Data Analysis Project: First Steps Intervention Program to Reduce Risk of Low Birth Weight

In this project, you will use the tools discussed in class to analyze infants' birthweights from a sample in King County after the introduction of the First Steps (FS) program. ***Your data analyses should be systematic statistically and scientifically and statistically appropriate***. You are assessing whether participation in the First Steps program increases birthweight, and you will make a recommendation to the board of Public Health King County based on your results. ***Write your report as though you are addressing the board of Public Health. Include an abstract, very brief Introduction, Results and Discussion, along with your recommendations. A partial template for the report is given in the guide below, but you are NOT restricted to this verbage.***

The data file is BirthweightsKingCounty2001-Biost514-517-2022.txt. We first recommend examining the data in Excel to make sure the data are not obviously corrupted, that each column has a header, and to see how many rows/observations to expect. You should find N=2500. Read the data into R and check the resulting matrix for size.

The overall plan is to first check univariate summary statistics to assess for outliers and potential confounders. Then, make appropriate plots with the potential confounders to assess their impact. Make note of any plots that will be useful to include in the report to support your analysis decisions and/or help explain the results. You will have a final estimate (with 95% CI) for the difference in birthweights between babies whose mothers participated in First Steps and babies whose mothers didn't participate in First Steps.

What follows are a list of questions we would like you to consider and some suggestions for analyses that we would like you to conduct. While there is no one correct way to analyze this dataset, we strongly recommend that you carefully read through everything we have written and ensure that your analyses can answer all the questions we have asked you to consider.

The value of each question is reported in red, e.g., "(4)" means four points will be given for getting the right answer.

Results

To start, create numeric variables for the binary variables that are coded using characters, such as the baby's sex. Next, make a table of descriptive statistics, stratifying on "firstep." All the other variables given in the dataset have been shown to affect birthweight (BW), meaning that any variable that is also associated with "firstep" is a potential confounder. Variables that are more imbalanced between First Step groups tend to have a larger impact on confounding. Remember that when choosing whether to adjust for a particular variable as a confounder, you must consider whether that variable is in the causal pathway.

Gather your descriptive statistics in a table labeled Table 1 and write at least one paragraph describing Table 1 (4). In your paragraph, we would like you to answer the following questions:

- Do you see any evidence of outliers? (1)
- When looking at birthweights, is there evidence of a skew? (1) If so, in which direction is the skew, and what do you think is causing the skew? (1) If not, what makes you think that there is no skew? (Hint: it might be a good idea to create a histogram for birthweight to answer this).

Next, we would like you to give a preliminary answer to the question of interest: what is the overall difference in birthweight between babies whose mothers were enrolled in First Steps and babies whose mothers were not enrolled in First Steps? Report your answer with a 95% confidence interval. (4)

Having given a first answer to the question, we would like you to take a more nuanced approach by considering variables that may affect this difference and their potential for intervention. Based on your information from Table 1 and any further analyses as necessary, we would like you to consider the following:

- How many participants were enrolled in First Steps? How many participants in the study were not enrolled in First Steps? (1)
- Of the participants in First Steps, what percentage were single? How does this compare to the non-participants? (1)
- Of the participants in First Steps, what percentage were on welfare? How does this compare to the non-participants? (1)
- Of the participants in First Steps, what percentage were smokers? How does this compare to the non-participants? (1)
- Were participants in First Steps older than non-participants? (1)
- Were participants in First Steps more educated than non-participants? (1)
- What is the racial distribution among participants? How about for non-participants? (1)
- Is there a difference in mean weight gain between participants and non-participants? (1)

Create another table (Table 2) that shows the effects on birthweight of potentially modifiable dichotomous variables that differ in proportion between participants and non-participants in the overall sample. (4) (Your answers to the questions above will likely be very relevant here!) Create a figure that investigates whether the effects of these risk factors in this sample are larger than the overall difference in BW between participants and non-participants. Comment on both the table and figure. (4) We would like you to answer the following questions:

- Of the risk factors that you have investigated, which do you think is the most amenable for modification? In other words, which risk factor do you think is easiest to get people in the population to limit their exposure to? (1)
- Report the treatment effect based on the risk factor that you identified as amenable for modification. In other words, report on the difference in birthweight between participants and non-participants among mothers who were exposed to the risk factor, and then do the

same for mothers who were not exposed to the risk factor. How do the two compare? Has your answer to the question of interest changed? (Hint: a t-test would be useful here.) (2)

- Next, make a figure that stratifies on race. In other words, graphically compare the difference in birthweight among participants and nonparticipants for mothers of different races. What trends do you notice? Does the effect of participation in First Steps vary with race? Comment on what you see. Describe which races seem to benefit from participation in First Steps and which races do not seem to benefit from participation. (4)

Low Birthweight and Very Low Birthweight

We have one final issue we would like you to consider. The American Academies of Obstetrics and the American Academy of Pediatrics have defined a birthweight of below 2,500 grams as "low birthweight". Birthweights of below 1,500 grams have been identified as "very low birthweight." For the First Steps group, what percentage of the births were officially low BW? How does this compare to the birthweights of non-participants? Repeat these calculations for very low birthweight, as well. Write up your findings in a table or paragraph – whatever you think communicates the results best. (4)

Feel free to include an additional result of your choice. (up to 4)

Discussion

Based on your analyses of the data, answer the question of interest: what is the overall effect of the First Steps program in this sample? It would be a good idea to mention the selection bias that you know about (i.e. which mothers were more likely to be recruited in the intervention group?) and how that is likely to influence the results.

Also consider: What are the mechanisms by which the various risk factors exert their effects? A successful intervention would address these mechanisms. Some are likely immutable. Should the "final answer" be adjusted for risk factors? What recommendations do you make? (4 for content)

Up to 4 points for writing style, readability, and clarity.

50 points total. Up to an additional 5 points extra credit for particularly informative analyses, but reports should **not** exceed 5 pages.