Universidade de São Paulo
Faculdade de Filosofia, Letras e Ciências Humanas
Departamento de Ciência Política

FLS-6183
Métodos Quantitativos de Pesquisa II
2º semestre / 2019

Lorena G. Barberia

Lab 1 / Class 2

Review of Hypothesis Tests
Difference of means test vs. coefficient test in a bivariate regression model

Answer Key

Part I. Univariate Hypothesis Test Review

1) Please summarize the hypothesis test (univariate) on y as described in the lab? After running the hypothesis test, what can we conclude?

Ho: y=0.5
Ha: y does not equal 0.5 (two-tailed hypothesis test)

$Pr(|T| > |t|) = 0.0000$

We can reject the null hypothesis that y=0.5 with 99% confidence.

Have students draw the Student t-distribution and illustrate the null hypothesis.

Part II. Bivariate Hypothesis Test Review

1) First, let's carry out a hypothesis test where the dependent variable, y, is a continuous variable (y) which differs between two groups with a difference of means test. Please summarize the result of the difference of means hypothesis test on y? After running the hypothesis test, is there a difference in the mean conditional on x?

$Ho : x_1 - x_2 = 0$
$Ha : x_1 - x_2 \neq 0$

$Pr(|T| > |t|) = 0.0000$

2) Next, let's carry out the same test using a bivariate regression model. Please summarize the result of the difference of means hypothesis test on y? After running the hypothesis test, is there a difference in the mean conditional on x?

$$y = 0.84 + 2.25x + \varepsilon$$

If we look at the results of the null hypothesis that b=0, we can reject the null with 99% confidence.

3) Do you think there is a preferred method in this particular case for these type of data (y= continuous and x=dummy)?

Both methods arrive at the same result. However, there are some important differences.

4) How is uncertainty captured by both methods?

Both tests captured the uncertainty of the coefficients by measuring t-statistics, p-value and the confidences interval. These measures help us understanding how the values obtained by our sample relate to the populational ones, allowing conclusions about the statistical significance of the values.

5) Are there any advantages of using a regression in this case? Please explain.

Since X is a categorical variable, the difference of means test is sufficient to verify the relation between the two variables. However, the advantage of the regression model is that it also captures the stochastic variation in Y. In other words, it captures how much Y varies based on changes in X and due to random variation. If X was a continuous variable, the linear regression model would provide us with more information and thus would be preferable.