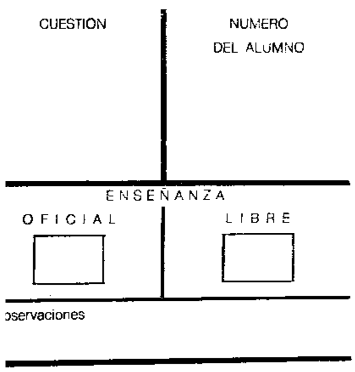
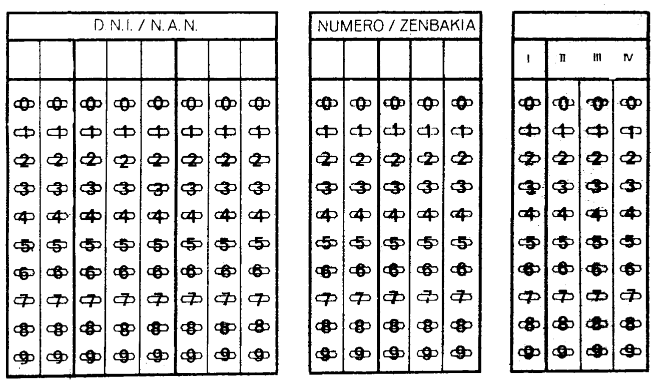
INSTRUCTIONS

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.
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 45 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 weight in kilograms for calves that are sold in a given farm follows a normal $N(250, \sigma^2 = 36)$ distribution. Calves weighing at least 257.68 kilograms are considered suitable to be sold in the farm. We assume independence between the weights for the different calves. **Remark**: When computing the probability of a calf being suitable to be sold in the farm, you should round this probability up to one decimal place.

2. If a random sample of 15 calves is taken from that farm, the probability that at least 4 of them are suitable to be sold in the farm is:

   (A) 0.9873 (B) 0.9444 (C) 0.0556 (D) 0.1285 (E) 0.0127

3. In the same sample of 15 calves, the probability that exactly 11 calves **are not** suitable to be sold in the farm is:

   (A) 0.0429 (B) 0.9873 (C) 0 (D) 1 (E) 0.0127

4. If we now take a random sample of 50 calves from that farm, the approximate probability that at most 6 of them are suitable to be sold in the farm is:

   (A) 0.3840 (B) 0.6160 (C) 0.7622 (D) 0.8666 (E) 0.2378

5. If we now take a random sample of 200 calves from that farm, the approximate probability that at least 25 of them are suitable to be sold in the farm is:

   (A) 0.5793 (B) 0.8810 (C) 0.4207 (D) 0.1446 (E) 0.8554

Questions 6 and 7 refer to the following exercise:

Let $X_1, \ldots, X_{20}$ be independent and identically distributed r.v. having a Poisson distribution and such that their modes are 4 and 5.

6. $P(X_1 > 6)$ is equal to:

   (A) 0.2378 (B) 0.8893 (C) 0.6160 (D) 0.1107 (E) 0.7622

7. If we define the random variable $Y = \sum_{i=1}^{20} X_i$, then $P(Y < 110)$ is, approximately equal to:

   (A) 0.17 (B) 1 (C) 0.83 (D) 0.68 (E) 0.32

8. If we have a r.v. X having a Poisson distribution with variance equal to 6, we have that:

   (A) $P(X = 6) = P(X = 7)$ (B) $P(X = 6) < P(X = 7)$ (C) All false
   (D) $P(X = 5) > P(X = 6)$ (E) $P(X = 6) = P(X = 5)$

Questions 9 and 10 refer to the following exercise:

Let X and Y be independent r.v. such that $X \in \chi^2_5$ and $Y \in \chi^2_{15}$.

9. The probability that the r.v. $Z = X + Y$ takes on values smaller than 23.8 is:

(A) 0.95 (B) 0.50 (C) 0.90 (D) 0.75 (E) 0.25

10. If we define the r.v. $V = \frac{Y}{3X}$. The value of k such that $P(V > k) = 0.90$ is equal to:

(A) 2.27 (B) 4.56 (C) 3.24 (D) 0.31 (E) 0.44

11. Let X be a random variable having a Student's t distribution with n degrees of freedom, $t_{\overline{n}|}$. We have that $P(t_{\overline{n}|1-\frac{\alpha}{2}} < X < t_{\overline{n}|\frac{3\alpha}{2}})$ is equal to:

(A) $1 - 2\alpha$ (B) $1 - 3\alpha$ (C) 2α (D) $1 - \alpha$ (E) α

12. Let X be a normal random variable with mean zero and variance equal to 2. We have that $P(X^2 < 7.68)$ is equal to:

(A) 0.05 (B) 0.75 (C) 0.995 (D) 0.005 (E) 0.95

Questions 13 to 17 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:

$$P(X = 1) = P(X = 2) = 2\theta, \quad P(X = 3) = 1 - 4\theta$$

13. The method of moments estimator of θ is:

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

14. Is this an unbiased estimator of θ?

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

15. Is this a consistent estimator of θ?

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

16. In order to be able to obtain an estimate of the parameter θ, a r.s. has been taken providing the following results: 1, 1, 1, 2, 2, 2, 3, 3, 3, 3. A method of moments estimate of θ is equal to:

(A) 0.10 (B) 0.85 (C) 0.20 (D) 0.25 (E) 0.15

17. For the same sample in the previous question, a maximum likelihood estimate of θ is equal to:

(A) 0.15 (B) 0.85 (C) 0.10 (D) 0.25 (E) 0.20

Questions 18 to 21 refer to the following exercise:

Let X be a r.v. having a gamma $\gamma(\frac{1}{\theta}, 4)$ distribution, so that its probability density function is given by

$$f(x, \theta) = \begin{cases} \frac{x^3}{6\theta^4}\, e^{-\frac{x}{\theta}} & \text{if } x \geq 0,\ \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

We wish to estimate the parameter θ and, thus, a r.s. of size n, $X_1, X_2, \ldots, X_n$, has been taken.

18. The maximum likelihood estimator of θ, $\hat{\theta}_{ML}$, will be:

(A) $\overline{X}$ (B) $\frac{\overline{X}}{4}$ (C) $\frac{\overline{X}}{2}$ (D) $4n\overline{X}$ (E) $4\overline{X}$

19. The method of moments estimator of θ, $\hat{\theta}_{MM}$, will be:

(A) $4\overline{X}$ (B) $\frac{\overline{X}}{4}$ (C) $\overline{X}$ (D) $4n\overline{X}$ (E) $\frac{\overline{X}}{2}$

20. Is the method of moments estimator an unbiased estimator of θ?

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

21. Is the method of moments estimator a consistent estimator of θ?

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

Questions 22 and 23 refer to the following exercise:

Let X be a r.v. with probability density function given by:

$$f(x,\theta) = \frac{1}{\theta}\, e^{-\frac{x}{\theta}}, \quad x \geq 0, \quad \theta > 0$$

Based on the information provided by a r.s. of size $n = 1$, X_1, we wish to test the null hypothesis $H_0 : \theta = 1$ against the alternative hypothesis $H_1 : \theta = 2$.

22. The most powerful critical region for this test, and for X_1, will be of the form:

(A) $[C, \infty)$ (B) $[C_1, C_2]^C$ (C) All false (D) $[C_1, C_2]$ (E) $[0, C]$

23. At the $\alpha = 0.05$ significance level, the power for this test is, approximately equal to:

(A) 0.7764 (B) 0.9512 (C) 0.2236 (D) 0.0488 (E) 0.3678

Questions 24 and 25 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with parameter λ. In order to test $H_0 : \lambda = 0.50$ against $H_1 : \lambda > 0.50$, a r.s. of size $n = 8$ has been taken and the test statistic $Z = \sum_{i=1}^{8} X_i$ is used.

24. At the $\alpha = 0.10$ significance level, the test will reject the null hypothesis if:

(A) $Z \geq 6$ (B) $Z \geq 8$ (C) $Z \geq 7$ (D) $Z \leq 7$ (E) $Z \leq 8$

25. For the previous critical region and $\lambda = 1$, the power for this test is equal to:

(A) 0.4075 (B) 0.3134 (C) 0.5925 (D) 0.4530 (E) 0.5470

Questions 26 and 27 refer to the following exercise:

The professors teaching a given course are interested in knowing if the mean grades obtained by their students are different for males and females. In order to do so, two r.s. of students from each gender are taken and average grades students obtained in that course are computed, providing the following results: for males ($n_M = 16$): $\bar{x}_M = 5.2$ and $s_M = 0.8$, whereas for females ($n_F = 16$): $\bar{x}_F = 5.8$ and $s_F = 1$. We assume normality for the distribution of grades and independence between grades obtained by males and females.

26. A 90% confidence interval for the ratio of variances σ_M^2/σ_F^2 is:

(A) $(0.2667, 1.5360)$ (B) $(0.1818, 2.2528)$ (C) $(0.4456, 1.7565)$ (D) $(0.1538, 1.0456)$ (E) $(0.3249, 1.2608)$

27. If, at the 10% significance level, we wish to test the null hypothesis that the variances for the grades obtained by males and females are equal, the decision for the test will be:

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

Questions 28 and 29 refer to the following exercise:

In a given city we wish to estimate the mean price of houses. We have evidence that this variable may behave in a different way for different neighborhoods and that, within each of these neighborhoods, house prices are homogeneous. Let A and B be two specific neighborhoods under study, each having the same number of houses; that is, $N_A = N_B$. In addition, it is known that, from information obtained from a previous study carried out in these neighborhoods, $\sigma_A^2 > \sigma_B^2$.

28. If we use proportional allocation, the sample sizes n_A and n_B will be such that:

(A) $n_A < n_B$ (B) All false (C) No decision can be adopted (D) $n_A > n_B$ (E) $n_A = n_B$

29. If we use n-optimal allocation, the sample sizes n_A and n_B will be such that:

(A) $n_A = n_B$ (B) All false (C) $n_A < n_B$ (D) $n_A > n_B$ (E) No decision can be adopted

30. A publishing firm wishes to estimate the proportion of libraries that, in a population of 5000 libraries, would be interested in buying a new dictionary of specialized Economic terms. The firm wishes to estimate this proportion with a 95% confidence level and an absolute error of ± 0.03. What would be the minimum number of libraries that would need to be selected if simple random sampling with replacement is used?

(A) 880 (B) 543 (C) 1068 (D) 733 (E) 147

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

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

$$f(x, \alpha, \theta) = \frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x^{\frac{\alpha}{4}-1} \qquad 0 < x \leq \theta, \qquad \theta, \alpha > 0,$$

and such that its mean is $m = \frac{\theta\alpha}{(\alpha+4)}$.

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

ii) If it is assumed that θ is known, obtain, **providing all relevant details**, the maximum likelihood estimator of the parameter α.

iii) If it is assumed that α is known, obtain, **providing all relevant details**, the maximum likelihood estimator of the parameter θ.

B. (10 points, 25 minutes)

A sample of 500 firms has been classified according to the time they have been in business (more than 10 years and 10 or fewer years) and their size (number of employees smaller than or equal to 50 and larger than 50), providing the following information:

	≤ 50 employees	> 50 employees
≤ 10 years	100	105
> 10 years	95	200

Based on this information, and at the 5% significance level, carry out a test of independence between the variables time in business and firm size.

C. (10 points, 25 minutes)

The following table includes information on the probability mass function a discrete r.v. X has under the null hypothesis ($P_0(x)$) and under the alternative hypothesis ($P_1(x)$).

X	1	2	3	4	5	6
$P_0(x)$	0.05	0.45	0	0.10	0.40	0
$P_1(x)$	0.15	0	0.35	0.20	0	0.30

A random sample of size $n = 1$ will be used to test the null hypothesis $H_0 : P(x) = P_0(x)$ against the alternative hypothesis $H_1 : P(x) = P_1(x)$.

i) Would you include the points $X = \{3, 6\}$ in the critical region? Explain why or why not.

ii) Would you include the points $X = \{2, 5\}$ in the critical region? Explain why or why not.

iii) At the 10% significance level and providing all relevant details used to obtain the requires response, find the most powerful critical region for this test. **Remark**: Before providing an answer to this item, take into account your responses to the previous items in this exercise.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

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

$$f(x,\alpha,\theta)=\frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x^{\frac{\alpha}{4}-1}\qquad 0<x\leq\theta,\qquad \theta,\alpha>0,$$

i) If θ is known, in order to be able to obtain the method of moments estimator of the parameter α, we need to equate the first population moment $\alpha_1=\mathrm{E}(X)=m=\frac{\theta\alpha}{(\alpha+4)}$ to the first sample moment. That is,

$$\alpha_1=\mathrm{E}(X)=a_1\Longrightarrow\frac{\alpha\theta}{(\alpha+4)}=\overline{X}\Longrightarrow\alpha(\theta-\overline{X})=4\overline{X}\Longrightarrow\hat{\alpha}_{\mathrm{MM}}=\frac{4\overline{X}}{(\theta-\overline{X})}$$

ii) If θ is known, in order to be able the maximum likelihood estimator of the parameter α, we have that likelihood function is given by:

$$L(\vec{x},\alpha)=f(x_1,\alpha)\ \ldots\ f(x_n,\alpha)=\left(\frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x_1^{\frac{\alpha}{4}-1}\right)\cdots\left(\frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x_n^{\frac{\alpha}{4}-1}\right)=\frac{\alpha^n}{4^n\theta^{\frac{n\alpha}{4}}}\left(\prod_{i=1}^{n}x_i\right)^{\frac{\alpha}{4}-1}$$

If we take its natural logarithm, we have that:

$$\ln L(\vec{x},\alpha)=n\ln(\alpha)-n\ln(4)-\left(\frac{n\alpha}{4}\right)\ln(\theta)+\left(\frac{\alpha}{4}-1\right)\ln\left[\prod_{i=1}^{n}x_i\right]$$

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

$$\frac{\partial\ln L(\vec{x},\alpha)}{\partial\alpha}=\frac{n}{\alpha}-\frac{n}{4}\ln(\theta)+\frac{\ln\left[\prod_{i=1}^{n}x_i\right]}{4}=0$$

so that,

$$\frac{n}{\alpha}=\left(\frac{n}{4}\right)\ln(\theta)-\frac{\ln\left[\prod_{i=1}^{n}x_i\right]}{4}\Longrightarrow\hat{\alpha}_{\mathrm{ML}}=\frac{4n}{\{n\ln(\theta)-\ln\left[\prod_{i=1}^{n}x_i\right]\}}$$

iii) If α is known, in order to be able to obtain the maximum likelihood estimator of the parameter θ, we have that the likelihood function is given by:

$$L(\vec{x},\theta)=f(x_1,\theta)\ \ldots\ f(x_n,\theta)=\left(\frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x_1^{\frac{\alpha}{4}-1}\right)\cdots\left(\frac{\alpha}{4\theta^{\frac{\alpha}{4}}}\ x_n^{\frac{\alpha}{4}-1}\right)=\frac{\alpha^n}{4^n\theta^{\frac{n\alpha}{4}}}\left(\prod_{i=1}^{n}x_i\right)^{\frac{\alpha}{4}-1},$$

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

Exercise B

The information available for this exercise is included in the following table:

	≤ 50 employees	> 50 employees	
≤ 10 years	100	105	205
> 10 years	95	200	295
	195	305	$n = 500$

We should perform a test of independence. More specifically, we have to test the null H_0: the time the firms have been in business and their size are independent variables, against the alternative hypothesis H_1: They are not. First of all, we have to estimate the probabilities for each one of the classes for each one of the two variables whose independence we wish to test. That is:

$$P(\leq 10) = 205/500 = 0.41, \quad P(> 10) = 295/500 = 0.59$$

$$P(\leq 50) = 195/500 = 0.39, \quad P(> 50) = 305/500 = 0.61$$

Under the null hypothesis of independence, we have that the joint probability will be equal to the product of its corresponding marginal probabilities. That is, $H_0 : p_{ij} = p_{i\bullet} \times p_{\bullet j}$. Therefore, with the information provided by the sample and the estimated marginal probabilities above, we build the table that will allow us to perform the test.

Class	n_{ij}	$\hat{p}_{ij}$	$n\hat{p}_{ij}$	$\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$
(≤ 10 years, ≤ 50 employees)	100	$0.41 \times 0.39 = 0.1599$	79.95	5.0282
(≤ 10 years, > 50 employees)	105	$0.41 \times 0.61 = 0.2501$	125.05	3.2147
(> 10 years, ≤ 50 employees)	95	$0.59 \times 0.39 = 0.2301$	115.05	3.4942
(> 10 years, > 50 employees)	200	$0.59 \times 0.61 = 0.3599$	179.95	2.2340
	$n = 500$	1	$n = 500$	$z = 13.9711$

We know that, under the null hypothesis, the test statistic Z converges to a $\chi^2_{(I-1)(J-1)}$ distribution, where I is the number of rows in the table (i.e., $I = 2$) and J is the number of columns in the table (i.e., $J = 2$). At the 5% significance level, the decision will be to reject the null hypothesis if

$$z > \chi^2_{1,0.05}$$

Given that

$$z = 13.9711 > \chi^2_{1,0.05} = 3.84,$$

we reject the null hypothesis and, thus, the time firms have been in business and their size are not independent variables.

Exercise C

We wish to test the null hypothesis that X is a discrete r.v. with probability mass function $P_0(x)$ against the alternative hypothesis that its probability mas function is $P_1(x)$:

X	1	2	3	4	5	6
$P_0(x)$	0.05	0.45	0	0.10	0.40	0
$P_1(x)$	0.15	0	0.35	0.20	0	0.30

We have taken a r.s. of size $n = 1$; that is, we observe X.

i) Would you include the points $X \in \{3, 6\}$ in the critical region?

Given that, for the probability mass function $P_0(x)$ under the null hypothesis, these points have probability zero, the r.v. can not take on these values under the null hypothesis. Therefore, the points $X \in \{3, 6\}$ are points that lead to the rejection of H_0 and, thus, they should **always** be included in the critical region.

ii) Would you include the points $X \in \{2, 5\}$ in the critical region?

Given that, for the probability mass function $P_1(x)$ under the alternative hypothesis, these points have probability zero, the r.v. can not take on these values under the alternative hypothesis, but it can take on them under the null hypothesis. Therefore, the points $X \in \{2, 5\}$ are points that lead to the no rejection of H_0 and, thus, they should **never** be included in the critical region.

iii) At the $\alpha = 0.10$ significance level and **using the conclusions we have arrived at in the previous items of this exercise**, we have that the two possible critical regions for this test are $\mathrm{CR}_1 = \{1, 3, 6\}$ and $\mathrm{CR}_2 = \{3, 4, 6\}$. This is because, for the type I error for each of them, we have that:

$$P(X \in \mathrm{CR}_1 | P_0) = P(X = 1, 3, 6 | P_0) = 0.05 + 0 + 0 = 0.05 \leq \alpha = 0.10$$
$$P(X \in \mathrm{CR}_2 | P_0) = P(X = 3, 4, 6 | P_0) = 0 + 0.10 + 0 = 0.10 \leq \alpha = 0.10$$

In order to see which one of these two critical regions is the most powerful one for this test, we compute their corresponding powers:

$$\mathrm{Power}_1 = P(X \in \mathrm{CR}_1 | P_1) = P(X = 1, 3, 6 | P_1) = 0.15 + 0.35 + 0.30 = 0.80$$
$$\mathrm{Power}_2 = P(X \in \mathrm{CR}_2 | P_1) = P(X = 3, 4, 6 | P_1) = 0.35 + 0.20 + 0.30 = 0.85$$

From the above, we conclude that, at the $\alpha = 0.10$ significance level, the critical region CR_2 is the most powerful one for this test.