STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2019-20
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
Second Call. July 6, 2020

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 30 minutes)

1. FREE-QUESTION. The capital of Spain is:

   (A) Paris (B) Sebastopol (C) Madrid (D) London (E) Pekin

Questions 2 to 5 refer to the following exercise:

The probability that a chess player wins a game is equal to 0.6. We assume independence between the different games being played. If in a given tournament 20 games are played:

2. The probability that the player wins exactly 5 games is:

   (A) 0.9997 (B) 0.0003 (C) 0.0013 (D) 0 (E) 0.0746

3. The probability that the player wins at least 8 games is:

   (A) 0.0355 (B) 0.0210 (C) 0.9790 (D) 0.0565 (E) 0.9435

4. The probability that the player **loses** more than five games, but fewer than ten games is:

   (A) 0.6297 (B) 0.8728 (C) 0.5053 (D) 0.8215 (E) 0.7469

5. If 50 games are played, the approximate probability that the player wins at least 40 games is:

   (A) 0.00135 (B) 0.0047 (C) 0.0031 (D) 0.0212 (E) 0.99865

6. Let X be a r.v. having a Poisson distribution such that $P(4) = P(3)$. The probability $P(X = 2)$ is:

   (A) 0.2381 (B) 0.4232 (C) 0.1465 (D) 0.2308 (E) 0.2240

7. Let Y be a r.v. having a Poisson distribution with variance equal to 2. The probability $P(2 \leq Y \leq 4)$ is:

   (A) 0.2707 (B) 0.3068 (C) 0.5413 (D) 0.5895 (E) 0.5774

8. Let X_1 and X_2 be two independent r.v. having Poisson distributions, the former with mean equal to 2 and the latter with mean equal to 4. If we define the r.v. $Z = X_1 + X_2$, the probability $P(Z \geq 5)$ is:

   (A) 0.2851 (B) 0.8270 (C) 0.7149 (D) 0.5543 (E) 0.1730

9. If we let $Z = X_1 + \ldots + X_{90}$, where the X_i's are independent Poisson r.v. with mean equal to 1, then $P(Z \leq 69)$ is, approximately:

   (A) 0.0154 (B) 0.4129 (C) 0.9846 (D) 0.0179 (E) 0.9821

Questions 10 to 12 refer to the following exercise:

Let X be a r.v. with a $\gamma(\frac{1}{2}, 1)$ distribution.

10. The mean and variance of the r.v. X are, respectively:

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

11. The probability $P(X > 4.61)$ is:

(A) 0.10 (B) 0.25 (C) 0.75 (D) 0.01 (E) 0.90

12. If we define the r.v. $Y = \frac{1}{4}X$, then the distribution of the r.v. Y is:

(A) χ_8^2 (B) $\exp(\lambda = 0.125)$ (C) $\gamma(\frac{1}{2}, 4)$ (D) $\gamma(\frac{1}{2}, \frac{1}{4})$ (E) $\exp(\lambda = 2)$

Questions 13 to 15 refer to the following exercise:

Let V, X, Y and Z be four independent r.v. such that their distributions are as follows: $V \in N(0,4)$, $X \in N(0,1)$, $Y \in \chi_4^2$ and $Z \in \gamma(\frac{1}{2}, 5)$.

13. If we define the r.v. $W_1 = Y + Z$, then the probability that the r.v. W_1 takes on values between 10.2 and 21.1 is:

(A) 0.95 (B) 0.35 (C) 0.65 (D) 0.10 (E) 0.75

14. If we define the r.v. $W_2 = \frac{V}{\sqrt{Y}}$, then the probability that the r.v. W_2 takes on values smaller than or equal to -2.13 is:

(A) 0.05 (B) 0.90 (C) 0.95 (D) 0.10 (E) 0.20

15. The distribution of the r.v. $W_3 = \frac{2\,(X^2+Y)}{Z}$ is:

(A) $F_{2,1}$ (B) $F_{5,10}$ (C) $F_{1,2}$ (D) $F_{10,5}$ (E) All false

16. Let the r.v. $W_4 \in F_{10,5}$, then the value of k such that $P(W_4 < k) = 0.10$ is, approximately:

(A) 0.10 (B) 0.17 (C) 0.90 (D) 0.30 (E) 0.40

Questions 17 and 18 refer to the following exercise:

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

$$P(0) = \theta \qquad P(1) = \frac{\theta}{2} \qquad P(2) = 1 - \frac{3\theta}{2}$$

In order to estimate the parameter θ a r.s. of size $n = 10$ has been taken, providing the following results: 0, 0, 1, 0, 2, 0, 2, 0, 0, 1.

17. The method of moments estimate of θ is:

(A) 0.20 (B) 0.80 (C) 0.40 (D) 0.56 (E) 0.27

18. The maximum likelihood estimate of θ is:

(A) 0.40 (B) 0.60 (C) 0.27 (D) 0.53 (E) 0.20

Questions 19 to 20 refer to the following exercise:

Let $X_1, \ldots, X_n$ be a r.s. from a r.v. X having probability density function given by:

$$f(x,\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

19. The maximum likelihood estimator of the parameter θ is:

(A) $\sum_{i=1}^{n} X_i$ (B) $\frac{n}{\prod_{i=1}^{n} \ln X_i}$ (C) $\frac{-n}{\ln\left(\prod_{i=1}^{n} X_i\right)}$ (D) $\prod_{i=1}^{n} X_i$ (E) $\ln\left(\prod_{i=1}^{n} X_i\right)$

20. The method of moments estimator of the parameter θ is:

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

Questions 21 and 23 refer to the following exercise:

Let X be a r.v. having a uniform $U(2\theta-2, 2\theta+2)$ distribution. It is known that the mean and the variance of this r.v. are 2θ and $\frac{4}{3}$, respectively. 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.

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

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

22. The bias of the method of moments estimator of θ is:

(A) 0 (B) $\theta + 4$ (C) $\theta^2 + 4$ (D) $\theta + 2$ (E) θ

23. The variance of the method of moments estimator of θ is:

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

Questions 24 and 25 refer to the following exercise:

We wish to test the null hypothesis that the course first call passing percentage is at least 50% against the alternative hypothesis that it is smaller than 50%. In order to do so, a r.s. of size $n = 5$ has been taken, and the null hypothesis is rejected if the number of students passing the course is smaller than 2.

24. The significance level for this test is:

(A) 0.1875 (B) 0.05 (C) 0.8125 (D) 0.1562 (E) 0.0313

25. The probability of type II error, when the true proportion value is equal to 0.4, is:

(A) 0.3370 (B) 0.7408 (C) 0.6630 (D) 0.1760 (E) 0.6826

Questions 26 and 27 refer to the following exercise:

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

$$f(x, \theta) = \theta(\theta+1)x(1-x)^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0$$

We wish to test the null hypothesis $H_0 : \theta = 1$ against the alternative hypothesis $H_1 : \theta = 2$. In order to do so, a r.s. of size $n = 1$, X_1, has been taken. The null hypothesis is rejected if $X_1 \leq C$.

26. At the $\alpha = 0.04$ significance level, the critical region for X_1 will be:

(A) $(0, 0.20]$ (B) $[0, 0.96]^C$ (C) $(0, 0.04]$ (D) $[0.04, 0.96]$ (E) $[0.20, 1)$

27. If we decide to reject the null hypothesis if $X_1 \leq 0.40$, the probability of type II error for this test is, approximately:

(A) 0.65 (B) 0.40 (C) 0.35 (D) 0.60 (E) 0.10

Questions 28 and 29 refer to the following exercise:

We wish to test the null hypothesis that the distribution, stratified by age, of the individuals for 4 large phone companies is the same. In order to do so, four r.s. of 600, 500, 700 and 800 clients, respectively, from those companies were taken. These clients were classified as a function of their ages in four different age intervals: younger than 30 years, between 30 and 44 years, between 45 and 60 years and older than 60 years of age.

28. The most appropriate test to be carried out is:

(A) Homogeneity

(B) Independence

(C) Equality of means

(D) Goodness of fit to a totally specified distribution

(E) All false

29. Under the null hypothesis H_0, the distribution of the test statistic for this test is:

(A) χ^2_{16} (B) χ^2_6 (C) χ^2_9 (D) t_{16} (E) N(0,1)

30. We wish to find out if the proportion of primary education teachers that typically assign homework to their students is the same in two given regions, say A and B. In order to do so, a r.s. of 250 teachers is taken in each one of the two regions, with the result that 190 teachers from region A and 204 from region B responded positively to this question. A 95% confidence interval for the difference of the proportion of teachers that typically assign homework to their students, $(p_A - p_B)$ is, approximately:

(A) $(-0.127,\ 0.015)$ (B) $(-0.116,\ 0.004)$ (C) $(-0.127,\ 0.127)$ (D) $(0,\ 0.004)$ (E) $(0,\ 0.015)$

EXERCISES (Time: 60 minutes)

A. (10 points, 20 minutes)

Let X be a continuous r.v. having a $\gamma(\frac{1}{\theta}, 3)$ distribution, so that its probability density function is given by:

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

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

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

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

iii) Choose one of the two estimators of θ above. Is the chosen estimator unbiased for θ? Is it consistent? You must justify all of the responses provided here.

B. (10 points, 20 minutes)

It is known that, for region A, the annual electricity expense per family, in euros, follows a normal distribution. A r.s. of 15 families in this region is taken, providing a sample mean and variance equal to 710 and 2300, respectively.

i) Find the 95% confidence interval for the mean annual electricity expense per family for region A.

ii) At the 5% significance level, carry out the test of the null hypothesis that the mean annual electricity expense per family is of at least 750 euros.

iii) Find the 95% confidence interval for the variance of the annual electricity expense per family for region A.

iv) It is known that, in region B, the annual electricity expense per family, in euros, follows a normal distribution with variance equal to that in region A. A r.s. of 15 families in this region is taken, providing a sample mean and variance equal to 800 and 2500, respectively. At the 5% significance level, carry out the test of the null hypothesis that the mean annual electricity expense in the two regions is the same.

C. (10 points, 20 minutes)

Managers for a given touristic destination wish to design their next promotion campaign. In order to do so, they are interested in testing if the probability distribution modelling the clients main reason for the trip is as follows:

Reason for the trip	Sun and beach	Cultural	Sports practice	Other
Probabilities	$3\,\theta$	θ	$2\,\theta$	$(1-6\,\theta)$

In order to be able to test the null hypothesis that this distribution is the one above, a r.s. of 500 clients is taken, and the question about their main reason for their trip is asked. Results were as follows: 200 clients were looking for a sun and beach destination, 100 for a cultural one, 150 for a sports practice destination, and the remaining 50 for other type of destination.

i) Obtain the maximum likelihood estimate for the parameter θ.

ii) At the 5% significance level, test the managers' hypothesis that the probability distribution modelling the clients' reason for their trip is the one following their beliefs. You should provide all relevant details about the specific procedure selected to test this hypothesis.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

Given that the r.v. $X \in \gamma(\frac{1}{\theta}, 3)$, its probability density function, mean and variance are, respectively:

$$f(x,\theta) = \frac{1}{2\theta^3}\, x^2 e^{-\frac{x}{\theta}}, \quad x > 0, \qquad \mathrm{E}(X) = \frac{3}{1/\theta} = 3\theta \quad \text{y} \quad \mathrm{Var}(X) = \frac{3}{(1/\theta)^2} = 3\theta^2$$

In order to estimate the parameter θ, a r.s. of size n has been taken.

i) Method of moments estimator. To obtain the method of moments estimator, we make the first population and sample moments equal. That is,

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

Given that the first population moment of the r.v. $X \in \gamma(\frac{1}{\theta}, 3)$ is $\mathrm{E}(X) = 3\theta$:

$$\alpha_1 = \mathrm{E}(X) = 3\theta = a_1 = \overline{X} \Longrightarrow \hat{\theta}_{\mathrm{MM}} = \frac{\overline{X}}{3}$$

ii) Maximum likelihood estimator. The corresponding likelihood function for the sample is:

$$L(\vec{x},\theta) = f(x_1,\theta)\ \ldots\ f(x_n,\theta) = \left[\frac{1}{2\theta^3}\, x_1^2 e^{-x_1/\theta}\right] \cdots \left[\frac{1}{2\theta^3}\, x_n^2 e^{-x_n/\theta}\right] = \frac{1}{2^n\theta^{3n}} \left(\prod_{i=1}^n x_i^2\right) e^{-\sum_{i=1}^n x_i/\theta}$$

To obtain the maximum likelihood estimator, we have to maximize the logarithm of the likelihood function. In this way, the natural logarithm of the likelihood function will be:

$$\ln L(\vec{x},\theta) = -n\ln 2 - 3n\ln\theta + \ln\left(\prod_{i=1}^n x_i^2\right) - \frac{\sum_{i=1}^n x_i}{\theta}$$

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

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta} = -\frac{3n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} = 0$$

so that,

$$\frac{\sum_{i=1}^n x_i}{\theta} = 3n \Longrightarrow \hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^n X_i}{3n} = \frac{\overline{X}}{3}$$

iii) We have that $\hat{\theta}_{\mathrm{MM}} = \hat{\theta}_{\mathrm{ML}} = \frac{\overline{X}}{3}$

Unbiasedness. In order to check if the estimator is unbiased, we need to verify if $\mathrm{E}(\hat{\theta}_{\mathrm{MV}}) = \theta$ holds. In this case,

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}) = \mathrm{E}\left(\frac{\overline{X}}{3}\right) = \frac{1}{3}\,\mathrm{E}(\overline{X}) = \frac{1}{3}\,\mathrm{E}(X) = \frac{1}{3}\,(3\theta) = \theta$$

Therefore, it is unbiased

Consistency. In order to verify if the estimator is consistent, we compute its variance.

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MV}}) = \mathrm{Var}\left(\frac{\overline{X}}{3}\right) = \left(\frac{1}{3}\right)^2 \mathrm{Var}(\overline{X}) = \left(\frac{1}{9}\right)\left(\frac{\mathrm{Var}(X)}{n}\right) = \left(\frac{1}{9}\right)\left(\frac{3\theta^2}{n}\right) = \frac{\theta^2}{3n}$$

Given that it is an unbiased estimator and that its variance tends to zero as n goes to infinity, the two sufficient conditions for consistency hold and, therefore, we can state that it is a consistent estimator for θ.

Exercise B

Let X = annual electricity expense (in euros) per family for region A, $X \in N(m_1, \sigma_1^2)$

Y = annual electricity expense (in euros) per family for region B, $Y \in N(m_2, \sigma_2^2)$.

X and Y are independent r.v. and $\sigma_1^2 = \sigma_2^2$.

Samples: $n_1 = 15,\ \bar{x} = 710,\ s_1^2 = 2300, \quad n_2 = 15,\ \bar{y} = 800,\ s_2^2 = 2500.$

i) Given that $X \in N(m_1, \sigma_1^2)$, with unknown variance, the corresponding confidence interval for the mean m_1 is:

$$\mathrm{CI}_{1-\alpha}(m_1) = \left(\overline{x} \pm t_{\overline{n_1-1}|\frac{\alpha}{2}} \frac{s_1}{\sqrt{n_1-1}}\right) \qquad t_{\overline{14}|\frac{0.05}{2}} = 2.14$$

$$\mathrm{CI}_{0.95}(m_1) = \left(710 \pm 2.14 \cdot \sqrt{\frac{2300}{14}}\right) = (\mathbf{682.57},\ \mathbf{737.43})$$

ii) We have to test $H_0 : m_1 \geq 750$ against $H_1 : m_1 < 750$

Under H_0: $\quad \dfrac{\bar{X} - 750}{S_1/\sqrt{n_1-1}} \in t_{\overline{n_1-1}|}$

At the α significance level, the decision rule states that we reject H_0 if: $\quad \dfrac{\bar{x} - 750}{s_1/\sqrt{n_1-1}} \leq -t_{\overline{n_1-1}|\alpha}$

In this case: $\quad \dfrac{710 - 750}{\sqrt{2300/14}} = -3.12 \leq -1.76 = t_{\overline{14}|0.05}$

Therefore, at the 5% significance level, we reject the null hypothesis that the annual mean electricity expense per family in region A is of at least 750 euros.

iii) Given that $X \in N(m_1, \sigma_1^2)$, with unknown mean, the confidence interval for σ_1^2 is:

$$\mathrm{CI}_{1-\alpha}(\sigma_1^2) = \left(\frac{n_1 s_1^2}{\chi^2_{\overline{(n_1-1)}|\alpha/2}},\quad \frac{n_1 s_1^2}{\chi^2_{\overline{(n_1-1)}|1-\alpha/2}}\right)$$

$$\mathrm{CI}_{0.95}(\sigma_1^2) = \left(\frac{15 \cdot 2300}{\chi^2_{\overline{14}|0.025}},\ \frac{15 \cdot 2300}{\chi^2_{\overline{14}|0.975}}\right) = \left(\frac{34500}{26.1},\ \frac{34500}{5.63}\right) = (\mathbf{1321.84},\ \mathbf{6127.89})$$

iv) We have to test:

$$H_0 : m_1 = m_2 \qquad \equiv \qquad m_1 - m_2 = 0$$

$$H_1 : m_1 \neq m_2 \qquad \equiv \qquad m_1 - m_2 \neq 0$$

X and Y are independent normal r.v. with $\sigma_1^2 = \sigma_2^2$. Thus, under H_0, $\quad \dfrac{(\bar{X} - \bar{Y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\ \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2}}} \in t_{\overline{n_1+n_2-2}|}$

At the α significance level, the decision rule states that we reject H_0 if:

$$\left| \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\ \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}}} \right| \geq t_{\overline{n_1+n_2-2}|\frac{\alpha}{2}}$$

In this case: $\quad \left| \dfrac{(710 - 800) - 0}{\sqrt{\frac{1}{15} + \frac{1}{15}}\ \sqrt{\frac{15 \cdot 2300 + 15 \cdot 2500}{15+15-2}}} \right| = |-4.86| \ \geq 2.05 = t_{\overline{28}|\frac{0.05}{2}}$

Therefore, at the 5% significance level, we reject the null hypothesis that the annual mean electricity expense is the same for the two regions A and B.

Exercise C

Reason for the trip	Sun and beach (I)	Cultural (II)	Sports practice (III)	Other (IV)
Probabilities	$3\,\theta$	θ	$2\,\theta$	$(1-6\,\theta)$
n_i	200	100	150	50

i) The corresponding likelifood function is:

$$L(\theta) = (3\theta)^{200}\theta^{100}(2\theta)^{150}(1-6\theta)^{50} = 3^{200}\,2^{150}\theta^{450}(1-6\theta)^{50}$$

If we take natural logarithm, we have that:

$$\ln L(\theta) = 200\ln(3) + 150\ln(2) + 450\ln(\theta) + 50\ln(1-6\theta)$$

If we take the derivative with respect to θ and make it equal to zero to maximize the likelihood function, we have that:

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{450}{\theta} + \frac{50\,(-6)}{(1-6\theta)} = 0 \Longrightarrow 450(1-6\theta) = 300\,\theta \Longrightarrow 450 = 3000\,\theta \Longrightarrow \hat{\theta}_{\text{MV}} = \frac{450}{3000} = 0.15$$

ii) Since the parameter θ is unknown, this corresponds to a **goodness of fit test to a partially specified distribution.**

H_0 : The managers' probability distribution is correct

H_1 : The managers' probability distribution is not correct

First of all, we have that the estimated probabilities, $\hat{p}_i$, for each one of the reasons for the trip are:

$$\hat{P}(\text{I}) = 3(0.15) = 0.45; \qquad \hat{P}(\text{II}) = 0.15; \qquad \hat{P}(\text{III}) = 2(0.15) = 0.30; \qquad \hat{P}(\text{IV}) = 1-6(0.15) = 0.10$$

Given that we have estimated the parameter θ, we have that $h = 1$. In addition, as we have four possible reasons for the trip, $k = 4$. With this information, we build the corresponding table:

	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$
Sun and beach (I)	200	0.45	225	2.78
Cultural (II)	100	0.15	75	8.33
Sports practice (III)	150	0.30	150	0.00
Other (IV)	50	0.10	50	0.00
	$n = 500$	1	$n = 500$	$z = 11.11$

Under the null hypothesis that the managers' probability distribution is the correct one, the test statistic $Z = \sum_i \frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} \xrightarrow{d} \chi^2_{k-h-1}$, where the number of classes is $k = 4$ and the number of estimated parameters is $h = 1$.

At the $\alpha = 0.05$ significance level, the decision rule states that we reject H_0 if:

$$z \geq \chi^2_{k-h-1,\,0.05} = \chi^2_{2,0.05}$$

In this case:

$$z = 11.11 \geq 5.99 = \chi^2_{2,0.05},$$

so that, at the 5% significance level, the null hypothesis that the probability distribution modelling the clients' reason for their trip is the one following the managers' beliefs is rejected.