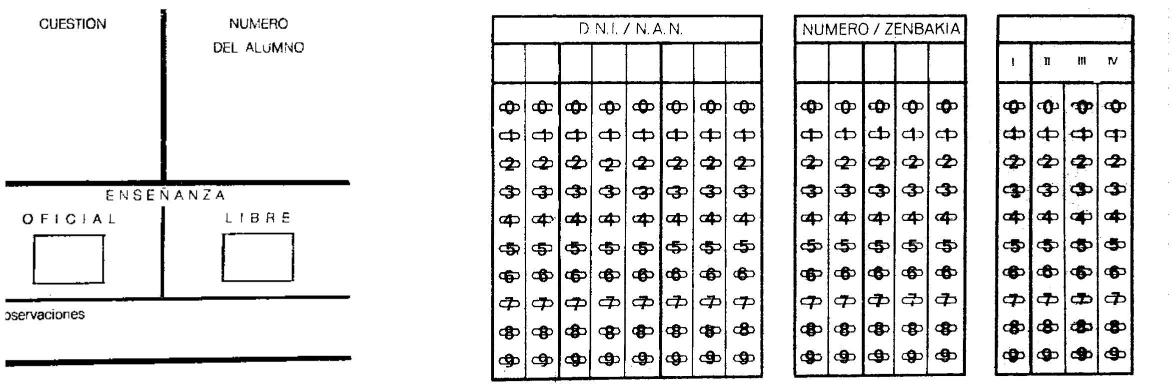
STATISTICS APPLIED TO BUSINESS ADMINISTRATION - Second Year Academic Year 2012-13
May 24, 2013

## 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 12 points in each part of the exam 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 30 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 lifetime or duration (in years) of the bulbs that are used in a given residential area follows an exponential distribution with parameter $\lambda = 0.50$. Bulbs considered defective or faulty will be those having a lifetime or duration smaller than 1 year. It is assumed that the lifetimes for the different bulbs are independent from each other. **Remark**: When computing the probability that a bulb is defective, you should round this probability up to one decimal place.

2. If a random sample of 10 bulbs of this type is taken, the probability that at least 2 of them will be defective is:

   (A) 0.3823 (B) 0.1673 (C) 0.0464 (D) 0.8327 (E) 0.9536

3. In the same sample of 10 bulbs of this type, the probability that exactly 6 bulbs **are not** defective is:

   (A) 0.2508 (B) 0.3669 (C) 0.6331 (D) 0.1114 (E) 0.9452

4. In the same sample of 10 bulbs of this type, the expected number of defective bulbs is:

   (A) 2 (B) 8 (C) 6 (D) 4 (E) 5

5. If we now take a random sample of 200 bulbs of this type, the approximate probability that no more than 72 of them are defective is, approximately:

   (A) 0.8599 (B) 0.5636 (C) 0.1401 (D) 0.4364 (E) 0.2581

**Questions 6 to 8 refer to the following exercise:**

Let $X$ be a r.v. with characteristic function given by $\Psi_X(u) = \mathrm{e}^{3(\mathrm{e}^{\mathrm{i}u}-1)}$.

6. $P(X > 2)$ is:

   (A) 0.1991 (B) 0.8009 (C) 0.5768 (D) 0.6472 (E) 0.4232

7. $P(X = 3)$ is:

   (A) 0.6472 (B) 0.4232 (C) 0.2240 (D) 0.4481 (E) 0.1680

8. Let $X_1, \ldots, X_{40}$ be a r.s. from this random variable. If we define the r.v. $Y = \sum_{i=1}^{40} X_i$, then $P(Y < 129)$ is approximately equal to:

   (A) 0.5319 (B) 0.7823 (C) 0.6802 (D) 0.4681 (E) 0.2177

**Questions 9 to 11 refer to the following exercise:**

Let $X$, $Y$ and $Z$ be three independent r.v. such that $X \in N(0,1)$, $Y \in \chi^2_{15}$ and $Z \in \gamma(\frac{1}{2}, 5)$.

9. The probability that the r.v. $W_1 = Y + Z$ takes on values smaller than 19.9 is:

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

10. The probability that the r.v. $W_2 = \dfrac{X}{\sqrt{Y/15}}$ takes on values smaller than 0.536 is:

(A) 0.70 (B) 0.60 (C) 0.20 (D) 0.35 (E) 0.30

11. If we define the r.v. $W_3 = \dfrac{1.5Z}{Y}$. The approximate value of $k$ such that $P(W_3 > k) = 0.90$ is equal to:

(A) 2.24 (B) 4.56 (C) 0.45 (D) 0.485 (E) 2.06

12. Let $X$ be a random variable having a Student's $t$ distribution with $n$ degrees of freedom, $t_{\overline{n}|}$. We have that $P(-t_{\overline{n}|\frac{\alpha}{4}} < X < t_{\overline{n}|\frac{\alpha}{4}})$ is equal to:

(A) $\frac{\alpha}{2}$ (B) $1-\alpha$ (C) $1-\frac{\alpha}{4}$ (D) $\frac{\alpha}{4}$ (E) $1-\frac{\alpha}{2}$

13. Let $Y$ be a normal r.v. with mean zero and variance equal to 4. We have that $P(Y^2 > 20.08)$ is equal to:

(A) 0.05 (B) 0.975 (C) 0.025 (D) 0.95 (E) 0.10

**Questions 14 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:

$$P(X=0) = P(X=1) = \theta, \quad P(X=2) = 1-2\theta$$

14. The method of moments estimator of $\theta$ is:

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

15. Is this an unbiased estimator of $\theta$?

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

16. In order to be able to obtain an estimate of the parameter $\theta$, a r.s. of size $n = 10$ has been taken providing the following results: 0, 0, 0, 1, 1, 1, 2, 2, 2, 2. A method of moments estimate of $\theta$ is equal to:

(A) 0.20 (B) 0.30 (C) 0.60 (D) 0.70 (E) 0.40

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

(A) 0.20 (B) 0.40 (C) 0.30 (D) 0.70 (E) 0.60

**Questions 18 to 21 refer to the following exercise:**

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

$$f(x,\theta) = \begin{cases} \frac{1}{4}\, \mathrm{e}^{-\frac{(x-\theta)}{4}} & \text{if } x \geq \theta; \\ 0 & \text{otherwise,} \end{cases}$$

We know that the mean and variance of this r.v. are $\mathrm{E(X)} = \theta + 4$ and $\mathrm{Var(X)} = 16$.

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

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

(A) $\overline{X}$ (B) $\min(X_i)$ (C) $\overline{X} - 4$ (D) $\max(X_i)$ (E) $4\overline{X}$

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

(A) $4\overline{X}$ (B) $\min(X_i)$ (C) $\overline{X}$ (D) $\max(X_i)$ (E) $\overline{X} - 4$

20. Is the method of moments estimator an unbiased estimator of $\theta$?

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

21. Is the method of moments estimator a consistent estimator of $\theta$?

(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:

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

Based on the information provided by a r.s. of size $n = 1$, $X_1$, we wish to test the null hypothesis $H_0 : \theta = 2$ against the alternative hypothesis $H_1 : \theta = 3$.

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

(A) $[C, 1)$ (B) $[C_1, C_2]^C$ (C) All false (D) $[C_1, C_2]$ (E) $(0, C]$

23. At the $\alpha = 0.10$ significance level, the power for this test is, approximately equal to

(A) 0.8538 (B) 0.9487 (C) 0.1462 (D) 0.0513 (E) 0.2710

**Questions 24 and 25 refer to the following exercise:**

Let $X$ be a r.v. having a Poisson distribution with parameter $\lambda$. In order to test $H_0 : \lambda = 0.80$ against $H_1 : \lambda < 0.80$, a r.s. of size $n = 10$ has been taken and the test statistic $Z = \sum_{i=1}^{10} X_i$ is used.

24. At the $\alpha = 0.05$ significance level, the test will reject the null hypothesis if:

(A) $Z \leq 3$ (B) $Z \leq 4$ (C) $Z \leq 5$ (D) $Z \geq 3$ (E) $Z \geq 4$

25. For the previous critical region and $\lambda = 0.40$, the power for this test is equal to:

(A) 0.4335 (B) 0.5665 (C) 0.2381 (D) 0.3712 (E) 0.6288

**Questions 26 and 27 refer to the following exercise:**

An individual is interested in buying a Blackberry. Before doing so, he decides to ask for its price at 26 different stores, obtaining a mean sample price of 450 euros with a sample standard deviation of 40 euros. We assume normality.

26. At the 90% confidence level, we can state that the Blackberry's mean price is contained in the interval:

(A) (436.32, 463.68) (B) (439.44, 460.56) (C) (434.56, 465.44) (D) (442.0, 458.0) (E) (433.52, 466.48)

27. At the 90% confidence level, we can state that the standard deviation of the price for the Blackberry is contained in the interval:

(A) (30.23, 63.03) (B) (27.59, 71.23) (C) (25.88, 65.34) (D) (34.78, 50.21) (E) (33.22, 53.38)

**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 students that, in a population of 15000 students, would be demanding specific services from the University Library. The firm wishes to estimate this proportion with a 95% confidence level and an absolute error of $\pm 0.01$. What would be the minimum number of students that would need to be selected if simple random sampling with replacement is used?

(A) 98 (B) 6724 (C) 82 (D) 5726 (E) 9604

**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:

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

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

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

B. (10 points, 25 minutes)

The Tourism Office from a given country claims that the probability that one of its adult inhabitants is bilingual is smaller than or equal to 0.85. In order to verify this claim, the country's Education Ministry takes a r.s. of 400 adult and obtains the result that 375 of them are bilingual.

i) Compute the approximate 0.95 confidence interval for the proportion of adult inhabitants that, in that specific country, are bilingual.

ii) Based on the information provided by the sample and at the approximate 5% significance level, what would be the country's Education Ministry decision about the country's Tourism Office claim?

**C.** (10 points, 25 minutes)

We wish to investigate if the distributions for the grades students have in a given course follows the theoretical model professors propose, under which $P(\text{Failing}) = 0.40$, $P(\text{Passing}) = 0.35$, $P(\text{Good}) = 0.20$, $P(\text{Very Good}) = 0.03$ and $P(\text{Outstanding}) = 0.02$. In order to do so, a r.s. of size 400 has been taken, providing the following results: out of the 400 students in the sample, 180 obtained a Failing grade, 130 obtained a Passing Grade, 70 obtained Good grade, 14 obtained a Very good Grade and only 6 obtained an Outstanding Grade.

i) What type of test would you perform to test the hypothesis of interest? Justify your response.

ii) At the 5% significance level, what is the decision on the basis of the result of the test?

iii) If, before actually carrying the test out, the professors had decided to fix a significance level smaller than the 5% from the previous item, what would had been the decision of the test?

**SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)**

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

## SOLUTIONS TO EXERCISES

**Exercise A**

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

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

**i)** In order to be able to obtain the method of moments estimator of the parameter $\theta$, we need to equate the first population moment $\alpha_1 = \mathrm{E}(X) = m$ to the first sample moment $a_1 = \overline{X}$. Before doing so, we need to compute $\alpha_1 = \mathrm{E}(X) = m$:

$$\alpha_1 = \mathrm{E}(X) = \int_0^1 x f(x,\theta) dx = \int_0^1 x\ \left(\theta\ x^{\theta-1}\right) dx$$

$$\alpha_1 = \int_0^1 \theta\ x^{\theta} dx = \left(\frac{\theta}{\theta+1}\right)\left[x^{\theta+1}\right|_0^1 = \left(\frac{\theta}{\theta+1}\right)$$

Therefore, we have that:

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

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

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

If we take its natural logarithm, we have that:

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

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

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

so that,

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

**Exercise B**

**i**) We wish to obtain the approximate 95% confidence interval for the proportion of adult inhabitants that, in that specific country, are bilingual. Given that the convergence conditions from the binomial to the normal distribution hold, we will have that:

$$\mathrm{CI}_{1-\alpha}(p) \simeq \left(\frac{z}{n} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{z(n-z)}{n^3}}\right)$$

In our specific case, and given that $n = 400$, $z = 375$ and $t_{0.025} = 1.96$, we have that

$$\mathrm{IC}_{0.95}(p) = \left(\frac{375}{400} \pm (1.96)\sqrt{\frac{375 \cdot (400-375)}{(400)^3}}\right) = (0.9375 \pm 0.0237) = (0.9138, 0.9612)$$

**ii**) We wish to test the null hypothesis $H_0 : p \leq 0.85 (= p_0)$, against the alternative hypothesis $H_1 : p > 0.85$. In order to do so, we use the test statistic given by

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

In this case and at the $\alpha$ significance level, the null hypothesis is rejected if:

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > t_\alpha$$

For $\alpha = 0.05$, the null hypothesis is rejected if:

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > t_{0.05} = 1.64$$

Given that in this specific test of hypotheses we have that $p_0 = 0.85$, $z = 375$ and $n = 400$, we can compute the test statistic value as

$$\frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{375}{400} - 0.85}{\sqrt{\frac{0.85(1-0.85)}{400}}} = 4.90$$

Therefore, at the approximate 5% significance level, the null hypothesis is rejected. That is, the claim made by the country's Tourism Office is not correct.

**Exercise C**

i) This corresponds to a goodness of fit test to a completely specified distribution, the one from the theoretical model the professors have proposed, and where no parameters need to be estimated.

ii) Under the null hypothesis of the theoretical model professors have proposed, we have that $P(\text{Failing}) = 0.40$, $P(\text{Passing}) = 0.35$, $P(\text{Good}) = 0.20$, $P(\text{Very Good}) = 0.03$ and $P(\text{Outstanding}) = 0.02$. That is, we initially have $K = 5$ classes and we do not need to estimate any parameter in the model ($h = 0$). Therefore, the degrees of freedom for the test statistic will be $K - 1 = 5 - 1 = 4$. Using the information obtained from the sample, we can build the table containing the required data that will allow us to perform the test of interest:

| Class | $n_i$ | $p_i$ | $np_i$ | $\frac{(n_i - np_i)^2}{np_i}$ |
|---|---|---|---|---|
| Failing | 180 | 0.40 | 160 | 2.50 |
| Passing | 130 | 0.35 | 140 | 0.71 |
| Good | 70 | 0.20 | 80 | 1.25 |
| Very Good | 14 | 0.03 | 12 | 0.33 |
| Outstanding | 6 | 0.02 | 8 | 0.50 |
| Total | 400 | 1 | 400 | $z = 5.29$ |

The test statistic $\sum \frac{(n_i - np_i)^2}{np_i}$ follows, under the null hypothesis, a $\chi^2_{K-1} = \chi^2_4$ distribution, with $K$ being the number of different classes or categories in which the grades for the specific course have been divided.

In this specific case:

$$z = 5.29 < 9.49 = \chi^2_{4\,,\,0.05},$$

so that, at the 5% significance level, the null hypothesis of the theoretical model proposed by the professors is not rejected.

**ii**) Given that we know that, in general, $\chi^2_{K-1\,,\,0.05} < \chi^2_{K-1\,,\,\alpha}$, for $\alpha < 0.05$, the decision of the test would have been the same.