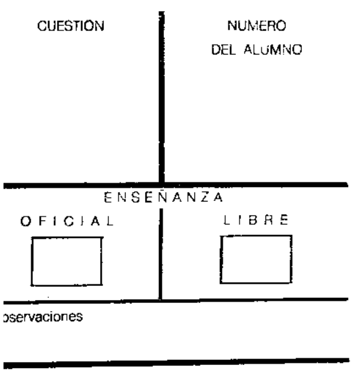
STATISTICS APPLIED TO BUSINESS ADMINISTRATION - Second Year Academic Year 2013-14
First Call. June 3, 2014

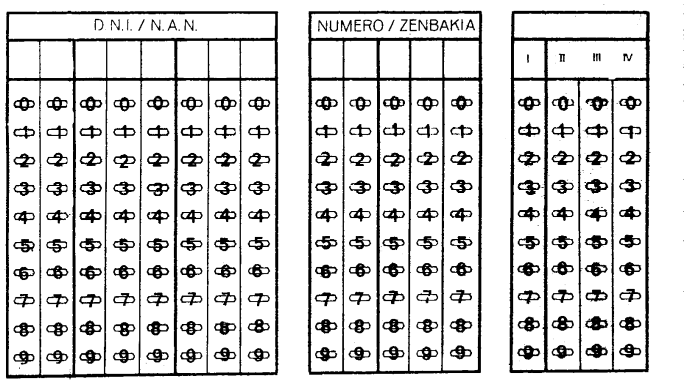
INSTRUCTIONS

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. If you have participated in the on-going evaluation process, you will need to obtain 12 points in each part of the exam to pass it. Otherwise, 15 points points in each part of the exam are required to pass it.
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:

12545 PEREZ, Ernesto Exam type 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 4 refer to the following exercise:

The probability that an individual has a given illness is 0.08. We assume independence between the different individuals in the population.

2. If a r.s. of 5 individuals is taken, the probability that only one of them has the illness is:

   (A) 0.0573 (B) 0.9427 (C) 0.3249 (D) 0.2866 (E) 0.7134

3. If a r.s. of 70 individuals is taken, the approximate probability that three of them have the illness is:

   (A) 0.1082 (B) 0.3506 (C) 0.6494 (D) 0.8918 (E) 0.2742

4. If a r.s. of 500 individuals is taken, the approximate probability that at most 51 of them have the illness is:

   (A) 0.9713 (B) 0.0409 (C) 0.0287 (D) 0.9591 (E) 0.8413

Questions 5 and 6 refer to the following exercise:

Let Z be a r.v. such that it follows a $b(p, n)$ binomial distribution, with mean 6 and variance 3.6.

5. The distribution of the r.v. Z is:

   (A) $b(0.6, 10)$ (B) $b(0.6, 15)$ (C) $b(0.4, 15)$ (D) $b(0.4, 10)$ (E) $b(0.3, 20)$

6. The probability $P(Z = 4)$ is equal to:

   (A) 0.2173 (B) 0.0074 (C) 0.1268 (D) 0.0634 (E) 0.9981

7. Let X be a r.v. having a Poisson distribution such that $P(X = 0) = 0.3679$. The value of the variance of the r.v. X is:

   (A) 1 (B) 2.4 (C) 2 (D) 1.5 (E) 1.8

Questions 8 to 10 refer to the following exercise:

In a given retail store, the number of clients entering the store each hour follows a Poisson distribution with parameter $\lambda = 5$.

8. The probability that, in a given hour, exactly 4 clients enter the store is:

   (A) 0.4405 (B) 0.2650 (C) 0.6160 (D) 0.1755 (E) 0.1404

9. The probability that, in a given hour, more than 5 clients enter the store is:

   (A) 0.2378 (B) 0.3840 (C) 0.5595 (D) 0.6160 (E) 0.4405

10. If we assume independence between the arrival of clients at different hours, the approximate probability that, in a 20-hour period, there are exactly 5 hours in which more than 5 clients enter the store is:

(A) 0.1408 (B) 0.6160 (C) 0.0903 (D) 0.2378 (E) 0.0453

11. Let X be a r.v. such that its characteristic function is given by $\Psi_X(u) = (1 - iu)^{-1}$. In this case, we have that the value of $P(-1 < X < 2)$ is equal to:

(A) 0.8647 (B) 0.7675 (C) 0.1353 (D) 0.4483 (E) 0.2325

12. Let X be a r.v. having a Student's t distribution with $n = 10$ degrees of freedom, $t_{\overline{10|}}$. For this r.v., $P(-2.23 < X < 1.81)$ is equal to:

(A) 0.075 (B) 0.900 (C) 0.850 (D) 0.150 (E) 0.925

13. Let Y be a normal r.v. with mean zero and variance 4. The value of $P(0.408 < Y^2 < 10.84)$ is equal to:

(A) 0.90 (B) 0.65 (C) 0.75 (D) 0.35 (E) 0.10

14. Let X be a r.v. having a Snedecor's $\mathcal{F}$ distribution with 5 and 10 degrees of freedom, respectively. The value of $P(0.3030 < X < 2.52)$ is equal to:

(A) 0.80 (B) 0.10 (C) 0.20 (D) 0.70 (E) 0.90

Questions 15 to 17 refer to the following exercise:

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

$$f(x,\theta) = \begin{cases} \frac{2(\theta-x)}{\theta^2} & \text{for } 0 \le x \le \theta; \\ 0 & \text{otherwise,} \end{cases}$$

It is known that the variance of this r.v. is $\mathrm{Var(X)} = \frac{\theta^2}{18}$.

We wish to estimate the parameter θ. In order to do so, a r.s. of size n, $X_1, X_2, \ldots, X_n$, is taken.

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

(A) $3\overline{X}$ (B) $\frac{\overline{X}}{3}$ (C) $\overline{X}$ (D) $\frac{1}{3\overline{X}}$ (E) $\frac{1}{\overline{X}}$

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

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

17. The variance of the method of moments estimator of θ is:

(A) $\frac{\theta^2}{2n}$ (B) $\frac{\theta^2}{n}$ (C) $\frac{\theta^2}{18n}$ (D) $\frac{\theta^2}{9n}$ (E) $\frac{\theta^2}{36n}$

Questions 18 to 21 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:

$$P(X=3) = P(X=-3) = \frac{\theta}{3}, \quad P(X=0) = 1 - \frac{2\theta}{3}$$

18. The method of moments estimator of θ is:

(A) $\frac{\Sigma_{i=1}^{n} X_i^2}{n}$ (B) $\frac{\overline{X}}{6}$ (C) All false (D) $\frac{\Sigma_{i=1}^{n} X_i^2}{6n}$ (E) $\overline{X}$

19. Is $\hat{\theta} = \frac{\Sigma_{i=1}^{n} X_i^2}{n}$ an unbiased estimator of θ?

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

20. In order to be able to obtain an estimate of the parameter θ, a r.s. of size $n = 10$ has been taken providing the following results: 0, 0, 0, 0, 3, 3, 3, -3, -3, -3. A method of moments estimate of θ is equal to:

(A) 0.90 (B) 0.30 (C) 0.60 (D) 0.70 (E) 0.10

21. In order to be able to obtain an estimate of the parameter θ, another r.s. of size $n = 11$ has been taken providing the following results: 0, 0, 0, 0, 0, 3, 3, 3, -3, -3, -3. A maximum likelihood estimate of θ is equal to:

(A) 0.75 (B) 0.91 (C) 0.82 (D) 0.55 (E) 0.62

Questions 22 and 23 refer to the following exercise:

The City Council of a given city wishes to test the null hypothesis that the probability that one of the inhabitants of one of its neighborhoods is an immigrant is $p = 0.20$, against the alternative hypothesis that it is $p = 0.40$. In order to test these hypotheses, a r.s. of $n = 20$ inhabitants has been taken, with the result that 5 of them were immigrants.

22. At the $\alpha = 0.05$ significance level, the most powerful test for the test statistic $Z = \Sigma_{i=1}^{20} X_i$ will reject the null hypothesis if:

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

23. The decision of the test will be:

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

Questions 24 to 26 refer to the following exercise:

An individual is interested in buying a Sarriko telephone. Before doing so, s/he decides to ask for its price at 31 different stores, obtaining a mean sample price of 750 euros with a sample standard deviation of 60 euros. We assume normality.

24. At the 98% confidence level, we can state that the Sarriko telephone mean price is contained in the interval:

(A) (723.05, 776.95) (B) (727.65, 772.35) (C) (745.08, 754.92)
(D) (731.38, 768.62) (E) (719.88, 780.12)

25. For the aforementioned confidence interval we wish to decrease its confidence level, the width of the new interval will be:

(A) Greater (B) Smaller (C) The same
(D) It cannot be determined (E) All false

26. At the 95% confidence level, we can state that the **standard deviation** of the Sarriko telephone price is contained in the interval:

(A) (48.73, 81.50) (B) (2192.53, 5417.48) (C) (39.57, 110.71)
(D) (46.82, 73.60) (E) (2374.47, 6642.86)

Questions 27 to 30 refer to the following exercise:

In a given city we wish to estimate the monthly mean expense a family has. Researchers suspect that this variable will have a very different behavior for the different neighborhoods and that, within each neighborhood, the behavior of this expense is very homogeneous. Let A and B be two of the neighborhoods under study, having each $N_A = 15000$ and $N_B = 10000$ families, respectively. It is also known that, from a previous study, the corresponding quasivariances are $\sigma_A^{*2} = 7000$ and $\sigma_B^{*2} = 6000$. It has been decided that stratified random sampling will be used, and that a sample of size $n = 1200$ will be taken.

27. If proportional allocation is used, the sample sizes n_A and n_B will be, respectively:

(A) 720 and 480 (B) 700 and 500 (C) 600 and 600
(D) 675 and 525 (E) 750 and 450

28. Under proportional allocation, the variance of the mean expense estimator will be:

(A) 3.332 (B) 2.288 (C) 1.825 (D) 5.236 (E) 1.904

29. If optimal allocation is used, the sample sizes n_A and n_B will be, respectively:

(A) 742 and 458 (B) 684 and 516 (C) 600 and 600
(D) 764 and 436 (E) 764 and 436

30. Under optimal allocation, the variance of the mean expense estimator will be:

(A) 3.228 (B) 5.228 (C) 3.332 (D) 5.236 (E) 2.000

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

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

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

It is known that the mean and the variance for the r.v. X are, respectively, $\mathrm{E}(X) = \theta + 4$ and $\mathrm{Var}(X) = \theta^2$.

i) Obtain, **providing all relevant details**, the method of moments estimator of the parameter θ.

ii) Obtain, **providing all relevant details**, the maximum likelihood estimator of the parameter θ.

iii) Is the maximum likelihood estimator of the parameter θ an unbiased estimator? Is it consistent? **You should appropriately justify all of your answers for this specific item**.

B. (10 points, 25 minutes)

Let X be a r.v. such that it follows an exponential distribution with parameter $1/\theta$. We wish to test the null hypothesis $H_0 : \theta = 3$, against the alternative hypothesis $H_1 : \theta = 1$. In order to do so, a r.s. of size $n = 1$ has been taken; that is, we observe X.

i) At the $\alpha = 0.10$ significance level, obtain, **providing all relevant details**, the most powerful critical region for this test.

ii) Obtain the power for this test.

C. (10 points, 25 minutes)

The r.v. X represents the time, in minutes, between two consecutive buying orders in the stock market for a given common stock. The following table includes the values obtained from a r.s. of size $n = 5000$, together with their sample absolute observed frequencies, for a one month-period.

X	n_i
$(0, 1]$	2000
$(1, 3]$	1900
$(3, 5]$	700
$(5, +\infty)$	400

i) Without performing any computation, which distribution would you think it is the most appropriate one for this data, a normal or an exponential distribution? **You should appropriately justify all of your answers for this specific item**

ii) Without paying any attention to the answer you have provided to the previous item, and at the 5% significance level, test the hypothesis that the distribution of the r.v. X follows an exponential distribution. You can use the maximum likelihood estimate of λ, given by $\hat{\lambda} = 1/\overline{X} = 0.50$. In addition, you can also use the fact that, for this case, $F_X(x) = P(X \leq x) = 1 - e^{-\lambda x},\ x \geq 0$.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

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

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

i) In order to be able to obtain the method of moments estimator of the parameter θ, we need to equate the first population moment $\alpha_1 = \mathrm{E}(X) = m$ to the first sample moment $a_1 = \overline{X}$. In this way, we have that:

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

ii) In order to be able to obtain the maximum likelihood estimator of the parameter θ, we have that likelihood function is given by:

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

If we take its natural logarithm, we have that:

$$\ln L(\vec{x},\theta) = -n\ln(\theta) - \frac{\sum_{i=1}^{n}(x_i-4)}{\theta}$$

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

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n}(x_i-4)}{\theta^2} = 0$$

so that,

$$\frac{n}{\theta} = \frac{\sum_{i=1}^{n}(x_i-4)}{\theta^2} \Longrightarrow \hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^{n}(X_i-4)}{n} = (\overline{X} - 4)$$

iii) Unbiasedness:

$$\mathrm{E}\left(\hat{\theta}_{\mathrm{ML}}\right) = \mathrm{E}\left(\overline{X} - 4\right) = \mathrm{E}\left(\overline{X}\right) - 4 = \mathrm{E}\left(X\right) - 4 = (\theta + 4) - 4 = \theta$$

Therefore, it is an unbiased estimator of θ.

Consistency:

We compute the estimator's variance.

$$\mathrm{Var}\left(\hat{\theta}_{\mathrm{ML}}\right) = \mathrm{Var}(\overline{X} - 4) = \mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\theta^2}{n}$$

Given that it is an unbiased estimator of θ, and that its variance tends to zero as n goes to infinity, we can state that it is a consistent estimator of θ.

Exercise B

We have that X is a r.v. following an exponential distribution with parameter $1/\theta$; that is, with probability density function given by:

$$f_X(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \; x > 0, \theta > 0$$

i) We wish to test the null hypothesis $H_0 : \theta = 3$, against the alternative hypothesis $H_1 : \theta = 1$. In order to do so, a r.s. of size $n = 1$ has been taken; that is, we observe X. The most powerful critical region for a given significance level can be obtained from the likelihood ratio test. That is,

$$\frac{L(x|H_0)}{L(x|H_1)} \leq k$$

In this case, given that the likelihood function is given by:

$$L(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}},$$

the most powerful critical region can be obtained from:

$$\frac{L(x|H_0)}{L(x|H_1)} = \frac{\frac{1}{3}e^{-\frac{x}{3}}}{e^{-x}} = \frac{1}{3}\, e^{\frac{2x}{3}} \leq k \Longrightarrow e^{\frac{2x}{3}} \leq k_1 \Longrightarrow \frac{2}{3}x \leq k_2 \Longrightarrow X \leq C$$

Therefore, we reject the null hypothesis if $X \leq C$, so that the critical region for the test statistic X is $\mathrm{CR} = (0, C]$. Given that $\alpha = 0.10$, we have that:

$$\alpha = 0.10 = P(X \leq C|H_0) = P(X \leq C|\theta = 3) = \int_0^C \frac{1}{3}e^{-\frac{x}{3}}dx = \left[-e^{-\frac{x}{3}}\right|_0^C = 1 - e^{-\frac{C}{3}} \Longrightarrow 0.10 = 1 - e^{-\frac{C}{3}}$$

$$\Longrightarrow e^{-\frac{C}{3}} = 0.90 \Longrightarrow C = -3\ln(0.90) \simeq 0.3161$$

Therefore, we reject the null hypothesis if $X \leq 0.3161$, so that the most powerful critical region will be given by $\mathrm{CR} = (0, 0.3161]$.

ii) The value for the power for this test will be:

$$\text{Power} = P(X \leq 0.3161|H_1) = P(X \leq 0.3161|\theta = 1) = \int_0^{0.3161} e^{-x}dx = \left[-e^{-x}\right|_0^{0.3161} = 1 - e^{-0.3161} = 0.2710$$

Exercise C

i) The normal distribution can take on negative values, whereas the exponential distribution only takes on positive values. The observed frequencies for the variable X are all positive for $X > 0$; thus, X does not take on negative values, and it would not make any sense to think that a normal distribution would be appropriate for this population. However, the exponential distribution would indeed make perfect sense.

ii) We have a goodness of fit test to a partially specified distribution, for which we are testing the null hypothesis $H_0 : X \in \exp(\lambda)$, against the alternative hypothesis $H_1 : X \notin \exp(\lambda)$. Moreover, we have $K = 4$ classes and, in addition, we have estimated one parameter ($h = 1$), so that the degrees of freedom for the test statistic will be $K - h - 1 = 4 - 1 - 1 = 2$. In this case, the corresponding table for the observed and expected frequencies, as well as the computation of the test statistic value for the chi-square test is the following one:

Class	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\dfrac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
$(0, 1]$	2000	0.3935	1967.5	0.5368
$(1, 3]$	1900	0.3834	1917.0	0.1508
$(3, 5]$	700	0.1410	705.0	0.0355
$(5, +\infty)$	400	0.0821	410.5	0.2686
Total	$n = 5000$	1	$n = 5000$	$z = 0.9917$

In order to be able to compute anyone of the estimated $\hat{p}_i$ probabilities, we use $P(a < X \leq b) = F_x(b) - F_X(a)$, where $F_X(x)$ is the cumulative distribution function for an exponential variable with parameter $\lambda = 0.50$. In this way,

$$P(0 < X \leq 1) = F_X(1) - F_X(0) = \left(1 - e^{-0.5\times 1}\right) - \left(1 - e^{-0.5\times 0}\right) = e^{-0.5\times 0} - e^{-0.5\times 1} = 1 - e^{-0.5\times 1} = 0.3935.$$

$$P(1 < X \leq 3) = F_X(3) - F_X(1) = \left(1 - e^{-0.5\times 3}\right) - \left(1 - e^{-0.5\times 1}\right) = e^{-0.5\times 1} - e^{-0.5\times 3} = 0.3834.$$

$$P(3 < X \leq 5) = F_X(5) - F_X(3) = \left(1 - e^{-0.5\times 5}\right) - \left(1 - e^{-0.5\times 3}\right) = e^{-0.5\times 3} - e^{-0.5\times 5} = 0.1410.$$

$$P(5 < X < \infty) = F_X(\infty) - F_X(5) = \left(1 - e^{-0.5\times \infty}\right) - \left(1 - e^{-0.5\times 5}\right) = e^{-0.5\times 5} - e^{-\infty} = e^{-0.5\times 5} = 0.0821.$$

The test statistic $\sum \dfrac{(n_i - np_i)^2}{np_i}$ follows, under the null hypothesis, a $\chi^2_{K-h-1} = \chi^2_2$ distribution. In this case:

$$z = 0.9917 < 5.99 = \chi^2_{2\,,\,0.05},$$

so that, at the 5% significance level, we do not reject the null hypothesis indicating that the theoretical model is an exponential distribution with parameter $\lambda = 0.50$.