Note: Assume that samples used in all the exercises are simple random samples.

1. In order to study the monthly wages of workers in a particular city, a sample of 70 workers has been obtained. The data on wages in the last month of these workers (in euros) can be seen on sheet "Exercise 1" of the attached file Problems2Data.xlsx. Let \( \mu \) and \( \sigma^2 \) the population mean and the population variance of the monthly wages of those workers in this city. Compute:

(a) An unbiased estimate of \( \mu \).
(b) An unbiased estimate of \( \sigma^2 \).
(c) An unbiased estimate of the variance of the sample mean.
(d) An unbiased estimate of the proportion of workers in the city with monthly wages below 2000 euros.

2. Let \( X_1, X_2, X_3 \) be a sample of a random variable \( X \) with mean \( \mu \) and variance \( \sigma^2 \). In order to estimate \( \mu \), the following 3 estimators are considered:

\[
\hat{\mu}_1 = \frac{X_1 + 4X_2 - 2X_3}{3}; \quad \hat{\mu}_2 = 0.4X_1 + 0.3X_2 + 0.3X_3; \quad \hat{\mu}_3 = \frac{1}{3}\hat{\mu}_1 + \frac{2}{3}\hat{\mu}_2.
\]

(a) Show that these three estimators are unbiased.
(b) Analyze which of these three estimators is more efficient.

3. In a survey regarding a particular issue, the interviewee can reply 1 (I disagree), 2 (indifferent) or 3 (I agree). It is known that the probability that a randomly selected person replies 1 is \( 2\theta \), the probability of replying 2 is \( \theta \) and replying 3 is \( 1 - 3\theta \), where \( \theta \) is an unknown parameter which lies in the interval \( (0, 1/3) \). In order to estimate \( \theta \) we have access to a sample with \( n \) replies of interviewees.

(a) One researcher (denoted as R1) estimates \( \theta \) by using \( (3 - \bar{X})/5 \) as an estimator, where \( \bar{X} \) is the sample mean. Show that this estimator is unbiased.
(b) Another researcher (denoted as R2) estimates \( \theta \) by using \( \hat{p}/2 \) as estimator, where \( \hat{p} \) is the proportion of people in the sample replying 1. Show that this estimator is unbiased.
(c) Show that the estimator used by R1 is more efficient than the estimator used by R2.
(d) Suppose that \( n = 20 \). In this sample there are nine replies with 1, four replies with 2 and seven replies with 3.

i. Compute the estimate of \( \theta \) obtained by researcher R1.

ii. Compute the estimate of \( \theta \) obtained by researcher R2.

iii. Taking into account the result in part (c), researcher R1 claims that the true value of \( \theta \) is closer to the estimate obtained in subpart (i) than to the estimate obtained in subpart (ii); is this statement true?

4. Traffic managers are analysing the maximum speed (in \( \text{km/h} \)) to reach the vehicles passing through a dangerous area. One sample of 7 cars provided the following maximum speeds:

\[
79.2 \quad 73.9 \quad 68.3 \quad 77.8 \quad 86.4 \quad 71.3 \quad 69.5
\]

This data can be found on sheet “Exercise 4” of the attached file **Problems2Data.xlsx**.

Suppose that random variable \( X \) denotes the maximum speed of a car in this area and follows a normal distribution with unknown mean \( \mu \) and variance 6.25.

(a) Compute a confidence interval for \( \mu \) with a confidence level of 90% and determine the width of this interval.

(b) Without any computation, explain whether the confidence interval for \( \mu \) one would obtain in the following situations are wider or more narrow than the interval computed in part (a):

i. With sample as in part (a), but with a confidence level of 95%.

ii. With the same level of confidence of part (a), but with a sample with 17 observations.

iii. With the same sample and the same confidence level of part (a), but assuming that the variance of \( X \) is 9 instead.

5. The monthly benefits (in thousands of euros) of a firm are a random variable \( X \) following a normal distribution. In order to study the characteristics of \( X \), we have access to a sample of size 6 of \( X \), which provides the following information:

\[
\text{Benefits:} \quad 23.4 \quad 23.2 \quad 24.1 \quad 19.3 \quad 20.1 \quad 22.2
\]

This data can be found on sheet “Exercise 5” in the attached file **Problems2Data.xlsx**.

(a) Compute a confidence interval for the mean benefit of the firm with a confidence level of 95% and determine the width of this interval.

(b) A researcher claims that, using this sample but with a different confidence level, he obtains a width of the confidence interval for the mean benefit of 4.74. Which is the confidence interval obtained by this researcher? Which is the confidence level of this interval?
6. A big company performs a test to analyse the ability of its employees. A sample of 352 employees is evaluated with this test; it is known that the mean and the standard deviation of the marks of this test in the sample were 60.41 and 11.28, respectively.

(a) Compute a confidence interval for the mean mark of the test with a confidence level of 99%, if the marks are assumed to follow a normal distribution with standard deviation equal to 11.

(b) Compute a confidence interval for the mean mark of the test with a confidence level of 99%, if the marks are assumed to follow a normal distribution with unknown standard deviation.

(c) Compute a confidence interval for the mean mark of the test with a confidence level of 99%, if the marks are assumed to follow distribution with standard deviation equal to 11 which is not normal.

(d) Compute a confidence interval for the mean mark of the test with a confidence level of 99%, if the marks are not normally distributed and its standard deviation is unknown.

7. There are two candidates (the candidate from the liberal party and the candidate from the democrat party) in the next elections. Denote $p$ as the proportion of voters in the population supporting the democrat candidate. In order to estimate this value, 120 voters are interviewed; 63 of these voters said that they will vote the liberal candidate, the rest of them said that they will vote the democrat candidate. Given these results, an analyst clams that the confidence interval for $p$ with an approximate confidence level of 95% only contains values larger than 0.5 and, therefore, the candidate of the democrat party will win the elections with an approximate confidence level of 95%. Is this statement true?

8. In order to know whether firms take into account a university degree at the moment of hiring, 344 human resources managers have been asked whether they consider that a candidate for a job position holds a university degree; 261 managers replied positively and the rest replied that they did not consider it was important. Let $p$ denote the population proportion of firms considering a university degree is important.

(a) Compute the confidence interval for $p$ with an approximate confidence level of 95%.

(b) Compute the confidence interval for $p$ with an approximate confidence level of 90%. The width of this interval, is it larger or smaller than the width of the interval in part (a)? Would you have been able to reply to this question without computing the interval?

(c) With the same data, the confidence interval obtained for $p$ by a researcher is $(0.729, 0.789)$. What is the approximate confidence level used by this researcher?
(d) Compute a confidence interval for \( 1 - p \) with an approximate confidence level of 90%.

9. In order to study the consumption (in liters per 100 km) of a new automobile, a sample on the consumption of 64 cars of this new model has been extracted. The results of this sampling can be seen on sheet “Exercise 9” of the attached file Problems2Data.xlsx. We assume that the consumption of an automobile of this new model is a normally distributed random variable. Compute:

(a) A confidence interval for the mean consumption with a confidence level of 95%.
(b) A confidence interval for the mean consumption with a confidence level of 99%.
(c) A confidence interval for the variance of the consumption with a confidence level of 95%.
(d) A confidence interval for the variance of the consumption with a confidence level of 99%.

10. Analyse whether the following statements are true or false. If they are true, prove them and if they are false, justify the reason why.

(a) Let \( X_1, ..., X_n \) be a sample of a random variable \( X \) following a binomial distribution \( Bi(1, p) \) and let \( \bar{X} \) be the sample mean. If we denote \( \theta = p^2 \) and \( \hat{\theta} = \bar{X}^2 \), then \( \hat{\theta} \) is an unbiased estimator of \( \theta \).

(b) The formula defining the bounds of a confidence interval for parameter \( \theta \) can never contain parameter \( \theta \).

(c) If \( X_1, X_2, X_3 \) are three independent random variables such that \( X_1 \sim N(1, 2), \ X_2 \sim N(1, 1), \ X_3 \sim N(2, 1) \), and we denote

\[
Y = \frac{(X_1 - 1)/\sqrt{2}}{\sqrt{(X_2 - 1)^2 + (X_3 - 2)^2}}
\]

then the probability that \( Y \) is below 1.89 is 0.90.

(d) If, for a given sample, the confidence interval for the population mean with confidence level 95% is (21.2, 24.5); then, the probability that the population mean lies in the interval (21.2, 24.5) is 0.95.

(e) If we do not know the population variance of a normal distribution, we can compute a confidence interval for the population mean with an approximate confidence level, but not with an exact confidence level.