AMS 315, Fall 2021

Midterm Examination 3

Solutions and Penalties

November 18, 2021

AMS 315 F2021 Midterm 3 Grading Penalties

1. -10 for each incorrect partial correlation; -20 for correct partial correlations that are mislabeled.
2. 30 points for a correct answer; 0 for an incorrect answer.
3. Standard anova table penalties: -15 for an incorrect degrees of freedom; -10 for an incorrect entry. Only penalize an error when it is made; do not penalize subsequent correct uses of an incorrect entry. No decision or inconsistent decision about the hypothesis test: -35.
4. -5 points for each incorrect SS; -10 points for incorrect Scheffe confidence interval.
5. Standard anova table penalties: -15 for an incorrect degrees of freedom; -10 for an incorrect entry. Only penalize an error when it is made; do not penalize subsequent correct uses of an incorrect entry. No decision or inconsistent decision about the hypothesis test: -25.
6. Standard anova table penalties: -15 for an incorrect degrees of freedom; -10 for an incorrect entry. Only penalize an error when it is made; do not penalize subsequent correct uses of an incorrect entry. No decision or inconsistent decision about the hypothesis test: -35. Incorrect or missing optimal setting: -10.
7. +10 for each correct and directed step
8. a. Correct 10; incorrect -10.
   b. +10 for each correct and directed step. No further credit for a major mistake in matrix computations. For example, $(X^TV^{-1}X)^{-1} \neq X^{-1}(V^{-1})^{-1}(X^T)^{-1}$ because X^{-1} does not exist as the matrix X is not square.

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

A. 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 behavior at age 22 (with a higher number indicating greater anti-social behavior). The variable x was a measure of the participant's depression at age 18 (higher values meant greater depression); and the variable w was a measure of the participant's rebelliousness at age 15 (higher values meant greater rebelliousness). They observed values of y, x, and w on $n = 534$ subjects. The mean and variance of Y (using $n-1$ as divisor) were 18.4 and 35.8 respectively. The mean and variance of x were 21.1 and 12.6 respectively. The mean and variance of w were 41.8 and 64.9 respectively. The correlation between Y and w was 0.58, the correlation between Y and x was 0.23; and the correlation between x and w was 0.39.

1. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$. This is worth 20 points.
2. 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. This question is worth 30 points.
3. 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 question is worth 50 points.

End of application of common information

1. $$r_{Yx\cdot w} = \frac{0.23-(0.39)(0.58)}{\sqrt{(1-.39^2)(1-.58^2)}} = \frac{0.0038}{\sqrt{(.8479)(0.6636)}}$$

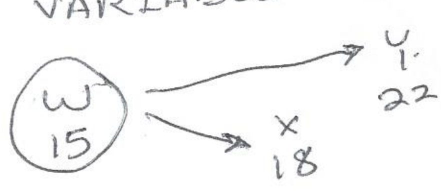
$$= \frac{0.0038}{\sqrt{0.56267}} = \frac{0.0038}{0.7501} = 0.00507.$$

$$r_{Yw\cdot x} = \frac{0.58-(0.23)(0.39)}{\sqrt{(1-.23^2)(1-.39^2)}} = \frac{0.4903}{\sqrt{(0.9471)(0.8479)}}$$

$$= \frac{0.4903}{\sqrt{0.80305}} = \frac{0.4903}{0.89613} = 0.54713$$

$$se(r_{Yx\cdot w}) = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{531}} = 0.043$$

2. KEY VARIABLE IS w!



EXPLANATION MODEL

A3. $TSS = (n-1)\, SD_Y^2 = 533(35.8) = 19{,}081.4$

$SSREG(x) = (r_{YX})^2\, TSS = (0.23)^2\, TSS$

$= 1009.4.$

$TSS - SSREG(x) = 18071.99.$

$SSREG(w|x) = (r_{Yw \cdot x})^2 (TSS - SSREG(x))$

$= (0.54713)^2 (TSS - SSREG(x)) = 5409.9.$

$SSERR = TSS - SSREG(x) - SSREG(w|x)$

$= 12662.1$ ON 531 DF

$MSE = 23.85.$

ANOVA TABLE

SOURCE	DF	SS	MS	F
REG(x)	1	1009.4	1009.4	226.9.
REG(w|x)	1	5409.9	5409.9	
ERROR	531	12,662.1	23.85	
TOTAL	533	19,081.4		

$$F_{w|x} = \frac{MS\ REG(w|x)}{MSE} = \frac{5409.9}{23.85} = 226.9$$

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

α	$F(1,531)$		$F(1,\infty)$.
.10	2.715	R	2.71
.05	3.859	R	3.84
.01	6.683	R.	6.63

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

B. 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 behavior at age 22 (with a higher number indicating greater anti-social behavior). 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 15 (higher values meant greater rebelliousness). They observed values of y, x, and w on $n = 618$ subjects. The mean and variance of Y (using $n-1$ as divisor) were 28.4 and 25.8 respectively. The mean and variance of x were 31.1 and 42.6 respectively. The mean and variance of w were 21.8 and 16.9 respectively. The correlation between Y and w was 0.44, the correlation between Y and x was 0.58; and the correlation between x and w was 0.71.

1. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$. This is worth 20 points.
2. 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. This question is worth 30 points.
3. 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 question is worth 50 points.

End of application of common information

1. $$r_{Yx\cdot w} = \frac{0.58 - 0.44(0.71)}{\sqrt{(1-.44^2)(1-.71^2)}} = \frac{0.2676}{\sqrt{(0.8064)(0.4959)}}$$

$$= \frac{0.2676}{\sqrt{0.39989}} = \frac{0.2676}{0.63237} = 0.42317.$$

$$r_{Yw.x} = \frac{0.44 - (0.58)(0.71)}{\sqrt{(1-.58^2)(1-.71^2)}} = \frac{0.0282}{\sqrt{(.6636)(0.4959)}}$$

$$= \frac{0.0282}{\sqrt{0.32908}} = \frac{0.0282}{0.57365} = 0.04916$$

$$se(r_{Yw.x}) \cong \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{615}} = 0.0403$$

2. THE KEY VARIABLE IS x.

$$\underset{15}{w} \longrightarrow \underset{18}{x} \longrightarrow \underset{22}{Y}$$

THIS IS A MEDIATION MODEL

B3. $TSS = (n-1)\,SD_Y^2 = 617\,(25.8) = 15918.6$

$SSREG(x) = r_{Yx}^2 \; TSS = (0.58)^2\,TSS = 5355.0$

$TSS - SSREG(x) = 10563.6$

$SSREG(w|x) = r_{rw.x}^2\,(TSS - SSREG(x)) = (0.04916)^2\,(10{,}563.6)$

$= 25.5.$

$SSERR = TSS - SSREG(x) - SSREG(w|x)$

$= 10538.1$ ON 615 DF

$MSE = 17.1.$

ANOVA TABLE

SOURCE	DF	SS	MS
REG(x)	1	5,355.0	
REG(w|x)	1	25.5	25.5
ERROR	615	10,538.1	17.1
TOTAL	617	15,918.6	

$$F_{w|x} = \frac{MSREG(w|x)}{MSE} = \frac{25.5}{17.1} = 1.49.$$

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

α	F (1, 615)		F (1, ∞).
.10	2.714	A	2.71
.05	3.857	A	3.84
.01	6.676	A	6.63

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

C. A research team studied how Y, the protein production of a laboratory animal, could be minimized by choice of dosage of a medicine. They used four doses: 1, 2, 3, and 4 units. They randomly assigned 18 animals to 1 unit, 18 to 2 units, 18 to 3 units, and 18 to 4 units. The average values of Y at each dosage were $y_{1\bullet} = 570$, $y_{2\bullet} = 536$, $y_{3\bullet} = 462$, and $y_{4\bullet} = 444$, where $y_{i.}$ was the average of the observations taken with dosage $i = 1,2,3,4$. The within dosage variances were $s_1^2 = 10{,}564$, $s_2^2 = 12{,}832$, $s_3^2 = 8{,}466$, and $s_4^2 = 11{,}102$. They found that $y_{\bullet\bullet} = 503$ and that the average s_i^2 was 10,741. The total sum of squares was 923,708. The coefficients of the linear contrast were $-3,-1,1,3$;, and $\hat{\lambda}_{Lin} = -452.0$ The coefficients of the quadratic contrast were $1,-1,-1,1$; and $\hat{\lambda}_{Quad} = 16.0$. The coefficients of the cubic contrast were $-1,3,-3,1$; and $\hat{\lambda}_{Cubic} = 96.0$.

4. What are the values of the sum of squares due to the linear contrast, the sum of squares dues to the quadratic contrast, and the sum of squares due to the cubic contrast? What is the 99% Scheffe confidence interval for λ_{Lin}? (25 points)
5. Find the analysis of variance table for the linear regression of Y on dosage, using the sum of squares due to the linear contrast as the sum of squares for the regression of Y on dosage. Test the null hypothesis that there is no linear association at the 0.10, 0.05, and 0.01 levels of significance. (40 points)
6. Test the null hypothesis that the linear model is adequate at the 0.10, 0.05, and 0.01 levels of significance. Report the analysis of variance table including the sum of squares due to lack of fit. What dosage appears to be optimal. (50 points).

End of application of common information

4. LINEAR, $\sum a_i^2 = 20$, $SS_{LIN} = \dfrac{(\hat{\lambda}_{LIN})^2}{20/18} = \dfrac{(-452.0)^2}{20/18} = 183{,}873.6$

QUAD, $\sum a_i^2 = 4$, $SS_{QUAD} = \dfrac{(\hat{\lambda}_{QUAD})^2}{4/18} = \dfrac{(16.0)^2}{4/18} = 1152.0$

CUBIC, $\sum a_i^2 = 20$, $SS_{CUBE} = \dfrac{(\hat{\lambda}_{CUBE})^2}{20/18} = \dfrac{(96.0)^2}{20/18} = 8294.4.$

$SS_{LIN} + SS_{QUAD} + SS_{CUB} = 193{,}320.$

$SS_{TREAT} = SS_{TOTAL} - 68\,MSE = 923{,}708 - 68(10{,}741) = 193{,}320$

RESULTS CHECK OUT. 99% CI FOR λ_{LIN}: $-452 \pm \sqrt{3(4.083)}\sqrt{10{,}741\,\frac{20}{18}}$

$= -452 \pm 3.50(109.2) = -452 \pm 382.4$

5 ANOVA TABLE LINEAR REGRESSION ON DOSE

SOURCE	DF	SS	MS	F
REGRESSION	1	183,873.6	183,873.6	17.4.
ERROR	70	739,834.4	10,569.1	
TOTAL	71	923,708.0		

α	$F(1,70)$	
.10	2.779	R
.05	3.978	R
.01	7.011	R

REJECT H_0: NO LINEAR ASSOCIATION AT $\alpha = .01$ (AND $\alpha = .05$ AND $\alpha = .10$).

C6

ANOVA TABLE LACK OF LINEAR FIT

SOURCE	DF	SS	MS
LINEAR REGRESSION	1	183,873.6	
LACK OF LINEAR FIT	2	9,446.4	4,723.2
PURE ERROR	68	730,388	10,741.0
TOTAL	71	923,708.0	

$SS_{LOF} = SS_{QUAD} + SS_{CUBIC} = 1152.0 + 8294.4 =$

$F_{LOF} = \dfrac{4723.2}{10,741.0} = 0.44.$ ON (2, 68)

α	F(2,68)	
.10	2.382	A
.05	3.132	A
.01	4.932	A.

ACCEPT H_0: LINEAR MODEL ADEQUATE AT $\alpha = .10$

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

DOSE 4 OR HIGHER IS OPTIMAL, THE ASSOCIATION

IS STRONGLY NEGATIVE AND LINEAR

DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD

D. A research team studied how Y, the protein production of a laboratory animal, could be maximized by choice of dosage of a medicine. They used four doses: 1, 2, 3, and 4 units. They randomly assigned 23 animals to 1 unit, 23 to 2 units, 23 to 3 units, and 23 to 4 units. The average values of Y at each dosage were $y_{1\bullet} = 372,\ y_{2\bullet} = 482,\ y_{3\bullet} = 458,$ and $y_{4\bullet} = 364$, where $y_{i.}$ was the average of the observations taken with dosage $i = 1,2,3,4$. The within dosage variances were $s_1^2 = 10{,}026, s_2^2 = 13{,}048,\ s_3^2 = 9{,}054,$ and $s_4^2 = 10{,}208$. They found that $y_{\bullet\bullet} = 419$ and that the average s_i^2 was 10,584. The total sum of squares was 1,178,044. The coefficients of the linear contrast were $-3,-1,1,3;$, and $\hat{\lambda}_{Lin} = -48.0$ The coefficients of the quadratic contrast were $1,-1,-1,1;$ and $\hat{\lambda}_{Quad} = -204.0$. The coefficients of the cubic contrast were $-1,3,-3,1;$ and $\hat{\lambda}_{Cubic} = 64.0$.

4. What are the values of the sum of squares due to the linear contrast, the sum of squares dues to the quadratic contrast, and the sum of squares due to the cubic contrast? What is the 99% Scheffe confidence interval for λ_{Quad}? (25 points)
5. Find the analysis of variance table for the linear regression of Y on dosage, using the sum of squares due to the linear contrast as the sum of squares for the regression of Y on dosage. Test the null hypothesis that there is no linear association at the 0.10, 0.05, and 0.01 levels of significance. (40 points)
6. Test the null hypothesis that the linear model is adequate at the 0.10, 0.05, and 0.01 levels of significance. Report the analysis of variance table including the sum of squares due to lack of fit. What dosage appears to be optimal? (50 points).

End of application of common information

4. LINEAR, $\sum a_i^2 = 20$, $SS_{LIN} = \dfrac{(\hat{\lambda}_L)^2}{20/23} = \dfrac{(-48.0)^2}{20/23} = 2649.6$

QUADRATIC, $\sum a_i^2 = 4$, $SS_{QUAD} = \dfrac{(\hat{\lambda}_{QUAD})^2}{4/23} = \dfrac{(-204.0)^2}{4/23} = 239{,}292$

CUBIC, $\sum a_i^2 = 20$, $SS_{CUBIC} = \dfrac{(\hat{\lambda}_{CUBIC})^2}{20/23} = \dfrac{(64.0)^2}{20/23} = 4710.4.$

CHECK CALCULATIONS:

$SS_{LIN} + SS_{QUAD} + SS_{CUBIC} = 2649.6 + 239{,}292.0 + 4{,}710.4 = 246{,}652$

$SS_{TREAT} = SS_{TOTAL} - 88(MSPE) = 1{,}178{,}044 - 88(10{,}584) = 246{,}652.$

RESULTS CHECK OUT.

99% SCHEFFE CI FOR λ_{QUAD}: $-204.0 \pm \sqrt{3(4.012)}\sqrt{10{,}584\left(\frac{4}{23}\right)}.$

$= -204.0 \pm 3.469(42.90) = -204.0 \pm 148.8 = -352.8$ TO $-55.2.$

5D ANALYSIS OF VARIANCE TABLE LINEAR REGRESSION

SOURCE	DF	SS	MS.	F
LINEAR REGRESSION	1	2,649.6	2649.6	0.20
ERROR	90	1,175,394.4	13,059.9	
TOTAL	91	1,178,044		

α	F(1, 90)	
.10		A
.05		A
.01		A

ACCEPT H_0: NO LINEAR ASSOCIATION BETWEEN DOSE AND Y_i AT $\alpha = .10$ (AND $\alpha = .05$ AND $\alpha = .01$).

6D ANALYSIS OF VARIANCE TABLE LACK OF FIT.

SOURCE	DF	SS	MS
LINEAR	1	2,649.6	
LACK OF FIT	2	244,002.4	122,001.2
PURE ERROR	88	931,392.0	10,584
TOTAL	91	1,178,044	

$$SS_{LOF} = SS_{QUAD} + SS_{CUBIC} = 239,292 + 4,710.4 =$$

$$F_{LOF} = \frac{MSLOF}{MSPE} = \frac{122,001.2}{10,584} = 11.53$$

α	F(2, 88)	
.10	2.364	R
.05	3.100	R
.01	4.855	R

REJECT H_0 LINEAR MODEL IS ADEQUATE AT $\alpha = .01$ (AND $\alpha = .05$ AND $\alpha = .10$).

THERE IS A STRONG QUADRATIC PATTERN CONVEX DOWNWARD. THE LARGEST OBSERVED MEAN WAS FOR DOSAGE 2. BEST SETTING MAY BE BETWEEN 2 AND 3.

EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE

7. The random variable $Y > 0$ has mean μ_Y and standard deviation σ_Y such that
$$\ln(\sigma_Y) = \beta_0 + \beta_1 \ln(\mu_Y).$$
Let f be a differentiable function with continuous first derivative. For $W = f(Y)$, find the approximate value of $var(W)$. Prove your result. This problem is worth 60 points.

$$VAR(w) \cong [f'(\mu_Y)]^2 \, VAR(Y)$$

$$\sigma_Y = e^{[\beta_0 + \beta_1 \ln(\mu_Y)]} = e^{\beta_0} e^{\ln[(\mu_Y)^{\beta_1}]}$$

$$= e^{\beta_0} \mu_Y^{\beta_1}.$$

$$VAR(w) \cong f'(\mu_Y)^2 e^{2\beta_0} (\mu_Y^{\beta_1})^2$$

FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF

8. The random vector Y is $n \times 1$, with $Y = X\beta + \epsilon$, 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 $E(\epsilon) = 0$ and $vcv(\epsilon) = V$, where V is a positive definite symmetric $n \times n$ matrix. That is, V is not necessarily proportional to $I_{n \times n}$, and V^{-1} exists. Let $T = X(X^T V^{-1} X)^{-1} X^T V^{-1} Y$
   a. Find $E(T)$. This part is worth 10 points.
   b. Find $vcv(T)$. This part is worth 50 points.

End of Examination

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

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

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

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

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

$= M V M^T$

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

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

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