EDUC 643 Assignment 03 Key

1. Do schools that educate a larger proportion of students with financial need spend more per student on their education?

1.1. Review the variables you have at your disposal and select a set of substantively sensible continuous covariates that might explain schools’ per-pupil expenditure and help clarify the relationship between *ppe* and *frpl*. (1 point)

There are six additional continuous variables available to us in our NERD$ data: total school enrollment (*enroll*), a composite SES index of families residing within the school district (*sesavgall*), the median income of families residing in the district (*inc50avgall*), the rate of adults in families holding four-year or higher degrees in the district (*baplusavgall*), the unemployment rate for families residing in the district (*unempavgall*), and the rate of SNAP-receipt for families residing within the district (*snapavgall*). Each of these variables could serve to explain per-pupil expenditure in schools in Oregon and would be helpful candidates to help clarify the nature of the relationship between the rate of free- and reduced-price lunch receipt in a school and its expenditure per-pupil.

1.2. Construct a correlation matrix and/or correlation heatmap to assess for potential multicollinearity issues in your selection of covariates and use it to decide which covariates you will select for your postulated model. (2 points)

In Table and Figure 1, we display the Pearson product-moment correlations between our question predictor (*frpl*), our outcome (*ppe*), and a set of candidate covariates that may serve to clarify our relationship of interest. We focus on the extent to which any of the covariates are highly correlated with our question predictor to assess for the potential of multicollinearity which could result in biased coefficients or inflated standard errors. All candidate covariates are only modestly correlated with our question predictor. Only *sesavgall* exceeds a ±0.5 correlation with *frpl*. Although some potential covariates (e.g., *snapavgall* and *sesavgall*) are highly correlated with each other, we are relatively unconcerned with this source of multicollinearity as it will not affect our question predictor. Nevertheless, we can test whether our results are sensitive to the inclusion or exclusion of *sesavgall*.

**Table 1**

Pearson correlation coefficients of the characteristics of Oregon public schools and the communities in which they are located, 2018-19.

|   | ppe | enroll | sesavgall | baplusavgall | unempavgall | snapavgall | frpl | inc50avgall |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ppe | 1 | . | . | . | . | . | . | . |
| enroll | -.078 | 1 | . | . | . | . | . | . |
| sesavgall | -.053 | .137 | 1 | . | . | . | . | . |
| baplusavgall | .025 | .222 | .694 | 1 | . | . | . | . |
| unempavgall | -.016 | -.050 | -.793 | -.457 | 1 | . | . | . |
| snapavgall | .017 | -.081 | -.957 | -.587 | .754 | 1 | . | . |
| frpl | .184 | -.129 | -.529 | -.419 | .352 | .515 | 1 | . |
| inc50avgall | -.132 | .193 | .887 | .599 | -.579 | -.804 | -.475 | 1 |

**Figure 1**

*Pearson correlation heatmap of the characteristics of Oregon public schools and the communities in which they are located, 2018-19.*

**

1.3. Write a formal linear model that describes the relationship between school-level per-pupil expenditure and the school-level average receipt of free- or reduced-price lunch, adjusting for the covariates you have selected (at least one required). Interpret each of the terms in this model. (1 point)

$$PPE\_{j}=β\_{0}+β\_{1}FRPL\_{j}+β\_{2}ENROLL\_{j}+β\_{3}INC50\_{j}+β\_{4}BAPLUS\_{j}+β\_{5}UNEMP\_{j}+ β\_{6}SNAP\_{j}+ε\_{j}$$

*PPEj* is our outcome and represents school-level (*j*) per-pupil expenditure in Oregon schools. *β0*, our intercept, represents the population parameter for predicted per-pupil expenditure in schools with no FRPL-students, that enroll no students, whose families have median income of $0, and whose 4-year degree attainment, unemployment and SNAP-receipt rates are all 0. *β1* represents the relationship between the proportion of FRPL students in a school and its per-pupil expenditure, adjusting for differences in school enrollment, median income, BA-attainment, unemployment and SNAP-receipt. *β2* through *β6* represent the relationship between each of the covariates and per-pupil expenditure, adjusting for all other variables in the model. Finally, *εj* is our school-level idiosyncratic error term or residual.

1.4. State your null hypothesis about the relationship between per-pupil expenditure and the proportion of students receiving free- or reduced-price lunch, accounting for the covariates in your model. (1 point)

Adjusting for schools’ total enrollment, the median income, college-degree attainment, unemployment rate and SNAP receipt rate of district residents, there is no relationship between per-pupil expenditure and the proportion of students receiving free- or reduced-price lunch, on average in the population of Oregon schools.

1.5. Formally test your hypothesis using an Ordinary Least Squares estimation strategy. Report a set of results in a formatted table in which you compare the bivariate relationship between *ppe* and *frpl* with the multivariate relationship you have just estimated. (2 points)

**Table 2**

Taxonomy of Ordinary Least Squares estimates of the relationship between school-level receipt of free- and reduced-price lunch (FRPL) and per-pupil expenditure (PPE) in Oregon, 2018-19

|  | (1) | (2) | (3) | (4) |
| --- | --- | --- | --- | --- |
| (Intercept) | 12,226.90\*\*\* | 12,592.77\*\*\* | 23,860.53\*\*\* | 21,276.08\*\*\* |
|  | (271.68) | (392.08) | (2,802.02) | (2,581.38) |
| Receiving FRPL (0-1) | 2,634.33\*\*\* | 2,533.13\*\*\* | 3,229.26\*\*\* | 3,321.17\*\*\* |
|  | (427.86) | (437.35) | (527.16) | (525.74) |
| Total enrollment |  | -0.66+ | -0.53 | -0.52 |
|  |  | (0.37) | (0.36) | (0.37) |
| Median income ($) |  |  | -0.07 | -0.13\*\*\* |
|  |  |  | (0.04) | (0.03) |
| BA+ holders (0-1) |  |  | 8,244.02\*\*\* | 5,674.43\*\*\* |
|  |  |  | (1,571.49) | (1,039.17) |
| Unemployment rate (0-1) |  |  | -38,341.73+ | -5,857.66 |
|  |  |  | (21,385.20) | (13,990.89) |
| SNAP receipt rate (0-1) |  |  | -49,538.91\*\*\* | -24,256.55\*\* |
|  |  |  | (13,368.27) | (7,835.34) |
| SES composite |  |  | -4,282.51\* |  |
|  |  |  | (1,999.29) |  |
| Num.Obs. | 1193 | 1193 | 1193 | 1193 |
| R2 | 0.034 | 0.037 | 0.096 | 0.091 |
| F-statistic | 41.85 (*df*=1191) | 22.81 (*df*=1190) | 18.02 (*df*=1185) | 19.74 (*df*=1186) |
| *Notes:* + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001. Cells report coefficients and heteroscedastic-robust standard errors in parentheses. Each observation is one school. |

1.6. Interpret the results of your test in 1-2 sentences. (2 points)

In Table 2, we find that the proportion of students receiving FRPL is positively related to the school’s per-pupil expenditure, on average in the population of Oregon schools. In particular, we predict that schools that differ in the proportion of FRPL students enrolled by 10 percentage points will differ in their PPE by between $253 and $323 per-student (*p<*0.001*)*. Our estimates are robust to the inclusion of various covariates related to school- and community-characteristics.

1.7. Assess the quality of your model fit and make appropriate inferences about your overall model, using relevant statistics. (2 points)

Our unconditional model explains roughly 3.4 percent of the variation in PPE. The inclusion of school enrollment increases this explanatory power marginally to 3.7 percent. The addition of the characteristics of families residing in the neighborhoods around schools (such as median income, 4-year degree status, SNAP-receipt rate, unemployment rate, and composite SES score) substantially increases our predictive power and in Model 3, we can explain 10 percent of the outcome variation. In all of our multi-variate regression (Models 2-4), our *F*-statistic exceeds the threshold of statistical significance (at conventional alpha=0.05 levels), and we conclude that these full models explain more of the outcome variation than would be expected as a result of idiosyncratic sampling alone.

1.7. Imagine you are writing a piece for The Oregonian detailing whether schools with greater levels of student financial need receive more money. Create a plot that illustrates the multivariate-adjusted relationship between the proportion of free- and reduced-price lunch recipients and per-pupil expenditure and prototypical values of the other variables in your model. Present the plot and a short paragraph reporting the results of your analysis to your readers **(2 point)**

Oregon schools with greater levels of student need, as measured by the proportion of students receiving meal subsidies, do spend more on their students. A school with 10 percentage points more low-family-income students would, on average, be expected to spend around $325 more per student than another school with fewer low-family-income students. This pattern holds true even when we account for various school and community factors that might also contribute to differences in spending across schools, such as the size of the school and the socio-economic status of the families residing in the community. However, although student family-income levels are related to expenditure, there are many other factors that are also predictive of school spending. For example, as we show in Figure 2, the average education level in a community is also highly predictive of spending levels. In fact, schools located in highly-educated communities and with few low-family-income students, on average, spend more per-pupil than schools located in communities where relatively few people hold four-year degrees, even if many low-family-income students attend the school. Thus, while, on average, schools with higher levels of student need do spend more per student, Oregon’s funding system may bear further scrutiny to ensure that schools across all communities have the sufficient level of resources to educate their students, particularly when student need is more acute in some communities than others.

**Figure 2**

*Relationship between rate of school-level free-and reduced-price lunch (FRPL) receipt and predicted per-pupil expenditure (PPE) at prototypical values of community Bachelor’s and above degree holding rates, adjusting for other school and community factors.*



*Notes:* predictions adjust for school enrollment, median income, SNAP-receipt rate, unemployment rate, and composite SES score and are set to the mean value of these variables in the data