Universidade de São Paulo Faculdade de Filosofia, Letras e Ciências Humanas

Departamento de Ciência Política

FLS-6183 Métodos Quantitativos de Pesquisa II

2º semestre / 2017

Lorena G. Barberia

Lista 1

Answer Key

Data de entrega: 31/8/2017

PART 1. Review of Types of Variables, Descriptive Statistics, Hypothesis Testing and the Normal Distribution

1. **For the cases below, please fill in the table with the descriptive statistics and graphs that would be most appropriate for each situation.**

Variables:	**Measures of Central Tendency and Dispersion**	**Types of Graphs**
Continuous	Mean Standard Deviation	Histogram, density plot, boxplot
Nominal	Mode Range of values Frequencies and Percentages	Discrete histogram
Ordinal	Mode Range of Values Frequencies and Percentages	Discrete histogram

2. **Classify the following variables as nominal, ordinal or continuous.**

Variable	**Answer**
GDP Growth (%)	Continuous
Presidential Approval (Yes or No)	Nominal
Presidential Approval (%)	Continuous
Age	Ordinal
Religion (Catholic, Protestant, etc.)	Nominal
Race (White, Black, etc.)	Nominal
Percent of Asian population in a certain country	Continuous

3. For the following types of cases, which are the types of statistical tests that are most appropriate?

Variables:	**Null Hypothesis**	**Statistical test:**
Continuous (y)	Ho: y=0	Univariate Mean test
Two categorical variables: x1 and x2 **(E.g.** Do men and women differ in their political party preferences (Republican or Democrat in a specific city?).	Ho: Men and women do not differ in their political preferences HA: Men and women do differ in their political preferences.	Chi-square
A continuous and a categorical variable (E.g. Does the percentage of Democrats differ in two cities?)	Ho: z1 –z2=0	Difference of Means Test (t-distribution) with equal or unequal variances
Two continuous variables (y and v).(E.g. Are increases in x associated with increases in y?)	$H_o : \frac{\Delta y}{\Delta x} \rangle 0$	Correlation Regression

4. Normal distribution and Student´s T-distribution. Please respond whether each sentence below is true or false. Justify your response if false.

Sentence:	**True or False?**
a) In a normal distribution, the mean and the median are equal (μ), but they differ from the mode.	F. They are all the same.
b) If the distribution is normal, at least 99% of the observations must be in the range [μ-3σ, μ+3σ]. When we have 2σ, 95%. For only one σ, 68%.	T.
c) If a distribution is skewed, it's not a normal distribution.	T.
d)The student´s t-distribution tends to a normal as the number of observations decreases	F. Tends as n increases.

PART 2. Correlation and Linear Regression

4. Please review the Lab 2 do-file. What is the null and alternative hypothesis being tested in the correlation test between x and y? In the specific case, what can we conclude?

In this case, the null hypothesis is that there is no correlation between x and y. The alternative hypothesis is that there is a correlation between X and Y. The p-value is less than 1% (our cutoff). The null hypothesis is rejected with 99% confidence.

5. What is the interpretation of the coefficient and the constant term in the following regression model y= $a + bx$.

   The coefficient tells us that for every one unit increase in x, y will increase by b.

   The constant (or intercept) tells us the expected value of y when x is equal to zero. We can think of the constant as the counterfactual in this bivariate regression.

6. Under what situations would we prefer a hypothesis of correlation between two variables as opposed to a regression model?

A correlation coefficient (partial correlation) may be useful when studying relationships involving more than two variables. In addition, there are instances where the correlation coefficient can help us to understand the relative strength of relationships across observations. For example, in time series analysis, we often examine if there is correlation across time.