n-1$ as divisor) were 30.4 and 49.6 respectively. The mean and variance of x were 19.7 and 50.2 respectively. The mean and variance of w were 27.6 and 42.2 respectively. The correlation between Y and w was 0.55; the correlation between Y and x was 0.18; and the correlation between x and w was 0.31.

1. Compute the partial correlation coefficients $r_{Yw\cdot x}$ and $r_{Yx\cdot w}$. (20 points)

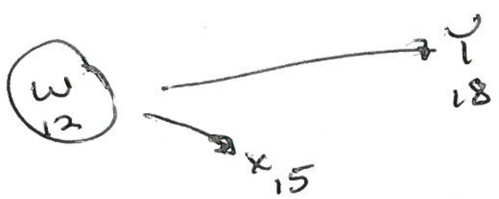
$$r_{Yw.x} = \frac{0.55 - (0.18)(0.31)}{\sqrt{(1-0.18^2)(1-0.31^2)}} = \frac{0.4942}{\sqrt{0.9676(0.9039)}}$$

$$= \frac{0.4942}{\sqrt{0.87461}} = 0.52844.$$

$$r_{Yx.w} = \frac{0.18 - (0.55)(0.31)}{\sqrt{(1-0.55^2)(1-0.31^2)}} = \frac{0.0095}{\sqrt{0.6975(0.9039)}}$$

$$= \frac{0.0095}{\sqrt{0.63047}} = 0.01196.$$

$$\sigma(r_{Yx.w}) \cong \frac{1}{\sqrt{614}} = 0.0404.$$



EXPLANATION:

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

2. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on w and the sum of squares due to the regression on x after including w. Test the null hypothesis that $\beta_1 = 0$ at the 0.10, 0.05, and 0.01 levels of significance. (50 points)

TOTAL SS = 616 (VAR(Y)) = 616 (49.6) = 30,553.6.

REG SS(W) = $(r_{Yw})^2$ TOTAL SS = $(0.55)^2$ SSTOTAL = 9242.46

TOTAL SS − REG SS(W) = 21,311.136.

REGSS(X|W) = $(r_{Yx \cdot w})^2$ (TOTAL SS − REG SS(W))

= $(0.01196)^2$ (21,311.136) = 3.048.

SSE = 30,553.6 − 9242.46 − 3.048 = 21,308.09.

ANOVA TABLE.

SOURCE	DF	SS	MS	F
REG (W)	1	9242.46		
REG (X|W)	1	3.05	3.05	0.088.
ERROR	614	21,308.09	34.70	
TOTAL	616			

α	F(1, 614)	F(1, ∞).	
.10	2.7137	2.71	A
.05	3.8566	3.84	A
.01	6.6763	6.63	A.

ACCEPT $H_0: \beta_1 = 0$ VS $H_1: \beta_1 \neq 0$ AT $\alpha = .10$ (+ .05 + .01).

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

3. Is a mediation model or an explanation model a better explanation of the observed results? You must support your choice with results from your analyses to receive credit for this question. (20 points).

EXPLANATION -

4. What changes in w and Y, if any, will be associated with a change in x given the information in the common section? Circle the letter of your answer. (20 points).
   a. No change in w and no change in Y.
   b. Non-zero change in w and no change in Y.
   c. Non-zero change in Y but no change in w.
   d. Non-zero change in both w and Y.

End of Application of Common Information

A.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Common Information for Questions 1, 2, 3, and 4

A research team seeks to estimate the model $(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y is a measure of the participant's academic performance at age 12 (where a larger number indicates better academic performance). The variable x is a measure of the participant's academic performance at age 8 (where a larger number indicates better academic performance). The variable w is the value of supplemental educational resources invested in the participant at age 6 (a larger number indicates more resources invested). They observed values of y, x, and w for 943 participants. The mean and variance of Y (using $n - 1$ as divisor) were 125.4 and 249.6 respectively. The mean and variance of x were 109.6 and 205.2 respectively. The mean and variance of w were 3.6 and 4.2 respectively. The correlation between Y and w was 0.31; the correlation between Y and x was 0.58; and the correlation between x and w was 0.52.

1. Compute the partial correlation coefficients $r_{Yw \cdot x}$ and $r_{Yx \cdot w}$. (20 points)

$$r_{Yw \cdot x} = \frac{0.31 - (0.58)(0.52)}{\sqrt{(1 - .58^2)(1 - .52^2)}} = \frac{.0084}{\sqrt{(.6636)(.7296)}}$$

$$= \frac{0.0084}{\sqrt{0.48416}} = 0.01207$$

NOTE $\sigma(r_{Yw \cdot x}) \cong \frac{1}{\sqrt{940}} = 0.0326.$

$$r_{Yx \cdot w} = \frac{0.58 - (0.31)(0.52)}{\sqrt{(1 - 0.31^2)(1 - 0.52^2)}} = \frac{0.4188}{\sqrt{(0.9039)(0.7296)}}$$

$$= \frac{0.4188}{\sqrt{0.65949}} = 0.51571$$

$$\underset{6}{w} \longrightarrow \underset{8}{x} \longrightarrow \underset{12}{Y} \qquad \text{MEDIATION}$$

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

2. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on w and the sum of squares due to the regression on x after including w. Test the null hypothesis that $\beta_1 = 0$ at the 0.10, 0.05, and 0.01 levels of significance. (50 points)

$$SS\,TOTAL = (n-1)\,VAR(Y) = 942(249.6)$$
$$= 235,123.2$$

$$SSREG(w) = (r_{yw})^2\, SS\,TOTAL = (0.31)^2\, SS\,TOTAL$$
$$= 22595.34$$

$$SS\,TOTAL - SS\,REG(w) = 212,527.86$$

$$SSREG(x|w) = (r_{yx\cdot w})^2\,(SS\,TOTAL - SSREG(w))$$
$$= (0.51571)^2\; 212,527.86 = 56,523.23.$$

$$SSERR = SS\,TOTAL - SSREG(w) - SSREG(x|w)$$
$$= 156,004.63$$

ANOVA TABLE

SOURCE	DF	SS	MS
REG(w)	1	22,595.34	22,595.34
REG(x|w)	1	56,523.23	56,523.23
ERROR	940	156,004.63	165.96

$$F_{x|w} = \frac{56,523.23/1}{156,004.63/940} = 340.58$$

α	F (1,940)	F(1,∞)	
.10	2.7109	2.71	R
.05	3.8514	3.84	R
.01	6.6619	6.63	R.

REJECT H_0 $\beta_1=0$ vs H_1 $\beta_1 \neq 0$ AT $\alpha = .01$ (+ .05 + .10).

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. Is a mediation model or an explanation model a better explanation of the observed results? You must support your choice with results from your analyses to receive credit for this question. (20 points).

MEDIATION MODEL.

4. What changes in x and Y, if any, will be associated with a change in w given the information in the common section? Circle the letter of your answer. (20 points)
   a. No change in x and no change in Y.
   b. Non-zero change in x and no change in Y.
   c. Non-zero change in Y and no change in x.
   d. Non-zero change in both x and Y.

End of Application of Common Information

D.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5. A research team exposed a total of 32 animals randomly assigned to four settings of dosages of a supplemental diet and observed the increase in weight response Y. The research team sought to find the dosage that maximized the response variable. Eight animals were given one unit of dosage with observed average and sample variance (unbiased estimate) with $y_{1.} = 47.0$ and $s_1^2 = 42.2$; eight were given two units of dosage with $y_{2.} = 52.5$ and $s_2^2 = 36.2$; eight were given three units of dosage with $y_{3.} = 49.6$ and $s_3^2 = 49.2$; and eight were given four units of dosage with $y_{4.} = 54.9$ and $s_4^2 = 43.2$ Complete the analysis of variance table including the degrees of freedom, sums of squares, mean squares, and test statistic. Test the null hypothesis that the mean increases in weight are equal for the four dosages at the 0.10, 0.05, and 0.01 levels of significance. What is the optimal setting of dosage? (60 points)

DOSE	J_i	$y_{i.}$	$y_{i.} - y_{..}$	s_i^2
1	8	47.0	-4.0	42.2
2	8	52.5	1.5	36.2
3	8	49.6	-1.4	49.2
4	8	54.9	3.9	43.2
		$y_{..} = 51$	0	42.7

$SS_{TREAT} = 8\,[-4.0^2 + 1.5^2$
$(-1.4)^2 + (3.9)^2]$
$= 8\,(35.42)$
$= 283.36$

ANOVA TABLE

SOURCE	DF	SS	MS	F
DOSE	3	283.36	94.453	2.212
ERROR	28	1195.60	42.7	
TOTAL	31	1478.96		

α	F(3,28)	
.10	2.291	A
.05	2.947	A
.01	4.568	A

ACCEPT H_0 ALL TREATMENT MEANS EQUAL AT $\alpha = .10$ (+ .05 + .01).

THERE IS NO OPTIMAL DOSE

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

6. The random variable $Y>0$ is a member of a class of random variables such that $E(Y) = \theta, \theta > 0$, and $\mathrm{var}(Y) = \theta^2$. Let $W = Y^p, p \neq 0$.
   a. What is the approximate value of $E(W)$? (10 points)
   b. What is the approximate value of var (W)? (40 points)

A. $E(W) \cong f(E(Y)) = \theta^p$

$E(W) \cong \theta^p$.

B. $VAR(W) \cong (f'(EY))^2 \; VAR(Y)$.

HERE $f(\theta) = \theta^p$. ; $f'(\theta) = p\theta^{p-1}$.

$VAR(W) \cong (p\theta^{p-1})^2 \cdot \theta^2$.

$\cong p^2 \theta^{2p-2+2}$

$\cong p^2 \theta^{2p}$

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

7. The random vector Y is $n \times 1$ with expected value $E(Y) = X_1\beta_1 + X_2\beta_2$. The matrix X_1 is an $n \times p_1$ matrix of constants with rank $p_1 \geq 2$; that is, $(X_1^T X_1)^{-1}$ exists. The matrix X_2 is an $n \times p_2$ matrix of constants with rank $p_2 \geq 1$. The vector β_1 is a $p_1 \times 1$ vector of unknown constant parameters. The vector β_2 is a $p_2 \times 1$ vector of unknown constant parameters. The variance-covariance matrix of Y is the symmetric positive definite $n \times n$ matrix $V \neq \sigma^2 I_{n\times n}$, where $I_{n\times n}$ is the $n \times n$ identity matrix.
   a. What is $E[(X_1^T X_1)^{-1} X_1^T Y]$? (30 points)
   b. What is the variance-covariance matrix of $(X_1^T X_1)^{-1} X_1^T Y$? (30 points)

End of Examination

A. $E((X_1^T X_1)^{-1} X_1^T Y) = (X_1^T X_1)^{-1} X_1^T E(Y)$

$= (X_1^T X_1)^{-1} X_1^T (X_1\beta_1 + X_2\beta_2)$

$= (X_1^T X_1)^{-1} X_1^T X_1 \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2$

$= \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2$.

B. $vcv((X_1^T X_1)^{-1} X_1^T Y) = vcv(MY)$ WHERE $M = (X_1^T X_1)^{-1} X_1^T$

$M^T = X_1 (X_1^T X_1)^{-1}$

$vcv(MY) = M \, vcv(Y) M^T = [(X_1^T X_1)^{-1} X_1^T] V X_1 (X_1^T X_1)^{-1}$.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBR

5. A research team exposed a total of 120 animals randomly assigned to four settings of dosages of a supplemental diet and observed the cholesterol level response Y in each animal. The research team sought to find the dosage that minimized the response variable. Thirty animals were given one unit of dosage with observed average and sample variance (unbiased estimate) with $y_{1\cdot} = 30.0$ and $s_1^2 = 315.4$; thirty were given two units of dosage with $y_{2\cdot} = 24.2$ and $s_2^2 = 286.2$; thirty were given three units of dosage with $y_{3\cdot} = 19.0$ and $s_3^2 = 269.2$; and thirty were given four units of dosage with $y_{4\cdot} = 13.6$ and $s_4^2 = 343.2$ Complete the analysis of variance table including the degrees of freedom, sums of squares, mean squares, and test statistic. Test the null hypothesis that the mean increases in weight are equal for the four dosages at the 0.10, 0.05, and 0.01 levels of significance. What is the optimal setting of dosage? (60 points)

DOSE	J_i	$y_{i\cdot}$	$y_{i\cdot} - y_{\cdot\cdot}$	s_i^2
1	30	30.0	8.3	315.4
2	30	24.2	2.5	286.2
3	30	19.0	-2.7	269.2
4	30	13.6	-8.1	343.2
		$y_{\cdot\cdot} = 21.7$	0	MSE = 3035

$SS_{TREAT} = 30(8.3^2 + 2.5^2 + (-2.7)^2 + (-8.1)^2 = 30(148.04)$

$= 4441.2$ ON 3 DF.

ANOVA TABLE.

SOURCE	DF	SS	MS	F
DOSE	3	4441.2	1480.4	4.878
ERROR	116	35,206.0	3035	
TOTAL	119	39,647.2		

α	$F(3,116)$
.10	2.132 R
.05	2.683 R
.01	3.955 R.

$F(3,120) = 2.13, (\alpha=.10)$
$2.68 (\alpha=.05)$
$3.95 (\alpha=.01)$

REJECT H_0 ALL MEANS EQUAL AT $\alpha = .01$ (AND .05 + .10)

LSD 99% $= t_{.995,116} \sqrt{303.5(\frac{1}{30} + \frac{1}{30})} = 2.619\sqrt{20.233} = 11.78.$

DOSAGE 4 APPEARS OPTIMAL AS IT WAS SMALLEST ON AVERAGE, AND THERE WAS A MONOTONIC DECREASE IN AVERAGES. PROTECTED T CI (99%) SHOWS ONLY DOSAGE 4 + 1 DIFFER

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

6. The random variable $Y>0$ is a member of a class of random variables such that $E(Y)=\theta, \theta>0$, and $\operatorname{var}(Y)=\theta^3$. Let $W=Y^p, p \neq 0$.
   a. What is the approximate value of $E(W)$? (10 points)
   b. What is the approximate value of var (W)? (40 points)

$W = Y^P$, $W = f(Y)$, WHERE $f(Y) = Y^P$.

$f(\theta) = \theta^P$.

$f'(\theta) = p\theta^{P-1}$.

A. $E(W) \cong f(\theta)$

$E(W) \cong \theta^P$.

B. $VAR(W) \cong (f'(\theta))^2 \; VAR(Y)$

$VAR(W) \cong (p\theta^{p-1})^2 \theta^3 = p^2 \theta^{2p-2+3}$

$VAR(W) \cong p^2 \theta^{2p+1}$.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

SAME AS A FORM.

7. The random vector $\boldsymbol{Y}$ is $n \times 1$ with expected value $E(Y) = X_1\beta_1 + X_2\beta_2$. The matrix X_1 is an $n \times p_1$matrix of constants with rank $p_1 \geq 2$; that is, $(X_1^T X_1)^{-1}$ exists. The matrix X_2 is an $n \times p_2$ matrix of constants with rank $p_2 \geq 1$. The vector β_1 is a $p_1 \times 1$ vector of unknown constant parameters. The vector β_2 is a $p_2 \times 1$ vector of unknown constant parameters. The variance-covariance matrix of $\boldsymbol{Y}$ is the symmetric positive definite $n \times n$ matrix $V \neq \sigma^2 I_{n\times n}$, where $I_{n\times n}$ is the $n \times n$ identity matrix.
   a. What is $E[(X_1^T X_1)^{-1} X_1^T Y]$? (30 points)
   b. What is the variance-covariance matrix of $(X_1^T X_1)^{-1} X_1^T Y$? (30 points)

End of Examination

A. $E[(X_1^T X_1)^{-1} X_1^T Y] = (X_1^T X_1)^{-1} X_1^T (EY)$

$= (X_1^T X_1)^{-1} X_1^T (X_1\beta_1 + X_2\beta_2)$

$= (X_1^T X_1)^{-1} X_1^T X_1 \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2$

$= \beta_1 + (X_1^T X_1)^{-1} (X_1^T X_2)\beta_2$

B. $VCV((X_1^T X_1)^{-1} X_1^T Y) = VCV(MY)$ WHERE $M = (X_1^T X_1)^{-1} X_1^T$

$VCV(MY) = M\, VCV(Y)\, M^T = M V M^T$

$= (X_1^T X_1)^{-1} X_1^T V [(X_1^T X_1)^{-1} X_1^T]^T$

$= (X_1^T X_1)^{-1} (X_1^T V X_1)(X_1^T X_1)^{-1}$.