EDUC 641 Assignment 04 Key

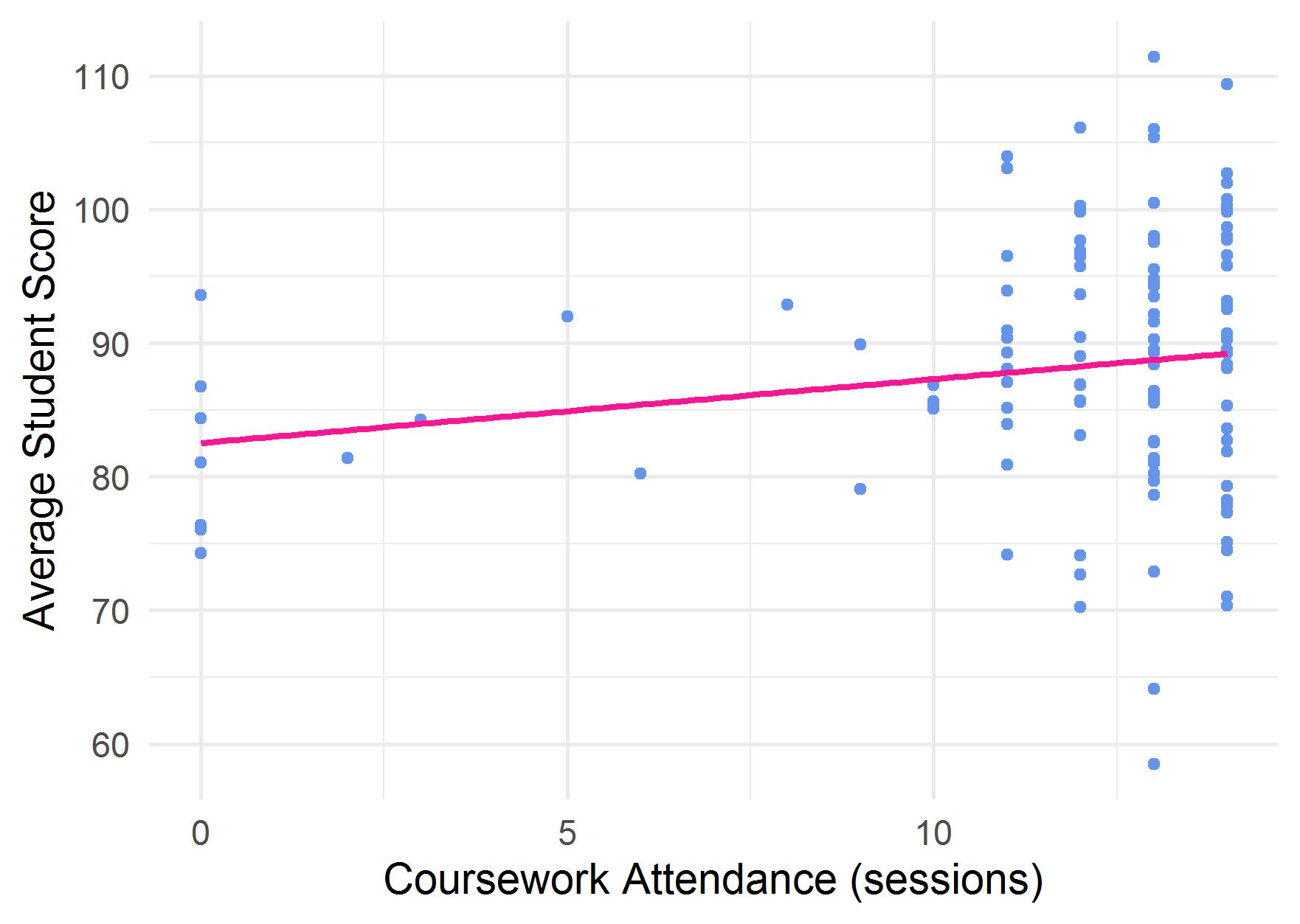
1. Descriptive statistics

1.1. Summarize the dataset. Specifically, create a table for the cont.csv dataset to show the means and standard deviations of the continuous variables and the counts and percentages for each category of the categorical variables. Write 2-3 sentences to report and interpret these statistics. (2 points)

|  |  |
| --- | --- |
|  | *N* = 126 |
| Control | 60 (47.6%) |
| Treatment | 66 (52.4%) |
| Coursework | 11.7 (3.5) |
| Vocabulary score | 88.1 (9.6) |
| *Notes:* Top two rows report counts and proportions in parentheses. Bottom two rows report means and standard deviations in parentheses. | |

There are 126 teacher-level observations in the data; none of the variables have any missing values. 60 teachers (47.6 percent) were in the control group and 66 (52.4 percent) were in the treatment group. On average, teachers attended 11.7 sessions of coursework with a standard deviation of 3.5 sessions, and their students’ average vocabulary score was 88.1, with a teacher-level standard deviation of 9.6 points.

1.2 Create a plot (label the x and y axis) to visualize the relationship between the variable *VOCABULARY* and *COURSEWORK*. Include a line of best fit on this display. (2 points)



***Figure 1.*** Relationship between teacher-level average student vocabulary scores on the PPVT-3 and number of professional development sessions attended

1.3 Write one sentence to interpret this visualized relationship. (1 point)

In Figure 1, on average, it appears that teachers who attended more course sessions have students who scored higher on the PPVT-3 vocabulary test, though there is substantial variation in students’ scores at all levels of coursetaking.

1.4 Can you conclude whether teachers who attended more coursework have higher student score? Why or why not? (1 point)

No, I cannot conclude that teachers who attended more coursework had higher student test scores because the slightly positive relationship in my sample could be due to random sampling idiosyncrasy, and I am not able to conclude that this is statistically different from zero.

#### 2. Research question

2.1. Write your null hypothesis about the relationship between coursework dosage and teachers’ students’ average vocabulary scores. (1 point)

There is no relationship between coursework dosage and teachers’ students’ average vocabulary scores, on average in the population.

2.2 Formally test your hypothesis using an Ordinary Least Squares estimation strategy. Conduct the test and report your results in a formatted table. **(2 points)**

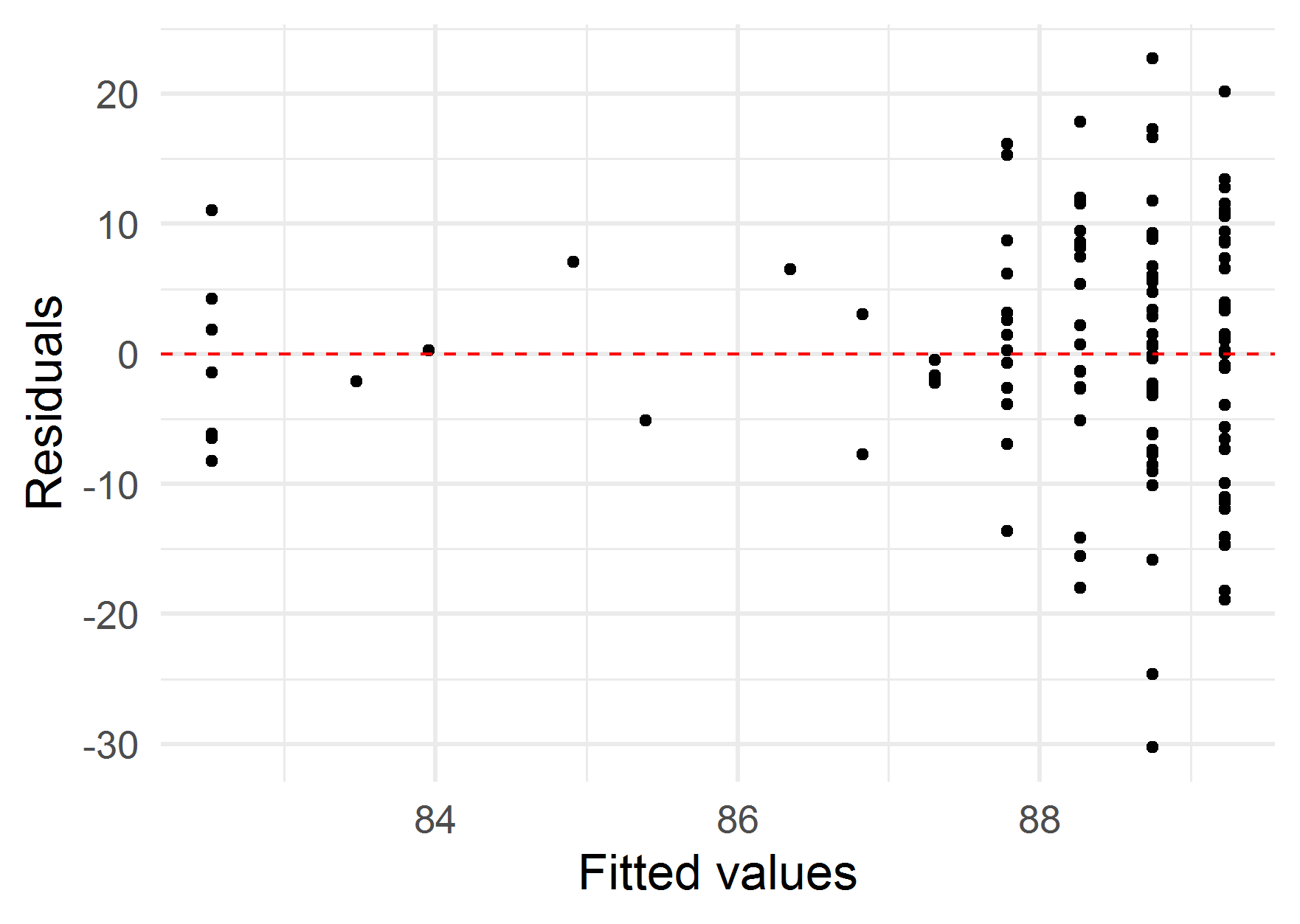
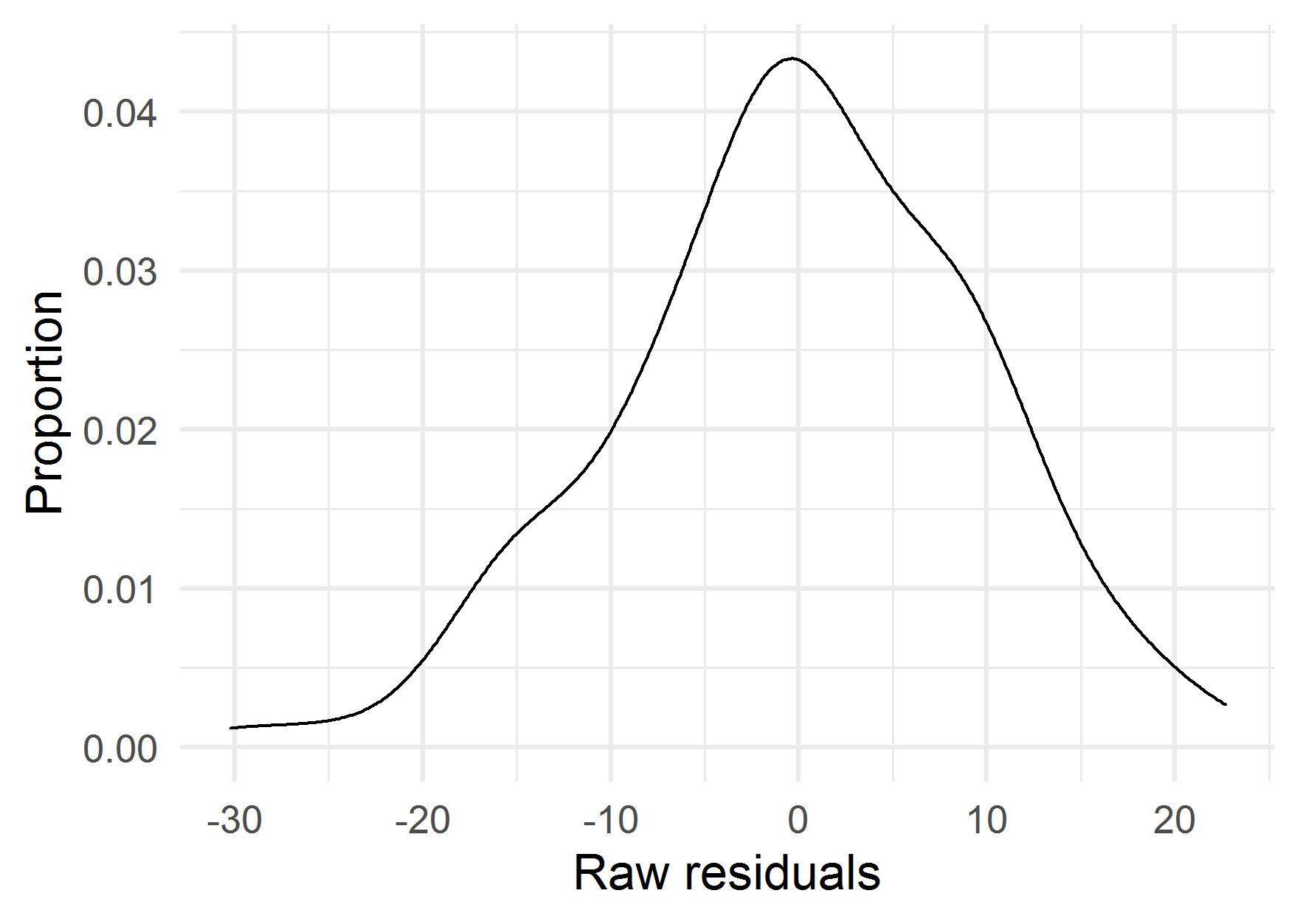
|  | Model 1 |
| --- | --- |
| (Intercept) | 82.520\*\*\* |
|  | (2.990) |
| Coursework Attendance (sessions) | 0.479+ |
|  | (0.245) |
| Num.Obs. | 126 |
| R2 | 0.030 |
| Cells represent coefficients and standard errors in parentheses. | |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | |

2.3 Substantively interpret the results of your test. **(2 point)**

On average, students whose teachers attended one more course session scored 0.479 points higher on the PPVT-3. However, I pre-established an alpha-threshold of 0.05, and the *p*-value on my estimate for coursework attendance exceeds this threshold. Thus, I fail to reject the null hypothesis that there is no relationship between coursework and vocabulary scores, on average in the population.

2.4. What statistical assumption(s) have you made in modeling the relationship between *VOCABULARY* and *COURSEWORK* in this way? Test whether your estimates satisfy these assumptions. **(2 points)**

In modeling the relationship between vocabulary and coursework, I have assumed that the residuals generated from my ordinary-least-squares regression fit are normally distributed. To test this assumption, I plot the density of these residuals in Panel A of Figure 2 and find that that their distribution is approximately normal (mound shaped and roughly symmetrical, with no long tails). To further test this, in Panel B, I plot the residuals against their fitted (predicted) values. I observe that the residuals are symmetrically distributed around the line of best fit at all predicted values. Thus, I conclude that the assumptions I have made hold in this context.



(a) Density plot of residuals (b) Residuals vs. fitted plot

***Figure 2***. Tests of regression assumptions

2.5. What is your answer to this research question?

The number of professional development course sessions that teachers attended is, on average, unrelated to their students’ average scores on a test of vocabulary knowledge. While it is true that in my sample there is a slight positive relationship between coursework and vocabulary scores, I cannot rule out that this relationship is due to idiosyncrasies of sampling (*p* = 0.053) and that no relationship exists in the population.