STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2020-21
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
Second Call. July 8, 2021

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 45 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 length, in millimeters, of the screws that are manufactured in a given specialized factory follows a $N(20, \sigma^2 = 4)$ distribution. Screws appropriate for sale are only those with length included in the interval $(17.92, 22.08)$. We assume independence between the lengths of the different screws (you need to round up to two decimal places the probability that the screw is appropriate to be sold).

2. If a batch of 10 screws is sold, the probability that 9 of them are appropriate for sale is:

   (A) 0.9718 (B) 0.8789 (C) 0.4035 (D) 0.0282 (E) 0.1211

3. If a batch of 10 screws is sold, the probability that at most 5 of them are appropriate for sale is:

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

4. If a batch of 100 screws is sold, what is the approximate probability that at least 74 of them are appropriate for sale?

   (A) 0.7764 (B) 0.2236 (C) 0.8849 (D) 0.1635 (E) 0.8365

Questions 5 and 6 refer to the following exercise:

Let X be r.v. having a Poisson distribution with parameter λ. It is known that $P(5) = 0.5P(4)$.

5. The probability $P(X = 3)$ is:

   (A) 0.2138 (B) 0.1404 (C) 0.7862 (D) 0.8596 (E) .01804

6. We have another r.v. Y, independent of the previous one, having the same probability distribution. The probability $P(X + Y \leq 6)$ is:

   (A) 0.7622 (B) 0.8893 (C) 0.1107 (D) 0.2378 (E) 0.6160

Questions 7 to 9 refer to the following exercise:

The number of clients arriving each hour to park their car at a given parking facility in an entertainment place follows a Poisson distribution with mean equal to 4. We assume independence between client arrivals at different hours.

7. The probability that, in a two-hour period, at least five clients arrive at the parking facility is:

   (A) 0.0996 (B) 0.8088 (C) 0.1912 (D) 0.9004 (E) 0.6866

8. The approximate probability that, in a ten-hour period, fewer than 43 clients arrive at the parking facility is:

   (A) 0.2877 (B) 0.7123 (C) 0.7642 (D) 0.3446 (E) 0.6554

9. What is the minimum number of free parking spots that the parking facility should have so that the probability that all the clients arriving at a given hour are able to park is at least 0.94?

(A) 7 (B) 6 (C) 8 (D) 5 (E) 4

Questions 10 and 11 refer to the following exercise:

Let $X_1, \ldots, X_4$ be four independent r.v. each having a $\gamma(2,2)$ distribution.

10. The mean and the variance of the r.v. $Y = X_1 + X_2 + X_3 + X_4$ are, respectively:

(A) $m = 4$, $\sigma^2 = 2$ (B) $m = 1$, $\sigma^2 = \frac{1}{8}$ (C) $m = 1$, $\sigma^2 = 1$ (D) $m = 1$, $\sigma^2 = \frac{1}{2}$ (E) $m = 4$, $\sigma^2 = 4$

11. If we define the r.v. $Z = 4X_1$, the probability $P(Z > 7.78)$ is:

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

Questions 12 to 14 refer to the following exercise:

The lifetime or time of duration of a given home appliance, in **thousands of hours**, follows an exponential distribution with mean equal to 2. We assume independence between the different home appliances.

12. The probability that a home appliance of this type has a lifetime of more than three thousand hours is:

(A) 0.223 (B) 0.632 (C) 0.777 (D) 0.002 (E) 0.368

13. If the lifetime of a given home appliance is smaller than its established warranty period (i.e., the one established by the firm selling it), the firm has to replace it with a brand new one. What should this warranty period be so that the probability that the firm has to replace a randomly selected home appliance is equal to 0.05?

(A) 5991 hours (B) 4000 hours (C) 102.6 hours (D) 210.7 hours (E) 155.9 hours

14. If the firm establishes the aforementioned warranty period and, if in a given month it sells a total of 200 of those home appliances, the expected number of home appliances that the firm should replace is:

(A) 20 (B) 5 (C) 1 (D) 2 (E) 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: $X \in N(2, \sigma^2 = 4)$, $Y \in \chi^2_9$ and $Z \in \gamma(\frac{1}{2}, 3)$.

15. If we define the r.v. $V_1 = \left(\frac{X-2}{2}\right)^2 + Z$, then the probability $P(V_1 \in (9.04, 14.10))$ is:

(A) 0.25 (B) 0.05 (C) 0.7 (D) 0.20 (E) 0.80

16. If we define the r.v. $V_2 = \dfrac{3\,(X-2)}{2\sqrt{Y}}$, the value of k such that $P\left(V_2 < k\right) = 0.05$ is:

(A) -2.82 (B) 2.26 (C) -1.83 (D) 1.83 (E) -2.26

17. If we define the r.v. $V_3 = \dfrac{9Z}{6Y}$, then the probability $P\left(0.1253 \leq V_3 \leq 3.37\right)$ is:

(A) 0.85 (B) 0.99 (C) 0.80 (D) 0.94 (E) 0.90

Questions 18 and 19 refer to the following exercise:

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

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

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

18. The method of moments estimate of θ is:

(A) 0.64 (B) 0.16 (C) 0.80 (D) 0.20 (E) 0.32

19. The maximum likelihood estimate of θ is:

(A) 0.80 (B) 0.32 (C) 0.64 (D) 0.16 (E) 0.20

Questions 20 and 21 refer to the following exercise:

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

$$f(x,\theta) = \begin{cases} (\theta+2)x^{\theta+1} & \text{for } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

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 of this distribution is $m = \frac{\theta+2}{\theta+3}$.

20. The method of moments estimator of θ is:

(A) $\dfrac{\overline{X}+2}{\overline{X}+3}$ (B) $\dfrac{3\overline{X}-2}{1-\overline{X}}$ (C) $-\dfrac{2n}{\ln\left(\Pi_{i=1}^{n} X_i\right)}$ (D) $-\dfrac{n}{\ln\left(\Pi_{i=1}^{n} X_i\right)} - 2$ (E) $\dfrac{\overline{X}+2}{1-\overline{X}}$

21. The maximum likelihood estimator of θ is:

(A) $\dfrac{\overline{X}+2}{1-\overline{X}}$ (B) $\dfrac{3\overline{X}-2}{1-\overline{X}}$ (C) $\dfrac{\overline{X}+2}{\overline{X}+3}$ (D) $-\dfrac{n}{\ln\left(\Pi_{i=1}^{n} X_i\right)} - 2$ (E) $-\dfrac{2n}{\ln\left(\Pi_{i=1}^{n} X_i\right)}$

Questions 22 and 23 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with unknown parameter λ. We wish to estimate the parameter λ and, in order to do so, a r.s. of size n has been taken and the estimator $\hat{\lambda} = \dfrac{2X_1 + X_2 + \cdots + X_{n-1} + 2X_n}{n+2}$ is proposed.

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

(A) $\dfrac{4\lambda}{n+2}$ (B) 0 (C) $\dfrac{(n-2)\lambda}{n+2}$ (D) $\dfrac{2\lambda}{n+2}$ (E) $\dfrac{\lambda}{n+2}$

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

(A) $\dfrac{\lambda}{(n+2)}$ (B) $\dfrac{(n+4)\lambda}{(n+2)^2}$ (C) $\dfrac{(n+6)\lambda}{(n+2)^2}$ (D) $\dfrac{(n+6)\lambda}{n^2+4}$ (E) $\dfrac{(n+4)\lambda}{n^2+4}$

Questions 24 to 26 refer to the following exercise:

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

$$f(x;\theta) = \begin{cases} 1+\theta\left(x^3 - \frac{1}{4}\right) & \text{if } x \in (0,1) \\ \\ 0 & \text{otherwise} \end{cases}$$

We wish to test the null hypothesis $\theta = 1$ against the alternative hypothesis $\theta = 0$. In order to do so, a r.s. of size 1, X_1, has been taken.

24. The form of the most powerful critical region for x_1 will be:

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

25. If we decide to reject the null hypothesis when $x_1 \leq 0.1$, the significance level for this test is:

(A) 0.6510 (B) 0.9250 (C) 0.8475 (D) 0.1525 (E) 0.0750

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

(A) 0.10 (B) 0.90 (C) 0.50 (D) 0.05 (E) 0.95

27. The number of delayed flights departing each hour from a given airport follows a Poisson distribution. We wish to test the null hypothesis that the mean of the distribution is at most 3 against the alternative hypothesis that it is larger than this value. In order to verify this hypothesis a r.s. of 3 hours is taken and the test statistic, z, the total number of delayed flights in those three hours, is considered. At the 5% significance level, the decision rule is to reject the null hypothesis if:

(A) $z \geq 14$ (B) $z \geq 16$ (C) $z \geq 15$ (D) $z \geq 12$ (E) $z \geq 13$

Questions 28 to 30 refer to the following exercise:

We wish to estimate the mean olive oil consumption, in liters, per family and year, in a given city. In order to do so, a r.s. of 41 families has been taken, providing the following results: $\overline{x} = 45$ and $s^2 = 100$. It is known that the olive oil consumption distribution follows a normal distribution.

28. A 95% confidence interval for the mean olive oil consumption per family and year is:

(A) $(41.806, 48.194)$ (B) $(42.551, 47.499)$ (C) $(40.078, 49.922)$
(D) $(40.438, 49.562)$ (E) $(42.344, 47.656)$

29. We wish to test the null hypothesis that the mean is at least 48 against the alternative hypothesis that it is smaller than this value. At the 5% significance level, the test decision will be to reject the null hypothesis if:

(A) $\dfrac{45-48}{10/\sqrt{40}} \leq -2.02$ (B) $\dfrac{45-48}{10/\sqrt{40}} \leq -1.68$ (C) $\dfrac{45-48}{10/\sqrt{40}} \geq 1.68$
(D) $\dfrac{45-48}{10/\sqrt{40}} \geq -1.68$ (E) $\dfrac{45-48}{10/\sqrt{40}} \geq 2.02$

30. A 90% confidence interval for the **standard deviation** of the olive oil consumption per family and year is:

(A) $(11.48, 24.16)$ (B) $(7.35, 15.47)$ (C) $(6.91, 16.80)$ (D) $(7.83, 14.07)$ (E) $(8.57, 12.44)$

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

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

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

It is known that the mean of this r.v. is $m = \theta\sqrt{\frac{\pi}{2}}$ and its variance is $\sigma^2 = \left(\frac{4-\pi}{2}\right)\theta^2$. We wish to estimate the parameter θ and, in order to do so, a r.s. of size n, $X_1, \ldots, X_n$, has been taken.

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

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

iii) Is the method of moments estimator unbiased? Consistent? Provide all relevant details to appropriately justify your answers.

iv) A r.s. of size 5 provided the folllowing results: $3.2, 4.3, 2.5, 5.3, 6.7$. Obtain the estimate of the parameter θ by using one of the two methods included in the previous items.

B. (10 points, 25 minutes)

A given illness can be treated with two different drugs, A and B. There is the suspicion that the side effects associated to those medicines can have different effects on patients receiving them. In order to further investigate this issue, two r.s. were taken, one of size $n_A = 300$, from the group of patients receiving treatment A, and another one of size $n_B = 200$ from the group of patients receiving treatment B. The results obtained from those two samples are summarized in the following table.

	No effect	Moderate effect	Strong effect	Total
Drug A	250	45	5	300
Drug B	170	20	10	200
Total	420	65	15	500

At the 1% significance level, test the null hypothesis that the distribution of the side effects produced by the two drugs are the same. Make sure to write down the null and alternative hypotheses, the test statistic, together with its distribution under the null hypothesis, as well as the test decision rule.

C. (10 points, 25 minutes)

A firm has a machine that manufactures objects or pieces of a given specification. The firm wishes to estimate the proportion of defective pieces the machine produces. In order to do so, a r.x. of 200 pieces is taken, with the result that 22 of them were defective.

i) Obtain a 90% confidence interval for the proportion of defective pieces.

ii) The firm considers that the production is appropriate if the proportion of defective pieces is at most 0.07. At the 5% significance level, and using the results obtained from the previous sample, what is the firm's conclusion with regard to the issue of the appropriateness of the production?

iii) With the purpose of reducing the proportion of defective pieces, the firm is considering the possibility of changing the machine that manufactures the aforementioned pieces. In order to do so, the firm decides to rent a new machine during a month and, if it really considers that the proportion of defective pieces has decreased, it would buy the new machine. To be able to make a decision on this matter, a r.s. of 300 pieces manufactured by the new machine, with the result that 18 of them were defective. At the 5% significance level, test the null hypothesis that the new machine produces the same proportion of defective pieces against the alternative hypothesis that it reduces this proportion. Moreover, from the results obtained, will the firm finally buy the new machine?

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

The probability density function for the r.v. X is:

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

with corresponding mean and variance given by $m = \theta\sqrt{\frac{\pi}{2}}$ and $\sigma^2 = \left(\frac{4-\pi}{2}\right)\theta^2$.

i) The sample likelihood function is as follows:

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

We compute its natural logarithm:

$$\ln L(\vec{x};\theta) = \ln\left(\Pi_{i=1}^n x_i\right) - 2n\ln\theta - \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2$$

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

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

$$\hat{\theta}_{\mathrm{ML}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}}$$

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

$$\alpha_1 = E(X) = a_1 = \overline{X} \Rightarrow \theta\sqrt{\frac{\pi}{2}} = \overline{X}$$

$$\hat{\theta}_{\mathrm{MM}} = \overline{X}\sqrt{\frac{2}{\pi}}$$

iii) We need to verify if the method of moments estimator is unbiased and consistent.

Unbiased: The method of moments is unbiased if $\mathrm{E}\left(\hat{\theta}_{\mathrm{MM}}\right) = \theta$.

$$\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \mathrm{E}\left[\overline{X}\sqrt{\frac{2}{\pi}}\right] = \sqrt{\frac{2}{\pi}}\mathrm{E}\left(\overline{X}\right) = \sqrt{\frac{2}{\pi}}\mathrm{E}\left(X\right) = \sqrt{\frac{2}{\pi}}\theta\sqrt{\frac{\pi}{2}} = \theta$$

Therefore, it is unbiased.

Consistent: An estimator is consistent if it verifies the sufficient conditions for consistency. That is,

a) The estimator is unbiased or asymptotically unbiased.

b) $\lim_{n\to\infty} \text{Var}\left(\hat{\theta}_{\text{MM}}\right) = 0$

As $\hat{\theta}_{\text{MM}}$ is unbiased, the first condition holds.

We now compute the estimator's variance.

$$\text{Var}(\hat{\theta}_{\text{MM}}) = \text{Var}\left[\overline{X}\sqrt{\frac{2}{\pi}}\right] = \left(\frac{2}{\pi}\right) \text{Var}\left(\overline{X}\right) = \left(\frac{2}{\pi}\right)\left[\frac{\text{Var}(X)}{n}\right] = \left(\frac{2}{\pi}\right)\left(\frac{4-\pi}{2}\right)\frac{\theta^2}{n} = \left(\frac{4-\pi}{\pi}\right)\frac{\theta^2}{n}$$

As we have that:

$$\lim_{n\to\infty} \text{Var}\left(\hat{\theta}_{\text{MM}}\right) = \lim_{n\to\infty}\left(\frac{4-\pi}{\pi}\right)\frac{\theta^2}{n} = 0,$$

the second condition also holds and, therefore, the method of moments estimator is consistent.

iv) From the resulting sample, parameter estimates for the two methods considered here are:

a) Maximum likelihood method:

$$\hat{\theta}_{\text{ML}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{2n}} = \sqrt{\frac{3.2^2 + 4.3^2 + 2.5^2 + 5.3^2 + 6.7^2}{10}} = 3.286$$

b) Method of moments:

$$\hat{\theta}_{\text{MM}} = \overline{X}\sqrt{\frac{2}{\pi}} = \left[\frac{3.2 + 4.3 + 2.5 + 5.3 + 6.7}{5}\right]\sqrt{\frac{2}{\pi}} = 3.511$$

Exercise B

It corresponds to a test of homogeneity in which we have two samples ($I = 2$) and three categories for the variable under study ($J = 3$).

Therefore, the null and alternative hypotheses are, respectively:

H_0: the distribution of the variable is homogeneous in both populations

H_1: the distribution of the variable is not homogeneous in both populations

The data are summarized in the following table:

	No effect	Moderate effect	Strong effect	Total
Drug A	250	45	5	300
Drug B	170	20	10	200
Total	420	65	15	500

Theoretical probabilities $\hat{p}_j$ are estimated from the information provided in the table. In this way,

$$\hat{p}_{\text{NE}} = \frac{420}{500} = 0.84, \qquad \hat{p}_{\text{ME}} = \frac{65}{500} = 0.13, \qquad \hat{p}_{\text{SE}} = \frac{15}{500} = 0.03.$$

We now build the corresponding table:

	n_{ij}	$\hat{p}_{\bullet j}$	$n_i\hat{p}_{\bullet j}$	$\frac{(n_{ij}-n_i\hat{p}_{\bullet j})^2}{n_i\hat{p}_{\bullet j}}$
(A, NE)	250	0.84	252	0.0159
(A, ME)	45	0.13	39	0.9231
(A, SE)	5	0.03	9	1.7778
(B, NE)	170	0.84	168	0.0238
(B, ME)	20	0.13	26	1.3846
(B, SE)	10	0.03	6	2.6667
	500	2	500	$z = 6.7919$

Under the null hypothesis of homogeneity, the test statistic $\sum_{i,j} \frac{(n_{ij}-n_i\hat{p}_j)^2}{n_i\hat{p}_j} \sim \chi^2_{(I-1)(J-1)}$, where I is th number of samples ($I = 2$) and J is the number of categories in which the variable under study has been divided ($J = 3$).

At the α significance level, the decision rule will be to reject the null hypothesis if, after taking the samples,

$$z = \sum_{i,j} \frac{(n_{ij} - n_i\hat{p}_j)^2}{n_i\hat{p}_j} \geq \chi^2_{(I-1)(J-1),\alpha}$$

In this case:

$$6.7919 < 9.21 = \chi^2_{(2-1)(3-1)\,,\,0.01} = \chi^2_{2,0.01},$$

so that, at the 1% significance level, the null hypothesis of homogeneity is not rejected. That is, we do not reject that the distribution of the side effects produced by both drugs are the same.

Exercise C

i) The $(1-\alpha)$ approximate confidence interval is:

$$\text{CI}_{1-\alpha}(p) = \left(\frac{z}{n} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{z(n-z)}{n^3}} \right)$$

In this case, given that $n = 200$ and $z = 22$, the 90% approximate confidence interval will be:

$$\text{CI}_{0.90}(p) = \left(\frac{22}{200} \pm 1.64 \sqrt{\frac{22 \cdot 178}{200^3}} \right) = (0.11 \pm 0.036) = (0.074, 0.146)$$

ii) We wish to test the null hypothesis that the proportion of defective pieces is at most 0.07 against the alternative hypothesis that it is larger than this value; that is,

$$H_0 : p \leq 0.07$$
$$H_1 : p > 0.07$$

At the α significance level, the decision rule will be to reject the null hypothesis if $\dfrac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \geq t_\alpha$.

In this case, at the 0.05 significance level,

$$\frac{\frac{22}{200} - 0.07}{\sqrt{\frac{0.07 \cdot 0.93}{200}}} = 2.217 \geq 1.64$$

Therefore, at the 5% significance level, the null hypothesis establishing that $p \leq 0.07$ is rejected.

iii) We wish to test the null hypothesis that the new machine manufactures the same proportion of defective pieces against the alternative hypothesis that it reduces this proportion. That is,

$$H_0 : p_1 = p_2 \quad \equiv \quad p_1 - p_2 = 0$$
$$H_1 : p_1 > p_2 \quad \equiv \quad p_1 - p_2 > 0$$

At the α significance level, the decision rule rejects H_0 if:

$$\frac{\frac{z_1}{n_1} - \frac{z_2}{n_2}}{\sqrt{\frac{z(n-z)}{n \cdot n_1 \cdot n_2}}} \geq t_\alpha$$

In this case:

$$\frac{\frac{22}{200} - \frac{18}{300}}{\sqrt{\frac{40 \cdot 460}{500 \cdot 200 \cdot 300}}} = 2.019 > t_{0.05} = 1.64$$

Therefore, at the 5% significance level, the null hypothesis indicating that the proportion of defective pieces is the same with the new machine is rejected. As a consequence of this decision, the firm will buy the new machine.