STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year   Academic Year 2021-22
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
Second Call. June 29, 2022

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:

The distribution of the daily family water consumption (in liters) in a given community follows a $N(680, \sigma^2 = 2500)$ distribution. We assume independence between the consumption for different families. We consider that the family consumption is excessive if it is greater than 762 liters (you need to round up the probability that a family has an excessive water consumption up to two decimal places).

2. If in a given day six families are randomly selected, the probability that only one of them has an excessive water consumption is:

   (A) 0.300 (B) 0.039 (C) 0.354 (D) 0.059 (E) 0.232

3. If in a given day a group of one hundred families is randomly selected, how many of them are expected to have an excessive water consumption?

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

4. If in a given day a group of one hundred families is randomly selected, what is the approximate probability that at least 9 of them have an excessive water consumption?

   (A) 0.968 (B) 0.986 (C) 0.032 (D) 0.932 (E) 0.068

Questions 5 to 8 refer to the following exercise:

The yearly number of accidents a car driver who uses his/her car for approximately one hour per day follows a Poisson distribution. It is known that $P(2) = P(1)$, and that the distributions of the number of accidents for different years are independent from each other.

5. The mean and variance of the distribution of the yearly number of accidents the car driver has are, respectively:

   (A) 2 y 4 (B) 1 y 1 (C) 1 y 2 (D) 2 y 2 (E) 1.5 y 1.5

6. What is the probability that, in a given year, the car driver has at most 3 accidents?

   (A) 0.143 (B) 0.667 (C) 0.537 (D) 0.857 (E) 0.323

7. If the car driver uses his/her car, with the same daily frequency, for a four-year period, what is the probability that, during that period, the car driver has more than 7 accidents?

   (A) 0.407 (B) 0.547 (C) 0.687 (D) 0.313 (E) 0.453

8. If the car driver uses his/her car under the same conditions for a ten-year period, what would be the approximate probability that, during that period, the car driver has fewer than 18 accidents?

   (A) 0.71 (B) 0.42 (C) 0.29 (D) 0.85 (E) 0.54

Questions 9 and 10 refer to the following exercise:

Let X_1 and X_2 be two independent r.v. having the same $\gamma(0.5, 1)$ probability distribution.

9. The probability $P(X_1 + X_2 \geq 1.92)$ is:

   (A) 0.38 (B) 0.25 (C) 0.62 (D) 0.75 (E) 0.15

10. The probability $P(2X_1 \geq 1.92)$ is:

    (A) 0.62 (B) 0.25 (C) 0.54 (D) 0.75 (E) 0.38

Questions 11 and 12 refer to the following exercise:

The time of delay (in minutes) for a given bus with respect to its established schedule follows an exponential distribution with mean equal to 2.

11. The probability that the delay for a given bus is less than three minutes is:

    (A) 0.487 (B) 0.003 (C) 0.997 (D) 0.223 (E) 0.777

12. The bus firm compensates its clients if the delay time is larger than a given value. Approximately, what should this value be so that the probability that the firm has to compensate its clients is equal to 0.05?

    (A) 4 minutes (B) 2 minutes (C) 6 minutes (D) 5 minutes (E) 3 minutes

Questions 13 to 15 refer to the following exercise:

Let X, Y and Z be three independent r.v. such that their distributions are as follows: $X \in N(2, \sigma^2 = 1)$, $Y \in \chi^2_3$ and $Z \in \gamma(\frac{1}{2}, 6)$.

13. If we define the r.v. $V_1 = Y + Z$, then the probability $P(-8.55 < V_1 < 8.55)$ is:

    (A) 0.1 (B) 0.8 (C) 0.2 (D) 0.9 (E) 0.5

14. If we define the r.v. $V_2 = \dfrac{\sqrt{3}\,(X-2)}{\sqrt{Y}}$, then the value of k such that $P\left(V_2 \leq k\right) = 0.1$ is:

    (A) -3.18 (B) 1.64 (C) -2.35 (D) 2.35 (E) -1.64

15. If we define the r.v. $V_3 = \dfrac{4Y}{Z}$, then the value of k such that $P\left(V_3 \leq k\right) = 0.1$ is:

    (A) 0.38 (B) 5.22 (C) 2.61 (D) 0.19 (E) 3.49

Questions 16 and 17 refer to the following exercise:

Let X be a discrete r.v. 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 θ a r.s. of size $n = 10$ has been taken, providing the following results: 0, 0, 0, 1, 1, 1, 1, 2, 2, 2.

16. The method of moments estimate of θ is:

(A) $\frac{5}{7}$ (B) $\frac{7}{25}$ (C) $\frac{3}{25}$ (D) $\frac{2}{7}$ (E) $\frac{3}{7}$

17. The maximum likelihood estimate of θ is:

(A) $\frac{3}{7}$ (B) $\frac{7}{25}$ (C) $\frac{5}{7}$ (D) $\frac{2}{7}$ (E) $\frac{3}{25}$

Questions 18 to 21 refer to the following exercise:

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

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

In order to estimate the parameter θ, a r.s. of size n, , $X_1, \ldots, X_n$, has been taken. It is known that the mean and variance of this distribution are, respectively, $m = \frac{1}{2} + \theta$ and $\sigma^2 = \frac{1}{4}$.

18. The method of moments estimator of θ, $\hat{\theta}_{\mathrm{MM}}$, is:

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

19. The maximum likelihood estimator of θ, $\hat{\theta}_{\mathrm{ML}}$, is:

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

20. We propose to use $\hat{\theta} = \overline{X}$ as the estimator for the parameter θ. The bias for this estimator is:

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

21. The mean square error for the proposed estimator, $\hat{\theta} = \overline{X}$, is:

(A) $\frac{1}{4n}$ (B) $\frac{1}{4n} - \frac{1}{2}$ (C) $\frac{1}{4}$ (D) $\frac{1}{4n} + \frac{1}{2}$ (E) $\frac{1}{4n} + \frac{1}{4}$

Questions 22 to 24 refer to the following exercise:

In order to be able to test the null hypothesis: $H_0 : f(x) = \begin{cases} \frac{3}{8}x^2 & \text{for } x \in (0,2) \\ 0 & \text{otherwise} \end{cases}$

against the alternative hypothesis: $H_1 : f(x) = \begin{cases} \frac{3}{8}(2-x)^2 & \text{for } x \in (0,2) \\ 0 & \text{otherwise} \end{cases}$

a r.s. of size $n = 1$, X_1, has been taken.

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

(A) $(0, c]$ (B) $[c_1, c_2]$ (C) All false (D) $[c_1, c_2]^C$ (E) $[c, 2)$

23. If we decide that the significance level for this test is equal to 0.10, the critical region will be:

(A) $[0.928, 2)$ (B) $[0.737, 0.928]$ (C) $(0, 0.928]$ (D) $[0.737, 0.928]^C$ (E) $(0, 0.735]$

24. For the aforementioned significance level, the power for this test will be:

(A) 0.154 (B) 0.567 (C) 0.928 (D) 0.846 (E) 0.072

Questions 25 to 27 refer to the following exercise:

The daily number of cars that visits a given service station for its car wash service follows a Poisson distribution. The station owner considers that maintaining this car wash service is profitable if the average number of cars visiting it is at least 9. S/he wishes to test the null hypothesis that this car wash service is profitable against the alternative hypothesis that it is not.

25. At the 0.05 significance level, and if we only have information about one day of service, the most adequate decision rule will be to reject the null hypothesis if the number of cars that have used the car wash service is:

(A) ≥ 3 (B) ≤ 4 (C) ≤ 3 (D) ≥ 4 (E) ≥ 5

26. If the decision rule would be to reject the null hypothesis is $x \leq 5$, what would be the approximate power for this test for a value of $\lambda = 6$?

(A) 0.555 (B) 0.955 (C) 0.715 (D) 0.446 (E) 0.285

27. If we had information about 4 days of service and the decision rule is to reject the null hypothesis if $x_1 + x_2 + x_3 + x_4 \leq 24$, then the approximate probability of a type II error when $\lambda = 6$ is:

(A) 0.460 (B) 0.618 (C) 0.540 (D) 0.382 (E) 0.224

Questions 28 to 30 refer to the following exercise:

We wish to estimate the mean Christmas lottery expense per family (in euros) in a given province. In order to do so, a r.s. of 31 families from that specific province has been taken, asking them about the money they have spent in Christmas lottery, providing a sample mean of 188 euros and a standard deviation of 40 euros. We assume that the expense follows a normal distribution.

28. The 0.95 confidence interval for the mean lottery expense is, approximately:

(A) (188 ± 14.90) (B) (188 ± 14.08) (C) (188 ± 10.24) (D) (188 ± 9.57) (E) (188 ± 12.42)

29. We wish to test the null hypothesis that the mean expense is equal to 200 euros against the alternative hypothesis that it is different from this amount. At the α significance level, the decision rule will be to reject the null hypothesis if:

(A) $\left|\frac{188-200}{\frac{40}{\sqrt{30}}}\right| \geq t_{\overline{30}|\frac{\alpha}{2}}$ (B) $\frac{188-200}{\frac{40}{\sqrt{30}}} \geq t_{\overline{30}|\frac{\alpha}{2}}$ (C) $\left|\frac{188-200}{\frac{40}{\sqrt{30}}}\right| \geq t_{\overline{30}|\alpha}$

(D) $\frac{188-200}{\frac{40}{\sqrt{30}}} \geq t_{\overline{30}|\alpha}$ (E) $\frac{188-200}{\frac{40}{\sqrt{30}}} \leq -t_{\overline{30}|\alpha}$

30. If we now wish to test the null hypothesis that the mean expense has been at most 200 euros against the alternative hypothesis that it has been larger than this amount, at the α significance level, the decision rule will be to reject the null hypothesis if :

(A) $\left|\frac{188-200}{\frac{40}{\sqrt{30}}}\right| \geq t_{\overline{30}|\alpha}$ (B) $\frac{188-200}{\frac{40}{\sqrt{30}}} \geq t_{\overline{30}|\frac{\alpha}{2}}$ (C) $\frac{188-200}{\frac{40}{\sqrt{30}}} \leq -t_{\overline{30}|\alpha}$

(D) $\frac{188-200}{\frac{40}{\sqrt{30}}} \geq t_{\overline{30}|\alpha}$ (E) $\left|\frac{188-200}{\frac{40}{\sqrt{30}}}\right| \geq t_{\overline{30}|\frac{\alpha}{2}}$

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

Let X be a r.v. with probability density function: $f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ for $x > 0$ and $\theta > 0$.

For this r.v., it is known that $\mathrm{E}(X) = \theta$ and that $\mathrm{Var}(X) = \theta^2$.

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:

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

i) Compute the bias and the variance for both estimators. Are they unbiased? Are they consistent?

ii) Obtain the Cramer-Rao lower bound for a regular unbiased estimator of θ. Is any of the above estimators efficient?

Remark: The Cramer-Rao lower bound for a regular unbiased estimator of θ is: $L_c = \dfrac{1}{n\mathrm{E}\left[\frac{\partial \ln(f(x,\theta))}{\partial \theta}\right]^2}$

B. (10 points, 25 minutes)

A firm wishes to promote a new product in the market and it is interested in knowing if it has to market this new product to the general public or if it should concentrate on a specific age group. In order to make a decision about this matter, the firm takes a r.s. of 500 individuals so that they can try out its new product and asks them their opinion on it and their age. Results are summarized as follows.

From the 200 individuals with ages between 18 and 30 years old, 150 indicated that they liked the product and 50 that they did not like it. From the 150 individuals with ages between 31 and 50 years old, 100 indicated they they liked the product and 50 that they did not like it. From the 150 individuals with ages over 50 years old, 90 indicated that they liked the product and 60 that they did not like it.

At the 5% significance level, test the null hypothesis that the preference distribution for this product is independent from age. Make sure to write the null and alternative hypotheses, the test statistic together with its distribution under the null hypothesis, and the decision rule for this test.

C. (10 points, 25 minutes)

A firm has offered two alternative services to its clients this year: a premium and a standard service. We assume that the distributions of the clients expense (in euros) for each of the two services follow a normal distribution with variances equal to 1600, and that they are independent from each other.

i) Obtain, **providing all relevant details**, the $(1-\alpha)\%$ confidence interval for the population mean, m, for the distribution of the premium service expense, based on a r.s. of size n from the clients that asked for this specific service.

ii) The firm wishes to estimate the clients mean expense for the premium service. In order to do so, a r.s. of size 200 among the clients asking for this service has been taken, providing a sample mean of 250 euros. Obtain the 95% confidence interval for the clients mean expense asking for for this service.

iii) The premium service is considered profitable if the clients mean expense for this service is at least 254 euros. Because of this, the firm wishes to test the null hypothesis that this expense is larger than or equal to 254 against the alternative hypothesis that it is smaller that this amount. At the 5% significance level, and from the results provided in the sample, what would be the firm's decision?

iv) The firm wishes to test the null hypothesis, $H_0 : m(\text{premium}) - m(\text{standard}) \leq 50$, against the alternative hypothesis, $H_1 : m(\text{premium}) - m(\text{standard}) > 50$. If a r.s. of size $n = 200$ from clients asking for the standard service is taken, providing a sample mean of 190 euros, at the 5% significance level, what would be the test decision?

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

i)

Biases: The bias for a given estimator is defined as$\text{b}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$.

In order to be able to compute the bias for the proposed estimators, we have to obtain its mathematical expectations.

$$\text{E}(\hat{\theta}_1) = \text{E}(\overline{X}) = \text{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\text{E}(X_1) + \cdots + \text{E}(X_n)}{n} = \frac{n\theta}{n} = \theta$$

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

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

Therefore, their corresponding biases will be:

$$\text{b}(\hat{\theta}_1) = \text{E}(\hat{\theta}_1) - \theta = \theta - \theta = 0$$

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

Variances: We now compute their corresponding variances.

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

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

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

$$= \frac{8\theta^2 + (n-2)\theta^2}{(n+2)^2} = \frac{(n+6)\theta^2}{(n+2)^2}$$

Are they unbiased?

An estimator is unbiased if its bias is equal to 0. Therefore, **both estimators are unbiased**.

Are they consistent?

Given that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators and that their corresponding variances tend to zero as n goes to infinity, the sufficient condition for consistency holds and, therefore, we can state that **both estimators are consistent for** θ.

ii) We obtain the Cramer-Rao lower bound for a regular unbiased estimator of θ:

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

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

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

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

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

We replace this expectation in the Cramer-Rao lower bound formula, so that:

$$L_c = \frac{1}{n\left(\frac{1}{\theta^2}\right)} = \frac{\theta^2}{n} = \mathrm{Var}(\hat{\theta}_1)$$

Therefore, the estimator $\hat{\theta}_1 = \overline{X}$ sí is efficient for θ.

Exercise B

This test corresponds to an independence test. More specifically, the null and alternative hypotheses are:

H_0: Age and product preference are independent variables

H_1: Age and product preference are nor independent variables

The summarized data are as follows:

	S/he liked it	S/he did not like it	Total
18-30	150	50	200
31-50	100	50	150
> 50	90	60	150
Total	340	160	500

The corresponding probabilities $\hat{p}_{i\bullet}$ and $\hat{p}_{\bullet j}$ are estimated from the data, so that,

$\hat{p}_{\bullet \text{Liked it}} = \frac{340}{500} = 0.68$ $\hat{p}_{\bullet \text{Did not like it}} = \frac{160}{500} = 0.32$

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

We now build the corresponding 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, Liked it	150	$0.4 \times 0.68 = 0.272$	136	1.441
18-30, Did not like it	50	$0.4 \times 0.32 = 0.128$	64	3.063
31-50, Liked it	100	$0.3 \times 0.68 = 0.204$	102	0.039
31-50, Did not like it	50	$0.3 \times 0.32 = 0.096$	48	0.083
> 50, Liked it	90	$0.3 \times 0.68 = 0.204$	102	1.412
> 50, Did not like it	60	$0.3 \times 0.32 = 0.096$	48	3.000
	500	1	500	$z = 9.038$

Under the null hypothesis of independence, the test statistics $Z = \sum_{i=1}^{k'}\sum_{j=1}^{k''} \frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \sim \chi^2_{(k'-1)(k''-1)}$, where k' and k'' are the number of classes in which every one of the two variables was divided ($k' = 3$, $k'' = 2$).

At the approximate 5% significance level, the decision rule is to reject the null hypothesis if:

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

In this specific case,

$$9.038 > 5.99 = \chi^2_{2\,,\,0.05}\,,$$

so that, at the 5% significance level, the null hypothesis of independence is rejected. Therefore, product preferences do vary with age.

Exercise C

i) We wish to obtain, providing all relevant details, the form of the $(1-\alpha)\%$ confidence interval for the mean of the distribution of the variable X: yearly expense for the premium service. We know that $X \in N(m, \sigma^2 = 1600)$

Confidence Interval

Let $X_1, \dots, X_n$ be a r.s. of size n taken from the $N(m, \sigma^2 = 1600)$ distribution, so that we have that:

$$\overline{X} \in N\left(m, \frac{\sigma^2}{n}\right) \qquad \Rightarrow \qquad \frac{\overline{X} - m}{\sigma/\sqrt{n}} \in N(0,1)$$

$$P\left(-t_{\frac{\alpha}{2}} < \frac{\bar{X} - m}{\sigma/\sqrt{n}} < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - m < t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(-\bar{X} - t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < -m < -\bar{X} + t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < m < \bar{X} + t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Therefore, the $(1 - \alpha)\%$ confidence interval for m is:

$$\text{CI}_{1-\alpha} = \left(\bar{x} \pm t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = \left(\bar{x} \pm t_{\frac{\alpha}{2}} \frac{40}{\sqrt{n}}\right)$$

ii) We obtain the 0.95 confidence interval for the mean using the information provided from the specific sample.

$$\text{CI}_{1-\alpha}(m) = \left(\overline{x} \pm t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$t_{\frac{0.05}{2}} = 1.96$$

$$\text{CI}_{0.95}(m) = \left(250 \pm 1.96 \cdot \frac{40}{\sqrt{200}}\right) = (250 \pm 5.54) =$$

$$=(244.46\ ,\ 255.54)$$

iii) We have to test the null hypothesis: $H_0 : m \geq 254$ against the alternative hypothesis: $H_1 : m < 254$.

Under H_0, $\dfrac{\bar{X} - 254}{\sigma/\sqrt{n}} \in N(0,1)$

At the α significance level, the decision rule for this test is to reject H_0 if: $\dfrac{\bar{X} - 254}{\sigma/\sqrt{n}} \leq -t_\alpha$.

In this case, and given that $\alpha = 5\%$,

$$\frac{250 - 254}{40/\sqrt{200}} = -1.414 \ > \ -1.64 = -t_{0.05}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis and, therefore, the firm believes that the premium service is profitable.

iv) We have to test:

$$H_0 : m_p - m_s \leq 50$$

$$H_1 : m_p - m_s > 50$$

Under H_0, $$\frac{\left(\bar{X}_p - \bar{X}_s\right) - 50}{\sqrt{\frac{\sigma_p^2}{n_p} + \frac{\sigma_s^2}{n_s}}} \in N(0,1)$$

At the α significance level, the decision rule is to reject H_0 if:

$$\frac{(\bar{x}_p - \bar{x}_s) - 50}{\sqrt{\frac{\sigma_p^2}{n_p} + \frac{\sigma_s^2}{n_s}}} \geq t_\alpha$$

In this case, and given that $\alpha = 5\%$,

$$\frac{(250 - 190) - 50}{\sqrt{\frac{1600}{200} + \frac{1600}{200}}} = 2.5 \ \geq \ 1.64 = t_{0.05}$$

Therefore, at the 5% significance level, we reject the null hypothesis.