

Homework Assignment 6: Applied Probabilistic Models

Statistical Tests

5273

1 Introduction

For this work, data is collected on the official website of Instituto Nacional de Estadística y Geografía (INEGI) [3]. The chosen section is Growth Domestic Product, and within this section, a national macroeconomic indicator is selected: Global Index of Economic Activity (IGAE) for the data analysis. Data obtained from INEGI website are in `csv` format, edited in order to work with the general values of the three main representative activities each month, which the aforementioned indicator is based on.

For the analysis it is used the R software in its version 4.0.2 [7] and the code used is available on the GitHub repository of [4]. This work is run on a MacBook Air with an Intel Core i5 CPU @ 1.8 GHz and 8 GB RAM.

2 Theoretical Aspects

This section covers several aspects of statistical and tests and when it is appropriate to use them and their interpretation. Characteristics of hypothesis tests are treated as well.

2.1 Hypothesis and statistical tests

A statistical test is a form of evaluating the evidence that data provide. The objective is to prove a hypothesis which is denominated as the null hypothesis (H_0). Typically, H_0 establishes equality (between means, variances, or correlation coefficient to cite some examples). H_0 usually opposed to a hypothesis denominate alternative (H_1) and implies difference (between means, for example).

2.2 Rejecting the null hypothesis

If data do not provide sufficient evidence against H_0 , it is not rejected. If, contrary, show strong evidence against H_0 it is rejected, and H_1 it is considered as true with a quantified risk (low) of being wrong.

2.3 Interpretation of a statistical test

The statistical test produces a number denominated p -value, which limits are 0 and 1. The p -value is the probability of obtaining data or extreme data below the null hypothesis.

In more practical terms, p -value should be compared with alpha (see subsection 2.4) and it can be interpreted follow:

- If $p < \alpha$, H_0 is rejected, and H_1 is accepted with a proportional risk of the p -value of being wrong.
- If $p > \alpha$, H_0 is not rejected, but this does not imply it is necessary to be accepted. It means H_0 is true, but the experiment and the statistical test are not “strong” enough to produce a p -value less than alpha.

2.4 Selecting alpha

When a study is designed, a risk threshold must be specified, above which H_0 should not be rejected. This threshold is known as significance level, also denoted as alpha or α , and should be between 0 and 1. Low alpha values are more conservative [9].

The selection of alpha should depend on how dangerous is reject H_0 in case it is true. For example, in a study to prove the benefits of medical treatment, alpha should be low. On the other hand, when revising effects of many attributes in the appreciation of a product, alpha could be moderate. In most cases, alpha is set at 0.05, 0.01, or 0.001.

2.5 Common mistakes when interpreting p -value

A mistake that can be made when interpreting p -value is if a study is realized two times and $p \leq 0.05$ in one case and $p > 0.05$ in another, its results are in conflict. This is a mistake to assume because most of the experimental series should throw some null result. Another one is when $p \leq 0.05$ assures that 95% of effects will be true. The true probability that a significative effect reflects the existence of a real effect is given by the Positive Predictive Value (PPV), and it depends on the statistical power, plausibility of alternative hypotheses, and other questionable research practices. Another one is that a $p > 0.05$ implies there is no effect. This may be, but the truth is that there is no power enough to detect it. The other one is that $p \leq 0.05$ implies a found effect. The valid use of p -value is to control the rate of error type 1. The p -value does not say anything about concrete cases (current experiment) it only provides a general rule of behaviour [5].

2.6 Statistical power

When various statistical tests are available to be used, it is critical to choose the most convenient and powerful for the experimental design to be used. For that reason, it is necessary a selection criterion [2]. One of them is the statistical power which is the probability of rejecting the null hypothesis when it is false. Therefore, it can be considered a statistical test as appropriate when the probability of rejecting the null hypothesis (H_0) is small when is true; and big enough the probability of rejecting it when is false.

2.7 Parametric and non-parametric statistical tests

Parametric tests are for numerical data and, in general, are based on the normal or gaussian distribution. Possible tests to apply are t -student, the Pearson correlation coefficient, linear regression,

Table 1: Map to select the appropriate test

		<i>Data type</i>			
		Numeric (gaussian)	Ordinal or numeric (non gaussian)	Numeric (outliers)	Nominal (binary)
<i>Objective</i>	Compare 2 independent groups	<i>t</i> -student test for 2 independent samples	Mann-Whitney Test	Yuen test for independent samples	Fisher or Chi-Squared (for bigger samples)
	Compare 2 related groups	<i>t</i> -student test for 2 related samples	Wilcoxon Test for related samples	Yuen Test for related samples	McNemar Test
	Compare 3 or more independent groups	ANOVA one-way for independent samples	Kruskall-Wallis Test	Robust ANOVA one-way for independent smples	Chi-Squared Test
	Compare 3 or more related groups	ANOVA one-way for related samples	Friedman Test	Robust ANOVA one-way for related samples	<i>Q</i> of Cochran Test
	Associate 2 variables	Pearson correlation	Spearman or Kendall correlation	Robust correlation	Cramer <i>V</i> coefficient

one-way Analysis of variance (One-way ANOVA), factorial analysis of variance (ANOVA), and Analysis of covariance (ANCOVA). Descriptive statistics such as standard deviation, mode, median and mean are used.

On the other hand, non-parametric tests are used with nominal and ordinal variables. They do not assume a particular distribution and the requirements regarding sample size are not as many as the parametric ones. Most used tests are Chi-Squared, Correlation and independency coefficients for cross-tabulation and Spearman's and Kendall's rank correlation coefficient.

2.8 Guide to find the needed statistical test

To begin with, the objective of the analysis needs to be clear. It could be about association or comparison. Both seek to establish relations (similarities or differences) between elements, but tests of comparison, in contrast with a test of association, evaluate relations between one or various groups. The type of variables has to be noticed (numeric or nominal) as well. Then, it needs to be distinguished if samples are independent or related. The next step corresponds to the fulfilment of the classical assumptions (Normality, homogeneity of variance, independency), this allows to choose between parametric, non-parametric and robust tests. Then, it can be chosen the appropriate test with the help of Table 1. Once the test is selected, the next step is to perform hypothesis testing, and finally it corresponds to the interpretation and if possible graph the results.

2.9 Assumptions to apply parametric techniques

To apply parametric techniques the following assumptions are needed:

- Observations are independent of each other.
- Populations should be normally distributed.
- These populations must have the same variance.
- Variables must be measured on at least an interval scale so that arithmetic operations can be used.

3 Application of Statistical Tests

In this section, several of the most used statistical tests are performed. These tests are applied considering data of the IGAE obtained from the INEGI website. The IGAE is an index that approximates the calculation of the generated wealth in the country monthly. It is considered a trend index and marks the path that the national economic activity is reporting in the given month [8]. The method used to calculate the IGAE consists of monthly indexes of the physical volume of production for each of the selected classes, with a fixed base in 2013 [1]. In the calculation of the IGAE, three fundamental areas or activities are involved: primary, secondary, and tertiary which are the ones analyzed in this work.

3.1 One Sample t -Test

This parametric test is used to test if the mean of a sample from a normal distribution could reasonably be a specific value [6]. In this case, data belong to the primary activities. The p -value is 0.4594, so the null hypothesis that the mean=90 cannot be rejected.

data.txt

```
One Sample t-test

data: pri
t = 0.74072, df = 329, p-value = 0.4594
alternative hypothesis: true mean is not equal to 90
95 percent confidence interval:
 88.54626 93.20969
sample estimates:
mean of x
 90.87798
```

3.2 Wilcoxon Signed Rank Test

This test can be considered as an alternative to t -Test, especially when the data sample is not assumed to follow a normal distribution and it is a non-parametric method [6]. For this test, data belong to secondary activities. The p -value is less than the significance level of 0.05, so the null hypothesis is rejected and accept the alternative that true mean is not equal to 90.

data.txt

```
Wilcoxon signed rank test with continuity correction

data: as.numericsec
V = 33775, p-value = 0.0001926
alternative hypothesis: true location is not equal to 90
95 percent confidence interval:
 91.41908 94.14020
sample estimates:
pseudomedian
 92.88731
```

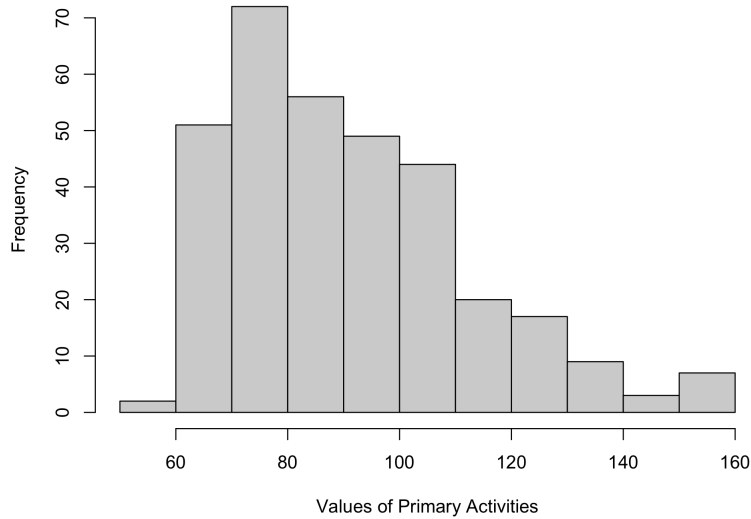


Figure 1: Histogram of the values of Primary Activiites

3.3 Two Sample *t*-Test and Wilcoxon Rank Sum Test

Here both *t*-Test and Wilcoxon rank test are used to compare the mean of 2 samples, passing the numeric vectors of primary and secondary activities. The *p*-value is 0.9976, so the null hypothesis cannot be rejected.

data.txt

Wilcoxon rank sum test with continuity correction

```
data: as.numericpri and as.numericsec
W = 47533, p-value = 0.9976
alternative hypothesis: true location shift is greater than 0
```

3.4 Shapiro Test

This is used to test if a sample follows a normal distribution. Data belong to primary activities. In this case, *p*-value is 3.29×10^{-10} , so the null hypothesis is rejected, which means data are not normally distributed. Figure 1 shows a histogram of the values reached by primary activities.

data.txt

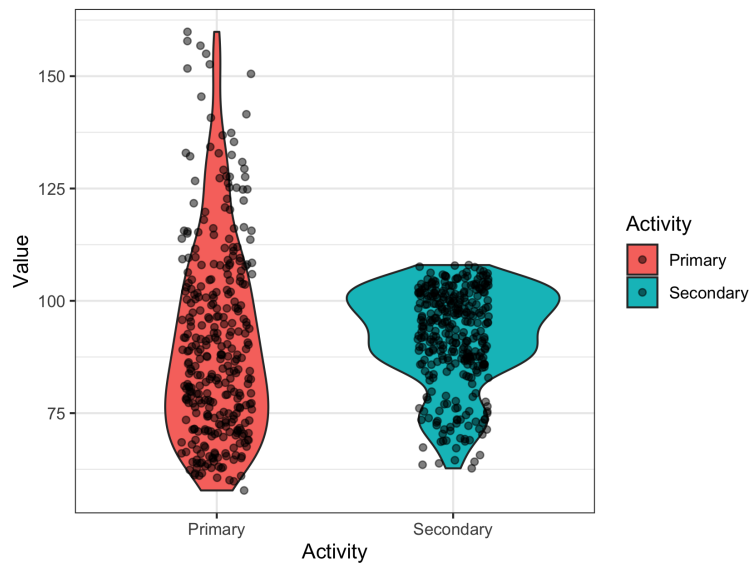


Figure 2: Violinplot of Primary and Secondary Activities

Shapiro-Wilk normality test

```
data: as.numericpri
W = 0.94082, p-value = 3.294e-10
```

3.5 Kolmogorov-Smirnov Test

This test is used to check whether two samples follow the same distribution. In this case, data belong to primary and secondary activities; the p -value is 3.70×10^{-10} , so the null hypothesis is rejected, which means data do not come from the same distribution. A violin plot is generated in Figure 2 to show how both groups are distributed.

data.txt

Two-sample Kolmogorov-Smirnov test

```
data: as.numericpri and as.numericsec
D = 0.26061, p-value = 3.695e-10
alternative hypothesis: two-sided
```

3.6 Fisher's F -Test

Fisher's F -test can be used to check if two samples have the same variance [6]. For this case, data of secondary and tertiary activities are taken as samples. The p -value is 2.20×10^{-16} , so the null hypothesis is rejected, which means the two samples do not have the same variance.

data.txt

```
F test to compare two variances

data:  as.numericsec and as.numericter
F = 0.341, num df = 329, denom df = 329, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2746199 0.4234194
sample estimates:
ratio of variances
 0.3409977
```

3.7 Chi-Squared Test

It can be used to test if two categorical variables are dependent, by means of a contingency table [6]. In this case, the experiment is performed to evaluate if the results of the level of activities that conform to the IGAE is independent to the months in the year 1993.

For this test, there are two ways to tell if the variables are independent. The first one is by looking at the p -value, which in this case is 0.9931, as is greater than 0.05, the null hypothesis that the two variables are independent can not be rejected. Also, if look at the calculated Chi-Squared value, it is 9.0381, which is less than 33.9244 (critical value), so it also indicates the null hypothesis can not be rejected. In conclusion, as the two approaches lead to the same conclusion, it can be said there is no evidence to reject that the two variables are independent.

data.txt

```
Pearson's Chi-squared test

data:  data_trimmed
X-squared = 9.0381, df = 22, p-value = 0.9931

Critical Value:  33.92444
```

3.8 Correlation

Correlation is used to test the linear relationship of two continuous variables. In this case, data come from primary and secondary activities. The p -value is 2.20×10^{-16} , which indicates the null hypothesis that there is a true correlation between the two variables is rejected. Figure 3 a scatter plot is generated in order to show how data is distributed.

data.txt

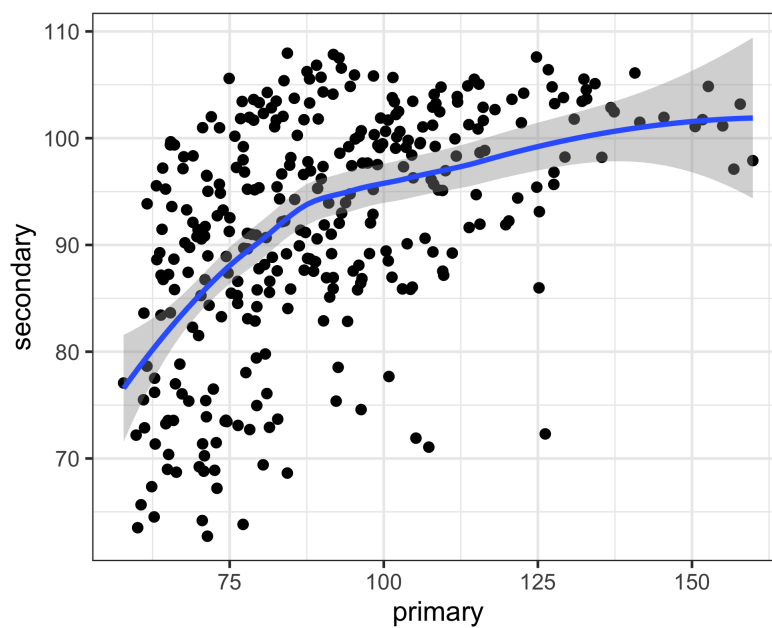


Figure 3: Scatterplot of Primary and Secondary Activities

```

Pearson's product-moment correlation

data: as.numericpri and as.numericsec
t = 10.528, df = 328, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4172422 0.5791088
sample estimates:
      cor 
0.5025669

```

References

- [1] Instituto Nacional de Estadística y Geografía. Sistema de cuentas nacionales de México. Fuentes y metodologías. año base 2013. Indicador Global de la Actividad Económica, 2017. https://www.inegi.org.mx/contenidos/programas/igae/2013/metodologias/SCNM_Metodo_IGAE_B2013.pdf, Last accessed on 2020-09-05.
- [2] EcuRed. Pruebas estadísticas, 2019. https://www.ecured.cu/Pruebas_estad%C3%ADsticas, Last accessed on 2020-10-07.
- [3] INEGI. Datos, 2020. <https://www.inegi.org.mx/temas/igae/>, Last accessed on 2020-09-05.
- [4] Oscar Alejandro Hernandez Lopez. Probability in R. <https://github.com/oscaralejandro1907/probability-in-R/blob/master/assignment1/t1.R>, 2020.
- [5] José C. Perales. Errores comunes en la interpretación del valor p , 2018. <https://masalladelpvalor.wordpress.com/2018/12/04/unidad-3-errores-comunes-en-la-interpretacion-del-valor-p/>, Last accessed on 2020-10-08.
- [6] Selva Prabhakaran. Statistical tests, 2016. <http://r-statistics.co/Statistical-Tests-in-R.html>, Last accessed on 2020-10-11.
- [7] The R Foundation. The R Project for Statistical Computing. <https://www.r-project.org/>, 2020.
- [8] Eduardo Torreblanca. ¿En qué consiste el IGAE?, 2016. <https://mvsnoticias.com/noticias/economia/opinion-en-que-consiste-el-igae-610/>, Last accessed on 2020-09-05.
- [9] XLSTAT. ¿Qué es una prueba estadística?, 2019. <https://help.xlstat.com/s/article/que-es-una-prueba-estadistica?language=es#:~:text=Una%20prueba%20estad%C3%ADstica%20es%20una,nula%2C%20y%20suele%20denominarse%20H0.&text=H0%20normalmente%20se%20opone%20a,alternativa%2C%20denominada%20H1%20o%20Ha>, Last accessed on 2020-10-07.