Methods Homework 9

Tim Vigers

11/26/2018

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?

There is a significant difference in average IQ for those who haven't been exposed to lead versus those who have (p = 0.0084). On average, IQ is 7.77 points (95% CI: 13.51 to 2.03 points) lower for those exposed compared to unexposed.

B. Adjusting for the effect of race, what is the adjusted estimate for the association between IQ and lead exposure?

After adjusting for race, there is no longer a significant difference in average IQ for those who haven't been exposed to lead versus those who have (p = 0.086). On average, IQ is 7.47 points (95% CI: 16.02 to 1.07 points) lower for those exposed compared to unexposed.

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

I think that race fits all of the requirements for a confounding variable. It's associated with the exposure and is a risk factor in the sense that IQ tests are known to be racially biased. Also, race is not affected by the exposure or the dependent variable. Finally, from an operational point of view it is a confounder because the betas are different in the crude model and adjusted model. However, you could also argue that the adjusted and crude models are still pretty close, so it might depend on your threshold. If you don't need the parameter estimates to be exactly the same, then you could argue that race actually isn't a confounder.

Ideally, I think it would be good to report both models if possible, but if you have to pick one the adjusted model is a more accurate reflection of the data. However, I think you could make the argument that the models are close enough, and that race is not really a confounder. As long as you're explicit about what you did and why you won't count race as a confounder, I don't think there's anything wrong with reporting the crude model.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1747.07448	1747.07448	7.17	0.0084
Error	122	29715	243.56576		
Corrected Total	123	31462			

Root MSE	15.60659	R-Square	0.0555
Dependent Mean	99.82258	Adj R-Sq	0.0478
Coeff Var	15.63433		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	102.70513	1.76710	58.12	<.0001	99.20698	106.20328
expose	Lead exposure group	1	-7.77035	2.90130	-2.68	0.0084	-13.51376	-2.02693

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1749.21830	874.60915	3.56	0.0314
Error	121	29713	245.56098		
Corrected Total	123	31462			

Root MSE	15.67039	R-Square	0.0556
Dependent Mean	99.82258	Adj R-Sq	0.0400
Coeff Var	15.69824		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	102.73167	1.79692	57.17	<.0001	99.17420	106.28914
expose	Lead exposure group	1	-7.47285	4.31551	-1.73	0.0859	-16.01655	1.07085
race	Race	1	-0.41404	4.43127	-0.09	0.9257	-9.18690	8.35882

2. Use linear regression to examine the relationship between IQ (iq) and the distance of the current residence from the smelter (miles).

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).

$$\hat{IQ} = 98.45 + 0.59 * \text{miles} + -19.61 * \text{first2y} + 17.65 * \text{miles} * \text{first2y}$$

$\beta_0 = 98.45 =$ Theoretical average IQ for those who live 0 miles from the smelter, but were not in the residence during the first two years of life.

$\beta_1 = 0.59 =$ Slope for those not living in the residence during the first two years of life.

$\beta_2 = -19.61 =$ The difference in IQ at 0 miles, between those living in the residence during the first two years of life and those not.

$\beta_3 = 17.65 =$ The difference in slope between those living in the residence during the first two years of life and those not.

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.

The relationship between IQ and the distance from the smelter does depend on whether or not the participant lived in the residence in the first two years of life, because the interaction term between distance from the smelter and first2y is significant (p = 0.0016).

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?

For those who lived in the residence during the first two years of life:

$$\hat{IQ} = 98.45 + 0.59 * \text{miles} + -19.61 * 1 + 17.65 * \text{miles} * 1 =$$

$$78.84 + (0.59 + 17.65) * \text{miles} = 78.84 + 18.24 * \text{miles}$$

For those who did not live in the residence during the first two years of life:

$$\hat{IQ} = 98.45 + 0.59 * \text{miles} + -19.61 * 0 + 17.65 * \text{miles} * 0 =$$

$$98.45 + 0.59 * \text{miles}$$

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3106.32999	1035.44333	4.38	0.0058
Error	120	28356	236.29806		
Corrected Total	123	31462			

Root MSE	15.37199	R-Square	0.0987
Dependent Mean	99.82258	Adj R-Sq	0.0762
Coeff Var	15.39931		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	98.45133	4.38619	22.45	<.0001	89.76699	107.13568
miles	Distance from smelter (miles)	1	0.59134	2.10468	0.28	0.7792	-3.57577	4.75845
first2y	Exposed during first 2yrs	1	-19.61160	7.85829	-2.50	0.0139	-35.17046	-4.05275
milesf2y	miles*first2y	1	17.65461	5.48109	3.22	0.0016	6.80244	28.50678

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3106.32999	1035.44333	4.38	0.0058
Error	120	28356	236.29806		
Corrected Total	123	31462			

Root MSE	15.37199	R-Square	0.0987
Dependent Mean	99.82258	Adj R-Sq	0.0762
Coeff Var	15.39931		

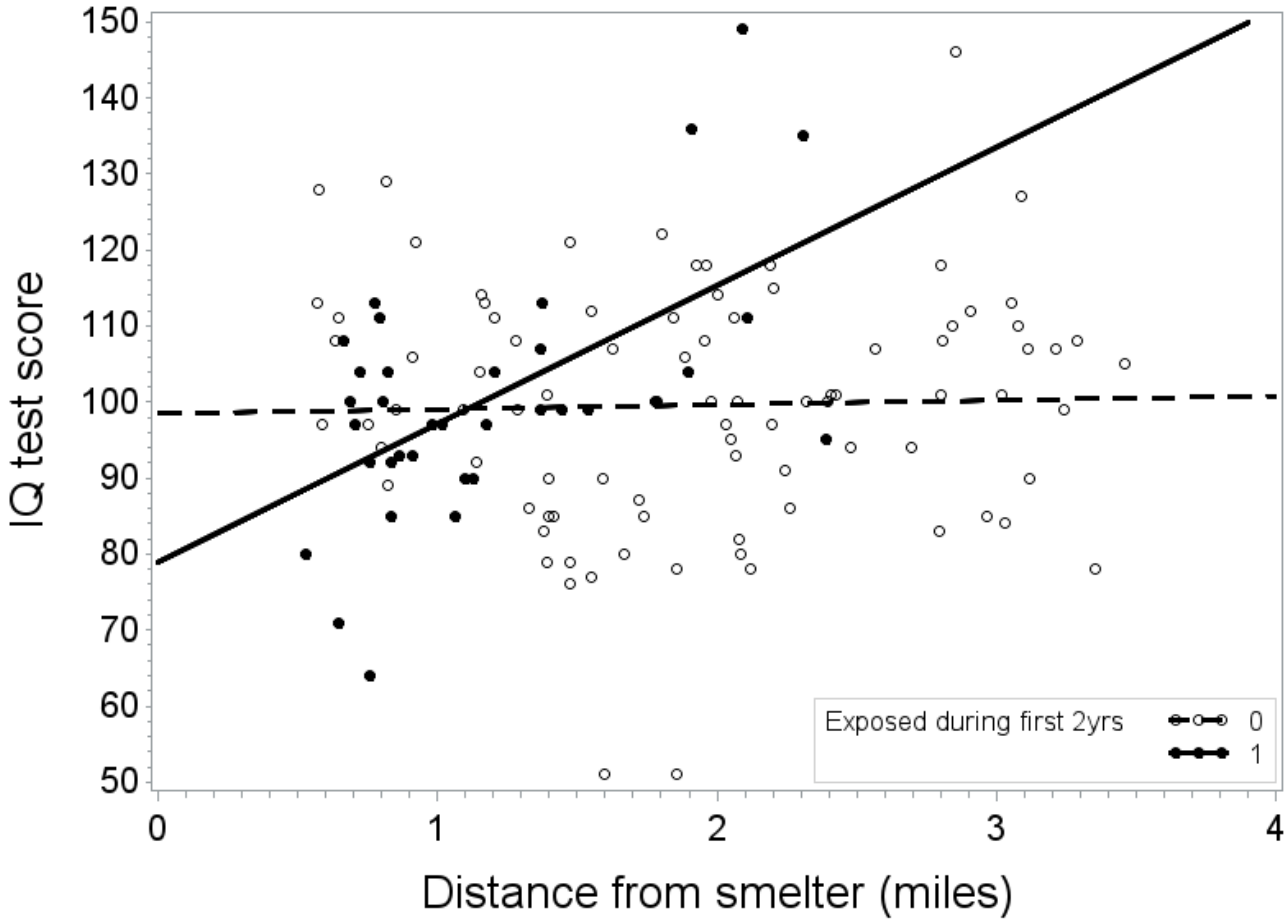
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	78.83973	6.52028	12.09	<.0001	65.93003	91.74943
miles	Distance from smelter (miles)	1	18.24595	5.06089	3.61	0.0005	8.22574	28.26617
notfirst2y	Not exposed first 2years	1	19.61160	7.85829	2.50	0.0139	4.05275	35.17046
milesNf2y	miles*notfirst2y	1	-17.65461	5.48109	-3.22	0.0016	-28.50678	-6.80244

D. Provide a brief, but complete, 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 those not exposed during the first two years of life, the average IQ is 98.45, and this increases by 0.59 points on average (95% CI: -3.58 to 4.76 points), for each mile away from the smelter the participant lives, although the relationship is not significant ($p = 0.78$). However, whether or not the participant was living in the residence during the first two years of life has a significant impact on the relationship between IQ and distance from the smelter ($p = 0.014$). For those who were living in the residence during the first two years of life, the average IQ is 78.84, and increases by 18.24 points on average (95% CI: 8.22574 to 28.26617 points), for each mile away from the smelter the participant lives. This relationship is significant ($p = 0.0005$), which means there is an interaction between distance from the smelter and whether or not the participant was exposed in the first two years of life.

All of these effects are fairly apparent when you look at the plot of IQ vs. miles from the smelter for each group (exposed in the first two years or not exposed). The intercept for non-exposed participants is close to 100, and is close to 80 for the exposed group. As the distance increases, the non-exposed group's predicted IQ stays about the same, only increasing a tiny bit as you'd expect with a slope of 0.59. However, the exposed group's IQ increases much more quickly as distance increases. This all suggests that children under 2 should not live within about 1 mile of the smelter (since the trend lines cross at about the 1 mile mark).


IQ test score
150
140
130
120
110
100
90
80
70
60
50
0
1
2
3
4
Distance from smelter (miles)
Exposed during first 2yrs
0
1

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.

The reduced model is a simple regression of IQ vs. miles, and the full model includes first2y and the interaction term. Since there are two additional coefficients in the full model, we set k = 2. Using this, we calculate an F statistic in order to test the null hypothesis (that all of our betas are equal to 0).

$$F = \frac{SS_{\text{model}}(full) - SS_{\text{model}}(reduced)/k}{MS_{\text{error}}(full)} = \frac{(3106.32999 - 443.85069)/2}{236.29806} =$$

$$\frac{1331.24}{236.29806} = 5.63$$

Because the F statistic is larger than F2,120 = 3.07, we know that the addition of the two variables does improve the model (p = 0.0046).

The SAS System

The REG Procedure
Model: MODEL1

Test first_miles Results for Dependent Variable iq				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1331.23965	5.63	0.0046
Denominator	120	236.29806		

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3106.32999	1035.44333	4.38	0.0058
Error	120	28356	236.29806		
Corrected Total	123	31462			

Root MSE	15.37199	R-Square	0.0987
Dependent Mean	99.82258	Adj R-Sq	0.0762
Coeff Var	15.39931		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	98.45133	4.38619	22.45	<.0001
miles	Distance from smelter (miles)	1	0.59134	2.10468	0.28	0.7792
first2y	Exposed during first 2yrs	1	-19.61160	7.85829	-2.50	0.0139
milesf2y	miles*first2y	1	17.65461	5.48109	3.22	0.0016

The SAS System

The REG Procedure
Model: MODEL2
Dependent Variable: iq IQ test score

Number of Observations Read	124
Number of Observations Used	124

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	443.85069	443.85069	1.75	0.1889
Error	122	31018	254.24792		
Corrected Total	123	31462			

Root MSE	15.94515	R-Square	0.0141
Dependent Mean	99.82258	Adj R-Sq	0.0060
Coeff Var	15.97349		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	95.68282	3.44488	27.78	<.0001	88.86334	102.50231
miles	Distance from smelter (miles)	1	2.40366	1.81921	1.32	0.1889	-1.19765	6.00497