STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2021-22
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. May 18, 2022

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 in which the selected answer is located appropriately**. Please make sure you know the answer you wish to mark before doing it. You can always cancel your mark by simply filling the rectangle below the answer you wish to cancel, and then fill the rectangle in which the new selected answer is located appropriately. 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, whereas each question incorrectly answered will not penalize your grade in any form. Therefore, you must answer all of the questions included in the exam**.
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 15 and 12 points in each part of the exam to pass it. Otherwise, 18 and 15 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 4 refer to the following exercise:

The daily expense, in euros, for a family in a given city follows a uniform $U(20, 120)$ distribution. It is established that the expense for a given family may be considered excessive if it is larger than 80 euros. We assume independence in the expenses for the different families in that city.

2. If a random sample of 10 families in that city is taken, the probability that at least 8 of them have an excessive daily expense is:

   (A) 0.0123 (B) 0.0017 (C) 0.0548 (D) 0.9983 (E) 0.9877

3. In the same sample of 10 families, the probability that exactly 5 families **do not** have an excessive daily expense is:

   (A) 0.8338 (B) 0.1114 (C) 0.2007 (D) 0.2508 (E) 0.9452

4. If we now take a random sample of 100 families in that city, the approximate probability that at least 30 of them have an excessive daily expense is:

   (A) 0.9838 (B) 0.6700 (C) 0.8413 (D) 0.3300 (E) 0.0162

Questions 5 and 6 refer to the following exercise:

Let Z be a r.v. having a binomial distribution with characteristic function $\psi_Z(u) = (0.35 + 0.65e^{iu})^n$ and variance equal to 4.55.

5. $P(Z = 14)$ is:

   (A) 0.1712 (B) 0.1272 (C) 0.1844 (D) 0.1532 (E) 0.1112

6. $P(Z \geq 1)$ is:

   (A) 0 (B) 0.9844 (C) 0.9755 (D) 0.0345 (E) 1

Questions 7 to 10 refer to the following exercise:

The number of clients arriving, **every half an hour**, at a home appliances store follows a Poisson distribution with $2P(X = 4) = P(X = 3)$. We assume independence between the arrivals of the different clients at the store.

7. The probability that, in a half an hour period, at least 4 clients arrive at the store is:

   (A) 0.0166 (B) 0.0527 (C) 0.1429 (D) 0.9473 (E) 0.8571

8. The probability that, in a two-hour period, at most 7 clients arrive at the store is:

   (A) 0.9489 (B) 0.4530 (C) 0.9989 (D) 0.5470 (E) 0.0511

9. The most likely number of clients expected to arrive at the store in a **one-hour period** is:

(A) 3 (B) 4 (C) 3 and 4 (D) 4 and 5 (E) 5 and 6

10. The approximate probability that, in a ten-hour period, more than 30 clients arrive at the store is:

(A) 0.5948 (B) 0.0668 (C) 0.8413 (D) 0.4052 (E) 0.9332

11. Let X be a random variable having a Student's t distribution with n degrees of freedom, $t_{\overline{n}|}$. We then have that $P(t_{\overline{n}|1-\frac{\alpha}{4}} < X < t_{\overline{n}|1-\frac{\alpha}{2}})$ is:

(A) $1-\frac{\alpha}{4}$ (B) $1-\alpha$ (C) $\frac{\alpha}{2}$ (D) $\frac{\alpha}{4}$ (E) $1-\frac{\alpha}{2}$

12. Let X be a r.v. such that $X \in \gamma(2,4)$. If we define the r.v. $Y = 4X$, then the value of $P(Y > 5.07)$ is:

(A) 0.75 (B) 0.95 (C) 0.90 (D) 0.10 (E) 0.25

Questions 13 and 14 refer to the following exercise:

The lifetime or duration for an halogen, **in hundreds of hours**, follows an exponential distribution with mean equal to 4. We assume independence between the different halogens.

13. The probability than an halogen of this type lasts less than one thousand hours is:

(A) 0.0821 (B) 0.0001 (C) 0.9179 (D) 0.9999 (E) 0.9505

14. If the manufacturer for these halogens claims that his/her products have an established minimum lifetime, what should this established lifetime be so that the probability that the manufacturer's claim is correct for a randomly selected halogen is equal to 0.98?

(A) 8.08 hours (B) 0.02 hours (C) 10.2 hours (D) 2.02 hours (E) 0.08 hours

Questions 15 to 17 refer to the following exercise:

Let X, Y and Z be three independent r.v. such that their distributions are as follows: $X \in N(0, \sigma^2 = 4)$, $Y \in \gamma(\frac{1}{2}, 4)$ and $Z \in \chi^2_7$.

15. If we define the r.v. $W_1 = \frac{X^2}{4} + Y$, then $P(4.17 < W_1 < 14.7)$ is:

(A) 0.90 (B) 0.80 (C) 0.10 (D) 0.75 (E) 0.85

16. If we define the r.v. $W_2 = \dfrac{\sqrt{2}X}{\sqrt{Y}}$, the value of k such that $P(W_2 < k) = 0.95$ is:

(A) 1.86 (B) −2.31 (C) 1.40 (D) 2.31 (E) −1.86

17. If we define the r.v. $W_3 = \frac{8Z}{7Y}$, then $P(0.2681 < W_3 < 3.50)$ is:

(A) 0.95 (B) 0.85 (C) 0.90 (D) 0.80 (E) 0.99

Questions 18 and 19 refer to the following exercise:

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

$$P(X=0)=\frac{1}{2\theta};\quad P(X=1)=\frac{2}{\theta};\quad P(X=2)=1-\frac{5}{2\theta}$$

In order to estimate the parameter θ a r.s. of size $n=15$ has been taken, providing the following sample values :

X	0	1	2
frequency	6	5	4

18. The maximum likelihood estimate of θ is:

(A) $\frac{75}{22}$ (B) $\frac{60}{22}$ (C) $\frac{25}{4}$ (D) $\frac{60}{38}$ (E) $\frac{75}{38}$

19. The method of moments estimate of θ is:

(A) $\frac{45}{43}$ (B) $\frac{30}{17}$ (C) $\frac{45}{17}$ (D) $\frac{45}{13}$ (E) $\frac{75}{22}$

Questions 20 and 21 refer to the following exercise:

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

$$f(x)=\begin{cases}\dfrac{\theta\, 2^{\theta}}{x^{\theta+1}}, & x\geq 2,\quad \theta>1\\ 0, & \text{otherwise}\end{cases},$$

for which we have that $\mathrm{E}(X)=\frac{2\theta}{(\theta-1)}$. In order to estimate the parameter θ, a r.s. of size n, $X_1,\ldots,X_n$, has been taken.

20. The method of moments estimator of θ is:

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

21. The maximum likelihood estimator of θ is:

(A) $\frac{n}{\ln(\prod_{i=1}^{n}X_i)-n\ln(2)}$ (B) $\frac{n}{\ln(\prod_{i=1}^{n}X_i)+n\ln(2)}$ (C) $\frac{n}{\ln(\prod_{i=1}^{n}X_i)}$ (D) $\frac{n}{\ln(\prod_{i=1}^{n}X_i)-\ln(2)}$ (E) All false

Questions 22 and 23 refer to the following exercise:

Let X be a r.v. having a normal $N(\theta,\sigma^2=1)$ distribution. In order to estimate the parameter θ, a r.s. of size n, $X_1,\ldots,X_n$, has been taken and $\hat{\theta}=(5X_1+3X_2+\ldots+3X_{n-1}+3X_n)/(3n+2)$ is proposed as an estimator for θ.

22. The bias of the proposed estimator is:

(A) 0 (B) $\dfrac{2\theta}{(3n+2)}$ (C) $-\dfrac{5\theta}{(3n+2)}$ (D) $\dfrac{2\theta}{3n}$ (E) $-\dfrac{3n\theta}{(3n+2)}$

23. The variance of the proposed estimator is:

(A) $\dfrac{(8n+2)}{(3n+2)^2}$ (B) $\dfrac{1}{n}$ (C) $\dfrac{(9n+16)}{(3n+2)^2}$ (D) $\dfrac{(9n+25)}{(3n+2)^2}$ (E) $\dfrac{16}{(3n+2)^2}$

Questions 24 and 25 refer to the following exercise:

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

$$f(x,\theta) = \frac{(\theta+1)}{4^{\theta+1}}\, x^{\theta}, \qquad 0 \le x \le 4, \qquad \theta > 0$$

In order to test the null hypothesis $\theta = 0$ against the alternative hypothesis $\theta = 1$, a r.s. of size $n = 1$ is taken.

24. The most powerful critical region for that observation and for a given significance level is of the form:

(A) $X \le C$ (B) $X \in (C_1, C_2)^c$ (C) All false (D) $X \in (C_1, C_2)$ (E) $X \ge C$

25. At the $\alpha = 0.10$ significance level, we reject H_0 if:

(A) $X \ge 0.40$ (B) $X \ge 3.60$ (C) $X \le 0.40$ (D) $X \le 3.60$ (E) $X \in (0.40, 3.60)$

Questions 26 and 27 refer to the following exercise:

Let X be a random variable having a binary distribution with parameter p. In order to test the null hypothesis $H_0 : p \le 0.30$ against the alternative hypothesis $H_1 : p > 0.30$, a r.s. of size $n = 10$ has been taken, and $Z = \sum_{i=1}^{10} X_i$ is used as test statistic.

26. At the $\alpha = 0.05$ significance level, we reject the null hypothesis if:

(A) $Z \ge 5$ (B) $Z \ge 6$ (C) $Z \ge 7$ (D) $Z \le 5$ (E) $Z \le 6$

27. For $p = 0.60$, the probability of type II error for this test is:

(A) 0.3669 (B) 0.9527 (C) 0.6331 (D) 0.8497 (E) 0.8338

Questions 28 and 29 refer to the following exercise:

A given firm buys ten backpacks from the M1 brand and thirteen from the M2 brand. We assume that the backpack prices follow a normal distribution, and that these distributions are independent from each other. These samples provided the corresponding sample mean prices and standard deviations as follows: $\overline{x}_{M1} = 85.70$, $\overline{x}_{M2} = 92.40$, $s_{M1} = 10.3$ and $s_{M2} = 9.70$.

28. A 90% confidence interval for the ratio of the variances, $\sigma^2_{M1}/\sigma^2_{M2}$ is:

(A) $(0.4130, 3.5503)$ (B) $(1.1564, 3.2381)$ (C) $(0.1000, 0.9000)$ (D) $(0.4331, 3.3653)$ (E) $(0.4130, 3.2381)$

29. If we assume that the variances are equal, a 90% confidence interval for the difference of the mean prices for the two brands of the backpacks under study, $m_{M1} - m_{M2}$, is,

(A) (-6.7 ± 4.39) (B) (-6.7 ± 9.12) (C) (-6.7 ± 7.55) (D) (6.7 ± 7.55) (E) (6.7 ± 4.39)

30. If we wish to estimate the mean of a given normal population by confidence interval, and increase the interval's confidence level, without changing its sample size, the confidence interval will be:

(A) Narrower (B) It will not change (C) All false (D) It cannot be determined (E) Wider

EXERCISES (Time: 70 minutes)

A. (10 points, 25 minutes)

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

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(\theta+1)}\, x^{(\theta-1)/(\theta+1)} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$, has been taken. **Remark**: It is known that $\mathrm{E}(X) = \frac{2\theta}{3\theta+1}$.

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

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

B. (10 points, 25 minutes)

In a given study on immigration in Spain, the following classification has been established using as a basis the immigrant's origin: African countries, Eastern European countries, South American countries, and rest of the world. In this study, the starting hypothesis is that the probabilities that an immigrant comes from Eastern European countries and from the rest of the world are each equal to $\frac{1}{8}$, whereas the probability that s/he comes from African countries and from South American countries are each equal to $\frac{3}{8}$. In order to test this hypothesis a r.s. of size 400 has been taken providing the following observed frequencies:

Immigrant's origin	n_i
Africa	145
Eastern Europe	45
South America	158
Rest of the world	52

At the 5% significance level and using this information, carry out the corresponding test. Write down the null and alternative hypotheses, the test statistic equation, together with its distribution under the null hypothesis, and the decision rule for this test

C. (10 points, 20 minutes)

Let $X_1, \ldots, X_6$ be a r.s. taken from a population having a Poisson distribution with parameter λ. We wish to test the null hypothesis $H_0 : \lambda = 0.50$ against the alternative hypothesis $H_1 : \lambda = 1.50$.

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

ii) At the 10% 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.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

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

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(\theta+1)}\ x^{(\theta-1)/(\theta+1)} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$, has been taken and it is known that $\mathrm{E}(X) = \frac{2\theta}{3\theta+1}$.

i) Method of moments estimator

We need to make the first population and sample moments equal. That is,

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

Therefore, as we know that $\mathrm{E}(X) = \frac{2\theta}{3\theta+1}$, we have that:

$$\alpha_1 = \mathrm{E}(X) = \frac{2\theta}{3\theta+1} = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \Longrightarrow 2\theta = 3\overline{X}\theta + \overline{X}$$

$$\Longrightarrow \theta(2-3\overline{X}) = \overline{X} \Longrightarrow \hat{\theta}_{\mathrm{MM}} = \frac{\overline{X}}{(2-3\overline{X})}$$

b) Maximum likelihood estimator

The likelihood function for the sample is:

$$\begin{aligned} L(\theta) &= f(x_1;\theta)\ \ldots\ f(x_n;\theta) \\ &= \left[\frac{2\theta}{(\theta+1)}\ x_1^{(\theta-1)/(\theta+1)}\right] \ldots \left[\frac{2\theta}{(\theta+1)}\ x_n^{(\theta-1)/(\theta+1)}\right] \\ &= \frac{2^n\theta^n}{(\theta+1)^n}\ \left[\prod_{i=1}^n x_i\right]^{(\theta-1)/(\theta+1)} \end{aligned}$$

We compute its natural logarithm, so that:

$$\ln L(\theta) = n\ln 2 + n\ln\theta - n\ln(\theta+1) + \left(\frac{\theta-1}{\theta+1}\right)\ \ln\left[\prod_{i=1}^n x_i\right]$$

If we now take derivatives with respect to θ and make it equal to zero, we have that:

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{n}{\theta} - \frac{n}{(\theta+1)} + \ln\left[\prod_{i=1}^n x_i\right]\left[\frac{(\theta+1)(1)-(\theta-1)(1)}{(\theta+1)^2}\right] = 0$$

$$\Longrightarrow \frac{n}{\theta} - \frac{n}{(\theta+1)} + \left[\frac{2}{(\theta+1)^2}\right]\ln\left[\prod_{i=1}^n x_i\right] = 0,$$

$$\Longrightarrow \left[\frac{2}{(\theta+1)^2}\right] \ln\left[\prod_{i=1}^{n} x_i\right] = -\frac{n}{\theta} + \frac{n}{(\theta+1)} = \left[\frac{n\theta - n(\theta+1)}{\theta(\theta+1)}\right] = -\frac{n}{\theta(\theta+1)}$$

$$\Longrightarrow \left[\frac{2}{(\theta+1)}\right] \ln\left[\prod_{i=1}^{n} x_i\right] = -\frac{n}{\theta}$$

$$\Longrightarrow 2\theta \ln\left[\prod_{i=1}^{n} x_i\right] = -n - n\theta \Longrightarrow -n = \theta\left\{n + 2\ln\left[\prod_{i=1}^{n} x_i\right]\right\}$$

so that,

$$\hat{\theta}_{\text{ML}} = -\frac{n}{\{n + 2\ln[\prod_{i=1}^{n} X_i]\}} = -\frac{n}{[n + 2\sum_{i=1}^{n}\ln(X_i)]}$$

Exercise B

We have to perform a χ^2 goodness-of-fit test to a completely specified distribution. More specifically, we have to test

H_0 : P(Eastern Europe) = P(Rest of the world) $= 0.125 = \frac{1}{8}$, P(Africa)= P(South America)=0.375=$\frac{3}{8}$

$H_1 : H_0$ is not true

We build the following table

	n_i	p_i	np_i	$\frac{(n_i-np_i)^2}{np_i}$
Africa	145	0.375	150	0.1667
Eastern Europe	45	0.125	50	0.5000
South America	158	0.375	150	0.4267
Rest of the World	52	0.125	50	0.0800
	400	1	400	$z = 1.1734$

We know that, under the null hypothesis, the test statistic $Z = \sum_{i=1}^{K} \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{(K-1)}$, where K is the number of categories in which the variable immigrant's origin has been divided ($K = 4$). At the approximate 5% significance level, the decision rule is to reject the null hypothesis if:

$$z > \chi^2_{(4-1)\,,\,0.05} = \chi^2_{3,0.05}$$

In this case:

$$z = 1.1734 < 7.81 = \chi^2_{3,0.05}$$

so that the null hypothesis is not rejected. Therefore, we conclude that the probabilities for the origin of the immigrants in Spain are those established in the study on immigration.

Exercise C

Let $X_1, \ldots, X_6$ be a r.s. of size $n = 6$ taken from a population having a Poisson distribution with parameter λ. We wish to test the null hypothesis $H_0 : \lambda = 0.50$ against the alternative hypothesis $H_1 : \lambda = 1.50$.

i) We use the Neyman-Pearson' theorem to obtain the form of the most powerful critical region for this test. In this way, the corresponding likelihood functions under the null and alternative hypotheses would be respectively given by:

$$L(\vec{x};\lambda_0) = L(\vec{x};\lambda = 0.50) = \frac{\left[e^{-0.50n}(0.50)^{\sum_{i=1}^{n} x_i}\right]}{\prod_{i=1}^{n} x_i!},$$

and

$$L(\vec{x};\lambda_1) = L(\vec{x};\lambda=1.50) = \frac{\left[e^{-1.5n}(1.50)^{\sum_{i=1}^n x_i}\right]}{\prod_{i=1}^n x_i!}$$

Therefore, if we apply Neyman-Pearson's theorem, we will have that:

$$\frac{L(\vec{x};\lambda_0)}{L(\vec{x};\lambda_1)} = \frac{\left[e^{-0.50n}(0.50)^{\sum_{i=1}^n x_i}\right] / \prod_{i=1}^n x_i!}{\left[e^{-1.50n}(1.50)^{\sum_{i=1}^n x_i}\right] / \prod_{i=1}^n x_i!} \leq K, \; K > 0$$

$$\Longrightarrow e^n \left(\frac{0.50}{1.50}\right)^{\sum_{i=1}^n x_i} \leq K \Longrightarrow \left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i} \leq K_1$$

If we now take natural logarithms, we will have that:

$$\left(\sum_{i=1}^n x_i\right) \ln\left(\frac{1}{3}\right) \leq K_2, \; K_2 > 0$$

Given that $\frac{1}{3} < 1$, the natural logarithm is negative and, thus, the inequality will change, so that the decision rule will be to reject the null hypothesis if $T = \sum_{i=1}^n X_i \geq C$.

ii) At the $\alpha = 0.10$ significance level, and taking into account that $T = \sum_{i=1}^n X_i \in \mathcal{P}(n\lambda)$ and that $n = 6$, we will have that:

$$\alpha = 0.10 \geq P[Z \geq C | H_0] = P[Z \geq C | Z \in \mathcal{P}(3)] = 1 - F_Z(C-1)$$

$$\Longrightarrow F_Z(C-1) \geq 0.90 \Longrightarrow (C-1) = 5 \Longrightarrow C = 6.$$

That is, we reject the null hypothesis if $T = \sum_{i=1}^6 X_i \geq 6$.

iii) To compute the power for this test, we will have that:

$$\text{Power} = P[Z \geq 6 | H_1] = P[Z \geq 6 | Z \in \mathcal{P}(9)] = 1 - F_Z(5) = 1 - 0.115690 = 0.884310.$$