Assignment 03

Correlation and Standardized Regression

This goal of this assignment is to give you more experience fitting and interpreting regression models. Submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and also should have an appropriate caption. This assignment is worth 10 points. (Each question is worth 1 point unless otherwise noted.)

Should more money be spent on public schools or should that money be spent elsewhere? Both sides of this ongoing public debate have been argued passionately, using a multitude of anecdotal evidence. Although we will not settle this debate, we will examine data akin to the types of data that policy makers use to make funding decisions. Specifically, we will examine whether teacher salaries are related to SAT scores at the state level. For this assignment, you will use the file *state-education.csv* (see the data codebook).

Part I: Correlation

Before carrying out any analyses, create a predictor called `salary_thousand` that indicates the average state salary in thousands of dollars (e.g., `salary` = 52143; `salary_thousand` = 52.143). This variable (not `salary`) should be used in all analyses in Part I.

1. Compute and report the Pearson correlation coefficient between SAT scores and teacher salaries.

2. Is the Pearson correlation coefficient an appropriate summary measure of the relationship? Explain. (Hint: Pay attention to the structural form in the data!)

Part II: Regression with a Mean Centered Predictor

Center the `salary_thousand` predictor by subtracting the mean teacher salary from each value. Call this new variable `center_salary_thousand` This variable should be used in all analyses in Part II.

3. Compute and report (a) the mean centered teacher salary, (b) the standard deviation of centered teacher salaries, and (c) the Pearson correlation coefficient between SAT scores and centered teacher salaries. How do these values compare to the values you computed in the previous assignment using the uncentered teacher salaries?

4. Report and interpret the value of Minnesota's centered teacher salary.

5. Explain how the slope of the uncentered regression you found in the previous assignment will compare to the slope of the regression if we regressed SAT scores on the centered teacher salaries by making reference to the values in the mathematical formula for slope, which is:

$$\hat{\beta}_1 = r_{XY} \times \left(\frac{SD_Y}{SD_X} \right)$$

6. Explain how the intercept of the uncentered regression you found in the previous assignment will compare to the intercept of the regression if we regressed SAT scores on the centered teacher salaries by making reference to the values in the mathematical formula for intercept, which is:

$$\hat{\beta}_0 = \bar{Y} - \beta_1(\bar{X})$$

7. Regress SAT scores on the centered teacher salaries. Write the *fitted equation* using Equation Editor (or some other program that correctly types mathematical expressions). (You should be able to now check your previous response as well.)

Part III: Standardized Regression

Convert the uncentered teacher salaries (`salary_thousand`) into z-scores by subtracting the mean salary and dividing by the standard deviation. Call this new variable `z_salary`. Also convert the SAT scores into z-scores and call that variable `z_sat`. Regress the `z_sat` variable on the `z_salary` variable.

8. Using the mathematical formula for intercept from Question #6, explain why the intercept from regressing z_{Outcome} on $z_{\text{Predictor}}$ will be 0.

9. Using the mathematical formula for slope from Question #5, explain why the slope from regressing z_{Outcome} on $z_{\text{Predictor}}$ will be the correlation coefficient between the predictor and outcome.

10. Compute and report the residual for Minnesota from the fitted standardized regression equation. Explain what the sign and magnitude of the residual value tells you about how Minnesota's mean SAT score compares to the mean SAT score for other states having the same average teacher salary?