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

SURNAMES:		NAME		
DNI: Group:		B:PROBLEM SOLVING (60% weight)	POINTS	STUDENT
A: MULTIPLE CHOICE (40% weight)		Exercise 1	2	
		Exercise 2	4.5	
RIGHT (+ 1/10)		Exercise 3	2.5	
WRONG (-0,2/10)	-	Exercise 4	1	
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 5-15) 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 (mark with X)

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

SECTION A: TEST

1.- A grade in an exam is a random variable following a N(6;1).Under these circumstances, what is the approximated probability of a student obtaining a grade higher than 8?

a) 0,95
b) 0,05
c) 0,025
d) 0

2.- Consider a random variable uniform $U[0;\theta]$ and a s.r.s. of size n taken from that population. Then the General Method of Moments estimator of the parameter $\hat{\theta}_{GMM}$ is equal to:

a) $\bar{x}$

b) $\frac{1}{\bar{x}}$

c) $2\bar{x}$

d) $\frac{\bar{x}}{2}$

3.- Let a random variable with density function $f(x;\theta)$ and a s.r.s. of size n obtained from it. Which of the following statements will be true regarding an unbiased estimator $\hat{\theta}$?

a) Its application to a s.r.s. of size n will produce the parameter θ exactly
b) Its application to a s.r.s. of size n will produce the parameter θ exactly, only when the Central Limit Theory applies
c) Its application to a s.r.s. of size n will produce the parameter θ exactly, only in normal variables
d) Its application to a s.r.s. of size n will produce the parameter θ on average

4.- Which of the following statements is true regarding a Pearson's Chi square goodness of fit test?

a) It is a two sided test
b) It has a big β when the sample size is small
c) It compares cumulative relative frecuencies in the sample with the cumulative distribution function
d) It only applies to discrete random variables

5.- Consider the following statistic coming from a s.r.s. of size n obtained from a random variable (σ^2 is the population variance and s_1^2 is the bias corrected sample variance):

$$\frac{(n-1)s_1^2}{\sigma^2} \sim \chi_{n-1}^2$$

Choose the right choice regarding the distribution followed necessarily by that random variable:
a) t_{n-1}
b) $P(\lambda)$
c) $N(\mu,\sigma)$
d) $B(m;p)$

6.- The number of patients on average entering an emergency unit in a hospital is 2 per minute. Under these conditions, which probability distribution will follow the number of patients entering that unit in 50 minutes?

a) $N(100; 10)$
b) $N(50; 7.07)$
c) $P(50)$
d) $B(50; p)$

7.- Choose which one of the following features correspond to a s.r.s.:

a) The population from which the sample has been selected is heterogeneus
b) Each element in the population has the same probability of been selected
c) Each element in the sample follows a $N(\mu;\sigma)$ when the sample size is big enough
d) Each element in the sample follows a $N(\mu;\sigma)$

8.- A manager of a given company has collected a s.r.s. of 100 days in order to estimate the daily income produced. With a 99% level of confidence the interval estimation obtained for the population mean has been [15;25] in thousands of euros. If we had used a 95% level of confidence instead, what would have been the consequence?

a) The sampling error would have been higher so the precision lower
b) The sampling error would have been higher so the precision higher
c) The probability for the interval including the parameter would have been lower
d) Provided that the sample size would have been the same the interval wouldn't have changed

9.- Choose the right option regarding the situtation arising from accepting the null hypothesis in a test:

a) It is the best situation because the type II error is our primary concern in a test
b) It is the best situation because the type I error is low
c) It is a situation where the power of the test is unknown
d) It is the best situation because β is low

10.- Which of the following tool is a secondary source of information?

a) Eurostat (the statistical office of the European Union)
b) A survey
c) An interview to an expert
d) A parameter

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 solar panel manufacturer is studying the efficiency of a new panel the company is about to launch to the market. Aimed at that the engineer in charge of the project has collected a s.r.s. of 30 days in Summer, measuring in kws (kilowatts) the energy produced by that panel per day. Assuming that the variable follows a N(μ;σ), these are the results obtained:

$$\sum x_i = 115.81 \qquad \sum x_i^2 = 473.71$$

First of all the previous assumption has to be tested using a Kolmogorov-Smirnov test with Lilliefors correction. With that purpose, take into accout the information included in the next table:

X: Energy per day	$F_{Oi}(x_i)$	$F_{Oi}(x_{i-1})$	$F(x_i)$	d_i:	d_{i-1}:
1,9	0,0333	0,0000	0,0209	0,0124	0,0209
2,5	0,0667	0,0333	0,0773	0,0107	0,0440
2,5	0,1000	0,0667	0,0837	0,0163	0,0170
2,6					
2,6	0,1667	0,1333	0,0960	0,0707	0,0374
2,8	0,2000	0,1667	0,1334	0,0666	0,0333
3,1	0,2333	0,2000	0,2024	0,0309	0,0024
3,3	0,2667	0,2333	0,2943	0,0276	0,0610
3,4	0,3000	0,2667	0,3260	0,0260	0,0594
3,5	0,3333	0,3000	0,3464	0,0131	0,0464
3,5	0,3667	0,3333	0,3693	0,0026	0,0360
3,6	0,4000	0,3667	0,4012	0,0012	0,0346
3,7					
3,7	0,4667	0,4333	0,4385	0,0282	0,0052
3,9	0,5000	0,4667	0,5038	0,0038	0,0371
3,9	0,5333	0,5000	0,5196	0,0137	0,0196
3,9	0,5667	0,5333	0,5359	0,0308	0,0025
4,0					
4,0	0,6333	0,6000	0,5673	0,0661	0,0327
4,0	0,6667	0,6333	0,5717	0,0949	0,0616
4,2	0,7000	0,6667	0,6219	0,0781	0,0447
4,3	0,7333	0,7000	0,6743	0,0590	0,0257
4,7	0,7667	0,7333	0,8232	0,0566	0,0899
4,8					
4,9	0,8333	0,8000	0,8597	0,0264	0,0597
5,0	0,8667	0,8333	0,8752	0,0085	0,0418
5,1	0,9000	0,8667	0,9023	0,0023	0,0356
5,1	0,9333	0,9000	0,9088	0,0245	0,0088
5,4	0,9667	0,9333	0,9472	0,0194	0,0139
5,8	1,0000	0,9667	0,9762	0,0238	0,0096

Where $F_{oi}(x_i)$ represents the relative cumulative frequency distribution at the observation x_i, $F(x_i)$ the cumulative distribution function of the normal variable at x_i and d_i the difference between both functions expressed in absolute terms.

Answer:

a) Fill in the blanks, justifying the calculus realized. (1 point)

b) Run the corresponding test, with $\alpha = 5\%$ including the hypothesis, the test statistic, the critical region and the decision taken. Also solve using the *p-value*. (1 point)

Exercise 2 (4.5 points)

The engineer asserts that the panel produces a mean of 4 kws (kilowatts) per day in sunny days. She wishes to show her position based on the sample used in exercise 1.

Answer:

a) Elaborate the corresponding hypothesis test for the mean amount of energy produced, under the engineer's point of view, with $\alpha = 5\%$, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value. (1.5 points)

b) Give (without justifying) the maximum likelihood point estimator of the population mean of question a) and justify if it verifies the properties of unbiasness and consistency. (1.25 points)

c) The technical specifications of the panel indicate that the average dispersion in the amount of energy produced by the panel per day, measured through the standard deviation, is equal to 0.5 kws. Using the s.r.s. collected in exercise 1, run the corresponding test with $\alpha = 1\%$, formulating the hypothesis, the test statistic, the critical region and the decision made. Also solve using the p-value. (1.75 points)

Exercise 3 (2.5 points)

The new panels have been elaborated in two plants, X and Y, and the engineer wants to know if there are significative differences between the proportion of panels having defects depending on the plant where they have been made. Having taken two s.r.s. of 1000 units from each plant, the number of panels with some problems have been 20 in X and 15 in Y.

Answer:

a) Elaborate a 95% confidence interval for the difference in the proportion of panels with defect between both plants, including the pivot statistic and the statistics defining the lower and upper limits of the interval. Based on that respond the engineer's question (1.5 points)

b) What would be the sample size required on each s.r.s. obtained in order to incrase the level of confidence till 99% keeping the sampling error unchanged? Justify. (1 point)

Exercise 4 (1 point)

The engineer wants to know if the distribution of observations in the s.r.s. of exercise 1 were randomly distributed in relation with the plant where they were made: X or Y. That sample of 30 panels regarding this issue is as it follows:

X	X	X	X	Y	Y	Y	Y	X	X
Y	Y	X	X	Y	Y	X	X	X	X
Y	X	X	X	Y	Y	Y	Y	X	X

Answer:

Elaborate the Wald Wolfowitz runs test with $\alpha = 5\%$, formulating the hypothesis, the test statistic, the critical region and the decision made.

DRAFT