EDUC 641 Assignment 04 Key

1. Descriptive statistics (40% point)

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.

|  |  |
| --- | --- |
|  | *N* = 126 |
| treat |  |
| 0 | 60 / 126 (48%) |
| 1 | 66 / 126 (52%) |
| coursework | 11.7 (3.5) |
| vocabulary | 88 (10) |

* There are 1,26 teacher-level observations in the data. No missing value on any variables.
* The summary statistics for the categorical variable treat show that there are 60 (48%) teachers in the control group and 66 (52%) in the treatment group.
* The summary statistics for the two continuous variables shou that on average, the sample teachers took 11.7 sessions of coursework with a standard deviation of 3.5 and their average student vocabulary score is 88 with a standard deviation of 10.

1.2. Summarize the outcome variable *vocabulary* by one of the two predictor variables, *treat*. Specifically, create a table and a plot (label the x and y axis) to show the mean and standard deviation of *vocabulary* separately for the treatment and control groups.

|  |  |  |
| --- | --- | --- |
|  | Control, *N* = 60 | Treatment, *N* = 66 |
| vocabulary | 89 (9) | 88 (10) |

Chart, box and whisker chart

Description automatically generated

* 1.2.1. Write 2-3 sentences to report and interpret these statistics.
  + The table shows that on average, the average student score is 89 (*SD* = 9) and 88 (*SD* = 10) for control and treatment groups.
  + The boxplot shows that the distribution of outcome variable (vocabulary) is approximately normal for both groups and the means are about the same. Also, there seem to be a couple of outliers in the treatment group.
* 1.2.2. Compare the mean of *vocabulary* for the consultancy treatment group and control group. Can you decide which is larger? Why or why not?
  + The mean for control group is one score higher than that of the treatment group but we don’t know whether this difference comes from random chance or the real difference between these two groups.

1.3 Create a plot (label the x and y axis) to visualize the relationship between *vocabulary* and another predictor variable *coursework*.

Chart, scatter chart

Description automatically generated

* 1.3.1. Write one sentence describe your observation.
  + There doesn’t seem to be a pattern of average student score across teacher coursework attendance.
* 1.3.2. Can you conclude whether teachers who attended more coursework have higher student score? Why or why not?
  + We cannot.

#### 2. Research question (a) (30% point)

2.1. Write your null hypothesis and the assumption(s) you make in investigating this research question.

* The students in treatment group scored the same on vocabulary test with the students in control group

2.2. Test your hypothesis

* 2.2.1. Which hypothesis testing you choose to answer this question? Why?
  + A two-sample *t*-test to compare the means of treatment and control groups.
* 2.2.2. Run the test then report and interpret your result.
  + On average, treatment group students score about one point less that students in control group but this difference is not significant (*t*(123.96) = 0.481, *p* = 0.631). I failed to reject the null hypothesis.

2.3. What is your answer to this research question?

* + In this random-controlled trial experiment, the consultancy PD intervention did not significantly improve student vocabulary score.

#### 3. Research question (b) (30% point)

3.1. Write your null hypothesis and the assumption(s) you make in investigating this research question.

* There was no relationship between PD intervention dosage and student vocabulary score

3.2. Test your hypothesis

\* 3.2.1. Which hypothesis testing you choose to answer this question? Why?

* I fit an ordinary-least-squares regression of student vocabulary score on PD intervention dosage to see whether the slope of the estimated trend-line is significantly different from zero since zero suggests that the two variables have no relationship in the population.

\* 3.2.2. Run the test then report and interpret your result.

* The *p*-value of the slope was 0.053, which is above the significance level of 5%, suggesting that the relationship between these two variables was not statistically significant. I failed to reject the null hypothesis.

3.3. What is your answer to this research question?

* In this random-controlled trial experiment, the dosage of the coursework PD intervention had no significant impact on student vocabulary score.