Quiz 6

1. (Short answer) Taylor has a continuous outcome Y and continuous predictor of interest X. She wants to estimate "the mean difference in Y comparing groups differing by one unit in X." She has access to a third variable W, which is not associated with X in the sample and is known to causally affect Y. Charlie suggests that Taylor should include W in her regression, and he claims it will give her more precision in estimating the mean difference in Y comparing groups differing by one unit in X.

Is Charlie's claim correct? Why or why not?

1. (True/False) The regression equation for adding potential confounders is the same as the regression equation for adding potential effect modifiers.
2. (True/False) Given an outcome Y and covariates X and Z, if the linearity assumptions holds for the simple linear regression of Y on X AND for the simple linear regression of Y on Z, then it holds for the multiple linear regression of Y on X and Z.
3. (Multiple choice) Taylor is studying the association between continuous quantitative variables X (predictor of interest) and Y (outcome), and also has information on a binary variable W. Taylor has prior scientific knowledge suggesting that W causally affects Y. She plots Y against X, and colors the points according to W. She fits stratified linear regression within strata of W. These fitted models are also shown below.

Among the following, which is the most reasonable conclusion to draw from this plot?

Chart, scatter chart

Description automatically generated

1. Since W appears to be associated with X, it is likely a precision variable.
2. Since W appears to be associated with X, it is likely to be a confounder.
3. Since W does not appear to be associated with X, it is unlikely to be a confounder.
4. Since W does not appear to be associated with X, it is unlikely to be a precision variable.
5. (Multiple choice) Which of the following characteristics of an estimate is not related to the others?
6. Precision of the estimate
7. Standard error of the estimate
8. Width of the confidence interval around the estimate
9. Accuracy of the estimate
10. (Multiple choice) Which of the following best describes the major pitfall of using training data to evaluate the predictive ability of a model?
11. We risk overfitting.
12. We risk violating the classical linear regression assumptions.
13. We risk reducing the precision of our regression coefficient estimates.
14. We risk reducing the accuracy of our regression coefficient estimates.
15. (True/False) Mean-squared error can take on any value between - and .
16. (Short answer) Charlie is studying the association between sepal length and species in iris flowers. He runs a regression of sepal length (in centimeters) on the species (a categorical variable with three levels: setosa, virginica, versicolor) and sees the following output:

Text

Description automatically generated

Use this output to produce an estimate of the mean difference in sepal length (in centimeters) comparing flowers of species versicolor and flowers of species setosa.

1. (Multiple choice) Which of the following statements regarding statistical inference for a population mean is true?
2. The p-value does not depend on the alternative hypothesis.
3. The confidence interval does not depend on the null hypothesis.
4. The p-value does not depend on the null hypothesis.
5. The confidence interval depends on the null hypothesis, but not the alternative hypothesis.

(Everyone got credit for this question because it was confusingly worded)

1. (Short answer) Taylor is building a prediction model to predict the number of hours her cat Alice will sleep on a given night. In her model, she includes the following predictors:

* Number of minutes Alice played during the day
* Number of hours of daylight
* Outside temperature
* Moon phase (full, gibbous, quarter, crescent, new)

Charlie says that Taylor must also include a variable indicating whether or not Alice received a bedtime snack, since this (a) is associated with the number of minutes Alice played, and (b) may causally affect the number of hours Alice sleeps. He claims that this makes the bedtime snack variable a potential confounder, and that it must be included in Taylor's prediction model.

Is Charlie correct that Taylor should include the bedtime snack variable because it is a potential confounder? Explain your rationale.