BUSINESS STATISTICS - Second Year Academic Year 2007-08
February 4, 2008

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:

$\underbrace{\text{12545}}$ 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) Londres (E) Pekin

Questions 2 to 4 refer to the following exercise:

In a shipping company it is known that about 80% of its deliveries are addressed to its habitual clients.

2. If the company performs 15 deliveries, the probability that exactly seven of them are addressed to its habitual clients is:

   (A) 0.9958 (B) 0.0034 (C) 0.8358 (D) 0.0001 (E) 0.0139

3. If the company performs 15 deliveries, the probability that at least ten of them are addressed to its habitual clients is:

   (A) 0.8358 (B) 0.9819 (C) 0.9389 (D) 0.1031 (E) 0.1876

4. If the company perform 100 deliveries, the approximate probability that no more than 76 of them are addressed to its habitual clients is:

   (A) 0.8809 (B) 0.1894 (C) 0.1020 (D) 0.8413 (E) 0.8106

Questions 5 and 6 refer to the following exercise:

Let X be a r.v. having a Poisson distribution and such that $P(X = 3) = P(X = 4)$.

5. The mode(s) of this distribution is (are):

   (A) 4 (B) 3 and 4 (C) 3 (D) 2 and 3 (E) 4 and 5

6. $P(X > 6)$ is approximately equal to:

   (A) 0.11 (B) 0.79 (C) 0.05 (D) 0.21 (E) 0.89

7. If we have a r.s. $X_1, \ldots, X_{50}$ taken from a Poisson distribution having variance equal to 4 and we define the r.v. $Y = \sum_{i=1}^{50} X_i$, then $P(Y > 180)$ is approximately equal to:

   (A) 0.9474 (B) 0.0838 (C) 0.5080 (D) 0.0526 (E) 0.9162

Questions 8 to 10 refer to the following exercise:

Let $X_1, X_2, \cdots, X_{50}$ be i.i.d. r.v. having a $\gamma(0.50, 1)$ distribution.

8. The probability that the r.v. X_3 takes on values larger than 2 is:

   (A) 0.3679 (B) 0.1353 (C) 0.6065 (D) 0.8647 (E) 0.6321

9. We define the r.v. $Y = X_1 + X_2 + \cdots + X_{50}$. The distribution of the r.v. Y is:

   (A) $\gamma(50, 50)$ (B) χ^2_{50} (C) $\gamma(0.50, 100)$ (D) $\gamma(0.50, 50)$ (E) χ^2_{25}

10. Using the central limit theorem, the approximate probability that the r.v. Y takes on values larger than 130 is:

(A) 0.5025 (B) 0.9830 (C) 0.4207 (D) 0.5793 (E) 0.0170

11. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of r.v. having a $N(0,\sigma^2 = 1+1/n^2)$ distribution. It is known that the characteristic function of a $N(m,\sigma^2)$ r.v. is given by $\psi_n(u) = e^{ium-\frac{\sigma^2 u^2}{2}}$. The sequence will converge:

(A) In distribution to a $N(0,1)$ r.v.

(B) In distribution to $X = 1$

(C) In distribution to a $N(0,2)$ r.v.

(D) In distribution to $X = 0$

(E) All false

12. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of r.v. with probability mass function given by:

$$P_n(x) = \begin{cases} \frac{3}{4} - \frac{1}{n}, & \text{if } x = -\frac{1}{n^2} \\ \frac{1}{n}, & \text{if } x = 0 \\ \frac{1}{4}, & \text{if } x = \frac{1}{n^2} \end{cases}$$

The sequence will converge:

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

(B) In distribution and probability to $X = \frac{3}{4}$

(C) In distribution and probability to $X = 0$

(D) Only in distribution to $X = \frac{3}{4}$

(E) All false

13. Let X be a r.v. having a $\gamma(1,5)$ distribution. The distribution of the r.v. $Y = 2X$ is:

(A) $\gamma(2,10)$ (B) $\gamma(\frac{1}{2},10)$ (C) $\gamma(2,5)$ (D) χ^2_5 (E) χ^2_{10}

14. Let X be a r.v. having an exponential distribution with mean equal to 2. Then, $P(X \leq 4)$ is:

(A) 0.1353 (B) 0.0003 (C) 0.0183 (D) 0.9997 (E) 0.8647

15. Let X be a r.v. having a t_n distribution; that is, a Student's t distribution with n degrees of freedom. Then, we have that:

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

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

$$f(x,\theta) = \frac{1}{2^{\frac{\theta}{2}}\Gamma(\frac{\theta}{2})}\, x^{\frac{\theta}{2}-1}\, e^{-\frac{x}{2}},\ x > 0,\ \theta > 0$$

In order to estimate the parameter θ, a r.s. of size n, $X_1,\ldots,X_n$, has been taken. A sufficient statistic for the parameter θ is:

(A) $\sum_{i=1}^n X_i$ (B) $\prod_{i=1}^n X_i$ (C) $\prod_{i=1}^n \ln X_i$ (D) $\prod_{i=1}^n \left(\frac{1}{X_i}\right)$ (E) $\sum_{i=1}^n X_i^2$

Questions 17 and 18 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:

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

Assuming that the parameter a is known and, in order to estimate the parameter r, a r.s. of size n has been taken.

17. The method of moments estimator of r, $\hat{r}_{MM}$, is:

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

18. Is this estimator unbiased?

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

Questions 19 and 20 refer to the following exercise:

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

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

In order to estimate the parameter θ, a r.s. of size n has been taken, for which three zeroes were obtained.

19. The maximum likelihood estimate of θ is:

(A) $\frac{3}{n}$ (B) $\frac{3}{2n}$ (C) $\frac{n-3}{2n}$ (D) $\frac{n-3}{n}$ (E) $\frac{1}{n}$

20. The method of moments estimate of θ is:

(A) $\frac{n-3}{2n}$ (B) $\frac{3}{2n}$ (C) $\frac{1}{n}$ (D) $\frac{n-3}{n}$ (E) $\frac{3}{n}$

Questions 21 to 23 refer to the following exercise:

Let $X_1, \ldots, X_n$ be a r.s. from a binary $b(\theta)$ population. In order to estimate the parameter θ, we propose to use the following estimator:

$$\hat{\theta} = \frac{2X_1 + X_2 + \cdots + X_{n-1} + 2X_n}{n+2}$$

21. The bias of the estimator $\hat{\theta}$ is:

(A) $n\theta$ (B) $\frac{1}{n}$ (C) $-\frac{\theta}{n}$ (D) $\frac{2\theta}{n}$ (E) 0

22. The variance of the estimator $\hat{\theta}$ is:

(A) $\frac{\theta(1-\theta)}{n}$ (B) $\frac{1}{(n+2)^2}$ (C) $\frac{(n+6)\theta(1-\theta)}{(n+2)^2}$ (D) $\frac{2\theta(1-\theta)}{n}$ (E) $\frac{n\theta(1-\theta)}{(n+2)^2}$

23. Is $\hat{\theta}$ a consistent estimator of θ?

(A) No (B) It cannot be determined (C) Yes (D) - (E) -

Questions 24 to 26 refer to the following exercise:

We wish to test the null hypothesis that the probability density function of a given population is $\gamma(a = 2, r)$, against the alternative hypothesis that it is $\gamma(a = 4, r)$; that is, the parameter r is common to both distributions. We recall that the probability density function for a $\gamma(a, r)$ distribution is given by:

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

In order to carry out this test, a random sample of size $n = 1$ has been taken from that population (that is, we observe X).

24. The most powerful critical region for X is of the form:

    (A) $[K, +\infty]$ (B) $[0, K]$ (C) All false (D) $[K_1, K_2]^c$ (E) $[K_1, K_2]$

25. If $r = 1$ and $x = 0.05$, what would be the decision at the 0.05 significance level?

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

26. What would be the approximate power for this specific case and specific significance level?

    (A) 0.10 (B) 0.95 (C) 0.85 (D) 0.05 (E) 0.90

Questions 27 to 30 refer to the following exercise:

An individual is interested in buying a Nintendo DS Light. Before doing so, he decides to ask for its price at 31 different stores, obtaining a mean sample price of 150 euros with a sample standard deviation of 10 euros. We assume normality.

27. At the 95% confidence level, we can state that the Nintendo DS Light mean price is contained in the interval:

    (A) (146.90, 153.10) (B) (147.50, 152.50) (C) (146.28, 153.72)
    (D) (145.70, 154.30) (E) (144.60, 155.40)

28. At the 95% confidence interval, we can state that the variance of the price for the Nintendo DS Light is contained in the interval:

    (A) (65.96, 184.52) (B) (70.78, 167.57) (C) (31.55, 92.15.)
    (D) (63.83, 178.57) (E) (6.60, 18.45)

29. If, at the 5% significance level, we wish to test the null hypothesis that the Nintendo DS Light mean price is $m = 140$, the test result will be:

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

30. If, at the 5% significance level, we wish to test the null hypothesis that the variance of the price for the Nintendo DS light is $\sigma^2 = 170$, the test result will be:

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

EXERCISES (Time: 75 minutes)

A. (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 $P(\text{F}) = 0.40$, $P(\text{C}) = 0.35$, $P(\text{B}) = 0.20$, $P(\text{A}) = 0.03$ and $P(\text{A+}) = 0.02$. 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 F, 130 obtained C, 70 obtained B, 14 obtained A and only 6 obtained A+.

i) What type of test would you perform to test the hypothesis of interest?

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

B (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a set of independent r.v. such that $X_1 \in N(k_1\theta, \sigma^2)$, $X_2 \in N(k_2\theta, \sigma^2), \ldots, X_n \in N(k_n\theta, \sigma^2)$, with known variance $\sigma^2 > 0$, and such that the $k_i > 0$'s $i = 1, \ldots, n$ are known constants.

It is known that the probability density function for a $N(m, \sigma^2)$ random variable is given by:

$$f(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

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

ii) Is this estimator unbiased? What is/are the required condition(s) that would make of this estimator a consistent estimator of θ? Remark: In order to answer the latter question, you must compute the variance of the estimator.

C. (10 points, 25 minutes)

The following table describes the probability mass function for the discrete random variable X under the null ($P_0(x)$) and alternative ($P_1(x)$) hypotheses.

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

In order to test the null hypothesis $H_0 : P(x) = P_0(x)$ against the alternative hypothesis $H_1 : P(x) = P_1(x)$, a random sample of size $n = 1$ has been taken.

i) Would you include the point $X = 1$ in the critical region? Provide all relevant details.

ii) Would you include the point $X = 3$ in the critical region? Provide all relevant details.

iii) At the 10% significance level, and providing all relevant details that lead us to the required answer, obtain the most powerful critical region for this test. **Hint**: It is very important that before answering this item you remember your answers to the previous ones.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A)

i) This corresponds to a goodness of fit test to a completely specified distribution.

ii) Under the null hypothesis of the theoretical model professors have proposed, we have that $P(\text{F}) = 0.40$, $P(\text{C}) = 0.35$, $P(\text{B}) = 0.20$, $P(\text{A}) = 0.03$ and $P(\text{A+}) = 0.02$. 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	$\dfrac{(n_i - np_i)^2}{np_i}$
F	180	0.40	160	2.50
C	130	0.35	140	0.71
B	70	0.20	80	1.25
A	14	0.03	12	0.33
A+	6	0.02	8	0.50
Total	400	1	400	$z = 5.29$

The test statistic $\sum \dfrac{(n_i - np_i)^2}{np_i}$ 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.

Exercise B

Given that $X_i \in N(k_i\theta, \sigma^2)$, with known variance $\sigma^2 > 0$, we then have that

$$f(x_i;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x_i-k_i\theta)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

i) Thus, the likelihood function will be given by

$$L(\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$

$$L(\theta) = \left[\frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x_1-k_1\theta)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x_2-k_2\theta)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x_n-k_n\theta)^2}{2\sigma^2}}\right]$$

$$L(\theta) = (2\pi)^{-\frac{n}{2}} \sigma^{-n}\; e^{-\sum_{i=1}^{n} \frac{(x_i-k_i\theta)^2}{2\sigma^2}}$$

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

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x_i - k_i\theta)^2}{2\sigma^2}$$

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

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - k_i\theta)(k_i) = 0$$

Therefore,

$$\sum_{i=1}^{n} k_i x_i - \sum_{i=1}^{n} k_i^2\theta = 0,$$

so that

$$\hat{\theta}_{\text{ML}} = \frac{\sum_{i=1}^{n} k_i X_i}{\sum_{i=1}^{n} k_i^2}$$

ii) The estimator will be unbiased if $\text{E}(\hat{\theta}_{\text{ML}}) = \theta$.

$$\text{E}(\hat{\theta}_{\text{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)}\text{E}\left(\sum_{i=1}^{n} k_i X_i\right) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)}\sum_{i=1}^{n} k_i\text{E}(X_i)$$

$$\text{E}(\hat{\theta}_{\text{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)}\sum_{i=1}^{n} k_i(k_i\theta) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)}\left(\sum_{i=1}^{n} k_i^2\right)\theta = \theta$$

Therefore, $\hat{\theta}_{\text{ML}}$ is an unbiased estimator of θ. In order to be able to establish the conditions under which $\hat{\theta}_{\text{ML}}$ is consistent, assuming that it can actually be consistent, we can check if the two sufficient conditions hold:

a) $\lim\limits_{n\to\infty} \text{E}(\hat{\theta}_{\text{ML}}) = \theta$

b) $\lim\limits_{n\to\infty} \text{Var}(\hat{\theta}_{\text{ML}}) = 0$

As $\hat{\theta}_{\text{ML}}$ is an unbiased estimator of θ, condition a) holds. In addition, we have that

$$\text{Var}(\hat{\theta}_{\text{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\text{Var}\left(\sum_{i=1}^{n} k_i X_i\right) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\sum_{i=1}^{n}\text{Var}(k_i X_i)$$

$$\text{Var}(\hat{\theta}_{\text{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\sum_{i=1}^{n} k_i^2\text{Var}(X_i) = \frac{\left(\sum_{i=1}^{n} k_i^2\right)}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} k_i^2}$$

Thus, if the condition $\lim_{n\to\infty}\sum_{i=1}^{n} k_i^2 = +\infty$ holds, we would then have that the two conditions for consistency would hold and, as a result, $\hat{\theta}_{ML}$ would be a consistent estimator of θ.

Exercise C)

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

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

A random sample of size $n = 1$ has been taken; that is, we observe X.

i) Would you include the point $X = 1$ in the critical region?

Given that, under the probability mass function in the null hypothesis $P_0(x)$, this point has zero probability, the random variable X cannot take on this value under the null hypothesis. Therefore, $X = 1$ is a rejection point for H_0 and, thus, it **should always** be included in the critical region for this test.

ii) Would you include the point $X = 3$ in the critical region?

Given that, under the probability mass function in the alternative hypothesis $P_1(x)$, this point has zero probability, the random variable X cannot take on this value under the alternative hypothesis, but it can clearly take on this value under the null hypothesis. Therefore, $X = 3$ is not a rejection point for H_0 and, thus, it **should never** be included in the critical region for this test.

iii) At the $\alpha = 0.10$ significance level and recalling the answers provided in earlier items for this exercise, we have that there are only three possible critical regions for this test: $\mathrm{CR}_1 = \{0, 1\}$, $\mathrm{CR}_2 = \{1, 2\}$ and $\mathrm{CR}_3 = \{1, 4\}$. This is due to the fact that:

$$\begin{aligned}
\alpha_1 &= P(X \in \mathrm{CR}_1|P_0) = P(X = 0, 1|P_0) = 0.10 + 0 = 0.10 \leq \alpha = 0.10 \\
\alpha_2 &= P(X \in \mathrm{CR}_2|P_0) = P(X = 1, 2|P_0) = 0 + 0.05 = 0.05 \leq \alpha = 0.10 \\
\alpha_3 &= P(X \in \mathrm{CR}_3|P_0) = P(X = 1, 4|P_0) = 0 + 0.10 = 0.10 \leq \alpha = 0.10
\end{aligned}$$

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

$$\begin{aligned}
\mathrm{Power}_1 &= P(X \in \mathrm{CR}_1|P_1) = P(X = 0, 1|P_1) = 0.30 + 0.20 = 0.50 \\
\mathrm{Power}_2 &= P(X \in \mathrm{CR}_2|P_1) = P(X = 1, 2|P_1) = 0.20 + 0.15 = 0.35 \\
\mathrm{Power}_3 &= P(X \in \mathrm{CR}_3|P_1) = P(X = 1, 4|P_1) = 0.20 + 0.20 = 0.40
\end{aligned}$$

From these calculations, we conclude that, at the $\alpha = 0.10$ significance level, the most powerful critical region for this test is CR_1.