

## 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 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:**

In a factory manufacturing screws it is known that the distribution of their lengths (in millimeters) is  $N(20, \sigma^2 = 1)$ . It can be assumed that the distributions of the different screws are independent from each other. Screws with a length larger than 22 millimeters are considered not adequate for their sale (round up to two decimal places the probability that a screw is not adequate for its sale).

2. If the factory manufactures 20 screws, the probability that exactly one of them is not adequate for its sale is:

- (A) 0.392      (B) 0.020      (C) 0.332      (D) 0.668      (E) 0.272

3. If the factory manufactures 100 screws, the approximate probability that at most three of them are not adequate for their sale is:

- (A) 0.332      (B) 0.857      (C) 0.180      (D) 0.677      (E) 0.143

4. If the factory manufactures 1000 screws, the approximate probability that more than 22 of them are not adequate for their sale is:

- (A) 0.633      (B) 0.712      (C) 0.367      (D) 0.212      (E) 0.288

**Questions 5 and 6 refer to the following exercise:**

Let  $X$  be a r.v. having a binomial distribution with mean  $m = 3$  and variance  $\sigma^2 = 2.4$

5. The probability  $P(X \leq 6)$  is:

- (A) 0.9389      (B) 0.0430      (C) 0.9819      (D) 0.0611      (E) 0.6080

6. The probability  $P(X \geq 4)$  is:

- (A) 0.6482      (B) 0.3518      (C) 0.8358      (D) 0.1642      (E) 0.9389

**Questions 7 and 8 refer to the following exercise:**

Let  $X$  be a r.v. having a Poisson distribution. It is known that  $P(4) = \frac{5}{8}P(3)$ .

7. The mean of this r.v. is:

- (A) 2.5      (B) 1.25      (C) 10      (D) 5      (E) 0.625

8. The most likely value for this r.v. is:

- (A) 2      (B) 1      (C) 2.5      (D) 1.25      (E) 0

**Questions 9 to 11 refer to the following exercise:**

The number of cars that visits a given car washing facility each hour follows a Poisson distribution with mean  $m = 3.5$ . Distributions for the different hours are assumed to be independent from each other.

9. The probability that, in a one-hour period, exactly four cars visit the car washing facility is:  
 (A) 0.6288      (B) 0.4335      (C) 0.8153      (D) 0.2381      (E) 0.1888
10. The probability that, in a two-hour period, between 5 and 9 cars, including both, visit the car washing facility is:  
 (A) 0.7487      (B) 0.5561      (C) 0.5298      (D) 0.6575      (E) 0.9182
11. The approximate probability that, in a ten-hour period, at least 33 cars visit the car washing facility is:  
 (A) 0.3372      (B) 0.4681      (C) 0.5948      (D) 0.6628      (E) 0.5319

**Questions 12 to 14 refer to the following exercise:**

Let  $X_1, X_2, X_3$  and  $X_4$  be i.i.d. r.v. with a  $\gamma(a = 2, r = 3)$  distribution.

12. The mean and variance of each one of these random variables are, respectively:  
 (A)  $\frac{3}{2}, \frac{3}{4}$       (B)  $\frac{2}{3}, \frac{4}{3}$       (C)  $\frac{2}{3}, \frac{2}{9}$       (D)  $\frac{2}{3}, \frac{4}{9}$       (E)  $\frac{3}{2}, \frac{9}{4}$
13. The distribution of the r.v.  $Y = \frac{(X_1+X_2+X_3+X_4)}{4}$  is:  
 (A)  $\gamma(a = \frac{2}{4}, r = \frac{3}{4})$       (B)  $\gamma(a = 8, r = 12)$       (C)  $\gamma(a = 2, r = 12)$   
 (D)  $\gamma(a = \frac{2}{4}, r = 3)$       (E)  $\gamma(a = \frac{2}{4}, r = 12)$
14. The probability  $P(4X_1 > 10.6)$  is:  
 (A) 0.25      (B) 0.90      (C) 0.05      (D) 0.95      (E) 0.10

**Questions 15 to 17 refer to the following exercise:**

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

15. If we define the r.v.  $V_1 = \left(\frac{X^2}{4}\right) + Z$ , then the value of  $k$  such that  $P(5.58 < V_1 < k) = 0.85$  is:  
 (A) 13.7      (B) 17.3      (C) 24.7      (D) 7.58      (E) 19.7
16. If we define the r.v.  $V_2 = \frac{X}{\sqrt{Y}}$ , then  $P(-2.13 < V_2 < 1.53)$  is:  
 (A) 0.20      (B) 0.80      (C) 0.95      (D) 0.85      (E) 0.90
17. If we define the r.v.  $V_3 = \frac{5X^2}{2Z}$ , then  $P(V_3 \leq 0.0166)$  is:  
 (A) 0.05      (B) 0.95      (C) 0.01      (D) 0.10      (E) 0.90

**Questions 18 and 19 refer to the following exercise:**

Let  $X_1, \dots, X_n$  be a r.s. taken from from a population with probability mass function given by:

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

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

18. The method of moments estimate of  $\theta$  is:

- (A) 0.12                      (B) 0.32                      (C) 0.25                      (D) 0.20                      (E) 0.15

19. A different r.s. of the same size was taken, providing the following results: 0, 0, 0, 0, 0, 1, 1, 1, 2, 2. In this case, the maximum likelihood estimate of  $\theta$  is:

- (A) 0.20                      (B) 0.12                      (C) 0.32                      (D) 0.15                      (E) 0.25

**Questions 20 and 21 refer to the following exercise:**

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

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

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

20. The method of moment estimator of  $\theta$  is:

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

21. The maximum likelihood estimator of  $\theta$  is:

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

**Questions 22 to 24 refer to the following exercise:**

Let  $X$  be a r.v. with mean  $m = \theta + 1$  and variance  $\sigma^2 = \theta$ . In order to estimate the parameter  $\theta$ , a r.s. of size  $n$  has been taken, and the estimator  $\hat{\theta} = \bar{X} + 1$  is defined.

22. The bias of the estimator  $\hat{\theta}$  is:

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

23. The variance of the estimator  $\hat{\theta}$  is:

- (A)  $\frac{\theta}{n} + 1$                       (B)  $\frac{\theta}{n} + \frac{1}{n}$                       (C)  $\frac{\theta}{n}$                       (D)  $\theta$                       (E)  $\frac{\theta}{(n+1)}$

24. The mean square error of the estimator  $\hat{\theta}$  is:

- (A)  $\frac{\theta}{n}$                       (B)  $\frac{\theta}{n} + \frac{1}{n}$                       (C)  $\frac{\theta}{(n+1)}$                       (D)  $\frac{\theta}{n} + 4$                       (E)  $\frac{\theta}{n} + 1$

**Questions 25 to 27 refer to the following exercise:**

Let  $X$  be a r.v. having an exponential  $\exp(\lambda)$  distribution. We wish to test the null hypothesis  $H_0 : \lambda = 2$  against the alternative hypothesis  $H_1 : \lambda = 1$ . In order to do so, a random sample of size one,  $X_1$ , has been taken.

25. The form of the most powerful critical region for the test statistic  $X_1$  is:

- (A)  $[c_1, c_2]^C$                       (B)  $[c, \infty)$                       (C)  $[c_1, c_2]$                       (D)  $(0, c]$                       (E)  $(-\infty, 0]$

26. At the 5% significance level, and after taken the sample, we decide to reject the null hypothesis if:  
 (A)  $x_1 \leq 0.026$     (B)  $x_1 \in (0.372, 1.498)$     (C)  $x_1 \geq 1.844$     (D)  $x_1 \leq 0.372$     (E)  $x_1 \geq 1.498$
27. For the same significance level, the power of this test is:  
 (A) 0.3106    (B) 0.2236    (C) 0.1024    (D) 0.1582    (E) 0.0257

**Questions 28 to 30 refer to the following exercise:**

We wish to estimate the proportion of families that is subscribed to a given digital platform. In order to do so, a r.s. of 200 families has been taken, and the result indicated that 46 of them are subscribed to it.

28. The 95% confidence interval for the proportion of families that are subscribed to this digital platform is, approximately:

- (A) (0.1717, 0.2883)    (B) (0.2283, 0.2317)    (C) (0.1477, 0.3123)  
 (D) (0.1655, 0.2945)    (E) (0.1812, 0.2788)

29. If we wish to test the null hypothesis that this proportion is equal to 0.30 against the alternative hypothesis that it is different from that value, at the  $\alpha$  significance level, the decision would be to reject the null hypothesis if:

- (A)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \right| \geq t_{\frac{\alpha}{2}}$     (B)  $\frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \geq t_{\alpha}$     (C)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \right| \geq t_{\alpha}$   
 (D)  $\frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.23) \cdot (0.77)}{200}}} \geq t_{\frac{\alpha}{2}}$     (E)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.23) \cdot (0.77)}{200}}} \right| \geq t_{\frac{\alpha}{2}}$

30. If we wish to test the null hypothesis that this proportion is smaller than or equal than 0.30 against the alternative that it is larger than this value, at the  $\alpha$  significance level, the decision would be to reject the null hypothesis if:

- (A)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \right| \geq t_{\alpha}$     (B)  $\frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \geq t_{\alpha}$     (C)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.23) \cdot (0.77)}{200}}} \right| \geq t_{\frac{\alpha}{2}}$   
 (D)  $\frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.23) \cdot (0.77)}{200}}} \geq t_{\frac{\alpha}{2}}$     (E)  $\left| \frac{\frac{z}{n} - 0.30}{\sqrt{\frac{(0.30) \cdot (0.70)}{200}}} \right| \geq t_{\frac{\alpha}{2}}$

### EXERCISES (Time: 65 minutes)

A. (10 points, 20 minutes)

Let  $X$  be a r.v. having a normal  $N(m, \sigma^2 = 4)$  distribution. In order to estimate the mean of this distribution, a r.s. of size  $n$ ,  $X_1, \dots, X_n$ , has been taken and the following estimators are proposed:

$$\hat{m}_1 = \frac{3X_1 + X_2 + X_3 + \dots + X_{n-2} + X_{n-1} + 3X_n}{n+4} \qquad \hat{m}_2 = \frac{X_1 + X_2 + \dots + X_{n-1} + X_n}{n} + \frac{1}{n}$$

- i) Obtain the bias for these estimators. Are they unbiased? Provide all the relevant details to justify your answers.
- ii) Compute the variance for these estimators. Are they consistent? Provide all the relevant details to justify your answers.
- iii) It is known that the Cramer-Rao lower bound for a regular unbiased estimator of  $m$  is  $L_c = \frac{4}{n}$ . Are any of the proposed estimators efficient? Provide all the relevant details to justify your answers.

B. (10 points, 20 minutes)

Let  $X$  be a r.v. having a  $\gamma(\theta, 4)$  distribution. That is, with probability density function  $f(x) = \frac{\theta^4}{6} x^3 e^{-\theta x}$  for  $x > 0$ ,  $\theta > 0$ . In order to test the null hypothesis  $H_0 : \theta = \frac{1}{2}$  against the alternative hypothesis  $H_1 : \theta = 1$ , a r.s. of size one,  $X$ , is taken.

- i) Providing all relevant details, obtain the form of the critical region obtained from the likelihood ratio test for the test statistic  $X$ .
- ii) For a 5% significance level, obtain the specific critical region.
- iii) For  $X$ , we propose the alternative critical region given by the interval  $(2.73, 3.49)$ . Obtain the significance level for this new critical region and, without performing any additional computation of their corresponding powers, indicate which one of the two proposed critical regions is better.

C. (10 points, 25 minutes)

A firm devoted to recreational activities wishes to study if the mean monthly expense in leisure activities for families having children or not having them are different from each other. In order to do so, the firm takes two random samples: a sample of size 21 for families with children, providing a mean of 130 euros and a standard deviation of 20 euros; and a sample of size 41 for families without children, providing a mean of 141 euros and a standard deviation of 24 euros. It is known that the distributions for these two variables are normal and independent from each other.

- i) Obtain the 0.90 confidence interval for the ratio of the variances (i.e., variance of the families with children and that of families without children).
- ii) At the 10% significance level, carry out the test of the null hypothesis that the variances are equal against the alternative hypothesis that they are different.
- iii) Under the assumption that the variances are equal and at the 5% significance level, test the null hypothesis that the mean expenses are equal against the alternative hypothesis that they are different.
- iv) Given that there is the belief that families without children are the ones having a larger mean expense, at the 5% significance level and also under the assumption that the variances are equal, test the null hypothesis that the mean expenses are equal against the alternative hypothesis that families without children have a larger mean expense than those with children.

**SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)**

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

## SOLUTIONS TO EXERCISES

### Exercise A

i) The bias for an estimator  $\hat{\theta}$  for the parameter  $\theta$  is defined as  $\text{Bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$ .

To compute the bias for each of the proposed estimators, we obtain their mathematical expectations.

$$\begin{aligned} \text{E}(\hat{m}_1) &= \text{E}\left(\frac{3X_1 + X_2 + X_3 + \cdots + X_{n-2} + X_{n-1} + 3X_n}{n+4}\right) = \left[\frac{3\text{E}(X_1) + \text{E}(X_2) + \cdots + \text{E}(X_{n-1}) + 3\text{E}(X_n)}{n+4}\right] = \\ &= \frac{6m + (n-2)m}{n+4} = \frac{(n+4)m}{n+4} = m \\ \text{E}(\hat{m}_2) &= \text{E}\left(\frac{X_1 + X_2 + \cdots + X_{n-1} + X_n}{n} + \frac{1}{n}\right) = \left(\frac{\text{E}(X_1) + \cdots + \text{E}(X_n)}{n} + \frac{1}{n}\right) = \frac{nm}{n} + \frac{1}{n} = m + \frac{1}{n} \end{aligned}$$

Therefore, their corresponding biases are:

$$\text{Bias}(\hat{m}_1) = \text{E}(\hat{m}_1) - m = m - m = 0$$

$$\text{Bias}(\hat{m}_2) = \text{E}(\hat{m}_2) - m = m + \frac{1}{n} - m = \frac{1}{n}$$

Thus, given that its bias is equal to zero, the estimator  $\hat{m}_1$  is unbiased, and the estimator  $\hat{m}_2$  is biased, with bias equal to  $\frac{1}{n}$ .

ii) We now compute the proposed estimators' variances.

$$\begin{aligned} \text{Var}(\hat{m}_1) &= \text{Var}\left(\frac{3X_1 + X_2 + \cdots + X_{n-1} + 3X_n}{n+4}\right) = \\ &= \frac{1}{(n+4)^2} [9\text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_{n-1}) + 9\text{Var}(X_n)] = \\ &= \frac{18\sigma^2 + (n-2)\sigma^2}{(n+4)^2} = \frac{(n+16)\sigma^2}{(n+4)^2} = \frac{4(n+16)}{(n+4)^2} \\ \text{Var}(\hat{m}_2) &= \text{Var}\left(\frac{X_1 + \cdots + X_n}{n} + \frac{1}{n}\right) = \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \\ &= \frac{1}{n^2} [\text{Var}(X_1) + \cdots + \text{Var}(X_n)] = \frac{n}{n^2} \text{Var}(X) = \frac{\sigma^2}{n} = \frac{4}{n} \end{aligned}$$

Are these estimators consistent?

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

a. It is unbiased.

$$\text{b. } \lim_{n \rightarrow \infty} \text{Var}(\hat{m}_1) = \lim_{n \rightarrow \infty} \left[\frac{4(n+16)}{(n+4)^2}\right] = 0$$

The estimator  $\hat{m}_2$  is also consistent because the two sufficient conditions hold:

a. It is asymptotically unbiased because:  $\lim_{n \rightarrow \infty} \text{E}(\hat{m}_2) = \lim_{n \rightarrow \infty} \text{E}\left(m + \frac{1}{n}\right) = m$

$$\text{b. } \lim_{n \rightarrow \infty} \text{Var}(\hat{m}_2) = \lim_{n \rightarrow \infty} \frac{4}{n} = 0$$

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



iii) A regular unbiased estimator is efficient if its variance is equal to the Cramer-Rao lower bound.

The estimator  $\hat{m}_1$  is not efficient because  $\text{Var}(\hat{m}_1) = \frac{4(n+16)}{(n+4)^2} \neq \frac{4}{n} = L_c$ .

The estimator  $\hat{m}_2$  is not efficient because it is not unbiased.

### Exercise B

$X \in \gamma(\theta, 4)$ ; that is,  $f(x) = \frac{\theta^4}{6} x^3 e^{-\theta x}$  for  $x > 0$ ,  $\theta > 0$ . For a sample of size one,  $X$ , we wish to test, using the likelihood ratio test, the null hypothesis  $H_0 : \theta = \frac{1}{2}$  against the alternative hypothesis  $H_1 : \theta = 1$ .

i) We carry out the likelihood ratio test. The corresponding likelihood functions under the null and alternative hypotheses are as follows:

$$L(x; \theta_0) = \frac{\left(\frac{1}{2}\right)^4}{6} x^3 e^{-\frac{x}{2}}, \quad \text{and} \quad L(x; \theta_1) = \frac{1^4}{6} x^3 e^{-x},$$

Therefore,

$$\frac{L(x; \theta_0)}{L(x; \theta_1)} = \frac{\left(\frac{1}{2}\right)^4 x^3 e^{-\frac{x}{2}}}{\frac{1^4}{6} x^3 e^{-x}} = \left(\frac{1}{2}\right)^4 e^{\frac{x}{2}} \leq k, \quad k > 0$$

$$e^{\frac{x}{2}} \leq k' \implies \frac{x}{2} \leq k''$$

$$x \leq C$$

That is, for the test statistics  $X$ , the most powerful critical region is of the form  $\text{CR} = (0, C]$ .

ii) For an  $\alpha = 0.05$  significance level, and taking into account that, under the null hypothesis, we have that:

$$X \in \gamma\left(\frac{1}{2}, 4\right) \equiv \chi_{8|}^2,$$

we must have that:

$$\alpha = 0.05 = P[X \in \text{CR} | H_0] = P[X \leq C | H_0] \implies C = \chi_{8|0.95}^2 = 2.73$$

That is, we reject the null hypothesis if  $x \leq 2.73$ .

iii) We start by computing the significance level for the new critical region (2.73, 3.49).

Given that, under the null hypothesis,

$$X \in \gamma\left(\frac{1}{2}, 4\right) \equiv \chi_{8|}^2,$$

and that  $2.73 = \chi_{8|0.95}^2$  and  $3.49 = \chi_{8|0.90}^2$ ,

$$\alpha = P[X \in \text{CR} | H_0] = P[X \in (2.73, 3.49) | H_0] = 0.95 - 0.90 = 0.05$$

iv) Given that both critical regions have the same significance level, we know that the power for the critical region in the second item is larger than that of the proposed new critical region because the Neyman Pearson Theorem states that the likelihood ratio test provides the most powerful critical region. Therefore, the first test is better than the second one.

### Exercise C

$X$ : Monthly expense (families with children)  $X \in N(m_1, \sigma_1^2)$ ,  $n_1 = 21$ ,  $\bar{x} = 130$ ,  $s_1 = 20$   
 $Y$ : Monthly expense (families without children)  $Y \in N(m_2, \sigma_2^2)$ ,  $n_2 = 41$ ,  $\bar{y} = 141$ ,  $s_2 = 24$

i) We obtain the confidence interval for the ratio of variances,  $\frac{\sigma_1^2}{\sigma_2^2}$ .

$$CI_{1-\alpha} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = \left( \frac{n_1 s_1^2 (n_2 - 1)}{n_2 s_2^2 (n_1 - 1)} \frac{1}{F_{n_1-1, n_2-1 | \alpha/2}}, \frac{n_1 s_1^2 (n_2 - 1)}{n_2 s_2^2 (n_1 - 1)} \frac{1}{F_{n_1-1, n_2-1 | 1-\alpha/2}} \right)$$

In our case, we have that:

$$IC_{90\%} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = \left( \frac{(21)(20^2)(40)}{(41)(24^2)(20)} \frac{1}{F_{20,40|0.05}}, \frac{(21)(20^2)(40)}{(41)(24^2)(20)} \frac{1}{F_{20,40|0.95}} \right)$$

We now have that:

$$F_{20,40|0.05} = 1.84$$
$$F_{20,40|0.95} = \frac{1}{F_{40,20|0.05}} = \frac{1}{1.99}$$

Therefore,

$$CI_{90\%} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = (0.3866, 1.4156)$$

ii) We wish to test the null hypothesis  $H_0 : \sigma_1^2 = \sigma_2^2$  against the alternative hypothesis  $H_1 : \sigma_1^2 \neq \sigma_2^2$ , which is equivalent to testing the null hypothesis  $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$  against the alternative hypothesis  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ .

Given that it is a bilateral test and that the significance level  $\alpha = 10\%$  is complementary to the confidence level  $1 - \alpha = 0.90$ , we can use the confidence interval to perform this test.

More specifically, as the value of the ratio under the null hypothesis  $H_0$ ,  $\frac{\sigma_1^2}{\sigma_2^2} = 1 \in CI_{90\%} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = (0.3866, 1.4156)$ , at the 10% significance level, we do not reject the null hypothesis that both variances are equal.

iii) At the 5% significance level, we wish to test the null hypothesis  $H_0 : m_1 = m_2$  against the alternative hypothesis  $H_1 : m_1 \neq m_2$ .

This is equivalent to testing the null hypothesis  $H_0 : m_1 - m_2 = 0$  against the alternative hypothesis  $H_1 : m_1 - m_2 \neq 0$ .

The test statistic for this specific test of hypotheses is:

$$\frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}}$$

which, under the null hypothesis  $H_0$ , follows a  $t_{n_1+n_2-2}$  distribution.

As it is a bilateral test, the decision rule indicates that, at the  $\alpha = 0.05$  significance level, we reject the null hypothesis if:

$$\left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} \right| \geq t_{n_1 + n_2 - 2} | \frac{0.05}{2}$$

In our case, we have that:

$$\left| \frac{(130 - 141)}{\sqrt{\frac{1}{21} + \frac{1}{42}} \sqrt{\frac{(21)(20^2) + (41)(24^2)}{21 + 41 - 2}}} \right| = 1.7745 \not\geq 2 = t_{60} | \frac{0.05}{2}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis that the mean expenses are equal.

iv) At the 5% significance level, we wish to test the null hypothesis  $H_0 : m_1 = m_2$  against the alternative hypothesis  $H_1 : m_1 < m_2$ .

This is equivalent to testing the null hypothesis  $H_0 : m_1 - m_2 = 0$  against the alternative hypothesis  $H_1 : m_1 - m_2 < 0$

The test statistic for this specific test of hypotheses is the same as the one used in the previous item:

$$\frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}},$$

which, under  $H_0$  follows a  $t_{n_1 + n_2 - 2}$  distribution.

This corresponds to a unilateral test and, therefore, the decision rule states that, at the  $\alpha = 0.05$  significance level, we should reject the null hypothesis if:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} \leq -t_{n_1 + n_2 - 2} | 0.05$$

In our case,

$$\frac{(130 - 141)}{\sqrt{\frac{1}{21} + \frac{1}{42}} \sqrt{\frac{(21)(20^2) + (41)(24^2)}{21 + 41 - 2}}} = -1.7745 \leq -1.67 = t_{60} | 0.05$$

Therefore, at the 5% significance level, we reject the null hypothesis that the mean expenses are equal and conclude that the mean expense for families without children is larger than that of families with children.