STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year   Academic Year 2015-16
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. May 27, 2016

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.
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. If you have participated in the on-going evaluation process, you will need to obtain 12 points in each part of the exam to pass it. Otherwise, 15 points points in each part of the exam are required 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.

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:

In a given town, 20 retired workers have been randomly selected. It is known that, out of the 20 randomly selected, the number of retired workers that goes to the summer excursions offered by the IMSERSO follows a binomial distribution $Z \in b(p, n)$, so that its mean is equal to 12. We assume independence between the different retired workers in that town.

2. The probability that exactly 2 of the retired workers from that town go to the summer excursions is:

   (A) 0 (B) 0.0023 (C) 0.0031 (D) 0.0013 (E) 1

3. The probability that at least 8 of the retired workers from that town go to the summer excursions is:

   (A) 0.0355 (B) 0.0210 (C) 0.9790 (D) 0.0565 (E) 0.9435

4. The probability that at most 15 of the retired workers from that town go to the summer excursions is:

   (A) 0.9490 (B) 0.0350 (C) 0.1256 (D) 0.0746 (E) 0.0510

5. The characteristic function of the r.v. Z is:

   (A) $\psi_Z(u) = e^{12(e^{iu}-1)}$ (B) $\psi_Z(u) = e^{24(e^{iu}-1)}$ (C) $\psi_Z(u) = (0.60 + 0.40e^{iu})^{20}$
   (D) $\psi_Z(u) = (0.40 + 0.60e^{iu})$ (E) $\psi_Z(u) = (0.40 + 0.60e^{iu})^{20}$

Questions 6 and 7 refer to the following exercise:

Let Z be a random variable having a Poisson distribution with modes equal to 6 and 7.

6. The probability $P(Z = 2)$ is:

   (A) 0.0296 (B) 0.0620 (C) 0.0174 (D) 0.0223 (E) 0.0446

7. The probability $P(Z \geq 5)$ is:

   (A) 0.2851 (B) 0.8270 (C) 0.7179 (D) 0.1730 (E) 0.5543

Questions 8 to 10 refer to the following exercise:

Let $X_1, \ldots, X_{60}$ be independent and identically distributed r.v. having a Poisson distribution with variance equal to 2.

8. $P(X_{10} < 3)$ is:

   (A) 0.6767 (B) 0.9473 (C) 0.1429 (D) 0.3233 (E) 0.8571

9. $P(2 < X_{10} < 6)$ is:

   (A) 0.1263 (B) 0.3068 (C) 0.3188 (D) 0.1383 (E) 0.9955

10. If we define the r.v. $Y = \sum_{i=1}^{60} X_i$, then $P(Y \leq 91)$ is approximately equal to:

(A) 0.5948 (B) 0.0047 (C) 0.1583 (D) 0.9953 (E) 0.4052

11. Let X be a r.v. with characteristic function given by $\psi_X(u) = (1-0.5iu)^{-1}$. If we define the r.v. $Y = 4X$, then the distribution of the r.v. Y is:

(A) exp(1) (B) exp(2) (C) χ_1^2 (D) All false (E) χ_2^2

Questions 12 to 15 refer to the following exercise:

Let X, Y, Z and V be independent r.v. so that their probability distributions are: $X \in N(0,4)$, $Y \in N(0,1)$, $Z \in \gamma(\frac{1}{2}, 3)$ and $V \in \chi_{10}^2$.

12. The probability that the r.v. $W_1 = Z + V$ takes on values between 7.96 and 26.3 is:

(A) 0.90 (B) 0.05 (C) 0.975 (D) 0.10 (E) 0.95

13. The probability that the r.v. $W_2 = \frac{\sqrt{10}X}{2\sqrt{V}}$ takes on values smaller than or equal to 1.37 is:

(A) 0.20 (B) 0.90 (C) 0.80 (D) 0.10 (E) 0.95

14. If we define the r.v. $W_3 = \frac{10(Z+Y^2)}{7V}$, then the value of k such that $P(W_3 \geq k) = 0.90$ is:

(A) 0.55 (B) 2.70 (C) 0.41 (D) 2.41 (E) 0.37

15. The value of k such that $P(6.74 < V < k) = 0.50$ holds is:

(A) 16.0 (B) 18.3 (C) 9.34 (D) 12.5 (E) 13.7

16. Let X be a r.v. such that $X \in t_1$. The value of k such that $P(X \leq k) = 0.20$ holds is:

(A) -6.31 (B) -1.376 (C) 3.08 (D) -3.08 (E) 1.376

Questions 17 and 18 refer to the following exercise:

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

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

In order to estimate the parameter θ, a r.s. of size n, $X_1, X_2, \ldots, X_n$, has been taken.

17. The method of moments estimator of θ is:

(A) $\overline{X}$ (B) $\frac{2-\overline{X}}{7}$ (C) $\frac{2(1-\overline{X})}{7}$ (D) $\frac{7}{2(1-\overline{X})}$ (E) $\frac{1-2\overline{X}}{7}$

18. In order to be able to obtain an estimate of the parameter θ, a random sample of size $n = 10$ has been taken providing the following results: -2, 2, 0, 2, -2, 0, 0, -2, 2, 0. The maximum likelihood estimate of θ is equal to:

(A) 0.41 (B) 0.35 (C) 0.29 (D) 0.65 (E) 0.50

Questions 19 to 22 refer to the following exercise:

Let X be a r.v. having a uniform $U(2\theta - 1, 2\theta + 1)$ distribution. It is known that the mean and variance of this r.v. are equal to 2θ and $\frac{1}{3}$, respectively. We wish to estimate the parameter θ and, thus, in order to do so, a r.s. of size n, $X_1, \ldots, X_n$, has been taken.

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

(A) $\frac{\overline{X}}{2}$ (B) $4\overline{X}$ (C) $\overline{X}$ (D) $\frac{\overline{X}}{4}$ (E) $2\overline{X}$

20. Is the method of moments estimator of θ unbiased?

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

21. The bias for the method of moments estimator of θ is:

(A) θ (B) $\theta + 4$ (C) 0 (D) $\theta + 2$ (E) $\theta^2 + 4$

22. The variance for the method of moments estimator of θ is:

(A) $\frac{1}{12n}$ (B) $\frac{3}{4n}$ (C) $\frac{1}{12}$ (D) $\frac{3}{2n}$ (E) $\frac{1}{6n}$

Questions 23 and 24 refer to the following exercise:

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

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

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

23. At the $\alpha = 0.04$ significance level, the most powerful critical region for the test statistic X_1 will be:

(A) $(0, 0.20]$ (B) $[0.20, 1)$ (C) $(0, 0.04]$ (D) $[0.04, 0.96]^C$ (E) $[0.04, 0.96]$

24. For the same significance level, the probability of type II error for this test is approximately equal to:

(A) 0.90 (B) 0.15 (C) 0.10 (D) 0.96 (E) 0.80

Questions 25 and 26 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with parameter λ, $\mathcal{P}(\lambda)$, $\lambda > 0$. In order to test the null hypothesis $H_0 : \lambda = 0.5$ against the alternative hypothesis $H_1 : \lambda = 0.9$, a r.s. of size $n = 10$ has been taken, and $Z = \sum_{i=1}^{10} X_i$ is used a test statistic for this specific test.

25. At the $\alpha = 0.08$ significance level, the null hypothesis is rejected if:

(A) $Z \leq 8$ (B) $Z \geq 9$ (C) $Z \geq 8$ (D) $Z \leq 9$ (E) $Z \geq 7$

26. The probability of type II error for the previous critical region is:

(A) 0.4126 (B) 0.3239 (C) 0.4557 (D) 0.5874 (E) 0.5443

Questions 27 and 28 refer to the following exercise:

Let X and Y be two independent r.v. with corresponding distributions given by: $X \in N(m_X, \sigma_X^2)$ and $Y \in N(m_Y, \sigma_Y^2)$. In order to test the null hypothesis $H_0 : \sigma_X^2 = \sigma_Y^2$ against the alternative hypothesis $H_1 : \sigma_X^2 \neq \sigma_Y^2$, two r.s., each of size 25, have been taken from these populations, providing sample variances equal to $s_X^2 = 10.4$ and $s_Y^2 = 9.5$, respectively

27. A 90% confidence interval for (σ_X^2/σ_Y^2) is approximately given by:

    (A) (0.55, 2.17) (B) (0.64, 1.86) (C) (1.03, 1.16) (D) (0.59, 2.25) (E) (1.18, 2.18)

28. At the $\alpha = 10\%$ significance level, the decision of the test will be:

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

Questions 29 and 30 refer to the following exercise:

In a given town belonging to the Basque Autonomous Community (CAPV), in which there are 4800 officially registered inhabitants, we wish to estimate the proportion of inhabitants that would probably remain in the town during the summer season. We allow for a maximum error of ± 0.04 and assume a confidence level equal to 99%.

29. If we decide to take a random sample without replacement, the required sample size will be:

    (A) 716 (B) 534 (C) 1041 (D) 855 (E) 638

30. If in the above question, we wish to decrease or reduce the maximum error allowed for, the required sample size will be:

    (A) Larger (B) The same (C) More information is required (D) - (E) Smaller

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

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

$$f(x,\theta) = \begin{cases} \frac{3}{\theta^3}\, x^2 & \text{si } x \in [0,\theta] \\ 0 & \text{en otro caso} \end{cases}$$

We wish to estimate the parameter θ and, in order to do so, we have taken a r.s. of size n, $X_1, X_2, \ldots, X_n$.

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

ii) Is this estimator unbiased? If it is known that the variance of the r.v. X is $\sigma_X^2 = \dfrac{3}{80}\theta^2$, is this estimator consistent? Provide all relevant details about your answer to these questions.

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) Obtain, **providing all relevant details**, the form of the most powerful critical region for this test.

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

iii) For the above significance level and critical region, obtain the power for this test.

C. (10 points, 25 minutes)

A firm wishes to sell four types of multimedia devices and, for this purpose, it has been able to collect the following information:

Type	I	II	III	IV
Probabilities	θ^3	$(1-\theta)^3$	$3\theta(1-\theta)^2$	$3\theta^2(1-\theta)$

i) In order to be able to estimate these probabilities, a random sample of 80 individuals has been taken and has provided the following information: 12 individuals bought type I multimedia devices; 38 individuals bought type II multimedia devices; 16 individuals bought type III multimedia devices and 14 individuals bought type IV multimedia devices. Show **providing all relevant details**, that the maximum likelihood estimate for the parameter θ is $\hat{\theta}_{\text{ML}} = \frac{1}{3}$.

ii) At the 5% significance level, test the hypothesis that the probability distribution the firm has is the correct one. **Remark**: You should solve this exercise **without** grouping classes.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

The probability density function for the random variable X is:

$$f(x,\theta) = \begin{cases} \frac{3}{\theta^3}\ x^2 & \text{si } x \in [0,\theta] \\ \\ 0 & \text{en otro caso} \end{cases}$$

We wish to estimate the parameter θ and, in order to do so, a random sample of size n, $X_1, X_2, \ldots, X_n$, has been taken.

i) In order to be able to obtain the method of moments estimator of the parameter θ, we need to equate the first population moment $\alpha_1 = \mathrm{E}(X) = m$ to the first sample moment $a_1 = \overline{X}$. That is,

$$\alpha_1 = \mathrm{E}(X) = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}$$

Therefore, we need to compute the first population moment or mean for this r.v., α_1:

$$\alpha_1 = \mathrm{E}(X) = \int_0^\theta x f(x,\theta)dx = \int_0^\theta \frac{3}{\theta^3}\ x^3 dx = \frac{3}{4\theta^3}\left[x^4\right|_0^\theta = \frac{3\theta}{4}$$

From this result, we have that:

$$\alpha_1 = \mathrm{E}(X) = \frac{3\theta}{4} = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \Longrightarrow \hat{\theta}_{\mathrm{MM}} = \frac{4\overline{X}}{3}$$

ii) Unbiasedness:

In order to verify if this is an unbiased estimator of the parameter θ, we should have that $\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \theta$. In this case,

$$\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \mathrm{E}\left(\frac{4\overline{X}}{3}\right) = \frac{4}{3}\ \mathrm{E}(\overline{X}) = \frac{4}{3}\ \mathrm{E}(X) = \frac{4}{3}\ \left(\frac{3\theta}{4}\right) = \theta$$

Therefore, the method of moments estimator is unbiased for θ.

Consistency

In order to verify if this is a consistent estimator of the parameter θ, we have to compute the estimator's variance.

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) = \mathrm{Var}\left(\frac{4\overline{X}}{3}\right) = \left(\frac{4}{3}\right)^2 \mathrm{Var}(\overline{X}) = \left(\frac{16}{9}\right)\ \left(\frac{\mathrm{Var}(X)}{n}\right) = \left(\frac{16}{9}\right)\ \left(\frac{3\theta^2}{80n}\right) = \frac{\theta^2}{15n}$$

Given that it is an unbiased estimator whose variance tends to zero as n goes to infinity, the two required conditions for consistency hold and, therefore, we can state that the method of moments estimator is a consistent estimator for θ.

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:

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

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

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

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 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:

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

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

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

It is a **goodness of fit test to a partially specified distribution.** The information we have to perform this test is given below:

Type	I	II	III	IV
Probabilities	θ^3	$(1-\theta)^3$	$3\theta(1-\theta)^2$	$3\theta^2(1-\theta)$

i) In order to be able to estimate these probabilities, a random sample of 80 individuals has been taken and has provided the following information: 12 individuals bought brand A mobile telephones; 38 individuals bought brand B mobile telephones; 16 individuals bought brand C mobile telephones and 14 individuals bought brand D mobile telephones. With this information and to find the maximum likelihood estimator of θ, we have to write the likelihood function:

$$L(\theta) = \left[\theta^3\right]^{12} \left[(1-\theta)^3\right]^{38} \left[3\theta(1-\theta)^2\right]^{16} \left[3\theta^2(1-\theta)\right]^{14} = 3^{30}\theta^{80}(1-\theta)^{160}$$

We now compute its natural logarithm to obtain:

$$\ln L(\theta) = 30\ln(3) + 80\ln(\theta) + 160\ln(1-\theta)$$

If we take derivatives with respect to θ and equate this to zero, we will have that:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{80}{\theta} - \frac{160}{(1-\theta)} = 0 \Longrightarrow 80(1-\theta) = 160\theta \Longrightarrow 80 = 240\theta \Longrightarrow \hat{\theta}_{\text{ML}} = \frac{80}{240} = \frac{1}{3} = 0.3333$$

ii) Goodness of fit test to a partially specified distribution.

First of all, we have that the estimated probabilities, $\hat{p}_i$, for each one of the mobile telephone brands will be:

$$\hat{P}(\text{I}) = (0.3333)^3 = 0.0370; \qquad \hat{P}(\text{II}) = (1-0.3333)^3 = 0.2963$$
$$\hat{P}(\text{III}) = 3(0.3333)(1-0.3333)^2 = 0.4444; \qquad \hat{P}(\text{IV}) = 3(0.3333)^2(1-0.3333) = 0.2222$$

Given that we have estimated the parameter θ, we have that $h = 1$. Moreover, as we have four brands of mobile telephones, $k = 4$. With this information and in order to perform the test, we build the following table:

	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
Type I	12	0.0370	2.960	27.61
Type II	38	0.2963	23.704	8.62
Type III	16	0.4444	35.552	10.75
Type IV	14	0.2222	17.776	0.80
	$n = 80$	$\simeq 1$	$n \simeq 80$	$z = 47.78$

Under the null hypothesis that the probability distribution the firm has is the correct one, the test statistic $\sum_i \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \sim \chi^2_{k-h-1}$, where k is the number of types of multimedia devices ($k = 4$) and h is the number of estimated parameters ($h = 1$).

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

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

In this case:

$$z = 47.78 > 5.99 = \chi^2_{2,0.05}$$

so that, at a 5% approximate significance level, we reject the null hypothesis that the probability distribution the firm has for the different types of multimedia devices is the correct one.