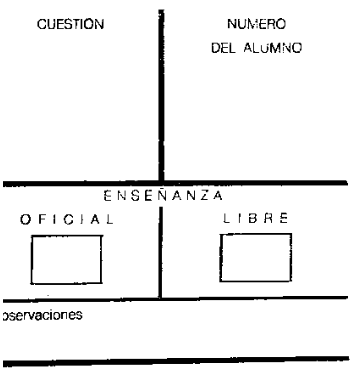
BUSINESS STATISTICS - Second Year Academic Year 2007-08
June 5, 2008

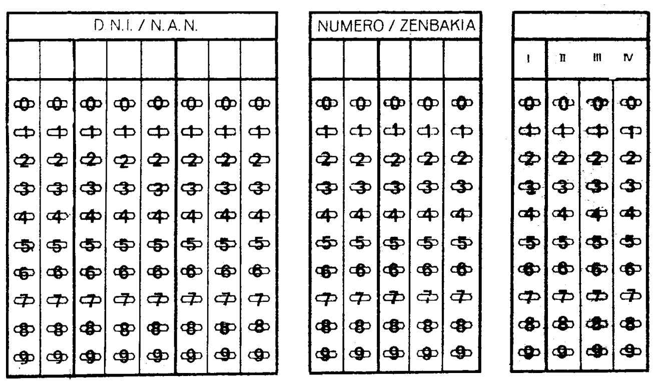
INSTRUCTIONS

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:

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





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 length (in centimeters) of the parts produced by a given factory follows a $N(10, \sigma^2 = 4)$ distribution. We consider as acceptable only those parts whose length belongs to the interval (8,12) centimeters.

2. If in a given period 100 parts have been produced by the factory, the exact distribution of the number of acceptable parts is:

   (A) $b(0.6826, 100)$ (B) $\mathcal{P}(\lambda = 68.26)$ (C) All false (D) $N(1000, \sigma^2 = 400)$ (E) $b(0.3174, 100)$

3. In that same period of time, the expected number of acceptable parts produced by the factory is:

   (A) 31.74 (B) 50.34 (C) 68.26 (D) 84.13 (E) 15.87

4. The approximate probability that, in that same period of time, more than 20 parts are **rejected** is:

   (A) 0.01 (B) 0.50 (C) 0.76 (D) 0.99 (E) 0.24

Questions 5 to 8 refer to the following exercise:

The number of clients that arrive per minute at a given gas station follows a Poisson distribution of mean equal to 0.80. We assume independence.

5. The mode(s) for this distribution is (are):

   (A) 0 and 1 (B) 1 (C) 0 (D) 1 y 2 (E) 2 y 3

6. The probability that more than two clients arrive at the gas station in a given minute is:

   (A) 0.0474 (B) 0.1912 (C) 0.4493 (D) 0.8088 (E) 0.9526

7. The probability that at most 5 clients arrive at the gas station in a 10-minute period is:

   (A) 0.8088 (B) 0.3841 (C) 0.0996 (D) 0.1912 (E) 0.6159

8. The approximate probability that at most 50 clients arrive at the gas station in a one-hour period is:

   (A) 0.6406 (B) 0.8106 (C) 0.9772 (D) 0.1894 (E) 0.3594

9. Let X and Y be two independent r.v. such that $X \in \gamma(a, 1)$ and $Y \in \gamma(b, 1)$. The distribution of the r.v. $X + Y$ is:

   (A) $\gamma(a + b, 1)$ (B) χ^2_1 (C) χ^2_{a+b} (D) All false (E) $\gamma(a, 2)$, if $a = b$

10. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables defined as:

$$X_n = \begin{cases} 2, & \text{with probability } \left(1-\frac{1}{n}\right) \\ 0, & \text{with probability } \frac{1}{n} \end{cases}$$

The sequence will converge:

(A) Only in distribution and probability to $X = 2$

(B) Only on distribution and probability to $X = 0$

(C) In distribution, probability and quadratic mean to $X = 2$

(D) In distribution, probability and quadratic mean to $X = 0$

(E) All false

11. Let $\{X_n\}_{n\in\mathcal{N}}$ be a sequence of random variables having a $N(1, \sigma^2 = 1/n)$ distribution. It is known that the characteristic function of a $N(m, \sigma^2)$ r.v. is given by $\psi_n(u) = e^{ium-\frac{\sigma^2 u^2}{2}}$. The sequence will converge:

(A) In distribution to a $N(1, 1)$ r.v.

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

(C) In distribution to a $N(1, 2)$ r.v.

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

(E) All false

12. Let X and Y be independent and identically r.v. having a $N(0, 1)$ distribution. Let $W = Y^2$ and $Z = \frac{X}{\sqrt{W}}$. The distribution of the r.v. Z is:

(A) χ_1^2 (B) $N(0, 1)$ (C) $Z \equiv 1$ (D) $\mathcal{F}_{1,1}$ (E) t_1

13. Let X be a r.v. having a t_n distribution; that is, a Student's t distribution with n degrees of freedom. Then, we have that:

(A) $t_{n,\alpha} > t_{n,\frac{\alpha}{4}}$ (B) $t_{n,\alpha} = -t_{n,1-\alpha}$ (C) $t_{n,\alpha} > t_{n,\frac{\alpha}{2}}$ (D) $t_{n,\frac{\alpha}{2}} > t_{n,\frac{\alpha}{4}}$ (E) $t_{n,\alpha} = t_{n,1-\alpha}$

14. Let X be a random variable having a Poisson distribution with parameter λ. In order to estimate the parameter λ, a r.s. of size n, $X_1, \ldots, X_n$, has been taken. A sufficient statistics for the parameter λ is:

(A) $\sum_{i=1}^n X_i$ (B) $\prod_{i=1}^n X_i$ (C) $\prod_{i=1}^n X_i!$ (D) $\prod_{i=1}^n \left(\frac{1}{X_i}\right)$ (E) $\sum_{i=1}^n X_i!$

Questions 15 and 16 refer to the following exercise:

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

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

In order to estimate the parameter θ, a r.s. of size n has been taken, for which four zeroes were obtained.

15. The maximum likelihood estimate of θ is:

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

16. The method of moments estimate of θ is:

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

Questions 17 to 20 refer to the following exercise:

We know that when we sample from a normal population we have that $\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$, where S^2 is the sample variance, and that the mean and the variance of a χ^2_p r.v. are p and $2p$, respectively.

17. The mean or expected value of S^2, $\mathrm{E}(S^2)$, is:

(A) σ^2 (B) All false (C) $\frac{(n-1)\sigma^2}{n}$ (D) $\frac{n\sigma^2}{(n-1)}$ (E) $\frac{n(n-1)}{\sigma^2}$

18. Is S^2 an unbiased estimator of σ^2?

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

19. The variance of S^2 is:

(A) $\frac{2n(n-1)}{\sigma^2}$ (B) All false (C) $\frac{2(n-1)\sigma^4}{n^2}$ (D) $\frac{2(n-1)\sigma^2}{n}$ (E) $\frac{2(n-1)n^2}{\sigma^4}$

20. Is S^2 a consistent estimator of σ^2?

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

Questions 21 to 23 refer to the following exercise:

We wish to test the null hypothesis that the probability density function of a given population is $\gamma(a = 5, r)$, against the alternative hypothesis that it is $\gamma(a = 2, r)$; that is, the parameter r is common to both distributions. We recall that the probability density function for a $\gamma(a, r)$ distribution is given by:

$$f(x; a, r) = \frac{a^r}{\Gamma(r)}\; x^{r-1} e^{-ax}, \quad x > 0, \; a, r > 0$$

In order to carry out this test, a random sample of size $n = 1$ has been taken from that population (that is, we observe X).

21. The most powerful critical region for X is of the form:

(A) $[K_1, K_2]$ (B) $[0, K]$ (C) $[K, +\infty]$ (D) $[K_1, K_2]^c$ (E) All false

22. If $r = 1$ and $x = 0.40$, what would be the decision at the 0.05 significance level?

(A) - (B) Do not reject H_0 (C) Reject H_0 (D) - (E) -

23. What would be the approximate power for this specific case and specific significance level?

(A) 0.70 (B) 0.60 (C) 0.30 (D) 0.40 (E) 0.95

Questions 24 and 25 refer to the following exercise:

Let X be a r.v. having a $U[\theta, 5]$ distribution. In order to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$, a random sample of size $n = 1$ has been taken and we reject H_0 if the sample observation is larger than or equal to 2.

24. If $\theta_0 = 0$, the significance level for this test is:

(A) 0.20 (B) All false (C) 0.60 (D) 0.05 (E) 0.40

25. If $\theta_1 = 1$, the power of the test is:

(A) 0.60 (B) All false (C) 0.25 (D) 0.75 (E) 0.40

Questions 26 and 27 refer to the following exercise:

We wish to test if a given coin has equal probability of obtaining heads and tails (legal or unbiased coin) or if, otherwise, the coin is biased, in which case we know that $P(\text{heads}) = 0.70$. In order to do so, we throw the coin 10 times.

26. If we let Z be the number of heads obtained after throwing the coin 10 times, at the 5% significance level, the decision will be to reject that the coin is an unbiased one if:

(A) $Z \geq 8$ (B) $Z \geq 9$ (C) $Z \geq 6$ (D) $Z = 0$ (E) $Z \geq 3$

27. The probability of not rejecting that the coin is an unbiased one if it is indeed biased is:

(A) 0.0282 (B) 0.8507 (C) 0.1493 (D) 0.0107 (E) 0.3828

Questions 28 to 30 refer to the following exercise:

An individual is interested in buying a calculator for his son. Before doing so, he decides to ask for its price at 16 different stores, obtaining a mean sample price of 178 euros with a sample standard deviation of 12 euros. We assume normality.

28. At the 90% confidence level, we can state that the calculator mean price is contained in the interval:

(A) (172.58, 183.42) (B) (173.85, 182.15) (C) (169.40, 186.60)
(D) (167.30, 188.70) (E) (171.40, 184.60)

29. If, at the 10% significance level, we wish to test the null hypothesis that the calculator mean price is $m = 180$, the test result will be:

(A) Do not reject the null hypothesis (B) - (C) Reject the null hypothesis (D) - (E) -

30. At the 95% confidence interval, we can state that the **standard deviation** of the price for the calculator is contained in the interval:

(A) (9.60, 17.81) (B) (9.15, 19.18) (C) (8.77, 16.30)
(D) (83.78, 368.05) (E) (102.16, 187.36)

EXERCISES (Time: 75 minutes)

A (10 points, 25 minutes)

Let $X_1, \ldots, X_n$ be a set of independent r.v. such that $X_1 \in N(k_1\theta, \sigma^2)$, $X_2 \in N(k_2\theta, \sigma^2), \ldots, X_n \in N(k_n\theta, \sigma^2)$, with known variance $\sigma^2 > 0$, and such that the $k_i > 0$'s $i = 1, \ldots, n$ are known constants.

It is known that the probability density function for a $N(m, \sigma^2)$ random variable is given by:

$$f(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

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

ii) Is this estimator unbiased? What is/are the required condition(s) that would make of this estimator a consistent estimator of θ? Remark: In order to answer the latter question, you must compute the variance of the estimator.

B. (10 points, 25 minutes)

The following table describes the probability mass function for the discrete random variable X under the null ($P_0(x)$) and alternative ($P_1(x)$) hypotheses.

X	0	1	2	3	4	5	6
$P_0(x)$	0	0	0.05	0.05	0.10	0.40	0.40
$P_1(x)$	0.20	0.20	0.10	0.10	0.40	0	0

In order to test the null hypothesis $H_0 : P(x) = P_0(x)$ against the alternative hypothesis $H_1 : P(x) = P_1(x)$, a random sample of size $n = 1$ has been taken.

i) Would you include the points $X = \{5, 6\}$ in the critical region? Provide all relevant details.

ii) Would you include the points $X = \{0, 1\}$ in the critical region? Provide all relevant details.

iii) At the 10% significance level, and providing all relevant details that lead us to the required answer, obtain the most powerful critical region for this test. **Hint**: It is very important that before answering this item you remember your answers to the previous ones.

C (10 points, 25 minutes) Let X be a $N(m, \sigma^2)$ random variable, with unknown variance we wish to estimate. In order to do so, a random sample of size n, $X_1, X_2, \cdots, X_n$, has been taken.

i) Find, **providing all relevant details**, the $(1 - \alpha)\%$ confidence interval for the population variance, σ^2.

ii) If $n = 30$ and, from the sample, we have that $s^2 = 25$, find the 95% confidence interval for the population variance σ^2.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

Exercise A)

Given that $X_i \in N(k_i\theta, \sigma^2)$, with known variance $\sigma^2 > 0$, we then have that

$$f(x_i;\theta) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x_i-k_i\theta)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

i) Thus, the likelihood function will be given by

$$L(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$$

$$L(\theta) = \left[\frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x_1-k_1\theta)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x_2-k_2\theta)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x_n-k_n\theta)^2}{2\sigma^2}}\right]$$

$$L(\theta) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\ e^{-\sum_{i=1}^n \frac{(x_i-k_i\theta)^2}{2\sigma^2}}$$

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

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \sum_{i=1}^n \frac{(x_i-k_i\theta)^2}{2\sigma^2}$$

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

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - k_i\theta)(k_i) = 0$$

Therefore,

$$\sum_{i=1}^n k_i x_i - \sum_{i=1}^n k_i^2\theta = 0,$$

so that

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

ii) The estimator will be unbiased if $\mathrm{E}(\hat{\theta}_{\mathrm{ML}}) = \theta$.

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{(\sum_{i=1}^n k_i^2)}\mathrm{E}\left(\sum_{i=1}^n k_i X_i\right) = \frac{1}{(\sum_{i=1}^n k_i^2)}\sum_{i=1}^n k_i \mathrm{E}(X_i)$$

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{(\sum_{i=1}^n k_i^2)}\sum_{i=1}^n k_i(k_i\theta) = \frac{1}{(\sum_{i=1}^n k_i^2)}\left(\sum_{i=1}^n k_i^2\right)\ \theta = \theta$$

Therefore, $\hat{\theta}_{\mathrm{ML}}$ is an unbiased estimator of θ. In order to be able to establish the conditions under which $\hat{\theta}_{\mathrm{ML}}$ is consistent, assuming that it can actually be consistent, we can check if the two sufficient conditions hold:

a) $\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_{\mathrm{ML}}) = \theta$

b) $\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_{\mathrm{ML}}) = 0$

As $\hat{\theta}_{\mathrm{ML}}$ is an unbiased estimator of θ, condition a) holds. In addition, we have that

$$\mathrm{Var}(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\ \mathrm{Var}\left(\sum_{i=1}^{n} k_i X_i\right) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\ \sum_{i=1}^{n} \mathrm{Var}(k_i X_i)$$

$$\mathrm{Var}(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\ \sum_{i=1}^{n} k_i^2 \mathrm{Var}(X_i) = \frac{\left(\sum_{i=1}^{n} k_i^2\right)}{\left(\sum_{i=1}^{n} k_i^2\right)^2}\ \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} k_i^2}$$

Thus, if the condition $\lim_{n\to\infty} \sum_{i=1}^{n} k_i^2 = +\infty$ holds, we would then have that the two conditions for consistency would hold and, as a result, $\hat{\theta}_{ML}$ would be a consistent estimator of θ.

Exercise B

We wish to test the null hypothesis that X is a discrete random variable with probability mass function $P_0(x)$ against the alternative hypothesis that its probability mass function is $P_1(x)$:

X	0	1	2	3	4	5	6
$P_0(x)$	0	0	0.05	0.05	0.10	0.40	0.40
$P_1(x)$	0.20	0.20	0.10	0.10	0.40	0	0

A random sample of size $n = 1$ has been taken; that is, we observe X.

i) Would you include the points $X = \{5, 6\}$ in the critical region?

Given that, under the probability mass function in the alternative hypothesis $P_1(x)$, these points have zero probability, the random variable X cannot take on these values under the alternative hypothesis, but it can clearly take on these values under the null hypothesis. Therefore, the points $X = \{5, 6\}$ are not rejection points for H_0 and, thus, they **should never** be included in the critical region for this test.

ii) Would you include the points $X = \{0, 1\}$ in the critical region?

Given that, under the probability mass function in the null hypothesis $P_0(x)$, these points have zero probability, the random variable X cannot take on these value under the null hypothesis. Therefore, the points $X = \{0, 1\}$ are rejection points for H_0 and, thus, they **should always** be included in the critical region for this test.

iii) At the $\alpha = 0.10$ significance level and **recalling the answers provided in earlier items for this exercise**, we have that there are only two possible critical regions for this test: $\mathrm{CR}_1 = \{0, 1, 2, 3\}$ and $\mathrm{CR}_2 = \{0, 1, 4\}$. This is due to the fact that:

$$\alpha_1 = P(X \in \mathrm{CR}_1 | P_0) = P(X = 0, 1, 2, 3 | P_0) = 0 + 0 + 0.05 + 0.05 = 0.10 \leq \alpha = 0.10$$

$$\alpha_2 = P(X \in \mathrm{CR}_2 | P_0) = P(X = 0, 1, 4 | P_0) = 0 + 0 + 0.10 = 0.10 \leq \alpha = 0.10$$

In order to see which one of these two critical regions is the most powerful one, we compute their respective powers:

$$\text{Power}_1 = P(X \in \text{CR}_1|P_1) = P(X = 0,1,2,3|P_1) = 0.20 + 0.20 + 0.10 + 0.10 = 0.60$$
$$\text{Power}_2 = P(X \in \text{CR}_2|P_1) = P(X = 0,1,4|P_1) = 0.20 + 0.20 + 0.40 = 0.80$$

From these calculations, we conclude that, at the $\alpha = 0.10$ significance level, the most powerful critical region for this test is CR_2.

Exercise C)

i) Let $\vec{X} = (X_1, X_2, \dots, X_n)$ be a random sample of size n taken from a $N(m, \sigma^2)$ population. Therefore, we will have that

$$\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$$

Confidence Interval

$$P\left(\chi^2_{n-1|1-\alpha/2} < \frac{nS^2}{\sigma^2} < \chi^2_{n-1|\alpha/2}\right) = 1-\alpha$$

$$P\left(\frac{1}{\chi^2_{n-1|\alpha/2}} < \frac{\sigma^2}{nS^2} < \frac{1}{\chi^2_{n-1|1-\alpha/2}}\right) = 1-\alpha$$

$$P\left(\frac{nS^2}{\chi^2_{n-1|\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{n-1|1-\alpha/2}}\right) = 1-\alpha$$

Therefore, the $(1-\alpha)\%$ confidence interval for the variance σ^2 is given by:

$$\text{CI}_{1-\alpha} = \left(\frac{ns^2}{\chi^2_{n-1|\alpha/2}}, \frac{ns^2}{\chi^2_{n-1|1-\alpha/2}}\right)$$

ii) More specifically, if it is given that $n = 30$ and $s^2 = 25$, the 95% confidence interval for σ^2 will be:

$$\text{CI}_{1-\alpha} = \left(\frac{750}{\chi^2_{n-1|\alpha/2}}, \frac{750}{\chi^2_{n-1|1-\alpha/2}}\right)$$

Given that $\chi^2_{29|0.05/2} = 45.7$ and $\chi^2_{29|1-0.05/2} = 16$, we will then have that

$$\text{CI}_{0.95} = (16.41, 46, 88)$$