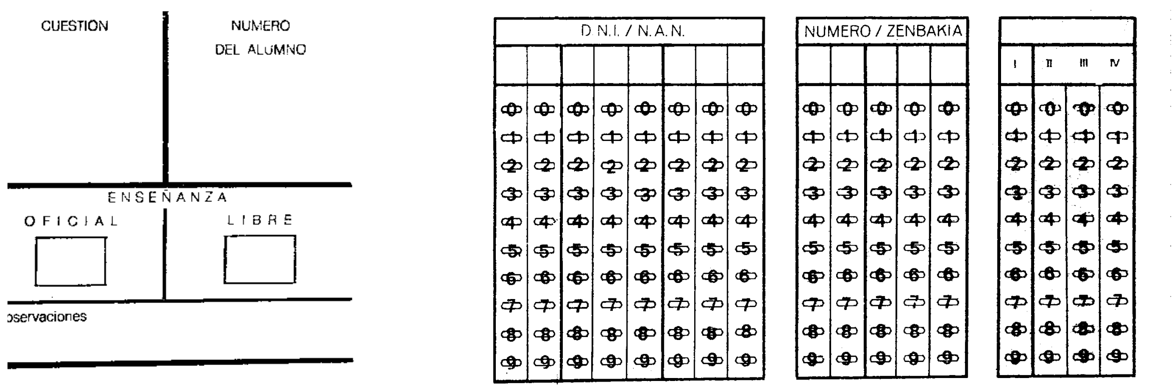
BUSINESS STATISTICS - Second Year Academic Year 2010-11
June 28, 2011

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) London (E) Pekin

Questions 2 and 3 refer to the following exercise:

In a given sports activity, the annual rate of accidents is of about 4 per thousand people. An insurance company has 3 000 insured clients that practice such sports activity.

2. A good approximation of the number of accident compensations the insurance company will have to pay in a given year is given by the distribution:

   (A) Poisson ($\lambda = 12$) (B) Binomial ($p = 0.4, n = 3\,000$) (C) Poisson ($\lambda = 48$)
   (D) N ($12, \sigma^2 = 12$) (E) Binomial ($p = 0.5, n = 3\,000$)

3. If all compensations were of 10 000 euros, the mean annual compensation payment that the insurance company will have to pay is:

   (A) 300 (B) 120 000 (C) 150 000 (D) 60 000 (E) 24 000

Questions 4 to 6 refer to the following exercise:

The number of clients that enter a given store each hour follows a Poisson distribution with mean 3.25. We assume independence between the different hours.

4. The probability that in a given hour exactly 5 clients enter the store is:

   (A) 0.8118 (B) 0.1172 (C) 0.9168 (D) 0.0252 (E) 0.7793

5. The most likely number of clients that would eventually enter the store in a given hour is:

   (A) 3 (B) 4 (C) 2 (D) 2.25 (E) 3.25

6. The approximate probability that in a 12-hour period fewer than 42 clients enter the store is:

   (A) 0.6255 (B) 0.6844 (C) 0.3745 (D) 0.3446 (E) 0.6554

Questions 7 to 9 refer to the following exercise:

The waiting time for clients in a given restaurant follows an exponential distribution with mean equal to 20 minutes. We assume independence between the different days.

7. If a client goes to that restaurant, what is the probability that the client has to wait less than 10 minutes to be seated?

   (A) 0.3624 (B) 0.6220 (C) 0.6065 (D) 0.6376 (E) 0.3935

8. If after waiting for 24 minutes in that restaurant without being seated a given client decides to leave, what is the probability that the client leaves?

   (A) 0.7 (B) 0.6 (C) 0.4 (D) 0.5 (E) 0.3

9. If a client goes to that restaurant 10 times, what is the probability that the client leaves more than 3 times without being seated ?

   (A) 0.9452 (B) 0.6496 (C) 0.9894 (D) 0.3504 (E) 0.0106

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

$$P(X_n = -2) = \frac{1}{3n}, \qquad P(X_n = 0) = 1 - \frac{2}{3n}, \qquad P(X_n = 2) = \frac{1}{3n}$$

   This sequence converges:

   (A) In distribution, probability and quadratic mean to $X = 2$

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

   (C) Only in distribution to $X = 2$

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

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

Questions 11 to 13 refer to the following exercise:

Let $X_1, X_2, \ldots, X_5$ five independent r.v., all having a $\gamma(0.25, 1)$ distribution.

11. The mean and variance of the random variable X_1 are, respectively:

   (A) 0.25 and 0.25 (B) 0.25 and 0.0625 (C) 4 and 4 (D) 4 and 16 (E) 4 and 0.25

12. The distribution of the r.v. $Z = X_1 + X_2 + \ldots + X_5$ is:

   (A) Exp$(\lambda = 0.25)$ (B) χ^2_{10} (C) All false (D) Exp$(\lambda = 1.25)$ (E) $\gamma(0.25, 1)$

13. What would be the distribution of the arithmetic mean $\overline{X} = \frac{X_1+\ldots+X_5}{5}$?

   (A) $\gamma(1.25, 5)$ (B) Exp$(\lambda = 1.25)$ (C) χ^2_{10} (D) $\gamma(0.25, 5)$ (E) Exp$(\lambda = 0.25)$

Questions 14 to 17 refer to the following exercise:

Let X_1, X_2, X_3 and X_4 be normally distributed independent random variables with respective means equal to -3, 0, -3 and 0, and variances equal to 1, 1, 4 and 4, respectively.

14. Let $Z = (X_1 + 3)^2 + \frac{1}{4}(X_3 + 3)^2$. Then, $P(Z < 4.61)$ is:

   (A) 0.1 (B) All false (C) 0.05 (D) 0.9 (E) 0.95

15. Let $W = \dfrac{X_2}{\sqrt{\frac{Z}{2}}}$. Then $P(W > 4.3)$ is:

   (A) 0.95 (B) 0.05 (C) 0.975 (D) 0.025 (E) 0.1

16. The value of k such that $P(|W| < k) = 0.6$ holds is equal to:

   (A) 0.617 (B) 2.92 (C) 1.061 (D) 1.89 (E) 0.142

17. The distribution of the random variable $Y = \dfrac{{X_4}^2}{2Z}$ is:

(A) t_1 (B) $F_{2,1}$ (C) χ^2_2 (D) $F_{1,2}$ (E) All false

Questions 18 and 19 refer to the following exercise:

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

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

In order to estimate the parameter θ, a r.s. of size $n = 10$ has been taken, providing the following results: three times the value -2, five times the value 0 and two times the value 2.

18. The maximum likelihood estimate of θ is:

(A) 1.4 (B) 0 (C) 0.75 (D) 0.25 (E) 0.37

19. The method of moments estimate θ is:

(A) 0.37 (B) 0 (C) 1.4 (D) 0.25 (E) 0.75

Questions 20 and 21 refer to the following exercise:

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

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

and mean equal to $2+\theta$. In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$ has been taken.

20. The maximum likelihood estimator of θ is:

(A) $\max\{x_i\}$ (B) $\overline{X}+2$ (C) $\overline{X}$ (D) $\overline{X}-2$ (E) $\min\{x_i\}$

21. The method of moments estimator of θ is:

(A) $\frac{1}{\overline{X}}$ (B) $\overline{X}$ (C) $\frac{\overline{X}}{2}$ (D) $\overline{X}+2$ (E) $\overline{X}-2$

Questions 22 and 23 refer to the following exercise:

We have a r.s. of size n from a Poisson distribution of parameter λ. In order to estimate λ, the estimator $\hat{\lambda} = \overline{X} + \frac{1}{n^2}$ is proposed.

22. This estimator of λ is:

(A) unbiased and asymptotically biased

(B) biased and asymptotically unbiased

(C) unbiased and asymptotically unbiased

(D) biased and asymptotically biased

(E) All false

23. Is this a consistent estimator of λ?

(A) No (B) – (C) Yes (D) – (E) –

Questions 24 and 25 refer to the following exercise:

Let X be a r.v. having a $N(m, \sigma^2 = 1)$ distribution. In order to be able to test the null hypothesis $H_0 : m = 5$, against the alternative hypothesis $H_1 : m > 5$, using $\overline{X}$ as the test statistic for it, we take the critical region as: $\text{CR} = (C, \infty)$.

24. For a significance level $\alpha = 0.10$, when the sample size has $n = 4$ observations, we would then have to take C equal to:

(A) 6.2800 (B) 5.8159 (C) 6.6500 (D) 5.6400 (E) 5.4079

25. For the value of C in the previous question, the power for $m = 5.5$ is equal to:

(A) 0.9406 (B) 0.6103 (C) 0.5714 (D) 0.3897 (E) 0.0594

Questions 26 to 28 refer to the following exercise:

In a marketing study we are interested in testing the null hypothesis H_0 indicating that al least 40% of the population is willing to buy a new product. In order to do so, a random sample of size $n = 20$ is taken, and the test statistic Z: number of people in the sample who are willing to buy the new product.

26. At the 10% significance level, the most appropriate critical region for this test is:

(A) $z \leq 4$ (B) $z \leq 3$ (C) All false (D) $z > 4$ (E) $z > 3$

27. The power of this test for $p = 0.3$ is:

(A) 0.8929 (B) 0.7625 (C) 0.1071 (D) 0.2375 (E) All false

28. If a r.s. of size $n = 100$ is taken and it is decided that the null hypothesis H_0 will be rejected if the number of people in the sample who are willing to buy the new product is smaller than 30, what is the approximate significance level for this new test?

(A) 0.052 (B) 0.016 (C) 0.026 (D) 0.979 (E) 0.337

Questions 29 and 30 refer to the following exercise:

We wish to test the null hypothesis that the benefits (in millions) of all firms in a given industry follow a normal distribution with mean 10 and variance 4. In order to do so, we have observed the benefits for 100 firms, from which 30 had benefits smaller than 8, 50 had benefits between 8 and 13, and the remaining 20 firms had benefits larger than 13.

29. Under the null hypothesis, approximately how many firms in the sample having benefits smaller than 8 would be expected?

(A) 25 (B) 16 (C) 84 (D) 0 (E) 32

30. Then, at the 5% significance level, the decision of the test will be:

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

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

A firm's manager wishes to determine if the probability that employees visit the firm's medical consulting room is the same for every working day in the week. Based on a random sample of four complete working weeks, we have obtained the following information:

Weekday	Number of visits to the medical consulting room
Monday	49
Tuesday	35
Wednesday	32
Thursday	39
Friday	45

At the 5% significance level, what would the firm's manager conclusion be about his/her initial hypothesis?

B. (10 points, 25 minutes)

It is known that the expense in computer equipment for firms having less than 250 employees is a normally distributed r.v. X. In order to estimate the mean expense, a r.s. of size $n = 15$ is taken, providing a mean expense of 6000 euros, and a standard deviation of 7500 euros.

i) Based on the distribution of $\overline{X}$, obtain, **providing all relevant details**, the theoretical expression for the $1 - \alpha$ confidence interval for the population mean expense, m, to be applied in this specific case.

ii) Using the information provided in this exercise, compute the 95% confidence interval for the population mean expense.

iii) At the 5% confidence level, test the null hypothesis $H_0 : m = 10000$ against alternative hypothesis $H_1 : m \neq 10000$.

C. (10 points, 25 minutes)

Let X be a r.v. having a $\gamma(a, 3)$ distribution; that is, with probability density function:

$$f(x; a) = \frac{a^3}{2} x^2 e^{-ax}, \qquad x > 0, \quad a > 0$$

We wish to test the null hypothesis $H_0 : a = \frac{1}{2}$ against the alternative hypothesis $H_1 : a = 2$. In order to do so, a random sample of size $n = 1$, X, is taken.

i) Obtain the form of the most powerful critical region for this test.

ii) Under the null hypothesis, what is the distribution of the test statistic X?

iii) At the 5% significance level, provide the specific most powerful critical region for this test.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A

It corresponds to a goodness of fit test to a completely specified distribution. Based on the information provided in the sample, we build the corresponding table as follows:

	n_i	p_i	np_i	$(n_i - np_i)$	$\frac{(n_i-np_i)^2}{np_i}$
Monday	49	0.2	40	+9	2.025
Tuesday	35	0.2	40	-5	0.625
Wednesday	32	0.2	40	-8	1.600
Thursday	39	0.2	40	-1	0.025
Friday	45	0.2	40	+5	0.625
	200	1	200	0	z =4.9

Under the null hypothesis of fit to the totally specified distribution stated by the firm's manager, we have that the test statistics $\sum_i \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{(K-1)}$, where K is the number of categories in which the sample has been divided (i.e., $K = 5$).

At the approximate 5% significance level, the decision rule will be to reject the null hypothesis if:

$$z > \chi^2_{(5-1)\,,\,0.05} = \chi^2_{4,0.05}$$

In this case, we have that:

$$z = 4.9 < 9.49 = \chi^2_{4,0.05}$$

so that, at the 5% significance level, we do not reject the null hypothesis of fit to the distribution specified by the firm's manager. That is, we can state that, based on the information provided in the sample, the probability that employees visit the firm's medical consulting room is the same for every working day in the week.

Exercise B

i) It corresponds to a confidence interval for the population mean in the case of a normal distribution and unknown variance.

We have a r.s. of size n: $X_1, X_2, \ldots, X_n$ taken from a $N(m, \sigma^2)$ distribution. Thus, we have that

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} \in N(0,1)$$

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

In addition, we have that the distributions for these two statistics are independent. We can now build the new test statistic

$$\frac{\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\left(\frac{nS^2}{\sigma^2}\right)/(n-1)}} = \frac{\bar{X} - m}{S/\sqrt{n-1}} \in t_{n-1},$$

so that:

$$P\left(-t_{\overline{n-1}|\alpha/2} < \frac{\bar{X} - m}{\frac{S}{\sqrt{n-1}}} < t_{\overline{n-1}|\alpha/2}\right) = 1 - \alpha$$

$$P\left(-t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}} < \bar{X} - m < t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

$$P\left(-\bar{X} - t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}} < -m < -\bar{X} + t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}} < m < \bar{X} + t_{\overline{n-1}|\alpha/2}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

Therefore, the $1 - \alpha$ confidence interval is given by:

$$\mathrm{CI}_{1-\alpha}(m) = \left(\bar{x} \pm t_{\overline{n-1}|\alpha/2}\frac{s}{\sqrt{n-1}}\right)$$

ii) In this case, we have that: $n = 15$, $\bar{x} = 6000$, $t_{\overline{n-1}|\alpha/2} = t_{\overline{14}|0.05/2} = 2.14$ and $s = 7500$. Therefore,

$$\begin{aligned}\mathrm{CI}_{0.95}(m) &= \left(6000 \pm 2.14\,\frac{7500}{\sqrt{14}}\right) = \\ &= (6000 \pm 4289.54) = \\ &= (1710.46\,,\ 10289.54)\end{aligned}$$

iii) Given that $m = 10\,000$ is contained in the aforementioned 95% confidence interval for the population mean, at the 5% significance level, we do not reject the null hypothesis H_0, in which the mean is equal to 10 000.

Exercise C

We have a r.v. X with probability density function given by:

$$f(x;a) = \frac{a^3}{2}\, x^2\, e^{-ax}, \qquad x > 0, \quad a > 0$$

We wish to test: $H_0 : a = 1/2$ against $H_1 : a = 2$. In order to do so, a r.s. of size $n = 1$ is taken.

i) In order to be able to determine the most powerful critical region (CR) for this test, we use the Neyman-Pearson Theorem, so that CR would be given by:

$$\frac{f(x; a = 1/2)}{f(x; a = 2)} \le K$$

$$\frac{\frac{1}{2}\,\left(\frac{1}{2}\right)^3\, x^2\; e^{-\frac{1}{2}x}}{\frac{1}{2}\,(2^3)\; x^2\; e^{-2x}} \le K$$

$$e^{-\left(\frac{1}{2}x - 2x\right)} \le K_1$$

$$\frac{3}{2}x \le K_2$$

$$x \le C$$

$$\mathrm{CR} = (0\,, C]$$

The most powerful critical region for this test will reject the null hypothesis if $X \le C,\ C > 0$.

ii) Under the null hypothesis, the distribution of the test statistic X is:

$$X \in \gamma\left(\frac{1}{2}, 3\right) \equiv \gamma\left(\frac{1}{2}, \frac{6}{2}\right) \equiv \chi^2_6$$

iii) For an $\alpha = 5\%$ significance level:

$$\begin{aligned} \alpha = 0.05 \ge P(\text{Reject } H_0 \mid H_0) &= \\ = P(X \le C \mid X \in \chi^2_6) & \\ \Rightarrow\ C = \chi^2_{6,\,0.95} = 1.64 & \\ \Rightarrow\ \mathrm{CR} = (0\,, 1.64] & \end{aligned}$$