How can we explain the level of happiness in certain countries ?

APPLIDED ECONOMETRICS , PARIS CITE UNIVERSITE

BENIGHIL Kenza & BENMADDI Yanis
2022/2023

Université Paris Cité

TABLE OF CONTENTS

INTRODUCTION

"The world may not be happier now, but the world is more about happiness and well-being than it was ten years ago", said Jeffrey Sachs

For the first time since the creation of the World Happiness Report created by the UN, France is in the Top 20 of this global happiness index, but what are the factors that really impact happiness in the world, how to measure it?

This is an existential question because every human being is looking for his happiness, which he can find in various ways (social, financial, family success, etc.) but can we define places where societies consider themselves happier than elsewhere?

The United States' declaration of independence established the pursuit of happiness as an inalienable human right. But you have to know what happiness is. However, it can have very varied meanings depending on the epochs, countries, cultures or even generations. Many of the proposed determinants are economic, such as GDP per capita, income, unemployment or major economic crises. But we will see throw our study that social factors are important determinants of happiness.

This is the subject of our study where we will try to explain the derived ranking of Gallup weights, which is the ranking of countries where people consider themselves most happy with a survey established in different countries in order to compare them according to economic and social factors

These totally subjective measures nevertheless reflect a strong psychological dimension in the persons concerned and thus make it possible to discuss the issue more in depth.

First of all, we will discuss the notion of happiness and how some variables can explain it thanks to our articles and research. Then we'll explain our database and our variables to make our study lighter and more comprehensive. After these presentations we will start our econometric work by the descriptive analysis of our variable score followed by a part where we'll speak about the relevance of each variable in our estimated model.

Next we'll make our progressive linear regression and talk about , make our multivariate analyze and see how our coefficients impact our explained variable.

It is for these reasons that the choice of our study focuses on the identification of the different factors influencing happiness in different countries in the world. The aim is to construct an econometric model to determine a score of happiness for each country studied.

I. LITERATURE REVIEW ON FACTORS IMPACTING GLOBAL HAPPINESS

➢ In a first step through our study we will seek to establish a causal link between variation in a country’s happiness and economic factors affecting well-being.

The literature found on this subject shows that economic factors can have a significant impact on happiness represented by the Gallup index, for example GDP per capita, income, unemployment or major economic crisis, are strongly correlated with happiness. (Dahan Jeremie , 2021)

➢In a second step we will prove that social and non-economic factor such as the social support, can have a real impact on happiness.

A second study, by Esteban Ospina and Max Roser in 2017, on happiness between countries, which takes as a determinant of happiness, income, life event, culture and society shows us that social factors influence happiness.

➢ Policy decisions at the organizational, corporate, and governmental levels should be more heavily influenced by issues related to well-being—people's evaluations and feelings about their lives (ed Dierner et Martin E.P Seligman 2004)

II. PRESENTATION OF THE DATABASE

Our database comes from the largest Data Science platform on the internet «Kaggle» is a Database for the analysis of the happiness score across 156 countries in the world, they are ranked from the happiest to the least happy.

Link to our Database below:

HTTPS://WWW.KAGGLE.COM/DATASETS/UNSDSN/WORLD-HAPPINESS?SELECT=2019.CSV

After consulting the necessary documentation concerning happiness in the world, the dependent variable on which our model will be built is the happiness score.

1. DESCRIPTIVE ANALYSIS OF VARIABLE Y: THE HAPPINESS SCORE

The score is calculated based on dystopia.
Dystopia is an imaginary country that has the world's least-happy people. The purpose in establishing Dystopia is to have a benchmark against which all countries can be favorably compared (no country performs more poorly than Dystopia) in terms of each of the six key variables, thus allowing each sub-bar to be of positive width. The lowest scores observed for the six key variables, therefore, characterize Dystopia. Since life would be very unpleasant in a country with the world's lowest incomes, lowest life expectancy, lowest generosity, most corruption, least freedom and least social support, it is referred to as "Dystopia," in contrast to Utopia.

Happiness score Y($score_i$)

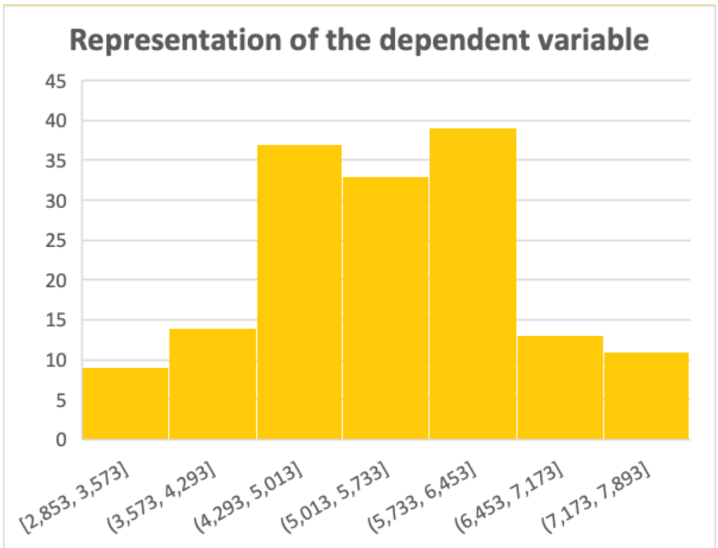
It is a quantitative variable that gives a score out of 10 of the country's level of happiness, it is obtained thanks to the data of the Gallup World Poll. It carries out surveys of representative samples in more than 160 countries and in 140 languages, on the basis of standardized questions. The main question asked is: "*Imagine a scale with bars numbered from zero at the bottom to ten at the top. The upper bar represents the best possible life for you, the lower bar the worst possible life for you. What bar do you think you stand at this moment in your life?*"
Known as the Cantril scale, this assessment of well-being, together with the precise formulation of the question, is a common standard for research on subjective well-being.

Table 01: Descriptive statistics of the dependent variable Y

Descriptive statistics of Y	
Moyenne	5,407096154
Erreur-type	0,089120915
Médiane	5,3795
Mode	5,208
Écart-type	1,113119869
Variance de l'échantillon	1,239035842
Kurstosis	-0,60837535
Coefficient d'asymétrie	0,011449949
Plage	4,916
Minimum	2,853
Maximum	7,769
Somme	843,507
Nombre d'échantillons	156

We can notice that the distribution of happiness scores is fairly homogeneous with a median and a fairly centered mean, we have a small variance of 1.23 so the difference is relatively small to the average.

Figure 01: Representation of the dependent variable



As we can see on this representation of the variable score, the repartition is almost following a Gaussian law and looks good with a variance and a kurtosis coefficient relatively low.

2. DESCRIPTIVE ANALYSIS OF EXPLANATORY VARIABLES

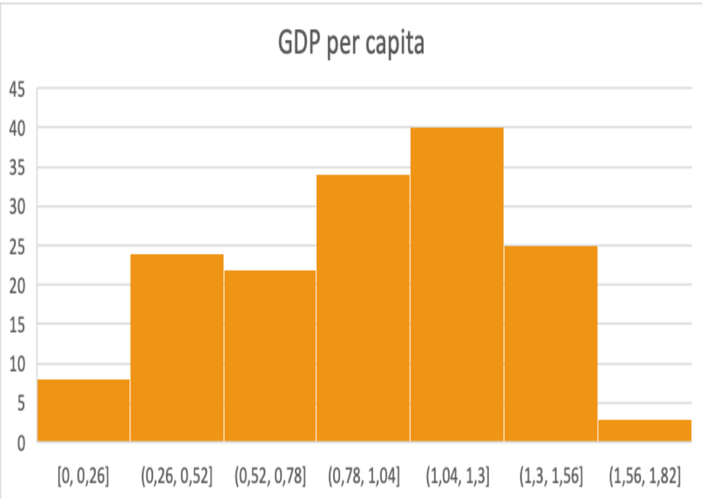
In our study, several variables will play the role of explanatory variables in order to determine happiness score. We have taken into account 2 explanatory variables based on what we could find relevant in the literature and 4 that we found interesting to analyze.

Variables from the literature:

- **GDP per capita**

According to the literature the work of the Stiglitz Commission GDP is a determinant of happiness, the richer a country is the happier it is, but Esterlin's paradox contradicts this hypothesis and is the subject of a real discussion. We will see in our model through the variable "*GDP per capita*" impact on happiness.

Figure 02: Representation of the GDP per capita variable



The first graph shows the distribution of the GDP per capita variable across each country surveyed. By graphic reading, we can see that this variable is fairly homogeneously distributed with a Gaussian look; the central values are quite strong compared to the extreme values.

- ***social support***

Our second variable chosen through the literature is *social support,* indeed it is a quantitative variable by definition it is the perception and actuality that one is cared for, has assistance available from other people, and most popularly, that one is part of a supportive social network that affects happiness

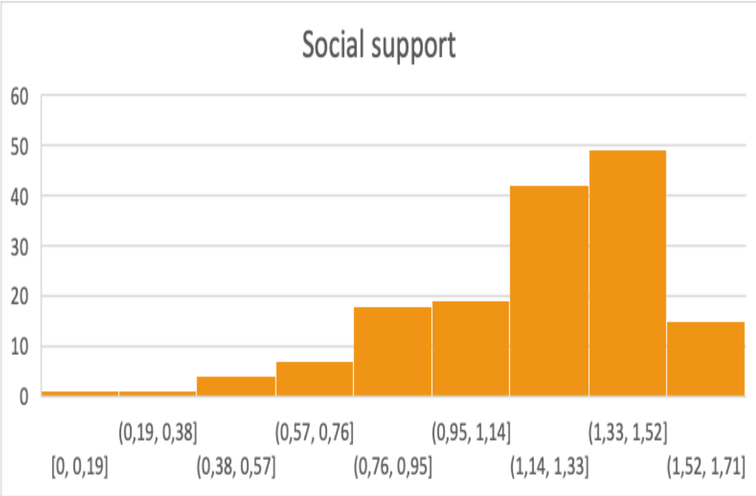


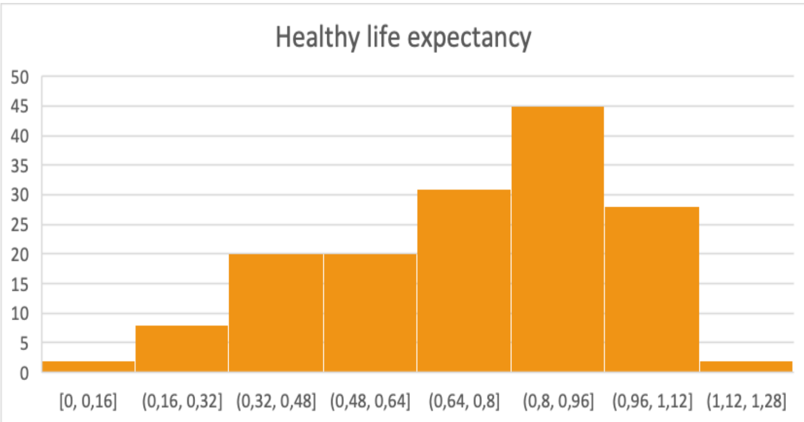
Figure 03: representation of the Social Support variable
This second graph shows the distribution of the variable Social Support for every country in the study. The distribution looks like a Gaussian distribution shift on the right with a peak between 1.44 and 1.52 with more than 80 countries.

<u>Variables that we have chosen:</u>

- **Life expectancy**

Our third variable is *Life expectancy* which resumes Life expectancy at birth represents the average lifespan of a fictitious generation subject to the age-specific mortality conditions of the year in question.

Figure 04: Representation of the healthy life expectancy variable

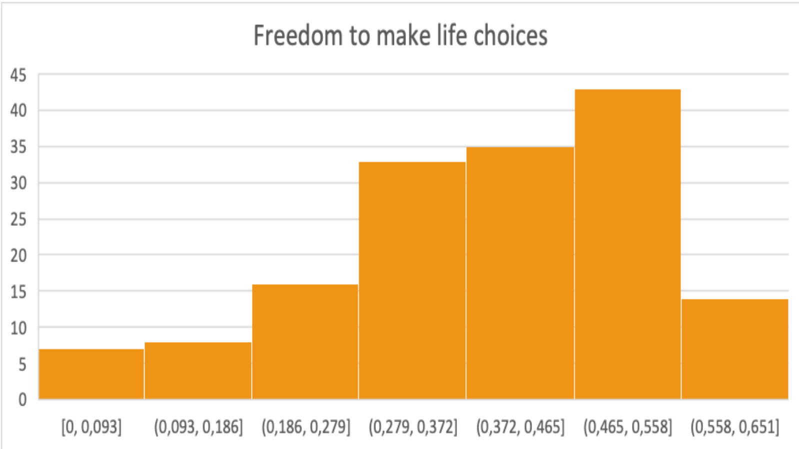


As the precedent graph we can see a similar distribution with same properties than Social Support graph. The repartition is more concentrated between 0,8 and 1,12 with more than 45 countries in this interval.

- **freedom to make life choices**

The fourth variable is the *freedom to make life choices*, it is a quantitative variable describing the freedom of individuals to make life choices freely, without difference of gender, number or social class

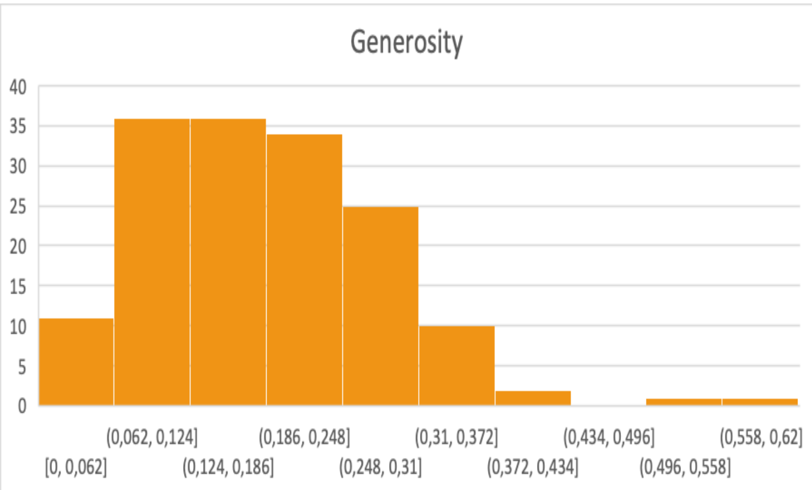
Figure 05: Representation of the Freedom to make life choices variable



- **Generosity**

Generosity describes the rate of action that is benevolent, lenient, forgiving in the country in question. The humanitarian goal rate in the country studied may be a criterion of this variable for example

Figure 06: Representation of generosity variable

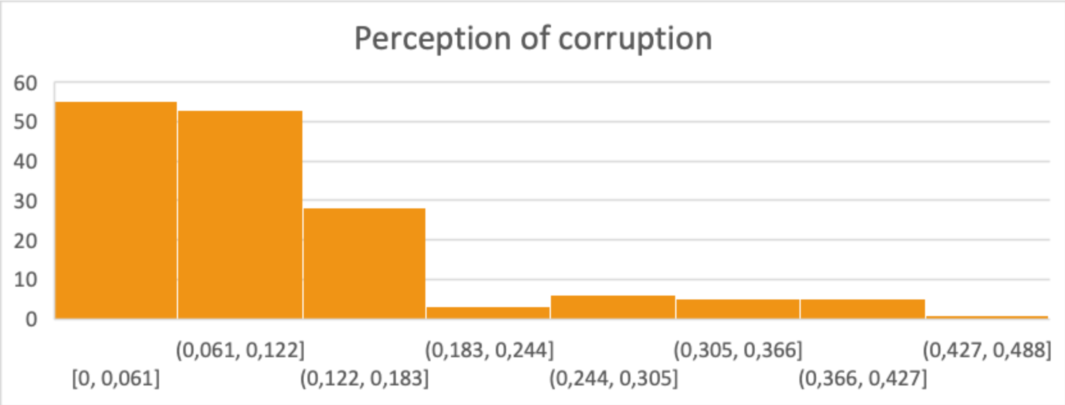


This distribution of the variable generosity also looks like a Gaussian distribution with a peak on the left means most countries are less generous than others but that's interesting because we can imagine countries with the best happiness score are also countries where people are more generous.

- **The perception of corruption**

The *perception of corruption is a variable* that describes whether the inhabitants of the studied country feel in a corrupt country or not.

Figure 07: representation of perception of corruption variable



This last graph is about how the distribution of the variable perception of corruption looks also like Gaussian but with a peak on the left that looks more understandable than the distribution of the variable generosity because we can imagine most countries have a less perception of corruption and a better happiness score that looks correlated.

We have done a table to summarize all the variables:

Table 02: Summary of our variables

Name	Nature	Unit	Expected signs
Score	Continuous quantitative	[0,10]	Positive
GDP per capita	Continuous quantitative	[0,10]	Positive
Social support	Continuous quantitative	[0,10]	Positive
Life expectancy	Continuous quantitative	[0,10]	Positive
Freedom to make life choice	Continuous quantitative	[0,10]	Positive
Generosity	Continuous quantitative	[0,10]	Positive
Perception of corruption	Continuous quantitative	[0,10]	Negative

The aim of our study from an econometric point of view is to explain the score by the various explanatory variables according to the following relation:

$$Y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \beta_4 x4 + \beta_5 x5 + \beta_6 x6 + u_i$$

Our model will be this following relation:

$$Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free + \beta_5 gener + \beta_6 percept + u_i$$

III. PRECONCEPTION REGRESSION

An initial exploratory analysis of our data highlights some potential relationships, the relevance of the choice of our variables and adjustments if necessary.

Table 03: table of the initial model

	Coefficients	*Erreur-type*	*Statistique t*	*Probabilité*
Constante	1,795220229	0,211073396	8,505194224	1,7676E-14
GDP per capita	0,775371626	0,218225354	3,553077636	0,0005103
Social support	1,124191579	0,236900074	4,745425193	4,8338E-06
Healthy life expectency	1,078142735	0,334538483	3,222776417	0,0015596
freedom to make choice	1,454832369	0,375337841	3,876061003	0,00015869
Generosity	0,489783351	0,497745477	0,98400362	0,32670886
Perception of corruption	0,972280221	0,542360729	1,792681825	0,07505259

Adjusted coefficient of determination $R^2 = 0,77$
Fisher P-value (overall model significance) =87.61

<u>Estimated Model Equation:</u>

$$Y(score_i) = 1{,}8 + 0{,}78 GDP + 1{,}12 social + 1{,}08 helth + 1{,}46 free + 0{,}49 gener + 0{,}97 precep + u_i$$

If we look at the significance of the coefficients estimated at the threshold of 5% we can notice that all the variables are relatively significant, there is little chance of being mistaken in rejecting the hypothesis H0 for the variables GDP, social support, Health and freedom, In addition, 0 is not included in the 95% confidence intervals of the four coefficients, so they are significantly different from 0 at the 5% threshold.

But the same cannot be said for the variable generosity and perceptions of corruption with a p-value of 0.33 and 0.08 which are > 0.05, these coefficients are not significant.

Following these observations we decided to modify the variables generosity and perception of corruption which are initially quantitative variables in binary quantitative variables , for this we have calculated the average of generosity and perception of corruption , countries with a rate of generosity and a perception of corruption above the average are considered generous and not corrupt.

We thus obtain our new estimated model:

Table 04: Regression of the model

	Coefficients	*Erreur-type*	*Statistique t*	*Probabilité*
Constante	1,840257464	0,20391399	9,02467485	8,4432E-16
GDP per capita	0,786057446	0,219287	3,58460577	0,00045679
Social support	1,094532906	0,23942859	4,57143771	1,0103E-05
Healthy life expectancy	1,128793783	0,33685257	3,3510024	0,00102041
Freedom to make life choices	1,611462667	0,36943004	4,3620239	2,3934E-05
Generosity quali	0,105587436	0,09400446	1,12321733	0,26315173
Perceptions quali	0,092460973	0,10286504	0,89885712	0,37017905

Adjusted coefficient of determination $R^2 = 0,77$
Fisher P-value (overall model significance) =85.47

Following the changes on the last two variables, we can see that the explanatory variables are better estimated than in the first case (table 03) for example for the variables freedom to make choice, the p-value is much smaller than in the first case, the coefficient is more significant at the 5% threshold. However, there is no real impact on the last two explanatory variables, they are not significant because the p-value >0.05.

V. Progressive linear regression (relationship between each X_i and Y)

1. GDP per capita variable

Figure 08: scatter plot of $Y(score_i) = \beta_0 + \beta_1 GDP + u_i$

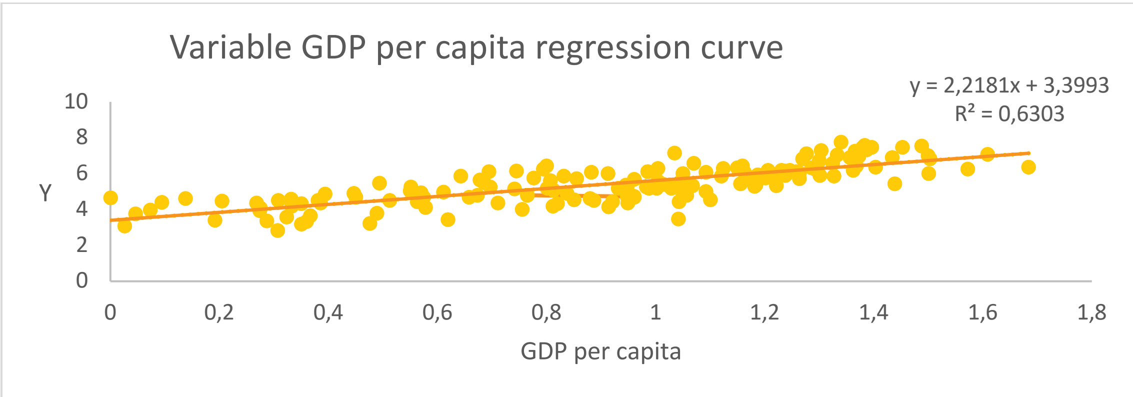


Table 05: Model regression $Y(score_i) = \beta_0 + \beta_1 GDP + u_i$

	Coefficients	Erreur-type	Statistic t	Probability
Constant	3,399345178	0,135323298	25,12017676	2,4398E-56
GDP per capita	2,218148001	0,136907682	16,20177899	4,3155E-35

R^2**=0.63**

Fisher P-value (overall model significance) =262,50

The regression gives significant results with an adjusted coefficient of determination R^2=0.63 (the GDP explains 63% of the variations of the variable Y of the score of happiness all other things being equal).

The GDP per capita variable is significant at 1%, so an increase of 1 unit of the GDP of the studied country increases the happiness score by 2.22 units the happiness score of this country, as seen previously in the literature.

If we want to know the overall significance of the model, the Fisher test gives us a coefficient equal to 262.5 and the table of Fisher's law (appendix 3) gives us 2.1; and as 262.5> 2.1 we reject H0 and we can globally conclude that the model is significant at the 5% level.

2. GDP per capita + social support

By adding the social support variable, to the variable Y we obtain the following regression:

Table 06: Model regression $Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + u_i$

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	2,329770638	0,211964551	10,99132201	4,43817E-21
GDP per capita	1,346478068	0,187480049	7,181980581	2,80444E-11
Social support	1,537510476	0,249639782	6,158916109	6,17724E-09

Adjusted coefficient of determination $R^2 = 0,60$

By adding the support social variable to our model which was initially bi-varied, we have a coefficient of determination which decreases slightly, which goes from 0.63 to 0.60, which means that the GDP and social support variables explain 60% of the variations in the happiness score, our model with the 2 variables remains relatively well explained.

Moreover, all the coefficients are significant because 0 is not included in the 99% confidence interval. We can also confirm that thanks to the Student p-value, there is little chance of being mistaken in rejecting the hypothesis H0 for the variables GDP and social support.

Thus, an increase of 1 social support unit increases the happiness score by 1.54 units and if GDP per capita and social support equal zero, the happiness score increases by 2,33 units, everything being equal.

For casting according to the literature, the GDP and social support variables are good determinants to explain happiness in the world.

3. GDP per capita + social support+ healthy life expectancy

Table 07: Model regression $Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + u_i$

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	2,135046981	0,211638497	10,0881787	1,222E-18
GDP per capita	0,809820385	0,235809748	3,43421081	0,000766123
Social support	1,321886267	0,24827968	5,32418227	3,58058E-07
Healthy life expectancy	1,297670988	0,366134165	3,54424993	0,000523454

Adjusted coefficient of determination $R^2 = 0,84$

For the moment it is the model that best explains the variations of the happiness score, the regression gives significant results with an adjusted coefficient of determination R^2=0.84 that is to say that the variables GDP, social support and health explain 84% of the variations of the variable Y of the score of happiness everything being equal.

We can notice that the *health* variable has a real impact and is an important variable in our model that explains well the happiness score because there is a gap of 21% with the previous model.
All variables are significant at 5%, so an increase of 1 unit of the variable healthy life expectancy increases the happiness score by 1.28 units.
It can be seen that the GDP per capita variable becomes less significant than the previous model seen before.

4. GDP per capita + social support + healthy life expectancy+ freedom to make choice

By adding the variable freedom to make choice to our model we obtain a coefficient of determination equal to 0.76, in this model, the explanatory variables explain 76% of the variations of the score of happiness

However, the significance of the coefficients is present as shown in the table below:

Table 08: Model regression $Y = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free + u_i$

	Coefficients	*Erreur-type*	*Statistique t*	*Probabilité*
Constant	1,89209537	0,19936358	9,49067711	4,802E-17
GDP per capita	0,81054939	0,21645095	3,74472553	0,000256291
Social support	1,0166104	0,23474755	4,33065392	2,69697E-05
Healthy life expectancy	1,1414486	0,33730887	3,38398635	0,000910305
Freedom to make life choices	1,84581569	0,34039145	5,42262655	2,28103E-07

Adjusted coefficient of determination $R^2 = 0,76$
Fisher P-value (overall model significance) =127,05

- Thus, increasing one unit of freedom to make choice increases the happiness score by 1.85 units all things being equal
- The Fisher P-value (overall model significance) =127,05 shows that it's model globally significant

5. GDP per capita + social support + healthy life expectancy + freedom du male choice +generosity

Table 09 : $Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free + \beta_5 gener + u_i$

	Coefficients	*Erreur-type*	*Statistique t*	*Probabilité*
Constante	1,833144057	0,20362988	9,002333346	9,23732E-16
GDP per capita	0,81944168	0,215980438	3,794055089	0,000214455
Social support	1,062506532	0,236611019	4,490520084	1,40857E-05
Healthy life expectancy	1,139663072	0,33641988	3,387621065	0,000900568
Freedom to make life choices	1,686462605	0,359653475	4,689131966	6,11816E-06
Generosity quali	0,123291697	0,091858977	1,342184531	0,181564141

Adjusted coefficient of determination $R^2 = 0,77$

By adding the generosity variable, the quality of the model stays sable with a coefficient of determination at 0.77.

however, this same variable is not significant with a p-value of 0.18, we can say that if the country is more generous than the average of all other countries in our observation then it does not have a real impact on its score of happiness

If the country is more generous than the average of all other observations(gener=1) then its happiness score increases by 0.12 points all things being equal.

We expected a higher value taken the literature into account (table 02)

6. **GDP per capita + social support healthy life expectancy + freedom du make choice +generosity + perception of corruption**

Table10:$Y(score_i) = \boldsymbol{\beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free +}$ $\boldsymbol{\beta_5 gener + \beta_6 perception + u_i}$

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	1,840257464	0,20391399	9,02467485	8,4432E-16
GDP per capita	0,786057446	0,219287	3,58460577	0,00045679
Social support	1,094532906	0,23942859	4,57143771	1,0103E-05
Healthy life expectancy	1,128793783	0,33685257	3,3510024	0,00102041
Freedom to make life choices	1,611462667	0,36943004	4,3620239	2,3934E-05
Generosity	0,105587436	0,09400446	1,12321733	0,26315173
Perceptions	0,092460973	0,10286504	0,89885712	0,37017905

Adjusted coefficient of determination $\boldsymbol{R^2 = 0,77}$

- As we saw previously in the progressive linear regression, the coefficient of the variable perception of corruption is not significant, there is a very strong chance of being wrong by rejecting the hypothesis H0 (p-value = 0.37).
- there is no real change in the fit quality of the model by adding the variable perception of corruption, the determination coefficient still is $R^2 = 0,77$

- Countries that have a higher than average perception of corruption (percp=1), increase its happiness score by 0.09 units, everything being equal, which is surprising, we expected more negative value. (table 02)

VI. Dummy variables study

We will try to see if there is an interaction between variables perception of corruption and generosity which are initially binary quantitative thanks to dummy variables

We will estimate this next model to identify possible interactions:

$$Y(score_i) = \beta_0 + \beta_1 gener + \beta_2 corruption\ + \beta_3\ generous * corruption + u_i$$

Table 11: results of estimation:

$Y(score_i) = \beta_0 + \beta_1 gener + \beta_2 corruption + \beta_3\ generous * corruption + u_i$

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	5,32146875	0,131245893	40,5457924	2,1028E-83
generosity	-0,373273628	0,21003344	-1,7772105	0,07753352
Perception of corruption	0,175766544	0,286486243	0,61352525	0,54044612
generosity* corruption	0,952626569	0,37601512	2,53347942	0,01230681

Adjusted coefficient of determination $R^2 = 0,11$

- <u>Interpretation of the interaction term:</u>

The interaction term is 5% significant with a p-value equal to 0.12306 which is a little bit lower than 0.05. So that means there is a significant interaction effect between generosity and perception of corruption.

Generosity's effect on the score depends on the perception of corruption and inversely so the interaction term is useful to explain the happiness score.

- <u>coefficient interpretation:</u>

The reference is for generosity = 0 and perception of corruption = 0 so a country is less generous than the average of all other countries and same for the perception of corruption.

The effect that the country is considered more corrupted (perception of corruption = 1) in a ungenerous country (generosity = 0) increases the score by 0.9 units of happiness.

The effect that the country is considered more corrupted (perception of corruption = 1) in a generous country (generosity = 1) increases the score by (0,176+0,953=1,129) 1,13 units of happiness.

The effect that the country is considered more generous (generosity = 1) in a country where there is a low perception of corruption (perception of corruption = 0) decreases the score by 0.37 units of happiness.

The effect that the country is considered more generous (generosity = 1) in a country where there is a high perception of corruption (perception of corruption = 1) decreases the score by (-0.373+0.952=0.579) 0.58 units of happiness.

VII. Reverse causality

In this part we will estimate a new model, we will through a bi-varied analysis try to explain the GDP by the score of happiness and thus inverted our variables of our basic model, We found it interesting to analyze this situation knowing the literature cited above, indeed the GDP could vary according to the happiness of the population. The dependent variable on which our model will be built is *GDP per capita,* and the explanatory variable is *score*

We will estimate this model to identify possible interactions:

$$Y(GDP_i) = \beta_0 + \beta_1 score + \text{ui}$$

Table 12: Reverse causality model

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	-0,6311894	0,09680105	-6,5204807	9,4989E-10
Score	0,2841334	0,01753718	16,201779	4,3155E-35

$$R^2 = 0{,}63$$

We obtained satisfactory results, with a model well explained indeed the score of happiness explains 63% of the variations of the GDP, the model to a correct quality of adjustment.

In terms of the significance of the coefficients, they are all significant at 1%, so an increase of one unit of the happiness score increases the GDP by 0.28 units and if the happiness score is zero then the GDP falls by 0.63 units everything being equal.

The coefficients are expected signs and are in agreement with the other studies carried out and cited in our literature We can say that the GDP plays a primordial role in our model is a variable that explains well the variations of the happiness score .

VIII.MULTIVARIATE ANALYSIS

1. The significance of the variables

Estimated (General) Model Equation:

$$Y(score_i) = 1{,}840257464 + 0{,}786057446GDP + 1{,}094532906social + 1{,}128793783health + 1{,}611462667free + 0{,}105587436generous + 0{,}105587436perception + u_i$$

Table 13: Estimation of the general model

	Coefficients	*Erreur-type*	*Statistique t*	*Probability*
Constant	1,840257464	0,20391399	9,02467485	8,4432E-16
GDP per capita	0,786057446	0,219287	3,58460577	0,00045679
Social support	1,094532906	0,23942859	4,57143771	1,0103E-05
Healthy life expectancy	1,128793783	0,33685257	3,3510024	0,00102041
Freedom to make life choices	1,611462667	0,36943004	4,3620239	2,3934E-05
Generosity	0,105587436	0,09400446	1,12321733	0,26315173
Perceptions	0,092460973	0,10286504	0,89885712	0,37017905

Thanks to table 13, we can see that GDP, social support, healthy life expectancy, freedom to make life choices and the constant are significant at the 99% level. Indeed, their T-stat is over

2.58 and their P-value is under 0.01. So we can reject the hypothesis H0 of non-significance for these 4 variables and the constant at 90%, 95% and 99% .

Generosity and Perception variables have no T-stat above or P-value below the relevant thresholds, so we cannot reject the H0 hypothesis of no significance

2. Interpretation and analysis of coefficients

All things being equal:

· If all explanatory variables are equal to zero, the happiness score increase by 1.84 units

· An increase of 1 unit of country's GDP increase happiness score by 0.78 units

· An increase of 1 unit of Social Support increase happiness score by 1.09 units

· An increase of 1 unit of Healthy Life Expectancy increase happiness score by 1.13 units

· An increase of 1 unit of Freedom to make life choices increase happiness score by 1.61 units

· If the country is considered generous his score increase by 0.1 units

· If inhabitants feel like living in a corrupted country the happiness score increases by 0.09 units.

In our final model all coefficients positively influence the happiness score which is surprising as we expected a negative coefficient for the last variable, the perception of corruption influences happiness positively but very weakly

3. Fisher and overall model significance:

The following hypothesis are used:

H0 -> the model is not significant

H1 -> the model is significant

If we calculate the coefficient of Fisher F=$(R^2/\text{k})/[(1\text{-}R^2)/(\text{N-(k+1)}]$ we get 85.47 (which corresponds to our analysis table of variance of our regression) and $F_{1-\alpha}(k, N-1-k)$ we get 2.10.

here $\text{F} > F_{1-\alpha}(k, N-1-k)$ we accept H0

According to the Fisher test, after comparing the F of the model and the table in (appendix 3), we find that we reject the hypothesis H0 meaning that the model is poorly estimated. Thus we conclude here that the model is globally significant

IX – LES TESTS

1-test de multi colinéarité

Tableau 14: correlation matrix

	Score	*GDP per capita*	*Social support*	*Health*	*Freedom*	*Generosity*	*Perceptions*
Score	1						
GDP per capita	0,79388287	1					
Social support	0,77705779	0,754905727	1				
Health	0,77988315	0,835462115	0,719009459	1			
Freedom	0,56674183	0,379079072	0,447333164	0,390395	1		
Generosity	0,04564942	-0,079157446	-0,095163554	-0,056417	0,2510867	1	
Perceptions	0,29922192	0,267431552	0,145092332	0,242428	0,33916177	0,25930763	1

Let remind that this table shows the correlation links between variables on the scale from 0 to 1, 0 showing an absence of link and 1 showing a perfect collinearity. Negative values having the same scale but an inverse ratio, that is, when one variable increases, the other decreases

- There is a strong correlation between the score and GDP as we see in the study of the precedent model
- We can see also a strong correlation between variables GDP and Social Support (with a correlation coefficient equal to 0.84) this result is predictable thanks to the literature
- However, GDP and generosity have a negative correlation which is understandable in our study, there is a close link between perception of corruption and generosity of inhabitants which is a social factor and GDP which is a variables.

Université Paris Cité

2-Klein test and VIF

- **Klein test**

The approach of this test is as follows:

We compare the R^2 calculated from the regression of the general model and the auxiliary regressions

the regression of our general model is : $Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free + \beta_5 gener + \beta_6 percept$

Auxiliary regressions are:

$$Y(GDP) = \gamma_0 + \gamma_1 social + \gamma_2 health + \gamma_3 free + \gamma_4 gener + \gamma_5 percept + v_i$$

$$Y(social) = \gamma_0 + \gamma_1 GDP + \gamma_2 health + \gamma_3 free + \gamma_4 gener + \gamma_5 percept + v_i$$

$$Y(health) = \gamma_0 + \gamma_1 GDP + \gamma_2 social + \gamma_3 free + \gamma_4 gener + \gamma_5 percept + v_i$$

$$Y(free) = \gamma_0 + \gamma_1 GDP + \gamma_2 social + \gamma_3 health + \gamma_4 gener + \gamma_5 percept + v_i$$

$$Y(gener) = \gamma_0 + \gamma_1 GDP + \gamma_2 social + \gamma_3 health + \gamma_4 free + \gamma_5 percept + v_i$$

$$Y(percept) = \gamma_0 + \gamma_1 GDP + \gamma_2 social + \gamma_3 health + \gamma_4 gener + \gamma_5 gener + v_i$$

- **VIF**

For each auxiliary regression, the tolerance is calculated: TOL = 1 $-R^2$

Then we calculate VIF = 1/TOL

Finally, all higher values of VIF (>10) indicate the presence of multi colinearity.

Tableau 15: multicollinearity tests:

	R^2	Tol	VIF
General model: score =f(GDP per Capita, Family, Life Expectancy, Freedom, Generosity, Trust Government Corruption)	0,77	0,23	4,35
Auxiliar model GDP per capita =f(social support, Life Expectancy, Freedom, Generosity, Trust Government Corruption)	0,75	0,25	4
social support =f(GDP per capita , Life Expectancy, Freedom, Generosity, Trust Government Corruption)	0,62	0,38	2,63
Life Expectancy,=f(GDP per capita ,social support, Freedom, Generosity, Trust Government Corruption)	0,71	0,29	3,45
freedom =f(GDP per capita ,social support, Life Expectancy, Generosity, Trust Government Corruption)	0,31	0,69	1,45
Generosity=f(GDP per capita ,social support, Life Expectancy , freedom, Generosity, Trust Government Corruption)	0,13	0,87	1,15
Trust Government Corruption=f(GDP per capita ,social support, Life Expectancy , freedom, Generosity,)	0,16	0,84	1,19

commentaries:

- R^2 of auxiliary models are lower than the Ra^2 so we can say there is no collinearity problems
- All the VIFs of the auxiliary models are strictly less than 10
- So, we can conclude that our model presents an absence of multicollinearity

3. Detection of heteroscedasticity

- **Breush-Pagan test**

The approach is :estimate the model used by the OLS : $Y(score_i) = \beta_0 + \beta_1 GDP + \beta_2 social + \beta_3 health + \beta_4 free + \beta_5 gener + \beta_6 percept$+$u_i$ and we take $u_i{}^2$

-we estimated the auxiliary model u_i^2 =$\lambda 0 + \lambda_1 GDP + \lambda_2 social + \lambda_3 health + \lambda_4 free + \lambda_5 gener + \lambda_6 percept + vi$

-We calulate the R^2 of the auxiliary model Ra^2

- We calculate Breush-Pagan stat : BPc=NxRa^2~Chi2(Ka); with Ka the number of explanatory variables of the auxiliary model, here we have 6 variables

- We conclude, at the 5% threshold if BPc<Chi2 we do not reject the null hypothesis of homoscedasticity

- In our case Bpc= 9,31468197 and Chi2=12,592 (appendix 2) , Chi2>Bpc therefore, H0 is not rejected, there is no risk of heteroscedasticity

CONCLUSION

Thanks to the documentation and our study, we can define some factors with more or less impact on happiness.

The study of correlation between variables in our econometric model allowed us to say that some of these variables can explain why there are some countries where the happiness score is higher than somewhere else. Indeed, in our third regression (table 07) for example we can see for the variables' combination GDP per capita, Social support and Healthy life expectancy a R2 = 0.84 which means the combination of these 3 variables explain 84% of the happiness score variation. We can also see in our correlation matrix a strong link between score and variable GDP per capita. That means some variables have a heavier impact on happiness score than others so countries with better GDP per capita, Social support and Healthy life expectancy have a better happiness score than others.

However it's also very important to remind ourselves there are a lot of other factors which can explain this score and this study can't define an universal ranking of "Where people are the happiest in the world ?" because every data of this study is based on the dystopia's concept which is subjective.

So is it really possible to measure happiness significantly and to compare with each other?

BIBLIOGRAPHY

1)Influence of Stock Market Indices on Citizen Happiness Dahan, J. (Author). June 2021

2)Esteban Ortiz-Ospina & Max Roser (2017) Happiness and life satisfaction , Our Word in Data

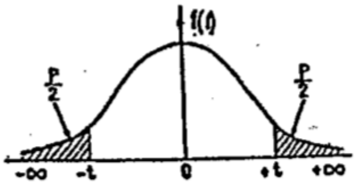
3)Diener, E., & Seligman, M. E. P. (2004). Beyond Money: Toward an Economy of Well-Being. *Psychological Science in the Public Interest*, *5*(1), 1–31.

APPENDIX

Appendix 01

TABLE DE LA LOI DE STUDENT

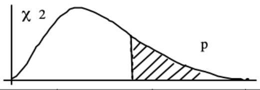
Valeurs de T ayant la probabilité P d'être dépassées en valeur absolue



ν	P=0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,336	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,12566	0,25335	0,38532	0,52440	0,67449	0,84162	1,03643	1,28155	1,64485	1,95996	2,32634	2,57582

Nota. — ν est le nombre de degrés de liberté.

TABLE DU CHI-DEUX : $\chi^2(n)$



n \ p	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,341
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Pour n > 30, on peut admettre que $\sqrt{2\chi^2} - \sqrt{2n-1} \approx N(0,1)$

Appendix 02

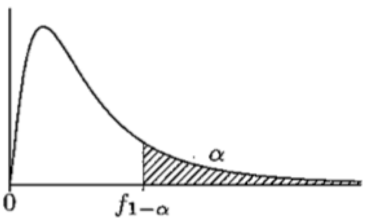
Appendix 3

Loi de Fisher–Snedecor ($\alpha = 0,05$)

Si F est une variable aléatoire suivant la loi de Fisher–Snedecor à (ν_1, ν_2) degrés de liberté, la table donne la valeur $f_{1-\alpha}$ telle que

$$\mathbb{P}\{F \geqslant f_{1-\alpha}\} = \alpha = 0{,}05.$$

Ainsi, $f_{1-\alpha}$ est le quantile d'ordre $1 - \alpha = 0{,}95$ de la loi de Fisher–Snedecor à (ν_1, ν_2) degrés de liberté.



$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	15	20	30	∞
1	161	200	216	225	230	234	239	242	246	248	250	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,94	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,85	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	2,72	2,65	2,57	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	2,62	2,54	2,47	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,20	2,12	2,04	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,15	2,07	1,98	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25	2,11	2,03	1,94	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,22	2,07	1,99	1,90	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,19	2,04	1,96	1,87	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	1,92	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03	1,87	1,78	1,69	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,84	1,75	1,65	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,95	1,79	1,70	1,60	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,93	1,77	1,68	1,57	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,67	1,57	1,46	1,00