STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2018-19
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
First Call. May 22, 2019

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, 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 12 points in each part of the exam to pass it. Otherwise, 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:

The probability that a client enters and purchases something at a given computer store is equal to 20%. It is known that 14 clients enter the computer store each day. We assume independence between the different clients' purchases.

2. The probability that, in a given day, 8 clients purchase something at the computer store is:

   (A) 0.9680 (B) 0.0092 (C) 0.9980 (D) 0.0020 (E) 0.0322

3. The probability that, in a given day, 6 clients **do not purchase** anything at the computer store is:

   (A) 0.044 (B) 0.0020 (C) 0.0322 (D) 0.9884 (E) 0.9996

4. The approximate probability that, in a 10-day period, at most 35 clients purchase something at the computer store is:

   (A) 0.9162 (B) 0.0571 (C) 0.9429 (D) 0.9505 (E) 0.0838

Questions 5 and 6 refer to the following exercise:

Let Z be a r.v. following a binomial $b(0.70, n)$ distribution with variance $\text{Var}(Z) = 2.1$.

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

   (A) 0.2668 (B) 0.1030 (C) 0.6496 (D) 0.8497 (E) 0.2001

6. $P(1 < Z \leq 6)$ is:

   (A) 0.8497 (B) 0.8401 (C) 0.1502 (D) 0.3503 (E) 0.9999

Questions 7 to 10 refer to the following exercise:

The number of clients arriving at a shoe repair store follows a Poisson distribution with mean equal to 2.5 clients per hour. We assume independence between the clients arriving at different hours.

7. The probability that, in a given hour, 6 clients arrive at the shoe repair store is:

   (A) 0.6680 (B) 0.1670 (C) 0.0101 (D) 0.0557 (E) 0.0278

8. The probability that, in a given hour, at most 1 client arrives at the shoe repair store is:

   (A) 0.2873 (B) 0.5438 (C) 0.3694 (D) 0.2565 (E) 0.0821

9. The probability that, in a two-hour period, at least 5 clients arrive at the shoe repair store is:

   (A) 0.4405 (B) 0.5595 (C) 0.6160 (D) 0.3840 (E) 0.8666

10. The approximate probability that, in a 10-hour period, at most 30 clients arrive at the shoe repair store is:

(A) 0.1357 (B) 0.1841 (C) 0.8643 (D) 0.8849 (E) 0.8159

Questions 11 and 12 refer to the following exercise:

Let X and Y be independent r.v. such that $X \in \gamma(\frac{1}{2}, \frac{9}{2})$ and $Y \in \exp(\frac{1}{2})$.

11. The probability that the r.v. Y takes on values smaller than 2 is:

(A) 0.6321 (B) 0.8647 (C) 0.3679 (D) 0.7500 (E) 0.1353

12. If we define the r.v. $Z = X + Y$, the value of k such that $P(k \leq Z \leq 17.3) = 0.65$ is:

(A) 7.58 (B) 8.44 (C) 3.82 (D) 5.58 (E) 6.74

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: $X \in N(0, \sigma^2 = 1)$, $Y \in N(0, \sigma^2 = 4)$, $Z \in \chi^2_9$ and $V \in \gamma(\frac{1}{2}, \frac{11}{2})$.

13. If we define the r.v. $W_1 = \dfrac{3(X+Y)}{\sqrt{5Z}}$, the value of k such that $P(W_1 < k) = 0.80$ is:

(A) 0.883 (B) -1.380 (C) -0.883 (D) 1.380 (E) 0.840

14. If we define the r.v. $W_2 = \frac{V}{11X^2}$, the approximate value of k such that $P(W_2 \leq k) = 0.05$ is:

(A) 4.84 (B) 4.57 (C) 5.05 (D) 0.38 (E) 0.21

15. If we define the r.v. $W_3 = Z + V$, then $P(W_3 < 9.59)$ is:

(A) 0.025 (B) 0.100 (C) 0.975 (D) 0.950 (E) 0.050

Questions 16 and 17 refer to the following exercise:

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

$$P(X = -2) = 4\theta \qquad P(X = 0) = (1 - 8\theta) \qquad P(X = 2) = 4\theta$$

In order to estimate the parameter θ, a r.s. of size n, $X_1, X_2, \ldots, X_n$, has been taken.

16. The method of moments estimator of θ is:

(A) $\dfrac{\overline{X}}{16}$ (B) $\dfrac{\sum_{i=1}^n X_i^2}{32n}$ (C) $\displaystyle\sum_{i=1}^n X_i^2$ (D) $\dfrac{\sum_{i=1}^n X_i^2}{16n}$ (E) $\dfrac{16}{\overline{X}}$

17. To obtain an estimate of the parameter θ, a random sample of size $n = 10$ has been taken, providing four sample values equal to zero. The maximum likelihood estimate of θ is equal to:

(A) 0.1500 (B) 0.0250 (C) 0.1154 (D) 0.0750 (E) 0.1875

Questions 18 and 19 refer to the following exercise:

Let X be a r.v. having a uniform distribution on the interval $(0, 2\theta)$. That is, $X \in U(0, 2\theta)$. 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.

18. The maximum likelihood estimator of θ, $\hat{\theta}_{\mathrm{ML}}$, will be:

(A) $2\min\{X_i\}$ (B) $\frac{\max\{X_i\}}{2}$ (C) $\max\{X_i\}$ (D) $\frac{\min\{X_i\}}{2}$ (E) $2\max\{X_i\}$

19. The method of moments estimator of θ, $\hat{\theta}_{\mathrm{MM}}$, will be:

(A) $2\overline{X}$ (B) $\frac{\overline{X}}{2}$ (C) $\overline{X}$ (D) $\frac{\overline{X}}{4}$ (E) $4\overline{X}$

Questions 20 to 22 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 $\hat{\theta} = (2X_1 + 4X_2 + \ldots + 4X_{n-1} + 6X_n)/4n$ is proposed as an estimator of θ.

20. The bias of the proposed estimator is:

(A) 0 (B) $-\dfrac{\theta}{n}$ (C) $\dfrac{1}{2n}$ (D) $\dfrac{\theta}{4n}$ (E) $-\dfrac{\theta}{4n}$

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) $\dfrac{\theta(16n+8)}{16n^2}$ (B) $\dfrac{\theta}{4n}$ (C) $\dfrac{\theta}{2n}$ (D) $\dfrac{\theta(n+2)}{16n^2}$ (E) $\dfrac{\theta}{n^2}$

Questions 23 and 24 refer to the following exercise:

A discrete r.v. X has the following probability mass function, if H_0 is true:

X	1	2	3	4	5	6
$P_0(X)$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$

and, if H_1 is true:

X	1	2	3	4	5	6
$P_1(X)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	0	0

After observing a single value X, we reject the null hypothesis H_0 if $X = 1$ or $X = 3$.

23. The probability of a Type I error for the test described above is:

(A) $\frac{2}{3}$ (B) $\frac{1}{6}$ (C) $\frac{1}{3}$ (D) 0 (E) 1

24. The power for the test described above is:

(A) $\frac{1}{3}$ (B) $\frac{5}{6}$ (C) 1 (D) $\frac{1}{6}$ (E) $\frac{2}{3}$

Questions 25 and 26 refer to the following exercise:

Let X be a r.v. having a Poisson $\mathcal{P}(\lambda)$ distribution. In order to test $H_0 : \lambda = 0.50$ against $H_1 : \lambda = 1$, a r.s. of size $n = 8$ has been taken and it is decided that $Z = \sum_{i=1}^{8} X_i$ is used as test statistic.

25. At the $\alpha = 0.10$ significance level, the null hypothesis is rejected if:

(A) $Z \leq 7$ (B) $Z \leq 8$ (C) $Z \geq 7$ (D) $Z \geq 8$ (E) $Z \geq 9$

26. For the critical region above, the probability of Type II error is:

(A) 0.4530 (B) 0.5470 (C) 0.6867 (D) 0.4071 (E) 0.5925

Questions 27 and 28 refer to the following exercise:

An individual is interested in buying a new car from a specifically selected brand and size. Before doing so, s/he decides to ask for its price at 15 car dealers in the Basque Country, obtaining a sample mean price of 10500 euros with a **sample standard deviation** of 750 euros. We assume normality.

27. At the 90% confidence level, we can state that the mean price for that specific car in the Basque Country is included in the interval:

(A) (10147.22, 10852.78) (B) (10085.59, 10914.41) (C) (10240.51, 10759.49)
(D) (10182.42, 10817.58) (E) (10120.45, 10879.55)

28. At the 5% significance level, we wish to test the null hypothesis that the above car mean price in the Basque Country is of at most 10000 euros. The decision for this test will be:

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

29. We wish to estimate the proportion of the 300000 inhabitants of a given city that like to go out for dinner at least once per month. The minimum sample size required for this proportion estimation with a confidence level of 95% and an error no larger than 3%, if we use simple random sampling without replacement, is:

(A) 1509 (B) 746 (C) 1068 (D) 1064 (E) 750

30. We wish to estimate the mean weekly ice cream consumption (in kilograms) in a given city considering a population stratified in two strata according to having or not underage children at home. It is known that the first stratum includes 2500 individuals, with quasi-standard deviation $\sigma_1^* = 10$, and that the second stratum includes 7500 individuals, with quasi-standard deviation $\sigma_2^* = 25$. We wish to take a sample of 200 individuals. If we decide to use proportional allocation, the corresponding sample size for the **first stratum** will be:

(A) 150 (B) 176 (C) 50 (D) 100 (E) 24

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

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

$$f_X(x;\theta) = \frac{2}{\theta}\, x^{\frac{2}{\theta}-1}, 0 < x < 1,\ \theta > 0$$

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. In order to do so, you **must show** that the mean of this r.v. is $m = \mathrm{E}(X) = \alpha_1 = \frac{2}{2+\theta}$.

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 $0.7, 0.8, 0.3, 0.5, 0.2, 0.3, 0.6, 0.4, 0.3, 0.2$, obtain an estimate of θ by any of the methods in the above items.

B. (10 points, 25 minutes)

The frequency table below includes information on the number of deaths that occurred in a secondary road during a period of 150 consecutive days.

Number of deaths	0	1	2	3	≥ 4
Frequency	36	45	41	14	14

At the 5% significance level, test the hypothesis that the number of deaths occuring in that road follows a Poisson distribution. For estimation purposes, you can consider that, in the days that 4 or more accidents have occurred, the number of accidents occurring is exactly 4.

C. (10 points, 25 minutes)

Let X be a r.v. having a $\gamma(a, 3)$ distribution. That is, having the following probability density function:

$$f_X(x;a) = \frac{a^3}{2} x^2 e^{-ax},\ x > 0,\ a > 0$$

We wish to test the null hypothesis $H_0 : a = \frac{1}{2}$ against the alternative hypothesis $H_1 : a = 2$. In order to do so, a r.s. of size $n = 5$, $X_1, \ldots, X_5$, has been taken.

i) Obtain the form of the most powerful critical region for this test. In order to do so, you **must show**, as part of your solution to this exercise, that we reject the null hypothesis if $S = \sum_{i=1}^{n} X_i \leq C,\ C > 0$.

ii) Under the null hypothesis, what is the distribution of the test statistic?

iii) At the $\alpha = 0.05$ significance level, obtain the specific most powerful critical region for this test.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

The probability density function for the random variable X is:

$$f_X(x;\theta) = \frac{2}{\theta}\ x^{\frac{2}{\theta}-1}, 0 < x < 1,\ \theta > 0$$

We wish to estimate the parameter θ and, in order to do so, a random sample of size n, $X_1, X_2, \ldots, X_n$, has been taken.

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 to the first sample moment. Therefore, we need to compute the first population moment or mean for this r.v., α_1:

$$\alpha_1 = \int_0^1 x f_X(x;\theta)dx = \int_0^1 x\left[\frac{2}{\theta}\ x^{\frac{2}{\theta}-1}\right]dx = \int_0^1 \frac{2}{\theta} x^{\frac{2}{\theta}}dx = \frac{2}{\theta}\left[\frac{x^{\frac{2}{\theta}+1}}{\left(\frac{2}{\theta}+1\right)}\right|_0^1 = \frac{2}{(2+\theta)}\left[x^{\frac{2}{\theta}+1}\right|_0^1 = \frac{2}{(2+\theta)}$$

From this result, we have that:

$$\alpha_1 = \mathrm{E}(X) = a_1 = \overline{X} \Longrightarrow \frac{2}{(2+\theta)} = \overline{X} \Longrightarrow 2 = 2\overline{X} + \theta\overline{X} \Longrightarrow 2(1-\overline{X}) = \theta\overline{X} \Longrightarrow \hat{\theta}_{\mathrm{MM}} = \frac{2(1-\overline{X})}{\overline{X}}$$

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

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

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:

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

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

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

so that,

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

iii) If a r.s. of size $n = 10$ has provided the values $0.7, 0.8, 0.3, 0.5, 0.2, 0.3, 0.6, 0.4, 0.3, 0.2$, a method of moments estimate of θ will be:

$$\hat{\theta}_{\mathrm{MM}} = \frac{2(1-\overline{X})}{\overline{X}} = \frac{2(1-0.43)}{0.43} = 2.6511,$$

and a maximum likelihood estimate of θ will be:

$$\hat{\theta}_{\mathrm{ML}} = \frac{-2\ln\left[\Pi_{i=1}^n x_i\right]}{n} = \frac{-2\ln\left(0.7\times 0.8\times 0.3\times 0.5\times 0.2\times 0.3\times 0.6\times 0.4\times 0.3\times 0.2\right)}{10} = 1.9062$$

Exercise B

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 that needs to be estimated by maximum likelihood, and

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

$$\hat{\lambda}_{\mathrm{ML}} = \overline{x} = \frac{(0 \times 36) + (1 \times 45) + (2 \times 41) + (3 \times 14) + (4 \times 14)}{150} = \frac{225}{150} = 1.50.$$

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

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

remembering that:

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

In this way, we have that:

$P(X = 0) = \frac{e^{-1.5}\ (1.5)^0}{0!} = e^{-1.5} = 0.2231$

$P(X = 1) = \left(\frac{1.5}{1}\right)\ P(X = 0) = 0.3347$

$P(X = 2) = \left(\frac{1.5}{2}\right)\ P(X = 1) = 0.2510$

$P(X = 3) = \left(\frac{1.5}{3}\right)\ P(X = 2) = 0.1255$

$P(X \geq 4) = 1 - F_X(3) = 1 - 0.9343 = 0.0657$

With this information and in order to perform the test, we build the following table:

	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
0	36	0.2231	33.465	0.1920
1	45	0.3347	50.205	0.5396
2	41	0.2510	37.650	0.2981
3	14	0.1255	18.825	1.2367
≥ 4	14	0.0657	9.855	1.7434
Total	$n = 150$	1	$n = 150$	$z = 4.0098$

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 has been divided ($k = 5$) and h is the number of estimated parameters ($h = 1$).

The decision rule is: at the approximate 5% significance level, reject the null hypothesis if

$$z \geq \chi^2_{(5-1-1)\,,\,0.05} = \chi^2_{3,0.05}$$

In this case, we have that

$$4.0098 < 7.81 = \chi^2_{(5-1-1)\,,\,0.05} = \chi^2_{3,0.05},$$

so that, at a 5% approximate significance level, we do not reject the null hypothesis that the number of deaths occurring in that secondary road follows a Poisson distribution.

Exercise C.

i) X is es a r.v. with probability density function given by:

$$f_X(x;a) = \frac{a^3}{2}x^2e^{-ax}, x > 0, a > 0$$

and we wish to the null hypothesis that $a = \frac{1}{2}$ against the alternative hypothesis that $a = 2$. That is, $H_0 : a = \frac{1}{2}$, $H_1 : a = 2$.

For a given significance level, the most powerful critical region is obtained from the likelihood ratio test. That is,

$$\frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)} \leq k$$

In this case, the likelihood function is:

$$L(\vec{x}, a) = \left[\frac{a^3}{2}x_1^2e^{-ax_1}\right] \dots \left[\frac{a^3}{2}x_n^2e^{-ax_n}\right] = \frac{a^{3n}}{2^n}\left(\Pi_{i=1}^n x_i^2\right)e^{-a\sum_{i=1}^n x_i},$$

and the critical region will then be obtained from:

$$\frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)} = \frac{\frac{\left(\frac{1}{2}\right)^{3n}}{2^n}\left(\Pi_{i=1}^n x_i\right)^2 e^{-\frac{1}{2}\left(\sum_{i=1}^n x_i\right)}}{\frac{2^{3n}}{2^n}\left(\Pi_{i=1}^n x_i\right)^2 e^{-2\left(\sum_{i=1}^n x_i\right)}}$$

$$\frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)} = \frac{\left(\frac{1}{2}\right)^{3n} e^{-\frac{1}{2}\left(\sum_{i=1}^n x_i\right)}}{2^{3n}e^{-2\left(\sum_{i=1}^n x_i\right)}} = \frac{1}{2^{6n}}e^{\frac{3}{2}\sum_{i=1}^n x_i} \leq k_1$$

$$\Longrightarrow \frac{3}{2}\sum_{i=1}^n x_i \leq k_2 \Longrightarrow \text{Reject } H_0 \text{ if } S = \sum_{i=1}^n X_i \leq C$$

Therefore, the general form of the most powerful critical region for the test statistic $S = \sum_{i=1}^n X_i$ is RC $= (0, C]$.

ii) Given that, under H_0, $X_i \in \gamma\left(\frac{1}{2}, 3\right)$, and that $n = 5$, if we use the convolution property, we have that,

$$S = \sum_{i=1}^5 X_i \in \gamma\left(\frac{1}{2}, 15\right) \equiv \chi^2_{30}$$

iii) We wish to find the value of C in the specific case where $\alpha = 5\%$. That is,

$$\alpha = 0.05 = P(S \leq C|H_0)$$

Given that, under the null hypothesis, $S \in \chi^2_{30|}$, the value of C for which this condition holds, is $C = \chi^2_{30|0.95} = 18.5$. Therefore, the specific form of the critical region for this significance level will be:

$$\text{RC} = (0, 18.5]$$