STATISTICS APPLIED TO BUSINESS ADMINISTRATION (BAD) - Second Year Academic Year 2015-16
STATISTICS APPLIED TO MARKETING (MD) - Second Year
STATISTICS APPLIED TO BUSINESS ADMINISTRATION (DD) - Third Year
Second Call. June 22, 2016

INSTRUCTIONS

1. The exam contains multiple choice questions that must be answered in the orange code sheet we have provided you with, together with three exercises that must be answered in detail in separate sheets of paper.
2. To select an answer, all you need to do is mark in the orange code sheet, **filling the rectangle over which the selected answer is located appropriately**. Please make sure you know the answer you wish to mark before doing it. Even though you can always erase your mark if you have used a pencil (number 2 or similar), any mark that has not been completely erased could be read by the machine. Therefore, we advice you to first mark your selected answers in the exam and to use only the last ten minutes or so from the time assigned to the multiple choice questions-part of the exam to copy them into the code sheet.
3. In the multiple choice questions-part of the exam there is always **only one correct answer** for every question. Every question correctly answered is worth 1 point, while each question incorrectly answered will penalize your grade by 0.2 points. Questions that have not been answered do not penalize your grade in any form.
4. Each one of the exercises, A, B and C, must be answered in a separate sheet of paper. We will collect the different parts of the exam at the indicated times and in this order: first, we will collect the code sheet for the multiple choice questions part of the exam and, then, and in this order, exercises A, B and C will be collected.
5. The exam has six numbered sheets, going from 0.1 to 0.6. Please make sure that you have all sheets and contact your professor if this is not the case. There are different exam types. This exam is of type 0. Mark a 0 in the column labelled with I in your code sheet.
6. The maximum final grade for each of the parts of the exam (i.e., the multiple choice questions part and the exercises part) is 30 points. You will need to obtain 15 points in each part of the exam to pass it.
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:

It is known that, in a given town, 40% of the inhabitants went to the movies more than three times last year.

2. If a r.s. of 12 inhabitants is taken, what is the probability that 10 of them had gone to the movies more than three times last year?

   (A) 0.0025 (B) 0 (C) 0.0005 (D) 1 (E) 0.9975

3. If a r.s. of 20 inhabitants is taken, what is the probability that at least 14 of them **had not gone** to the movies more than three times last year?

   (A) 0.7500 (B) 0.2500 (C) 0.9984 (D) 0.0016 (E) 0.1256

4. If a r.s. of 200 inhabitants is taken, what is the approximate probability that less than 88 of them had gone to the movies more than three times last year?

   (A) 0.86 (B) 0.89 (C) 0.20 (D) 0.11 (E) 0.80

5. The probability that a student does not pass a given exam is 0.015. If 200 students are randomly selected, the approximate probability that exactly 3 of them do not pass the exam is:

   (A) 0.3528 (B) 0.4232 (C) 0.6472 (D) 0.2240 (E) 0.5768

Questions 6 and 7 refer to the following exercise:

The number of units of a given product ordered per day follows a Poisson distribution with mean 3. We assume independence between the orders taking place in different days.

6. What should be the minimum available stock of that product in the store, k, such that with probability of at least 0.95 all of the ordered sales that specific day can be handled? That is, $F_X(k) \geq 0.95$.

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

7. What is the approximate probability that, in a ten-day period, more than 26 units of that product are ordered?

   (A) 0.82 (B) 0.26 (C) 0.78 (D) 0.63 (E) 0.74

8. It is known that the waiting time (in minutes) in a line for a given car dealer is a random variable following an exponential distribution with mean 25. What is the probability of waiting more than 30 minutes?

   (A) 0.5430 (B) 0.9994 (C) 0.0006 (D) 0.3012 (E) 0.6988

Questions 9 and 10 refer to the following exercise:

Let X be a random variable having a $\gamma(1, 5)$ distribution.

9. If $Y = 2X$, then the distribution of the r.v. Y is:

   (A) $\gamma(2, 5)$ (B) $\exp(\lambda = 10)$ (C) $\gamma(2, 10)$ (D) $\exp(\lambda = 2.5)$ (E) $\gamma(\frac{1}{2}, \frac{10}{2})$

10. The value of k such that $P(Y > k) = 0.25$ is:

   (A) 6.63 (B) 20.50 (C) 12.50 (D) 6.74 (E) 12.80

Questions 11 to 13 refer to the following exercise:

Let X_1, X_2, X_3 and X_4 be four independent r.v. with distributions such that: $X_1 \in N(0, \sigma^2 = 1)$, $X_2 \in N(3, \sigma^2 = 9)$, $X_3 \in N(0, \sigma^2 = 9)$ and $X_4 \in N(2, \sigma^2 = 4)$.

11. The probability that the r.v. $Y = \left(\frac{X_2 - 3}{3}\right)^2 + \left(\frac{X_4 - 2}{2}\right)^2$ belongs to the interval $(0.575,\ 5.99)$ is:

   (A) 0.05 (B) 0.70 (C) 0.75 (D) 0.80 (E) 0.30

12. We define the r.v. $Z = \dfrac{18X_1^2}{(X_2 - 3)^2 + X_3^2}$. The value of k such that $P(Z > k) = 0.05$ is:

   (A) 5.025 (B) 18.500 (C) 199 (D) -5.400 (E) 5.400

13. The probability that the r.v. $V = \dfrac{\sqrt{2}X_1}{\sqrt{\left(\frac{X_2-3}{3}\right)^2 + \left(\frac{X_4-2}{2}\right)^2}}$ is larger than 1.89 is:

   (A) 0.90 (B) 0.05 (C) 0.20 (D) 0.80 (E) 0.10

Questions 14 and 15 refer to the following exercise:

Let X be a discrete r.v. with probability mass function given by: $P(1) = \lambda$, $P(2) = \frac{\lambda}{4}$, $P(3) = 1 - \frac{5\lambda}{4}$. In order to estimate the parameter λ, a r.s. of size $n = 10$ has been taken, rendering the following results: 1, 1, 3, 1, 2, 1, 2, 1, 1, 3.

14. The method of moments estimate of λ is:

   (A) 0.170 (B) 0.622 (C) 0.325 (D) 0.444 (E) 0.640

15. The maximum likelihood estimate of λ is:

   (A) 0.622 (B) 0.640 (C) 0.170 (D) 0.444 (E) 0.325

16. A r.s. of size $n = 4$ is taken from a population having a uniform distribution on the interval $[\theta,\ 10]$, rendering the following results: 1, 5, 9 and 7. Under the maximum likelihood criterion, which one of the following distributions would we select as the one generating such a sample?

   (A) $U[0,\ 10]$ (B) $U[9,\ 10]$ (C) $N(1,\ 9)$ (D) $U[1,\ 10]$ (E) $N(0,\ 7)$

17. Let X be a r.v. with probability density function $f(x) = \theta x^{\theta-1}$, $x \in (0,\ 1)$. The maximum likelihood estimate of θ computed from a r.s. of size $n = 1$, rendering a result of $x = 0.5$ is:

   (A) 0.5 (B) $-\ln(0.5)$ (C) $\ln(0.5)$ (D) $\frac{1}{\ln(0.5)}$ (E) $\frac{-1}{\ln(0.5)}$

Questions 18 to 20 refer to the following exercise:

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

$$f(x,\theta) = \frac{2^2}{\theta^3}\, x^2\, e^{\frac{-2x}{\theta}}, \qquad x > 0, \qquad \theta > 0$$

It is known that the mean and variance of the r.v. X are $m = \frac{3\theta}{2}$ and $\sigma^2 = \frac{3\theta^2}{4}$. We wish to estimate the parameter θ and, thus, in order to do so, a r.s. of size n, $X_1, \ldots, X_n$, has been taken, and the following estimators are proposed:

$$\hat{\theta}_1 = \frac{X_1 + X_2 + X_3}{3} \qquad \hat{\theta}_2 = \frac{X_1 + X_2}{3}$$

18. With regard to $\hat{\theta}_1$ and $\hat{\theta}_2$, we can state that:

(A) Both are unbiased (B) $\hat{\theta}_1$ is unbiased and $\hat{\theta}_2$ is biased (C) -
(D) $\hat{\theta}_1$ is biased and $\hat{\theta}_2$ is unbiased (E) Both are biased

19. The variance of $\hat{\theta}_2$ is:

(A) $\frac{1}{2}\theta^2$ (B) $\frac{1}{6}\theta^2$ (C) $\frac{1}{12}\theta^2$ (D) $\frac{3}{4}\theta^2$ (E) $\frac{1}{4}\theta^2$

20. With regard to $\hat{\theta}_1$, we can state that:

(A) It is efficient (B) - (C) - (D) - (E) It is not efficient

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

$$f(x,\theta) = \theta^2\, x\, e^{-\theta x}, \qquad x > 0, \qquad \theta > 0$$

Based on a r.s. of size $n = 1$, we wish to test the null hypothesis that $\theta = 3$ against the alternative hypothesis that $\theta = 2$. The most powerful critical region for this single observation and for a given significance level is of the form:

(A) $X \in (C_1, C_2)^c$ (B) $X \leq C$ (C) $X \geq C$ (D) $X \in (C_1, C_2)$ (E) All false

Questions 22 to 24 refer to the following exercise:

A given machine manufactures engines, whose fuel consumption per hour follows an exponential distribution. If the machine works properly, the engine's mean fuel consumption is of 1 liter, and, if the machine does not work properly, the engine's mean fuel consumption is 2 liters instead. In order to be able to test, in a given moment in time, if the machine works properly, one of those engines is randomly selected and the following decision rule is used: if during a given hour, the engine's fuel consumption is larger than 1.2 liters, the hypothesis that the machine works properly is rejected, and, if otherwise, this hypothesis is not rejected. **Remark**: If $X \in \exp(\lambda)$, then $F_X(x) = 1 - e^{-\lambda x}, x \geq 0$.

22. The significance level for this test is:

(A) 0.301 (B) 0.465 (C) 0.289 (D) 0.684 (E) 0.698

23. The power for this test is:

(A) 0.907 (B) 0.284 (C) 0.549 (D) 0.361 (E) 0.212

24. If we now consider the alternative critical region $(0.163, 1.891)^C$, which has the same significance level than the aforementioned unilateral critical region, can we state that the unilateral test is better than this one?

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

25. We have a population for which we can classify each subject in one of four possible different classes. We wish to test the hypothesis that the probabilities of belonging to each one of those classes are p_1, p_2, p_3 and p_4 (**known**). In order to do so, a r.s. of size $n = 200$ is taken. The most appropriate test is:

(A) Test of independence

(B) A χ^2 goodness of fit test to a partially specified distribution

(C) Test of homogeneity

(D) A χ^2 goodness of fit test to a completely specified distribution

(E) All false

26. Let X be the number of cars sold per week in a given car dealer. It is known that X follows a Poisson distribution. In order to be able to test if the mean number of cars sold per week is at least 4, a random sample of 2 weeks is taken. At the 5% significance level, the decision rule is to reject the null hypothesis if z, the total number of cars sold during the two-week period, verifies that:

(A) $z \leq 3$ (B) $z \geq 12$ (C) $z \in (1, 12)^c$ (D) $z \geq 14$ (E) $z \leq 4$

Questions 27 and 28 refer to the following exercise:

A firm manufacturing tablets states that at most 11% of its products is defective and needs to be replaced. In order to be able to test the firm's statement, a r.s. of 200 of those tablets is taken with the result that 30 of them were defective.

27. The approximate 0.95 confidence interval for the proportion of defective tablets is:

(A) $(0.1255, 0.1755)$ (B) $(0.1125, 0.1585)$ (C) $(0.1005, 0.1995)$ (D) $(0.0915, 0.2095)$ (E) $(0.1355, 0.1655)$

28. At the 5% significance level, what will be the decision about the firm's statement?

(A) It is not rejected (B) - (C) - (D) - (E) It is rejected

Questions 29 and 30 refer to the following exercise:

In a study about leisure, we wish to estimate the monthly mean expense (in euros) in going to the movies. In order to do so, the population has been divided into two groups of people of different ages. The first age group is the one with individuals younger 35, and the second one is that of individuals 35 or older. The corresponding population sizes for these two groups are $N_1 = 20000$ and $N_2 = 30000$, with corresponding population quasivariances equal to 400 and 100, respectively. It has been decided that a stratified random sampling will be taken, with a total sample size of 1000 individuals.

29. If proportional allocation is used, the corresponding sample sizes, n_1 and n_2, will be:

(A) 200 and 800 (B) 400 and 600 (C) 500 and 500 (D) 727 and 273 (E) 571 and 429

30. If optimal allocation is used, the corresponding sample sizes, n_1 and n_2, will be:

(A) 257 and 743 (B) 400 and 600 (C) 571 and 429 (D) 273 and 727 (E) 200 and 800

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

In order to be able to test if the variables size of a firm and its profit level (in millions of euros) are independent variables, a random sample of 200 firms is taken, rendering the following results for firms grouped by their size (small or large), and their profit (less than 2 million, between 2 and 10 million, and larger than 10 million):

	< 2	$[2, 10]$	> 10
Small	20	50	10
Large	10	50	60

At the 1% significance level, carry out the aforementioned test of hypothesis.

B. (10 points, 25 minutes)

Let $X_1, \cdots, X_n$ be a r.s. from a population having probability density function given by:

$$f(x;\theta) = \frac{x^4\, e^{-\frac{x}{\theta}}}{24\,\theta^5} \qquad x > 0, \qquad \theta > 0$$

It is known that $\mathrm{E}(X) = 5\theta$ and $\sigma^2_X = 5\theta^2$

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

ii) Is this estimator unbiased? Is it consistent?

iii) Find the Cramer-Rao lower bound for regular and unbiased estimators of θ for this specific population. Is this estimator efficient?

Remark: The Cramer-Rao lower bound for regular and unbiased estimators of θ, obtained from a r.s. from the population under study is:

$$L_c = \frac{1}{n\mathrm{E}\left[\frac{\partial \ln f(X,\theta)}{\partial\theta}\right]^2} = \frac{1}{-n\mathrm{E}\left[\frac{\partial^2 \ln f(X,\theta)}{(\partial\theta)^2}\right]}$$

C. (10 points, 25 minutes)

A given firm manufacturing soft drinks wishes to estimate the mean energy expense per hour required for the bottling machine. In order to do so, a r.s. of 41 hours is taken, rendering a sample mean and variance of 20 euros and 25 euros^2, respectively. The electricity company suggests that the firm change its rate so that its energy expense can be reduced. In order to check that the energy expense has been reduced with the rate change, a new r.s. of 21 hours is taken, rendering a sample mean and variance of 18 euros and 16 euros^2, respectively. It is assumed that the energy expense per hour is normally distributed, and that there is independence between the energy expenses before and after the aforementioned rate change. In addition, is it also assumed that the population variance has not changed with the energy rate expense change.

i) Obtain the 95% confidence interval for the mean energy expense before the energy rate expense change.

ii) Obtain the 95% confidence interval for the difference of the mean energy expense before and after the rate change.

iii) The firm wishes to verify if the rate change has been indeed effective. In order to do so, it carries out the corresponding test of hypothesis, where the null hypothesis is that the mean expenses are the same against the alternative hypothesis that the initial mean expense is higher than the one after the rate change. At the 5% significance level, what will be the firm's decision? You should provide all relevant arguments supporting your conclusions.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

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

SOLUTIONS TO EXERCISES

Exercise A

It corresponds to a test of independence. The estimated probabilities $\hat{p}_i$ and $\hat{p}_j$ are computed from the information provided in the table. That is,

$$\hat{p}_S = \frac{80}{200} = 0.40 \qquad \hat{p}_L = \frac{120}{200} = 0.60$$

$$\hat{p}_{<2} = \frac{30}{200} = 0.15 \qquad \hat{p}_{[2,10]} = \frac{100}{200} = 0.50 \qquad \hat{p}_{>10} = \frac{70}{200} = 0.35$$

We are then able to build the table to carry out the corresponding test of independence:

Class	n_{ij}	$\hat{p}_{ii} = \hat{p}_i \hat{p}_j$	$n\hat{p}_{ij}$	$\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$
(S, <2)	20	$0.40 \cdot 0.15 = 0.06$	12	5.33
(S, $[2,10]$)	50	$0.40 \cdot 0.50 = 0.20$	40	2.50
(S, >10)	10	$0.40 \cdot 0.35 = 0.14$	28	11.57
(L, <2)	10	$0.60 \cdot 0.15 = 0.09$	18	3.56
(L, $[2,10]$)	50	$0.60 \cdot 0.50 = 0.30$	60	1.67
(L, >10)	60	$0.60 \cdot 0.35 = 0.21$	42	7.71
Total	$n = 200$	1	$n = 200$	$z = 32.34$

Under the null hypothesis, the test statistic $\sum_{i,j} \frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$ converges to a $\chi^2_{(k'-1)(k''-1)}$ distribution, where k' is the number of classes corresponding to the first variable, and k'' is the number of classes corresponding to the second variable.

In this case:

$$z = 32.34 > 9.21 = \chi^2_{(2-1)(3-1)\,,\,0.01} = \chi^2_{2,0.01},$$

so that, at the 1% significance level, we reject the null hypothesis of independence between the firms' size and profit level.

Exercise B

$$f(x;\theta) = \frac{x^4\, e^{-\frac{x}{\theta}}}{24\,\theta^5}, \qquad x > 0, \qquad \theta > 0$$

i) Maximum likelihood estimator

$$L(\vec{x};\theta) = f(x_1;\theta)\ldots f(x_n;\theta) = \frac{x_1^4\, e^{-\frac{x_1}{\theta}}}{24\,\theta^5} \cdots \frac{x_n^4\, e^{-\frac{x_n}{\theta}}}{24\,\theta^5} =$$

$$= \frac{\left(\prod_{i=1}^n x_i^4\right)\; e^{-\frac{\sum_{i=1}^n x_i}{\theta}}}{24^n\,\theta^{5n}}$$

$$\ln L(\vec{x};\theta) = \ln\left(\prod_{i=1}^n x_i^4\right) - \frac{\sum_{i=1}^n x_i}{\theta} - n\ln 24 - 5n\ln\theta$$

$$\frac{\partial \ln L(\vec{x},\theta)}{\partial \theta} = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{5n}{\theta} = 0$$

$$\sum_{i=1}^{n} x_i - 5n\theta = 0$$

$$\hat{\theta}_{\text{ML}} = \frac{\sum_{i=1}^{n} X_i}{5n} = \frac{\overline{X}}{5}$$

ii) Unbiasedness

$\mathrm{E}(\hat{\theta}_{MV}) = \mathrm{E}\left(\frac{\overline{X}}{5}\right) = \frac{1}{5}\mathrm{E}(\overline{X}) = \frac{1}{5}\mathrm{E}(X) = \frac{1}{5}\ (5\theta) = \theta.$

Therefore, $\hat{\theta}_{\text{ML}}$ is an unbiased estimator of θ.

Consistency

We compute the maximum likelihood estimator's variance.

$$\mathrm{Var}(\hat{\theta}_{\text{ML}}) = \frac{1}{5^2}\mathrm{Var}(\overline{X}) = \frac{1}{5^2}\frac{\mathrm{Var}(X)}{n} = \frac{1}{5^2}\left(\frac{5\theta^2}{n}\right) = \frac{\theta^2}{5n}$$

Given that $\hat{\theta}_{\text{ML}}$ is an unbiased estimator of θ, and that its variance tends to zero as n goes to infinity, we can state that $\hat{\theta}_{\text{ML}}$ is a consistent estimator of θ.

iii) Efficiency

In order to be able to verify if $\hat{\theta}_{\text{ML}}$ is an efficient estimator of θ, we compute the Cramer-Rao lower bound for regular and unbiased estimators.

$$L_c = \frac{1}{n\mathrm{E}\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2}$$

$$\ln f(x;\theta) = 4\ln x - \frac{x}{\theta} - \ln 24 - 5\ln\theta$$

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = \frac{1}{\theta^2}x - \frac{5}{\theta} = \frac{1}{\theta^2}\ (x - 5\theta)$$

$$\mathrm{E}\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2 = \mathrm{E}\left[\frac{1}{\theta^2}\ (X - 5\theta)\right]^2 = \frac{1}{\theta^4}\ \mathrm{E}(X-5\theta)^2 = \frac{1}{\theta^4}\ \mathrm{Var}(X) = \frac{1}{\theta^4}\ \left(5\theta^2\right) = \frac{5}{\theta^2}$$

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

$$L_c = \frac{1}{n\left(\frac{5}{\theta^2}\right)} = \frac{\theta^2}{5n} = \mathrm{Var}(\hat{\theta}_{\text{ML}})$$

Therefore, $\hat{\theta}_{\text{ML}}$ is an efficient estimator of θ

Exercise C

$X \in N(m_1, \sigma_1^2)$ $\quad n_1 = 41 \quad \bar{x} = 20 \quad s_1^2 = 25$

$Y \in N(m_2, \sigma_2^2)$ $\quad n_2 = 21 \quad \bar{y} = 18 \quad s_2^2 = 16$

In addition, we also assume that $\sigma_1^2 = \sigma_2^2$.

i)

$$\mathrm{CI}_{1-\alpha}(m_1) = \left(\overline{x} \pm t_{\overline{n-1|\frac{\alpha}{2}}} \frac{s_1}{\sqrt{n_1 - 1}}\right)$$

$$t_{\overline{40|\frac{0.05}{2}}} = 2.02$$

$$\mathrm{CI}_{0.95}(m_1) = \left(20 \pm 2.02 \cdot \sqrt{\frac{25}{40}}\right) =$$

$$=(18.4030 \ , \ 21.597)$$

ii)

$$\mathrm{CI}_{1-\alpha}(m_1 - m_2) = \left(\overline{x} - \overline{y} \pm t_{\overline{n_1+n_2-2|\frac{\alpha}{2}}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\right)$$

$$t_{\overline{60|\frac{0.05}{2}}} = 2.00$$

$$\mathrm{CI}_{0.95}(m_1 - m_2) = \left(20 - 18 \pm 2.00 \sqrt{\frac{1}{41} + \frac{1}{21}} \sqrt{\frac{41 \cdot 25 + 21 \cdot 16}{41 + 21 - 2}}\right) =$$

$$=(-0.5561 \ , \ 4.5561)$$

iii) We have to test the hypotheses:

$H_0 : m_1 = m_2 \quad \equiv \quad m_1 - m_2 = 0$

$H_1 : m_1 > m_2 \quad \equiv \quad m_1 - m_2 > 0$

Under H_0 :

$$\frac{(\bar{X} - \bar{Y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} \in t_{\overline{n_1+n_2-2|}}$$

The decision rule for this test, at the α significance level, rejects the null hypothesis if:

$$\frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} > t_{\overline{n_1+n_2-2|\alpha}}$$

In this case, we have that:

$$\frac{(20 - 18) - 0}{\sqrt{\frac{1}{41} + \frac{1}{21}} \sqrt{\frac{41 \cdot 25 + 21 \cdot 16}{41 + 21 - 2}}} = 1.5649 \ < \ 1.67 = t_{\overline{60|0.05}}$$

Therefore, at the 5% significance level, we do not reject the null hypothesis that the mean gas expenses are the same, or that the mean energy expense has not changed with the energy rate expense change.