STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2017-18
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. May 22, 2018

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 40 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 set of Law students from a given university, a random group of n students is selected. It is known that, among them, the number of students that will finish their studies within the established deadline follows a binomial distribution, $Z \in b(0.40, n)$, with variance $\text{Var}(Z) = 2.40$. We assume independence between the different Law students from that university.

2. The probability that exactly 6 students from that group finish their studies within the established deadline is:

   (A) 0.1114 (B) 0.0367 (C) 0.0425 (D) 0.8338 (E) 0.9452

3. The probability that more than 2 and no more than 7 students from that group finish their studies within the established deadline is:

   (A) 0.7779 (B) 0.5629 (C) 0.6054 (D) 0.8204 (E) 0.6160

4. If we now have that 85 students have been randomly selected, the number of students among them that is expected to finish their studies within the established deadline is:

   (A) 4 (B) 34 (C) 51 (D) 30 (E) 20.4

5. If we take the 85 students selected in the previous question, the approximate probability that at most 39 of them finish their studies within the established deadline is:

   (A) 0.8888 (B) 0.3974 (C) 0.8925 (D) 0.1112 (E) 0.6026

Questions 6 and 7 refer to the following exercise:

Let X be a r.v. having a Poisson distribution, so that $P(X = 2) = 0.224042$ and $P(X = 1) = 0.149361$.

6. $P(X = 5)$ is:

   (A) 0.1008 (B) 0.9161 (C) 0.0361 (D) 0.7851 (E) 0.9834

7. $P(3 \leq X < 7)$ is:

   (A) 0.3193 (B) 0.3409 (C) 0.5649 (D) 0.9955 (E) 0.5433

Questions 8 to 10 refer to the following exercise:

The number of clients arriving at the customer service section for a given firm follows a Poisson distribution with mean equal to 3.5 clients per hour. We assume independence between the clients arriving at different hours.

8. The probability that, in a given hour, 4 clients arrive at the firm's customer service section is:

   (A) 0.7552 (B) 0.1954 (C) 0.2534 (D) 0.1888 (E) 0.1145

9. The probability that, in a two-hour period, at least 8 clients arrive at the firm's customer service section is:

(A) 0.5987 (B) 0.7291 (C) 0.4013 (D) 0.2709 (E) 0.5503

10. The approximate probability that, in an eight-hour working day, at least 23 clients arrive at the firm's customer service section is:

(A) 0.1492 (B) 0.5793 (C) 0.8051 (D) 0.4207 (E) 0.8508

Questions 11 and 12 refer to the following exercise:

Let X and Y be independent r.v. such that $X \in \exp(\frac{1}{2})$ and $Y \in \gamma(\frac{1}{2}, 4)$.

11. The probability that the r.v. X takes on values between 0.211 and 5.99 is:

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

12. If we define the r.v. $Z = \frac{Y}{4X}$, the probability that the r.v. Z takes on values smaller than 9.37 is:

(A) 0.10 (B) 0.01 (C) 0.95 (D) 0.90 (E) 0.05

Questions 13 to 15 refer to the following exercise:

Let X, Y, Z and V be four independent r.v. so that their probability density functions are as follows: $X \in N(0,1)$, $Y \in N(0,9)$, $Z \in \chi^2_5$ and $V \in \gamma(\frac{1}{2}, 5)$.

13. If we define the r.v. $W_1 = \dfrac{3X}{\sqrt{Y^2}}$, the value of k such that $P(W_1 < k) = 0.10$ is:

(A) -3.08 (B) 6.31 (C) 3.08 (D) -6.31 (E) 12.71

14. If we define the r.v. $W_2 = \frac{2Z}{V}$, the approximate value of k such that $P(W_2 \leq k) = 0.10$ is:

(A) 3.30 (B) 2.52 (C) 0.48 (D) 0.40 (E) 0.30

15. If we define the r.v. $W_3 = Z + V$, then $P(W_3 < 18.2)$ is:

(A) 0.25 (B) 0.50 (C) 0.75 (D) 0.90 (E) 0.95

16. Let X be a r.v. with probability density function $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$. In order to estimate the parameter θ, a r.s. of size n, $X_1, X_2, \ldots, X_n$, has been taken. The method of moments estimator of θ, $\hat{\theta}_{\mathrm{MM}}$, is:

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

Questions 17 and 18 refer to the following exercise:

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

$$P(X=-1) = \theta \qquad P(X=0) = (1-2\theta) \qquad P(X=1) = \theta$$

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

18. To obtain an estimate of the parameter θ, a random sample of size $n = 10$ has been taken, providing the following results: 0, 0, -1, 1, -1, 0, 0, 1, 1, 1. The maximum likelihood estimate of θ is equal to:

(A) 0.30 (B) 0.15 (C) 0.45 (D) 0.75 (E) 0.60

Questions 19 to 21 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with parameter θ. In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$, has been taken, and $\hat{\theta} = (3X_1 + 4X_2 + \ldots + 4X_{n-1} + 3X_n)/4n$ is proposed as an estimator of θ.

19. The bias of the proposed estimator is:

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

20. The proposed estimator is:

(A) Unbiased (B) Efficient (C) Unbiased and asymptotically biased

(D) Biased and asymptotically biased (E) Biased and asymptotically unbiased

21. The variance of the proposed estimator is:

(A) $\frac{\theta}{n^2}$ (B) $\frac{\theta(n-1)}{n^2}$ (C) $\frac{\theta(4n-5)}{4n^2}$ (D) $\frac{\theta(8n-7)}{8n^2}$ (E) $\frac{\theta}{2n}$

Questions 22 and 23 refer to the following exercise:

Let X be a r.v. having a uniform distribution on the interval $(-4, \theta)$; that is, $X \in U(-4, \theta)$. Based on a r.s. of size $n = 1$, X_1, we wish to test the null hypothesis $H_0 : \theta = 0$ against the alternative hypothesis $H_1 : \theta = 2$, and reject the null hypothesis H_0 if $X_1 \geq C$.

22. At the $\alpha = 0.10$ significance level, we reject the null hypothesis if X_1 is larger than or equal to:

(A) -0.10 (B) 0.10 (C) -0.50 (D) -0.40 (E) 0.40

23. For the critical region above, the power for this test is:

(A) 0.40 (B) 0.30 (C) 0.80 (D) 0.60 (E) 0.50

Questions 24 and 25 refer to the following exercise:

Let X be a r.v. having a binary distribution $b(p)$. To test $H_0 : p = 0.50$ against $H_1 : p < 0.50$, a r.s. of size $n = 15$ has been taken and it is decided that $Z = \sum_{i=1}^{15} X_i$ is used as test statistic.

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

(A) $Z \geq 5$ (B) $Z \leq 5$ (C) $Z \leq 6$ (D) $Z \geq 4$ (E) $Z \leq 4$

25. For the critical region above and $p = 0.20$, the probability of type II error is:

(A) 0.1642 (B) 0.9389 (C) 0.8358 (D) 0.0611 (E) 0.0592

Questions 26 to 28 refer to the following exercise:

An individual is interested in buying a new computer. Before doing so, s/he decides to ask for its price at 21 stores in Bilbao (B), obtaining a sample mean price of 875 euros with a **sample standard deviation** of 75 euros. In addition, and independently, s/he decides to ask for its price at 16 stores in Vitoria (V), obtaining a sample mean price of 925 euros with a **sample standard deviation** of 80 euros. We assume normality and that the population variances for the prices in the two cities are unknown but equal.

26. At the 95% confidence level and taking into account that $t_{35,0.025} = 2.03$, $t_{35,0.05} = 1.69$, $t_{35,0.10} = 1.30$, $t_{0.025} = 1.96$, $t_{0.05} = 1.64$ and $t_{0.10} = 1.30$, we can state that the computer's mean price difference between Bilbao and Vitoria, $m_B - m_V$, is included in the interval:

(A) $(-103.47,\ 3.47)$ (B) $(-68.80,\ -31.20)$ (C) $(-95.82,\ -4.18)$ (D) $(-35.38,\ 64.62)$ (E) $(-56.26,\ -43.74)$

27. At the 90% confidence level, we can state that the ratio of the population variances for the computer's price in Bilbao and Vitoria, σ_B^2/σ_V^2, is included in the interval:

(A) $(0.40,\ 2.03)$ (B) $(0.37,\ 1.90)$ (C) $(0.40,\ 1.90)$ (D) $(0.37,\ 2.03)$ (E) $(0.45,\ 1.90)$

28. At the 95% confidence level, we can state that the computer's mean price in Vitoria is included in the interval:

(A) $(885.8,\ 964.2)$ (B) $(913.6,\ 936.4)$ (C) $(754.6,\ 1095.4)$ (D) $(915.2,\ 934.8)$ (E) $(881.0,\ 969.0)$

Questions 29 and 30 refer to the following exercise:

We wish to estimate the mean alcoholic drink consumption in a given city during the weekends, considering a population stratified in three strata according to age. It is known that the first stratum includes 600 individuals, with quasi-standard deviation $\sigma_1^* = 20$, that the second stratum includes 300 individuals, with quasi-standard deviation $\sigma_2^* = 30$, and that the third stratum includes 100 individuals, with quasi-standard deviation $\sigma_3^* = 10$. We wish to take a sample of 300 individuals.

29. If we decide to use uniform allocation, the corresponding sample size for the first stratum will be:

(A) 100 (B) 150 (C) 200 (D) 125 (E) 50

30. If we decide to use proportional allocation, the corresponding sample size for the first stratum will be:

(A) 90 (B) 150 (C) 30 (D) 180 (E) 120

EXERCISES (Time: 70 minutes)

A. (10 points, 20 minutes)

The following table includes information on the probability mass function a discrete r.v. X has under the null hypothesis ($P_0(x)$) and under the alternative hypothesis ($P_1(x)$).

X	0	1	2	3	4	5
$P_0(x)$	0.05	0	0.45	0	0.45	0.05
$P_1(x)$	0.15	0.30	0	0.30	0	0.25

A random sample of size $n = 1$ will be used to test the null hypothesis $H_0 : P(x) = P_0(x)$ against the alternative hypothesis $H_1 : P(x) = P_1(x)$.

i) Would you include the points $X = \{1, 3\}$ in the critical region? Explain why or why not.

ii) Would you include the points $X = \{2, 4\}$ in the critical region? Explain why or why not.

iii) At the 10% significance level and providing all relevant details used to obtain the required response, find the most powerful critical region for this test, and compute its probability of type II error. **Remark**: Before providing an answer to this item, take into account your responses to the previous items in this exercise.

B. (10 points, 25 minutes)

The frequency table displayed below shows the number of cars families have in a r.s. of 1200 families obtained in a given country.

Number of cars	0	1	2	3	≥ 4
Number of families	405	457	223	85	30

At the 10% significance level, test the hypothesis that the number of cars families have in that country follows a Poisson distribution. For estimation purposes, you can consider that, for families having 4 or more cars, the number of cars they have is exactly 4.

C. (10 points, 25 minutes)

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

$$f_X(x; \theta) = \frac{x}{\theta^2} e^{-(x^2/2\theta^2)}, \; x > 0, \; \theta > 0$$

Its is known that the mean and variabce for this r.v. are $\mathrm{E}(X) = \sqrt{\frac{2}{\pi}}\,\theta$ and $\mathrm{Var}(X) = \left(2 - \frac{\pi}{2}\right)\,\theta^2$. Let $X_1, \ldots, X_n$ be a r.s. from this r.v.

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

ii) Obtain, **providing all relevant details**, the maximum likelihood estimator of θ.

iii) If a r.s. of size $n = 10$ has been taken, providing the values $7, 8, 3, 5, 2, 1, 6, 4, 3, 2$, obtain a method of moments estimate of θ.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

We wish to test the null hypothesis that X is a r.v. 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
$P_0(x)$	0.05	0	0.45	0	0.45	0.05
$P_1(x)$	0.15	0.30	0	0.30	0	0.25

We have taken a r.s. of size$n = 1$; that is, we observe X.

i) Would you include the points $X \in \{1, 3\}$ in the critical region?

Given that, under the probability mass function for the null hypothesis $P_0(x)$, these points have probability zero, the r.v. cannot take these values under the null hypothesis. Therefore, the points $X \in \{1, 3\}$ are rejection points for H_0 and, thus, they should **always** be included in the critical region for this test.

ii) Would you include the points $X \in \{2, 4\}$ in the critical region?

Given that, under the probability mass function for the alternative hypothesis $P_1(x)$, these points have probability zero, the r.v. cannot take these value under the alternative hypothesis, but it can under the null hypothesis. Therefore, the points $X \in \{2, 4\}$ are points of no rejection for H_0 and, thus, they should **never** be included in the critical region for this test.

iii) At the $\alpha = 0.10$ significance level and **taking into account the responses we have provided to the previous items**, we have that the possible critical regions for this test are $\mathrm{CR}_1 = \{0, 1, 3\}$, $\mathrm{CR}_2 = \{1, 3, 5\}$ and $\mathrm{CR}_3 = \{0, 1, 3, 5\}$. This is so because the corresponding probabilities of type I error for these critical regions are:

$$P(X \in \mathrm{CR}_1|P_0) = P(X = 0, 1, 3|P_0) = 0.05 + 0 + 0 = 0.05 \leq \alpha = 0.10$$

$$P(X \in \mathrm{CR}_2|P_0) = P(X = 1, 3, 5|P_0) = 0 + 0 + 0.05 = 0.05 \leq \alpha = 0.10$$

$$P(X \in \mathrm{CR}_3|P_0) = P(X = 0, 1, 3, 5|P_0) = 0.05 + 0 + 0 + 0.05 = 0.10 \leq \alpha = 0.10$$

However, given that CR_3 includes the other two critical regions, we should only consider this CR, which will obviously be the most powerful one. In this case,

$$\text{Power} = P(X \in \mathrm{RC}_3|P_1) = P(X = 0, 1, 3, 5|P_1) = 0.15 + 0.30 + 0.30 + 0.25 = 1$$

$$\beta = P(\text{Error tipo II}) = 1 - \text{Potencia} = 1 - 1 = 0$$

Exercise B

In this case, and given that we have to estimate the parameter λ, we have a goodness of fit test to a partially specified distribution.

$H_0 : X \in \mathcal{P}(\lambda)$, where λ is an unknown parameter we need to estimate by maximum likelihood, and

$H_1 : X \notin \mathcal{P}(\lambda)$

$$\hat{\lambda}_{ML} = \overline{x} = \frac{(0 \times 405) + (1 \times 457) + (2 \times 223) + (3 \times 85) + (4 \times 30)}{1200} = \frac{1278}{1200} = 1.065$$

The estimated probabilities $\hat{p}_i$ can be computed from the probability mass function:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1 \ldots, \qquad \lambda > 0,$$

recalling that:

$$P(X = x) = \frac{\lambda}{x}\, P(X = x - 1)$$

In this way, we have that:

$P(X = 0) = \frac{e^{-1.065}\ (1.065)^0}{0!} = e^{-1.065} = 0.3447$

$P(X = 1) = \left(\frac{1.065}{1}\right)\ P(X = 0) = 0.3671$

$P(X = 2) = \frac{1.065}{2}\ P(X = 1) = 0.1955$

$P(X = 3) = \frac{1.065}{3}\ P(X = 2) = 0.0694$

$P(X \geq 4) = 1 - F(X = 3) = 1 - 0.9767 = 0.0233$

By using this information, we build the corresponding table:

	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
0	405	0.3447	413.64	0.1805
1	457	0.3671	440.52	0.6165
2	223	0.1955	234.60	0.5736
3	85	0.0694	83.28	0.0355
≥ 4	30	0.0233	27.96	0.1488
Total	$n = 1200$	1	$n = 1200$	$z = 1.5549$

Under the null hypothesis, the test statistic $\sum_i \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$ converges to a $\chi^2_{(k-h-1)}$ distribution, where k is the number of classes in which the sample is divided ($k = 5$) and h is the number of estimated parameters ($h = 1$).

The decision rule is: at the approximate 10% significance level, reject the null hypothesis if

$$z \geq \chi^2_{(5-1-1)\,,\,0.10} = \chi^2_{3,0.10}$$

In this case,

$$1.5549 < 6.25 = \chi^2_{(5-1-1)\,,\,0.10} = \chi^2_{3,0.10}$$

so that, at the approximate 10% significance level, we do not reject the null hypothesis that the number of cars families have in that country follows a Poisson distribution.

Exercise C

The probability density function for the r.v. X is

$$f_X(x;\theta) = \frac{x}{\theta^2} e^{-(x^2/2\theta^2)}, \ x > 0, \ \theta > 0$$

In order to estimate the parameter θ, we have that the r.v. $X_1, \ldots, X_n$ are independent and identically distributed (I.I.D.), having the probability density function provided above.

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 will have that:

$$\alpha_1 = \mathrm{E}(X) = \sqrt{\frac{2}{\pi}}\ \theta = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \Longrightarrow \hat{\theta}_{\mathrm{MM}} = \sqrt{\frac{\pi}{2}}\ \overline{X}$$

ii) To obtain the likelihood function, we have that:

$$L(\vec{x},\theta) = f(x_1,\theta)\ \ldots\ f(x_n,\theta) = \left[\frac{x_1}{\theta^2} e^{-(x_1^2/2\theta^2)}\right] \cdots \left[\frac{x_n}{\theta^2} e^{-(x_n^2/2\theta^2)}\right] = \left[\frac{\Pi_{i=1}^n x_i}{\theta^{2n}}\right] e^{-\left(\sum_{i=1}^n x_i^2/2\theta^2\right)}$$

In order to be able to obtain the maximum likelihood estimator of θ, we have to maximize the logarithm of the likelihood function. We now compute its natural logarithm to obtain:

$$\ln L(\vec{x},\theta) = \ln\left(\Pi_{i=1}^n x_i\right) - 2n\ln(\theta) - \left[\frac{\sum_{i=1}^n x_i^2}{2\theta^2}\right]$$

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

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta} = -\frac{2n}{\theta} + \frac{\left[\sum_{i=1}^n x_i^2\right]}{\theta^3} = 0 \Longrightarrow -2n + \frac{\left[\sum_{i=1}^n x_i^2\right]}{\theta^2} = 0$$

so that,

$$2n = \frac{\left[\sum_{i=1}^n x_i^2\right]}{\theta^2} \Longrightarrow \theta^2 = \frac{\left[\sum_{i=1}^n x_i^2\right]}{2n} \Longrightarrow \hat{\theta}_{\mathrm{ML}} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{2n}}$$

iii) If a r.s. of size $n = 10$ has provided the values $7, 8, 3, 5, 2, 1, 6, 4, 3, 2$, a method of moments estimate of θ will be:

$$\hat{\theta}_{\mathrm{MM}} = \sqrt{\frac{\pi}{2}}\ \overline{X} = \sqrt{\frac{\pi}{2}}\ \left[\frac{(7+8+3+5+2+1+6+4+3+2)}{10}\right] = \sqrt{\frac{\pi}{2}} \times \left[\frac{41}{10}\right] = \sqrt{3.14/2} \times (4.1) = 5.1373$$