STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2023-24
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
Second Call. July 4, 2024

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 5 refer to the following exercise:

The weight, in grams, for a pack of cookies of a given brand follows a $N(500, \sigma^2 = 9)$ distribution. We assume that the distributions of the different packs of cookies are independent from one another. Packs that have a weight lower than 496.1 grams are not adequate for their sale (round up to one decimal place the probability that a given pack of cookies is not adequate for its sale).

2. If 8 packs are randomly selected, the probability that exactly one of them is not adequate for its sale is:

   (A) 0.0478 (B) 0.1976 (C) 0.4783 (D) 0.2793 (E) 0.3826

3. If 20 packs are randomly selected, the probability that more than two of them are not adequate for their sale is:

   (A) 0.3917 (B) 0.6083 (C) 0.2852 (D) 0.6769 (E) 0.3231

4. If 200 packs are randomly selected, the approximate probability that more than 28 of them are not adequate for their sale is:

   (A) 0.977 (B) 0.039 (C) 0.281 (D) 0.719 (E) 0.023

5. In the latter case (i.e., 200 packs), what is the expected number of packs that are not adequate for their sale?

   (A) 20 (B) 50 (C) 10 (D) 40 (E) 30

Questions 6 and 7 refer to the following exercise:

Let X be a r.v. having a binomial $b(p = 0.4, n)$ distribution with variance $\sigma^2 = 2.4$.

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

   (A) 0.0510 (B) 0.6331 (C) 0.2508 (D) 0.3823 (E) 0.7492

7. The probability $P(3 \leq X \leq 7)$ is:

   (A) 0.6054 (B) 0.4159 (C) 0.8204 (D) 0.7841 (E) 0.2375

Questions 8 to 10 refer to the following exercise:

The number of clients arriving each hour at a given bank branch follows a Poisson distribution such that $P(6) = P(7)$. We assume independence among the distributions for the different hours.

8. The probability that, in a one-hour period, more than four clients arrive at the store is:

   (A) 0.918 (B) 0.827 (C) 0.629 (D) 0.082 (E) 0.173

9. The probability that, in a one-hour period, exactly 5 clients arrive at the store is:

   (A) 0.301 (B) 0.616 (C) 0.699 (D) 0.128 (E) 0.872

10. The probability that, in a six-hour period, more than 28 clients arrive at the store is, approximately:

(A) 0.1730 (B) 0.2851 (C) 0.0188 (D) 0.7149 (E) 0.9812

Questions 11 and 12 refer to the following exercise:

Let X_1 and X_2 be two independent r.v. having each a $\gamma(a = 3, r = 1)$ distribution.

11. The value of k such that $P(X_1 \geq k) = 0.05$ holds is:

(A) 1.365 (B) 0.017 (C) 0.383 (D) 2.321 (E) 0.999

12. If we define the r.v. $Y = 6X_1 + 6X_2$, then $P(Y \leq 5.39)$ is:

(A) 0.90 (B) 0.10 (C) 0.75 (D) 0.05 (E) 0.95

Questions 13 to 15 refer to the following exercise:

Let X, Y and Z be three independent r.v. having the following distributions: $X \in N(0, \sigma^2 = 1)$, $Y \in \chi^2_{10}$ and $Z \in \gamma(\frac{1}{2}, 3)$.

13. If we define the r.v. $V_1 = Y + Z$, the value of k such that $P(k < V_1 < 19.4) = 0.5$ holds is:

(A) 6.91 (B) 11.9 (C) 7.96 (D) 15.3 (E) 9.31

14. If we define the r.v. $V_2 = \dfrac{\sqrt{10}X}{\sqrt{Y}}$, the probability $P(V_2 \leq -1.37)$ is:

(A) 0.20 (B) 0.80 (C) 0.95 (D) 0.10 (E) 0.90

15. If we define the r.v. $V_3 = \dfrac{10(Z + X^2)}{7Y}$, the value of k such that $P(V_3 \geq k) = 0.95$ holds is:

(A) 0.318 (B) 3.14 (C) 2.78 (D) 0.275 (E) 3.64

Questions 16 and 17 refer to the following exercise:

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

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

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

16. The method of moments estimate of θ is:

(A) 1.1 (B) 0.8 (C) 0.5 (D) 0.6 (E) 0.9

17. The maximum likelihood estimate of θ is:

(A) 0.9 (B) 1.1 (C) 0.8 (D) 0.5 (E) 0.6

Questions 18 and 19 refer to the following exercise:

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

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

It is known that the mean of this r.v. is $m = \frac{2\theta}{3}$. In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$ has been taken.

18. The method of moments estimator of θ is:

(A) $\overline{X}$ (B) $\frac{2\overline{X}}{3}$ (C) $\max\{X_i\}$ (D) $\min\{X_i\}$ (E) $\frac{3\overline{X}}{2}$

19. The maximum likelihood estimator of θ is:

(A) $\frac{3\overline{X}}{2}$ (B) $\frac{2\overline{X}}{3}$ (C) $\overline{X}$ (D) $\min\{X_i\}$ (E) $\max\{X_i\}$

Questions 20 to 23 refer to the following exercise:

Let X be a r.v. having a Poisson distribution. In order to estimate the parameter λ, a r.s. of size n has been taken and the estimator $\hat{\lambda} = \frac{X_1+X_2+\cdots+X_n}{n+1}$ is defined.

20. The bias of the estimator $\hat{\lambda}$ is:

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

21. The variance of the estimator $\hat{\lambda}$ is:

(A) $\frac{\lambda^2}{n+1}$ (B) $\frac{n\lambda}{n+1}$ (C) $\frac{\lambda^2}{n}$ (D) $\frac{n\lambda}{(n+1)^2}$ (E) $\frac{\lambda}{n+1}$

22. The mean square error of the estimator $\hat{\lambda}$ is:

(A) $\frac{\lambda}{n+1}$ (B) $\frac{\lambda+\lambda^2}{n+1}$ (C) $\frac{(n+1)\lambda+\lambda^2}{(n+1)^2}$ (D) $\frac{\lambda^2}{n+1}$ (E) $\frac{n\lambda+\lambda^2}{(n+1)^2}$

23. For this estimator, we can state that it is:

(A) Unbiased and not consistent (B) Unbiased and consistent (C) Biased and consistent (D) Biased and not consistent (E) Not regular

Questions 24 to 27 refer to the following exercise:

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

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

We wish to test the null hypothesis $H_0 : \theta = 0$ against the alternative hypothesis $H_1 : \theta = 1$. In order to do so, a random sample of size $n = 1$ has been taken and we reject the null hypothesis H_0 if $x \leq 0.1$.

24. The significance level for this test is:

(A) 0.10 (B) 0.15 (C) 0.01 (D) 0.09 (E) 0.05

25. The probability of type II error for this test is:

(A) 0.99 (B) 0.01 (C) 0.90 (D) 0.10 (E) 0.05

26. At the $\alpha = 0.10\%$ significance level, the most powerful critical region for this test is:

(A) $[0.9, 1)$ (B) $(0, 0.1]$ (C) $(0, 0.01]$ (D) $(0, 0.05]$ (E) $[0.99, 1)$

27. For the above critical region, the power for the test is:

(A) 0.19 (B) 0.90 (C) 0.10 (D) 0.99 (E) 0.95

Questions 28 to 30 refer to the following exercise:

We wish to estimate the mean monthly urban public transportation expense, in euros, for residents in a given city. In order to do so, a r.s. of 31 people has been taken, proving a mean equal to 30 euros and a variance equal to 9 euros^2. We assume that the distribution of the above monthly expense is normal.

28. The 95% confidence interval for the mean expense is, approximately:

(A) $(29.791, 30.209)$ (B) $(28.944, 31.056)$ (C) $(28.883, 31.117)$ (D) $(26.566, 33.434)$ (E) $(26.832, 33.168)$

29. At the α significance level, if we wish to test the null hypothesis that the mean expense is of at least 33 euros against the alternative hypothesis that it is smaller than this amount, the decision will be to reject the null hypothesis if:

(A) $\dfrac{\overline{x} - 33}{\frac{3}{\sqrt{31}}} \leq t_\alpha$ (B) $\dfrac{\overline{x} - 33}{\frac{3}{\sqrt{30}}} \geq t_{\overline{30}|\alpha}$ (C) $\left|\dfrac{\overline{x} - 33}{\frac{3}{\sqrt{31}}}\right| \geq t_\alpha$ (D) $\dfrac{\overline{x} - 33}{\frac{3}{\sqrt{30}}} \leq -t_{\overline{30}|\alpha}$ (E) $\dfrac{\overline{x} - 33}{\frac{3}{\sqrt{31}}} \leq -t_\alpha$

30. The 95% confidence interval for the **standard deviation** of this specific expense is:

(A) $(0.355, 0.994)$ (B) $(2.803, 3.185)$ (C) $(1.979, 5.536)$ (D) $(2.436, 4.075)$ (E) $(1.855, 4.145)$

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

The number of car tow trucks requested, in a one-hour period, to a given road side assistance company follows a Poisson distribution with mean equal to 2. We assume that each tow truck can only perform its service once per hour and that there is independence among the distributions for the different hours.

i) If, in a given one-hour period, the road side assistance company has 5 tow trucks available, what is the probability that it can provide all of its requested services?

ii) If, in a one-hour period, the company wishes to be able to perform all of its requested services with probability of at least 92%, what is the minimum number of tow trucks the company should have available during this one-hour period?

iii) Recently, a new road side assistance company in the same area has been opened and the former company believes that this fact will have a clear effect on the number of two trucks requested to this company. To be able to arrive to a conclusion about their belief, they wish to test the null hypothesis that the mean number of tow trucks requested in a one-hour period remains the same (i.e., it is equal to 2) against the alternative hypothesis that it has become 1. In order to test these hypotheses, a sample of 4 hours of service has been taken. At the 5% significance level, obtain, **providing all relevant details**, the most powerful test for this specific test of hypotheses. Moreover, if it has been observed that, during this four-hour period under study, the company has performed 5 services, what will be the decision of the test?

B. (10 points, 25 minutes)

Let X be a r.v. having an exponential distribution with mean equal to θ. In order to be able to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$ has been taken and the following estimators are proposed:

$$\hat{\theta}_1 = \frac{X_1+X_2+\cdots+X_n}{n+2} \qquad \hat{\theta}_2 = \frac{2X_1+X_2+X_3\cdots+X_{n-1}+2X_n}{n+2}$$

i) Obtain the bias for both estimators. Are they unbiased? Provide the required relevant details to justify your answer to this item.

ii) Compute the variance for both estimators. Are they consistent? Provide the required relevant details to justify your answer to this item.

iii) Compute the Cramer-Rao lower bound for a regular unbiased estimator of θ. Are any of the proposed estimators above efficient? **Remark:** The Cramer-Rao lower bound for a regular unbiased estimator of θ is: $L_c = \dfrac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial\theta}\right]^2} = \dfrac{1}{-nE\left[\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right]}$

C. (10 points, 25 minutes)

The Well-Being Department in a given City Hall has put forward a campaign to promote regular physical exercise among people older than 70 years old.

In order to be able to assess if the campaign has been effective, a random sample of people with those characteristics has been taken before and after the campaign, where people were asked about the number of hours devoted to physical exercise during a given week. The first sample included 100 people, providing a mean of 5.3 weekly hours, whereas the second one included 200 people, providing a mean of 5.9 weekly hours. We assume that the distributions of the time devoted to physical exercise before and after the campaign are normal and independent from each other, having both a common variance $\sigma^2 = 4$.

i) Obtain the 0.95 confidence interval for the mean number of hours devoted to physical exercise **before** the campaign.

ii) At the 5% significance level, test the null hypothesis that the mean number of hours has remained the same against the alternative hypothesis that the mean has increased after the campaign. What is the final conclusion in relation to the objective the campaign had?

iii) If the objective of the campaign was to have a mean of at least 6 hours per week, at the 5% significance level, test the null hypothesis that the mean **after the campaign** is at least 6 hours against the alternative hypothesis that it is smaller than such an amount. What would be the test decision about the achievement or not of the proposed objective?

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

i) We have to compute the probability that the number of requested tow trucks is smaller than or equal to 5 because, if this were not the case, the company would no be able to provide the requested services. Let us denote by X the number of tow trucks requested in a one-hour period, and it is known that $X \in \mathcal{P}(\lambda = 2)$, so that:

$$P(X \leq 5) = F_X(5) = 0.9834$$

ii) We wish to be able to obtain the number of available tow trucks, k, that the company should have to be able to satisfactorily fulfill the requested service demand with a probability of at least 92%.

The company will be able to fulfill the requested service demand if $X \leq k$ holds. Therefore,

$$P(X \leq k) \geq 0.92 \Longrightarrow F_X(k) \geq 0.92$$

If we look for this value in the tables, it would be clear that the smallest value of k that satisfies this condition is $k = 4$.

iii) We wish to test the null hypothesis $H_0 : \lambda = 2$ against the alternative hypothesis $H_1 : \lambda = 1$.

In order to be able to obtain the most powerful test for this test of hypotheses we use the Neyman-Pearson Theorem. The likelihood functions under the null and alternative hypotheses will be respectively given by:

$$L(\vec{x};\lambda_0) = L(\vec{x};\lambda = 2) = \frac{\left[e^{-2n}\ 2^{\sum_{i=1}^n x_i}\right]}{\prod_{i=1}^n x_i!} \quad \text{y} \quad L(\vec{x};\lambda_1) = L(\vec{x};\lambda = 1) = \frac{\left[e^{-n}\ 1^{\sum_{i=1}^n x_i}\right]}{\prod_{i=1}^n x_i!}$$

Therefore, we apply the Neyman-Pearson Theorem, so that we have that:

$$\frac{L(\vec{x};\lambda_0)}{L(\vec{x};\lambda_1)} = \frac{\left[e^{-2n}\ 2^{\sum_{i=1}^n x_i}\right] / \prod_{i=1}^n x_i!}{\left[e^{-n}\ 1^{\sum_{i=1}^n x_i}\right] / \prod_{i=1}^n x_i!} \leq K,\ K > 0$$

$$\Longrightarrow e^{-n}\ 2^{\sum_{i=1}^n x_i} \leq K \Longrightarrow 2^{\sum_{i=1}^n x_i} \leq K_1 \Longrightarrow \left(\sum_{i=1}^n x_i\right) \ln 2 \leq K_2 \Longrightarrow \sum_{i=1}^n x_i \leq C$$

In our case, as $n = 4$, the decision rule will be to reject the null hypothesis if $\sum_{i=1}^4 X_i \leq C$.

At the $\alpha = 0.05$ significance level, and taking into account that, under the null hypothesis, $Z = \sum_{i=1}^4 X_i \in \mathcal{P}(8)$, we have that:

$$\alpha = 0.05 \geq P[Z \leq C|H_0] = P[Z \leq C|Z \in \mathcal{P}(8)] = F_Z(C)$$

$$\Longrightarrow F_Z(C) \leq 0.05 \Longrightarrow C = 3$$

That is, we reject the null hypothesis if $Z = \sum_{i=1}^4 X_i \leq 3$.

Given that the sample provided a total of 5 requests for service during the four-hour period under study, at the 5% significance level, the decision will be not to reject the null hypothesis.

Exercise B

i) The bias of an estimator $\hat{\theta}$ for the parameter θ is defined as$\mathrm{b}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$.

To obtain the bias for the proposed estimators, we compute their mathematical expectations:

$$\mathrm{E}(\hat{\theta}_1) = \mathrm{E}\left[\frac{X_1 + X_2 + \cdots + X_{n-1} + X_n}{(n+2)}\right] = \left[\frac{\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_{n-1}) + \mathrm{E}(X_n)}{(n+2)}\right] = \frac{n}{n+2}\theta$$

$$\mathrm{E}(\hat{\theta}_2) = \mathrm{E}\left[\frac{2X_1 + X_2 + X_3 + \cdots + X_{n-2} + X_{n-1} + 2X_n}{(n+2)}\right] = \left[\frac{2\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_{n-1}) + 2\mathrm{E}(X_n)}{(n+2)}\right] =$$

$$= \frac{4\theta + (n-2)\theta}{(n+2)} = \frac{(n+2)\theta}{(n+2)} = \theta$$

Therefore, the corresponding biases are:

$$\mathrm{b}(\hat{\theta}_1) = \mathrm{E}(\hat{\theta}_1) - \theta = \frac{n\theta}{(n+2)} - \theta = \frac{-2\theta}{(n+2)}$$

$$\mathrm{b}(\hat{\theta}_2) = \mathrm{E}(\hat{\theta}_2) - \theta = \theta - \theta = 0$$

Thus, $\hat{\theta}_1$ is a biased estimator and $\hat{\theta}_2$ is an unbiased one.

ii) We now compute the variance for the proposed estimators. Let us recall that the variance of an exponential distribution with mean $m = \theta$ is $\sigma^2 = \theta^2$.

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{(n+2)}\right] = \frac{1}{(n+2)^2}\left[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\right] = \frac{n}{(n+2)^2}\mathrm{Var}(X) = \frac{n\theta^2}{(n+2)^2}$$

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\left[\frac{2X_1 + X_2 + \cdots + X_{n-1} + 2X_n}{(n+2)}\right] =$$

$$= \frac{1}{(n+2)^2}\left[4\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_{n-1}) + 4\mathrm{Var}(X_n)\right] = \frac{8\theta^2 + (n-2)\theta^2}{(n+2)^2} = \frac{(n+6)\theta^2}{(n+2)^2}$$

Are they consistent estimators?

First of all, $\hat{\theta}_1$ is consistent because the required sufficient conditions are satisfied:

a. It is asymptotically unbiased because $\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_1) = \lim_{n\to\infty}\left(\frac{n\theta}{n+2}\right) = \theta$

b. $\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_1) = \lim_{n\to\infty} \frac{n\theta^2}{(n+2)^2} = 0$

With regard to $\hat{\theta}_2$, it is consistent because it also satisfies the required sufficient conditions:

a. Es insesgado

b. $\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_2) = \lim_{n\to\infty}\left[\frac{(n+6)\theta^2}{(n+2)^2}\right] = 0$

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

iii) We start by computing the Cramer-Rao lower bound $L_c = \dfrac{1}{n\text{E}\left[\frac{\partial \ln(f(x;\theta))}{\partial\theta}\right]^2}$

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

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

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

$$\text{E}\left[\frac{\partial \ln(f(x;\theta))}{\partial\theta}\right]^2 = \text{E}\left[\frac{X}{\theta^2} - \frac{1}{\theta}\right]^2 = \text{E}\left[\frac{1}{\theta^2}(X-\theta)\right]^2 = \frac{1}{\theta^4}\text{E}(X-\theta)^2 = \frac{1}{\theta^4}\text{Var}(X) = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2}$$

If we replace the expectation in the Cramer-Rao lower bound, we have that:

$$L_c = \frac{1}{n\text{E}\left[\frac{\partial \ln(f(x;\theta))}{\partial\theta}\right]^2} = \frac{1}{n\left(\frac{1}{\theta^2}\right)} = \frac{\theta^2}{n}$$

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

Therefore, $\hat{\theta}_1$ is not efficient because it is not unbiased.

With regard to $\hat{\theta}_2$ it is not efficient because $\text{Var}(\hat{\theta}_2) = \dfrac{(n+6)\theta^2}{(n+2)^2} \neq \dfrac{\theta^2}{n} = L_c$.

Exercise C

X: Weekly hour of exercise before the campaign. $X \in N(m_1, \sigma_1^2 = 4)$, $n_1 = 100$, $\overline{x} = 5.3$

Y: Weekly hours of exercise after the campaign. $Y \in N(m_2, \sigma_2^2 = 4)$, $n_2 = 200$, $\overline{y} = 5.9$

i) We obtain the 0.95 confidence interval for the mean number of hours of exercise people under study devoted before the campaign (m_1).

In this specific case, the distribution is normal with known variance. Therefore, the $(1-\alpha)$ confidence interval for m_1 is:

$$\text{CI}_{1-\alpha}(m_1) = \left(\bar{x} \pm t_{\frac{\alpha}{2}}\frac{\sigma_1}{\sqrt{n}}\right)$$

$$\text{CI}_{0.95}(m_1) = \left(5.3 \pm t_{0.025}\frac{2}{\sqrt{100}}\right) = \left(5.3 \pm 1.96\frac{2}{\sqrt{100}}\right) = (5.3 \pm 0.392) = (4.908, 5.692)$$

ii) 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 he null hypothesis $H_0 : m_1 - m_2 = 0$ against the alternative hypothesis $H_1 : m_1 - m_2 < 0$.

Under H_0, $$\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \in N(0, 1).$$

As it corresponds to a left-sided unilateral test, at the α significance level, the decision rule will be to reject H_0 if:

$$\frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -t_\alpha$$

In this case, as $\alpha = 0.05$,

$$\frac{5.3 - 5.9}{\sqrt{\frac{4}{100} + \frac{4}{200}}} = -2.45 \ \leq \ -1.64 = -t_{0.05}$$

Therefore, at the 5% significance level, we reject the null hypothesis. The conclusion is that the campaign has been clearly effective.

iii) We have to test the null hypothesis: $H_0 : m_2 \geq 6$ against the alternative hypothesis: $H_1 : m_2 < 6$.

Under H_0, $\dfrac{\bar{Y} - 6}{\sigma_2/\sqrt{n_2}} \in N(0, 1)$.

As it corresponds to a left-sided unilateral hypothesis, at the α significance level, the decision rule will be to reject H_0 if: $\dfrac{\bar{y} - 6}{\sigma_2/\sqrt{n_2}} \leq -t_\alpha$.

In this case, as $\alpha = 0.05$,

$$\frac{5.9 - 6}{2/\sqrt{200}} = -0.707 \ \not\leq \ -1.64 = -t_{0.05}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis. That is, the conclusion is that the campaign has been clearly effective.