BIOS 6611 Homework 9

Due Monday, November 26, 2018 <u>by midnight</u> to Canvas Assignment Basket

The data for this exercise (lead2.sas7bdat SAS dataset available on Canvas in data repository and on assignment page) are from a study of the neurological and psychological effects of environmental lead exposure in a population of children who had lived near a lead smelter in El Paso, Texas. The study was conducted because previous work suggested that blood-lead levels of 40-80 μg/100ml adversely affected the nervous system, and this level of absorption might be associated with behavior abnormalities. Thus, the study purpose was to look for a relationship between lead exposure and neurological damage or behavior problems in children.

The data set for this exercise comes from a group of 124 children who had participated in the survey of blood lead levels in the larger study and who had lived near the lead smelter for at least 12 of the 24 months preceding the study. Forty-six of these children (the "exposed group") had high blood-lead levels ($\geq$ 40μg/100ml). The other children (the "unexposed group") had blood-lead levels less than 40μg/100ml. The study also collected data on the distance each child lived from the smelter, how long the child lived at the residence, and whether the child lived in that residence during the first 2 years of life.

After obtaining parental permission, children were evaluated individually by examiners who did not know the child's blood-lead level or exposure group. Children first underwent complete medical and neurological evaluations. All children were given a battery of neurological tests including the Wechsler intelligence scale for children (WISC) or the Wechsler preschool and primary scale of intelligence (WPPSI) used to measure intelligence (IQ).

id	Subject ID number
age	Age in years
sex	Sex (1 = male; 2 = female)
race	Race (1=African American; 0=Non-Hispanic White)
resdur	Years lived at current residence
miles	Distance of current residence from the smelter (miles)
first2y	Did the child live at current residence during the first 2 years of life? (1=yes; 0=no)
expose	Lead exposure group, blood lead levels >40μg/100ml (0 = not exposed; 1 = exposed)
iq	IQ test score

1) Use linear regression to examine the relationship between IQ (*iq*) and lead exposure (*expose*):

   A) What is the unadjusted (crude) estimate for the association between IQ and lead exposure? Write a brief, but <u>complete</u>, summary of the relationship between IQ and lead exposure.

   B) Adjusting for the effect of race, what is the adjusted estimate for the association between IQ and lead exposure? Write a brief, but <u>complete</u>, summary of the relationship between IQ and lead exposure adjusting for race.

   C) Is race a confounder of the association between IQ and lead exposure? Should you report the results from (A) or (B)? Justify your answer.

2) Use linear regression to examine the relationship between IQ (*iq*) and the distance of the current residence from the smelter (*miles*). In this question you will examine whether the magnitude of the association between IQ (the response) and distance of the residence from the smelter (the primary explanatory variable) depends on whether the child was exposed during the first two years of life (i.e., if they lived in the current residence during the first two years of life).

   A) Write down the regression equation for the regression of *IQ* on *miles*, *first2y*, and the interaction between *miles* and *first2y*. Provide an interpretation for each of the coefficients in the model (including the intercept).

   B) Test whether the relationship between *IQ* and distance of the residence from the smelter (*miles)* depends on whether the child lived in the residence during the first two years of life.

   C) What is the regression equation for children who lived in the current residence during the first two years of life? What is the regression equation for children who didn't live in the residence during the first two years of life?

   D) Provide a brief, but <u>complete,</u> summary of the relationship between IQ and distance of the current residence from the smelter, accounting for any observed interaction with exposure during the first two years of life. For your summary, include a scatterplot of *IQ* versus *miles*, using different symbols and separate regression lines for children who lived in the residence during the first two years of life and for those who didn't live in the residence during the first two years of life. Be sure to comment on the graph in your summary.

3) By "hand", using the output from PROC REG, determine if the addition of both the covariate first2y and the interaction term miles*first2y in question 2A significantly contributes to the prediction of IQ, given the variable miles is included in the model 2A. *Don't just use the commands in PROC REG to specify the partial F test, although you can check your answer with it.* (i.e. what is the reduced and full model, the null hypothesis, and how do you test this?) [Hint: $F_{2,120} = 3.07$.]

```
*************************************;
***  BIOS 6611              ***;
***  Assignment #9 some SAS code ***;
***  Use at your own risk      ***;
*************************************;


/* Begin Question 2*/
/* New DATA step to create new variables */
DATA lead2;
SET lead; /* name of your original data set, in this case lead */
IF first2y = 1 THEN notfirst2y = 0;
IF first2y = 0 THEN notfirst2y = 1;

milesf2y = miles*first2y;
milesNf2y = miles*notfirst2y;

LABEL notfirst2y = 'Not exposed first 2years'
      milesf2y = 'miles*first2y'
      milesNf2y = 'miles*notfirst2y';
RUN;




/* Question 2d figures */
/* SGPlot Code for 2D */
PROC SGPLOT DATA=lead2;
     REG Y=iq X=miles / GROUP=first2y;
     FORMAT first2y;
RUN;


/* Plot for 2D */
PROC GPLOT DATA=lead2;
     PLOT iq*miles = first2y / VAXIS=axis1 HAXIS=axis2 LEGEND=legend1;
     SYMBOL1 I=rl VALUE=circle COLOR=black LINE=3 WIDTH=2;
     SYMBOL2 I=rl VALUE=dot COLOR=black LINE=1  WIDTH=2;
     AXIS1 LABEL = (FONT=ARIAL HEIGHT=2.5 ANGLE=90 POSITION=center  )
       VALUE=(FONT=ARIAL HEIGHT=2);
     AXIS2 LABEL = (FONT=ARIAL HEIGHT= 2.5 POSITION=center )
       VALUE=(FONT=ARIAL HEIGHT=2);
     LEGEND1 FRAME LABEL=(FONT=ARIAL HEIGHT= 1.5) VALUE=(FONT=ARIAL
HEIGHT=1.5)
         POSITION=(bottom inside right) ACROSS=1
          ;
     FORMAT first2y;
   RUN;
```