EDUC 641 Assignment 04 Key

1. Descriptive statistics (40% point)

1.1. Summarize the cont.csv dataset (leave out the two variables *tchid* and *vocabulary\_random*). Specifically, create a table to show the mean and standard deviation of each continuous variable and the count and percentage 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 126 teacher-level observations in the data. No missing value on any variables.
* The summary statistics of 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 of the two continuous variables show 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 was 88 with a standard deviation of 10.

1.2. Summarize the variable vocabulary by treat. Specifically, create a table to show the mean and standard deviation of vocabulary for the treatment and control groups separately. Also, create a boxplot (make sure to label the x- and y-axes) to visualize the distributions of vocabulary for the treatment and control groups separately (you can either use base R to produce two separate plots or use ggplot2 package to produce one plot).

|  |  |  |
| --- | --- | --- |
|  | Control, *N* = 60 | Treatment, *N* = 66 |
| vocabulary | 88.55 (9.26) | 87.73 (10.02) |

* Two boxplots from base R

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

* Alternatively, one boxplot using ggplot()

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 88.55 (*SD* = 9.26) and 87.73 (*SD* = 10.02) for control and treatment groups.
  + The boxplots show 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 observed mean of *vocabulary* for the consultancy treatment group and control group. Can you decide whether the treated students score higher or lower than their peers in the population? Why or why not?
  + The observed mean for the control group is 88.55, a bit higher than 87.73, the mean for the treatment group.
  + However, we don’t know about the population because we don’t know whether this difference comes from sampling idiosyncrasy or the “true” difference between these two groups.

1.3. Create a plot (make sure to label the x- and y- axes) to visualize the relationship between vocabulary and 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 change in 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 since we didn’t observe any change in student score when the number of coursework increases

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

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

* There was no relationship between consultancy PD intervention and student vocabulary outcome

2.2. Test your hypothesis

* 2.2.1. Which hypothesis testing you choose to answer this question? Why?
  + We should use
* 2.2.2. Run the test then report and interpret your result.

2.3. What is your answer to this research question?

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

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

3.2. Test your hypothesis

* 3.2.1. Which hypothesis testing you choose to answer this question? Why?
* 3.2.2. Run the test then report and interpret your result.

3.3. What is your answer to this research question?

* s not significantly different from the population mean.