STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2020-21
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. June 2, 2021

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 50 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 normal distribution $N(40, \sigma^2 = 25)$. It is established that the expense for a given family may be considered excessive if it is larger than 44.2 euros. We assume independence in the expenses for the different families in that city. **Remark**: When computing the probability that a family has an excessive daily mean expense you should round this number up to two decimal places.

2. If a random sample of 15 families in that city is taken, the probability that at least 5 of them have an excessive daily expense is:

   (A) 0.0177 (B) 0.0611 (C) 0.1642 (D) 0.8358 (E) 0.9389

3. In the same sample of 15 families, the probability that exactly 13 families **do not** have an excessive daily expense is:

   (A) 0.3980 (B) 0.3431 (C) 0 (D) 1 (E) 0.2309

4. If we now take a random sample of 200 families in that city, the approximate probability that at most 50 of them have an excessive daily expense is:

   (A) 0.0314 (B) 0.3745 (C) 0.8413 (D) 0.9686 (E) 0.6255

Questions 5 and 6 refer to the following exercise:

Let Z be a r.v. following a binomial $b(0.35, n)$ distribution with variance 2.275.

5. $P(Z = 3)$ is:

   (A) 0.1757 (B) 0.2377 (C) 0.2522 (D) 0.2616 (E) 0.5138

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

   (A) 0.07249 (B) 0.9865 (C) 0.0752 (D) 0.0860 (E) 0.0135

Questions 7 to 10 refer to the following exercise:

Let X be a r.v. with characteristic function given by $\Psi_X(u) = \mathrm{e}^{4(\mathrm{e}^{\mathrm{i}u}-1)}$.

7. $P(X > 3)$ is:

   (A) 0.5665 (B) 0.6288 (C) 0.4335 (D) 0.3712 (E) 0.1954

8. $P(2 \leq X < 7)$ is:

   (A) 0.8473 (B) 0.8577 (C) 0.7977 (D) 0.5423 (E) 0.6512

9. $P(4 < X < 9)$ is:

   (A) 0.1935 (B) 0.0893 (C) 0.3498 (D) 0.3630 (E) 0.2067

10. Let $X_1, \ldots, X_{50}$ be a r.s. from that variable. If we define the r.v. $Y = \sum_{i=1}^{50} X_i$, then $P(Y > 180)$ is, approximately:

(A) 0.0838 (B) 0.4602 (C) 0.9162 (D) 0.5000 (E) 0.5398

11. Let X be a normal r.v. with mean zero and variance 2. The value of $P(X^2 < 5.42)$ is:

(A) 0.975 (B) 0.025 (C) 0.95 (D) 0.10 (E) 0.90

12. Let X be a random variable having a Student's t distribution with n degrees of freedom, $t_{\overline{n}|}$. We then have that $P(t_{\overline{n}|1-\frac{\alpha}{2}} < X < t_{\overline{n}|\frac{3\alpha}{2}})$ is:

(A) 2α (B) $1 - 3\alpha$ (C) α (D) $1 - \alpha$ (E) $1 - 2\alpha$

Questions 13 to 15 refer to the following exercise:

Let X, Y, Z and V be four independent r.v. such that their distributions are as follows: $X \in N(0, \sigma^2 = 5)$, $Y \in N(0, \sigma^2 = 4)$, $Z \in \exp(\frac{1}{2})$ and $V \in \gamma(\frac{1}{2}, \frac{7}{2})$.

13. If we define the r.v. $W_1 = \frac{(X-Y)^2}{9}$, then $P(W_1 > 0.102)$ is:

(A) 0.25 (B) 0.50 (C) 0.75 (D) 0.90 (E) 0.85

14. If we define the r.v. $W_2 = \dfrac{\sqrt{2}Y}{2\sqrt{Z}}$, the value of k such that $P(W_2 < k) = 0.01$ is:

(A) 6.96 (B) -6.96 (C) -4.30 (D) -9.92 (E) 9.92

15. If we define the r.v. $W_3 = \frac{7Y^2}{4V}$, the value of k such that $P(W_3 > k) = 0.90$ is:

(A) 0.017 (B) 58.9 (C) 3.59 (D) 5.59 (E) 237

Questions 16 and 17 refer to the following exercise:

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

$$P(X = 0) = \frac{1}{\theta}; \quad P(X = 1) = \frac{5}{2\theta}; \quad P(X = 2) = \frac{(\theta - 4)}{\theta}; \quad P(X = 3) = \frac{1}{2\theta}$$

In order to estimate the parameter θ a r.s. of size $n = 20$ has been taken, providing the following sample values:

X	0	1	2	3
frequency	8	5	3	4

16. The maximum likelihood estimate of θ is:

(A) $\frac{80}{17}$ (B) $\frac{17}{80}$ (C) $\frac{60}{17}$ (D) $\frac{17}{40}$ (E) $\frac{40}{17}$

17. The method of moments estimate of θ is:

(A) $\frac{40}{17}$ (B) $\frac{17}{80}$ (C) $\frac{80}{17}$ (D) $\frac{17}{40}$ (E) $\frac{60}{17}$

Questions 18 and 19 refer to the following exercise:

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

$$f(x) = \begin{cases} \dfrac{x}{\theta}\; e^{-\frac{x^2}{2\theta}}, & x > 0, \quad \theta > 0 \\ 0, & \text{otherwise} \end{cases},$$

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

18. The maximum likelihood estimator of θ is:

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

19. The method of moments estimator of θ is:

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

Questions 20 to 22 refer to the following exercise:

Let X be a r.v. having a normal $N(\theta, \sigma^2 = 4)$ distribution. In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$, has been taken, and $\hat{\theta} = (3X_1 + 4X_2 + \ldots + 4X_{n-1} + 2X_n)/(4n+5)$ is proposed as an estimator of θ.

20. The bias of the proposed estimator is:

(A) 0 (B) $-\dfrac{5\theta}{4n}$ (C) $-\dfrac{2}{(4n+5)}$ (D) $\dfrac{3\theta}{(4n+5)}$ (E) $-\dfrac{8\theta}{(4n+5)}$

21. The proposed estimator is:

(A) Biased and asymptotically unbiased (B) – (C) Unbiased
(D) Biased and asymptotically biased (E) Unbiased and asymptotically biased

22. The variance of the proposed estimator is:

(A) $\dfrac{(64n-76)}{(4n+5)^2}$ (B) $\dfrac{64n}{16n^2}$ (C) $\dfrac{64}{16n^2}$ (D) $\dfrac{(64n-12)}{(4n+5)^2}$ (E) $\dfrac{(4n-12)}{(4n+5)^2}$

Questions 23 and 24 refer to the following exercise:

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

$$f(x, \theta) = 3^\theta\; x^{\theta-1} e^{-3x}, \qquad x > 0, \qquad \theta > 0$$

In order 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 taken.

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

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

24. At the $\alpha = 0.05$ significance level, we reject H_0 if:

(A) $X \leq 0.9986$ (B) $X \in (0.0171, 0.998)$ (C) $X \geq 0.0171$ (D) $X \leq 0.0171$ (E) $X \geq 0.9986$

Questions 25 and 26 refer to the following exercise:

Let X be a random variable having a Poisson distribution with parameter λ. In order to test the null hypothesis $H_0 : \lambda \leq 2$ against the alternative hypothesis $H_1 : \lambda > 2$, a r.s. of size $n = 3$ has been taken, and $T = \sum_{i=1}^{3} X_i$ is used as test statistic.

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

(A) $T \leq 10$ (B) $T \geq 11$ (C) $T \geq 10$ (D) $T \leq 11$ (E) $T \geq 9$

26. For $\lambda = 3$, the power for this test is:

(A) 0.8030 (B) 0.1970 (C) 0.7060 (D) 0.2940 (E) 0.4126

Questions 27 and 28 refer to the following exercise:

The professors for the course Statistics Applied to Business are interested in knowing the percentage p of students failing the course, and if this percentage decreases with the on-going evaluation students participate in during the semester. In order to further investigate this matter, they have the corresponding information for the June 2019 exam call. In that specific call, the number of students failing the course was 110 out of a total of 371 students taking the exam.

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

(A) $(0.264, 0.329)$ (B) $(0.243, 0.342)$ (C) $(0.258, 0.335)$ (D) $(0.295, 0.298)$ (E) $(0.250, 0.343)$

28. At the 5% significance level, the result for the test $H_0 : p \leq 0.40$ against $H_1 : p > 0.40$ is:

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

Questions 29 and 30 refer to the following exercise:

A given firm buys ten fluorescent tubes with type A material and ten with type B material. We assume that both tubes' durations follow a normal distribution, and that tubes of type A and type B are independent from each other. From these samples, we obtained the corresponding sample mean and standard deviations for the durations, given by: $\overline{x}_A = 1362.5$, $\overline{x}_B = 1225.8$, $s_A = 202.46$ y $s_B = 256.49$.

29. A 0.90 confidence interval for the ratio of variances, σ_A^2/σ_B^2 is,

(A) $(0.6577, 3.9160)$ (B) $(0.3000, 8.5862)$ (C) $(0.1959, 1.9814)$ (D) $(0.1000, 0.9000)$ (E) $(0.5088, 4.3355)$

30. A 0.90 confidence interval for the difference of the mean durations for the two types of tubes, $m_A - m_B$ is,

(A) (136.7 ± 188.44) (B) (136.7 ± 137.43) (C) (136.7 ± 297.59) (D) (136.7 ± 169.46) (E) (136.7 ± 176.54)

EXERCISES (Time: 70 minutes)

A. (10 points, 25 minutes)

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

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(1-\theta)}\, x^{(3\theta-1)/(1-\theta)} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

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 $\mathrm{E}(X) = \frac{2\theta}{\theta+1}$.

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 School of Economic and Business Sciences researchers wish to test if students' satisfaction with their courses is independent or not from the degree selection they made by their specific vocation. In order to do so, a random sample was taken among all of the students regularly attending class, providing the following results:

	Yes vocation	Indifferent	No vocation	Total
Yes satisfaction	223	60	43	326
Indifferent	127	94	45	266
No satisfaction	40	23	36	99
Total	390	177	124	691

At the 5% significance level and using this information, carry out the corresponding test.

C. (10 points, 20 minutes) For an exponential distribution with parameter λ, we wish to test the null hypothesis $H_0 : \lambda = \frac{1}{2}$ against the alternative hypothesis $H_1 : \lambda = 2$. In order to do so, a r.s. of size $n = 10$, $X_1, \ldots, X_{10}$, has been taken. At the 5% significance level, obtain, **providing all relevant details**, the most powerful critical region if the test statistic used is $Z = \sum_{i=1}^{10} X_i$. **Remark**: If $Y \in \gamma(1/2, n/2) \Longrightarrow Y \in \chi^2_n$.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

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

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(1-\theta)} \; x^{(3\theta-1)/(1-\theta)} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

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

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

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

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

b) Maximum likelihood estimator

The likelihood function for the sample is:

$$\begin{aligned} L(\theta) &= f(x_1;\theta) \;\ldots\; f(x_n;\theta) \\ &= \left[\frac{2\theta}{(1-\theta)} \; x_1^{(3\theta-1)/(1-\theta)}\right] \ldots \left[\frac{2\theta}{(1-\theta)} \; x_n^{(3\theta-1)/(1-\theta)}\right] \\ &= \frac{2^n\theta^n}{(1-\theta)^n} \; \left[\prod_{i=1}^{n} x_i\right]^{(3\theta-1)/(1-\theta)} \end{aligned}$$

We compute its natural logarithm, so that:

$$\ln L(\theta) = n\ln 2 + n\ln\theta - n\ln(1-\theta) + \left(\frac{3\theta-1}{1-\theta}\right) \; \ln\left[\prod_{i=1}^{n} x_i\right]$$

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

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

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

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

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

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

so that,

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

Exercise B

We have to perform a **test of independence**. The corresponding data is included in the following table:

	Yes vocation	Indifferent	No vocation	Total
Yes satisfaction	223	60	43	326
Indifferent	127	94	45	266
No satisfaction	40	23	36	99
Total	390	177	124	691

The probabilities $\hat{p}_{i\cdot}$ and $\hat{p}_{\cdot j}$ are estimated from the data. In this way, we have that:

$$\hat{p}_{\mathrm{YS},\bullet} = \frac{326}{691} \qquad \hat{p}_{\mathrm{IS},\bullet} = \frac{266}{691} \qquad \hat{p}_{\mathrm{NS},\bullet} = \frac{99}{691}$$

$$\hat{p}_{\bullet,\mathrm{YV}} = \frac{390}{691} \qquad \hat{p}_{\bullet,\mathrm{IV}} = \frac{177}{691} \qquad \hat{p}_{\bullet,\mathrm{NV}} = \frac{124}{691}$$

We can then build the corresponding table for this test, so that:

	n_{ij}	$\hat{p}_{i,j} = \hat{p}_{i,\bullet} \cdot \hat{p}_{\bullet,j}$	$n\hat{p}_{i,j}$	$\frac{(n_{ij}-n\hat{p}_{i,j})^2}{n\hat{p}_{i,j}}$
YS, YV	223	$326 \cdot 390/691^2$	183.994	8.269
YS, IV	60	$326 \cdot 177/691^2 =$	83.505	6.616
YS, NV	43	$326 \cdot 124/691^2$	58.500	4.107
IS, YV	127	$266 \cdot 390/691^2$	150.130	3.564
IS, IV	94	$266 \cdot 177/691^2$	68.136	9.818
IS, NV	45	$266 \cdot 124/691^2$	47.734	0.156
NS, YV	40	$99 \cdot 390/691^2$	55.875	4.511
NS, IV	23	$99 \cdot 177/691^2$	25.359	0.219
NS, NV	36	$99 \cdot 124/691^2$	17.765	18.716
	691	1	691	$z = 55.976$

Under the null hypothesis of independence, the test statistic $\sum_{i,j} \frac{(n_{ij}-n\hat{p}_{i,j})^2}{n\hat{p}_{i,j}} \sim \chi^2_{(I-1)(J-1)}$, where I is the number of the categories in which the variable satisfaction has been divided ($I = 3$) and J is the number of categories in which the variable vocation has been divided ($J = 3$). At the approximate 5% significance level, the decision rule is to reject the null hypothesis if:

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

In this case:

$$z = 55.9766 > 9.49 = \chi^2_{4,0.05},$$

so that, at the 5% significance level, the null hypothesis of independence is rejected. That is, we can state that students' satisfaction with their courses is statistically independent from the degree selection they made by their specific vocation.

Exercise C

We have an exponential distibution with parameter λ (i.e., $X \in \exp(\lambda)$), and we wish to test the null hypothesis $H_0 : \lambda = \lambda_0 = \frac{1}{2}$ against the alternative hypothesis $H_1 : \lambda = \lambda_1 = 2$. As we have a r.s. of size $n = 10$, the Neyman-Pearson theorem states that, for a given constant $k > 0$, the most powerful critical region for this test will be given by the one verifying the following inequality for the likelihood ratio functions $L(\cdot)$ under the null and alternative and hypotheses:

$$\frac{L(\lambda_0)}{L(\lambda_1)} \leq k$$

In this case, we will have that:

$$\frac{\lambda_0 e^{-\lambda_0 x_1} \cdots \lambda_0 e^{-\lambda_0 x_n}}{\lambda_1 e^{-\lambda_1 x_1} \cdots \lambda_1 e^{-\lambda_1 x_n}} \leq k \Longrightarrow \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{(\lambda_1-\lambda_0)\sum_{i=1}^n x_i} \leq k$$

Given that $\lambda_1, \lambda_0 > 0$, we will have that:

$$\Longrightarrow e^{(\lambda_1-\lambda_0)\sum_{i=1}^n x_i} \leq k_1 \quad (k_1 > 0)$$

If we take natural logarithms and keep in mind that $\lambda_1 > \lambda_0$, we will have that:

$$(\lambda_1 - \lambda_0)\sum_{i=1}^n x_i \leq k_2 \quad (k_2 > 0) \Longrightarrow \sum_{i=1}^n x_i \leq C \quad (C > 0)$$

Therefore, we will reject the null hypothesis if $\sum_{i=1}^n X_i \leq C$.

Under the null hypothesis, we have that $\lambda = \frac{1}{2}$, so that $X_i \in \exp(\frac{1}{2}) \equiv \gamma(\frac{1}{2}, 1)$. In this way, we will have that $\sum_{i=1}^n X_i \in \gamma(\frac{1}{2}, n) \equiv \chi^2_{2n}$. Therefore, for the specific case where $n = 10$, we will have that $\sum_{i=1}^{10} X_i \in \gamma(\frac{1}{2}, 10) \equiv \chi^2_{20}$. Thu, at the 5% significance level, we will have that:

$$\alpha = 0.05 = P\left(\sum_{i=1}^{10} X_i \leq C \,\middle|\, \lambda = \frac{1}{2}\right) \Longrightarrow 0.05 = P\left(\chi^2_{20} \leq C\right) \Longrightarrow C = \chi^2_{20,0.95} = 10.9,$$

so that, the most powerful critical region for this test will be $\mathrm{CR} = (0, 10.9]$.