STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2024-25
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. June 3, 2025

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 45 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:

In the "sales" section of a given department store it is known that the number of items that arrives daily from other sections in the store and that show some factory defect follows a binomial, $Z \in b(p,n)$, distribution with characteristic function given by $\Psi_Z(u) = (0.25 + 0.75e^{iu})^n$. We assume independence between the different items.

2. If in a given day 14 items arrive at the store "sales" section, the probability that there are 6 defective items is:

   (A) 0.0082 (B) 0.0022 (C) 0.0103 (D) 0.0280 (E) 0.0018

3. If we consider the same items described in the previous question, the probability that there are at least 12 defective items is:

   (A) 0.1802 (B) 0.2811 (C) 0.1010 (D) 0.0832 (E) 0.8990

4. If in another given day 50 items arrive at the store "sales" section, the probability that there are at most 40 defective items is, approximately:

   (A) 0.1635 (B) 0.3338 (C) 0.8365 (D) 0.6255 (E) 0.3745

Questions 5 and 6 refer to the following exercise:

Let Z be a r.v. having a binomial, $Z \in b(p, 15)$, distribution with mean $\mathrm{E}(Z) = 9$.

5. $P(Z \leq 8)$ is:

   (A) 0.3902 (B) 0.6098 (C) 0.7869 (D) 0.2131 (E) 0.9050

6. $P(5 \leq Z < 10)$ is:

   (A) 0.5875 (B) 0.5949 (C) 0.4125 (D) 0.2038 (E) 0.3809

Questions 7 to 10 refer to the following exercise:

The number of clients arriving, **every ten minutes**, at the parking of a given appliances store follows a Poisson distribution with $P(X = 3) = \frac{4}{3}P(X = 2)$. We assume independence between the arrivals of the different clients at the appliances store parking.

7. The probability that, in a **twenty-minute period**, at least 6 clients arrive at the appliances store parking is:

   (A) 0.2149 (B) 0.6866 (C) 0.7851 (D) 0.8088 (E) 0.1912

8. In a **twenty-minute period**, what should be minimum number of parking places in the appliances store parking so that all arriving clients can park with a probability of at least 0.90?

   (A) 10 (B) 8 (C) 12 (D) 11 (E) 13

9. The approximate probability that, in a **one-hour period**, at most 28 clients arrive at the appliances store parking is:

(A) 0.1788 (B) 0.5714 (C) 0.8212 (D) 0.4286 (E) 0.7190

10. The most likely number of clients expected to arrive at the appliances store parking in a **one-hour and a half period** is:

(A) 24 y 25 (B) 25 (C) 47 y 48 (D) 35 y 36 (E) 36

Questions 11 and 12 refer to the following exercise:

Let X be a r.v. with a $\gamma(a, r)$ distribution, having mean and variance equal to 2 and 8, respectively.

11. The distribution of the r.v. X is:

(A) $\gamma(\frac{1}{4}, \frac{1}{2})$ (B) $\gamma(\frac{1}{4}, 2)$ (C) $\gamma(4, 2)$ (D) $\gamma(4, \frac{1}{2})$ (E) $\gamma(\frac{1}{2}, \frac{1}{2})$

12. If we define the r.v. $Y = \frac{X}{2}$, the value of k such that $P(0.0039 < Y < k) = 0.90$ is:

(A) 3.84 (B) 5.99 (C) 6.63 (D) 4.61 (E) 2.71

Questions 13 to 15 refer to the following exercise:

Let X, Y and Z be three independent r.v. such that: $X \in N(0, \sigma^2 = 16)$, $Y \in \gamma(2, 4)$ and $Z \in \gamma(\frac{1}{2}, 3)$, respectively.

13. If we define the r.v. $V_1 = \dfrac{\sqrt{6}X}{4\sqrt{Z}}$, the value of k such that $P(V_1 < k) = 0.10$ is:

(A) 1.94 (B) 1.44 (C) -1.44 (D) -1.94 (E) -2.45

14. The value of k such that $P(4Y > k) = 0.75$ is:

(A) 5.07 (B) 10.2 (C) 3.45 (D) 7.72 (E) 1.92

15. If we define the r.v. $V_2 = \dfrac{8Z}{3X^2}$, then $P(V_2 < 0.2646)$ is:

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

Questions 16 and 17 refer to the following exercise:

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

$$P(X = 0) = 1 - \frac{3\theta}{2}; \quad P(X = 1) = \frac{3\theta}{4}; \quad P(X = -1) = \frac{3\theta}{4}$$

In order to be able to estimate the parameter θ, a r.s. of size $n = 10$ has been taken, providing 6 zeroes.

16. The method of moments estimate of θ is:

(A) $\frac{4}{5}$ (B) $\frac{2}{15}$ (C) $\frac{2}{5}$ (D) $\frac{4}{15}$ (E) $\frac{6}{15}$

17. The maximum likelihood estimate of θ is:

(A) $\frac{6}{15}$ (B) $\frac{2}{15}$ (C) $\frac{4}{5}$ (D) $\frac{4}{15}$ (E) $\frac{2}{5}$

18. Let X be a r.v. having a normal $N(0,\theta)$ distribution. Based on a r.s. of size n, we consider the following estimator of θ: $\hat{\theta} = k\sum_{i=1}^{n} X_i^2$. What should be the value of k so that $\hat{\theta}$ is an unbiased estimator of θ?

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

Questions 19 and 20 refer to the following exercise:

Let X be a r.v. having a Poisson $\mathcal{P}(\theta)$ distribution. We wish to estimate the parameter θ and, in order to do so, a r.s. of size n, $X_1, \ldots, X_n$, has been taken, where n is an even number, and the following estimator is proposed: $\hat{\theta} = (X_1 + X_3 + X_5 + \ldots + X_{n-3} + X_{n-1})/\left(\frac{n}{2}\right)$.

19. The proposed estimator is:

(A) Biased and asymptotically unbiased (B) Unbiased (C) Biased and asymptotically biased

(D) Unbiased and asymptotically biased (E) It cannot be determined

20. The variance of the proposed estimator is:

(A) $\dfrac{2\theta}{n}$ (B) $\dfrac{\theta}{n}$ (C) $\dfrac{4\theta}{n}$ (D) $\dfrac{\theta}{2n}$ (E) θ

Questions 21 and 22 refer to the following exercise:

The number of clients entering a specific bank branch per hour follows a Poisson distribution. The branch director considers the idea of opening a new bank counter for the public if the average number of clients entering the branch is at least 7. In order to test the hypothesis about the need to open a new bank counter in this branch, that is $H_0 : \lambda \geq 7$, against the alternative hypothesis $H_1 : \lambda < 7$, the director has information about the clients entering the branch for a given hour.

21. At the $\alpha = 0.10$ significance level, the most powerful decision rule will be to reject H_0 if the number of clients entering this specific bank branch per hour is:

(A) $X \leq 3$ (B) $X \geq 4$ (C) $X \geq 3$ (D) $X \leq 4$ (E) $X \leq 7$

22. For this test and for a value of $\lambda = 5$, the probability of type II error is:

(A) 0.1247 (B) 0.2650 (C) 0.1404 (D) 0.7350 (E) 0.8753

Questions 23 to 25 refer to the following exercise:

Let X be a r.v. having probability density function $f_X(x;\theta) = (\theta+1)x^{\theta}, 0 \leq x \leq 1, \theta > 0$. In order to be able to test the null hypothesis $H_0 : \theta = 1$ against the alternative hypothesis $H_1 : \theta = 2$, a r.s. of size $n = 1$ has been taking, thus, observing the value $X = x$.

23. At the α significance level, the form of the most powerful region for this test is:

(A) $[C_1, C_2]$ (B) $[C, 1]$ (C) $[0, C]$ (D) $[C_1, C_2]^c$ (E) All false

24. For $\alpha = 0.05$, the specific form of the critical region for this test is, approximately:

(A) $[0, 0.0253]$ (B) $[0.9747, 1]$ (C) $[0.75, 1]$ (D) $[0.0253, 0.9747]^c$ (E) $[0, 0.25]$

25. The power for this test is, approximately:

(A) 0.93 (B) 0.75 (C) 0.07 (D) 0.95 (E) 0.05

Questions 26 to 28 refer to the following exercise:

The professors of the course Statistics Applied to Business are interested in knowing the proportion of students failing the course, p, and if this proportion decreases for students participating in the on-going evaluation during the quarter. In order to be able to investigate this issue, they have data for one of the previous exam calls. In this specific exam call, the number of failing students was 122 out of a total of 270 students taking the exam.

26. A 95% confidence interval for the proportion of failing students is:

(A) $(0.402, 0.502)$ (B) (0.383,0.521) (C) (0.392, 0.511) (D) (0.374,0.530) (E) (0.450,0.454)

27. At the 5% significance level, the result of the test for $H_0 : p \leq 0.4$ against $H_1 : p > 0.4$ is:

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

28. In addition, it is known that, out of the 270 students, 200 participated in the on-going evaluation and 60 of them failed the course. Moreover, from the 70 students not participating in the on-going evaluation, 62 failed the course. At the 5%, significance level, can we state that the proportion of failing students decreases with the participation of the students in the on-going evaluation?

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

Questions 29 and 30 refer to the folllowing exercise:

It is known that the expense on computer equipment that medium size firms take on is a r.v. X having a normal distribution. In order to be able to estimate the mean expense, a r.s. of size $n = 15$ has been taken, obtaining a sample mean expense of 8000 euros with a sample standard deviation of 1500 euros.

29. A 95% confidence interval for the mean expense is:

(A) $(7495.67, 8504.33)$ (B) $(7214.25, 8785.75)$ (C) $(7142.09, 8857.91)$ (D) $(7770.71, 8229.29)$ (E) $(7294.43, 8705.57)$

30. At the 5% significance level, the result of the test for the null hypothesis $H_0 : m = 7500$ against the alternative hypothesis $H_1 : m \neq 7500$ will be:

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

EXERCISES (Time: 80 minutes)

A. (10 points, 30 minutes)

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

$$f(x;\theta) = \begin{cases} (\theta+2)\ 5^{\theta+2}\ x^{\theta+1} & \text{for } 0 < x < \frac{1}{5}; \\ 0 & \text{otherwise,} \end{cases}$$

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

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

ii) Obtain, **providing all relevant details**, the method of moments estimator of θ. For this purpose, you have to previously show that the mean of the r.v. X is $m = \frac{(\theta+2)}{5(\theta+3)}$.

B. (10 points, 30 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, 20 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:

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

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

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

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

$$f(x;\theta) = \begin{cases} (\theta+2)\ 5^{\theta+2}\ x^{\theta+1} & \text{para } 0 < x < \frac{1}{5}; \\ 0 & \text{otherwise,} \end{cases}$$

i) Maximum likelihood estimator

The likelihood function will be given by

$$L(\theta) = f(x_1;\theta)\cdots f(x_n;\theta) = \left[(\theta+2)\ 5^{\theta+2}\ x_1^{\theta+1}\right]\cdots\left[(\theta+2)\ 5^{\theta+2}\ x_n^{\theta+1}\right]$$

$$L(\theta) = (\theta+2)^n\ 5^{n(\theta+2)}\ \left[\Pi_{i=1}^n\ x_i\right]^{\theta+1}$$

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

$$\ln L(\theta) = n\ln(\theta+2) + [n(\theta+2)]\ln(5) + (\theta+1)\ln\left[\Pi_{i=1}^n\ x_i\right]$$

Taking the derivative with respect to θ, we will have that:

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

Therefore,

$$\frac{n}{(\theta+2)} + n\ln(5) + \ln\left[\Pi_{i=1}^n\ x_i\right] = 0 \Longrightarrow \frac{n}{(\theta+2)} = -\left\{n\ln(5) + \ln\left[\Pi_{i=1}^n\ x_i\right]\right\}$$

$$\frac{-n}{\left\{n\ln(5) + \ln\left[\Pi_{i=1}^n\ x_i\right]\right\}} = (\theta+2) \Longrightarrow \hat{\theta}_{\text{ML}} = \frac{-n}{\left\{n\ln(5) + \ln\left[\Pi_{i=1}^n\ x_i\right]\right\}} - 2$$

ii) Method of moments estimator

First of all, we have to compute the first population moment:

$$\alpha_1 = \mathrm{E}(X) = \int_0^{\frac{1}{5}} x f(x,\theta)dx = \int_0^{\frac{1}{5}} (\theta+2)\ 5^{\theta+2}\ x^{\theta+2}dx = \left[\frac{(\theta+2)}{(\theta+3)}\right]5^{\theta+2}\left[x^{\theta+3}\Big|_0^{\frac{1}{5}}\right. = \frac{(\theta+2)}{5(\theta+3)}$$

We need to make the first population and sample moments equal, so that:

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

Therefore, we have that:

$$\alpha_1 = \mathrm{E}(X) = \frac{(\theta+2)}{5(\theta+3)} = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \Longrightarrow (\theta+2) = 5\overline{X}(\theta+3)$$

$$\Longrightarrow \theta(1-5\overline{X}) = (15\overline{X}-2) \Longrightarrow \hat{\theta}_{\text{MM}} = \frac{(15\overline{X}-2)}{(1-5\overline{X})}$$

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:

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

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

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

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 $\text{CR} = [0, C]$.

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:

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

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

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 corresponds to a **goodness of fit test to a completely specified distribution**. The data is displayed in the following table:

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

Under the null hypothesis that the probability that employees visit the firm's medical consulting room is the same for every working day in the week, we have to test the null hypothesis $H_0 : p_i = 1/5 = 0.20,\ i = 1, 2, 3, 4, 5$, against the alternative hypothesis $H_1 : p_i \neq 1/5 \neq 0.20,\ i = 1, 2, 3, 4, 5$. Given that we have five days of the week or five classes, $k = 5$. Based on the information provided in the sample, we build the corresponding table as follows:

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 probability that employees visit the firm's medical consulting room is the same for every working day in the week, the test statistic $Z = \sum_i \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{k-1}$, where k is the number of days in the week.

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.