AEDE 2005 – Spring 2019

Group Assignment, Section A – Due 11:59 PM, Friday, Feb. 01, 2019

Describing and Exploring National School Lunch Participation in the U.S. (3 pages)

For this assignment, use the data set uploaded to Carmen entitled 'Section A Data.xlsx.' It is a Microsoft Excel file that contains one sheet with the data and a row of variable names at the top of each column. Note: any blank cell means the datum for this variable and this observation is either missing or not checked. The second sheet in this file explains the meaning of the variable names and how data were coded (e.g., yes = 1, no = 0). You will also find it useful to download and refer to 'Section A Survey,' which is the survey that parents and students received.

2 Notes about the data

- I trimmed some of the answers from the survey from the data set (so it wasn't' so large) and added a few items that aren't in the survey to the data set such as weights, id etc.
- Read the survey document "Parent-Student Survey and Codebook.pdf" first to understand a key pattern in the questionnaire. Parent and student answer together and individually depending on the question, and in a few occasions depending on a participant's answers, they answer different subsets of latter questions, which will be important for understanding the data set.

Task 1: (20%) Assess and communicate the quality of the data.

Calculate the number of missing observations for each variable and then construct a histogram of the number of missing observations. Are there any variables contain a large number (e.g., more than half) of missing entries? How many categorical variables contain values outside the allowed range (e.g., a '9' when the variable is only coded 1 through 8)? Are there any variables that have non-numeric characters or numbers that do not match the number format of the variable? How many variables contain outliers? If there are any outliers, do you think the numbers are simple errors when the data was entered or do you think these might be accurate data that is truly unusual? Please summarize whether you find this a 'clean' data set that is ready for analysis.

Grading Guide: an 'A' level response will document how outliers are defined, defend your answer for the reason for the outlier, document the number of outliers and missing entries for each variable and communicate all of this in a succinct paragraph along with the summary assessment concerning the 'cleanness' of the data.

Task 2: (20%) For the variable 'SO_4b' (how much is paid for a typical school lunch), summarize the variable using a graphical technique of your choosing and provide key summary statistics (at least one central tendency and one dispersion measure). Repeat the analysis for only the observations where the variable 'P/S_5' = 1 (students who ate school lunch during the most recent school day). Then repeat the analysis for three groups: students who attend public, private, and charter schools ('P/S_S_type'). Briefly discuss any differences or similarities that you observe between the entire sample and the subset of districts where 'P/S_5' = 1 and briefly discuss differences or similarities between public, private, and charter school students.

Grading Guide: an 'A' level response will defend your choice of graphical techniques and descriptive statistics and provide a succinct paragraph summarizing differences or similarities across the groups.

Task 3: (20%) Assess and communicate which continuous variables in the data set feature the greatest statistical association and speculate as to the basis for any association levels that are greater than 0.10 in absolute value.

Grading Guide: an 'A' level response will, in a succinct paragraph, defend the measure of association you chose, provide a succinct summary of the levels of association across all variables in the section, and defend your choice of the variable or variables with the greatest association. In another succinct paragraph present an intuitive argument for the basis of association or an argument from existing academic literature.

Task 4: (10%) Assess and communicate whether the continuous variables (in terms of their averages) are sensitive to using variable weights. *Note: only this task should be completed with the use of weights.*

Grading Guide: an 'A' level response will, in a succinct paragraph, explain how you calculated and compared weighted and unweighted averages and how you chose the variables identified as most sensitive.

Task 5: (20%) Create a cross-tabulation between the variable 'SO_6' (Self-assessed autonomy in lunch related decision making)' and variable 'P/S_age' (child age). Discuss if the pattern revealed in the cross tabulation comports with your expectation of the relationship between child's age and the autonomy in lunch related decisions (you are encouraged to look at other relevant academic literature or data to learn about parent child interaction in food choice behavior). Repeat the exercise by doing cross-tabulation between 'PO_2_16' (percentage of household income spent on food) and 'SC_3' (household income).

*Grading Guide: an 'A' level response will use a pivot table to formulate a cross tabulation table between the age categories and an appropriate number of bins for the '*Self-assessed autonomy in deciding lunch*' variable. The table should be easy to read with clear column and row labels. The discussion should draw upon economic logic to argue for a likely pattern and clearly discuss how the results either meet or deviate from this expectation. Same grading criteria applies for the second cross-tabulation.*

Task 6: (10%) What question or questions about students' participation in National School Lunch programs would you most like to see answered?

INSTRUCTIONS FOR SUBMITTING YOUR ASSIGNMENT

1. Each group will deliver:
    a. A document (MS Word or pdf) containing (1) the names of all group members and (2) responses to the tasks listed above, and
    b. One supporting spreadsheet with calculations and other work.
    c. Notes:
        i. Title both files you will upload starting with your group number.
            1. For example, if you are group 15, you should have "15_responses.docx" and "15_spreadsheet.xls"; the exact wording isn't crucial, but starting with the group number is.
        ii. Embed relevant graphs and charts into the document. This is more difficult as the formatting sometimes transfers imperfectly, but learning this skill will prove useful as you prepare more reports throughout your academic and professional careers.

iii. To be ensured full credit, deliver this no later than **11:59 PM, Friday, Feb. 01, 2019**.
iv. Choose one group member to be the 'uploader', and have him/her deliver all materials to the Carmen dropbox for the project.

d. Each member of your group must upload a group evaluation spreadsheet. If you don't complete this, you receive a 'zero' for the group project.

Good Luck!