SOLUTION

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAR

AMS 315, Examination 2

November 15, 2018

Name: **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 6:00 pm today. In this event, you are still on your honor not to give or receive assistance.

The last problem is number 6. The total value of the examination is 240 points. Your graded examination will be returned on Tuesday.

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

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Common Information for Questions 1, 2, and 3

A research team sought to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y was the systolic blood pressure measurement at age 40 (with a higher number indicating greater—more problematic—blood pressure). The variable x was a measure of the participant's blood glucose level at age 35 (higher values indicate greater—more problematic—blood glucose level); and the variable w was a measure of the participant's body mass index (BMI) at age 30 (higher values indicate greater—more problematic--obesity). They observed values of y, x, and w on $n = 842$ participants. The mean and **standard deviation** of Y (using $n-1$ as divisor) were 125.4 mmHg and 20.8 mmHg respectively. The mean and **standard deviation** of w were 24.7 and 10.1 respectively. The mean and **standard deviation** of x were 91.6 mg/dl and 15.7 mg/dl respectively. The correlation between Y and w was 0.41, the correlation between Y and x was 0.23; 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.

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

$$= \frac{0.23 - (0.41)(0.55)}{\sqrt{(1 - .41^2)(1 - .55^2)}} = \frac{.0045}{\sqrt{(.8319)(.6975)}}$$

$$= \frac{.0045}{\sqrt{.58025}} = \frac{.0045}{.76174} = .00591$$

$$r_{Yw \cdot x} = \frac{.41 - .23(.55)}{\sqrt{(1 - .23^2)(1 - .55^2)}} = \frac{.2835}{\sqrt{(.9471)(.6975)}}$$

$$= \frac{.2835}{\sqrt{.66060}} = \frac{.2835}{.81277} = 0.3488$$

+10 FOR EACH CORRECT PARTIAL

−10 FOR REVERSING PARTIALS.

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 the sum of squares due to the regression on x after including w, the error sum of squares, and the total sum of squares. Test the null hypothesis that $\beta_1 = 0$ against the alternative $\beta_1 \neq 0$. What is the correct test? What is your conclusion? This question is worth 50 points.

$$SSTOTAL = 841(20.8)^2 = 363850.24.$$

$$SS(w) = (.41)^2\ SSTOTAL = 61163.225$$

$$SSTOTAL - SS(w) = 302,687.0147$$

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

$$= (.00591)^2\ (302687.0147)$$

$$= 10.572.$$

$$SSERR = SSTOTAL - SS(w) - SS(x|w)$$

$$= 302676.44 \quad \text{ON } 839\text{DF}$$

SOURCE	DF	SS	MS
REG(w)	1	61163.225	61163.225
REG(x|w)	1	10.572	10.572
ERROR	839	302676.44	$360.759 = (18.99)^2$
TOTAL	841	363850.24	

$$F_{x|w} = \frac{SSREG(x|w)/1}{MSE} = \frac{10.572}{360.759} = 0.029$$

α	$F(1,839)$	
.10	2.712	ACCEPT $H_0: \beta_1 = 0$. AT $\alpha = .10$
.05	3.853	ACCEPT $H_0: \beta_1 = 0$ AT $\alpha = .05$
.01	6.665	ACCEPT $H_0: \beta_1 = 0$ AT $\alpha = .01$.

−20 WRONG SEQUENCE

−15 EACH INCORRECT ENTRY IN ANOVA TABLE

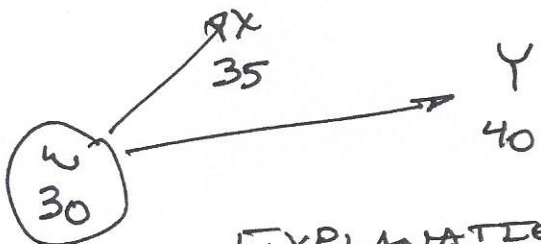
−40 INCONSISTENT OR NO

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. 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



EXPLANATION.

CORRECT OR NOT. NO PARTIAL CREDIT.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

Common Information for Questions 1, 2, and 3

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 686 participants. The variable Y is the participant's blood glucose level at age 40 (where a larger number indicates greater—more problematic--level); the variable x is the measure of the participant's educational attainment at age 30 (where a greater number indicates greater years of education completed); and the variable w is the participant's body mass index at age 35 (a larger number indicates more obesity). They observed values of y, x, and w for the participants. The mean and **standard deviation** of Y (using $n-1$ as divisor) were 129.4 mg/dl and 16.1 mg/dl respectively. The mean and **standard deviation** of x were 14.8 years and 2.9 years respectively. The mean and **standard deviation** of w were 26.8 and 8.9 respectively. The correlation between Y and w was 0.62; the correlation between Y and x was **-0.30**; and the correlation between x and w was **-0.48**.

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

$$r_{Yx \cdot w} = \frac{-.30 - (0.62)(-.48)}{\sqrt{(1-.62^2)(1-(-.48)^2}} = \frac{-.0024.}{\sqrt{(.6156)(.7696)}}$$

$$= \frac{-.0024}{\sqrt{.47377}} = \frac{-.0024}{.6883} = -.00349.$$

$$r_{Yw.x} = \frac{0.62 - (-0.30)(-0.48)}{\sqrt{(1-(-.30)^2)(1-(-0.48)^2)}} = \frac{0.476}{\sqrt{(0.9100)(.7696)}}$$

$$= \frac{0.476}{\sqrt{.70034}} = \frac{0.476}{0.83686} = 0.5688.$$

+10 FOR EACH CORRECT PARTIAL

-10 FOR REVERSING PARTIALS.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

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. Test the null hypothesis that $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. What is the correct test? What is your decision? Use levels of significance 0.10, 0.05, and 0.01. This question is worth 50 points.

$TSS = (n-1)(s_y)^2 = 685(16.1)^2 = 177{,}558.85$

$REG\,x = (r_{yx})^2\, TSS = (-0.30)^2\, TSS$

$= 15980.2965.$

$TSS - REG\,x = 161{,}578.555.$

$REG(w|x) = (r_{yw\cdot x})^2\,(TSS - REG\,x)$

$= (.5688)^2\,(161578.555)$

$= 52276.065$

$SSE = TSS - REG\,x - REG(w|x)$

$= 109304.4483$

SOURCE	DF	SS	MS
REG(x)	1	15980.2965	15980.2965
REG(w|x)	1	52276.065	52276.065
ERROR	683	109304.4483	$160.033 = (12.65)^2$
TOTAL	685	177,558.85	

$$F_{w|x} = \frac{52276.065}{160.033} = 326.658$$

α	$F(1,683)$	
.10	2.713	R
.05	3.855	R
.01	6.672	R

REJECT $H_0: \beta_2 = 0$ AT $\alpha = .01$.

−15 EACH INCORRECT ENTRY IN ANOVA TABLE

−40 INCONSISTENT OR NO CONCLUSION

−20 REVERSE SEQUENCE

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

3. 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

$X \longrightarrow W \longrightarrow Y$

30 35 40

MEDIATION

NO PARTIAL CREDIT. WRONG ANSWER -20

−15 EACH INCORRECT ENTRY IN ANOVA TABLE
−40 INCONSISTENT OR NO CONCLUSION
−10 NO OPTIMAL SETTING

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

4. 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 7 animals to dosage 0, 7 to dosage 1, 7 to dosage 2, and 7 to dosage 3. They observed that the average values of Y at each dosage were $y_{0\bullet} = 481.6, y_{1\bullet} = 388.7,\ y_{2\bullet} = 412.4,$ 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 $s_0^2 = 8250.9, s_1^2 = 9878.3,\ s_2^2 = 10630.4,$ 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.
   a. 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.
   b. 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? What is your conclusion?
   c. What is your recommendation for the optimal setting. The three parts of this question are worth 60 points.

DOSE	$y_{i\cdot}$	$y_{i\cdot}-y_{\cdot\cdot}$	$(y_{i\cdot}-y_{\cdot\cdot})^2$	s_i^2	J_i
0	481.6	62.6	3918.76	8250.9	7
1	388.7	−30.3	918.09	9878.3	7
2	412.4	−6.6	43.56	10630.4	7
3	393.3	−25.7	660.49	7164.5	7
$y_{\cdot\cdot}$	$\frac{1676}{4} = 419$	0	5540.9	$\frac{35924.1}{4}$	28

$MSE = \frac{35924.1}{4} = 8981.025$ ON 24DF $SSE = 24(MSE) = 215544.6$

$SSTREAT = J\sum(y_{i\cdot}-y_{\cdot\cdot})^2 = 7(5540.9) = 38786.3$ ON 3DF

SOURCE	DF	SS	MS	F
TREAT	3	38786.3	12928.767	1.4395
ERROR	24	215544.6	8981.025	
TOTAL	27			

α	$F(3,24)$	
.10	2.327	A
.05	3.009	A
.01	4.718	A

ACCEPT H_0 AT $\alpha = .10$
$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$.

C. THERE IS NO OPTIMAL DOSE. AT $\alpha = .10$

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

4. 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 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 $y_{0\bullet} = 71.6, y_{1\bullet} = 106.4, y_{2\bullet} = 98.7$, and $y_{3\bullet} = 83.3$ where $y_{i.}$ was the average of the observations taken with dosage $i = 0,\ldots,3$, respectively. They also observed that $s_0^2 = 1645.9, s_1^2 = 1278.3, s_2^2 = 1830.4$, and $s_3^2 = 1464.5$ where s_i^2 was the unbiased estimate of the variance for the observations taken with dosage $i = 0,\ldots,3$, respectively.
   a. 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.
   b. 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? What is your conclusion?
   c. What is your recommendation for the optimal setting. The three parts of this question are worth 60 points.

DOSE	$y_{i\cdot}$	$y_{i\cdot} - y_{\cdot\cdot}$	J_i	s_i^2	$(y_{i\cdot} - y_{\cdot\cdot})^2$
0	71.6	-18.4	20	1645.9	338.56
1	106.4	16.4	20	1278.3	268.96
2	98.7	8.7	20	1830.4	75.69
3	83.3	-6.7	20	1464.5	44.89
	$\Sigma =$ 360	0		6219.1	728.1

$y_{\cdot\cdot} = 90$

MSE = 1554.775 ON 76 DF
SSE = 118162.9
SSTR = 20(728.1) = 14562

SOURCE	DF	SS	MS
DOSE	3	14562.0	4854
(PURE) ERROR	76	118162.9	1554.775
TOTAL	89	132724.9	

$$F = \frac{4854}{1554.775} = 3.122$$

α		
.10	2.157	R
.05	2.725	R
.01	4.050	A

ACCEPT $H_0: \mu_0 = \mu_1 = \mu_2 = \mu_3$ AT $\alpha = .01$ REJECT AT $\alpha = .05, .10$

D) DOSAGE 1 HAS HIGHEST AVERAGE

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

95% PROTECTED T CI FOR $\mu_1 - \mu_2$ IS

$$106.4 - 98.7 \pm 1.992 \sqrt{1554.775} \sqrt{\tfrac{1}{20} + \tfrac{1}{20}}$$

$$7.7 \pm 1.992 \sqrt{155.475}$$

$$7.7 \pm 1.992 (12.47)$$

$$7.2 \pm 24.84.$$

THIS INCLUDES (μ_2 AND μ_1 APPEAR EQUAL);

BUT μ_1 AND μ_0 APPEAR DIFFERENT:

95% PROTECTED t FOR $\mu_1 - \mu_0$:

$$106 - \cancel{98.7}\ 71.6 \pm 24.84.$$

$$34.8 \pm 24.84$$ EXCLUDES 0.

95% PROTECTED t FOR $\mu_1 - \mu_3$

$$106.4 - 83.3 \pm 24.84$$

$$23.1 \pm \cancel{23.1}\ 24.84.$$

INCLUDES 0 (BARELY).

-15 EACH INCORRECT ENTRY IN ANOVA TABLE.

-40 INCONSISTENT OR NO CONCLUSION.

-10 NO OPTIMAL SETTING

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5. The random variable Y has $E(Y)=\theta$ and $\text{var}(Y)=\theta^2, \theta>0$. Let $W=Y^{1.5}$. Find the approximate expected value and variance of W. This question is worth 40 points.

$$f(y) = y^{1.5}$$

$$f'(y) = 1.5\, y^{0.5}$$

$$f'(\theta) = 1.5\,\theta^{0.5}.$$

$$E(W) \cong \theta^{1.5}.$$

$$\text{VAR}(W) \cong [f'(\theta)]^2\, \text{VAR}(Y)$$

$$\cong (1.5\,\theta^{0.5})^2 \cdot \theta^2$$

$$\cong 2.25\,\theta^3.$$

+10 FOR CORRECT APPX E(W).

+30 FOR CORRECT APPX VAR(W)

+10 PARTIAL CREDIT CORRECT f'(E(Y))

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

5. The correlation matrix of the random variables Y_1, Y_2, Y_3, Y_4 is $\begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & \tau & 1 \end{pmatrix}$, $0 < \rho, \tau < 1$, and each random variable has variance σ^2. Let $W_1 = -Y_1 - Y_2 + Y_3 + Y_4$, and let $W_2 = -Y_1 + Y_2 + Y_3 - Y_4$. Find the variance covariance matrix of (W_1, W_2). This question is worth 40 points.

$$M = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

$$M \, \mathrm{vcv}(Y) = \sigma^2 \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & \tau & 1 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} -1-\rho & -1-\rho & 1+\tau & 1+\tau \\ -1+\rho & -\rho+1 & 1-\tau & \tau-1 \end{bmatrix} M^T$$

$$= \sigma^2 \begin{bmatrix} -1-\rho & -1-\rho & 1+\tau & (1+\tau) \\ -1+\rho & 1-\rho & 1-\tau & -1+\tau \end{bmatrix} \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1+\rho+1+\rho+1+\tau+1+\tau & 1+\rho-1-\rho+1+\tau-(1+\tau) \\ -1+\rho-1+\rho+1-\tau-1+\tau & 1-\rho+1-\rho+1-\tau+1-\tau \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 4+2\rho+2\tau & 0 \\ 0 & 4-2\rho-2\tau \end{bmatrix}$$

+10 CORRECT M

−20 EACH INCORRECT VAR

−20 EACH INCORRECT COV

−10 NO σ^2

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

6. A research team will run a one-way analysis of variance with I settings of the treatment variable. They will collect J observations at the first setting, J observations at the second setting, …, and J observations at the Ith setting. That is, they will observe $I \times J$ outcome values $Y_{ij}, i = 1, \ldots, I; j = 1, \ldots, J$. Let the ith treatment mean be $Y_{i\bullet} = \frac{\sum_{j=1}^{J} Y_{ij}}{J}, i = 1, \ldots, I;$ and let the grand mean be $Y_{\bullet\bullet} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}}{IJ}$. What is the value of $\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet})$? Prove your answer. This question is worth 50 points.

End of the Examination

$\Sigma\Sigma\,(Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet}) = 0.$

+15 FOR ONE CORRECT MANIPULATION OF ABOVE

-15 FOR EACH SUBSTANTIVE ERROR

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet}) = \sum_{i=1}^{I}\left[(Y_{i\bullet} - Y_{\bullet\bullet})\Sigma(Y_{ij} - Y_{i\bullet})\right]$$

$$= \sum_{i=1}^{I}(Y_{i\bullet} - Y_{\bullet\bullet})0 = 0. \quad \text{SINCE } \sum_{j=1}^{J}(Y_{ij} - Y_{i\bullet}) = 0$$

BRUTE FORCE EXPANSION WORKS AS WELL

$$\Sigma\Sigma(Y_{ij}Y_{i\bullet} - Y_{i\bullet}Y_{i\bullet} - Y_{ij}Y_{\bullet\bullet} + Y_{i\bullet}Y_{\bullet\bullet})$$

$$= \sum_i \left(Y_{i\bullet}\sum_j Y_{ij}\right) - \Sigma\Sigma(Y_{i\bullet})^2 - IJ(Y_{\bullet\bullet})^2 + J\,Y_{\bullet\bullet}\,\Sigma Y_{i\bullet}$$

$$= \sum_i Y_{i\bullet}(J Y_{i\bullet}) - J\,\Sigma(Y_{i\bullet})^2 - IJ(Y_{\bullet\bullet})^2 + IJ(Y_{\bullet\bullet})^2 = 0$$

IT IS TRUE $\sum_{i=1}^{I}(Y_{i\bullet} - Y_{\bullet\bullet}) = 0$. THE ATTEMPT TO USE THIS BY REVERSING ORDER OF SUMMATION IS COMPLEX

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet}) = \sum_{j=1}^{J}\sum_{i=1}^{I}((Y_{ij}) - Y_{i\bullet})(Y_{i\bullet} - Y_{\bullet\bullet})$$

NOTE THAT $(Y_{i\bullet} - Y_{\bullet\bullet})$ HAS THE MULTIPLIER $(Y_{ij} - Y_{i\bullet})$