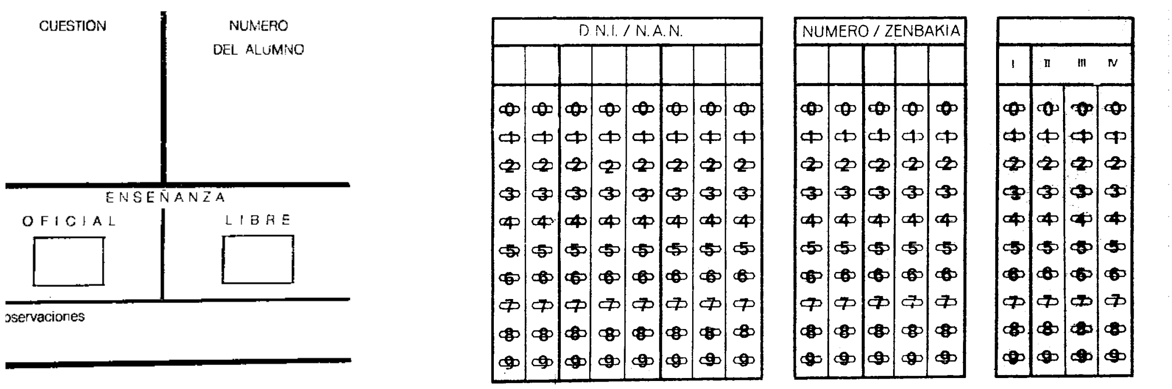
BUSINESS STATISTICS - Second Year Academic Year 2009-10
June 16, 2010

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, just as is illustrated in the example.
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. However, exams having a multiple choice question part with grades greater than or equal to 14 could, under special circumstances, be compensated with a good grade in the exercises part of the exam.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises. In "Resit" (column II) you will write the number of times you have registered for an exam in this course *not including this one.*

Example:

$\underbrace{\text{12545}}$ PEREZ, Ernesto Exam type $\underbrace{0}$ Resit



MULTIPLE CHOICE QUESTIONS (Time: 1 hour and 40 minutes)

1. FREE-QUESTION. The capital of Spain is:

   (A) Paris (B) Sebastopol (C) Madrid (D) London (E) Pekin

Questions 2 to 6 refer to the following exercise:

The percentage of clients of a given bank that shows interest for a new product after an advertising campaign is 40%. We assume independence between the different clients.

2. If 7 clients are randomly selected, the probability that exactly 2 of them show interest for this product is:

   (A) 0.0124 (B) 0.0036 (C) 0.0940 (D) 0.0774 (E) 0.2613

3. If 20 clients are randomly selected, the probability that less than 10 of them show interest for this product is:

   (A) 0.8725 (B) 0.2403 (C) 0.1597 (D) 0.7553 (E) 0.1275

4. If 20 clients are randomly selected, the probability that less than 15 of them show **no** interest for this product is:

   (A) 0.1256 (B) 0.9997 (C) 0.0016 (D) 0.8744 (E) 0.7984

5. If the bank has 36000 clients, how many clients are expected to show interest for this new product?

   (A) 144 (B) 86.4 (C) 8640 (D) 14400 (E) 5760

6. If the bank has 36000 clients, the approximate probability that more than 14500 of them show interest for this product is:

   (A) 0.8599 (B) 0.8632 (C) 0.1703 (D) 0.1401 (E) 0.1368

Questions 7 to 10 refer to the following exercise:

The number of bicycles borrowed during a quarter of an hour period by a given city council follows a Poisson distribution with variance equal to 1.75. We assume independence between the borrowing of bicycles occurring at different periods.

7. The probability that during a quarter of an hour period exactly 2 bicycles are borrowed is:

   (A) 0.1308 (B) 0.4145 (C) 0.5322 (D) 0.2661 (E) 0.2072

8. The most likely number of bicycles to be borrowed during a one hour period is:

   (A) 6 (B) 7 and 8 (C) 7 (D) 8 (E) 6 and 7

9. The probability that during a one hour period less than 5 bicycles are borrowed is:

   (A) 0.8088 (B) 0.1730 (C) 0.1912 (D) 0.8270 (E) 0.3134

10. The approximate probability that during a five hour period less than 35 bicycles are borrowed is:

    (A) 0.5319 (B) 0.4681 (C) 0.1701 (D) 0.4361 (E) 0.5639

11. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables with probability mass function given by:

$$P_n(x) = \begin{cases} 1 - \frac{2+5n}{10n^2} & \text{if } x = 1 \\ \frac{1}{2n} & \text{if } x = 3 \\ \frac{1}{5n^2} & \text{if } x = 5 \end{cases}$$

The sequence will converge:

(A) In distribution and probability to $X = 1$

(B) Only in distribution to $X = 3$

(C) Only in probability to $X = 3$

(D) In distribution and probability to $X = 3$

(E) Only in distribution to $X = 1$

12. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables with characteristic function given by $\psi_n(u) = \frac{n^2}{n^2-iu}$. This sequence of r.v. will converge:

(A) In distribution and probability to $X = 0$

(B) Only in probability to $X = 0$

(C) Only in distribution to $X = 0$

(D) Only in distribution to $X = 4$

(E) In distribution and probability to $X = 4$

Questions 13 to 15 refer to the following exercise:

The duration, in minutes, of the phone calls for a given individual is a random variable, X, having a uniform distribution on the interval $(0, 10)$

13. The probability that a given call lasts between 2 and 5 minutes is:

(A) 0.3 (B) 0.5 (C) 0.7 (D) 0.4 (E) 0.6

14. If during a one month period the individual makes 80 independent phone calls, the approximate distribution of the total duration of his/her phone calls is:

(A) $U(0, 800)$ (B) exp(1/400) (C) P(400) (D) $N(400,\ 666.67)$ (E) $U(0, 80)$

15. If each call has an associated cost of 0.2 euros for the call establishment and of an additional cost of 0.1 euros per minute, the corresponding mean price, in euros, for the 80 phone calls will be:

(A) 56 (B) 40.2 (C) 8.2 (D) 28 (E) 48

16. If X is a random variable having a $\gamma(4, 2)$ distribution, then its characteristic function is:

(A) $(1-4iu)^{-2}$ (B) $\left(1-\frac{iu}{2}\right)^{-4}$ (C) All false (D) $\left(1-\frac{iu}{4}\right)^{-2}$ (E) $(1-2iu)^{-4}$

17. Let X be a random variable having a $\gamma(1, 1)$ distribution. Then, $P(X > 2)$ is:

(A) 0.2706 (B) 0.1353 (C) 0.8647 (D) 0.4212 (E) 0.5788

Questions 18 and 19 refer to the following exercise:

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

$$P(X=-2)=\theta; \qquad P(X=0)=1-2\theta; \qquad P(X=2)=\theta.$$

In order to estimate the parameter θ, a random sample of size $n=10$ has been taken, providing the following results: -2 appeared five times, 0 appeared three times and 2 appeared two times.

18. The maximum likelihood estimate of θ is:

(A) 0.64 (B) 0.39 (C) 0.35 (D) 0.72 (E) 0.41

19. The method of moments estimate of θ is:

(A) 0.41 (B) 0.39 (C) 0.64 (D) 0.72 (E) 0.35

Questions 20 and 21 refer to the following exercise:

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

$$f(x;\theta)=\frac{1}{2\theta}e^{-\frac{1}{2\theta}x}, \quad x>0, \ \theta>0,$$

and a mean equal to 2θ. In order to estimate the parameter θ, a random sample of size n, $X_1,\dots,X_n$, has been taken.

20. The maximum likelihood estimator of θ is:

(A) $\sum_{i=1}^n \frac{X_i}{n}$ (B) $\sum_{i=1}^n \frac{X_i}{2n}$ (C) $\sum_{i=1}^n 2X_i$ (D) $\sum_{i=1}^n \frac{2X_i}{n}$ (E) $\sum_{i=1}^n \frac{X_i}{2}$

21. The method of moments estimator of θ is:

(A) $\sum_{i=1}^n \frac{X_i}{2}$ (B) $\sum_{i=1}^n \frac{X_i}{2n}$ (C) $\sum_{i=1}^n \frac{X_i}{n}$ (D) $\sum_{i=1}^n \frac{2X_i}{n}$ (E) $\sum_{i=1}^n 2X_i$

Questions 22 and 23 refer to the following exercise:

Let X be a r.v. having a $N(m,\sigma^2=9)$ distribution from which a r.s. of size $n=100$, $X_1,\dots,X_{100}$ has been taken. We propose two estimators for the parameter m: $\hat{m}_1=\frac{X_1+\ldots+X_{100}}{100}=\overline{X}$ and $\hat{m}_2=3\overline{X}$.

22. Are they unbiased estimators for m?

(A) Only $\hat{m}_1$ (B) Both (C) - (D) None (E) Only $\hat{m}_2$

23. The corresponding variances for $\hat{m}_1$ and $\hat{m}_2$ are, respectively:

(A) 0.09 and 0.81 (B) 9 and 81 (C) 0.09 and 0.27 (D) 9 and 27 (E) All false

24. Let X be a random variable having an $F_{3,5}$ distribution. Then, $P(X>0.1883)$ is:

(A) 0.90 (B) 0.80 (C) 0.10 (D) 0.95 (E) 0.05

Questions 25 and 26 refer to the following exercise:

Let X be a r.v. having a Poisson distribution with parameter θ. We wish to test the null hypothesis $H_0:\theta=3$ against the alternative hypothesis $H_1:\theta=4$. In order to do so, a random sample of size 2, X_1,X_2, has been taken. We decide to reject the null hypothesis if $X_1+X_2<3$.

25. The type I error probability for this test is:

(A) 0.4232 (B) 0.9380 (C) 0.0620 (D) 0.1512 (E) 0.5768

26. The type II error probability for this test is:

(A) 0.9862 (B) 0.2381 (C) 0.7619 (D) 0.0138 (E) 0.4335

27. We wish to estimate the proportion of individuals that have seen a given advertising campaign. In order to do so, a random sample of 2000 individuals has been taken, and 72% of them have seen the advertising. An approximate 0.95 confidence interval for the aforementioned proportion is:

(A) $(0.70, 0.74)$ (B) $(0.028, 0.044)$ (C) $(0.32, 0.82)$ (D) $(0.13, 0.16)$ (E) $(0.61, 0.83)$

28. Let X be a random variable having a $N(m, \sigma^2)$ distribution. A random sample of size 15 has been taken, providing a sample variance equal to 3.4. A 0.95 confidence interval for the population variance is:

(A) (1.95, 9.06) (B) (1.98, 11.96) (C) (1.57, 10.72) (D) (2.35, 7.96) (E) (2.04, 7.02)

Questions 29 and 30 refer to the following exercise:

We wish to test if the game of chance in which an individual spends his/her money is related to his/her age. In order to do so, **a r.s.** has been taken and individuals are classified according to their age ($\leq 30, 30-50, \geq 50$) and the type of game of chance they spend their money on (weekly state-run lottery–lotería primitiva, national-run lottery–lotería nacional, football pools coupon–quiniela, bingo).

29. The type of test that one should carry out is:

(A) Independence

(B) Goodness-of-fit to a partially specified distribution

(C) Homogeneity

(D) Goodness-of-fit to a totally specified distribution

(E) -

30. The number of degrees of freedom that the distribution of the test statistic used for this specific test is:

(A) 6 (B) 4 (C) 12 (D) 3 (E) 11

EXERCISES (Time: 75 minutes)

A (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a r.s. of size n taken from a r.v. X such that $X \in \gamma(\frac{1}{\theta}, 3)$ distribution. That is, its probability density function is given by:

$$f(x;\theta) = \frac{1}{2\,\theta^3}\, x^2\, e^{-\frac{x}{\theta}}, \qquad x > 0, \quad \theta > 0$$

i) Find the maximum likelihood estimator of the parameter θ.

ii) Find the method of moments estimator of the parameter θ.

iii) Is the method of moments estimator an unbiased estimator?, Is it consistent?, Is it efficient?

Remark: Remember that the Cramer-Rao lower bound for an unbiased estimator of θ, based on a r.s. of size n, is given by:

$$L_C = \frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$$

B. (10 points, 25 minutes)

On January 2, 2010, a newspaper article was published and it included the indexes of audience for the TV shows broadcasted during the New Year's evening. The article claimed that, from those watching these TV shows, 50.8% watched them in TVE1, 21.2% watched them in Tele 5, 12.9% watched them in the local autonomous TV channels, 3.3% watched them in TVE2, 3% watched them in Antena 3, 2% watched them in La sexta, 1.3% watched them in Cuatro and 5.5% watched them in the remaining TV channels.

In order to be able to test that the data provided in the newspaper article are true, a sample of 500 randomly selected individuals from those who watched these TV shows was taken, providing the following results: 200 of them watched them in TVE1, 110 watched them in Tele 5, 70 watched them in the local autonomous TV channels, 15 watched them in TVE2, 20 watched them in Antena 3, 15 watched them in La sexta, 50 watched them in Cuatro and 20 watched them in the remaining TV channels.

At the 5% significance level, test the hypothesis that the newspaper article provides an accurate information about the real indexes of audience for these TV shows.

C (10 points, 25 minutes) A driver used to check his/her car's mean gas consumption (in liters per 100 Km) on a weekly basis. One day he/she goes to a course on efficient driving and he/she wishes to know if his/her car's gas consumption has really decreased after attending this course. In order to do so, he continues to check on his/her car's mean gas consumption. He/she has data on a 25-week period before the course, providing a mean consumption of 8.4 liters per 100 Km and a standard deviation of 2.5 liters. In addition, he/she also has data on a 16-week period after the course, providing a mean consumption of 6.0 liters per 100 Km and a standard deviation of 2.1 liters.

Remark: We assume that the weekly data, before and after the course, are independent and that their consumption distributions are normal.

i) At the 10% significance level, test the hypothesis of equal variances for the mean gas consumptions after and before the course.

ii) At the 10% significance level, test the hypothesis that the mean gas consumption has not changed after the course versus the alternative hypothesis that it has really decreased after the course.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A)

Let $X_1, \dots, X_n$ be a r.s. of size n taken from a r.v. X that follows a $\gamma(\frac{1}{\theta}, 3)$ distribution. That is, with probability density function given by:

$$f(x;\theta) = \frac{1}{2\,\theta^3}\, x^2\, e^{-\frac{x}{\theta}}, \qquad x > 0, \quad \theta > 0$$

i) Therefore, the likelihood function will be given by

$$L(\vec{x};\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$

$$L(\vec{x};\theta) = \left(\frac{1}{2\,\theta^3}\, x_1^2\, e^{-\frac{x_1}{\theta}}\right)\left(\frac{1}{2\,\theta^3}\, x_2^2\, e^{-\frac{x_2}{\theta}}\right) \cdots \left(\frac{1}{2\,\theta^3}\, x_n^2\, e^{-\frac{x_n}{\theta}}\right)$$

$$L(\vec{x};\theta) = \frac{1}{2^n\,\theta^{3n}} \left(\prod_{i=1}^{n} x_i^2\right) e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$$

The maximum likelihood estimator is the value of θ that maximizes the likelihood function or, equivalently, its natural logarithm:

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

If we take derivatives with respect to θ, we have that:

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

Therefore,

$$\frac{3n\theta}{\theta^2} = \frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

$$\Rightarrow \qquad \hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^{n} X_i}{3n} = \frac{\overline{X}}{3}$$

ii) In order to be able to compute the method of moments estimator of θ, we have to find $\alpha_1 = \mathrm{E}(X)$. In this case, and given that X is a r.v. following a gamma distribution: $\gamma(a = \frac{1}{\theta}, r = 3)$, we have that

$$\alpha_1 = \mathrm{E}(X) = \frac{r}{a} = \frac{3}{(1/\theta)} = 3\theta$$

In this way, after we make the sample and population moments equal to each other; that is, $\alpha_1 = a_1$; we have that:

$$\alpha_1 = \mathrm{E}(X) = 3\theta = \overline{x} = a_1,$$

$$\Rightarrow \hat{\theta}_{\mathrm{MM}} = \frac{\overline{X}}{3}$$

iii) Unbiasedness:

In order to check if the method of moments estimator is unbiased we have to verify if $\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \theta$.

In this case,

$$\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) \;=\; \mathrm{E}\left(\frac{\overline{X}}{3}\right) \;=\; \frac{1}{3}\,\mathrm{E}(\overline{X}) \;=\; \frac{1}{3}\,\mathrm{E}(X) \;=\; \frac{1}{3}\,(3\,\theta) \;=\; \theta$$

Therefore, the method of moments estimator is unbiased.

Consistency

In order to see if the method of moments estimator is consistent we compute its variance.

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) \;=\; \mathrm{Var}\left(\frac{\overline{X}}{3}\right) \;=\; \frac{1}{9}\,\mathrm{Var}(\overline{X}) \;=\; \frac{1}{9}\left[\frac{\mathrm{Var}(X)}{n}\right]$$

We need to compute $\mathrm{Var}(X)$. In this case, and given that X is a r.v. following a gamma distribution, we have that:

$$\mathrm{Var}(X) = \frac{r}{a^2} = \frac{3}{(1/\theta^2)} = 3\theta^2$$

Therefore,

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) \;=\; \frac{3\theta^2}{9n} \;=\; \frac{\theta^2}{3n}$$

As the method of moments estimator is unbiased and, in addition, its variance tends to zero as n goes to infinity, the two sufficient conditions for consistency hold and, thus, we conclude that $\hat{\theta}_{\mathrm{MM}}$ is a consistent estimator of θ.

Efficiency:

In order to check if the method of moments estimator is efficient, we need to verify if its variance coincides with the Cramer-Rao lower bound. We now compute the Cramer-Rao lower bound for regular and unbiased estimators of θ:

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

In this case,

$$\ln f(x;\theta) = -\ln 2 - 3\ln\theta + 2\ln x - \frac{x}{\theta}$$

$$\frac{\partial f(x;\theta)}{\partial\theta} = -\frac{3}{\theta} + \frac{x}{\theta^2} \;=\; \frac{1}{\theta^2}\,(x - 3\theta)$$

$$\mathrm{E}\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 = \mathrm{E}\left(\frac{X-3\theta}{\theta^2}\right)^2 = \frac{\mathrm{E}\,(X-3\theta)^2}{\theta^4} = \frac{\mathrm{Var}(X)}{\theta^4} \;=\frac{1}{\theta^4}\;\left(3\theta^2\right) = \frac{3}{\theta^2}$$

Replacing this expected value into the formula for the Cramer-Rao lower bound, we have that:

$$L_c = \frac{1}{n\left(\frac{3}{\theta^2}\right)} = \frac{\theta^2}{3n} = \mathrm{Var}(\hat{\theta}_{\mathrm{MM}})$$

Therefore, as the estimator's variance is equal to the Cramer-Rao lower bound, we conclude that the method of moments estimator is efficient.

Exercise B)

This corresponds to a goodness of fit test to a completely specified distribution. In order to carry out this test, we need to build the following table:

	p_i	n_i	np_i	$(n_i - np_i)$	$\frac{(n_i-np_i)^2}{np_i}$
TVE1	0.508	200	254	-54	11.4803
Tele 5	0.212	110	106	+4	0.1509
Auto. TV	0.129	70	64.5	+5.5	0.4690
TVE2	0.033	15	16.5	-1.5	0.1364
Antena 3	0.030	20	15	+5	1.6667
La Sexta	0.020	15	10	+5	2.5000
Cuatro	0.013	50	6.5	+43.5	291.1154
Other Chan.	0.055	20	27.5	-7.5	2.0455
	1	$n = 500$	1	0	$z = 309.5642$

Under the null hypothesis for the goodness of fit test to a completely specified distribution, the test statistic

$\sum_i \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{(K-1)}$, where K is the number of categories into which the sample has been divided (i.e., $K = 8$).

The decision rule suggests that, at the approximate 5% significance level, the null hypothesis should be rejected if:

$$z > \chi^2_{(8-1)\,,\,0.05} = \chi^2_{7,0.05}.$$

In this case:

$$z = 309.5642 > 14.1 = \chi^2_{7,0.05},$$

so that, at the 5% significance level, the null hypothesis of goodness of fit to a completely specified distribution is rejected. That is, we can conclude that the data in the sample does not follow the distribution reported in the newspaper article for indexes of audience for the TV shows broadcasted during the New Years' evening.

Exercise C)

Let X and Y be the two r.v. representing the gas consumption before and after the efficient driving course, respectively. The two r.v. are independent $X \sim N(m_X, \sigma_X^2)$ and $Y \sim N(m_Y, \sigma_Y^2)$. Two random samples of sizes n_1 and n_2, $\vec{X} = (X_1, X_2, \ldots, X_{n_1})$ and $\vec{Y} = (Y_1, Y_2, \ldots, Y_{n_2})$ have been taken. The sample data are as follows:

$$n_1 = 25 \qquad n_2 = 16$$
$$\overline{x} = 8.4 \qquad \overline{y} = 6$$
$$s_X = 2.5 \qquad s_Y = 2.1$$

i) At the 10% significance level, test the hypothesis of equal variances for the mean gas consumptions after and before the course.

$$H_0 : \sigma_X^2 = \sigma_Y^2 \quad \Rightarrow \quad \frac{\sigma_X^2}{\sigma_Y^2} = 1$$

$$H_1 : \sigma_X^2 \neq \sigma_Y^2 \quad \Rightarrow \quad \frac{\sigma_X^2}{\sigma_Y^2} \neq 1$$

It is known that:

$$\frac{n_1 S_X^2}{\sigma_X^2} \in \chi^2_{n_1-1} \qquad \frac{n_2 S_Y^2}{\sigma_Y^2} \in \chi^2_{n_2-1},$$

so that:

$$\frac{\left(n_1 S_X^2/\sigma_X^2\right)/(n_1-1)}{\left(n_2 S_Y^2/\sigma_Y^2\right)/(n_2-1)} = \frac{n_1(n_2-1)S_X^2}{n_2(n_1-1)S_Y^2}\,\frac{1}{\underbrace{\left(\sigma_X^2/\sigma_Y^2\right)}_{=1 \text{ under } H_0}} \in \mathcal{F}_{n_1-1,n_2-1}$$

$$\text{CR} = \left\{\frac{s_X^2}{s_Y^2} \quad \text{such that} \quad \frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} \leq F_{\overline{n_1-1,n_2-1}|1-\alpha/2} \quad \text{or} \quad \frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} \geq F_{\overline{n_1-1,n_2-1}|\alpha/2}\right\}$$

In our case:

$$F_{\overline{n_1-1,n_2-1}|\alpha/2} = F_{\overline{25-1,16-1}|0.05} = 2.29$$

$$F_{\overline{n_1-1,n_2-1}|1-\alpha/2} = F_{\overline{25-1,16-1}|0.95} = \frac{1}{F_{\overline{15,24}|0.05}} = \frac{1}{2.11} = 0.4739$$

$$\frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} = \frac{25\,(16-1)\,(2.5)^2}{16\,(25-1)\,(2.1)^2} = \frac{2343.75}{1693.44} = 1.3840$$

Given that:

$$0.4379 \leq 1.3840 \leq 2.29,$$

at the 10% significance level, we do not reject the null hypothesis of equal variances for the gas consumptions after and before the course

ii) At the 10% significance level, test the hypothesis that the mean gas consumption has not changed after the course versus the alternative hypothesis that it has really decreased after the course.

Given that the result obtained in the previous item, we confirm that we have a test for independent normal distributions with equal unknown variances, so that

$$\frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} \in N(0,1)$$

$$\frac{n_1 S_X^2}{\sigma^2} \in \chi^2_{n_1-1} \qquad \frac{n_2 S_Y^2}{\sigma^2} \in \chi^2_{n_2-1} \qquad \Rightarrow \qquad \left(\frac{n_1 S_X^2 + n_2 S_Y^2}{\sigma^2}\right) \in \chi^2_{n_1+n_2-2}$$

$$\frac{\frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}}{\sqrt{\frac{\left(\frac{n_1 S_X^2+n_2 S_Y^2}{\sigma^2}\right)}{n_1+n_2-2}}} = \frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1 S_X^2+n_2 S_Y^2}{n_1+n_2-2}}} \in t_{n_1+n_2-2}$$

$H_0 : m_X = m_Y \qquad \equiv \qquad m_X - m_Y = 0$

$H_1 : m_X > m_Y \qquad \equiv \qquad m_X - m_Y > 0$

Under H_0 :

$$\frac{(\bar{X} - \bar{Y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{\frac{n_1 S_X^2 + n_2 S_Y^2}{n_1+n_2-2}}} \in t_{n_1+n_2-2}$$

$$\text{RC} = \left\{ \bar{x} \text{ e } \bar{y} \quad \text{such that} \quad \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{\frac{n_1 s_X^2 + n_2 s_Y^2}{n_1+n_2-2}}} \geq t_{\overline{n_1+n_2-2}|\alpha} \right\}$$

In this case

$t_{\overline{n_1+n_2-2}|\alpha} \; = \; t_{\overline{25+16-2}|0.10} \; = \; t_{\overline{39}|0.10} = 1.30$, and

$\frac{(8.4-6)-0}{\sqrt{\frac{1}{25}+\frac{1}{16}}\sqrt{\frac{25\,(2.5)^2+16\,(2.1)^2}{25+16-2}}} = 3.1085$

Given that:

$3.1085 \geq 1.30,$

at the 10% significance level, we reject the null hypothesis of equal means and, therefore, we conclude that the efficient driving course has decreased this driver's mean gas consumption.