Questions Module 5.3

Which of the following is a linear model? Select all that apply.

- [x] a. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- [x] b. $Y = \beta_0 + \beta_1 \sqrt{X_1} + \beta_2 Log\,(X_2) + \varepsilon$
- [x] c. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$
- [x] d. $Y = \beta_0 + \beta_1 X_1 + \varepsilon$
- [ ] e. none of the above

Which test statistic reported in the ANOVA table tests for the significance of the regression model as a whole?

- *a.* mean square model
- *b.* mean square error
- *c.* RSquare
- *d.* *F* ratio
- *e.* root mean square error

What does Cook's D measure?

- *a.* The influence of a point on the regression model
- *b.* Your grade on the quiz at the end of this module
- *c.* Whether a point is an outlier
- *d.* The slope of a regression line

Incorrect.
The correct answer is **a.** Cook's D influence, or Cook's distance, measures the influence of an observation on the regression model. For each observation, it measures how much the model changes if the model is refit without that observation.

You fit a model with a five-level categorical predictor. How many parameter estimates (coefficients) does the model have for this predictor?

- *a.* 5
- *b.* 4
- *c.* 3
- *d.* 2
- *e.* 1

Incorrect.
The correct answer is **b**. The model will have *k*-1, or 5-1=4, parameter estimates for this predictor.

Assuming that you have three continuous predictors, X1, X2, and X3, and you fit a model with all main effects and two-way interactions, how many two-way interactions are possible?

- *a.* 1
- *b.* 2
- *c.* 3
- *d.* 4

Incorrect.
The correct answer is **c**. There are three possible two-way interactions: X1*X2, X1*X3, and X2*X3. There is also one possible three-way interaction: X1*X2*X3.

For the Impurity scenario, recall that a team is tasked with reducing the impurity in a polymer. You used regression to identify important factors and interactions relative to **Impurity** in the previous video.

You learned that **Catalyst Conc** and **Temp** are the most significant factors, followed by the interaction between **Catalyst Conc** and **Reaction Time**. **Reactor** is significant, which indicates that there is a difference in impurity between the three reactors. There is also a significant interaction between **Reactor** and **Shift**, which indicates that the reactors perform differently, relative to Impurity, on the different shifts.

How could the team use these results toward achieving its goal?

There can be many possible answers.

You can use the Prediction Profiler to find the settings of these three predictors that minimize the predicted Impurity. For example, when **Temp**, **Catalyst Conc**, and **Reaction Time** are at their low levels, with Reactor 1 and Shift 1, the predicted **Impurity** is 3.56. These settings get the team close to the goal of 3% average impurity.

You could conduct an experiment to determine the optimal settings for these factors to minimize impurity.

You could study the reactors to determine what is causing this difference. At the low settings of the continuous predictors, the predicted impurity for **Reactor 1** on second shift is 2.89. You could study how the reactors are operated on the different shifts and address these differences or standardize best practices.

Which of the following is a definition of multicollinearity?

- *a.* when an observation is exerting influence on the model
- *b.* when two or more potential predictors are correlated with the response
- *c.* when none of the predictors in a regression model are significant
- *d.* when two or more predictors in a regression model are correlated with one another

---

Incorrect.
The correct answer is **d**. Multicollinearity is when two or more predictors in a regression model are correlated with one another.

Multicollinearity is the condition in multiple regression, in which two or more predictors are highly correlated with one another. Why is this a problem in explanatory modeling?

- *a.* Collinearity can make it difficult to separate the individual effects of correlated variables on the response.
- *b.* The coefficients might be inflated.
- *c.* The coefficients might have signs that don't make sense (for example, there might be a negative coefficient when the relationship is known to be positive.)
- *d.* The standard errors for these coefficients might be inflated.
- *e.* all of the above

Incorrect.
The correct answer is **e**. All of these are issues that are caused by severe multicollinearity.