BUSINESS STATISTICS - Second Year Academic Year 2006-07
June 25, 2007

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 5 refer to the following exercise:

The probability that a driver has an accident in the highway that goes from Bilbao to Burgos is 0.10. We assume independence between the different drivers going from Bilbao to Burgos in this specific highway.

2. If there are 15 drivers going from Bilbao to Burgos in this highway, the probability that at most three of them have an accident is:

   (A) 0.1285 (B) 0.9873 (C) 0.8159 (D) 0.9444 (E) 0.0556

3. The probability that, among the 15 drivers going from Bilbao to Burgos in this highway, exactly nine drivers **do not** have an accident is:

   (A) 0.1413 (B) 1 (C) 0.0019 (D) 0.3431 (E) 0.2059

4. If we now have 60 drivers going from Bilbao to Burgos in this highway, the approximate probability that at least four drivers have an accident is:

   (A) 0.8488 (B) 0.7149 (C) 0.2851 (D) 0.1339 (E) 0.1512

5. If we now have 360 drivers going from Bilbao to Burgos in this highway, the approximate probability that exactly 34 drivers have an accident is:

   (A) 0.0674 (B) 0.6026 (C) 0.6700 (D) 0.9326 (E) 0.1434

Questions 6 and 7 refer to the following exercise:

The number of labor accidents **per day** in a given region follows a Poisson distribution with mean 3. We assume independence between the labor accidents occurring in different days in that region.

6. The probability that in **two days** exactly six labor accidents occur in that region is:

   (A) 0.3937 (B) 0.6063 (C) 0.1606 (D) 0.4457 (E) 0.1377

7. The probability that in **10 days** fewer that 25 labor accidents occur in that region is, approximately:

   (A) 0.8212 (B) 0.1587 (C) 0.8413 (D) 0.1788 (E) 0.2061

8. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables having a $N(0, 1/n)$ distribution. If it is known that the characteristic function of a normal $N(m, \sigma^2)$ r.v. is $\psi_n(u) = e^{ium - \frac{\sigma^2 u^2}{2}}$, then the sequence will converge:

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

   (B) Only in distribution to $X = 1$

   (C) In distribution and probability to $X = 1$

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

   (E) All false

9. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables with probability mass function given by:

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

The sequence will converge:

(A) In distribution and probability to $X = \frac{1}{2}$

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

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

(D) Only in distribution to $X = \frac{1}{2}$

(E) All false

Questions 10 and 11 refer to the following exercise:

Let $X_1, X_2, \cdots, X_{75}$ be independent r.v. having a uniform $U(-1, 1)$ distribution. **Hint:** Note that $\mathrm{E}(X) = 0$ and that $\mathrm{Var}(X) = \frac{1}{3}$.

10. If we define the r.v. $Y = X_1 + X_2 + \cdots + X_{75}$, then $P(Y \geq 6)$ is, approximately:

(A) 0.1587 (B) 0.8413 (C) 0.1151 (D) 0.8849 (E) 0

11. If we define the r.v. $Z = (X_1 + X_2 + \cdots + X_{75})/75$, then $P(0 \leq Z \leq 0.10)$ is, approximately:

(A) 0.50 (B) 1 (C) 0.9332 (D) 0 (E) 0.4332

12. If W is a r.v. such that $W \in \gamma(1/2, 5)$, then $P(W \leq 4.87)$ is:

(A) 0.10 (B) 0.05 (C) 0.25 (D) 0.95 (E) 0.90

13. If Y is a r.v. such that $Y \in \mathcal{F}_{12,10}$, then $P(Y \leq 0.3636)$ is:

(A) 0.95 (B) 0.10 (C) 0.05 (D) 0.90 (E) 0.25

14. If Z is a r.v. such that $Z \in t_{15}$, then $P(-1.75 \leq Z \leq 0.691)$ is:

(A) 0.40 (B) 0.10 (C) 0.30 (D) 0.70 (E) 0.25

15. Let X be a random variable with probability density function given by:

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

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

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

Questions 16 and 17 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 random sample of size n has been taken, for which three zeroes were obtained.

16. The maximum likelihood estimate of θ is:

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

17. The method of moments estimate of θ is:

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

Questions 18 and 19 refer to the following exercise:

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

$$f(x)=\begin{cases}\frac{x}{\theta}e^{\frac{-x^2}{2\theta}}, & x>0,\quad \theta>0\\ 0, & \text{otherwise}\end{cases},$$

and for which we have that $\mathrm{E}(X)=\sqrt{\frac{\theta\pi}{2}}$. In order to estimate the parameter θ, a random sample of size n has been taken.

18. The maximum likelihood estimator of θ is:

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

19. The method of moments estimator of θ is:

(A) $\frac{\overline{X}^2}{\pi}$ (B) $\frac{2\overline{X}^2}{\pi}$ (C) $\pi\overline{X}^2$ (D) $\frac{\pi\overline{X}^2}{2}$ (E) All false

Questions 20 and 21 refer to the following exercise:

Let $X_1,\dots,X_n$ be a random sample from a $N(m,\sigma^2)$ population. In order to estimate the parameter m, we propose to use the estimator:

$$\hat{m}=\frac{X_1+X_2+\cdots+X_{n-1}}{n}$$

20. The bias of the estimator $\hat{m}$ is:

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

21. Is the estimator $\hat{m}$ consistent?

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

22. Let X be a r.v. having an exponential distribution with parameter $1/\theta$. If we take a random sample of size $n=3$ and, in addition, we propose to use the estimator $\hat{\theta}=(X_1+X_2+X_3)/2k$ to estimate the parameter θ, the value of k such that $\hat{\theta}$ is an unbiased estimator of θ is:

(A) $\frac{2}{3}$ (B) All false (C) $\frac{1}{3}$ (D) $\frac{3}{2}$ (E) 3

Questions 23 and 24 refer to the following exercise:

We wish to test the null hypothesis that the r.v. X has a probability density function $f(x) = 2x,\ x \in (0,1)$ against the alternative hypothesis that its probability density function is $f(x) = 2(1-x),\ x \in (0,1)$. In order to do so, a random sample of size $n = 1$ has been taken.

23. At a 5% significance level, the most powerful test indicates that the null hypothesis should be rejected if:

(A) $X \leq 2.24$ (B) $X \leq 0.224$ (C) $X \geq 0.224$ (D) $X \in (-0, 224, 0.224)^c$ (E) All false

24. The power of this test is:

(A) 0.224 (B) 0.398 (C) All false (D) 0.95 (E) 0.603

Questions 25 and 26 refer to the following exercise:

The random variable X follows a distribution with probability mass function given by:

$$P(X = 0) = 2\theta^2; \qquad P(X = 1) = 3\theta^2; \qquad P(X = 2) = 1 - 5\theta^2$$

In order to test the null hypothesis $\theta = 0.05$ against the alternative hypothesis $\theta = 0.40$, a random sample of size $n = 1$ has been taken and it is decided that the null hypothesis should be rejected if the resulting sample values are $X = 0$ ó $X = 1$.

25. The significance level for this test is:

(A) 0.25 (B) 0.75 (C) 0.0125 (D) 0.125 (E) 0.9875

26. The probability of type II error for this test is:

(A) 0.20 (B) 0.95 (C) 0.75 (D) 0.25 (E) 0.80

Questions 27 to 30 refer to the following exercise:

An individual is interested in buying a Play Station 2. Before doing so, he decides to ask its price at 11 different stores, obtaining a mean sample price of 125 euros with a sample standard deviation of 15 euros. We assume normality.

27. At the 90% confidence level, we can state that the Play Station 2 mean price is contained in the interval:

(A) (116.41, 133.59) (B) (118.50, 131.50) (C) (114.42, 135.58)
(D) (122.29, 127.72) (E) (120.36, 129.64)

28. At the 95% confidence interval, we can state that the variance of the price for the Play Station 2 is contained in the interval:

(A) (8.05, 50.77) (B) (135.25, 628.17) (C) (150.33, 892.44)
(D) (9.02, 41.88) (E) (120.73, 761.54)

29. If, at the 10% significance level, we wish to test the null hypothesis that the Play Station 2 mean price is $m = 130$, 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 Play Station 2 is $\sigma^2 = 800$, 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)

Let $X_1, X_2, \ldots, X_n$ be n independent r.v. such that, for each i, X_i follows a Poisson distribution with parameter $k_i\lambda$. That is, $X_i \in \mathcal{P}(k_i\lambda),\ i = 1, \ldots, n$, where the k_i's are positive known constants and $\lambda > 0$.

a) Find, **providing all relevant details**, the maximum likelihood estimator of λ.

b) Is it unbiased? Is it consistent? **Hint**: You can assume that $\lim_{n\to\infty} \sum_{i=1}^{n} k_i = +\infty$.

B (10 points, 25 minutes) The professor of a given course wished to find out if the number of students attending class is uniformly distributed among the five class days per week. The following data were obtained from a random sample of four complete weeks of classes:

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number of students	49	35	32	39	45

At the 5% significance level, is there any reason that leads the professor to believe that the number of students attending class can not be assumed to be uniformly distributed among the five class days per week?

C (10 points, 25 minutes) We wish to test the null hypothesis $H_0 : \lambda = \frac{1}{2}$ against the alternative hypothesis $H_1 : \lambda = 2$ for an exponential distribution with parameter λ, for which a random sample of size $n = 10$, $X_1, \ldots, X_{10}$, has been taken. At a 5% significance level, find, **providing all relevant details**, the most powerful critical region for this test, if the test statistic being used is $Z = \sum_{i=1}^{10} X_i$. **Hint**: If $Y \in \gamma(1/2, n/2) \Longrightarrow Y \in \chi^2_n$.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

A) Given that $X_i \in \mathcal{P}(k_i\lambda)$, we then have that

$$P(x_i) = \frac{e^{-k_i\lambda}(k_i\lambda)^{x_i}}{x_i!}, \quad x_i = 0, 1, 2, \cdots, k_i, \quad \lambda > 0.$$

a) In this way, the maximum likelihood function will be given by

$$L(\lambda) = P(x_1; k_1\lambda) \cdots P(x_2; k_2\lambda) \cdots P(x_n; k_n\lambda)$$

$$L(\lambda) = \frac{e^{-k_1\lambda}(k_1\lambda)^{x_1}}{x_1!} \cdots \frac{e^{-k_n\lambda}(k_n\lambda)^{x_n}}{x_n!}$$

$$L(\lambda) = \frac{e^{-\lambda \sum_{i=1}^n k_i} \lambda^{\sum_{i=1}^n x_i} \Pi_{i=1}^n k_i^{x_i}}{\Pi_{i=1}^n x_i!}$$

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

$$\ln L(\lambda) = -\lambda \sum_{i=1}^n k_i + \sum_{i=1}^n x_i \ln \lambda + \sum_{i=1}^n x_i \ln k_i - \sum_{i=1}^n \ln x_i!$$

Taking derivatives with respect to λ, we will have that:

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = 0$$

Therefore,

$$-\sum_{i=1}^n k_i + \sum_{i=1}^n x_i \frac{1}{\lambda} = 0,$$

so that

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

b) This estimator will be unbiased if $\mathrm{E}(\hat{\lambda}_{ML}) = \lambda$.

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

$$\mathrm{E}(\hat{\lambda}_{ML}) = \frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n k_i\lambda = \lambda$$

Therefore $\hat{\lambda}_{ML}$ is an unbiased estimator of λ. In order to verify if $\hat{\lambda}_{MV}$ is a consistent estimator of λ, we can see if these two sufficient conditions hold:

i) $\lim_{n\to\infty} \mathrm{E}(\hat{\lambda}_{ML}) = \lambda$

ii) $\lim_{n\to\infty} \mathrm{Var}(\hat{\lambda}_{ML}) = 0$

On the one hand, given that $\hat{\lambda}_{ML}$ i an unbiased estimator of λ, condition i) holds. On the other hand,

$$\lim_{n\to\infty} \mathrm{Var}(\hat{\lambda}_{ML}) = \lim_{n\to\infty} \frac{1}{\left(\sum_{i=1}^{n} k_i\right)^2} \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$\lim_{n\to\infty} \mathrm{Var}(\hat{\lambda}_{ML}) = \lim_{n\to\infty} \frac{1}{\left(\sum_{i=1}^{n} k_i\right)^2} \sum_{i=1}^{n} k_i\lambda = \lim_{n\to\infty} \frac{\left(\sum_{i=1}^{n} k_i\right)}{\left(\sum_{i=1}^{n} k_i\right)^2} \lambda = \lim_{n\to\infty} \frac{\lambda}{\sum_{i=1}^{n} k_i} = 0$$

Therefore, the two conditions for consistency hold and, thus, $\hat{\lambda}_{ML}$ is a consistent estimator of λ.

B) We have a **goodness of fit test to a completely specified distribution**. The data available for this specific test are as follows:

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number of students	49	35	32	39	45

Under the null hypothesis that the number of students attending class is uniformly distributed among the five class days per week (i.e., that it is the same), we have that $p_i = 1/5 = 0.20,\ i = 1, 2, 3, 4, 5$. Given that we have five class days per week or five classes, $k = 5$. With this information and in order to carry out the test, we build the following table:

Day	n_i	p_i	np_i	$\frac{(n_i-np_i)^2}{np_i}$
Monday	49	0.20	40	2.025
Tuesday	35	0.20	40	0.625
Wednesday	32	0.20	40	1.600
Thursday	39	0.20	40	0.025
Friday	45	0.20	40	0.625
	$n = 200$	1	$n = 200$	$z = 4.90$

Under the null hypothesis that the number of students attending class is uniformly distributed among the five class days per week, the test statistic $\sum_i \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{k-1}$, where k is the number of class days per week.

At the approximate $\alpha = 5\%$ significance level, the decision rule is to reject the null hypothesis if:

$$z > \chi^2_{k-1\,,\,0.05} = \chi^2_{4,0.05}$$

In this case, we have that:

$$z = 4.90 < 9.49 = \chi^2_{4,0.05},$$

so that, at the the approximate $\alpha = 5\%$ significance level, we do not reject the null hypothesis.

C) We have an exponential distribution with λ (i.e., $X \in \exp(\lambda)$), and we wish to test the null hypothesis $H_0 : \lambda = \lambda_0 = \frac{1}{2}$ against the alternative hypothesis $H_1 : \lambda = \lambda_1 = 2$. Given that we have a r.s. of size $n = 10$, by the Neyman-Pearson Lemma we have that the most powerful critical region for this test will be the one verifying, for a given constant $k > 0$, the following inequality between the likelihood functions $L(\cdot)$ under the null and alternative hypotheses:

$$\frac{L(\lambda_0)}{L(\lambda_1)} \leq k$$

In this case, we will have that:

$$\frac{\lambda_0 e^{-\lambda_0 x_1} \cdots \lambda_0 e^{-\lambda_0 x_n}}{\lambda_1 e^{-\lambda_1 x_1} \cdots \lambda_1 e^{-\lambda_1 x_n}} \leq k \Longrightarrow \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{(\lambda_1 - \lambda_0)\sum_{i=1}^n x_i} \leq k$$

Given that $\lambda_1, \lambda_0 > 0$:

$$\Longrightarrow e^{(\lambda_1 - \lambda_0)\sum_{i=1}^n x_i} \leq k_1, \quad (k_1 > 0)$$

After taking natural logarithms and using the fact that $\lambda_1 > \lambda_0$, we will have that:

$$(\lambda_1 - \lambda_0)\sum_{i=1}^n x_i \leq k_2, \quad (k_2 > 0) \Longrightarrow \sum_{i=1}^n x_i \leq C, \quad (C > 0)$$

Therefore, we will reject the null hypothesis if $\sum_{i=1}^n X_i \leq C$. Under the null hypothesis, we have that $\lambda = \frac{1}{2}$, so that $X_i \in \exp(\frac{1}{2}) \equiv \gamma(\frac{1}{2}, 1)$. In this way, we will have that $\sum_{i=1}^n X_i \in \gamma(\frac{1}{2}, n) \equiv \chi^2_{2n}$. Thus, for the specific case in which $n = 10$, we will have that $\sum_{i=1}^{10} X_i \in \gamma(\frac{1}{2}, 10) \equiv \chi^2_{20}$. Moreover, at the 5% significance level, we will have that:

$$\alpha = 0.05 = P\left(\sum_{i=1}^{10} X_i \leq C \,\middle|\, \lambda = \frac{1}{2}\right) \Longrightarrow 0.05 = P\left(\chi^2_{20} \leq C\right) \Longrightarrow C = \chi^2_{20,0.95} = 10.9,$$

and, from here, the most powerful critical region for this test will be $\mathrm{CR} = (0, 10.9]$.