Final Exam Review Questions

Due No due date **Points** 29 **Questions** 29 **Time Limit** None

Instructions

This is a set of practice questions that covers material from Lectures 16-26. You are welcome to work with others to answer the questions. SAS output is provided in a **PDF** to reference and answer questions on the practice questions. The general SAS code used to create these tables or figures is included in the **SAS file**.

On the actual exam you will receive a similar booklet of relevant output to use in answering the exam questions and the SAS code will be directly included with the relevant output (unlike here where it is in two separate files). Additionally, the exam will likely include interpretation and summary questions (i.e., the "complete" summary involves a point estimate, an interval estimate, a decision, and a measure of uncertainty in the decision; additionally it may be useful to include other summary statistics like the R^2 value), however those are not included on the practice exam since they are not easily scored automatically.

Completion of these questions is worth 5% extra credit on your overall final exam score, so you can earn extra credit in advance before taking the exam! For example, if you get 60% or 90% of the practice questions correct, you will still earn the 5% extra credit. After you complete the practice questions, you can review your questions and see the correct answers.

If you notice any typos or have questions about an answer please let Alex know via Canvas message, email, or a post on the discussion board.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	68 minutes	19.67 out of 29

Submitted Dec 9 at 11:46am

Question 1

0 / 1 pts

In class we have extensively used the FEV data from Rosner. In our lecture on confounding we examined if age was a confounder in the relationship between FEV and current smoking status.

Now we are restricting the FEV data set to non-smokers age 12 or older to examine if age is a confounder in the relationship between FEV and sex (coded as MALE for boys=1). Refer to the PDF of output to answer the following question:

Based on the first classical criterion for confounding (association between MALE and AGE), does AGE appear to be a confounder?

orrect Answer

No, there is no association as based on the t-test of difference in mean age between boys and girls (p=0.4568).

ou Answered

No, it doesn't make sense for age to somehow be associated with sex.

Yes, in the regression of FEV~MALE+AGE we have p<0.0001 for age.

While it might not make sense intuitively for age to be a confounder, the best response identifies a p-value to back up the claim. Using the regression output is inappropriate for our 1st criterion, but it is useful for our 2nd classical criterion.

Question 2

0 / 1 pts

In class we have extensively used the FEV data from Rosner. In our lecture on confounding we examined if age was a confounder in the relationship between FEV and current smoking status.

Now we are restricting the FEV data set to non-smokers age 12 or older to examine if age is a confounder in the relationship between FEV and sex (coded as MALE for boys=1). Refer to the PDF of output to answer the following question:

Based on the second classical criterion for confounding (association between FEV and AGE, holding MALE constant), does AGE appear to be a confounder?

orrect Answer

Yes, in the regression of FEV~MALE+AGE we have p<0.0001 for age implying an association when adjusting for MALE.

ou Answered

Yes, the ANOVA table for the FEV~MALE+AGE model has p<0.0001, indicating a significant association.

No, the mean for males is 10.01 L versus 9.84 L for females, indicating an association.

No, there is no association as based on the t-test of difference in mean age between boys and girls (p=0.4568).

For our 2nd classical criterion the output we are interested in examining is the regression parameter table for FEV~MALE+AGE. The beta coefficient for age is 0.13772 and has a significant p-value when holding MALE constant.

Using the ANOVA table is inappropriate because it tests the *overall* model, so if either age or male is significant. Comparing the means can be useful, however it is better to provide some estimate of the variability and statistical significance. Additionally the means represent a stratified analysis, whereas the regression model incorporates MALE as a predictor.

Question 3

0 / 1 pts

In class we have extensively used the FEV data from Rosner. In our lecture on confounding we examined if age was a confounder in the relationship between FEV and current smoking status.

Now we are restricting the FEV data set to non-smokers age 12 or older to examine if age is a confounder in the relationship between FEV and sex (coded as MALE for boys=1). Refer to the PDF of output to answer the following question:

Based on the third classical criterion for confounding (AGE on the causal pathway between MALE and FEV), does AGE appear to be a confounder?

ou Answered

- No. AGE is not on the causal pathway--sex does not "cause" age.
- Yes. AGE is on the causal pathway--sex "causes" age.
- No. AGE is on the causal pathway--sex "causes" age.

orrect Answer

- Yes. AGE is not on the causal pathway--sex does not "cause" age.

The third classical criterion can be tricky to describe without enough background knowledge. However, in this case, it is pretty apparent that being MALE does not "cause" AGE (at least not in the human population this study is based on). Since AGE is not on the causal pathway, we could consider it to potentially be a confounder (depending on what we conclude for classical criteria 1 and 2).

Question 4

1 / 1 pts

In class we have extensively used the FEV data from Rosner. In our lecture on confounding we examined if age was a confounder in the relationship between FEV and current smoking status.

Now we are restricting the FEV data set to non-smokers age 12 or older to examine if age is a confounder in the relationship between FEV and sex (coded as MALE for boys=1). Refer to the PDF of output to answer the following question:

Based on the operational criterion for confounding, does AGE appear to be a confounder?

Correct!

- No. The difference between our crude and adjusted estimates is about 4.3% (or 4.5%), which is less than 10%.

- Yes. The difference between our crude and adjusted beta estimates is only 0.04028 which is less than 0.10.

- No. The difference between our crude and adjusted beta estimates is only 0.04028 which is less than 0.10.

- Yes. The difference between our crude and adjusted estimates is about 4.3% (or 4.5%), which is less than 10%.

Depending on our discipline, we may calculate the percent difference in slightly different ways:

$\frac{\beta_{crude}-\beta_{adj}}{\beta_{crude}}$ or $\frac{\beta_{crude}-\beta_{adj}}{\beta_{adj}}$

We have generally similar responses:

$$\frac{\beta_{crude}-\beta_{adj}}{\beta_{crude}} = \frac{0.92869-0.88841}{0.92869} = 0.04337292 \rightarrow 4.3\%$$

$$\frac{\beta_{crude}-\beta_{adj}}{\beta_{adj}} = \frac{0.92869-0.88841}{0.88841} = 0.04533943 \rightarrow 4.5\%$$

Note, the responses looking at the absolute difference happen to be similar in this example, but the meaningful absolute difference can be different in each context, whereas a 10% or 20% threshold is generally accepted for the operational definition of confounding.

Question 5

1 / 1 pts

In class we have extensively used the FEV data from Rosner. In our lecture on confounding we examined if age was a confounder in the relationship between FEV and current smoking status.

Now we are restricting the FEV data set to non-smokers age 12 or older to examine if age is a confounder in the relationship between FEV and sex (coded as MALE for boys=1). Refer to the PDF of output to answer the following question:

Based on our exploration of classical and operational criterion for confounding, is AGE a confounder for the relationship between FEV and MALE?

- Yes. All criterion suggested it was a confounder.
- Yes. Since the 2nd classical criterion was met it is a confounder.

Correct!

No. Only the 2nd classical criterion was met.

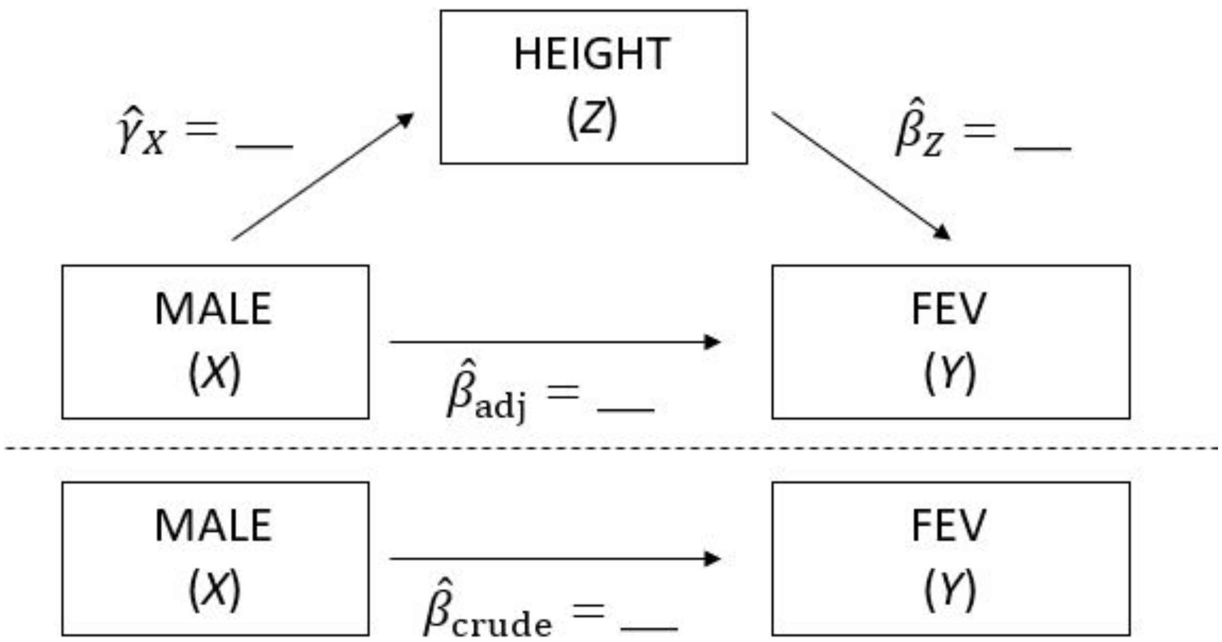
Since the 2nd classical criterion was the only consideration we actually met, it is unlikely we can claim AGE to be a confounder. However these considerations can be considered somewhat subjective and in situations with disagreeing information (especially if the classical criteria suggest something different from the operational criterion) and either conclusion could be appropriate depending on the arguments you put forth.

Question 6

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Based on the mediation framework, what is the estimate of $\hat{\beta}_{crude}$?



Correct!

0.9287

orrect Answers

0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.9287 (with margin: 0.01)

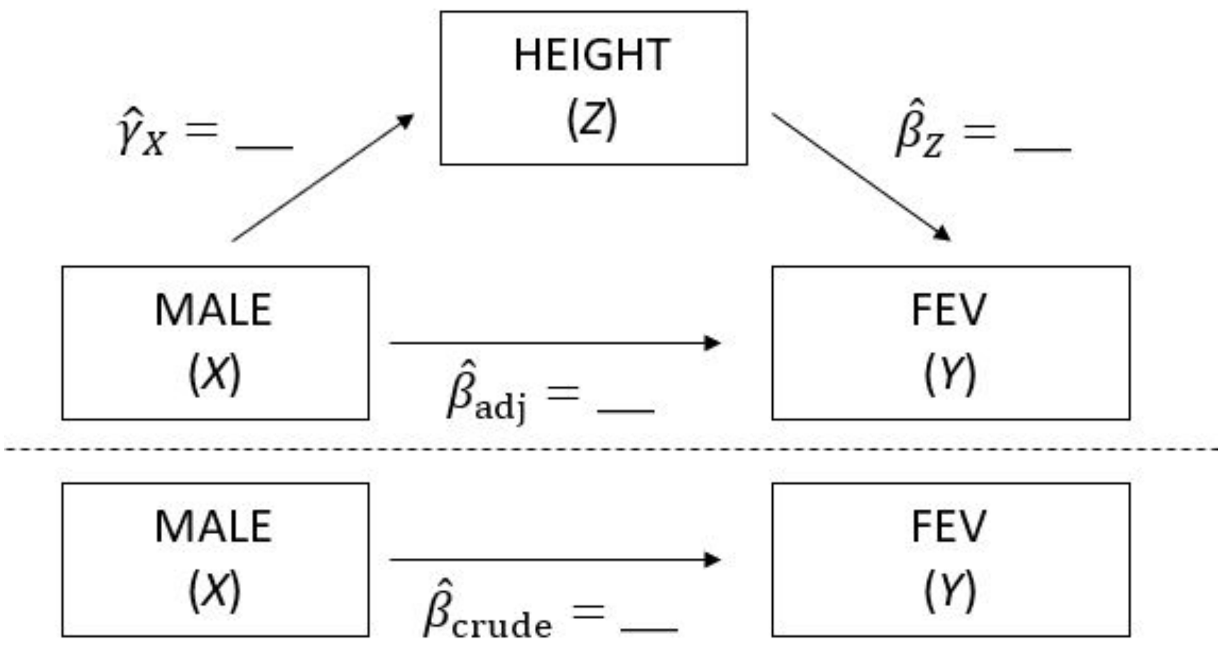
From the output for FEV on MALE, we know $\hat{\beta}_{crude}$ is 0.92869.

Question 7

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Based on the mediation framework, what is the estimate of $\hat{\beta}_{adj}$?



Correct!

0.2894

orrect Answers

0.2894 (with margin: 0.01)

0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)

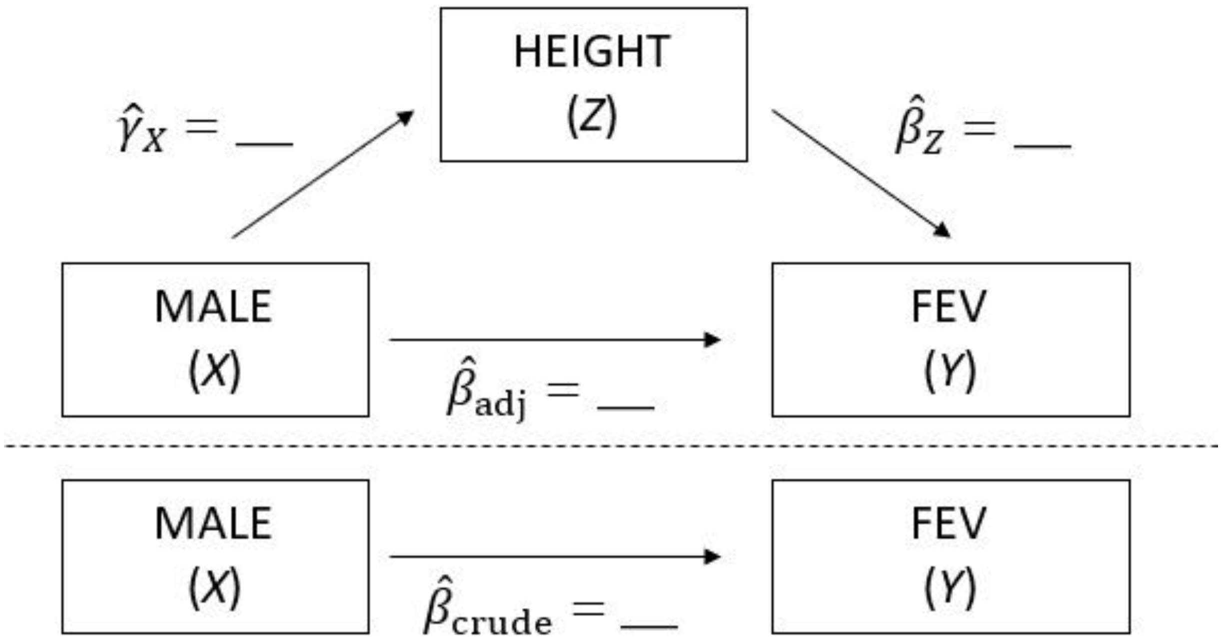
From our regression output for FEV on MALE and HEIGHT we know that the adjusted estimate for $\hat{\beta}_{adj}$ is 0.28940.

Question 8

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Based on the mediation framework, what is the estimate of $\hat{\gamma}_X$?



Correct!

4.7548

orrect Answers

0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
4.7548 (with margin: 0.05)

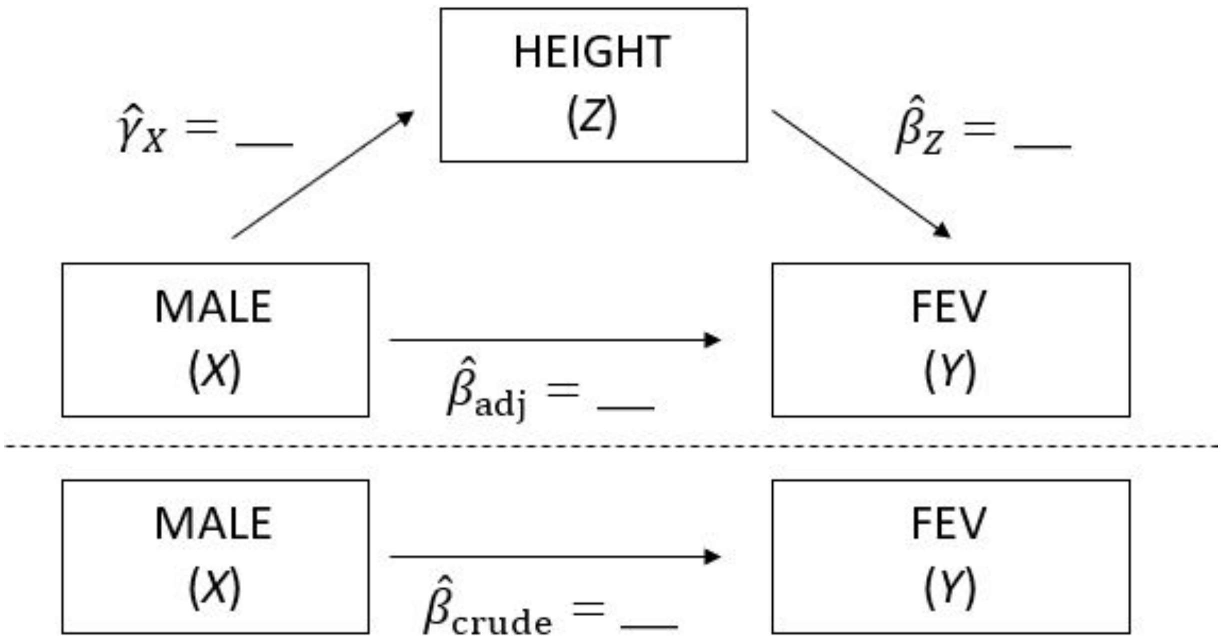
Based on our regression output for HEIGHT on MALE we know that $\hat{\gamma}_X$ is 4.75478.

Question 9

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Based on the mediation framework, what is the estimate of $\hat{\beta}_Z$?



Correct!

0.1344

orrect Answers

0.0 (with margin: 0.0)
0.1344 (with margin: 0.01)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)

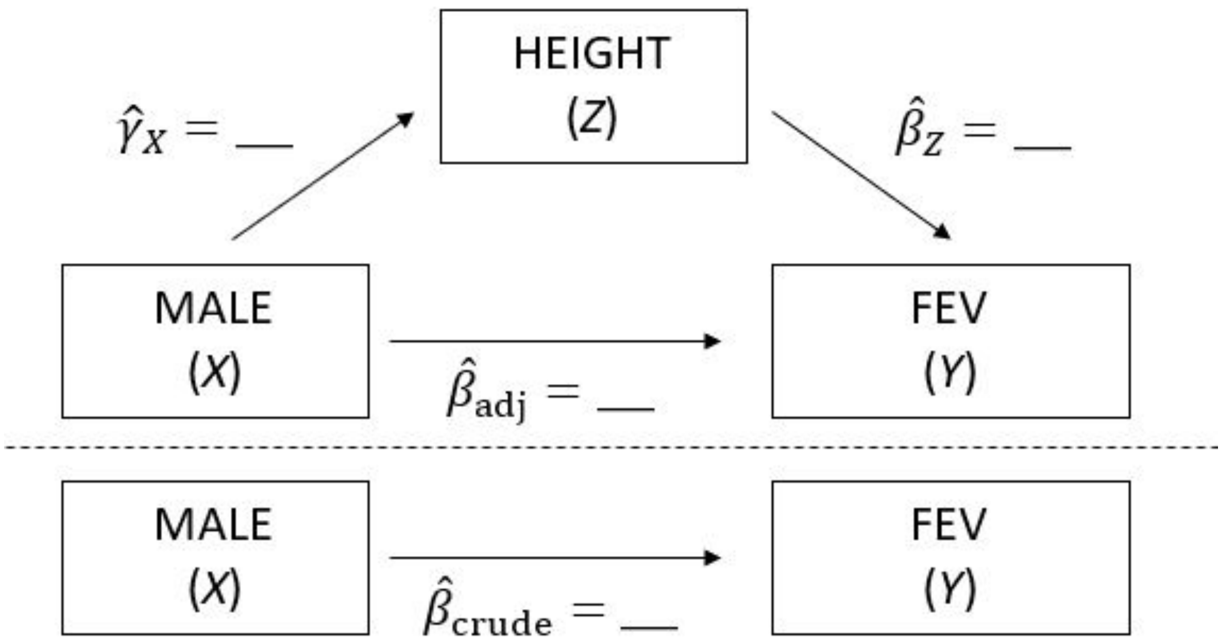
Based on our regression output for FEV on MALE and HEIGHT, we know that $\hat{\beta}_Z$ is 0.13445.

Question 10

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. What is the proportion mediated by HEIGHT (round to 3 decimal places for proportion or 1 decimal place for percentage, i.e., 0.853 or 85.3%)?



Correct!

68.8

Correct Answers

0.0 (with margin: 0.0)
68.8 (with margin: 1.0)
0.688 (with margin: 0.01)
0.0 (with margin: 0.0)

We can calculate the proportion mediated (or percent of total effect mediated) by:

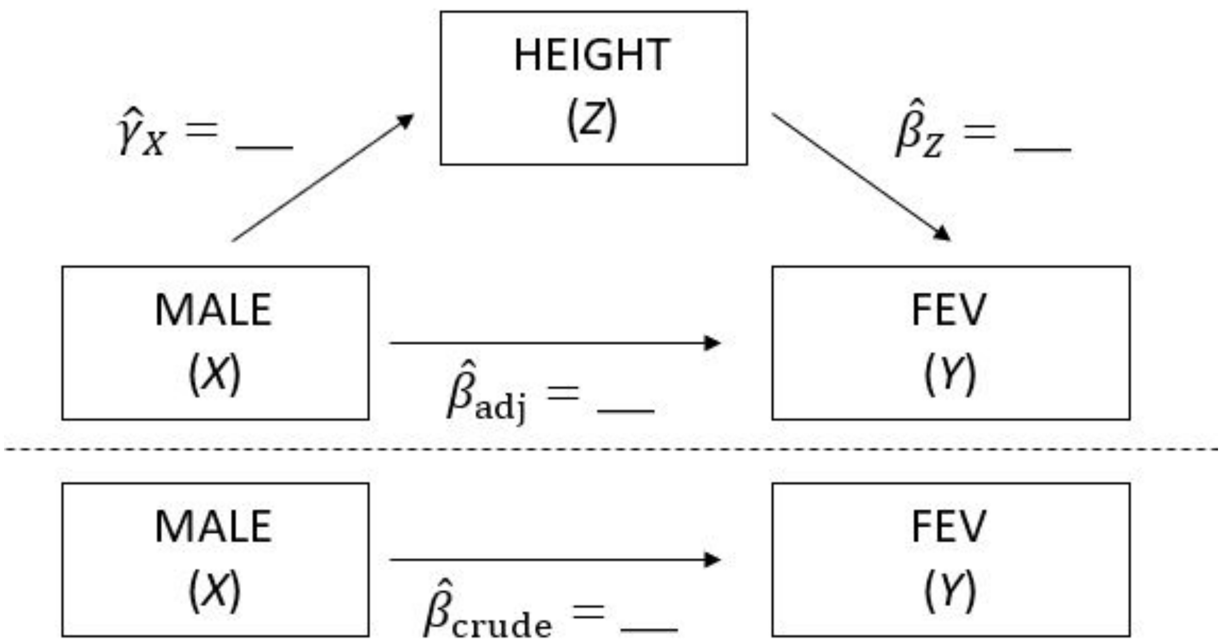
$$\frac{\hat{\beta}_{crude} - \hat{\beta}_{adj}}{\hat{\beta}_{crude}} = \frac{\hat{\gamma}_X \times \hat{\beta}_Z}{\hat{\beta}_{crude}} = \frac{0.63929}{0.92869} = 0.6883$$

Question 11

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. What is the standard error of our indirect effect (round to 3 decimal places)?



Correct!

0.105

orrect Answers

0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.105 (with margin: 0.005)
0.0 (with margin: 0.0)

We can calculate the standard error of our indirect effect as

$$SE(\hat{\gamma}_X \times \hat{\beta}_Z) = \sqrt{\hat{\gamma}_X^2 \times (SE(\hat{\beta}_Z))^2 + \hat{\beta}_Z^2 \times (SE(\hat{\gamma}_X))^2}$$

$$= \sqrt{(4.75478)^2(0.01621)^2 + (0.13445)^2(0.52733)^2}$$

$$= \sqrt{0.01096729} = 0.1047249$$

Question 12

0 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Calculate the 95% confidence interval for our proportion/percent mediated (note: use 1.96). What is the lower limit of our CI (round to 3 decimal places for proportion, 1 for percent).

You Answered

68.6

Correct Answers

0.467 (with margin: 0.02)
46.7 (with margin: 2.0)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)

Our 95% CI is

$$(\hat{\beta}_{crude} - \hat{\beta}_{adj}) \pm 1.96 \times SE(\hat{\beta}_{crude} - \hat{\beta}_{adj}) = 0.63929 \pm 1.96 \times 0.105 = (0.433, 0.845)$$

Note, this is the 95% CI for the indirect effect. We need to translate it to proportion/percent mediated:

$$\frac{0.433}{0.92869} = 0.4667758$$

$$\frac{0.845}{0.92869} = 0.9099807$$

Question 13

0 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. Calculate the 95% confidence interval for our proportion/percent mediated (note: use 1.96). What is the upper limit of our CI (round to 3 decimal places for proportion, 1 for percent).

You Answered

69

Correct Answers

0.0 (with margin: 0.0)

0.91 (with margin: 0.02)

0.0 (with margin: 0.0)

91.0 (with margin: 2.0)

Our 95% CI is

$$(\hat{\beta}_{crude} - \hat{\beta}_{adj}) \pm 1.96 \times SE(\hat{\beta}_{crude} - \hat{\beta}_{adj}) = 0.63929 \pm 1.96 \times 0.105 = (0.433, 0.845)$$

Note, this is the 95% CI for the indirect effect. We need to translate it to proportion/percent mediated:

$$\frac{0.433}{0.92869} = 0.4667758$$

$$\frac{0.845}{0.92869} = 0.9099807$$

Question 14

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Based on the calculations we have done so far (i.e., proportion mediated, 95% CI, Z statistic), is HEIGHT a significant mediator?

- Yes, height explains 68.8% of the effect of MALE, which is pretty significant.

Correct!

- Yes, the CI does not include 0 and Z=6.088 is large enough to likely be significant.
- No, the CI does not include 0, so height is not a mediator.
- No, the CI includes 0 and height is therefore not a mediator.

The best answer is to reference the CI not including 0 and/or the large test statistic makes it unlikely we have an insignificant mediator in HEIGHT.

Note, even a large seeming proportion mediated can be insignificant if the variability is large enough to make a wide CI or small test statistic.

Question 15

1 / 1 pts

In class we have extensively used the FEV data from Rosner. Still considering the relationship of FEV and MALE, let's work through the steps of determining if HEIGHT is a mediator.

Refer to the PDF of output to answer the following question. What is the test statistic (Z) for our indirect effect based on the SE you calculated previously (round to 3 decimal places).

Correct!

6.103

orrect Answers

6.088 (with margin: 0.1)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)
0.0 (with margin: 0.0)

$$Z = \frac{\hat{\gamma}_X \times \hat{\beta}_Z}{SE(\hat{\gamma}_X \times \hat{\beta}_Z)} = \frac{0.6392802}{0.105} = 6.088$$

Question 16

1 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the simple linear regression of YIELD and TEMP. Consider the diagnostic figures on page 7 of the output PDF. Does the assumption of **linearity** appear to be violated?

- No, the linearity assumption is valid based on these plots.

- Yes, the normal probability plot does not have a *perfect* fit to the diagonal line.

Correct!

- Yes, the scatterplot shows the regression fit does not fit the model well and the residual plot has a clear pattern that does not suggest a random scatter.

For our simple linear regression it appears our linearity assumption is violated because the scatterplot points do not appear to be well described by the straight-line fit and the residual plot has a non-random scatter.

Question 17

0 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the simple linear regression of YIELD and TEMP. Consider the diagnostic figures on page 7 of the output PDF. Does the assumption of **homoscedasticity** appear to be violated?

Yes, since linearity is violated the assumption of equal variances also must be violated.

No, the normal curve on the histogram suggests equal variances are met.

You Answered

Yes, the spread of the points may be similar but are over very distinct ranges of yield.

Correct Answer

No, the equal variance assumption appears to be generally true given the similar spread of the residual plot points.

The easiest way to roughly judge equal variances is to look at the residual plot since they work as a summary in both multiple and simple linear regression. For simple linear regression you may also consider the scatterplot since it contains both the outcome and only predictor and will be extremely similar to the residual plot.

The histogram and normal probability plot do not necessarily provide the appropriate information since we need to determine equal variances of Y at each value of X ($\hat{\sigma}^2_{Y|X}$).

Question 18

0 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80.

After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the simple linear regression of YIELD and TEMP. Consider the diagnostic figures on page 7 of the output PDF. Does the assumption of **normality** appear to be violated?

ou Answered

- Yes, the residuals do not form a normal curve in the residual plot.

orrect Answer

- No, the normal probability plot is generally following the diagonal line.
- Yes, I have literally never seen a more non-normal histogram in my life. I mean, look at it! LOOK!
- No, the scatterplot shows approximate normality within each temperature.

The two figures that are most useful are the histogram and the normal probability plot. In this case the small sample size makes evaluating the normal-looking nature of the histogram difficult, but the points do follow the normal probability plot diagonal line pretty well.

Question 19

1 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the simple linear regression of YIELD and TEMP. Consider the output on page 6. Is TEMP a significantly associated with YIELD?

Yes, p=0.0178 and suggests that TEMP is significantly associated with YIELD.

Correct!

No, p=0.3667 and suggests that TEMP is not significantly association with YIELD.

Yes, p=0.3667 and suggests that TEMP is significantly associated with YIELD.

No, p=0.0178 and suggests that TEMP is not significantly associated with YIELD.

Looking at either the parameter estimate table or ANOVA table for our simple linear regression we see that t=0.95 and F=0.89, respectively, resulting in p=0.3667 which suggests that TEMP is not significantly associated with YIELD.

However, based on the evaluation of the diagnostic plots we should probably explore the lack of fit (LOF) and if polynomial terms may be advantageous.

Question 20

0.67 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the simple linear regression of YIELD and TEMP. Using any of the relevant "Set 2" output, would a higher order polynomial improve our model and reduce lack of fit? (There may be multiple correct answers.)

No, the output on page 8 for a model including up to a third-order term for temp (i.e., TEMP^3) has no significant p-values for the Parameter Estimates table.

orrect Answer

Yes, the orthogonal polynomial results on page 11 indicate that "quad" is significant in the Parameter Estimates table (p=0.0021), but cubic and lin are not (p>0.05). This suggests we should include the quadratic (second-order polynomial) term and any lower order terms in our model.

Correct!

Yes, the LOF tables on page 8 suggest that "LOF linear" is not adequate (p=0.0061 and we reject the null hypothesis that at all higher order terms are 0) and "LOF quad" has p=0.4468 suggesting that a quadratic term is an adequate model.

No, the simple linear regression output on page 6 indicates that the model with only a linear term is not significant and it is useless to pursue more complex models.

Correct!

Yes, the results from the model of YIELD on TEMP and TEMP^2 on page 9 have a significant F value (p=0.0031), indicating that at least one included variable is a significant improvement. We could also formally write out a partial F test to formally evaluate this improvement.

In this case, given our output, there are many, many ways we could come to the same conclusion that a better model is one that includes the quadratic term. These include the LOF tests on page 8, the orthogonal polynomial parameter estimates table on page 11, or comparing the results of the model with TEMP and TEMP^2 on page 9 to the simple linear regression on page 6 (admittedly this last option should probably also include a formal comparison of a partial F test).

Question 21

1 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

Consider the linear regression of our outcome of YIELD with TEMP and TEMP^2 as covariates. Is the linearity assumption still violated?

- [ ] Yes, the scatterplot regression fit line does go through all points.

Correct!

- [x] No, the residuals are more randomly scattered (or as random as 12 points can be) on the residual plot.

Question 22

0 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

The researchers are interested in taking the natural log of the outcome YIELD to determine if that fixes any violations of the assumptions from the non-transformed model. They examine the simple linear regression of log(YIELD) as the outcome and TEMP as the predictor. Are any of the assumptions of linearity, homoscedasticity, or normality violated?

orrect Answer

- [ ]

Yes, linearity is still violated (see the scatterplot). In this case log transforming the outcome did not improve a fit to the linear model. Equal variances and normality appear to be generally acceptable (see residual plot and normal probability plot, respectively).

You Answered

- [x] Yes, all assumptions are violated (all plots are bad).

- [ ] No, no assumptions are violated.

Similar to our non-transformed outcome, the linearity assumption is violated. Based on the residual plot on page 13 it appears a quadratic term may also be appropriate.

Question 23

1 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

The researchers are interested in taking the natural log of the outcome YIELD to determine if that fixes any violations of the assumptions from the non-transformed model. Using any of the "Set 3" output, does a higher order term improve the lack of fit observed in the straight-line model? (There may be multiple correct answers to select.)

- [ ] No, based on the ANOVA table for our simple linear regression on page 12 our F value is 0.79 with p=0.3941, suggesting TEMP does not estimate YIELD in any meaningful way.

Correct!

- [x]

Yes, based on the orthogonal polynomial variable Parameter Estimates table we see quad is significant (p=0.0021), whereas lin and cubic are not (p>0.05).

Correct!

- [x] Yes, adding a quadratic term is significant based on the LOF quad table on either page 14 (simple polynomial) or page 15 (orthogonal polynomial).

- [ ] No, based on the orthogonal polynomial variable Parameter Estimates table we see lin is not significant (p=0.1681) and we should therefore not consider higher order terms.

Similar to our non-transformed model, a quadratic term improves our model fit significantly over a straight-line model. This can be identified from the LOF tests coded with the simple polynomial (page 14) or orthogonal polynomial (page 15) models, or with the parameter estimates table from the orthogonal polynomial model.

Question 24

0 / 1 pts

Researchers wanted to explore how temperature impacts yield of a certain crop. They designed a small study that controlled for moisture, daylight, and other confounders and included 4 temperature levels: 50, 60, 70, and 80. After 12 weeks they measured the yield to determine what, if any, relationship may be present.

The researchers are interested in taking the natural log of the outcome YIELD to determine if that fixes any violations of the assumptions from the non-transformed model. They examine the simple linear regression of log(YIELD) as the outcome and TEMP as the predictor. What is the interpretation of the $\hat{\beta}_1 = 0.00366$ coefficient on the original YIELD (non-logged) measure?

orrect Answer

On average, YIELD is 1.003667 times higher (0.37% higher) for a one unit increase in TEMP.

ou Answered

On average, log(YIELD) increases by 0.00366 for a one unit increase in TEMP.

On average, YIELD is 0.00366 times higher for a one unit increase in TEMP.

On average, YIELD increases by 0.00366 for a one unit increase in TEMP.

If we desire to interpret our output on the original, non-transformed scale we must exponentiate the beta coefficient. It is then, on the original scale, representative of the *geometric* mean (instead of the arithmetic mean we have been generally working with). This means we interpret the outcome in terms of "X times lower/higher" or "X% lower/higher".

$$\exp(0.00366) = 1.003667$$

We can note that technically the response "On average, log(YIELD) increases by 0.00366 for a one unit increase in TEMP." is correct, however it is on the log(YIELD) scale and not the original scale.

Question 25

1 / 1 pts

A study asked participants of their alcohol consumption habits and classified them into four groups: abstainers, social drinkers, binge drinkers, and heavy drinkers. Consider the following regression equation:

$$E(Y) = \beta_{abstain} I_{abstain} + \beta_{social} I_{social} + \beta_{binge} I_{binge} + \beta_{heavy} I_{heavy}$$

The researchers are interested in comparing the means for those who consume no alcohol to those who consume any alcohol.

Which contrast is (most) correct (assume equal group sizes and consider a contrast whose group means are in the order of the regression equation, i.e., $(\mu_{abstain}, \mu_{social}, \mu_{binge}, \mu_{heavy})$).

- (1, -1, -1, -1)

Correct!

- (3, -1, -1, -1)
- (1, -3, -3, -3)
- (3, 1, 1, 1)

While there are many ways to represent this contrast, the correct answer from those provided is (3, -1, -1, -1). Alternatively, one may consider something like (1, -1/3, -1/3, -1/3) if they wanted to calculate a contrast that compared the mean of abstainers to the mean of those who consume alcohol.

Note the incorrect options are all actually invalid contrasts, since the sum of the elements does not equal 0.

Question 26

1 / 1 pts

A study asked participants of their alcohol consumption habits and classified them into four groups: abstainers, social drinkers, binge drinkers, and heavy drinkers. Consider the following regression equation:

$$E(Y) = \beta_{abstain} I_{abstain} + \beta_{social} I_{social} + \beta_{binge} I_{binge} + \beta_{heavy} I_{heavy}$$

The researchers are interested in comparing the means for those who consume alcohol in social settings to those who consume alcohol to excess (i.e., binge or heavy).

Which contrast is (most) correct (assume equal group sizes and consider a contrast whose group means are in the order of the regression equation, i.e., $(\mu_{abstain}, \mu_{social}, \mu_{binge}, \mu_{heavy})$).

Correct!

- (0, 1, -1/2, -1/2)
- (0, -1, 0, 1)
- (-2, 0, 1, 1)
- (1, 1, -1, -1)

The most appropriate contrast presented considers the mean of social drinkers compared to the mean of binge and heavy drinkers combined.

The other contrasts are all valid contrasts, but answer different questions:

(-2,0,1,1) compares abstainers to excess drinkers.

(0,-1,0,1) compares social drinkers to heavy drinkers.

(1,1,-1,-1) compares abstain+social drinkers to binge+heavy drinkers.

Question 27

1 / 1 pts

A study asked participants of their alcohol consumption habits and classified them into four groups: abstainers, social drinkers, binge drinkers, and heavy drinkers. Consider the following regression equation:

$$E(Y) = \beta_{abstain} I_{abstain} + \beta_{social} I_{social} + \beta_{binge} I_{binge} + \beta_{heavy} I_{heavy}$$

The researchers are interested in comparing the means for binge and heavy drinkers.

Which contrast is (most) correct (assume equal group sizes and consider a contrast whose group means are in the order of the regression equation, i.e.,

$(\mu_{abstain}, \mu_{social}, \mu_{binge}, \mu_{heavy})$).

(1,-1,0,0)

(1,2,3,4)

Correct!

(0,0,1,-1)

(1,0,0,1)

This contrast is the "easiest" of the 3 considered. In this case we need to only compare the mean for binge and heavy drinkers without having to combined multiple groups, so (0,0,1,-1) is the correct choice.

Question 28

1 / 1 pts

A study asked participants of their alcohol consumption habits and classified them into four groups: abstainers, social drinkers, binge drinkers, and heavy drinkers. Consider the following regression equation:

$$E(Y) = \beta_{abstain} I_{abstain} + \beta_{social} I_{social} + \beta_{binge} I_{binge} + \beta_{heavy} I_{heavy}$$

The researchers are interested in comparing the means for those who consume no alcohol to those who consume any alcohol.

Is the following set of contrasts orthogonal?

(-3, 1, 1, 1); (0,-2,1,1); (0,0,-1,1)

Correct!

Yes

No

Yes, these contrasts are all pairwise orthogonal:

(-3)(0)+(1)(-2)+(1)(1)+(1)(1) = 0 - 2 + 1 + 1 = 0

(-3)(0)+(1)(0)+(1)(-1)+(1)(1) = 0 + 0 -1 + 1 = 0

(0)(0)+(-2)(0)+(1)(-1)+(1)(1) = 0 + 0 -1 + 1 = 0

Question 29

1 / 1 pts

A study asked participants of their alcohol consumption habits and classified them into four groups: abstainers, social drinkers, binge drinkers, and heavy drinkers. Consider the following regression equation:

$$E(Y) = \beta_{abstain} I_{abstain} + \beta_{social} I_{social} + \beta_{binge} I_{binge} + \beta_{heavy} I_{heavy}$$

The researchers are interested in comparing the means for those who consume no alcohol to those who consume any alcohol.

Is the following set of contrasts orthogonal?

(-1, 1, 0, 0); (1, 0, -1, 0); (0, 0, -1, 1)

- Yes

Correct!

- No

These three contrasts are not completely orthogonal. Consider the first and second contrasts:

(-1)(1) + (1)(0) + (0)(-1) + (0)(0) = -1 + 0 + 0 + 0 = -1 (not equal to 0, therefore not pairwise orthogonal)