Practice Midterm Questions:

Part 1 (ACED data)

1. All the data were taken from a single suburban middle school in New Jersey. Does that present a problem for generalizing to other middle schools?
2. The students were randomly assigned to study conditions, and no Control student has an EAP.sgp score (as they didn’t use ACED). Does this introduce a bias?
3. Is the distribution the same for everybody in the group? Are there differences by study condition (Cond\_code)? class level (Level\_Code)? class (Session)?
4. What at the distributions of (EAP.sgp) and (post\_scaled) like? High skeweness? High kurtosis? Outliers?
5. Is the relationship mostly linear? Will transforming or make it more linear?
6. Is the variance roughly the same for all values of ? Will transforming help?
7. Are the residuals roughly normal? Are there any outliers? Will transforming help?
8. Is there enough evidence in the sample to conclude that and are related? How strong is the relationship?
9. What is the equation of the regression line?
10. The ELL students had slightly different administration conditions. How sensitive are the conclusions to those students?
11. What can we conclude about the relationship between and ? In particular, both EAP.sgp and post\_scaled are measures of the students ability to solve geometric sequence problems? Are they measuring something similiar?

Part 2 (Teacher Pay data)

## Questions

1. The data are from 1986. How much will that tell us about 2023?
2. Are there any clusters in the data with different distributions?
3. What at the distributions of (Spend) and (Pay) like? High skeweness? High kurtosis? Outliers?
4. Is the relationship mostly linear? Frequently with money, log transformations help. Which model is better Pay ~ Spend or log(Pay) ~ log(Spend)
5. Is the variance roughly the same for all values of ? Will transforming help?
6. Are the residuals roughly normal? Are there any outliers? Will transforming help?
7. Are there any differences by region?
8. Is there enough evidence in the sample to conclude that and are related? How strong is the relationship?
9. What is the equation of the regression line?
10. Which states are outliers in the regression?
11. Which states are high leverage points?
12. What is the sensitivity of the results to the high leverage values? Is it of concern?
13. What can we conclude about the relationship between and ? In particular, does increasing spending cause an increase in pay? Increasing pay cause an increase in spending? Or is there some common factor?