Quiz 5

1. (Short answer) Suppose I am interested in the causal relationship between physical activity (predictor of interest) and risk of coronary heart disease (outcome). What role would the variable **age** play in this relationship (precision variable, confounder, or neither)? Explain your reasoning.
2. (Short answer) You are told that the sample mean of a variable X is 5, with 95% confidence interval (2, 8), based on a Normal distribution. Compute the standard error of the sample mean and show your work, or explain why the standard error cannot be calculated based on the information provided.
3. (Short answer) What is the name of the theorem that explains why we do not need the errors of linear regression to be normally distributed when dealing with a large sample size?
4. We are fitting a linear regression model to examine whether Z modifies the association between X and Y.

$E[Y]=\beta\_0+\beta\_1X+\beta\_2Z+\beta\_3X\*Z$

At a 5% significance level, the p-value for the interaction term is significant, but the p-values for the main effect terms of X and Z are not significant. Which of the following is true?

* 1. We can remove the X main effect term from our model, but not the Z main effect term
  2. We can remove the Z main effect term from our model, but not the X main effect term
  3. We can remove both the X and Z main effect terms from our model
  4. We cannot remove either main effect term from our model

1. (Multiple Choice) We are testing whether the mean of X is 0 using a t-test. Our p-value is 0.06. Which of the following is true?
   1. At a 5% significance level, we accept the null hypothesis.
   2. At a 10% significance level, we reject the null hypothesis.
   3. The 95% confidence interval for the mean of X will not contain 0.
   4. The 90% confidence interval for the mean of X will contain 0.
2. (Multiple Choice) I fit a linear regression model with heart rate as my outcome, whether someone is running or not (binary) as my predictor of interest, and whether they are in the park or on a track (binary):

$E[\text{Heartrate}] = \beta\_0 + \beta\_1\text{Running}+\beta\_2\text{Park}+\beta\_3\text{Running}\*\text{Park}$

Which of the following represents the difference in average heartrate between a two groups of people, both of whom are running, when one group is in a park and one group is on a track?

* 1. $\beta\_2+\beta\_3$
  2. $\beta\_1+\beta\_2+\beta\_3$
  3. $\beta\_1+\beta\_3$
  4. $\beta\_1+\beta\_2$

1. (Multiple Choice) Which of the following assumptions are automatically satisfied for simple linear regression with a binary predictor.
   1. Linearity
   2. Independence
   3. Normality
   4. Equal Variance
2. (True/False) A confounder cannot be an effect modifier.
3. (True/False) Data from a well-conducted randomized controlled trial can include precision variables.
4. (True/False) Which R function do we use to get the p-value for a test of whether a categorical variable with more than 2 levels is associated with an outcome?
   1. anova
   2. lm
   3. t.test
   4. prop.test