Quiz 5

1. (True/False) The two slope estimates from a stratified, linear regression analysis, where the stratification variable is binary, are necessarily different than the slope estimate you would get from fitting a linear regression model to the full data.
2. (Short answer) Suppose I am interested in the linear association between birth weight of children and age of parent, and in particular, whether this association differs for birth parents who have had previous children (parity > 0) vs. those who have not. I fit a stratified analysis and obtain the following output from R:

Text

Description automatically generated

Text

Description automatically generated

Interpret the slope coefficients from each stratified model in context.

1. (Multiple response) Select all of the causal pathways from F to C. Note that the options listed are not necessarily exhaustive.

Diagram

Description automatically generated

(In case it is difficult to see the arrows vs. lines, there are arrows from A->B, D->A, B->D, B->C, B->E, E->C, G->E, F->D, F->G, and lines between D-E, C-G, E-F)

* + FDABC
  + FEC
  + FDABEC
  + FGC
  + FGEC
  + FDABDABDABDABDABEC

1. (Short answer) Suppose I am interested in the association between birth weight (in grams) and years of education that the birth parent has had. There are two observations in my dataset who have zero years of education that I remove from my dataset so that I can log transform my education variable. I fit the following a linear regression model in R with birth weight as the outcome and log education in years as the predictor of interest, and obtain the following output:

Text

Description automatically generated

Using the output provided, interpret the coefficient that corresponds to my scientific question *using multiplicative statements* (i.e. you cannot talk about "one unit differences in log(X)"). Note that since the data was subset to include only birth parents with at least 1 year of education, your interpretation should be for birth parents with at least 1 year of education.

1. (True/False) If you conduct a stratified analysis and the slope estimates in each strata are *different*, this suggests that the stratification variable is an effect modifier.
2. (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.
3. (Short answer) Explain what is misleading about the following graph, and explain what you would do to change it to make it less misleading.

Chart

Description automatically generated

1. (Multiple choice) Suppose I gather data related to breast cancer diagnoses in women from a large medical database that contains the following variables for each person in my dataset:

* Has previously had or currently has breast cancer (Yes/No)
* History of breast cancer in family (Yes/No)
* Current age (years)

Each woman is only listed in my dataset only once. I am interested in whether having a family history of breast cancer (predictor of interest) increases the risk of getting breast cancer (outcome). What role would age play in this relationship (confounder, precision variable, neither)? For the purposes of this question, assume that if two variables are likely not associated in the population, then they are not associated in our dataset.

1. Confounder
2. Precision variable
3. Neither
4. (Multiple choice) Suppose I am interested in the linear association between weight prior to pregnancy (in pounds) and birth weight of children (in grams). I choose to adjust for marital status (married yes/no) and years of education a birth parent has had in my model. Fitting this model in R, I obtain the following results:

Text

Description automatically generated

Which of the following is the correct interpretation of the coefficient that corresponds to my statistical question?

1. We estimate that the difference in mean birth weight between two groups of unmarried birth parents with zero years of education, differing by one pound in pre-pregnancy weight is 3.23 grams, with the heavier group having higher birth weights
2. We estimate that the difference in mean birth weight between two groups of birth parents with the same marital status and years of education, differing by one pound in pre-pregnancy weight is 3.23 grams, with the heavier group having higher birth weights.
3. We estimate that the difference in mean birth weight between two groups of birth parents with the same marital status and pre-pregnancy weight, differing by one year in education is 9.75 grams, with the group with more education having higher birth weights.
4. We estimate that the difference in mean birth weight between two groups of unmarried birth parents who weighed zero pounds pre-pregnancy, differing by one year in education is 9.75 grams, with the group with more education having higher birth weights.
5. We estimate that the difference in mean birth weight between two groups of birth parents differing by one pound in pre-pregnancy weight is 3.23 grams, with the heavier group having higher birth weights.
6. (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.