AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AMS 315, Examination 2

April 19, 2018

Name: SOLUTION **ID:**

Directions: Write your name in the space provided. Work each problem in the space underneath the problem and on the back side of the page. You may use a calculator but not a computer or cell-phone. You may also use a single sheet of notes in your handwriting that is the size of the paper in this examination. You may use only the paper in this form. You are on your honor not to use any other assistance during this examination. Do not make marks on the tables given to you to work this examination. Turn in your paper, your notes, and your tables at the end of the examination. There will be no partial credit given for a problem unless you show your work. In the event of a fire alarm, please take your papers, exit the room, find a private place to work, and turn in your examination to me in my office (Math Tower 1-113) by 9:00 pm today. In this event, you are still on your honor not to give or receive assistance.

The last problem is number 13. The total value of the examination is 270 points. Your graded examination will be returned on Tuesday.

Since the course satisfies requirements for actuarial credentials, academic integrity standards will be enforced strictly.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

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

A research team sought to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y was a scale measuring anti-social behaviors at age 22 (with a higher number indicating greater anti-social behavior). The variable x was a measure of the participant's extent of substance use at age 17 (higher values meant greater substance use); and the variable w was a measure of the participant's rebelliousness at age 12 (higher values meant greater rebelliousness). They observed values of y, x, and w on $n = 621$ subjects. The mean and variance of Y (using $n-1$ as divisor) were 125.4 and 745.8 respectively. The mean and variance of w were 33.8 and 44.9 respectively. The mean and variance of x were 11.6 and 26.7 respectively. The correlation between Y and w was 0.37, the correlation between Y and x was 0.61; and the correlation between x and w was 0.55.

1. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$. This question is worth 20 points.

$$r_{Yx\bullet w} = \frac{0.61 - (0.37)(0.55)}{\sqrt{(1-.37^2)(1-.55^2)}} = \frac{.4065}{\sqrt{(.8631)(.6975)}}$$

$$= \frac{.4065}{\sqrt{.60201}} = \frac{.4065}{.77589} = .5239.$$

$$r_{Yw\bullet x} = \frac{0.37 - (.61)(.55)}{\sqrt{(1-.61^2)(1-.55^2)}} = \frac{.0345}{\sqrt{(.6279)(.6975)}}$$

$$= \frac{.0345}{\sqrt{.43796}} = \frac{.0345}{.66179} = 0.0521.$$

+10 EACH CORRECT ANSWER; −5 CORRECT FORMULA WITH COMP ERROR.

$$\underset{12}{w} \to \underset{17}{x} \to \underset{22}{Y}$$

MEDIATION.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

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 the sum of squares due to the regression on x after including w, the error sum of squares, and the total sum of squares. This question is worth 20 points.

$SS\ TOT = 620(745.8) = 462396.0$

$SS(w) = (r_{yw})^2\ SSTOT = (.37)^2\ 462396.0 = 63302.01$

$SS\ TOT - SS(w) = 399093.99.$

$SS(x|w) = (r_{yx \cdot w})^2\ (SSTOT - SS(w))$

$= 109539.81.$

ANOVA TABLE.

SOURCE	DF	SS	MS
w	1	63,302.01	63302.01
x|w	1	109,539.81	109,539.81
ERROR	618	289554.18	468.53
TOTAL	620	462396.0	

$F_{x|w} = \frac{109539.81}{468.53} = 233.79.$

−15 EACH INCORRECT ENTRY

−5 CORRECT FORMULA WITH COMP ERROR.

α	F(1, 618)	
.10	2.714	R.
.05	3.857	R.
.01	6.676	R.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. Test the null hypothesis that $\beta_1 = 0$ against the alternative $\beta_1 \neq 0$. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

$F_{X|W} = 233.79$ ON (1, 620).

   a. Accept the null hypothesis at the 0.10 level of significance.
   b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
   c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
   d. Reject the null hypothesis at the 0.01 level.

NO TEST OR INCORRECT TEST -20.

COMP ERROR ON CORRECT TEST -5 OR -10

ANOVA TABLE INCORRECT SEQUENCE

SOURCE	DF	SS	MS	F
x	1	172057.55	172057.55	367.23
W|X	1	784.27	784.27	
ERROR	618	289554.18	468.53	
TOTAL	620	462396.0		

-10 FOR INCORRECT SEQUENCE ON 2 AND +20 ON 3 FOR USING 367.23; -20 FOR $\frac{784.27}{468.53}$ (WRONG TEST).

4. Is a mediation model or an explanation model a better explanation of the observed results? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

a. An explanation model is better than a mediation model.
b. A mediation model is better than an explanation model.
c. Neither model is a good explanation of the observed results.

End of application of common information

+20 FOR CORRECT ANSWER

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

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

A research team seeks to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$ using data from a longitudinal sample of 915 participants. The variable Y is a measure of the participant's utilization of health services at age 30 (where a larger number indicates greater utilization); the variable x is the measure of the participant's substance usage at age 22 (where a greater number indicates greater usage of alcohol and illegal substances); and the variable w is a measure of the participant's level of depression at age 26 (a larger number indicates more extensive depression). They observed values of y, x, and w for the participants. The mean and variance of Y (using $n-1$ as divisor) were 85.4 and 528.5 respectively. The mean and variance of x were 11.6 and 56.7 respectively. The mean and variance of w were 33.8 and 74.9 respectively. The correlation between Y and w was 0.10; the correlation between Y and x was 0.24; and the correlation between x and w was 0.36.

1. Compute the partial correlation coefficients $r_{Yx \bullet w}$ and $r_{Yw \bullet x}$. This question is worth 20 points.

$$r_{Yx\cdot w} = \frac{r_{Yx} - r_{Yw} r_{xw}}{\sqrt{(1-r_{Yw}^2)(1-r_{xw}^2)}} = \frac{.24-(.10)(.36)}{\sqrt{(1-.10^2)(1-.36^2)}}$$

$$= \frac{.204}{\sqrt{(.99)(.8704)}} = \frac{.204}{\sqrt{.86170}} = \frac{.204}{.92828} = 0.2198.$$

$$r_{Yw\cdot x} = \frac{.10-(.24)(.36)}{\sqrt{(1-.24^2)(1-.36^2)}} = \frac{.0136}{\sqrt{(.9424)(.8704)}}$$

$$= \frac{.0136}{\sqrt{.82026}} = \frac{.0136}{.90568} = 0.0150$$

GRADING AS IN A FORM.

$$\begin{tikzcd} x \arrow[rr] \arrow[r] & w & Y \\ 22 & 26 & 30 \end{tikzcd}$$

EXPLANATION.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

2. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on x, the sum of squares due to the regression on w after including x, the error sum of squares, and the total sum of squares. This question is worth 20 points.

$$SSTOT = 914(528.5) = 483,049.$$

$$SS(x) = (r_{yx})^2 \, SSTOTAL = (.24)^2 \, 483.049$$

$$= 27823.6224.$$

$$SSTOT - SS(x) = 455225.3776.$$

$$SS(w|x) = (0.0150)^2 (SSTOT - SS(x)) = 102.43$$

ANOVA TABLE

SOURCE	DF	SS	MS
x	1	27823.62	
$w \mid x$	1	102.43	102.43
ERROR	912	455122.95	499.038
TOTAL	914	483049.0	

$$F_{w|x} = \frac{102.43}{499.038} = 0.205.$$

α	$F(1,914)$.	
.10	2.711	A
.05	3.852	A
.01	6.663	A.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

3. Test the null hypothesis that $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points. $F_{W|X} = 0.205$ ON (1, 912).

   a. Accept the null hypothesis at the 0.10 level of significance.
   b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
   c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
   d. Reject the null hypothesis at the 0.01 level.

SAME GRADING CRITERIA AS A FORM.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

4. Is a mediation model or an explanation model a better explanation of the observed results? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

   a. An explanation model is better than a mediation model.
   b. A mediation model is better than an explanation model.
   c. Neither model is a good explanation of the observed results.

End of application of common information

CORRECT OR NOT.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

Common Information for Questions 5 to 12

A research team studied Y, the protein production of a laboratory animal, and how Y was affected by the dose of medicine. The research team sought to minimize $E(Y)$. They used four doses of medicine: 0, 1, 2, and 3 units respectively. They randomly assigned 20 animals to dosage 0, 20 to dosage 1, 20 to dosage 2, and 20 to dosage 3. They observed that the average values of Y at each dosage were $y_{0\bullet} = 681.6, y_{1\bullet} = 412.4, y_{2\bullet} = 288.7,$ and $y_{3\bullet} = 393.3$ where $y_{i.}$ was the average of the observations taken with dosage $i = 0,\ldots,3$, respectively. They also observed that $s_0^2 = 36250.9, s_1^2 = 32878.3, s_2^2 = 38630.4,$ and $s_3^2 = 37164.5$ where s_i^2 was the unbiased estimate of the variance for the observations taken with dosage $i = 0,\ldots,3$, respectively.

DOSAGE	J_i	$y_{i\cdot}$	$y_{i\cdot} - y_{\cdot\cdot}$	$(y_{i\cdot}-y_{\cdot\cdot})^2$	s_i^2
0	20	681.6	237.6	56453.76	36250.9
1	20	412.4	−31.6	998.56	32878.3
2	20	288.7	−155.3	24118.09	38630.4
3	20	393.3	−50.7	2570.49	37164.5
SUM	80	1776.0	0	84140.9	144924.1
AVE		444.0			36231.025

SSTREAT = 20 (84 140.9) = 1,682,818

MS TREAT = SS TREAT / 3 = 560939.33 ON 3 DF

MSPE = 36 231.025 ON 76 DF

SSPE = 2,753,557.9

ANOVA TABLE

SOURCE	DF	SS	MS	F
DOSE	3	1,682,818	560,939.33	15.482
PURE ERROR	76	2,753,557.9	36,231.025	
TOTAL	79	4 436 375.9		

α	F(3,76)
.10	2.157 R
.05	2.725 R
.01	4.050 R

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

5. Complete the one way analysis of variance table for these results; that is, be sure to specify the degrees of freedom and sum of squares for the treatment source, the error source, and the total. Also include appropriate mean squares and test statistic. This question is worth 20 points.

−15 EACH INCORRECT ENTRY.

−5 CORRECT FORMULA WITH COMP ERROR

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

6. Test the null hypothesis that the mean production is the same for each dosage against the alternative that the mean production is different for at least one dosage. What is the correct test. Circle the answer below corresponding to your conclusion. This question is worth 20 points.

$F_{OVERALL} = 15.482$ ON $(3, 76)$.

   a. Accept the null hypothesis at the 0.10 level of significance.
   b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
   c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
   d. Reject the null hypothesis at the 0.01 level.

−20 NO TEST OR INCORRECT TEST

90% LSD = 100.22

95% LSD = 119.90

7. What is the optimal dosage? This question is worth 10 points.

99% LSD $= 2.642\sqrt{MSE\left(\frac{1}{20}+\frac{1}{20}\right)} = 159.03$

OPTIMAL DOSAGE IS 2.

99% CI FOR $E(\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot}) = 288.7 - 393.3 \pm 2.642\sqrt{36{,}231\left(\frac{1}{20}+\frac{1}{20}\right)}$

$= -104.6 \pm 159.03$

90% CI IS -104.6 ± 100.22

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

8. Find the estimate of the linear contrast and its sum of squares. The coefficients of the linear contrast are $-3,-1,1,3$. This question is worth 10 points.

$$\hat{\lambda}_{LIN} = -3(681.6) - 412.4 + 288.7 + 3(393.3) = -988.6$$

$$SS_{LIN} = \frac{(\hat{\lambda}_{LIN})^2}{20/20} = 977329.96$$

+5 FOR $\hat{\lambda}_{LIN}$.

+5 FOR SS_{LIN}.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

9. Report the analysis of variance table for the linear regression of *Y* on dosage. Use the sum of squares due to the linear contrast as the sum of squares for the regression of *Y* on dosage. Include the sum of squares due to error and the total sum of squares. This question is worth 20 points.

ANOVA TABLE

SOURCE	DF	TOTAL SS	MS	F
LINEAR	1	977,329.6	977, 329.6	22.038
ERROR	78	3,459,045.94	44,346.74	
TOTAL	79	4,436,375.9		

α		
.10	2.771	R.
.05	3.963	R
.01.	6.971	R

TSS IN #5.

−15 TOTAL SS NOT EQUAL TO TSS IN #5.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

10. Test the null hypothesis that there is no linear association between Y and dosage against the alternative that there is a linear association. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

$F_{LIN} = 22.038$ ON (1,78).

a. Accept the null hypothesis at the 0.10 level of significance.
b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
d. Reject the null hypothesis at the 0.01 level.

+20 IF CHOICE CONSISTENT WITH YOUR CALCULATION.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

11. Report the analysis of variance table including the sum of squares due to linear regression, the sum of squares due to lack of fit, the sum of squares due to pure error, and the total sum of squares. Include the degrees of freedom for each sum of squares. This question is worth 20 points.

LACK OF FIT ANOVA TABLE

SOURCE	DF	SS	MS
REG	1	977,329.6	977,329.6
LACK OF FIT	2	705488.04	352,744.02
PURE ERROR	76	2,753,557.9	36,231.025
TOTAL	79	4,436,375.9	

$$F_{LOF} = \frac{352{,}744.02}{36{,}231.025} = 9.736$$

α	F (2,76)	
.10	2.374	R
.05	3.117	R
.01	4.900	R

−15 IF TSS IN 11 DOES NOT MATCH TSS IN #5 OR #9.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

12. Test the null hypothesis that the linear model is adequate against the alternative that the linear model is not adequate. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

    $F_{LOF} = 9.736$ ON $(2, 76)$.

    a. Accept the null hypothesis at the 0.10 level of significance.
    b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
    c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
    d. Reject the null hypothesis at the 0.01 level.

End of application of common information

−20 NO OR WRONG TEST.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Common Information for Questions 5 to 12

A research team studied Y, the protein production of a laboratory animal, and how Y was affected by the dose of medicine. The research team sought to maximize $E(Y)$. They used four doses of medicine: 0, 1, 2, and 3 units respectively. They randomly assigned 6 animals to dosage 0, 6 to dosage 1, 6 to dosage 2, and 6 to dosage 3. They observed that the average values of Y at each dosage were
$y_{0\bullet} = 281.6, y_{1\bullet} = 309.4, y_{2\bullet} = 268.7$, and $y_{3\bullet} = 293.3$ where $y_{i.}$ was the average of the observations taken with dosage $i = 0,\ldots,3$, respectively.
They also observed that
$s_0^2 = 6250.9, s_1^2 = 6878.3, s_2^2 = 6630.4$, and $s_3^2 = 7164.5$ where s_i^2 was the unbiased estimate of the variance for the observations taken with dosage $i = 0,\ldots,3$, respectively.

DOSE	J_i	$y_{i\cdot}$	$y_{i\cdot} - y_{\cdot\cdot}$	$(y_{i\cdot} - y_{\cdot\cdot})^2$	s_i^2
0	J=6	281.6	-6.65	44.2225	6250.9
1	J=6	309.4	21.15	447.3225	6878.3
2	J=6	268.7	-19.55	382.2025	6630.4
3	J=6	293.3	5.05	25.5025	7164.5
SUM		1153.0	0.	899.25	26924.1
AVE		288.25			6731.025

SSTREAT = 6 x 899.25 = 5395.5 ON 3 DF.

MS TREAT = 5395.5 / 3 = 1798.5

ONE-WAY LAYOUT ANOVA

SOURCE	DF	SS	MS	F
DOSE	3	5,395.5	1798.5	0.267
PURE ERROR	20	134,620.5	6731.025	
TOTAL	23	140,016.0		

α	F(3,20)	
.10	2.380	A
.05	3.098	A
.01	4.938	A

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5. Complete the one-way analysis of variance table for these results; that is, be sure to specify the degrees of freedom and sum of squares for the treatment source, the error source, and the total. Also include appropriate mean squares and test statistic. This question is worth 20 points.

SOURCE	DF	SS	MS	F
DOSE	3	5 395.5	17985	0.267
PURE ERROR	20	134 620.5	6731.025	
TOTAL	23	140,016.0		

α	F(3, 20)	
.10	2.380	ACCEPT
.05	3.098	ACCEPT
.01	4.938	ACCEPT.

−15 EACH INCORRECT ANOVA ENTRY.

CORRECT FORMULA WITH COMP ERROR

−5 OR −10

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

6. Test the null hypothesis that the mean production is the same for each dosage against the alternative that the mean production is different for at least one dosage. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

   $F_{OVERALL} = 0.267$ ON $(3, 20)$.

   a. Accept the null hypothesis at the 0.10 level of significance.
   b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
   c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
   d. Reject the null hypothesis at the 0.01 level.

   WRONG TEST OR NO TEST −20

7. What is the optimal dosage? This question is worth 10 points.

   NO OPT DOSE, ALL MEANS APPEAR EQUAL

   HIGHEST AVERAGE WITH DOSAGE 1.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

8. Find the estimate of the linear contrast and its sum of squares. The coefficients of the linear contrast are $-3,-1,1,3$. This question is worth 10 points.

$$\hat{\lambda}_{LIN} = -3(281.6) - 309.4 + 268.7 + 3(293.3)$$

$$= -5.6.$$

$$SS_{LIN} = \frac{(\hat{\lambda}_{LIN})^2}{20/6} = 9.408$$

+5 FOR $\hat{\lambda}_{LIN}$

+5 FOR SS_{LIN}.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

9. Report the analysis of variance table for the linear regression of Y on dosage. Use the sum of squares due to the linear contrast as the sum of squares for the regression of Y on dosage. Include the sum of squares due to error and the total sum of squares. This question is worth 20 points.

ANOVA TABLE

SOURCE	DF	SS	MS
LINEAR CONTRAST	1	9.408	9.408
ERROR	22	140,006.592	6 363.936
TOTAL	23	140,016.0	

$$F_{LIN} = \frac{9.408}{6363.936} = + 0.0015$$

α	$F(1, 22)$	
.10	2.949	A
.05	4.301	A
.01	7.945	A

−15 TSS IN #9 DOES NOT MATCH TSS IN #5.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

10. Test the null hypothesis that there is no linear association between Y and dosage against the alternative that there is a linear association. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

$F_{LON} = 0.0015$ ON $(1, 22)$.

a. Accept the null hypothesis at the 0.10 level of significance.
b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
d. Reject the null hypothesis at the 0.01 level.

-20 NO OR WRONG TEST

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

11. Report the analysis of variance table including the sum of squares due to linear regression, the sum of squares due to lack of fit, the sum of squares due to pure error, and the total sum of squares. Include the degrees of freedom for each sum of squares. This question is worth 20 points.

ANOVA TABLE

SOURCE	DF	SS	MS
LINEAR	1	9.408	9.408
LACK OF FIT	2	5386.092	2693.046
PURE ERROR	20	134620.5	6731.025
TOTAL	23	140,016.0	

$$F_{LOF} = \frac{2693.046}{6731.025} = 0.400$$

α	$F(2,20)$
.10	2.589
.05	3.493
.01	5.849

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

12. Test the null hypothesis that the linear model is adequate against the alternative that the linear model is not adequate. What is the correct test? Circle the answer below corresponding to your conclusion. This question is worth 20 points.

$F_{LOF} = 0.400$ ON $(2, 20)$

a. Accept the null hypothesis at the 0.10 level of significance.
b. Accept the null hypothesis at the 0.05 level and reject it at the 0.10 level.
c. Accept the null hypothesis at the 0.01 level and reject it at the 0.05 level.
d. Reject the null hypothesis at the 0.01 level.

End of application of common information

−20 NO OR WRONG TEST

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

13. Let $Y = X\beta + \varepsilon$, where Y is an $n \times 1$ vector of random variables, X is an $n \times p$ matrix of known constants of full column rank p ($p<n$), β is a $p \times 1$ vector of unknown constants and ε is an $n \times 1$ vector of normally and independently distributed random variables with mean 0 and variance σ^2. That is, Y is described by the standard OLS model. Simplify $(Y - X(X^TX)^{-1}X^TY)^T(X(X^TX)^{-1}X^TY)$. Give the matrix dimensions of your answer. This problem is worth 50 points.

End of Examination

$$(Y - X(X^TX)^{-1}X^TY)^T(X(X^TX)^{-1}X^TY).$$

$$= Y^T(I - X(X^TX)^{-1}X^T)^T(X(X^TX)^{-1}X^T)Y.$$

$$= Y^T(I^T - (X(X^TX)^{-1}X^T)^T)(X(X^TX)^{-1}X^T)Y.$$

$$= Y^T(I - X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)Y.$$

$$= Y^T[X(X^TX)^{-1}X^T - X(X^TX)^{-1}\{X^T X(X^TX)^{-1}\}X^T]Y.$$

$$= Y^T[X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T]Y.$$

$$= Y^T[0_{n\times n}]Y = 0_{1\times 1}.$$

ERRORS: $X(X^TX)^{-1}X^T = X\frac{1}{(X^TX)}X^T = \frac{X}{X}\cdot\frac{X^T}{X^T}$ -50

+5 ANSWER IS 1×1.