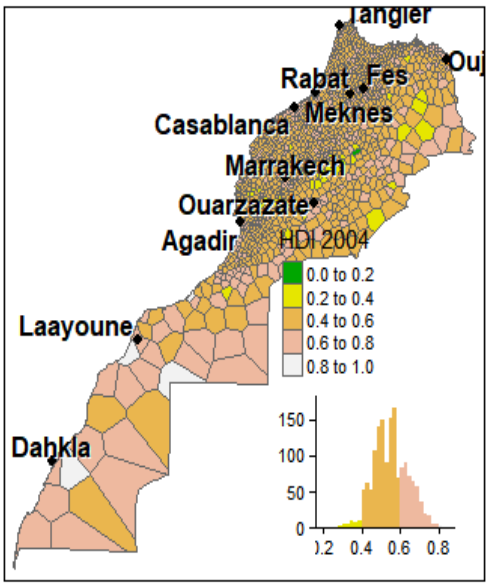
STATISTICAL REPORT

Statistical Analysis and Inference about HDI in Morocco

Hypothesis testing, Visualizations, Maps (LISA) and ANOVA analysis using R-studio.

(2004 and 2014 Excel official data)



Map using R-studio of HDI in 2004 of Morocco by communes

جامعة الأخوين
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵏ ⵉⵅⴰⵡⴰⵢⵏ
AL AKHAWAYN UNIVERSITY

Written by Yassine Fahim

Outline:

Abstract

I. Introduction

1. *What is HDI and why is it important?*
2. *What is our study area and why does it matter?*
3. *The data used in the study:*

II. Analysis

1. *Methodology:*
   - Descriptive statistics:
   - Definition of statistical tests used in the study:
   - Statistical analysis:
2. *Equations*
   - Hypothesis testing:
   - LISA clusters:
3. *Visualizations*

III. Conclusion about HDI in Morocco

IV. References

Abstract:

This report is a study about the Human Development Index (HDI) in 2004 and 2014 of Morocco and the inferences done after analyzing data in the framework of the quantitative methods class in Al Akhawayn University in Ifrane, Morocco during summer 2020. The statistical methods applied are hypothesis testing (t distribution testing, F distribution testing, ANOVA…). All these methods were implemented using R studio which is a powerful statistical tool that provides a wide range of statistical methods. The hypothesis testing was mainly about HDI. In addition, this study provides maps and various visualizations about HDI in Morocco, such as LISA (Local Indicators of Spatial Association) which shows different clusters of HDI in Morocco. The HDI is an index established by the United Nations Development Program (UNDP) that assesses the development of a country in many ways. The parameters included in this index are income, life expectancy, and knowledge. These dimensions were chosen by the United nations as the most important parameters to evaluate the development of a country and the well-being of individuals. The conclusions made in this study concerns only HDI and its dimensions; for a deeper analysis, one can analyze other indexes and more data to conclude and reflect about other parameters such as empowerment, poverty, political membership, gender inequalities… etc.

I. Introduction:

1. What is HDI and why is it important?

In 1990, the united nations development program established their first human development index trough an annual report (Stanton, 2007). HDI is an index that illustrates the human development level of a country. It takes into consideration the basic needs of an individual which are the access to health, knowledge, and goods (Stanton, 2007). United nations development program officials developed HDI by measuring important indexes computed from annual data and combined them into one index which is HDI. Those indexes reflect the three dimensions of HDI: health, knowledge, and income (Stanton, 2007). The health dimension is illustrated by the life expectancy at birth which is the life expectancy index. Knowledge is translated to the education index by measuring the expected years of schooling and the mean years of schooling. The income dimension is translated to the gross national income index by measuring the gross national income per capita. All these indexes combine in the HDI index.

HDI is important as it shows the human well-being in a certain country, region, or commune. We can compare different years data to see the change of human development of an area.

Analyzing HDI can be used to make important economic, or social conclusions about an area (Stanton, 2007).

2. What is our study area and why does it matter?

The area chosen in this study is Morocco. The data sample studied concerns 1528 communes in 12 different regions in Morocco. We are in a Moroccan university, so choosing Morocco as a target region is interesting. Analyzing and explaining various results in an accurate way would require a certain degree of acquaintance with the society and its behavioral aspects. Investigating in a known area can lead to interesting and accurate interpretations, yet statistical results are always tricky and very interesting without taking into consideration the fact that we have some knowledge about the studied area.

3. The data used in the study:

In this study, we used two different types of data, excel tables (.csv) and shapefile format (.shp) that are imported to R studio. The excel tables we used are data of different observations from Moroccan communes made in 2004 and 2014, the last two years where the Moroccan governmental statistics institution HCP ("Haut Commissariat au Plan") performed a national surveying and data collection campaign ("BDS", 2020). The observations are made for 1528 communes in 12 different regions of Morocco which is basically the sample we used to

perform our statistical analysis; each commune has a specific number called geo ID.

The observations in excel tables are the columns; each column representing a specific observation and each row representing a commune. The observations are:

- A random number assigned to each region
- The region of the commune
- The name of the commune
- Total population of the commune 2004
- Total population of the commune 2014
- The percentage of people that have four or less years old in 2004
- The percentage of people that have four or less years old in 2014
- The percentage of people that have ten to fourteen years old in 2004
- The percentage of people that have ten to fourteen years old in 2014
- The percentage of people that have between five to nine years 2004

- The percentage of people that have between five to nine years 2014
- The percentage of people who have ever got in a divorce in 2004
- The percentage of people who have ever got in a divorce in 2004
- illiteracy rate of the commune in 2004
- illiteracy rate of the commune in 2014
- The percentage of people who were economically active in 2004
- The percentage of people who were economically active in 2014
- The percentage of people who have a cellphone in 2004
- The percentage of people who have a cellphone in 2014
- The percentage of people who rent a house in 2004
- The percentage of people who rent a house in 2014
- The percentage of people who have electricity 2004
- The percentage of people who have electricity 2014
- Being part of the Moroccan Sahara (1) or not (0)
- Being 100 Km of the Algerian border (1) or not (0)
- The human development index of the commune 2004
- The human development index of the commune 2014
- The region of the commune

In addition, there're the shapefile format files which are geospatial vector data for geographic information system software. We used these files in R studio to design various maps that illustrated the distribution of HDI in Moroccan communes and regions.

V. Analysis

1. Methodology:

- Descriptive statistics:

In the study, we performed various statistical methods to establish connections and to explain the distribution of HDI values in Morocco. We started our study with descriptive statistics. Descriptive statistics are a way to summarize data, it can be tabular, graphical, or numerical (Anderson et al., n.d.). It consists of computing the HDI's sample mean in 20004, which was around 0.555, sample variance (0.008), sample standard deviation (0.09), sample co-variance with illiteracy (-1.21), which indicates a negative relationship with illiteracy, and sample correlation with illiteracy (-0.88), which indicates a strong negative relationship.

The results of the descriptive statistics can be summarized using R studio functions ("summary ()").

- Definition of statistical tests used in the study:

The next step of the analysis is to perform hypothesis testing, which is statistical inference. The procedure where we infer a population statistic through the sample statistic values by making two competing hypotheses, the null and the alternative (Anderson et al., n.d.). By rejecting the null, we

are sure that the null scenario is impossible within a significance level (alpha). The mean and the variance of the sample computed in the descriptive statistics part were used in the hypothesis testing analysis to make inference about the population statistics by performing tests of different distributions such as the Welch's and Students' t distribution test, Chi² distribution test, Fisher distribution test, and ANOVA.

- *The Student's t distribution test:*

The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. It can be used in a one-sample location test of whether the mean of a population is equal to a value specified in a null hypothesis (Sealey Gosset's, 1908).

- *The Welch's t distribution test:*

Welch's t-test, or unequal variances t-test, is a two-sample test which is used to test the hypothesis that two populations have equal means. It is named for its creator, Bernard Lewis Welch, and is an adaptation of Student's t-test (B.L. Welch,1947). It is more reliable when the two samples have unequal variances and/or unequal sample sizes. In addition, we assume that the two sample are independent and are normally distributed (B.L. Welch,1947).

- *The chi² distribution test:*

The Chi Square statistic can be used to test relationships between categorical variables. The null hypothesis of the χ2 test is that no relationship exists between the categorical variables in the population; they are independent (Pearson, 1900). In this study, we used χ2 to compare the variance of one sample to a known value. Test statistics that follow a χ2 distribution occur when the observations are independent and normally distributed (Pearson, 1900).

- *The F distribution test:*

It is called the F distribution, named after Sir Ronald Fisher, an English statistician (Rao. C, 1992). The F statistic is a ratio of the variance of one sample over the variance of another one. It is performed to test whether a sample has a greater variability than another one due to a different treatment. There are two sets of degrees of freedom; one for the numerator and one for the denominator (Rao. C, 1992).

- *ANOVA:*

Created by Ronald Fisher, the analysis of variance (ANOVA) can be used as an exploratory tool to explain observations (Rao. C, 1992). It is performed using F test ratio: the variance between

the samples over the variance within the samples. Anova is performed between different samples to see if the variability comes from the samples or from the treatment that differs a sample from another one. In addition, we assume that the samples have common variances and are independent and normally distributed (Rao. C, 1992).

- Statistical analysis:

The first hypothesis test that we performed was a one sample test checking if the Moroccan population mean HDI in 2004 was different than the global mean HDI in 2004 (0.65). We performed a student's t distribution test, as the global mean HDI is known, with a significance level alpha of 5%. The null hypothesis was that the Moroccan population mean HDI in 2004 equals the global mean HDI in 2004, and the alternative stated that they are different. We draw the conclusions based on three approaches: The p value, which is smaller than alpha thus we reject the null hypothesis, The confidence interval, which is an interval we built using the margin of error: we found that the value of Moroccan population mean HDI in 2004 was between 0.55 and 0.56 (95% sure). Finally, we used the t statistic value (-40) that is smaller than the t value of -alpha/2 (-1.96), thus we reject the null hypothesis.

The second hypothesis test was a two samples test comparing between the Moroccan population means HDI in 2004 and 2014.

The null hypothesis was that the two populations means are equal, and the alternative hypothesis stated that they are different. We performed a welch t distribution test which yielded with a p value smaller than alpha, and a confidence interval that excluded zero ($0.77 < mean\_04 - mean\_14 < 0.88$), and a t statistic higher than t alpha/2, thus we rejected the null hypothesis which means that the two populations means are different.

All these steps were performed in R studio using various functions such as "t.test ()", or by building our own model using the statistics tests formulas.

The third hypothesis test that we performed was the Chi^2 test. We tested whether the variability of the Moroccan HDI in 2004 was different than the global variability of HDI in 2004.

We assumed a known value (0.0275) for the global variability of HDI in 2004 that we compared to the Moroccan population variance of HDI in 2004. This value was computed from the HDI of all the countries in the world in 2004. We found that the p value is smaller than alpha and that the variability of Moroccan HDI in 2004 was 95% sure to be between 0.0077 and 0.0088 which does not include 0.0275, thus we rejected the null hypothesis. So, the Moroccan population variance of HDI in 2004 was smaller than the global variability of HDI in 2004.

The fourth test was a Fisher distribution test, as we have two unknown population variances. We compared the Moroccan variance of HDI in 2004 and 2014. We found that the p value was smaller than alpha and that the ratio of the 2004 HDI variance over the 2014 HDI variance was 95% sure to be between 1.88 and 2.29, which excludes 1, thus we reject the null hypothesis. So, the 2004 HDI variance is higher than the 2014 HDI variance in Morocco.

After that we introduced the region factor in our analysis. As stated before, the data set divides Moroccan communes into the 12 Moroccan regions. Basically, we divided our big sample into 12 subsamples.

The following test was the ANOVA test; we performed it to see whether the treatment (region) influences HDI in 2004 of Morocco. The null hypothesis was that the 12 populations' means are equal, the alternative states that at least two populations means are different. First, we arranged the data in R studion by regions, then we assumed that the variances of the populations are equal. We used this function: "aov (HDI_04 ~ region, data_shape)" to perform ANOVA.

We found out that the p value is smaller than alpha thus we reject the null hypothesis. So, at least two population means are different.

After further investigation, we found that the assumption we made is not correct, as there's a slight difference between the populations variances, thus we used another ANOVA, a one-way ANOVA that don't assume equal population variances: we found the same result, we rejected the null hypothesis.

Moreover, to test more about the effect of the region treatment on HDI we performed a pairwise test, which is basically a t test between all possible pairs of regions. The results are stored in a (12x12) table, each cell with the p value of the test. In the table, all the p values will be doubled because of the 12x12 arrangement, as there're repetitive pairs.

	1	2	3	4	5	6	7	8	9	10	11	12
Beni Mellal-Khenifra	0.0000000000	0.002211724	2.993120e-01	2.066289e-01	3.693260e-02	1.262711e-03	1.893923e-04	2.175939e-01	8.139811e-02	2.369256e-04	4.838891e-01	2.453673e-01
Dakhla-Oued Eddahab	0.0022117244	0.000000000	1.575111e-03	3.160179e-03	3.844796e-03	5.320162e-02	1.087294e-03	3.005170e-03	3.401985e-03	6.839036e-03	2.908376e-03	3.005170e-03
Draa-Tafilalt	0.2993119837	0.001575111	0.000000e+00	9.709183e-03	4.020953e-03	2.939964e-04	8.061748e-03	1.928407e-02	3.844796e-03	1.700076e-06	5.253410e-02	2.201198e-02
Fez-Meknes	0.2066288879	0.003160179	9.709183e-03	0.000000e+00	1.923332e-01	3.401985e-03	1.442809e-11	8.502096e-01	5.107560e-01	2.211724e-03	5.015601e-01	9.560550e-01
Guelmim-Oued Noun	0.0369326000	0.003844796	4.020953e-03	1.923332e-01	0.000000e+00	3.742133e-02	3.959726e-06	2.802834e-01	3.899232e-01	4.282261e-01	8.522638e-02	2.289278e-01
Laayoune Sakia El Hamra	0.0012627110	0.053201618	2.939964e-04	3.401985e-03	3.742133e-02	0.000000e+00	1.069031e-05	4.614578e-03	6.422742e-03	8.580813e-02	2.211724e-03	3.844796e-03
Marrakech Safi	0.0001893923	0.001087294	8.061748e-03	1.442809e-11	3.959726e-06	1.069031e-05	0.000000e+00	1.971217e-07	8.933205e-11	1.141993e-15	1.216236e-08	1.024011e-07
Oriental	0.2175938595	0.003005170	1.928407e-02	8.502096e-01	2.802834e-01	4.614578e-03	1.971217e-07	0.000000e+00	7.292153e-01	1.217473e-02	4.817151e-01	8.956483e-01
Rabat-Sale-Kenitra	0.0813981131	0.003401985	3.844796e-03	5.107560e-01	3.899232e-01	6.422742e-03	8.933205e-11	7.292153e-01	0.000000e+00	1.783368e-02	2.206948e-01	6.053997e-01
Settat-Casablanca	0.0002369256	0.006839036	1.700076e-06	2.211724e-03	4.282261e-01	8.580813e-02	1.141993e-15	1.217473e-02	1.783368e-02	0.000000e+00	4.610000e-04	6.460675e-03
Souss-Massa	0.4838891299	0.002908376	5.253410e-02	5.015601e-01	8.522638e-02	2.211724e-03	1.216236e-08	4.817151e-01	2.206948e-01	4.610000e-04	0.000000e+00	5.377279e-01
Tangier-Tetouan-Al Hoceima	0.2453673311	0.003005170	2.201198e-02	9.560550e-01	2.289278e-01	3.844796e-03	1.024011e-07	8.956483e-01	6.053997e-01	6.460675e-03	5.377279e-01	0.000000e+00

The table of p values resulted from the pairwise test in R studio

We found out that for most of the cases Dakhla-Oued eddahab region mean HDI is the highest mean HDI, apart from the Dakhla-Oued eddahab vs Laayoun-Sakia El hamra test which was

inconclusive. However, in the case of Marrakech-Safi region, we found that it is statistically speaking the region with the lowest HDI values.

So, this investigation concluded that we are 95% sure that the region treatment influences the HDI value of any commune.

2. Equations:

Hypothesis testing:

According to the statistics textbook, "Statistics for business & economics", the equations of the hypothesis test that we performed are the following (Anderson et al., n.d.):

1. The first test is a student's t distribution two-tailed test.

   - Hypotheses:

H_0: $\mu = 0.65$

H_a: $\mu \neq 0.65$ (With $\alpha = 0.05$, df = n-1)

   - T statistic is: $$t = \frac{(\bar{x} - 0.65)}{S/\sqrt{(n)}}$$

$\bar{x}$: sample mean
S: sample standard deviation
n: sample size

   - The confidence interval is: $$\bar{x} \pm t \times \frac{S}{\sqrt{n}}$$

Result: $0.55 < \mu < 0.56$ μ: population mean HDI in 2004

Thus, we are 95% sure that the mean HDI in 2004 is lower than the global HDI in 2004.

2. The second test is a Welch's t distribution two-tailed test.

- Hypotheses:

H_0: $\mu_{2004} - \mu_{2014} = 0$

H_a: $\mu_{2004} - \mu_{2014} \neq 0$ (With $\alpha = 0.05$)

- T statistic is: $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s^2_1}{n_1} - \frac{s^2_2}{n_2}}}$$

- Degrees of freedom:

$$df = \frac{(\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2})^2}{(\frac{s^4_1}{n^2_1.df_1} + \frac{s^4_2}{n^2_2.df_2})}$$

- Confidence interval:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df} \times \sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}$$

Result: $0.077 < (\mu_{2014} - \mu_{2004}) < 0.088$

Thus, we are 95% that the mean HDI increased between 2014 and 2004.

3. The third test performed is a chi² test; it compares the HDI's variance in 2004 of Morocco to the global variance of HDI in 2004:

- Hypotheses:

H_0: $\sigma^2_{2004} = 0.0275$

H_a: $\sigma^2_{2004} \neq 0.0275$ (With $\alpha = 0.05$)

- The confidence interval:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2_{2004} \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

- The χ^2 statistic:

$\chi^2 = \frac{(n-1)s^2}{\sigma_0}$; σ_0 is the variance to which we are comparing

$\chi^2 > \chi^2_{alpha}$ -> reject the null hypothesis

result: $0.0077 < \sigma^2_{2004} < 0.0088$

Thus, we are 95% sure that the population variance of HDI in 2004 of Morocco is different than the global variance of HDI in 2004.

4. The fourth test was the F test; it compares the variance of two samples: HDI in 2004 and 2014.

- Hypotheses:

H_0: $\sigma^2_{2004} - \sigma^2_{2014} = 0$

H_a: $\sigma^2_{2004} - \sigma^2_{2014} \neq 0$ (With $\alpha = 0.05$)

- F ratio:

$F = \frac{s^2_{2004}}{s^2_{2014}} > F_{\alpha/2}$ -> reject the null hypothesis.

Result: $1.88 < \frac{\sigma^2_{2004}}{\sigma^2_{2014}} < 2.29$ (the interval doesn't include 1)

Thus, the variability of HDI decreased from 2004 to 2014 in Morocco.

5. *ANOVA, the test compares the variability between samples and the variability within the samples:*

- The hypotheses:

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_{12}$

H_a: Not all regions' means are equal (With α=0.05)

$$SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 \quad ; \quad MSTR = \frac{SSTR}{k-1}$$

$$SSE = \sum_{j=1}^{k} (n_j - 1) s_j^{\,2} \quad ; \quad MSE = \frac{SSE}{n_T - k}$$

$F = \frac{MSTR}{MSE}$ -> variance between samples over the variance within the sample.

SSE: Sum of squares for the error

$SSTR$: Sum of square for the treatment

$MSTR$: Mean square for treatment

MSE: Mean square for error

$\bar{x}$: The overall sample Mean

n_j: number of observations for treatment j

$\bar{x}_j$: sample mean for treatment j

$s_j^{\,2}$: sample variance for treatment j

k: number of samples (12 regions)

($n_T = n_1 + n_2 + . . . + n_k$)

Result: $F > F_{\alpha/2}$ -> reject the null hypothesis.

Thus, we are 95% sure that two or more regions have different means.

- LISA clusters:

Maps are a way to represent data and to analyze it. LISA (Local Indicators of Spatial Association), provides an interesting way to analyze data by clusters. Established by Anselin in 1995, LISA is a way to assess significant local spatial clustering around an individual location and to indicate unexplained spatial patterns, or to suggest outliers or spatial regimes. It is a tool of ESDA (Explanatory spatial data analysis) (Anselin, 1995). It is based on the global Moran's "I" coefficient decomposed into a local level, which is the commune in this study. This spatial clustering procedure is repeated to all the communes (i) from all the regions (j). It produces a categorical outcome based on the relationship of a commune to other remaining communes within an area (Anselin, 1995). The equation representing the Moran's "I" coefficient is:

$$I = \frac{1}{S^2} \times \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}(\overline{Y_i} - \overline{Y})(\overline{Y_j} - \overline{Y})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}}$$

S^2: *is the overall mean rate*

ω_{ij}: *is the area weight indicator (1 for neighbor, 0 if otherwise)*

Y_i: *the local unit value of the variable*

$\overline{Y}$: *the global mean*

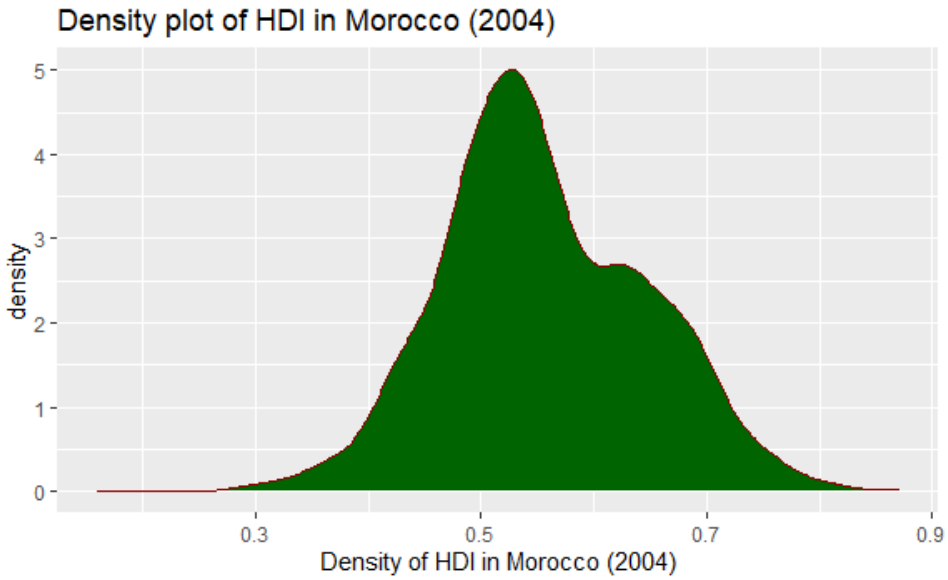
$\overline{Y_j}$: *the area average value of the variable*

This equation produces a statistic in which each commune's interaction with another one is taken into consideration, and when communes are statistically significant and similar, Morgan's "*I*" is positive, meaning that close communes tend to be more similar than the others far apart. Which yields to spatial clustering. Generally, LISA provides four types of clusters, high-high, low-low, high-low, low-high; each category concerns a certain spatial regime. A high-high cluster indicates the clustering of high values in an area, low-low cluster shows the clustering of low values in an area, whereas high-low shows the clustering of low values around a high value, and low-high cluster shows a clustering of high values around a low value. In this study, the data of HDI in communes shows only high-high and low-low regimes.

3. Visualizations:

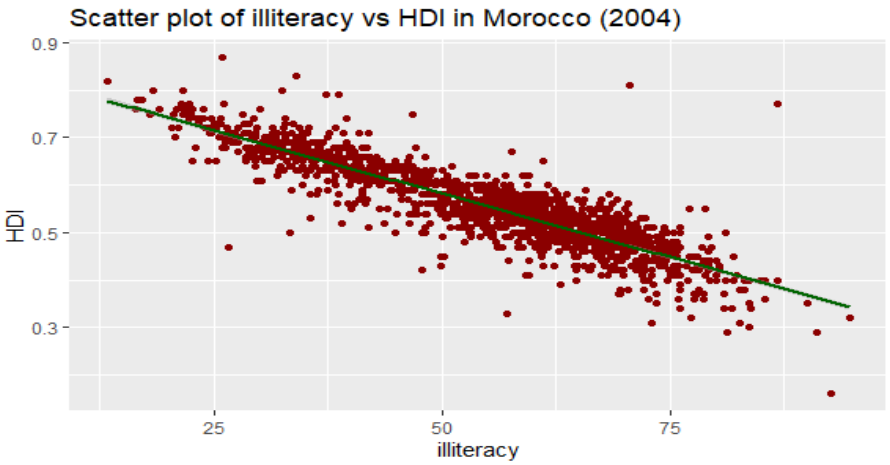
The visualizations that we created in this study were done using R studio. The main tool we used to create various plots is the ggplot function (Anselin, 1995).

To illustrate the distribution of HDI in 2004 of Moroccan communes, we used the standard density plot and the ggplot density using the ggplot function:



This distribution shows that HDI has not a perfect normal distribution, as the curve is skewed. However, its mean (0.555) is the value that mostly occur in the observation.

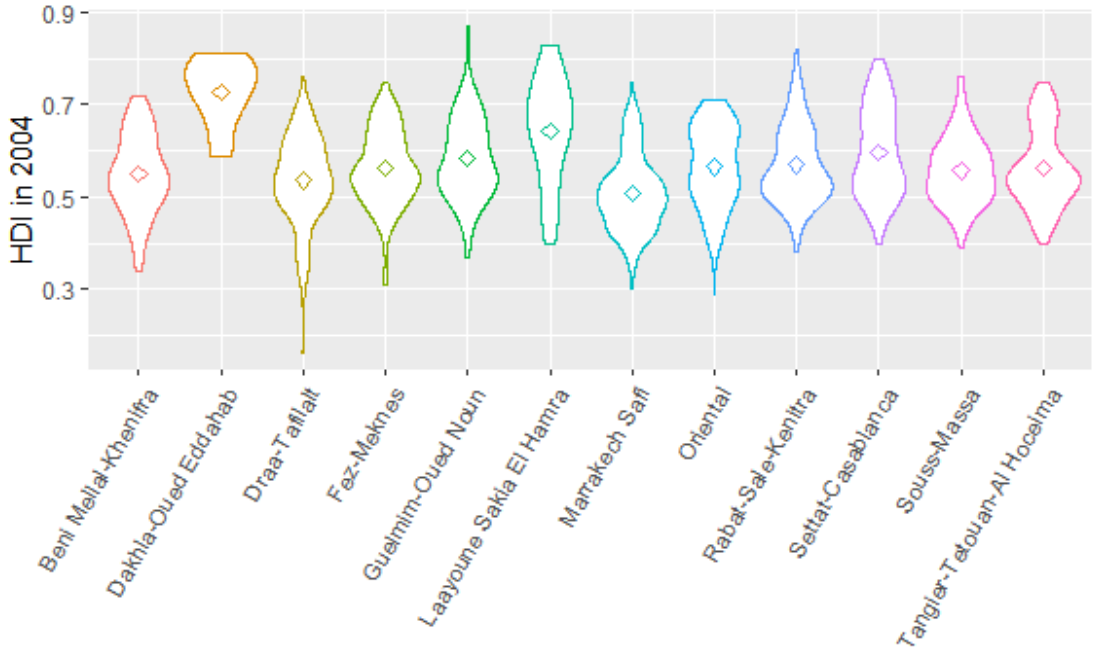
To illustrate the correlation between the HDI and the illiteracy rates in 2004 of Morocco, we used the standard scatter plot and the scatter ggplot:



Scatter ggplot using R studio of HDI vs Illiteracy in 2004

As you can see, the scatter plot indicates a negative relationship between HDI and Illiteracy, as when illiteracy increases, HDI goes down.

In order to illustrate the relationship between the region (treatment) and the HDI values, we used a violin plot using ggplot function. It is a very interesting plot as it gives a measure of the distribution of HDI in each region with its mean HDI which is the diamond at the center.

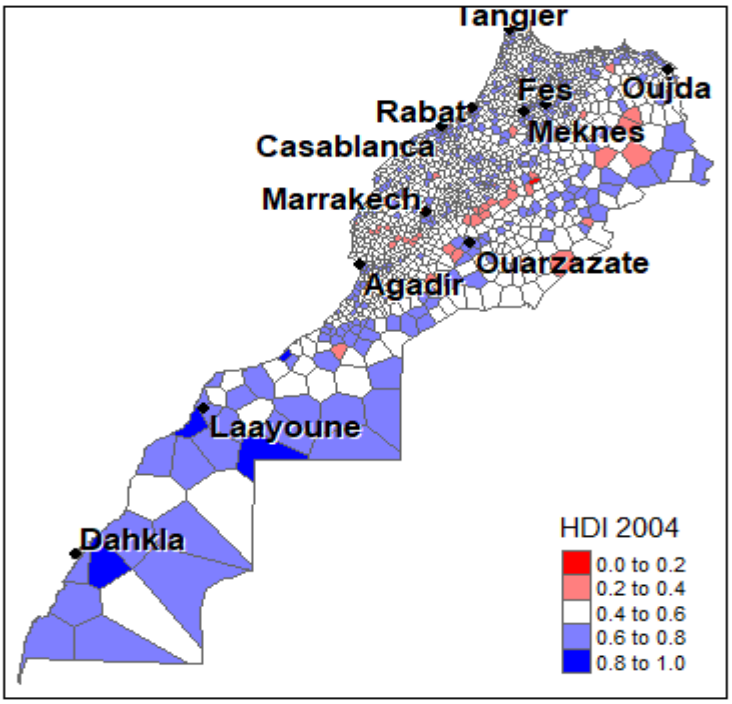


Violin plot using R studio of each region's HDI values in 2004

As you can notice, the region with the highest mean HDI value is Dakhla-Oued eddahab. However, the violin plot shows other high values, but the region with the biggest quantity of high HDI values is Dakhla-Oued eddahab. So, we can relate with the fact that statistically Dakhla-Oued eddahab is the region with the highest HDI values in 2004.

Another way to visualize data are maps. Thanks to R studio, we were able to plot the map of the Moroccan communes according to their HDI values in 2004. The HDI values in 2004 were divided into an interval, each interval represented by a color. The function used in R studio is "tm_shape" function.

We added layer by layer the different components of the map, which are the Moroccan communes, the cities, the HDI values.

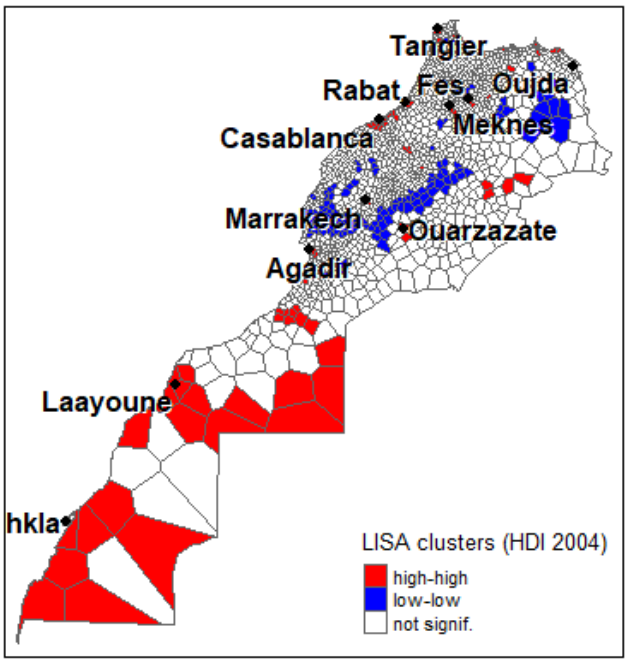


Map using R studio of HDI in 2004 of Moroccan communes

The map shows the HDI in 2004 represented by a colored interval of values. The highest values of HDI are in the cities of Dakhla, and Laayoun.

LISA clusters:

Using R studio, we mapped LISA clusters of HDI in 2004 of Moroccan communes.



Map using R studio of LISA clusters of HDI 2004 in Morocco

The map shows the different clusters that are present in the distribution of HDI of Moroccan communes in 2004. There're two types of clusters: high-high red cluster and low-low blue cluster. The high-high clusters indicates that the communes with a high HDI tend to gather around with other communes with high HDI, which is considered as a spatial regime. The low HDI communes tend to gather around with other communes with low HDI, which is another spatial regime as it is indicated by the low-low cluster. These regimes might be explained by several reasons that require further analysis.

IV. Conclusion about HDI in Morocco

This study was conducted on HDI of Morocco in 2004 and 2014, it presents various statistical results that are interpreted. We first compared the HDI mean in 2004 to the global mean HDI in 2004; Morocco has a lower HDI compared to the world, which means that, in 2004, Morocco is part of the countries who have well-being conditions for individuals below the global average. After that, we computed the correlation between HDI and illiteracy rates which is -0.88 then we plotted their respective scatter plot. The conclusion is that illiteracy strongly affects HDI in a negative direction. The more illiteracy we have in a commune, the lower its HDI. It is logic as the HDI includes the education dimension in its computation.

Another conclusion we made after comparing the HDI in 2004 and 2014 is that the population mean HDI in Morocco increased in the past decade. This might indicate that the indexes inside HDI has increased in this decade. All in all, this conclusion shows that the well-being conditions in Morocco have improved.

To investigate more about HDI, we compared the variance of HDI in 2004 of Morocco and the global variability of HDI in 2004. We found that the HDI variability in 2004 of Morocco is lower than the global HDI variability in 2004, which means that the Human development is more fairly distributed in Morocco

compared to the rest of the world. Moreover, we compared the variability of HDI in Morocco between 2004 and 2014. We found that the variability of HDI in 2004 is higher than the variability of HDI in 2014. It indicates that the human development gaps between communes have decreased during this decade. To analyze more about what affects HDI in Morocco, we included a new variable which is the region. Following ANOVA, we found that the region treatment influences HDI values in 2004 of Morocco. We found that HDI value ranges and means change from a region to another. This indicates that the human development is different from a region to another one and this can be explained by analyzing more deeply other variables.

In addition, we performed a spatial data analysis by computing the LISA clusters of HDI in 2004 in Morocco. We found that high HDI communes gather together in the same areas, and low HDI values communes gather together in the same areas. It indicates the specific regions with low and high HDI values, so analyzing data of other variables in these areas may explain the reason behind each spatial regime.

IV. References:

Stanton, E. (2007). The Human Development Index: A History. Working Paper Series, (127), 1-8. Retrieved 16 June 2020, from https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1101&context=peri_workingpapers.

Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. Statistics for business & economics.

Organisation territoriale du Maroc. Fr.wikipedia.org. (2020). Retrieved 16 June 2020, from https://fr.wikipedia.org/wiki/Organisation_territoriale_du_Maroc#:~:text=La%20loi%20no%2047,et%20place%20de%20la%20r%C3%A9gion.

BDS. Bds.hcp.ma. (2020). Retrieved 17 June 2020, from http://bds.hcp.ma/sectors.

Anselin, L. (1995), Local Indicators of Spatial Association—LISA. Geographical Analysis, 27: 93-115. doi:10.1111/j.1538-4632.1995.tb00338.x

B. L. WELCH, B.A., PH.D., THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED, *Biometrika*, Volume 34, Issue 1-2, January 1947, Pages 28–35, https://doi.org/10.1093/biomet/34.1-2.28

Sealey Gosset's, W. (1908). THE PROBABLE ERROR OF A MEAN [Ebook] (pp. 1-25). Biometrika. Retrieved 17 June 2020, from https://www.york.ac.uk/depts/maths/histstat/student.pdf.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling [Ebook] (50th ed., pp. 157-175.). Philosophical Magazine. Retrieved 17 June 2020, from https://doi.org/10.1080%2F14786440009463897.

Rao, C. (1992). R. A. Fisher: The Founder of Modern Statistics. Statistical Science, 7(1), 34-48. Retrieved June 17, 2020, from www.jstor.org/stable/2245989