BUSINESS STATISTICS - Second Year
February 10, 2011

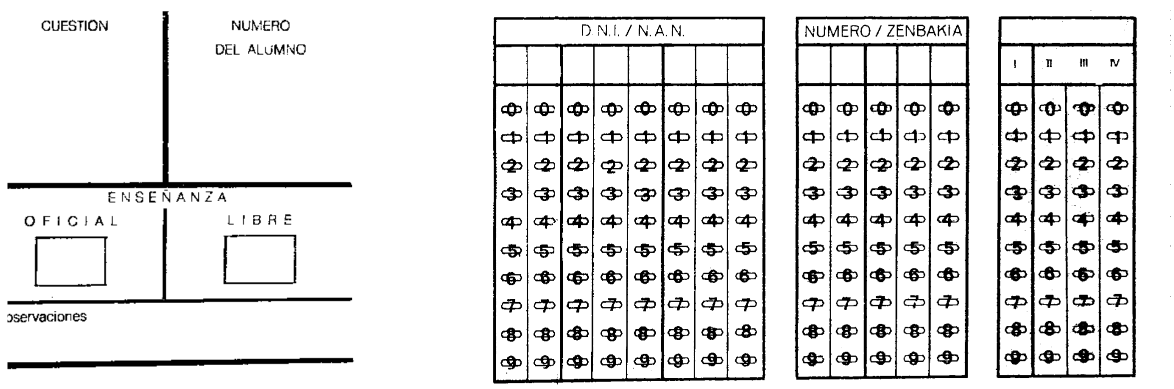
Academic Year 2010-11

INSTRUCTIONS

1. The exam contains multiple choice questions that must be answered in the orange code sheet we have provided you with, together with three exercises that must be answered in detail in separate sheets of paper.
2. To select an answer, all you need to do is mark in the orange code sheet, **filling the rectangle over which the selected answer is located appropriately**. Please make sure you know the answer you wish to mark before doing it. Even though you can always erase your mark if you have used a pencil (number 2 or similar), any mark that has not been completely erased could be read by the machine. Therefore, we advice you to first mark your selected answers in the exam and to use only the last ten minutes or so from the time assigned to the multiple choice questions-part of the exam to copy them into the code sheet.
3. In the multiple choice questions-part of the exam there is always **only one correct answer** for every question. Every question correctly answered is worth 1 point, while each question incorrectly answered will penalize your grade by 0.2 points. Questions that have not been answered do not penalize your grade in any form.
4. Each one of the exercises, A, B and C, must be answered in a separate sheet of paper. We will collect the different parts of the exam at the indicated times and in this order: first, we will collect the code sheet for the multiple choice questions part of the exam and, then, and in this order, exercises A, B and C will be collected.
5. The exam has six numbered sheets, going from 0.1 to 0.6. Please make sure that you have all sheets and contact your professor if this is not the case. There are different exam types. This exam is of type 0. Mark a 0 in the column labelled with I in your code sheet, just as is illustrated in the example.
6. The maximum final grade for each of the parts of the exam (i.e., the multiple choice questions part and the exercises part) is 30 points. You will need to obtain 15 points in each part of the exam to pass it. However, exams having a multiple choice question part with grades greater than or equal to 14 could, under special circumstances, be compensated with a good grade in the exercises part of the exam.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises. In "Resit" (column II) you will write the number of times you have registered for an exam in this course *not including this one.*

Example:

$\underbrace{12545}$ PEREZ, Ernesto Exam type $\underbrace{0}$ Resit



MULTIPLE CHOICE QUESTIONS (Time: 1 hour and 40 minutes)

1. FREE-QUESTION. The capital of Spain is:

   (A) Paris (B) Sebastopol (C) Madrid (D) London (E) Pekin

Questions 2 to 5 refer to the following exercise:

The probability that an appliance of a given brand, commonly sold in specialized stores, fails within its warranty period is 0.05. When this happens, the store salesman will have to replace the appliance. If the second one fails again within its warranty period, it is not replaced and the store salesman has to compensate the client with the amount of money he originally paid for the appliance. It is assumed that all appliances are independent from each other.

2. If the store salesman sells 10 of those appliances, the probability that he has to replace at least one of them is:

   (A) $(0.95)^{10}$ (B) $1-(0.95)^{10}$ (C) $1-(0.05)^{10}$ (D) $(0.05)(0.95)^9$ (E) $(0.05)^{10}$

3. If the store salesman sells 65 of those appliances, the probability that he has to replace exactly 2 of them is:

   (A) 0.79 (B) 0.42 (C) 0.21 (D) 0.58 (E) 0.80

4. If it is known that the store salesman has replaced 2 appliances, the probability that he has to compensate the client only for one of them is:

   (A) 0.9525 (B) 0.0475 (C) 0.9500 (D) 0.0500 (E) 0.0950

5. If the store salesman sells 500 of those appliances, the approximate probability that he has to replace at most 20 of them is:

   (A) 0.2475 (B) 0.6614 (C) 0.7525 (D) 0.8212 (E) 0.1788

Questions 6 to 8 refer to the following exercise:

Let $X_1, \ldots, X_4$ be independent and identically distributed random variables having a Poisson distribution with parameter $\lambda = 2$. We define another r.v. $Z = \sum_{i=1}^{4} X_i$.

6. We have that $P(Z \leq 6)$ is:

   (A) 0.3134 (B) 0.1912 (C) 0.4530 (D) 0.6866 (E) 0.1220

7. The mode(s) of the r.v. Z is (are):

   (A) Only 7 (B) 7 and 8 (C) Only 8 (D) 8 and 9 (E) Only 9

8. If we define the r.v. $\overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i$, then its exact distribution is:

   (A) $\mathcal{P}(2)$ (B) All false (C) $N(m=2, \sigma^2 = 2/\sqrt{4})$ (D) $\mathcal{P}(4)$ (E) $N(m=2, \sigma^2 = 2)$

9. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables associated to the result of throwing a regular coin; that is, $X_n = 1$ if we obtain heads, and $X_n = 0$ if we obtain tails, both events occurring with respective probabilities $p = q = \frac{1}{2}$. The sequence converges:

(A) In distribution to $X = \frac{1}{2}$

(B) In probability to $X = 0$

(C) In distribution to a binary r.v. of probability $\frac{1}{2}$

(D) In probability to $X = 1$

(E) All false

10. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables with cumulative distribution function given by

$$F_n(x) = \begin{cases} 0, & \text{if } x < 0 \\ nx, & \text{if } 0 \leq x < \frac{1}{n} \\ 1, & \text{if } x \geq \frac{1}{n} \end{cases}$$

Then, we have that the sequence converges:

(A) In distribution to $X = 2$ (B) In distribution to $X = 3$ (C) It does not converge

(D) In distribution to $X = 0$ (E) In distribution to $X = 1$

Questions 11 and 12 refer to the following exercise:

Let $X_1, \ldots, X_5$ be five independent and identically distributed r.v. having a $\gamma(0.25, 1)$ distribution.

11. The distribution of the r.v. $Z = X_1 + X_2 + X_3 + X_4 + X_5$ is:

(A) $\gamma(1.25, 5)$ (B) $\exp(\lambda = 1.25)$ (C) $\chi^2_{\overline{10}|}$ (D) $\gamma(0.25, 5)$ (E) $\exp(\lambda = 0.25)$

12. The distribution of the r.v. $\overline{X} = (X_1 + X_2 + X_3 + X_4 + X_5)/5$ is:

(A) $\chi^2_{\overline{10}|}$ (B) $\exp(\lambda = 1.25)$ (C) $\exp(\lambda = 0.25)$ (D) $\gamma(0.25, 5)$ (E) $\gamma(1.25, 5)$

13. Let X be a random variable having a Student's t distribution with n degrees of freedom, $t_{\overline{n}|}$. Then, $P(t_{\overline{n}|1-4\alpha} < X < t_{\overline{n}|3\alpha})$ is:

(A) $1 - 7\alpha$ (B) $1 - 3\alpha$ (C) 7α (D) $1 - 4\alpha$ (E) All false

14. Let X be a normal r.v. having mean zero and variance a, with $a > 0$. The value of $P(X^2 < 1.32a)$ is:

(A) 0.25 (B) 0.91 (C) 0.84 (D) 0.09 (E) 0.75

15. Let X be a random variable having an $\mathcal{F}_{9,12}$ distribution. Then, the approximate value of k, $k > 0$, such that $P(X > k) = 0.90$ is:

(A) 0.33 (B) 2.38 (C) 3.07 (D) 5.11 (E) 0.42

Questions 16 to 18 refer to the following exercise:

Let $X_1, \ldots, X_n$ be a r.s. taken from a population with probability mass function given by:

$$P(1) = P(2) = \theta, \quad P(3) = 1 - 2\theta$$

16. The method of moments estimator of θ is:

(A) $\overline{X}$ (B) $\frac{2\overline{X}-3}{6}$ (C) All false (D) $\frac{3-\overline{X}}{3}$ (E) $\frac{\overline{X}+1}{4}$

17. Is this estimator unbiased?

(A) No (B) - (C) Yes (D) - (E) -

18. Is this estimator consistent?

(A) - (B) - (C) No (D) Yes (E) -

Questions 19 to 22 refer to the following exercise:

Let X be a r.v. having a $\gamma(a, r)$ distribution; that is, with probability density function given by

$$f(x; a, r) = \frac{a^r}{\Gamma(r)} x^{r-1} e^{-ax}, \qquad x > 0, \qquad a, r > 0$$

and $X_1, \ldots, X_n$ be a r.s. of size n taken from this distribution.

19. If the value of a is known, the method of moments estimator of r is:

(A) $\frac{\overline{X}}{a}$ (B) $\frac{\overline{X}}{2}$ (C) $\overline{X}$ (D) $a\overline{X}$ (E) $\frac{a}{\overline{X}}$

20. If the value of r is known, the method of moments estimator of a is:

(A) $\overline{X}$ (B) $\frac{\overline{X}}{2}$ (C) $\frac{1}{\overline{X}}$ (D) $\frac{r}{\overline{X}}$ (E) $\frac{\overline{X}}{r}$

21. In the previous question, the expected value of the method of moments estimator of a is:

(A) $\mathrm{E}(\overline{X})$ (B) $r\mathrm{E}\left(\frac{1}{\overline{X}}\right)$ (C) $\frac{\mathrm{E}(\overline{X})}{r}$ (D) $r\mathrm{E}(\overline{X})$ (E) $\frac{\mathrm{E}(\overline{X})}{2}$

22. If the value of r is known, the maximum likelihood estimator of a is:

(A) $\overline{X}$ (B) $\frac{\overline{X}}{2}$ (C) $\frac{1}{\overline{X}}$ (D) $\frac{r}{\overline{X}}$ (E) $\frac{\overline{X}}{r}$

Questions 23 to 25 refer to the following exercise:

A board game (i.e., parchís) player believes that his opponent's die obtains more fives and sixes than expected in a regular die. In order to find some evidence about his suspicion, he decides to write down all results for the next 15 die-throws and to accuse him of cheating if he obtains a 5 or a 6 at least 8 times. We will use as null hypothesis the one indicating that we have a regular die and F_p to denote the cumulative distribution function for a binomial r.v. with parameter p, and a sample size of $n = 15$. In addition, we assume that each die-throw is a binary r.v. taking value one with probability p if the result is a 5 or a 6.

23. The probability of making a type I error is:

(A) $1 - F_{1/3}(7)$ (B) $1 - F_{1/6}(7)$ (C) All false (D) $F_{1/6}(8)$ (E) $F_{1/3}(8)$

24. If the die is not regular, so that $P(5) = 0.30$ and $P(6) = 0.20$, the probability of making a type II error is:

(A) All false (B) $1 - F_{0.5}(7)$ (C) $F_{0.3}(7)$ (D) $F_{0.2}(7)$ (E) $F_{0.5}(7)$

25. The power for $P(5) = p_5$ and $P(6) = p_6$ (with $p_5, p_6 > \frac{1}{3}$), is:

(A) $1 - F_{p_5+p_6}(7)$ (B) $1 - F_{p_5}(7)$ (C) All false (D) $1 - F_{p_6}(7)$ (E) $F_{p_5+p_6}(7)$

Questions 26 and 27 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with parameter λ. In order to test $H_0 : \lambda = 1$ against $H_1 : \lambda < 1$, a r.s. of size $n = 4$ has been taken and it is decided that the null hypothesis is rejected if $T = \sum_{i=1}^{4} X_i \leq 1$.

26. The significance level α for this test is:

(A) 0.0516 (B) 0.0750 (C) 0.3123 (D) 0.0916 (E) 0.0191

27. The power for $\lambda = 0.5$ is:

(A) 0.4060 (B) 0.3033 (C) 0.9098 (D) 0.6065 (E) 0.5940

28. In a manufacturing process, we wish to test the null hypothesis that the percentage of acceptable manufactured parts is 10% against the alternative that it is not. We have taken a random sample of size $n = 400$ observing 34 acceptable parts. At the approximate 10% significance level, the decision of the test will be:

(A) Reject the null hypothesis (B) - (C) -
(D) - (E) Do not reject the null hypothesis

29. Let X and Y be two independent r.v. such that $X \in N(m_X, \sigma_X^2 = 25)$ and $Y \in N(m_Y, \sigma_Y^2 = 25)$. In order to test $H_0 : m_X = m_Y$ against $H_1 : m_X \neq m_Y$, we have taken two separate r.s. of size 30 each, so that $\overline{x} = 120$ and $\overline{y} = 115$. At the 10% significance level, the decision of the test will be:

(A) Do not reject H_0 (B) - (C) Reject H_0 (D) - (E) -

30. It is known that the expense in the school bus per family in a given population follows a normal distribution and it is claimed that the variance of these expenses made by families having more than one child, σ_1^2, is smaller than or equal than the variance in these expenses made by families having only one child, σ_2^2. In order to test $H_0 : \sigma_1^2 \leq \sigma_2^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$, we have taken two separate r.s. of size 21 in each of the two populations, so that $s_1^2 = 55$ and $s_2^2 = 42$. At the 5% significance level, the decision of the test will be:

(A) Reject H_0 (B) - (C) - (D) - (E) Do not reject H_0

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

In a given study on immigration in Spain, the following classification has been established using as a basis for it the immigrant's origin: African countries, Eastern European countries, South American countries, and rest of the world. In this study, the starting hypothesis is that the probabilities that an immigrant comes from African countries, from Eastern European countries, and from the rest of the world are each equal to $\frac{1}{5}$, whereas the probability that he comes from South American countries is twice this probability, that is, $\frac{2}{5}$. In order to test this hypothesis a r.s. of size 500 has been taken providing the following observed frequencies:

Immigrant's origin	n_i
Africa	120
Eastern Europa	130
South America	160
Rest of the world	90

i) At the 5% significance level, test the starting aforementioned hypothesis stated in the study.

ii) From the information provided in the whole sample, obtain a 95% confidence interval for the proportion of immigrants coming from African countries.

B. (10 points, 25 minutos)

Let $X_1, \ldots, X_n$ be a r.s. taken from a population having an exponential distribution with parameter $\lambda > 0$. We wish to test the null hypothesis $H_0 : \lambda = \lambda_0$ against the alternative hypothesis $H_1 : \lambda = \lambda_1$, where $\lambda_1 < \lambda_0$. Find (providing all relevant details), using the likelihood ratio test, the best critical region for this test.

C. (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a r.s. of size n taken from a r.v. X having a probability density function given by:

$$f(x, \alpha, \theta) = \frac{\alpha}{\theta^\alpha}\, x^{\alpha-1} \qquad 0 < x \leq \theta, \qquad \theta, \alpha > 0$$

i) Assuming that θ is known, obtain, **providing all relevant details**, the method of moments estimator for the parameter α.

ii) Assuming that θ is known, obtain, **providing all relevant details**, the maximum likelihood estimator for the parameter α.

iii) Assuming that α is known, obtain, **providing all relevant details**, the maximum likelihood estimator for the parameter θ.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

1: C	11: D	21: B
2: B	12: E	22: D
3: C	13: A	23: A
4: E	14: E	24: E
5: E	15: E	25: A
6: A	16: D	26: D
7: B	17: C	27: A
8: B	18: D	28: E
9: C	19: D	29: C
10: D	20: D	30: E

SOLUTIONS TO EXERCISES

Exercise A

i) We have to carry out a χ^2 goodness of fit test to a completely specified distribution. More especifically, we have to test

H_0 : P(Africa) = P(Eastern Europe) = P(Rest of the World) = 0.2, P(South America)=0.4. Using the information available in the sample, we build the table

Immigrant's origin	n_i	p_i	np_i	$\frac{(n_i-np_i)^2}{np_i}$
Africa	120	0.2	100	4
Eastern Europe	130	0.2	100	9
South America	160	0.4	200	8
Rest of the World	90	0.2	100	1
	$n = 500$	1	$n = 500$	$z = 22$

We know that, under the null hypothesis, the test statistic Z converges to a χ^2_{k-1} distribution, where k is the number of classes in which the sample has been divided (i.e., $k = 4$). The decision rule will be that, at the approximate 5% significance level, we reject the null hypothesis if

$$z > \chi^2_{3,0.05}$$

Given that

$$z = 22 > \chi^2_{3,0.05} = 7.81,$$

the null hypothesis is rejected.

ii) From the table above, we have that the number of immigrants in the sample, whose origin was Africa, is $z = 120$ from a total of $n = 500$. Therefore, the approximate 95% confidence interval for the proportion of immigrants whose origin is Africa will be:

$$CI_{0.95}(p) = \left(\frac{z}{n} \pm t_{\alpha/2}\sqrt{\frac{z \times (n-z)}{n^3}}\right) = \left(\frac{120}{500} \pm 1.96\sqrt{\frac{120 \times 380}{500^3}}\right) = (0.203, 0.277)$$

Exercise B

Let $X_1, \ldots, X_n$ be a r.s. of size n taken from a population having an exponential distribution with parameter $\lambda > 0$. We wish to test the null hypothesis $H_0 : \lambda = \lambda_0$ against the alternative hypothesis $H_1 : \lambda = \lambda_1$, with $\lambda_1 < \lambda_0$.

In order to obtain the form of the best critical region for this test by using the likelihood ratio test, we will have that the corresponding likelihood functions under the null and alternative hypotheses will be given by:

$$L(\vec{x}; \lambda_0) = \left(\lambda_0 e^{-\lambda_0 x_1}\right) \cdots \left(\lambda_0 e^{-\lambda_0 x_n}\right) = \lambda_0^n \; e^{-\lambda_0 \sum_{i=1}^n x_i},$$

and

$$L(\vec{x}; \lambda_1) = \left(\lambda_1 e^{-\lambda_1 x_1}\right) \cdots \left(\lambda_1 e^{-\lambda_1 x_n}\right) = \lambda_1^n \; e^{-\lambda_1 \sum_{i=1}^n x_i},$$

respectively. Therefore, if we use the likelihood ratio test, we will have that:

$$\frac{L(\vec{x}; \lambda_0)}{L(\vec{x}; \lambda_1)} = \frac{(\lambda_0)^n \; e^{-\lambda_0 \sum_{i=1}^n x_i}}{(\lambda_1)^n \; e^{-\lambda_1 \sum_{i=1}^n x_i}} \leq K, \; K > 0$$

$$\Longrightarrow \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{(\lambda_1 - \lambda_0) \sum_{i=1}^n x_i} \leq K \Longrightarrow e^{(\lambda_1 - \lambda_0) \sum_{i=1}^n x_i} \leq K_1, \; K_1 > 0$$

If we take natural logarithms, we will have that:

$$(\lambda_1 - \lambda_0) \sum_{i=1}^n x_i \leq K_2, \; K_2 > 0$$

Now, as $\lambda_1 < \lambda_0$, we have that $(\lambda_1 - \lambda_0)$ is negative, so that the last inequality changes and, thus, we will then have that the decision rule will be to reject the null hypothesis if $Z = \sum_{i=1}^n X_i \geq C$. The corresponding best critical region for this test will be then given by $\mathrm{CR} = [C, +\infty)$.

Exercise C

The probability density function for the r.v. X is:

$$f(x, \alpha, \theta) = \frac{\alpha}{\theta^\alpha} \; x^{\alpha-1} \qquad 0 < x \leq \theta, \qquad \theta, \alpha > 0$$

i) If θ is known and in order to obtain the method of moments estimator for α, we have to compute the first population moment, $\alpha_1 = \mathrm{E}(X)$:

$$\alpha_1 = \mathrm{E}(X) = \int_0^\theta x \left(\frac{\alpha}{\theta^\alpha} \; x^{\alpha-1} \right) dx = \left(\frac{\alpha}{\theta^\alpha} \right) \left[\frac{x^{\alpha+1}}{(\alpha+1)} \right|_0^\theta = \frac{\alpha\theta}{(\alpha+1)}$$

We now make the first population moment equal to the first sample moment. That is,

$$\alpha_1 = \mathrm{E}(X) = a_1 \Longrightarrow \frac{\alpha\theta}{(\alpha+1)} = \overline{X} \Longrightarrow \alpha(\theta - \overline{X}) = \overline{X} \Longrightarrow \hat{\alpha}_{\mathrm{MM}} = \frac{\overline{X}}{(\theta - \overline{X})}$$

ii) If θ is known, in order to obtain the maximum likelihood estimator for α, we have that the corresponding likelihood function is given by:

$$L(\alpha) = f(x_1, \alpha) \; \ldots \; f(x_n, \alpha) = \left(\frac{\alpha}{\theta^\alpha} \; x_1^{\alpha-1} \right) \cdots \left(\frac{\alpha}{\theta^\alpha} \; x_n^{\alpha-1} \right) = \frac{\alpha^n}{\theta^{n\alpha}} \left[\prod_{i=1}^n x_i \right]^{\alpha-1}$$

If we take its natural logarithms, we have that:

$$\ln L(\alpha) = n \ln(\alpha) - n\alpha \ln(\theta) + (\alpha - 1) \ln \left[\prod_{i=1}^n x_i \right]$$

If we now take its derivative with respect to α, and make it equal to zero, we have that:

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha} - n \ln(\theta) + \ln \left[\prod_{i=1}^n x_i \right] = 0$$

so that,

$$\frac{n}{\alpha} = n \ln(\theta) - \ln \left[\prod_{i=1}^n x_i \right] \Longrightarrow \hat{\alpha}_{\mathrm{ML}} = \frac{n}{\{ n \ln(\theta) - \ln \left[\prod_{i=1}^n x_i \right] \}}$$

iii) If α is known, in order to obtain the the maximum likelihood estimator for θ, we have that the corresponding likelihood function is given by:

$$L(\alpha) = f(x_1, \alpha) \; \ldots \; f(x_n, \alpha) = \left[\frac{\alpha}{\theta^\alpha} \; x_1^{\alpha-1} \right] \cdots \left[\frac{\alpha}{\theta^\alpha} \; x_n^{\alpha-1} \right] = \frac{\alpha^n}{\theta^{n\alpha}} \left[\prod_{i=1}^n x_i \right]^{\alpha-1},$$

for $0 < x_i \leq \theta$, $i = 1, \ldots, n$. From the above, we have that the likelihood will be different from zero only if, for all $i = 1, \ldots, n$, $0 < x_i \leq \theta$, which holds if $0 < \max(x_i) \leq \theta$. Therefore, the maximum likelihood estimator of θ should verify that $\hat{\theta}_{\mathrm{ML}} \geq \max(x_i)$, so that the minimum value this estimator can take so that this holds is equal to $\max(x_i)$. Thus, the maximum likelihood estimator of θ will be $\hat{\theta}_{\mathrm{ML}} = \max(x_i)$.