EPIB697

Jenna Harrison-Peters

HW 8

HW 8

* Summarize the sampling design for Add Health (bullet points are acceptable).
* Use the AddHealth sample to test the hypothesis that sex  affects weight **with and without** accounting for complex design (weight and cluster variables). Please add the SAS code and results. You can test this hypothesis by calculating the 95% Confidence Intervals of the mean of each sex and then checking if they overlap or by using a linear regression model in which the physical weight of the person is the dependent variable and the gender is the independent variable. Note: In case that you are wondering where the survey weights and cluster information are (given that they are not in the same data set as the questionnaire data):  The survey weight and cluster information for Wave 1 is in the file 21600-0004-Data.stc (this file is provided in this module). You can see how I did this in the programs AddHealthSamplePrePROC.sas and AddHealthSampleBasicPROC.sas and modify them appropriately.
* Compare and contrast your results with and without accounting for complex design in writing. Do you get the exact same statistics in both cases?
  + Note 1: You can test this hypothesis by calculating the 95% Confidence Intervals of the mean of each sex and then checking if they overlap or by using a linear regression model in which the physical weight of the person is the dependent variable and the gender is the independent variable.
  + Note 2: The physical weight of the person is the variable *w\_w1r*while the survey weight is *GSWGT1* . You can see this in the programs that I have used for the preprocessing of the AddHealth data, e.g. *AddHealthSamplePrePROC.sas*.
  + Note 3: The options CIBASIC and CLM can help to produce the 95% CI for PROC UNIVARIATE and PROC MEANS, respectively.
* The data set [middle\_school2.sas7bdat](https://umd.instructure.com/courses/1284684/files/57156989/download?wrap=1) contains a the information corresponding to a sample (n=2030) from middle schools students that answered the 2011 Youth Risk Behavior Surveillance System ([YRBSSPreview the document](https://umd.instructure.com/courses/1284684/files/57156990/download?wrap=1)) in Montgomery County  (not really but assume that this is the case).  A preliminary analysis shows that the sample proportions for gender and ethnicity are not the same as those for the general population (Population Counts: Hispanic Females=352, Hispanic Males= 402, Non-Hispanic Females=6164, Non-Hispanic Males=6639 for a total population size of N=13,557). Based on this sample administrators from the school district what to know what percentage of their students use a helmet when they ride a bicycle, how many have had sex before and what percentage has been electronically bullied. Find this percentages taking into consideration that you will need to create the appropriate weights to make the sample representative of the gender and ethnicity proportions that exist in your population. Note.- SAS for Windows users might need to include the option*OPTIONS NOFMTERR;*before the start of their code.