AMS 315 S2021 EXAMINATION 2

GRADING DEDUCTIONS

1A −10 EACH INCORRECT ANOVA TABLE ENTRY.

DO NOT PENALIZE A CONSISTENCY ERROR.

−15 WRONG DV.

−20 NO DECISION OR INCONSISTENT DECISION.

1B −10 WRONG SLOPE.

−5 WRONG T VALUE.

1C −20 REVERSE PI AND CI FOR $\beta_0 + \beta_1 x$.

1D RIGHT OR WRONG. NO PARTIAL CREDIT.

2 +15 CORRECT F(p).

−15 INCORRECT $|z_\alpha|$

−15 INCORRECT $|z_\beta|$.

−25 FORGET TO SQUARE.

3 +15 CORRECT F(r).

+10 CORRECT CI FOR F(p).

−25 INCORRECT INVERSION.

4 A −10 EACH INCORRECT PARTIAL

−15 REVERSING PARTIALS.

B. NO PARTIAL CREDIT: RIGHT OR WRONG

C −10 EACH INCORRECT ANOVA ENTRY

−20 WRONG VARIABLE SEQUENCE.

−30 NO DECISION OR

INCONSISTENT DECISION

5 A RIGHT OR WRONG.

5B. −25 USE $\sigma^2 I$ AS VCV(Y), NOT $\sigma^2 V$.

−10 EACH INCORRECT

MATRIX OPERATION.

−20 DON'T USE VCV(MY) = M VCV(Y) M^T.

AMS 315, Examination 2, April 1, 2021

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

1. A research team studied the response of a participant to a dosage of medication. Dosages were randomly assigned to participants. The research team then measured each participant's response for $n = 842$ participants. The average response was 231.4 (higher is better), with an observed standard deviation of 57.4 (the divisor in the underlying variance calculation was $n-1$). The average dosage was 28.2, with an observed standard deviation of 13.9 (the divisor in the underlying variance calculation was also $n-1$). The correlation coefficient between the two variables was 0.34. The team sought to estimate the regression of participant response on the dosage of medication.
    a. Complete the analysis of variance table for the regression of participant response on the dosage of medicine given the participant. Test the null hypothesis that the slope of this regression is zero at levels of significance 0.10, 0.05, and 0.01. This part is worth 25 points.
    b. Find the estimated regression equation of participant response on dosage. Find the 95% confidence interval for the slope in this equation. This part is worth 15 points.
    c. Find the estimated regression equation of participant response on dosage. Use this equation to estimate the response for participants whose dosage was 50.0 What is the 95% confidence interval for this response score? This part is worth 20 points.
    d. Discuss whether the research team can conclude that the medicine causes the higher response or is only associated with the higher response. This part is worth 10 points.

A1A $TSS = 841(57.4)^2 = 2{,}770{,}893.16$

$SSREG = (.34)^2 TSS = 320{,}315.25$

$SSERR = (1-.34^2)TSS = 0.8844(TSS) = 2{,}450{,}577.91$

ANOVA TABLE

SOURCE	DF	SS	MS	F
REG	1	320,315.25	320,315.25	109.80
ERROR	840	2,450,577.91	2917.35	
TOTAL	841	2,770,893.16		

α	$F(1, 840)$	
.10	2.712	REJECT
.05	3.853	REJECT
.01	6.665	REJECT

REJECT H_0: NO LINEAR ASSOCIATION AT $\alpha = .01$ (AND .05 AND .01)

−10 EACH ANOVA TABLE ENTRY ERROR

−20 NO DECISION OR INCONSISTENT DECISION

A1B. $\sum(x_i - \bar{x}_n)^2 = 841(13.9)^2 = 162,489.61$

$$\hat{\beta}_1 = r\frac{SD_{DV}}{SD_{IV}} = 0.34\left(\frac{57.4}{13.9}\right) = 1.404$$

$$\hat{y}(x) = 231.4 + 1.404(x - 28.2) = 191.81 + 1.404x.$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{2917.35}{162,489.61}} = \sqrt{0.01795} = 0.134$$

95% CI FOR β_1: $1.404 \pm 1.963(0.134)$

$= 1.404 \pm 0.263 = 1.141$ TO 1.667

A1C 95% CI FOR $\beta_0 + 50\beta_1$

$$\hat{y}(50) = 231.4 + 1.404(50 - 28.2) = 231.4 + 1.404(21.8)$$

$= 262.0$

$$SE(\hat{y}(50)) = \sqrt{2917.35\left(\frac{1}{842} + \frac{(21.8)^2}{162,489.61}\right)}$$

$$= \sqrt{2917.35(.001188 + 0.002947)} = \sqrt{11.997}$$

$= 3.46.$

95% CI FOR $\beta_0 + 50\beta_1$: $262.0 \pm 1.963(3.46)$

$= 262.0 \pm 6.80 = 255.2$ TO 268.8

A1D: HIGHER DOSAGE OF MEDICINE CAUSES HIGHER RESPONSE. DOSAGE WAS RANDOMLY ASSIGNED TO PARTICIPANTS.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

1. A research team studied the response of a participant to dosage of medication in an observational study. The research team measured each participant's response $n = 234$ participants. The average response was 324.5, with an observed standard deviation of 61.7 (the divisor in the underlying variance calculation was $n-1$). The research team then extracted the participant's dosage from the participant's medical record. The average dosage was 168.3, with an observed standard deviation of 15.2 (the divisor in the underlying variance calculation was also $n-1$). The correlation coefficient between the two variables was 0.21. The team sought to estimate the regression of participant response on the dosage of medication.

   a. Complete the analysis of variance table for the regression of participant response on the dosage of medicine given the participant. Test the null hypothesis that the slope of this regression is zero at levels of significance 0.10, 0.05, and 0.01. This part is worth 25 points.

   b. Find the 99% confidence interval for the slope in this equation. This part is worth 15 points.

   c. Find the estimated regression equation of participant response on dosage. Use this equation to estimate the response for a future participant whose dosage will be 200.0 What is the 99% prediction interval for this participant's response? This part is worth 20 points.

   d. Discuss whether the research team can conclude that the medicine causes the higher response or is only associated with the higher response. This part is worth 10 points.

B1A: $TSS = (n-1)\,SD_{DV}^2 = 233(61.7)^2 = 887{,}005.37$

$\Sigma(x_i - \bar{x}_m)^2 = (n-1)\,SD_{IV}^2 = 233(15.2)^2 = 53{,}832.32.$

$REG\,SS = (r_{YX})^2\,TSS = (.21)^2 TSS = 39{,}116.94.$

$SSE = (1 - r_{YX}^2)TSS = (1 - .21^2)TSS = 0.9559\,TSS$

$= 847{,}888.43$ ON 232 DF

$MSE = \dfrac{SSE}{232} = 3654.69.$

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	SS	MS	F
REG	1	39,116.94	39,116.94	10.70
ERROR	232	847,888.43	3,654.69	
TOTAL	233	887,005.37		

α	$F(1,232)$	
.10	2.727	REJECT
.05	3.882	REJECT
.01	6.745	REJECT

REJECT H_0 NO ASSOCIATION BETWEEN DOSAGE AND RESPONSE AT $\alpha = .01$ (AND $\alpha = .05$ AND .10)

B1B. $\hat{\beta}_1 = r \frac{SD_{DV}}{SD_{IV}} = 0.21 \left(\frac{61.7}{15.2}\right) = 0.852.$

$SE(\hat{\beta}_1) = \sqrt{\frac{3654.69}{53,832.32}} = \sqrt{.06789} = 0.2606$

99% CI FOR β_1: $0.852 \pm 2.597(0.2606)$

$= 0.852 \pm 0.677 = 0.175$ TO 1.529.

B1C $\hat{y}(x) = 324.5 + 0.852(x - 168.3)$

$= 181.1 + 0.852x$

$\hat{y}(200) = 324.5 + 0.852(200 - 168.3)$

$= 324.5 + 0.852(31.7) = 351.5.$

99% PI FOR $Y_F(200)$:

$351.5 \pm 2.597\sqrt{3,654.69\left(1 + \frac{1}{234} + \frac{(31.7)^2}{53,832.32}\right)}$

$= 351.5 \pm 2.597\sqrt{3,654.69(1 + 0.00427 + 0.01867)}$

$= 351.5 \pm 2.597\sqrt{3,654.69(1.0229)}$

$= 351.5 \pm 2.597\sqrt{3,738.53} = 351.5 \pm 2.597(61.1)$

$= 351.5 \pm 158.8 = 192.7$ TO 510.3

B1D THE TEAM MAY ONLY CONCLUDE ASSOCIATION. THERE WAS NO RANDOM ASSIGNMENT OF DOSE TO PARTICIPANTS

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

2. A research team wishes to test the null hypothesis H_0: $\rho = 0$ at $\alpha = 0.025$ against the alternative H_1: $\rho > 0$ using Fisher's transformation of the Pearson product moment correlation coefficient as the test statistic. They have asked their consulting statistician for a sample size n such that $\beta = 0.05$ when $\rho = 0.15$. What is this value? This problem is worth 40 points.

3. A research team studied the association between dosage of a medicine and participant response for $n = 631$ participants. They observed a Pearson product moment correlation coefficient of 0.29. What is the 99% confidence interval for the population correlation coefficient. This problem is worth 50 points.

C2. $\sqrt{n-3} \geq \dfrac{1.960(1) + 1.645(1)}{0.1511} = 23.85$

$F(0.15) = \frac{1}{2} \ln\left(\frac{1+.15}{1-.15}\right) = \frac{1}{2} \ln(1.3529) = 0.1511$

$n-3 \geq 569.03$, $n \geq 573$.

-15 NO 1.960

-15 NO 1.645

-25 FORGET TO SQUARE.

+15 $F(0.15) = 0.154$.

C3. $F(0.29) = \frac{1}{2} \ln\left(\frac{1+.29}{1-.29}\right) = \frac{1}{2} \ln\left(\frac{1.29}{.71}\right) = \frac{1}{2} \ln(1.817) = 0.299$.

99% CI FOR $F(\rho)$: $0.299 \pm 2.576\sqrt{\frac{1}{628}} = 0.299 \pm 2.576(0.0399)$

$= 0.299 \pm 0.103 = 0.196$ TO 0.4018.

THE INVERSION FORMULA IS $\rho = \dfrac{e^{2F(\rho)} - 1}{e^{2F(\rho)} + 1}$.

LEFT END POINT: $e^{2(0.196)} = 1.4799$. $\rho_L = \dfrac{1.4799 - 1}{1.4799 + 1} = 0.194$.

RIGHT END POINT $e^{2(0.4018)} = 2.234$ $\rho_R = \dfrac{2.234 - 1}{2.234 + 1} = 0.381$.

THE 99% CI FOR ρ IS .194 TO 0.381

DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD

2. A research team wishes to test the null hypothesis H_0: $\rho = 0$ at $\alpha = 0.005$ against the alternative H_1: $\rho > 0$ using Fisher's transformation of the Pearson product moment correlation coefficient as the test statistic. They have asked their consulting statistician for a sample size n such that $\beta = 0.01$ when $\rho = 0.12$. What is this value? This problem is worth 40 points.

3. A research team studied the association between dosage of a medicine and participant response for $n = 248$ participants. They observed a Pearson product moment correlation coefficient of 0.79. What is the 95% confidence interval for the population correlation coefficient. This problem is worth 50 points.

D2. $F(0.12) = \frac{1}{2}\ln\left(\frac{1+.12}{1-.12}\right) = \frac{1}{2}\ln\left(\frac{1.12}{.88}\right) = \frac{1}{2}\ln(1.273)$

$= 0.1206$ +15 FOR $F(\rho)$.

$\sqrt{n-3} \geq \frac{2.576(1) + 2.326(1)}{0.1206} = 40.66.$

$n-3 \geq 1652.98$

$n \geq 1656.$

D3. $F(0.79) = \frac{1}{2}\ln\left(\frac{1+.79}{1-.79}\right) = \frac{1}{2}\ln\left(\frac{1.79}{0.21}\right) = 1.071$

$\frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{245}} = 0.0639.$

95% CI FOR $F(\rho)$: $1.071 \pm 1.96(0.0639)$

$= 1.071 \pm 0.125 = .9458$ TO 1.196

INVERSION FORMULA $\rho = \frac{e^{2F(\rho)} - 1}{e^{2F(\rho)} + 1}.$

LEFT END POINT $e^{2(.9458)} = 6.630$ $\rho_L = \frac{6.630-1}{6.630+1} = 0.738$

RIGHT END POINT $e^{2(1.196)} = 10.935$ $\rho_R = \frac{10.935-1}{10.935+1} = 0.832.$

THE 95% CI FOR ρ IS 0.738 TO 0.832.

EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE

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 adult responsibility at age 26 (with a higher number indicating greater responsibility). The variable x was a measure of the participant's educational achievement at age 22 (higher values meant greater achievement); and the variable w was a measure of the participant's responsibility at age 16 (higher values meant greater responsibility). They observed values of these three variables on $n = 532$ participants. The mean and variance of responsibility at age 26 (using $n-1$ as divisor) were 138.7 and 613.8 respectively. The mean and variance of educational achievement at age 22 were 12.3 and 27.3 respectively. The mean and variance of responsibility at age 16 were 101.8 and 481.2 respectively. The correlation between Y and w was 0.36, the correlation between Y and x was 0.19; and the correlation between x and w was 0.49.
   a. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$. This part is worth 15 points.
   b. Is a mediation model or an explanation model a better explanation of the observed results? This part is worth 15 points.
   c. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on x and the sum of squares due to the regression on w after including x. Test the null hypothesis that $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. Report whether the test is significant at the 0.10, 0.05, and 0.01 levels of significance. This part is worth 40 points.

E4A. $$r_{Yx.w} = \frac{0.19-(0.36)(0.49)}{\sqrt{(1-(.36)^2)(1-.49^2)}} = \frac{0.0136}{\sqrt{0.8704(0.7599)}}$$

$$= \frac{0.0136}{\sqrt{0.6614}} = \frac{0.0136}{0.8133} = 0.0167$$

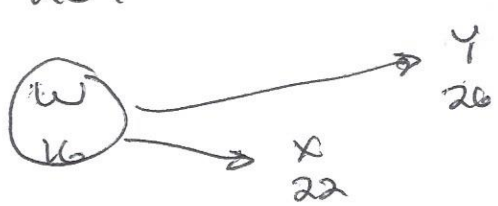
NOTE $\sqrt{\frac{1}{n-3}} = \sqrt{\frac{1}{529}} = 0.0435.$

$$r_{Yw.x} = \frac{0.36-0.19(0.49)}{\sqrt{(1-.19^2)(1-.49^2)}} = \frac{0.2669}{\sqrt{.9639(.7599)}}$$

$$= \frac{0.2669}{\sqrt{0.73247}} = \frac{0.2669}{0.8558} = 0.3119.$$

E4B. KEY VARIABLE IS w.

EXPLANATION:



E4C $TOTSS = (531)(613.8) = 325,927.8$

$REGSS(x) = (0.19)^2\ TOTSS = 11,766.0$

$SSE(x) = (1-(0.19)^2)\ TOTSS = (0.9639)\ TOTSS$

$= 314,161.8$

$REGSS(w|x) = (r_{Yw.x})^2\ SSE(x)$

$= (0.3119)^2\ SSE(x) = 30,562.2$

$SSE = TOTSS - REGSS(x) - REGSS(w|x) = 283599.6$

ANOVA TABLE

SOURCE	DF	SS	MS	
REG(x)	1	11,766.0	11766.0	
REG(w|x)	1	30,562.2	30,562.2.	57.0
ERROR	529	283,599.6	536.1.	
TOTAL	531	325,927.8		

α	$F(1,529)$.	
.10	2.715	REJECT
.05	3.859.	REJECT
.01	6.683	REJECT

REJECT $H_0\ \beta_2 = 0$ VS $H_1\ \beta_2 \neq 0$ AT $\alpha = .01$ (AND $\alpha = .05$ AND $\alpha = .10$).

$$F_{w|x} = \frac{30562.2}{536.1} = 57.0$$

FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF

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 adult criminality at age 22 (with a higher number indicating greater criminality). The variable x was a measure of the participant's delinquency at age 18 (higher values meant greater delinquency); and the variable w was a measure of the participant's rebelliousness at age 14 (higher values meant greater responsibility). They observed values of these three variables on $n = 487$ participants. The mean and variance of rebelliousness at age 14 (using $n-1$ as divisor) were 18.7 and 63.8 respectively. The mean and variance of delinquency at age 18 were 22.3 and 87.3 respectively. The mean and variance of criminality at age 22 were 42.8 and 118.2 respectively. The correlation between Y and w was 0.44, the correlation between Y and x was 0.59; and the correlation between x and w was 0.71.
   a. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$. This part is worth 15 points.
   b. Is a mediation model or an explanation model a better explanation of the observed results? This part is worth 15 points.
   c. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on x and the sum of squares due to the regression on w after including x. Test the null hypothesis that $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. Report whether the test is significant at the 0.10, 0.05, and 0.01 levels of significance. This part is worth 40 points.

F4A. $r_{yx\cdot w} = \dfrac{0.59-(0.44)(0.71)}{\sqrt{(1-.44^2)(1-.71^2)}} = \dfrac{0.2776}{\sqrt{(.8064)(0.4958)}}$

$= \dfrac{0.2776}{\sqrt{0.4000}} = \dfrac{0.2776}{0.63237} = 0.4390$

$r_{yw\cdot x} = \dfrac{0.44 - 0.59(0.71)}{\sqrt{(1-.59^2)(1-.71^2)}} = \dfrac{0.0211}{\sqrt{0.6519(.4958)}}$

$= \dfrac{0.0211}{\sqrt{0.32328}} = \dfrac{0.0211}{0.56857} = 0.0371.$

$\dfrac{1}{\sqrt{n-3}} = \dfrac{1}{\sqrt{484}} = 0.0455$

F4B. KEY VARIABLE IS X MEDIATION MODEL

$w \longrightarrow x \longrightarrow Y$

14 18 22

F4C $TSS = (n-1)\,VAR(Y) = 486(118.21)$

$= 57445.2$

$REGSS(x) = (r_{Yx})^2\,TSS = (.59)^2\,TSS = 19,996.7$

$SSE(x) = (1 - r^2_{Yx})\,TSS = 0.6519 \times TSS = 37,448.5$

$REGSS(w|x) = (r^2_{Yw.x})\,SSE(x) = (0.0371)^2\,SSE(x)$

$= 51.5.$

$SSE(x,w) = SSE(x) - REGSS(w|x) = 37,397.0$

ON. 484 DF

$MSE(x,w) = 77.3.$

ANOVA TABLE

SOURCE	DF	SS	MS	F
REG(x)	1	19,996.7		
REG(w|x)	1	51.5	51.5	0.67.
ERROR	484	37,397.0	77.3	
TOTAL	486	57,445.2		

$$F_{w|x} = \frac{MS\,REG(w|x)}{MSE} = \frac{51.5}{77.3} = 0.67$$

α	$F(1,484)$	
.10	2.716	ACCEPT
.05	3.861	ACCEPT
.01	6.688	ACCEPT.

ACCEPT $H_0: \beta_2 = 0$ VS.

$H_1\ \beta_2 \neq 0$ AT $\alpha = .10$

(AND $\alpha = .05$ AND $\alpha = .01$).

GGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

5. The random vector Y is $n \times 1$, with $Y = X\beta + \varepsilon$, where β is a $p \times 1$ vector of (unknown) constants, X is an $n \times p$ matrix of known constants with $rank(X) = p$, ε is an $n \times 1$ vector of random variables with $E(\varepsilon) = 0$ and $vcv(Y) = vcv(\varepsilon) = \sigma^2 V$, where V is a symmetric positive definite $n \times n$ matrix. The matrix V is not an identity matrix. The ordinary least squares estimate of β is $\hat{\beta} = (X^T X)^{-1} X^T Y$.
   a. Find $E(\hat{\beta})$. This part is worth 10 points.
   b. Find the variance-covariance matrix of $\hat{\beta}$. This part is worth 40 points.

End of the Examination

G5A $E(\hat{\beta}) = E[(X^T X)^{-1} X^T Y]$

$= (X^T X)^{-1} X^T (EY)$

$= (X^T X)^{-1} X^T (X\beta) = \beta.$

G5B. $vcv(\hat{\beta}) = vcv(MY)$, WHERE $M = (X^T X)^{-1} X^T$.

$vcv(MY) = M\, vcv(Y)\, M^T$

$= (X^T X)^{-1} X^T [\sigma^2 V][(X^T X)^{-1} X^T]^T$

$= (X^T X)^{-1} X^T (\sigma^2 V) X (X^T X)^{-1}$

$= \sigma^2 (X^T X)^{-1} (X^T V X)(X^T X)^{-1}.$