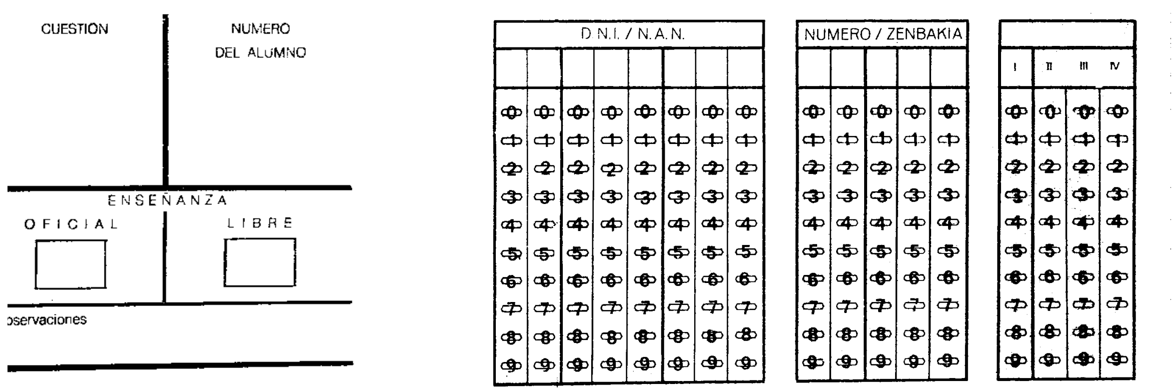
BUSINESS STATISTICS - Second Year Academic Year 2008-09
June 8, 2009

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 30 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 width (in milimeters) of the components of an electrical device produced by a given factory follows a uniform $U(5,15)$ distribution. We consider as acceptable only those components whose width belongs to the interval (8,12) milimeters.

2. If in a given period 20 components have been produced, the probability that at least 8 of them are considered as acceptable is:

   (A) 0.4044 (B) 0.4159 (C) 0.2447 (D) 0.5841 (E) 0.5956

3. In that same period of time, the expected number of acceptable components produced by the factory is:

   (A) 10 (B) 6 (C) 8 (D) 12 (E) 2

4. If in another period of time 100 components have been produced, the approximate probability that, for this new period of time, less than 51 components are **rejected** is:

   (A) 0.0613 (B) 0.9750 (C) 0.0262 (D) 0.9838 (E) 0.025

Questions 5 to 7 refer to the following exercise:

The number of people that arrive per minute at a given governmental office follows a Poisson distribution of variance equal to 3. We assume independence between the arrivals occurring in different minutes.

5. The probability that at least 3 people arrive at that governmental office in a given minute is:

   (A) 0.4232 (B) 0.8153 (C) 0.3528 (D) 0.6472 (E) 0.5768

6. The probability that at most 7 people arrive at that governmental office in a three-minute period is:

   (A) 0.7440 (B) 0.4557 (C) 0.9881 (D) 0.3239 (E) 0.6761

7. The approximate probability that at most 118 people arrive at that governmental office during a thirty-minute period is:

   (A) 0.014 (B) 0.9750 (C) 0.9987 (D) 0.025 (E) 0.0019

8. Let X and Y be two independent r.v. such that $X \in \gamma(1/2, 1)$ and $Y \in \gamma(1/2, 1)$. The distribution of the r.v. $X + Y$ is:

   (A) χ^2_2 (B) χ^2_4 (C) $\gamma(1, 2)$ (D) All false (E) $\gamma(1/2, 4)$

9. Let $\{X_n\}_{n\geq 2}$ be a sequence of random variables defined as:

$$X_n = \begin{cases} -n, & \text{with probability } \frac{1}{n} \\ 0, & \text{with probability } 1-\frac{2}{n} \\ n, & \text{with probability } \frac{1}{n} \end{cases}$$

The sequence will converge:

(A) Only in probability to $X = 0$

(B) Only in distribution to $X = 0$

(C) Only in distribution and probability to $X = 0$

(D) In distribution, probability and quadratic mean to $X = 0$

(E) All false

10. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables having a $N(4, \sigma^2 = 2/n^2)$ distribution. The sequence will converge:

(A) In distribution, probability and quadratic mean to $X = 4$

(B) Only in probability to $X = 4$

(C) Only in distribution to $X = 4$

(D) Only in distribution and probability to $X = 4$

(E) All false

11. Let X and Y be independent and identically distributed r.v. having a $N(0, 1)$ distribution. Let $Z = \frac{X}{\sqrt{Y^2}}$. The distribution of Z is:

(A) t_1 (B) $N(0, 1)$ (C) χ^2_1 (D) $N(0, 2)$ (E) $Z \equiv 1$

12. Let X be a random variable having a t_n distribution, a Student's t distribution with n degrees of freedom. Then, we have that:

(A) $t_{n,\alpha} > t_{n,\frac{\alpha}{4}}$ (B) $t_{n,\alpha} < t_{n,\frac{\alpha}{2}}$ (C) $\mathrm{E}(X) = n$ (D) $t_{n,\frac{\alpha}{2}} > t_{n,\frac{\alpha}{4}}$ (E) $t_{n,\alpha} = t_{n,1-\alpha}$

13. Let X be a random variable having a $\mathcal{F}_{10,20}$ distribution. The value of k such that $P(X > k) = 0.90$ is:

(A) 0.45 (B) 0.52 (C) 2.20 (D) 0.10 (E) 1.94

Questions 14 and 15 refer to the following exercise:

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

$$P(X = 0) = \theta^2; \quad P(X = 1) = 2\theta(1-\theta); \quad P(X = 2) = (1-\theta)^2.$$

In order to estimate the parameter θ, a random sample of size $n = 100$ has been taken, providing the following results: 22 zeros, 53 ones and 25 twos.

14. The maximum likelihood estimate of θ is:

(A) 0.485 (B) 0.125 (C) 0.776 (D) 0.225 (E) 0.554

15. The method of moments estimate of θ is:

(A) 0.554 (B) 0.125 (C) 0.485 (D) 0.225 (E) 0.776

Questions 16 to 19 refer to the following exercise:

Let us assume that we have a random variable having a uniform $U[\theta, \theta+1]$ distribution and that, in order to estimate the parameter θ, we have taken a r.s. of size n, $X_1, \ldots, X_n$, from it. We propose to use the estimator $\hat{\theta} = \overline{X} - \frac{1}{2}$. In addition, we know that the variance of a $U[a, b]$ r.v. is $\sigma^2 = (b-a)^2/12$.

16. Is this estimator unbiased?

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

17. The variance of this estimator is:

(A) $\frac{1}{12n}$ (B) $\frac{1}{12n} + 1$ (C) $\frac{1}{12}$ (D) $\frac{1}{12n} + \frac{1}{2}$ (E) All false

18. The mean square error of this estimator is:

(A) $\frac{1}{12}$ (B) $\frac{1}{12n} + 1$ (C) All false (D) $\frac{1}{12n} + \frac{1}{2}$ (E) $\frac{1}{12n}$

19. Is this estimator consistent?

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

Questions 20 to 22 refer to the following exercise:

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

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

We wish to test the null hypothesis $H_0 : \theta = 1$ against the alternative hypothesis $H_1 : \theta = 2$. In order to do so, a random sample of size $n = 1$ has been taken (that is, we observe X).

20. The most powerful critical region for X will be of the form:

(A) $(K, 1)$ (B) (K_1, K_2) (C) All false (D) $(K_1, K_2)^{\mathrm{c}}$ (E) $(0, K)$

21. If $\alpha = 0.05$, the most powerful critical region will be:

(A) $(0, 0.05)$ (B) $(0.05, 0.95)$ (C) $(0.95, 1)$ (D) $[0.05, 0.95]^c$ (E) $(0.90, 1)$

22. The power of this test is:

(A) 0.9025 (B) 0.5487 (C) 0.0975 (D) 0.4513 (E) 0.0025

Questions 23 to 26 refer to the following exercise:

The number of clients that arrive per minute at a given banking branch follows a Poisson distribution with parameter λ. In order to test $H_0 : \lambda = 1$ against $H_1 : \lambda = 2$, we have taken a r.s. of 4 minutes providing the following result: 3, 2, 1 and 5 clients arrived during those 4 minutes.

23. If we let Z be the total number of clients arriving at the banking branch during those 4 minutes, the most powerful critical region for this test will be of the form:

    (A) $Z \leq K$ (B) $Z \in [K_1, K_2]$ (C) $Z \geq K$ (D) $Z \in [K_1, K_2]^c$ (E) All false

24. If $\alpha = 0.05$, the most powerful critical region will be:

    (A) $Z \geq 9$ (B) $Z \in [3, 9]$ (C) All false (D) $Z \in [3, 9]^c$ (E) $Z \leq 3$

25. The power of this test is:

    (A) 0.5925 (B) 0.7166 (C) 0.0424 (D) 0.2834 (E) 0.4075

26. The decision of this test will then be:

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

Questions 27 and 28 refer to the following exercise:

In a given manufacturing process we wish to test the null hypothesis that the proportion of defective parts produced is no larger than 10%, against the alternative that it is actually larger than 10%. We have taken a r.s. of size $n = 200$, observing 25 defective parts.

27. With an approximate 90% confidence, we can state that the proportion of defective parts produced falls in the interval:

    (A) (0.05, 0.12) (B) (0.11, 0.18) (C) (0.15, 0.22) (D) (0, 0.05) (E) (0.09, 0.16)

28. 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

Questions 29 to 30 refer to the following exercise:

We wish to estimate the lottery mean expense per family (in euros) in a given province. In order to do so, a random sample of 61 families has been taken in that province, asking them how much money they have spent in lottery. The sample provided the following results: $\overline{x} = 188$ and $s^2 = 10000$. We assume normality.

29. The 95% confidence interval for the lottery mean expense per family is:

    (A) (166.44, 209.56) (B) (154.33, 204.56) (C) (162.18, 213.82) (D) (178.56, 198.65) (E) (180.33, 235.77)

30. We wish to test the null hypothesis that the lottery mean expense per family has been of 200 euros against the alternative hypothesis that it is different from it. At the 5% significance level, the decision of the test will be:

    (A) Reject the null hypothesis (B) It cannot be determined (C) - (D) - (E) Do not reject the null hypothesis

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:

$$f(x;\theta) = (\theta + 4) \;\; x^{\theta+3}, \quad 0 < x < 1, \; \theta > 0$$

i) Find (providing all relevant details) the maximum likelihood estimator of the parameter θ.

ii) Find (providing all relevant details) the method of moments estimator of the parameter θ.

B. (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a r.s. of size $n = 15$ taken from a binary $b(p)$ population. We wish to test the null hypothesis $H_0 : p = 0.50$ against the alternative hypothesis $H_1 : p = 0.30$.

i) Find (providing all relevant details) the form of the most powerful critical region for this test.

ii) At the 5% significance level, compute the critical region for this test.

iii) What is the power for this test?

C (10 points, 25 minutes) Let X be a $N(m, \sigma^2)$ random variable, with unknown variance we wish to estimate. In order to do so, a random sample of size n, $X_1, X_2, \cdots, X_n$, has been taken.

i) Find, **providing all relevant details**, the $(1 - \alpha)\%$ confidence interval for the population variance, σ^2.

ii) If $n = 30$ and, from the sample, we have that $s^2 = 25$, find the 95% confidence interval for the population standard deviation σ.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A)

We know that $X_i,\ i = 1, \ldots, n$, has a probability density function given by:

$$f(x_i; \theta) = (\theta + 4)\ \ x_i^{\theta+3}, \quad 0 < x_i < 1,\ \theta > 0$$

i) Thus, the likelihood function will be given by

$$L(\vec{x}; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

$$L(\vec{x}; \theta) = \left[(\theta + 4)\ \ x_1^{\theta+3}\right]\left[(\theta + 4)\ \ x_2^{\theta+3}\right] \cdots \left[(\theta + 4)\ \ x_n^{\theta+3}\right]$$

$$L(\vec{x}; \theta) = (\theta + 4)^n \left(\prod_{i=1}^{n} x_i\right)^{(\theta+3)}$$

The maximum likelihood estimator of θ is the value that maximizes the likelihood function or, equivalently, its natural logarithm:

$$\ln L(\vec{x}; \theta) = n \ln(\theta + 4) + (\theta + 3) \ln\left(\Pi_{i=1}^{n} x_i\right)$$

If we take derivatives with respect to θ, we have that:

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

Therefore,

$$\frac{n}{(\theta + 4)} = -\ln\left(\prod_{i=1}^{n} x_i\right),$$

so that

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

ii) In order to find the method of moments estimator of θ, we have to compute $\alpha_1 = \mathrm{E}(X)$. In this specific case,

$$\alpha_1 = \mathrm{E}(X) = \int_0^1 x f(x; \theta) dx = \int_0^1 x\left[(\theta + 4) x^{\theta+3}\right] dx = \int_0^1 (\theta + 4) x^{\theta+4} dx$$

$$\Longrightarrow \alpha_1 = \mathrm{E}(X) = (\theta + 4)\left[\frac{x^{\theta+5}}{(\theta + 5)}\right|_0^1 = \left(\frac{\theta + 4}{\theta + 5}\right)$$

Finally, to compute the method of moments estimator of θ, we make the sample and population moments equal to each other. That is, $\alpha_1 = a_1$, so that

$$\alpha_1 = \mathrm{E}(X) = \left(\frac{\theta+4}{\theta+5}\right) = \overline{X} = a_1,$$

and, thus:

$$\theta + 4 = \overline{X}(\theta+5) \Longrightarrow \theta(1-\overline{X}) = 5\overline{X} - 4 \Longrightarrow \hat{\theta}_{MM} = \left(\frac{5\overline{X}-4}{1-\overline{X}}\right)$$

Exercise B

Let $X_1, \ldots, X_n$ be a random sample of size $n = 15$ taken from a binary $b(p)$ population. We wish to test the null hypothesis $H_0 : p = 0.50$ against the alternative hypothesis $H_1 : p = 0.30$.

i) To obtain the form of the most powerful critical region for this test, we use the Neyman-Pearson theorem. In this way, the likelihood functions under the null and alternative hypotheses will be given by:

$$L(\vec{x}; p_0) = L(\vec{x}; p = 0.50) = (0.50)^{\Sigma_{i=1}^n x_i}(1-0.50)^{n-\Sigma_{i=1}^n x_i},$$

and

$$L(\vec{x}; p_1) = L(\vec{x}; p = 0.30) = (0.30)^{\Sigma_{i=1}^n x_i}(1-0.30)^{n-\Sigma_{i=1}^n x_i},$$

respectively. Therefore, if we use the Neyman-Pearson theorem, we will have that:

$$\frac{L(\vec{x}; p_o)}{L(\vec{x}; p_1)} = \frac{(0.50)^{\Sigma_{i=1}^n x_i}(1-0.50)^{n-\Sigma_{i=1}^n x_i}}{(0.30)^{\Sigma_{i=1}^n x_i}(1-0.30)^{n-\Sigma_{i=1}^n x_i}} \leq K,\ K > 0$$

$$\Longrightarrow \left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right]^{\Sigma_{i=1}^n x_i} \quad \left[\frac{(1-0.50)}{(1-0.30)}\right]^n \leq K$$

$$\Longrightarrow \left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right]^{\Sigma_{i=1}^n x_i} \leq K_1,\ K_1 > 0$$

If we now take natural logarithms, we have:

$$\left(\sum_{i=1}^n x_i\right) \ln\left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right] \leq K_2,\ K_2 > 0$$

Now, given that $0.50 > 0.30$ and that, in addition, $(1-0.30) > (1-0.50)$, the natural logarithm is positive, so that we conclude that the decision rule will be to reject the null hypothesis if $\sum_{i=1}^n X_i \leq C$. Thus, the form of the most powerful critical region for the test statistic $Z = \sum_{i=1}^n X_i$ will be $\mathrm{CR} = [0, C]$.

ii) At the $\alpha = 0.05$ significance level and taking into account that $Z = \sum_{i=1}^n X_i \in b(p, 15)$, we will have that:

$$\alpha = 0.05 \geq P[Z \in \mathrm{CR}|H_0] = P[Z \leq C | Z \in b(0.50, 15)] = F_Z(C)$$

$$\Longrightarrow F_Z(C) \leq 0.05 \Longrightarrow C = 3 \Longrightarrow \mathrm{RC} = [0, 3].$$

That is, we reject the null hypothesis if $Z = \sum_{i=1}^n X_i \leq 3$.

iii) To compute the power of this test, we will have that:

$$\text{Power} = P[Z \in \text{CR}|H_1] = P[Z \leq 3|Z \in b(0.30, 15)] = F_Z(3) = 0.2969.$$

Exercise C)

i) Let $\vec{X} = (X_1, X_2, \ldots, X_n)$ be a random sample of size n taken from a $N(m, \sigma^2)$ population. Therefore, we will have that

$$\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$$

Confidence Interval

$$P\left(\chi^2_{n-1|1-\alpha/2} < \frac{nS^2}{\sigma^2} < \chi^2_{n-1|\alpha/2}\right) = 1 - \alpha$$

$$P\left(\frac{1}{\chi^2_{n-1|\alpha/2}} < \frac{\sigma^2}{nS^2} < \frac{1}{\chi^2_{n-1|1-\alpha/2}}\right) = 1 - \alpha$$

$$P\left(\frac{nS^2}{\chi^2_{n-1|\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{n-1|1-\alpha/2}}\right) = 1 - \alpha$$

Therefore, the $(1 - \alpha)\%$ confidence interval for the variance σ^2 is given by:

$$\text{CI}_{1-\alpha} = \left(\frac{ns^2}{\chi^2_{n-1|\alpha/2}}, \frac{ns^2}{\chi^2_{n-1|1-\alpha/2}}\right)$$

ii) More specifically, if it is given that $n = 30$ and $s^2 = 25$, the 95% confidence interval for σ^2 will be:

$$\text{CI}_{1-\alpha} = \left(\frac{750}{\chi^2_{n-1|\alpha/2}}, \frac{750}{\chi^2_{n-1|1-\alpha/2}}\right)$$

Given that $\chi^2_{29|0.05/2} = 45.7$ and $\chi^2_{29|1-0.05/2} = 16$, we will then have that the 95% confidence interval for the variance will be:

$$\text{CI}_{0.95} = (16.41, 46, 88),$$

so that, the 95% confidence interval for the standard deviation will be:

$$CI_{0.95} = (\sqrt{16.41}, \sqrt{46, 88}) = (4.05, 6.85).$$