URJC – GADE BILINGÜE - CORPORATE STATISTICS II
June 2018 Exam
(model X)

<table>
<tr><td colspan="2">SURNAMES:</td><td colspan="3">NAME</td></tr>
<tr><td colspan="2">DNI:
Group:</td><td>B:PROBLEM SOLVING (60% weight)</td><td>POINTS</td><td>STUDENT</td></tr>
<tr><td colspan="2">A: MULTIPLE CHOICE (40% weight)</td><td>Exercise 1</td><td>2.25</td><td></td></tr>
<tr><td colspan="2"></td><td>Exercise 2</td><td>3.5</td><td></td></tr>
<tr><td>RIGHT (+ 1/10)</td><td></td><td>Exercise 3</td><td>2.5</td><td></td></tr>
<tr><td>WRONG (-0,2/10)</td><td>-</td><td>Exercise 4</td><td>1.75</td><td></td></tr>
<tr><td>MC GRADE out of 10</td><td></td><td>PS GRADE</td><td>out of 10</td><td></td></tr>
<tr><td>40 %</td><td></td><td>60%</td><td></td><td></td></tr>
<tr><td colspan="3"></td><td>FINAL GRADE</td><td></td></tr>
</table>

EXAM DURATION: 110 MINUTES

The exam has two sections:

- Multiple Choice section (pages 2-4): 10 questions, weighting 40% of the final grade. **ONLY THOSE ANSWERS MARKED IN THE MASK WILL BE CONSIDERED.** At the end of this section there is space to carry out calculations if needed.
- A correct answer adds 1 point. A wrong answer substracts 0.2 points. A question not answered adds 0 points. **A minimum grade of 4 points in this section is required for assessing section B.**
- Problem Solving section (pages 5-14): 4 exercises weighting 60% of the total grade. **A minimum grade of 5 points is required to pass**.
- The final grade will be the result of adding 40% of the MC and 60% of the PS. **A minimum final grade of 5 points is required to pass**.

MULTIPLE CHOICE ANSWERS

	1	2	3	4	5	6	7	8	9	10
A					X					
B	X		X					X	X	
C		X					X			X
D				X		X				

SECTION A: TEST

1.- The period of time for a bus covering a certain trip is a random variable following a N(40;2). Under these circumstances, what is the approximated probability for a bus doing a given trip between 34 and 46 minutes?

a) 0,95
b) 0,99
c) 0,05
d) 0

2.- Suppose a variable X following a $N(\mu_x;\sigma_x)$ and a variable Y a $N(\mu_y;\sigma_y)$. A s.r.s. from each variable has been taken, being both samples independent. Then the following estimation with a 95% of confidence has been obtained:

$$\frac{\sigma_x^2}{\sigma_y^2}\epsilon[0.4;1.1]$$

Choose the right option at that level of confidence:

a) X has higher dispersion than Y
b) Y has higher dispersion than X
c) There aren't significative differences between the dispersions of both variables
d) We cannot make any statement unless the Central Limit Theorem applies

3.- Let ξ be a random variable N(μ;σ) and a s.r.s. with size n obtained from it in order to estimate μ. Consider now the following point estimator:

$$\hat{\mu}=\frac{\sum x_i}{n+1}$$

Choose the right option regarding that point estimator:

a) It is unbiased
b) It is biased but asymptotically unbiased
c) Its bias is equal to $\frac{1}{n+1}\mu$
d) The three previous statements are false

4.- In the previous question, choose the right choice regarding the variance of the point estimator:

a) $\frac{\sigma^2}{(n+1)^2}$

b) $\frac{\sigma^2}{n+1}$

c) $\frac{\sigma^2}{n^2}$

d) The three previous statements are false

5.- Choose the right choice in relation with the situation arising in a hypothesis test when it has a low power:

a) The probability of accepting the null hypothesis being false is high
b) The probability of accepting the alternative hypothesis being false is high
c) The probability of rejecting the null hypothesis being right is high
d) The three previous statements are false

6.- Let ξ be a random variable and a s.r.s. of size n obtained from it in order to estimate μ. In which of the following situations a Student's t distribution will be used?

a) As long as the distribution of the variable be unknown and the sample small
b) As long as the Central Limit Theorem be applied
c) As long as the variable follows a $N(\mu;\sigma)$ and the population variance be known
d) The three previous statements are false

7.- Let a B(m;p) where m is large (at least 30) and p equal to 0.3. Choose the right answer regarding the distribution than can be approximated:

a) t_{n-1}
b) $F_{n-1,m-1}$
c) $N(mp;\sqrt{mp(1-p)}$
d) $P(\lambda = mp)$

8.- Which of the following is a feature of a simple random sample (s.r.r.)?

a) It is applied when the population is heterogeneus
b) Randomness plays a key role in the selection process
c) When the sample size is high, each element in the sample follows a normal distribution
d) It is a non probabilistic sampling technique

9.- Choose the right answer regarding the maximum likelihood estimators:

a) They are unbiased, efficient and normal variables
b) They are unbiased, efficient and normal variables only in large simple random samples
c) They are always the same estimators than those obtained by the general method of moments
d) They arise when maximizing the density function of the variable

10.- The field of Statistics allowing to draw conclusions over populations based on data coming from samples is referred to as:

a) Probability
b) Descriptive Statistics
c) Statistical Inference
d) Trigonometry

REMEMBER TO PASS YOUR ANSWERS TO THE MASK

SPACE FOR YOUR CALCULATIONS FOR THE MULTIPLE CHOICE SECTION. IF NEEDED.

PROBLEM SOLVING SECTION

Exercise 1 (2.25 points)

An independent chemical laboratory is carrying out a research on the quality of the air in Madrid. It has started studying the amount of NO_2 (nitrogen dioxide) pollution measured in micrograms/m^3. Aimed at that the chemical engineer in charge of the project has taken a s.r.s. of 100 measures realized in different locations of the city in February 2018. Assuming that the variable follows a N(μ;σ) these have been the results obtained:

$$\sum x_i = 4469.6 \qquad \sum x_i^2 = 202051.07$$

We wish to run a Pearson's Chi square goodness of fit test to match that sample with a N(μ;σ) distribution. To achieve that you must take into account that the sample has been divided in 5 classes with the same probability, apart from the information provided in the next table:

Classes	Oi	Pi	Ei	$\frac{(O_i - E_i)^2}{E_i}$
	16			
	26			
	21			
	20			
	17			
	100			

Where Oi is referred to as the observed frequency in each class, Pi is the probability assigned by the normal distribution and Ei is the expected frequency under such distribution.

Answer:

a) Complete the table. (1.25 points)

b) Run the corresponding test, with α = 5%, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value. (1 point)

Exercise 2 (3.5 points)

The maximum exposure limit for NO_2 compatible with health is established in 40 micrograms/m^3. Having that in mind the engineer assures that the mean amount of NO_2 existing in Madrid air in February was higher than that limit (the variable corresponds to that one considered in exercise 1 and it is assumed to follow a $N(\mu;\sigma)$). Based on the s.r.s. of 100 observations obtained in exercise 1 answer:

a) Run the appropriate test under the engineer's point of view at a 5% significant level, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve by the p-value. (1.5 points)

b) By another side the engineer wants to verify that the standard deviation of the variable studied is equal to 5 micrograms/m^3. Run the appropriate test under the engineer's point of view at a 5% significant level, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve by the p-value. (2 points)

Exercise 3 (2.5 points)

Then the engineer has mixed the NO_2 pollution with ozone (O3) and particulates (PM10) in order to come up with an air quality index. When such index is above certain amount it means that the air quality is bad. With the purpose of knowing if the air quality is good or bad, the engineer mixed the pollutants in the 100 locations that she chose at random in February, conforming a s.r.s. After that she computed 40 locations with bad air. Answer:

a) Elaborate a confidence interval estimation at a 99% level of confidence for the corresponding population proportion, including the pivot statistic and the statistics defining the lower and upper limits of the interval. Consider the maximum uncertainty case. The engineer affirms that more than a half of the city bore a bad air in February. Is she right? Justify. (1.75 points)

b) Give the sampling distribution of the point estimator used in question 3a). (0.75 points)

Exercise 4 (1.75 points)

Finally the engineer wishes to know if the quality air is similar along the city or not. With that purpose she has divided the city in three sectors: A, B, and C. Then supported by the s.r.s. with 100 locations of the previous exercise the distribution of places regarding the quality of the air was:

	A	B	C
Good air	18	24	18
Bad air	15	8	17

Run an independence test at a 5 % significance level, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value.