URJC – GADE BILINGÜE - CORPORATE STATISTICS II

June 2016 Exam

(model X)

SURNAMES:		**NAME**		
DNI: Group:		B:PROBLEM SOLVING (60% weight)	POINTS	STUDENT
A: MULTIPLE CHOICE (40% weight)		Exercise 1	2	
		Exercise 2	2	
RIGHT (+ 1/10)		Exercise 3	4	
WRONG (-0.2/10)	-	Exercise 4	2	
MC GRADE out of 10		PS GRADE	out of 10	
40 %		60%		
			FINAL GRADE	

EXAM DURATION: 100 MINUTES

The exam has two sections:

- Section A: Multiple Choice (pages 2-4) with 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 subtracts 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 and for passing the exam.**

- Section B: Problem Solving (pages 5-12) with 4 exercises weighting 60% of the total grade. **A minimum grade of 5 points in this section is required to pass**.

- The final grade will be the result of adding 40% of section A and 60% of section B. **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								X	
B										X
C		X					X	X		
D			X	X	X	X				

SECTION A: TEST

1.- Choose the right answer regarding a hypothesis testing in which the null hypothesis is rejected with α = 5% but accepted with α = 1%:

a) The p-value is higher than 1% but lower than 5%
b) The p-value is higher than 5%
c) The p-value is lower than 1%
d) The situation formulated in the statement can never take place

2.- Suppose we want to compare the population variances of two normal variables $X \sim N(\mu_x; \sigma_x)$ and $Y \sim N(\mu_y; \sigma_y)$. Then we take two independent s.r.s. of a given size, one from each variable, and obtain a 95% confidence interval estimate for the quotient of variances with this result:

$$\frac{\sigma_x^2}{\sigma_y^2} \in [0{,}8 - 1{,}3]_{\gamma=0{,}95}$$

Choose the right answer at that level of confidence:

a) $\sigma_x^2 > \sigma_y^2$
b) $\sigma_x^2 < \sigma_y^2$
c) $\sigma_x^2 = \sigma_y^2$
d) We can only get conclusions when we use a 100% level of confidence interval estimate

3.- Choose the right answer regarding the sampling distribution of a point estimator:

a) This means that the point estimator is an unbiased estimator
b) This means that the point estimator is an efficient estimator
c) This means that the point estimator is a consistent estimator
d) This means that the point estimator is a random variable

4.- Choose the right answer regarding a hypothesis test consisting in comparing the observed (absolute) frequencies in the sample with those expected under the theoretical distribution being tested:

a) It is the goodness of fit Kolmogorov-Smirnov test
b) It is the goodness of fit Shapiro-Wilks test
c) It is a parametric test for the mean in a normal variable
d) It is the goodness of fit Pearson's Chi square test

5.- Which one is the maximum likelihood (ML) point estimator of the population proportion in a B(1;p) once a s.r.s. of size *n* has been obtained?

a) The first element in the sample x_1
b) The last element in the sample x_n
c) The inverse of the sample mean $\frac{1}{\bar{x}}$
d) The sample mean $\bar{x}$

6.- Which of the following features correspond to a simple random sample?

a) Each element in the population have the same probability to be included in the sample
b) Each element in the sample is independent in probability from other ones when replacement is considered
c) Each element in the sample has the same probability distribution than the variable under study
d) The three previous statements are true

7.- Let a B(m;p) where m is large (bigger than 30) and p is not small (higher than 0.1). 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.- When we have a large s.r.s. selected from a random variable, we say that the sample mean $\bar{x}$ follows a $N(\mu; \frac{\sigma}{\sqrt{n}})$ where *n* is the sample size. Choose the right answer:

a) This is true only when the variable under study follows a $N(\mu; \sigma)$ with σ known
b) This is true only when the variable under study follows a $N(\mu; \sigma)$ regardless of knowing σ
c) This is true in any case due to the Central Limit Theory
d) This is true only when the variable under study follows a $P(\lambda)$

9.- Choose the right answer in relation with the Fisher-Snedecor $F_{m,n}$ distribution (where t_n denotes the student's t distribution):

a) $F_{1,n} \sim t_n^2$
b) $F_{2,n} \sim t_n^2$
c) It is symmetric
d) It fits many phenomenon in the real life

10.- Choose the right answer regarding the power of the test in a hypothesis test:

a) It is the probability of rejecting the null hypothesis when it is true
b) It is the probability of rejecting the null hypothesis when it is false
c) It is equal to 1-α
d) It is equal to α-1

REMEMBER TO PASS YOUR ANSWERS TO THE MASK

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

PROBLEM SOLVING SECTION

Exercise 1 (2 points)

A pharmaceutical company has launched a new medicine for certain disease. The company assures that patients recover in less than 5 days. In order to test this hypothesis a group of doctors have applied this medicine to a s.r.s. of 20 patients in a given hospital. Data concerning the recovery time measured in days (variable X) obtained from the sample has been:

$$\sum x_i = 83{,}11 \qquad \sum x_i^2 = 362{,}42$$

Assuming that X follows a $N(\mu_x;\sigma)$, answer:

a) Elaborate the corresponding test under the point of view of the pharmaceutical company with $\alpha = 1\%$, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value. (1,5 points)

b) Discuss the kind of error which could be incurred arising from the decision made and the associated probability. (0,5 points)

Exercise 2 (2 points)

In the context of the previous exercise, the group of doctors want to compare the effectiveness of the new medicine to that achieved with the medicine they were using so far. Calling Y the variable consisting in the recovery time measured in days for those patients being treated with the old medicine, the doctors have collected the following data from Y based on a s.r.s. of 20 different patients:

$$\sum y_i = 85{,}75 \qquad \sum y_i^2 = 380{,}54$$

Assuming that Y follows a $N(\mu_y;\sigma)$, both samples are independent and both variables X and Y have the same unknown standard deviation σ, answer:

a) Run a test to check if there are significative differences between the population means μ_x and μ_y at a 1% level of significance, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value. (1,5 points)

b) Discuss the kind of error which could be incurred arising from the decision made and the associated probability. (0,5 points)

Exercise 3 (4 points)

The group of doctors are also concerned with the proportion of patients having adverse reactions to the new medicine.

a) As a consequence they propose two point estimators of the proportion of patients suffering adverse reactions in the population, coming from a s.r.s. of size *n*. Those point estimators are:

$$\hat{p}_1 = \frac{x_1 - 2x_2}{3} \qquad\qquad \hat{p}_2 = \bar{x}$$

Compare both estimators studying the properties of unbiasness, efficiency and consistency and choose the best one. (1,5 points)

b) In order to quantify this effect they have broaden the s.r.s. with patients of other hospitals who have been also treated with the new medicine. Over all they have computed 16 patients with adverse reactions out of 100. Based on this, elaborate a 99% level of confidence interval estimation for the population proportion, including the pivot statistic and the statistics defining the lower and upper limits of the interval. The pharmaceutical company assures that only 7% of the patients experience adverse reactions. Discuss the last statement under the results obtained. (1,5 points)

c) What will be the sample size required in order to have a sampling error of 3% keeping the level of confidence unchanged? (1 point)

Exercise 4 (2 points)

Finally, the group of doctors want to know if the appearance of adverse reactions does not depend on the age of the patient. For this purpose they have considered three age stratus: 18-45; 46-70; beyond 70. Therefore supported by the s.r.s. of the previous exercise the distribution of patients by age has been:

	18-45	46-70	beyond 70
Patient having adverse reactions	4	4	8
Patient not having adverse reactions	29	31	24

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.