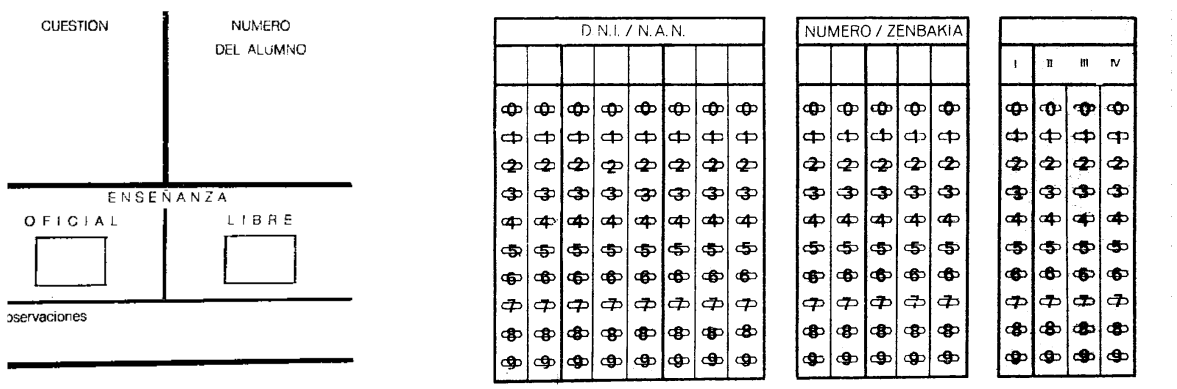
INSTRUCTIONS

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, while each question incorrectly answered will penalize your grade by 0.2 points. Questions that have not been answered do not penalize your grade in any form.
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, just as is illustrated in the example.
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 15 points in each part of the exam to pass it. However, exams having a multiple choice question part with grades greater than or equal to 14 could, under special circumstances, be compensated with a good grade in the exercises part of the exam.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises. In "Resit" (column II) you will write the number of times you have registered for an exam in this course *not including this one.*

Example:

12545 PEREZ, Ernesto Exam type 0 Resit



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 4 refer to the following exercise:

A candidate for a government job post has to answer a specific questionnaire including 10 multiple choice questions. The probability of selecting the correct answer for each one of the questions in the exam is 0.3.

2. What is the probability that the candidate does not obtain any correct answer in the questionnaire?

   (A) 1 (B) 0 (C) 0.3 (D) 0.9718 (E) 0.0282

3. What is the probability that the candidate obtains at least 5 correct answers in the questionnaire?

   (A) 0.8497 (B) 0.0473 (C) 0.9527 (D) 0.1503 (E) 0.1030

4. If the questionnaire has 100 questions, what would be the approximate probability that the candidate obtains at most 22 correct answers in the questionnaire?

   (A) 0.05 (B) 0.85 (C) 0.25 (D) 0.15 (E) 0.95

Questions 5 to 7 refer to the following exercise:

The number of people using a given teller machine each hour follows a Poisson distribution with mean 4. It is assumed that the distributions for the use of the teller machine for the different hours are independent.

5. The probability that at least 6 people use the teller machine in a given hour is:

   (A) 0.2149 (B) 0.8893 (C) 0.7851 (D) 0.1107 (E) 0.4327

6. The probability that at most 10 people use the teller machine in a two-hour period is:

   (A) 0.8159 (B) 0.7166 (C) 0.5824 (D) 0.2834 (E) 0.1841

7. The approximate probability that at most 50 people use the teller machine in a ten-hour period is:

   (A) 0.0485 (B) 0.9515 (C) 0.7824 (D) 0.3728 (E) 0.5428

8. Let $\{X_n\}_{n\in N}$ be a sequence of r.v. having a $N(0, \sigma^2 = 1/n)$ distribution. The sequence will converge to $X = 0$:

   (A) only in distribution and probability.

   (B) only in quadratic mean.

   (C) only in probability.

   (D) only in distribution.

   (E) in distribution, probability and quadratic mean.

9. Let $\{X_n\}_{n\in N}$ be a sequence of r.v. with characteristic function given by $\psi_n(u) = \frac{1}{n}\left(1 - e^{iu}\right) + e^{iu}$. The sequence will converge:

(A) Only in distribution to $X = 0$.

(B) Only in probability to $X = 0$.

(C) Only in distribution to $X = 1$.

(D) In probability and distribution to $X = 1$.

(E) In probability and distribution to $X = 0$.

Questions 10 and 11 refer to the following exercise:

Let X_1 and X_2 be two independent and identically distributed r.v. having a $\gamma(0.5, 0.75)$ distribution.

10. The distribution of the r.v. $X_1 + X_2$ is:

(A) $\gamma(0.5, 0.75)$ (B) $\gamma(1, 0.75)$ (C) All false (D) $\gamma(1, 1.5)$ (E) $\gamma(0.5, 1.5)$

11. The distribution of the r.v. $2X_1$ is:

(A) $\gamma(0.5, 1.5)$ (B) $\gamma(0.25, 0.75)$ (C) $\gamma(1, 0.75)$ (D) $\gamma(0.5, 0.75)$ (E) All false

12. Let X be a r.v. having a $\gamma(0.5, 4)$ distribution. The mean and variance for this r.v. are, respectively:

(A) 2 and 1 (B) 0.5 and 4 (C) 8 and 8 (D) 2 and 2 (E) 8 and 16

Questions 13 to 15 refer to the following exercise:

Let X, Y, Z and W four independent r.v. with distributions given by $X \in N(0, \sigma^2 = 1)$, $Y \in N(2, \sigma^2 = 4)$, $Z \in N(-1, \sigma^2 = 9)$ and $W \in N(4, \sigma^2 = 4)$.

13. The probability that the r.v. defined as $X^2 + \left(\frac{Y-2}{2}\right)^2 + \left(\frac{Z+1}{3}\right)^2$ takes on values greater than 4.11 is:

(A) 0.10 (B) 0.75 (C) 0.90 (D) 0.50 (E) 0.25

14. The probability that the r.v. defined as $\dfrac{\frac{W-4}{2}}{\sqrt{\frac{X^2+\left(\frac{Y-2}{2}\right)^2+\left(\frac{Z+1}{3}\right)^2}{3}}}$ takes on values smaller than 2.35 is:

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

15. The probability that the r.v. $\dfrac{X^2 + \left(\frac{Y-2}{2}\right)^2}{\left(\frac{Z+1}{3}\right)^2 + \left(\frac{W-4}{2}\right)^2}$ takes on values smaller than or equal to 19 is:

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

Questions 16 and 17 refer to the following exercise:

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

$$P(0) = p, \quad P(1) = 0.2, \quad P(2) = 1 - p - 0.2$$

In order to estimate the parameter p, a r.s. of size $n = 9$ has been taken, providing the values 0,0,0,2,2,0,1,0,2.

16. The maximum likelihood estimate of p is:

(A) 1/2 (B) 2/3 (C) 5/9 (D) 3/4 (E) 1/3

17. The method of moments estimate of p is, approximately:

(A) 0.32 (B) 0.64 (C) 0.51 (D) 0.71 (E) 0.22

Questions 18 and 19 refer to the following exercise:

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

$$f(x;\theta) = 2e^{-2(x-\theta)}, \quad x > \theta$$

It is known that the mean of this r.v. is $m = \frac{1}{2} + \theta$. In order to estimate the parameter θ, a random sample of size n, $X_1, X_2, \ldots, X_n$, has been taken.

18. The method of moments estimator of θ is:

(A) $\overline{X}$ (B) $\overline{X} - \frac{1}{2}$ (C) $\min\{X_i\}$ (D) $\max\{X_i\}$ (E) $\overline{X} + \frac{1}{2}$

19. The maximum likelihood estimator of θ is:

(A) $\overline{X} + \frac{1}{2}$ (B) $\overline{X} - \frac{1}{2}$ (C) $\overline{X}$ (D) $\max\{X_i\}$ (E) $\min\{X_i\}$

Questions 20 to 23 refer to the following exercise:

In order to estimate the mean of a normal distribution with variance $\sigma^2 = 36$, a random sample of size n has been taken and the estimator given by $\left(\dfrac{n}{n-1}\right)\overline{X}$ is proposed.

20. Is this an unbiased estimator?

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

21. The variance of this estimator is:

(A) $\dfrac{36}{(n-1)}$ (B) $\dfrac{36n}{(n-1)^2}$ (C) $\dfrac{36}{n}$ (D) $\dfrac{36n^2}{(n-1)^2}$ (E) All false

22. Is this a consistent estimator?

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

23. Is this an efficient estimator?

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

Questions 24 to 26 refer to the following exercise:

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

$$f(x;\theta) = \begin{cases} (\theta+1)x^{\theta} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

We wish to test the null hypothesis $\theta = 1$ against the alternative hypothesis $\theta = 2$. In order to do so, a random sample of size $n = 1$ has been taken.

24. Suppose we propose the following rejection rule: reject the null hypothesis if $x \in (0.25, 0.5)$. The level of significance for this test is:

(A) 0.1875 (B) 0.8906 (C) 0.1094 (D) 0.8125 (E) 0.25

25. For this rejection rule, the probability of type II error is:

(A) 0.75 (B) 0.1875 (C) 0.8125 (D) 0.1094 (E) 0.8906

26. The rejection region for the most powerful test for the observed value x is of the form:

(A) $(0, k_2)$ (B) (k_1, k_2) (C) All false (D) $(k_1, k_2)^C$ (E) $(k_1, 1)$

Questions 27 to 30 refer to the following exercise:

A car manufacturer claims that one of the models he produces (model A) has a lower gas consumption that the model (model B) produced by a competitor. In order to test his claim, two different random samples (one of each car models) of size 31 were taken and the gas consumption (in liters) for a 100 Km. distance was observed. The resulting measures were: $\overline{x}_A = 8.2$, $\overline{x}_B = 7.9$, $s_A^2 = 1.44$, $s_B^2 = 1.21$. It is assumed that the gas consumption distributions for the two models are normal, independent and have a common variance.

27. The 95% confidence interval for the difference of the mean gas consumption for the two car models, $m_A - m_B$ is, approximately:

(A) $(0, 0.64)$ (B) $(-0.32, 0.32)$ (C) $(-0.29, 0.89)$ (D) $(-0.12, 0.72)$ (E) $(-0.18, 0.78)$

28. The firm wished to test the null hypothesis that the mean gas consumption for his model A is **not larger than** the one for model B. At the 5% significance level, the decision will be:

(A) -

(B) No decision can be adopted.

(C) Do not reject the null hypothesis.

(D) Reject the null hypothesis.

(E) -

29. The 95% confidence interval for the variance of the gas consumption for model A is:

(A) $(1.21, 1.65)$ (B) $(1.32, 1.59)$ (C) $(1.12, 1.96)$ (D) $(0.95, 2.66)$ (E) $(1.01, 1.49)$

30. The firm wished to test the null hypothesis that the variance for the gas consumption for model A is smaller than or equal to 1.3 against the alternative hypothesis that it is larger than 1.3. At the 5% significance level, the decision will be:

(A) Reject the null hypothesis. (B) No decision can be adopted. (C) -

(D) - (E) Do not reject the null hypothesis.

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

It is known that the number of daily labor absences taking place in a given firm follows a Poisson distribution for which the value of its parameter λ is unknown. The labor union representative claims that this value is $\lambda = 1$, while the firm's personnel director claims that it is $\lambda = 2$ instead. In order to test the null hypothesis that the labor union representative's claim is correct against the alternative hypothesis that the firm's personnel director's claim is correct, a random sample of four days (i.e., $n = 4$) have been taken.

i) At the 5% significance level, find the most powerful critical region for this test.

ii) What is the power for this test?

iii) If the sample provides the values 1, 0, 2 and 2, what would be the firm's decision about the value of the parameter λ?

B. (10 points, 25 minutes)

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

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

In addition, it is known that:

$$E(X) = \theta - 1$$

$$\text{Var}(X) = (\theta - 1)^2$$

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

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

ii) If we propose to estimate θ with the estimator $\hat{\theta} = \overline{X} + 1$. Is this an unbiased estimator of θ? Is it consistent? Is it efficient? Provide all relevant details to justify your answers.

C. (10 points, 25 minutes)

A firm wishes to commercialize a new product and it is interested in knowing whether its product should be sold to the general public or, on the contrary, to a more specific age group. In order to decide on this issue, a random sample of 500 individuals was taken, recording both their age and their views on the product. The results of this study can be summarized as follows.

Out of the 200 individuals with an age in the range between 18 and 30 years old, 130 claim that they liked the product, 40 indicated that they did not like the product and the remaining 30 did not have a clear opinion on it.

Out of the 150 individuals with an age in the range between 31 and 50 years old, 85 claimed that they liked the product, 40 indicated that they did not like the product and the remaining 25 did not have a clear opinion on it.

Out of the 150 individuals with an age over 50 years old, 75 claimed that they liked the product, 35 indicated that they did not like the product and the remaining 40 did not have a clear opinion on it.

At the 5% significance level, test the null hypothesis that preferences for that product do not change with age.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A)

$$H_0 : \lambda = 1$$
$$H_1 : \lambda = 2$$

i) We obtain the most powerful critical region from the likelihood ratio test and the Neyman-Pearson Theorem.

$$\frac{L(\vec{x}; \lambda = 1)}{L(\vec{x}; \lambda = 2)} \leq k \quad \text{for} \quad k > 0$$

In this case, the likelihood function will be:

$$L(\vec{x}; \lambda) = \frac{e^{-n\lambda}\lambda^{x_1+x_2+x_3+x_4}}{x_1!x_2!x_3!x_4!},$$

so that, the most powerful critical region will be obtained from the inequality:

$$\frac{\left(\frac{e^{-n}1^{x_1+x_2+x_3+x_4}}{x_1!x_2!x_3!x_4!}\right)}{\left(\frac{e^{-2n}2^{x_1+x_2+x_3+x_4}}{x_1!x_2!x_3!x_4!}\right)} \leq k$$

$$e^n \left(\frac{1}{2}\right)^{x_1+x_2+x_3+x_4} \leq k$$

$$\left(\frac{1}{2}\right)^{x_1+x_2+x_3+x_4} \leq k_1$$

$$x_1 + x_2 + x_3 + x_4 \geq C$$

Therefore, the decision rule will be to reject H_0 if:

$$X_1 + \ldots + X_4 \geq C$$

To be able to determine C, we fixed the significance level to $\alpha = 0.05$.

Under H_0, $X_1 + \ldots + X_4 \in \mathcal{P}(\lambda = 4)$, so that

$$\begin{aligned}
\alpha = 0.05 \geq P(\text{reject } H_0 \mid \lambda = 1) &= \\
&= P(X_1 + \ldots + X_4 \geq C \mid \lambda = 1) = \\
&= 1 - F_{\mathcal{P}(\lambda=4)}(C-1) \\
F_{\mathcal{P}(\lambda=4)}(C-1) &\geq 0.95 \\
C - 1 &= 8 \\
C &= 9
\end{aligned}$$

That is, the decision rule will be to reject H_0 if:

$$X_1 + \ldots + X_4 \geq 9$$

ii) Under H_1, $X_1 + \ldots + X_4 \in \mathcal{P}(\lambda = 8)$, so that,

$$\begin{aligned}
\text{Power}(\lambda = 2) &= P\left(\text{reject } H_0 \mid \lambda = 2\right) \\
&= P(X_1 + \ldots + X_4 \geq 9 \mid \lambda = 2) = \\
&= 1 - F_{\mathcal{P}(\lambda=8)}(8) = \\
&= 1 - 0.5925 = \\
&= 0.4075
\end{aligned}$$

iii) If we use the results provided by the sample, we have that

$$x_1 + \ldots + x_4 = 1 + 0 + 2 + 2 = 5$$

Now, given that $5 < 9$ and at the 5% significance level, we do not reject the null hypothesis. Therefore, we decide that the labor union representative's claim is the correct one.

Exercise B

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

$$\text{E}(X) = \theta - 1$$

$$\text{Var}(X) = (\theta - 1)^2$$

i) **Maximum likelihood estimator**

$$\begin{aligned}
L(\vec{X};\theta) = f(X_1;\theta)\ldots f(X_n;\theta) &= \frac{1}{\theta - 1}\; e^{-\frac{1}{\theta-1}X_1} \cdots \frac{1}{\theta - 1}\; e^{-\frac{1}{\theta-1}X_n} = \\
&= \frac{1}{(\theta - 1)^n}\; e^{-\frac{1}{\theta-1}\sum_{i=1}^n X_i}
\end{aligned}$$

$$\ln L(\vec{X};\theta) = -n\ln(\theta - 1) - \frac{1}{\theta - 1}\sum_{i=1}^{n} X_i$$

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

$$\frac{n}{\theta - 1} = \frac{\sum_{i=1}^n X_i}{(\theta - 1)^2}$$

$$n(\theta - 1) = \sum_{i=1}^{n} X_i$$

$$\hat{\theta}_{\text{ML}} = \frac{\sum_{i=1}^n X_i}{n} + 1 = \overline{X} + 1$$

ii) We wish to check if the estimator $\hat{\theta} = \overline{X} + 1$ is unbiased, consistent and efficient.

Unbiasedness.

$$\mathrm{E}\left(\hat{\theta}\right) = \mathrm{E}\left(\overline{X}+1\right) =$$
$$= \mathrm{E}(\overline{X}) + 1 = \mathrm{E}(X) + 1 = (\theta - 1) + 1 = \theta$$

Thus, the estimator is unbiased.

Consistency. This is a consistent estimator because the two sufficient conditions hold. That is,

1) θ is an unbiased estimator and

2) $\lim_{n\to\infty}\left(\mathrm{Var}\left(\hat{\theta}\right)\right) = \lim_{n\to\infty}\frac{(\theta-1)^2}{n} = 0$, because:

$$\mathrm{Var}\left(\hat{\theta}\right) = \mathrm{Var}\left(\overline{X}+1\right) = \mathrm{Var}\left(\overline{X}\right) = \frac{\mathrm{Var}(X)}{n} = \frac{(\theta-1)^2}{n}$$

Efficiency. To be able to verify if this estimator is efficient, we compute the Cramer-Rao lower bound.

$$Lc = \frac{1}{nE\left[\frac{\partial \ln f(X,\theta)}{\partial\theta}\right]^2}$$

$$f(X;\theta) = \frac{1}{\theta-1}\ e^{-\frac{1}{\theta-1}X}$$

$$\ln f(X;\theta) = -\ln(\theta-1) - \frac{1}{\theta-1}X$$

$$\frac{\partial \ln f(X;\theta)}{\partial\theta} = -\frac{1}{\theta-1} + \frac{X}{(\theta-1)^2}$$

$$E\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2 = E\left[-\frac{1}{\theta-1} + \frac{X}{(\theta-1)^2}\right]^2 =$$

$$= E\left[\frac{1}{(\theta-1)^2}(X-(\theta-1))\right]^2 = \frac{1}{(\theta-1)^4}E[(X-(\theta-1))]^2 =$$

$$= \frac{1}{(\theta-1)^4}\ \mathrm{Var}(X) = \frac{1}{(\theta-1)^4}\ (\theta-1)^2 = \frac{1}{(\theta-1)^2}$$

$$Lc = \frac{(\theta-1)^2}{n}$$

The variance of the estimator coincides with the Cramer-Rao lower bound, thus implying that this is an efficient estimator.

Exercise C

This corresponds to a test of independence between two variables. The data provided by the sample is summarized as follows:

	Yes	No	No clear opinion	Total
18-30	130	40	30	200
31-50	85	40	25	150
> 50	75	35	40	150
Total	290	115	95	500

Therefore and to be able to carry out the requested test, we can build the table:

	n_{ij}	$\hat{p}_{ij} = \hat{p}_{i\cdot}\hat{p}_{\cdot j}$	$n\hat{p}_{ij}$	$\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$
18-30, Yes	130	$0.4 \times 0.58 = 0.232$	116	1.6897
18-30, No	40	$0.4 \times 0.23 = 0.092$	46	0.7826
18-30, No clear opinion	30	$0.4 \times 0.19 = 0.076$	38	1.6842
31-50, Yes	85	$0.3 \times 0.58 = 0.174$	87	0.0460
31-50, No	40	$0.3 \times 0.23 = 0.069$	34.5	0.8768
31-50, No clear opinion	25	$0.3 \times 0.19 = 0.057$	28.5	0.4298
> 50, Yes	75	$0.3 \times 0.58 = 0.174$	87	1.6552
> 50, No	35	$0.3 \times 0.23 = 0.069$	34.5	0.0072
> 50, No clear opinion	40	$0.3 \times 0.19 = 0.057$	28.5	4.6404
	500	1	500	$z = 11.8119$

The estimated probabilities $\hat{p}_{i\cdot}$ and $\hat{p}_{\cdot j}$ can be obtained from the information provided by the data in the sample, so that,

$$\hat{p}_{\bullet \text{Yes}} = \frac{290}{500} = 0.58 \qquad \hat{p}_{\bullet \text{No}} = \frac{115}{500} = 0.23 \qquad \hat{p}_{\bullet \text{No clear opinion}} = \frac{95}{500} = 0.19$$

$$\hat{p}_{18-30,\bullet} = \frac{200}{500} = 0.4 \qquad \hat{p}_{31-50,\bullet} = \frac{150}{500} = 0.3 \qquad \hat{p}_{>50,\bullet} = \frac{150}{500} = 0.3$$

Under the null hypothesis of independence, the test statistic $\sum_{i=1}^{k'} \sum_{j=1}^{k''} \frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \sim \chi^2_{(k'-1)(k''-1)}$, where k' and k'' are the number of classes in which each of the two variables under study has been divided.

In this case,

$$11.8119 > 9.49 = \chi^2_{(3-1)(3-1)\,,\,0.05},$$

so that, at the 5% significance level, the null hypothesis of independence is rejected. That is to say that we can confirm that clients' preferences for that product do change with age.