

Statistics I Final Exam, 17th May 2017.

Degrees in ADE, DER-ADE, ADE-INF, FICO, ECO, ECO-DER, ADE-EEII.

EXAM RULES: 1) Use separate booklets for each problem.

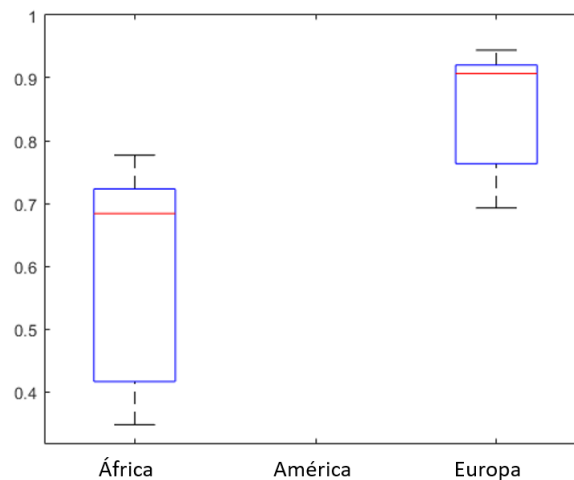
- 2) Perform the calculations with at least two significant decimal places.
- 3) You may not leave the exam during the first 30 minutes.
- 4) You are not allowed to leave the classroom without handing in the exam.

1. (2,5 points) The following table shows the values of the Human Development Index (HDI) for different countries of Africa, America and Europe in the year 2015.

Africa	0,348	0,411	0,413	0,416	0,419	0,646	0,666	0,684	0,69	0,698	0,721	0,724	0,736	0,772	0,777
America	0,483	0,666	0,679	0,714	0,715	0,772	0,78	0,783	0,785	0,79	0,793	0,827	0,847	0,919	0,923
Europe	0,693	0,751	0,754	0,761	0,771	0,899	0,907	0,907	0,908	0,916	0,916	0,922	0,923	0,93	0,944

Answer to the following questions:

- (a) (0,6 points) Find the three quartiles for each of the three continents and decide if there are any outliers in the data of each continent.
- (b) (0,7 points) Draw the box-plot of the American data in the following picture. Determine the shape of each distribution and compare them. Which measures of centrality and variability are more appropriate in each case? Do not calculate them.



- (c) (0,6 points) Justify the truthfulness or falseness of the following statements. Apply the quartiles to justify your answers.
 - 1) 50% of African countries in the table have an HDI that is below the level reached by any of the European countries in the table.

2) 75% of American countries in the table have an HDI that is above the level reached by any African countries in the table.

(d) (0,6 points) The HDI can be classified as follows: Very High $[0.8, 1)$; High $[0.7, 0.8]$; Medium $[0.55, 0.7)$ and Low $[0, 0.55]$. The contingency table for the variable *continent* (X) and the variable *HDI* in categories (Y) is depicted below:

X / Y	Low, (0; 0, 55)	Medium, [0, 55; 0, 7)	High, [0, 7; 0, 8)	Very High, [0, 8; 1)
África	5	5	5	0
América	1	2	8	4
Europa	0	1	4	10

What percentage of countries with high or very high HDI belongs to Europe? And what percentage of countries with an HDI of less than 0, 55 belongs to the African continent?

2. (3 points) According to the data published by the INE, the distribution of households in terms of the number of members is the one shown in the following table. The third column shows the percentage of households owning an e-book:

Households	Number	Percentage of households owning an e-book
1 or 2 members	7.899.565	18, 60%
3 members	3.847.663	23, 90%
4 or more members	4.282.595	29, 30%

Answer the following questions defining clearly the events and the results used.

- (a) (0,75 points) Calculate the probability that a Spanish household randomly selected does not own an e-book.
- (b) (0,75 points) If a Spanish household is selected randomly and it turns out that it owns an e-book, what is the probability of having only 1 or 2 members?
- (c) (0,75 points) Choosing a household at random, what is the probability of it having less than 4 members and not owning an e-book?
- (d) (0,75 points) An e-book company designs an e-book promotion campaign and selects a simple random sample of 20 households, what is the probability that more than 85% of selected households will not have an e-book?

3. (2,25 points) In a production chain, the occurrence of an error that can be critical in the chain is usually checked. Let X be a random variable that represents the relative magnitude of that error. Let $f(x)$ be its density function:

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) (0,75 points) Draw the density function and the distribution function of X .
- (b) (0,75 points) Calculate the probability that the error is greater than half its maximum value.
- (c) (0,75 points) Calculate the median and the expected value of X .
4. (2,25 points) The results from a survey carried out by the CIS about “*Attitude and Innovative Behaviour in the Spanish Society*” are shown below (Table 1). The surveyed were asked about their opinion on the importance of innovation in various areas (energy sources, infrastructure, environment, education, corporations and social services). Additionally, the following global indicator about the importance of innovation has been constructed based on their answers:

$$INNOVA_i = \frac{X1_i + X2_i + X3_i + X4_i + X5_i + X6_i + X7_i + X8_i + X9_i}{9}$$

The basic descriptive analysis of $INNOVA$ is also shown below (Table 2):

	Media	Desviación típica
Las fuentes de energía (X_1)	8,47	1,66
Las infraestructuras (X_2)	7,86	1,77
El medio ambiente (X_3)	8,30	1,90
La medicina (X_4)	9,41	1,09
La alimentación (X_5)	7,60	2,31
La Administración pública (X_6)	7,19	2,27
La enseñanza (X_7)	8,55	1,76
Las empresas (X_8)	8,30	1,76
Los servicios sociales (X_9)	7,95	2,03

Table 1

Mean	8,20528133
Standard error	0,02518517
Median	8,33333333
Mode	10
Standard deviation	1,24252491
Sample variance	1,54386814
Kurtosis	1,13235098
Skewness	-0,72792436
Range	9,22222222
Minimum	0,77777778
Maximum	10
Sum	19971,6548
Count	2434

Table 2

- (a) (0,6 puntos) Explain how you think the estimate of the mean of variable $INNOVA$, that is, $\overline{innova} \approx 8,20$, was obtained.

- (b) (0,7 points) Calculate a 95% confidence interval for the average level of importance assigned to innovation, μ_{INNOVA} .
- (c) (0,7 points) Comment the adjustment of variable *INNOVA* to a normal distribution. Justify your answer based on the following graphs.

