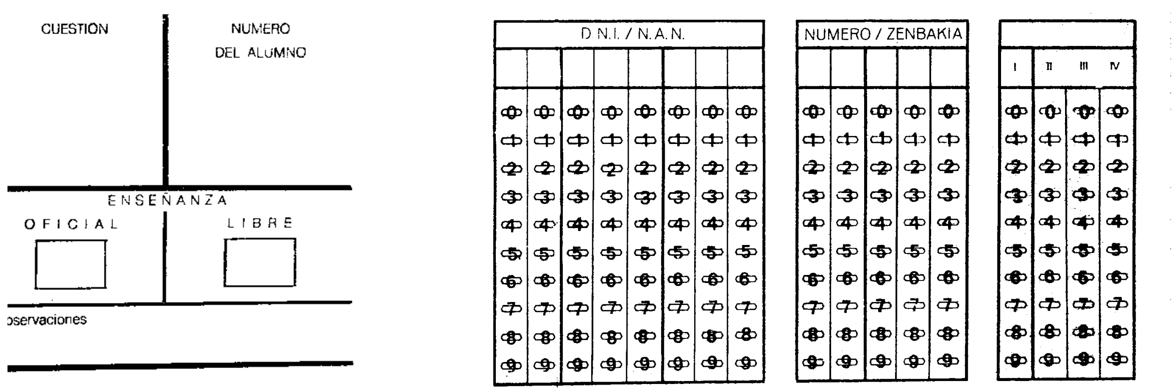
BUSINESS STATISTICS - Second Year Academic Year 2006-07
January 26, 2007

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:

A door-to-door selling editorial firm knows that about 20% of its potential clients or homes that were visited by the firm salesmen buy some of its products.

2. If the firm salesman visits 15 potential clients, the probability that he successfully sells something to more than two potential clients is:

   (A) 0.6020 (B) 0.2309 (C) 0.1671 (D) 0.8329 (E) 0.3980

3. The daily mean sales for a salesman visiting 15 potential clients per day is:

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

4. If in a given month the salesman visits 300 potential clients, the approximate probability that he successfully sells something to a maximum of 52 potential clients is:

   (A) 0.3572 (B) 0.8599 (C) 0.1401 (D) 0.5672 (E) 0.4298

Questions 5 and 6 refer to the following exercise:

The number of daily labor absences in a given firm follows a Poisson distribution with parameter $\lambda = 4$. It is assumed that the distributions of the labor absences for the different days are independent.

5. If we consider the distribution of the labor absences that occur in two given days, the probability that in those two days there are more than 6 absences is:

   (A) 0.6866 (B) 0.1912 (C) 0.3134 (D) 0.8088 (E) 0.4529

6. If we consider the distribution of the labor absences that occur in ten given days, the approximate probability that in those ten days there are at most 37 absences is:

   (A) 0.34 (B) 0.48 (C) 0.32 (D) 0.52 (E) 0.66

Questions 7 to 9 refer to the following exercise:

Let $X_1, X_2, \cdots, X_{100}$ be independent random variables, each one having a $\gamma(2,1)$ distribution.

7. The mean and variance of each one of those random variables are, respectively:

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

8. The probability that the random variable X_1 takes on values greater than or equal to 1 is:

   (A) 0.1353 (B) 0.2706 (C) 0.4729 (D) 0.7294 (E) 0.8647

9. If we define the random variable $Y = X_1 + X_2 + \cdots + X_{100}$, the approximate probability that the random variable Y takes on values in the interval $(48, 52)$ is:

(A) 0.689 (B) 0.952 (C) 0.311 (D) 0.048 (E) 0.437

10. Let $\{X_n\}_{\{n \geq 1\}}$ be a sequence of random variables with probability mass function given by:

$$X_n = \begin{cases} -(\frac{1}{n}), & \text{with probability } \frac{1}{4}; \\ \frac{1}{n}, & \text{with probability } \frac{1}{4}; \\ 1, & \text{with probability } \frac{1}{2}. \end{cases}$$

This sequence of random variables will converge:

(A) In distribution to a degenerate random variable in 0

(B) In distribution to a binary random variable with $p = \frac{1}{2}$

(C) In probability to a $N(0,1)$ random variable

(D) In distribution to a degenerate random variable in $\frac{1}{2}$

(E) All false

Questions 11 to 13 refer to the following exercise:

Let X_1, X_2, X_3 and X_4 be independent random variables with distributions: $X_1 \in N(2, \sigma^2 = 1)$, $X_2 \in N(5, \sigma^2 = 4)$, $X_3 \in N(4, \sigma^2 = 9)$ and $X_4 \in N(8, \sigma^2 = 4)$.

11. If we define the random variable $Y = (X_1 - 2)^2 + \left(\frac{X_2 - 5}{2}\right)^2$, then $P(2.77 \leq Y \leq 5.99)$ is:

(A) 0.05 (B) 0.20 (C) 0.75 (D) 0.80 (E) 0.25

12. If we define the random variable $Z = \frac{(X_1 - 2)^2 + \left(\frac{X_2-5}{2}\right)^2}{\left(\frac{X_3-4}{3}\right)^2 + \left(\frac{X_4-8}{2}\right)^2}$, then $P(Z \leq 9)$ is:

(A) 0.95 (B) 0.01 (C) 0.05 (D) 0.10 (E) 0.90

13. If we define the random variable $V = \frac{\sqrt{2}\,(X_1 - 2)}{\sqrt{\left(\frac{X_3-4}{3}\right)^2 + \left(\frac{X_4-8}{2}\right)^2}}$, then $P(V \leq 2.92)$ is:

(A) 0.90 (B) 0.01 (C) 0.95 (D) 0.10 (E) 0.05

Questions 14 and 15 refer to the following exercise:

The lifetime (in thousands of hours) for some specific light bulbs follows a probability distribution that depends on the parameters α and γ. It is known that $E(X) = \alpha\gamma$ and $E(X^2) = \alpha\gamma(1 + \alpha)$. In order to estimate both parameters, a random sample of four light bulbs has been taken and the resulting lifetimes were: 4.1, 4.5, 3.9 and 5.

14. The method of moments estimate of α is:

(A) $\hat{\alpha} = 9.13$ (B) $\hat{\alpha} = 1.54$ (C) $\hat{\alpha} = 4.42$ (D) $\hat{\alpha} = 3.42$ (E) $\hat{\alpha} = 4.82$

15. The method of moments estimate of γ is:

(A) $\hat{\gamma} = 0.27$ (B) $\hat{\gamma} = 1.63$ (C) $\hat{\gamma} = 1.28$ (D) $\hat{\gamma} = 0.20$ (E) $\hat{\gamma} = 4.38$

Questions 16 and 17 refer to the following exercise:

Let X be a random variable with probability density function given by:

$$f(x) = \begin{cases} (\theta+2)x^{-(\theta+3)}, & x > 1,\ \theta > 0 \\ 0, & \text{otherwise} \end{cases},$$

and let $X_1, \cdots, X_n$ be a random sample of size n from this distribution. It is known that $m = \frac{\theta+2}{\theta+1}$.

16. The method of moments estimator of the parameter θ is:

(A) $\left(\dfrac{\overline{X}-2}{1-\overline{X}}\right)$ (B) $(\overline{X}-2)$ (C) All false (D) $(2-\overline{X})$ (E) $\left(\dfrac{\overline{X}-2}{1+\overline{X}}\right)$

17. The maximum likelihood estimator of the parameter θ is:

(A) $\dfrac{n-1}{\ln\left(\prod_i X_i\right)}$ (B) $\dfrac{n}{\ln\left(\prod_i X_i\right)} - 2$ (C) $\dfrac{n}{\ln\left(\prod_i X_i\right)}$ (D) $\dfrac{-n}{\ln\left(\prod_i X_i\right)}$ (E) All false

Questions 18 to 21 refer to the following exercise:

We have a normal distribution with unknown mean. In order to estimate its unknown mean, a random sample of size n has been taken and, based on it, two estimators are proposed:

$$\hat{\theta}_1 = \frac{X_1 + X_2 + \cdots + X_n}{n}, \qquad \hat{\theta}_2 = \frac{X_1 + X_2 + \cdots + X_n}{n+1}$$

18. Which one of these estimators is unbiased?

(A) $\hat{\theta}_1$ (B) Both estimators are unbiased (C) - (D) None of them is unbiased (E) $\hat{\theta}_2$

19. Which one of them have the smallest variance?

(A) $\hat{\theta}_2$ (B) They have the same variance (C) $\hat{\theta}_1$ (D) It depends on the sample values (E) -

20. Is any of these estimators consistent?

(A) $\hat{\theta}_1$ (B) Both estimators are consistent (C) - (D) None of them is consistent (E) $\hat{\theta}_2$

21. If it is known that the Cramer-Rao lower bound for an unbiased estimator of m is $L_c = \frac{\sigma^2}{n}$, is any of these estimators efficient?

(A) $\hat{\theta}_2$ (B) Both estimators are efficient (C) $\hat{\theta}_1$ (D) None of them is efficient (E) -

Questions 22 and 23 refer to the following exercise:

We have a random variable that takes on values in the interval $(0,1)$, and we wish to test the null hypothesis $H_0 : f(x) = 2x$ against the alternative hypothesis $H_1 : f(x) = 2-2x$. In order to do so, a random sample of size $n = 1$ has been taken and, thus, we consider to use the test statistic X.

22. For a given significance level, the most powerful critical region for this test is of the form:

(A) $(0, C_1)$ (B) $(C_2, 1)$ (C) All false (D) $(C_1, C_2)^C$ (E) (C_1, C_2)

23. At the $\alpha = 0.05$ significance level, the most powerful critical region for this test is:

(A) $(0, 0.328)$ (B) $(0.776, 1)$ (C) $(0, 0.224)$ (D) $(0.672, 0.776)^C$ (E) $(0.672, 1)$

Questions 24 and 25 refer to the following exercise:

Let $\overline{X}$ be the mean of a random sample of size $n = 36$ that has been taken from a $N(m, \sigma^2 = 9)$ population. The decision rule to test $H_0 : m \leq 50$ against $H_1 : m > 50$ is to reject H_0 if $\overline{X} \geq 50.8$.

24. The significance level for this test is:

(A) 0.9452 (B) 0.2981 (C) 0.4853 (D) 0.0548 (E) 0.7020

25. The power of this test for $m = 50.8$ is:

(A) 0.9452 (B) 0.50 (C) 0 (D) 0.0548 (E) 0.7324

Questions 26 and 27 refer to the following exercise:

We have a normal population with variance $\sigma^2 = 25$. In order to estimate its mean, a random sample of size n has been taken.

26. If the sample size is $n = 100$, a 95% confidence interval for the mean is:

(A) $(\overline{x} \pm 0.98)$ (B) $(\overline{x} \pm 1.35)$ (C) $(\overline{x} \pm 4.47)$ (D) $(\overline{x} \pm 3.28)$ (E) $(\overline{x} \pm 1.64)$

27. If we wish to test the null hypothesis $H_0 : m \leq 20$ against the alternative hypothesis $H_1 : m > 20$ and, from the sample, we have that $\overline{x} = 22$, at the 5% significance level, the decision would be:

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

Questions 28 and 29 refer to the following exercise:

An apple growing firm wants to introduce its products into a new market. This would be a successful operation if the variance of the weight their apples have is at most $50gr^2$, and, otherwise, it would not be successful. In order to make a decision on this issue, a random sample of size $n = 10$ has been taken and, from the sample, we have that $s^2 = 53$. It is assumed that apple weights are normally distributed.

28. A 95% confidence interval for the population variance is:

(A) $(27.89, 196.30)$ (B) $(38.92, 162.12)$ (C) $(43.37, 127.21)$ (D) $(57.34, 138.19)$ (E) $(31.36, 159.16)$

29. If $H_0 : \sigma^2 \leq 50gr^2$, at the $\alpha = 5\%$ significance level, the firm would make the decision of:

(A) Not rejecting H_0 (B) - (C) Rejecting H_0 (D) - (E) -

30. We wish to test if the population attitude towards a given product is the same for individuals in three different age intervals (18 to 30, 31 to 45 and older than 45). In order to do so, three random samples of sizes 100, 150 and 80 individuals have been taken, and the percentage of individuals that use the product regularly, sporadically and never, in each of the samples, was studied. The most appropriate test for this specific study is:

(A) An homogeneity test

(B) A test of difference of proportions

(C) A test of equality of variances

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

(E) A test of independence

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

We wish to investigate if there exists a gender effect on the preferences individuals have with respect to the specific television programs they watch. In order to do so, a random sample of 1000 individuals has been taken and individuals were asked about which one of the following types of TV program they prefer to watch: sports, contests or movies. Individuals were then classified according to their gender and their TV program preference, resulting in the following table:

Gender	**Sports**	**Contests**	**Movies**	**Totals**
Male	210	140	130	480
Female	150	160	210	520
Totals	360	300	340	1000

At the 5% significance level, what is the test resulting decision?

B (10 points, 25 minutes) Let X be a random variable with probability mass function given by:

$$P(x) = e^{-(\lambda/2)} \left(\frac{\lambda}{2}\right)^x \left(\frac{1}{x!}\right) \qquad x = 0, 1, 2, \cdots, \qquad \lambda > 0$$

In order to be able to estimate the parameter λ, a random sample of size n, $X_1, X_2, \cdots, X_n$, has been taken. It is known that both the mean and the variance of this distribution are equal to $\frac{\lambda}{2}$.

i) Find, **providing all relevant details**, the method of moments and maximum likelihood estimators of the parameter λ.

ii) If we use the estimator $\hat{\lambda} = 2\overline{X}$. Is this an unbiased estimator of λ? Is it consistent? Is it efficient?

C (10 points, 25 minutes) Let X be a $N(m, \sigma^2)$ random variable, with known variance and an unknown mean that 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 mean, m.

ii) If $n = 25$, $\sigma^2 = 16$ and, from the sample, we have that $\overline{x} = 20$, find the 95% confidence interval for the population mean, m.

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS

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

SOLUTIONS TO EXERCISES

Exercise A)

We have a test of independence to investigate if there is a gender effect on the preferences individuals have with respect to the specific television programs they watch.

Under the null hypothesis of independence, we have that $p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$.

Given that we do not know $p_{i\bullet}$ or $p_{\bullet j}$, we estimate them from the information in the table as follows:

so that

$$\hat{p}(\text{Male}) = 0.48, \qquad \hat{p}(\text{Female}) = 0.52$$

$$\hat{p}(\text{Sports}) = 0.36, \qquad \hat{p}(\text{Contests}) = 0.30, \qquad \hat{p}(\text{Movies}) = 0.34$$

Clase	n_{ij}	$\hat{p}_{ij}$	$n\hat{p}_{ij}$	$\dfrac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$
Male-Sports	210	0.1728	172.8	8.008
Male-Contests	140	0.1440	144	0.111
Male-Movies	130	0.1632	163.2	6.754
Female-Sports	150	0.1872	187.2	7.392
Female-Contests	160	0.1560	156	0.103
Female-Movies	210	0.1768	176.8	6.234
	1000	1	1000	$z = 28.602$

Therefore, under the null hypothesis of independence, the test statistic $\sum \dfrac{(n_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}}$ follows a $\chi^2_{(k'-1)\cdot(k''-1)}$ distribution, where k' and k'' stand for the number of classes in which each of the two characteristics under study have been divided.

More specifically:

$$28.602 > 5.99 = \chi^2_{(2-1)(3-1)\,,\,0.05} = \chi^2_{2\,,\,0.05},$$

so that, at the 5% significance level, the null hypothesis of independence is rejected.

Exercise B

$$P(x, \lambda) = e^{-(\lambda/2)} \left(\frac{\lambda}{2}\right)^x \left(\frac{1}{x!}\right) \qquad x = 0, 1, 2, \cdots, \qquad \lambda > 0$$

$$m(X) = \text{Var}(X) = \frac{\lambda}{2}$$

i)

Maximum likelihood estimator

$$L(\vec{X}; \lambda) = P(X_1; \lambda) \dots P(X_n; \lambda) = e^{-n(\lambda/2)} \left(\frac{\lambda}{2}\right)^{X_1 + X_2 + \cdots + X_n} \left(\frac{1}{X_1! X_2! \cdots X_n!}\right)$$

$$\ln L(\vec{X}; \lambda) = -n\left(\frac{\lambda}{2}\right) + \sum_{i=1}^{n} X_i \ln\left(\frac{\lambda}{2}\right) - \ln(X_1! X_2! \cdots X_n!)$$

$$\frac{\partial \ln L(\vec{X},\lambda)}{\partial \lambda} = -\frac{n}{2} + \frac{\sum_{i=1}^{n} X_i}{\lambda} = 0$$

$$\hat{\lambda}_{ML} = \frac{2\sum_{i=1}^{n} X_i}{n} = 2\overline{X}$$

Method of moments estimator

In order to find the method of moments estimator of λ, we need to equate the first population moment to the first sample moment. That is,

$$a_1 = \alpha_1$$

Given that, in this case, $\alpha_1 = \frac{\lambda}{2}$, and $a_1 = \overline{X}$, we have that: $\dfrac{\lambda}{2} = \overline{X}$, so that,

$$\hat{\lambda}_{MM} = 2\overline{X}$$

ii)

Unbiasedness. The estimator is unbiased because

$$\mathrm{E}\left(\hat{\lambda}\right) = \mathrm{E}\left(2\overline{X}\right) = 2\mathrm{E}(\overline{X}) = 2\left(\frac{\lambda}{2}\right) = \lambda$$

Consistency. The estimator is consistent because it satisfies the two sufficient conditions for sufficiency. That is,

1) $\hat{\lambda}$ is an unbiased estimator, and, in addition,

2) $\lim\limits_{n\to\infty}\left(\mathrm{var}\left(\hat{\lambda}\right)\right) = \lim\limits_{n\to\infty} \mathrm{var}\left(2\overline{X}\right) = \lim\limits_{n\to\infty} 4\dfrac{\mathrm{var}(X)}{n} = \lim\limits_{n\to\infty} 4\left(\dfrac{\lambda}{2n}\right) = \lim\limits_{n\to\infty}\left(\dfrac{2\lambda}{n}\right) = 0,$

Efficiency. In order to find out if the estimator is an efficient one, we need to calculate the Cramer-Rao lower bound for the variance of an unbiased estimator of λ.

$$Lc = \frac{1}{n\mathrm{E}\left[\dfrac{\partial \ln P(X,\lambda)}{\partial\lambda}\right]^2}$$

$$\ln P(X,\lambda) = -\frac{\lambda}{2} + X\ln\left(\frac{\lambda}{2}\right) - \ln X!$$

$$\frac{\partial \ln P(X,\lambda)}{\partial\lambda} = -\frac{1}{2} + \frac{X}{\lambda} = \frac{2X-\lambda}{2\lambda} = \frac{X-(\lambda/2)}{\lambda}$$

$$\mathrm{E}\left[\frac{\partial \ln P(X,\lambda)}{\partial\lambda}\right]^2 = \frac{1}{\lambda^2}\mathrm{E}\left(X-(\lambda/2)\right)^2 = \frac{1}{\lambda^2}\mathrm{var}(X) = \left(\frac{1}{\lambda^2}\right)\left(\frac{\lambda}{2}\right) = \frac{1}{2\lambda}$$

Therefore,

$$Lc = \frac{1}{n\left(\frac{1}{2\lambda}\right)} = \frac{2\lambda}{n}$$

Given that the variance of the estimator coincides with the Cramer-Rao lower bound, the estimator is efficient.

Exercise C

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

$$\bar{X} \in N\left(m, \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} \in N(0,1)$$

Confidence Interval

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

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

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

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

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

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

ii) More specifically, if it is given that $n = 25$, $\sigma^2 = 16$, and $\overline{x} = 20$, the 95% confidence interval for m will be:

$$\mathrm{CI}_{0.95} = \left(20 \pm t_{0.025} \frac{4}{5}\right)$$

Given that $t_{0.025} = 1.96$, we will then have that

$$\mathrm{CI}_{0.95} = \left(20 \pm 1.96 \frac{4}{5}\right) = (20 \pm 1.568) = (18.432, 21.568)$$