DATA 557: HW Assignment 2

Hriday Baghar

February 11, 2022

Data: 'lead.csv'

The data are from a study of the association between exposure to lead and IQ. The study was conducted in an urban area around a lead smelter. A random sample of 124 children who lived in the area was selected.

Each study participant had a blood sample drawn in both 1972 and 1973 to assess blood concentrations of lead. The children were grouped based on their blood concentrations as follows: - Group 1: concentration < 40 mg/L in both 1972 and 1973 - Group 2: concentration > 40 mg/L in both 1972 and 1973 or > 40 mg/L in 1973 alone (3 participants) - Group 3: concentration > 40 mg/L in 1972 but < 40 mg/L in 1973

Each participant completed an IQ test in 1973. (A subset of the IQ scores from this study were used in HW 1, Question 3.) The variables in the data set are listed below. - ID: Participant identification number - SEX: Participant sex (1=M or 2=F) - GROUP: As described above (1, 2, or 3) - IQ: IQ score

1. The first goal is to compare the mean IQ scores for males and females. Use a 2-sample t-test for this comparison. What is the p-value?

2. State the conclusion from your test.

3. Are the independence assumptions valid for the t-test in this situation? Give a brief explanation.

4. The second goal is to compare the mean IQ scores in the 3 groups. State in words the null hypothesis for this test.

5. State in words the alternative hypothesis for this test.

6. What method should be used to perform the test?

7. Perform the test. Report the p-value.

8. State your conclusion about the evidence for an association between lead exposure and IQ.

9. Are there strong reasons to believe that the assumptions of this test are not met? Briefly justify your answer.

10. Conduct all pairwise comparison of group means. Report the p-values.

11. What conclusion about the association between lead and IQ would you draw from the pairwise comparisons of group means? Does it agree with the conclusion in Q8? (Consider the issue of multiple testing in your answer.)

12. Now we wish to compare the 3 group means for males and females separately. Display some appropriate descriptive statistics for this analysis.

13. Perform tests to compare the mean IQ scores in the 3 groups for males and females separately. Report the p-values from the two tests.

14. What can you conclude about the association between lead and IQ from these tests? Does it agree with the result in Q8 and Q11? (Consider multiple testing.)

15. Now perform all 3 pairwise comparisons of groups for males and females separately. Report the p-values from these tests?

16. What do you conclude about the association between lead and IQ from the results in Q13? Does your conclusion change from previous conclusions made in Q8, Q11 and Q14?