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
Second Call. July 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. You will need to obtain at least 18 and 15 points in each part of the exam 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 probability of developing an illness when visiting a given country is 0.6.

2. If a group of 20 people decide to travel to that specific country, what is the probability that more than 50% of them develop the illness?

   (A) 0.7553 (B) 0.1275 (C) 0.9750 (D) 0.8725 (E) 0.2447

3. What is the probability that, in the same group of people, fewer than 80% of them develop the illness?

   (A) 0.1756 (B) 0.0160 (C) 0.9490 (D) 0.0510 (E) 0.8744

4. If the group had 200 people instead, what would be the approximate probability that at most 108 of them develop the illness?

   (A) 0.0485 (B) 0.5948 (C) 0.4052 (D) 0.9515 (E) 0.3420

Questions 5 and 6 refer to the following exercise:

A tax civil servant selects a r.s. of 10 tax returns from people having a given professional career.

5. If 2 or more of their tax returns have some type of issue, all of the tax returns from people having that specific professional career will be reviewed. What is the probability that all of those tax returns are reviewed if 30% of the professionals having that specific career have presented tax returns having some type of issue?

   (A) 0.9752 (B) 0.8507 (C) 0.1211 (D) 0.1493 (E) 0.0248

6. If, at the 5% significance level, the specific tax civil servant wishes to test the null hypothesis $H_0 : p \leq 0.3$ against the alternative hypothesis $H_1 : p > 0.3$, where p is the proportion of tax returns having some type of issue among the people having that specific professional career, using the r.s. of the 10 tax returns, what would be the critical region that should be used for the number of tax returns having some type of issue in the sample?

   (A) $[5,\ 10]$ (B) $[0,\ 5]$ (C) $[7,\ 10]$ (D) $[6,\ 10]$ (E) $[0,\ 6]$

7. In a given firm we consider that workers' punctuality is acceptable if they arrive late to work at most 10% of the working days. In order to be able to penalize the lack of punctuality the firm decides to establish the following procedure: during 20 randomly selected days during the year, they check the arrival time for a given worker. If it turns out that s/he has arrived late for more than 3 days, s/he will be considered as having a clear lack of punctuality and his/her salary will be reduced. What is the maximum probability that a worker having an acceptable punctuality record has his/her salary reduced?

   (A) 0.0190 (B) 0.0430 (C) 0.9568 (D) 0.8670 (E) 0.1330

Questions 8 to 10 refer to the following exercise:

The number of daily labor accidents occurring in a given firm follows a Poisson distribution with parameter $\lambda = 0.5$. We assume independence between the number of accidents occurring in different days.

8. If we assume that a working week for this specific firm has 5 days, what is the probability that, in a **two-week** period, there are fewer than 3 accidents?

   (A) 0.8754 (B) 0.9197 (C) 0.0803 (D) 0.1247 (E) 0.2650

9. The probability that, in the 20 working days for a given month, there are no accidents is, approximately:

   (A) 0 (B) 0.607 (C) 1 (D) 0.393 (E) 0.09

10. Let Y be the random variable representing the number of days, in the 20 working days for a given month, where there is at least one accident. The probability distribution of Y is:

    (A) $N(m = 0.5,\ \sigma^2 = 0.5)$ (B) $\mathcal{P}(\lambda = 10)$ (C) $N(m = 10,\ \sigma^2 = 10)$
    (D) $b(p = 0.393,\ n = 20)$ (E) $b(p = 0.09,\ n = 20)$

11. In a given hospital section, the number of sick patients that requires hospitalization follows a Poisson distribution with parameter $\lambda = 3$. What if the minimum number of beds required so that, with a probability of at least 0.96, no sick patients should be send to another hospital?

    (A) 5 (B) 6 (C) 7 (D) 3 (E) 4

12. The time between two consecutive phone calls to an emergency service is a random variable following an exponential distribution with mean equal to 2 minutes. The probability that the time between two consecutive calls is less than 1 minute is:

    (A) 0.6065 (B) 0.1353 (C) 0.6988 (D) 0.3935 (E) 0.8647

Questions 13 to 15 refer to the following exercise:

Let X_1, X_2 and X_3 be three independent random variables so that: $X_1 \in N(3, \sigma^2 = 4)$, $X_2 \in N(2, \sigma^2 = 1)$ and $X_3 \in \gamma(1, 3)$.

13. If we define the r.v. $Y = 2X_3$, the probability distribution for Y is:

    (A) $\gamma\left(\frac{1}{2}, 3\right)$ (B) $\gamma(2, 3)$ (C) $\gamma\left(\frac{1}{2}, 6\right)$ (D) $\gamma(2, 6)$ (E) $\gamma\left(\frac{1}{2}, \frac{3}{2}\right)$

14. If we define the r.v. $Z = \dfrac{\sqrt{3}(X_2 - 2)}{\sqrt{X_3}}$, then $P\left[Z \notin (-2.45,\ 0.265)\right]$ is:

    (A) 0.850 (B) 0.425 (C) 0.225 (D) 0.150 (E) 0.575

15. The probability $P\left[\left(\frac{X_1 - 3}{2}\right)^2 + 2X_3 < 12\right]$ is:

    (A) 0.10 (B) 0.50 (C) 0.975 (D) 0.90 (E) 0.025

16. If we let $\mathcal{F}_{\overline{n_1, n_2}|\alpha} = z$, we then have that:

    (A) $\mathcal{F}_{\overline{n_2, n_1}|\alpha} = 1/z$ (B) $\mathcal{F}_{\overline{n_2, n_1}|\alpha} = z$ (C) $\mathcal{F}_{\overline{n_2, n_1}|1-\alpha} = z$
    (D) $\mathcal{F}_{\overline{n_1, n_2}|\alpha} < \mathcal{F}_{\overline{n_1, n_2}|\alpha/2}$ (E) $\mathcal{F}_{\overline{n_1, n_2}|1-\alpha} = z$

17. Let X be a r.v. following a Student's t distribution with 10 degrees of freedom. The value of k such that $P(X > k) = 0.8$ is:

(A) -0.879 (B) 1.37 (C) 0.879 (D) -1.37 (E) 0.26

Questions 18 and 19 refer to the following exercise:

Let X be a r.v. having a $N(0,1)$ distribution and let $Y = 2X^2 - 1$

18. The variance of Y is:

(A) 1 (B) 2 (C) 3 (D) 8 (E) 7

19. $P(Y > 1.64)$ is:

(A) 0.05 (B) 0.25 (C) 0.95 (D) 0.75 (E) 0.50

Questions 20 to 22 refer to the following exercise:

Let X be a r.v. such that:

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

In order to be able to estimate p, a r.s. of size $n = 10$ has been taken, resulting in 3 zeroes.

20. The method of moments estimate of p is:

(A) 0.35 (B) 0.15 (C) 0.50 (D) 0.70 (E) 0.30

21. The maximum likelihood estimate of p is:

(A) 0.30 (B) 0.15 (C) 0.35 (D) 0.70 (E) 0.50

22. The maximum likelihood estimate of the variance of the r.v. X is:

(A) 0.85 (B) 0.30 (C) 0.50 (D) 0.65 (E) 0.70

Questions 23 to 25 refer to the following exercise:

Let $X \in N(m, \sigma^2)$, with m and σ^2 both unknown. In order to be able to estimate σ^2, from a r.s. of size n, we consider the following estimator: $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$. In addition, it is known that the mean and variance of a r.v. having a χ^2_k distribution are k and $2k$, respectively, and that $\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$.

23. The expected value or mean of S^2 is:

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

24. The variance of S^2 is:

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

25. With regard to the estimator of S^2, we can state that:
   (A) It is unbiased and not consistent
   (B) It is biased and consistent
   (C) It is unbiased and consistent
   (D) It is biased and not consistent
   (E) -

Questions 26 and 27 refer to the following exercise:

Let X be a r.v. having a uniform $U[\theta, 5]$ distribution. In order to be able to test $H_0 : \theta = 0$ against $H_1 : \theta = 1$, a r.s. of size $n = 1$ is taken and we decide to reject H_0 of the sample observation, X, is larger than or equal to 2.

26. The significance level for this test is:
   (A) 0.20 (B) 0.50 (C) 0.05 (D) 0.60 (E) 0.40

27. The power for this test is:
   (A) 0.60 (B) 0.75 (C) 0.25 (D) 0.40 (E) 0.50

Questions 28 and 29 refer to the following exercise:

We wish to estimate the family Christmas mean lottery expense (in euros) for a given province. In order to be able to do so, a r.s. of 121 families living in that province is taken, providing the following information about their Christmas lottery expense: $\overline{x} = 188$ euros and $s = 100$ euros. We assume that the family Christmas lottery expense follows a normal distribution.

28. The 95% confidence interval for the family Christmas mean lottery expense is, approximately:
   (A) (188 ± 9.86) (B) (188 ± 11.66) (C) (188 ± 18.07)
   (D) (188 ± 21.54) (E) (188 ± 15.15)

29. We wish to test the null hypothesis that the mean expense has been equal to 200 euros against the alternative hypothesis that it has been different from that amount. At the 5% significance level, the decision will be:
   (A) - (B) - (C) Reject the null hypothesis
   (D) Do not reject the null hypothesis (E) It cannot be determined

30. We have a population where we have to classify each individual in one of four different classes. We wish to test the hypothesis that the probabilities of belonging to each o the classes are p_1, p_2, p_3 and p_4 (**all known**), respectively. A r.s. of size 200 has been taken. The most appropriate test will be:
   (A) Test of homogeneity
   (B) χ^2 goodness-of-fit test to a completely specified distribution
   (C) χ^2 goodness-of-fit test to a partially specified distribution
   (D) Test of independence
   (E) All false

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ $(n > 3)$ be a r.s. from a population having a Poisson distribution with parameter λ. We consider the following estimators for the parameter λ.

$$\widehat{\lambda}_1 = \frac{X_1 + X_2 + \ldots + X_n}{n} = \overline{X} \qquad\qquad \widehat{\lambda}_2 = \frac{3X_1 + X_2 + \ldots + X_n}{n+2}$$

i) Obtain, **providing all relevant details**, the bias for these estimators.

ii) Obtain, **providing all relevant details**, the variance for these estimators.

iii) Is any of the estimators consistent?

iv) In any of the estimators efficient?

v) Which one of these estimators has a smaller variance?

Remark: The Cramer-Rao lower bound for a regular and unbiased estimator of λ obtained from a r.s. of size n is:

$$L_c = \frac{1}{n\mathrm{E}\left[\frac{\partial \ln \mathrm{P}(\mathrm{X},\lambda)}{\partial \lambda}\right]^2} = \frac{1}{-n\mathrm{E}\left[\frac{\partial^2 \ln \mathrm{P}(\mathrm{X},\lambda)}{\partial \lambda^2}\right]}$$

B. (10 points, 25 minutes)

The following table includes the probability mass function of the discrete r.v. X under the null $(P_0(x))$ and alternative $(P_1(x))$ hypotheses.

X	1	2	3	4	5	6
$P_0(x)$	0	0.10	0.10	0.30	0.40	0.10
$P_1(x)$	0.30	0	0.25	0.05	0.10	0.30

We have a random sample of size $n = 1$ to be able 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 point $X = \{2\}$ in the critical region? You need to appropriately justify your response.

ii) Would you include the point $X = \{1\}$ in the critical region? You need to appropriately justify your response.

iii) At the 10% significance level, and providing all relevant details to justify your response, obtain the most powerful critical region for this test. In addition, compute its probability of type II error.

Remark: Before providing any response to this item, you should take into account your responses to the previous items.

C. (10 points, 25 minutes)

A given investor will only invest in a specific firm if the firm mean benefit is at least 3 and if its dispersion, measured with its variance, is at most 1. In order to be able to facilitate his/her decision, s/he looks at the firm's financial history, obtaining a r.s. of size 26, which results in $\overline{x} = 2.9$ and $s^2 = 1.2$. It is known that firm benefits follow a normal distribution.

i) At the 5% level, test the hypothesis that the firm mean benefit is at least 3.

ii) At the 5% level, test the hypothesis that the dispersion is at most 1.

iii) Based on the results obtained in the previous items, would you recommend him/her to invest in this specific firm?

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

The two estimators being considered here are:

$$\widehat{\lambda}_1 = \frac{X_1 + X_2 + \ldots + X_n}{n} = \overline{X} \qquad \widehat{\lambda}_2 = \frac{3X_1 + X_2 + \ldots + X_n}{n+2}$$

In addition, as X is a r.v. having a Poisson distribution, we have that:

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

$$\mathrm{E(X) = Var(X)} = \lambda$$

i) **Biases**

The bias of a given estimator $\hat{\theta}$ for the parameter θ is defined as $\mathrm{b}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$.

In order to be able to compute the bias for the proposed estimators, we have to obtain their expected values.

$$\mathrm{E}(\hat{\lambda}_1) = \mathrm{E}(\overline{\mathrm{X}}) = \lambda$$

$$\mathrm{E}(\hat{\lambda}_2) = \mathrm{E}\left(\frac{3X_1 + X_2 + \cdots + X_n}{n+2}\right) = \left(\frac{3\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n)}{n+2}\right) = \frac{3\lambda + (n-1)\lambda}{n+2} = \frac{\lambda(n+2)}{n+2} = \lambda$$

Therefore, both estimators are unbiased and, thus, their corresponding biases are equal to 0:

$$\mathrm{b}(\hat{\lambda}_1) = \mathrm{E}(\hat{\lambda}_1) - \lambda = 0$$

$$\mathrm{b}(\hat{\lambda}_2) = \mathrm{E}(\hat{\lambda}_2) - \lambda = 0$$

ii) **Variances**

We compute the variances for the proposed estimators.

$$\mathrm{Var}(\hat{\lambda}_1) = \mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\lambda}{n}$$

$$\mathrm{Var}(\hat{\lambda}_2) = \mathrm{Var}\left(\frac{3X_1 + X_2 + \cdots + X_n}{n+2}\right) = \frac{9\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)}{(n+2)^2} = \lambda\left[\frac{(n+8)}{(n+2)^2}\right]$$

iii) **Consistency**

The estimator $\hat{\lambda}_1$ is consistent because the required sufficient conditions hold:

a. It is unbiased.

b. $\lim\limits_{n\to\infty} \mathrm{Var}(\hat{\lambda}_1) = \lim\limits_{n\to\infty} \frac{\lambda}{n} = 0$

The estimator $\hat{\lambda}_2$ is consistent because the required sufficient conditions hold:

a. It is unbiased.

b. $\lim_{n\to\infty} \text{Var}(\hat{\lambda}_2) = \lim_{n\to\infty} \left[\lambda \frac{(n+8)}{(n+2)^2}\right] = 0$

Therefore, we can state that both estimators are consistent for λ.

iv) **Efficiency.** In order to be able to find out if the estimator is efficient, we compute the Cramer-Rao lower bound.

$$Lc = \frac{1}{n\text{E}\left[\dfrac{\partial \ln \text{P}(\text{X}, \lambda)}{\partial \lambda}\right]^2}$$

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\ln P(x,\lambda) = -\lambda + x\ln(\lambda) - \ln x!$$

$$\frac{\partial \ln P(x,\lambda)}{\partial \lambda} = -1 + \frac{x}{\lambda} = \frac{1}{\lambda}(x-\lambda)$$

$$\text{E}\left[\frac{\partial \ln \text{P}(\text{X},\lambda)}{\partial \lambda}\right]^2 = \text{E}\left[\frac{1}{\lambda}(\text{X}-\lambda)\right]^2 = \frac{1}{\lambda^2}\ \text{E}(\text{X}-\lambda)^2 = \frac{1}{\lambda^2}\ \text{Var}(X) = \frac{1}{\lambda^2}\ (\lambda) = \frac{1}{\lambda}$$

Therefore, we have that:

$$Lc = \frac{1}{n\left(\frac{1}{\lambda}\right)} = \frac{\lambda}{n}$$

As the variance of the estimator $\hat{\lambda}_1$ coincides with the Cramer-Rao lower bound, this is an efficient estimator.

As the variance of the estimator $\hat{\lambda}_2$ does not coincide with the Cramer-Rao lower bound, this is not an efficient estimator.

v) **Which one of these estimators has a smaller variance?**

Given that $\hat{\lambda}_1$ in an efficient estimator because its variance coincides with the Cramer-Rao lower bound, we cannot have any other estimator (regular and unbiased) having a smaller variance than that of $\hat{\lambda}_1$.

That is, $\text{Var}(\hat{\lambda}_1) < \text{Var}(\hat{\lambda}_2)$

Exercise B

We wish to test the null hypothesis that the discrete r.v. X has a probability mass function given by $P_0(x)$ against the alternative hypothesis that its probability mass function is $P_1(x)$:

X	1	2	3	4	5	6
$P_0(x)$	0	0.10	0.10	0.30	0.40	0.10
$P_1(x)$	0.30	0	0.25	0.05	0.10	0.30

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

i) Would you include the point $X = \{2\}$ in the critical region?

Given that, under the probability distribution for the alternative hypothesis $P_1(x)$, this point has probability zero, the r.v. cannot take on this value under the alternative hypothesis, but it can under the null hypothesis. Therefore, the point $X = \{2\}$ is a point of no rejection for H_0 and, thus, it should **never** be included in the critical region.

ii) Would you include the point $X = \{1\}$ in the critical region?

Given that, under the probability distribution for the null hypothesis $P_0(x)$, this point has probability zero, the r.v. cannot take on this value under the null hypothesis. Therefore, the point $X = \{1\}$ is a point of rejection for H_0 and, thus, it should **always** be included in the critical region.

iii) At the $\alpha = 0.10$ significance level, and **taking into account the responsed in the previous items**, we have that the possible critical regions for this test are $\text{CR}_1 = \{1, 3\}$ and $\text{CR}_2 = \{1, 6\}$. Moreover, this is because the type I probabilities for these critical regions are:

$$P(X \in \text{CR}_1 | P_0) = P(X = 1, 3 | P_0) = 0 + 0.10 = 0.10 \leq \alpha$$
$$P(X \in \text{CR}_2 | P_0) = P(X = 1, 6 | P_0) = 0 + 0.10 = 0.10 \leq \alpha$$

We now obtain the corresponding powers for these critical regions:

$$\text{Power} = P(X \in \text{CR}_1 | P_1) = P(X = 1, 3 | P_1) = 0.30 + 0.25 = 0.55$$
$$\text{Power} = P(X \in \text{CR}_2 | P_1) = P(X = 1, 6 | P_1) = 0.30 + 0.30 = 0.60$$

Therefore, the most powerful critical region for this test is CR_2.

Finally, the probability of type II error for CR_2 is:

$$\beta = P(\text{Type II error}) = 1 - \text{Power} = 1 - 0.60 = 0.40$$

Exercise C

i) **Test if mean benefit is at least 3**

We wish to test:

$H_0 : m \geq 3 = m_0$

$H_1 : m < 3$

Given that we have a normal distribution with unknown variance, at the α significance level, we reject the null hypothesis if:

$$\frac{\overline{x} - m_0}{\frac{s}{\sqrt{n-1}}} \leq -t_{\overline{n-1}|\alpha}$$

In our specific case, we have that:

$$\frac{\overline{x} - m_0}{\frac{s}{\sqrt{n-1}}} = \frac{2.9 - 3}{\sqrt{\frac{1.2}{25}}} = -0.4564 > -1.71 = t_{\overline{25}|0.05}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis.

ii) **Test if the dispersion of the benefits is at most 1**

$H_0 : \sigma^2 \leq 1 = \sigma_0^2$

$H_1 : \sigma^2 > 1$

At the α significance level, we reject the null hypothesis if:

$$\frac{ns^2}{\sigma_0^2} > \chi^2_{\overline{n-1}|\alpha}$$

In our specific case, we have that:

$$\frac{(26) \cdot (1.2)}{1} = 31.2 < 37.7 = \chi^2_{\overline{25}|0.05}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis.

iii) **Invest or not?**

Yes, at the 5% significance level, we would recommend the investor to invest in this firm.