Statistics 101 – Lab Report 5

Questions

1) These questions are based on your reading of the article. You don’t need to use JMP.

a. What are the experimental units (the subjects under investigation)? 

b. What are the two treatment conditions?

c. Is this an observational study or a randomized experiment?

d. Are there background characteristics (not treatment or response variables) that differ substantially in the two groups? If so, state which one(s).
2) The main analysis of the paper compares mean performance IQ scores (W.I.S.C. + W.P.P.S.I.) for the high lead and low lead groups.

a. Report the means and SDs you found.

High lead group: ________________________________

Low lead group: ________________________________

b. Do the means and standard deviations for performance IQs for the high and low groups match the means and standard deviations reported in the paper?

__________

c. Exclude the four outliers and compare the new means and SDs to those in Question 2a. Report the new means and SDs in the spaces provided.

High Lead Level Group

New Mean: ____________________  New SD: ____________________

Low Lead Level Group

New Mean: ____________________  New SD: ____________________

d. Are any of the changes big enough that the authors should have mentioned the effect of the outliers in their article?

____________________________________________

____________________________________________

e. A comparison of means and standard deviations might be inadequate. For example, suppose one group has a right-skewed distribution, and the other group has a left-skewed distribution. Just reporting means and standard deviations does not capture such structure.

Compare the distributions of performance IQ of the high and low lead groups. Describe succinctly any differences between the two groups' distributions of performance IQ in two
3) The authors chose to evaluate blood lead level as a categorical variable rather than as a continuous variable.

a. Report either the correlation or $R^2$. 

b. Is there a strong linear relationship between performance IQ and the blood level in 1972 measured on a continuous scale?

4) Compare the distributions of age in months for the high and low lead groups.

a. Report what difference you found between the two groups. (one sentence)

b. Does it agree with what was reported in the paper?

c. Is age a possible confounding factor?

5) One of the main analyses of the paper is a regression of finger-wrist tapping speed on age in months, which was calculated separately for the high and low lead level groups (bottom right corner of page 710). Replicate these regressions in order to check their validity.

a. Report the regression equations that you found.
b. Do your regression equations match (within rounding) the equations on the bottom of page 710? 

________________

c. Compare the typical deviations from the regression line for the high and low lead level regressions. Report the ratio of the two values. (Hint: If the ratio is near 1, i.e. less than 1.5, it is reasonable to say that they are of similar magnitude.)

______________________________________________________________________________

d. Are they of roughly similar magnitude? __________________

e. Does your analysis verify that the regression lines are nearly parallel? _____________

f What is the difference in predicted average tapping speed between a kid in the low lead group and a kid of the same age in the high lead group?

__________________________________________________________________________

g. Would your answer to this question change if the lines were not parallel? Explain why or why not.
h. Examine the plots of residuals versus the predictors. In no more than two sentences, explain any patterns that would cause you to worry about the validity of the regression assumptions.