Statistical inference
Semester 4
2024/2025

Worksheet n°1

Exercise 1. The table below shows the distribution of a university's 10,000 undergraduate students by level of study and gender:

Level / Gender	S_1	S_2	S_3	S_4	S_5	S_6
Male	2000	1600	1000	900	400	250
Female	1500	900	600	500	200	150

We propose to carry out a survey of 1000 students to obtain information on the quality of teaching at this university. How can this sample be distributed among the different strata of the 1000 students, taking into account level and gender?

Exercise 2. On a population of 15,000 companies studied according to sales (variable X) and number of employees (variable Y). We obtain the following results:

X (10^5 DZD)	Number	Y	Number
[0,50[	5500	[0,10[	7050
[50,100[	4500	[10,50[	3700
[100,200[	1750	[50,100[	2500
[200,500[	2000	[100,500[	1500
[500,1000[	500	[500,--- [	250
[1000,--- [	750	-	-
Total	15000	Total	15000

Consider a sample of 1,500**.** What method can be used to determine the sample? Assume that the institute responsible for the survey has 30 interviewers.
Explain your approach and the concepts used.

Exercise 3. You're doing an internship at a car dealership. Your manager asks you to carry out a study on the dynamics of this market.

1. What should the sample size be if you wish to have a sampling error sampling error of no more than 3%?
2. You have finally interviewed 800 people. 8% told you they would like to change their vehicle in the next 6 months. What is the margin of error of this result?

Exercise 4.

1. The candidate in a municipal election asks you what the sample size should be to gather voting intentions in his municipality with a margin of error of 10%.
2. What should the sample size be if he wishes to double the precision of his estimate?
3. The poll indicates that 53% of the population (sample of 400) is ready to vote for Mr. X. Can we say (with certainty) that Mr. X will be elected?
4. On election day, Mr. X receives only 48% of the vote and criticizes the reliability of your poll. How do you respond?

Exercise 5. You have a budget of 600,000 DA. The corresponding costs are as follows: questionnaire design: 130,000 DA, interviewer remuneration: 500 DA/questionnaire, survey input and processing: 100 DA/questionnaire.

1. How many questionnaires can you administer with this budget?
2. If you wish to carry out a survey of the Algerian population with an error of 3%, can you do so? If not, what budget will you need?

Exercise 6. Let T_1 and T_2 be two different estimators of the same parameter θ. Assume that $E[T_1]= \theta+b_1$ and $E[T_2]= \theta+b_2$ (b_1 and b_2 are two known numerical values). Let $T=\alpha T_1+\beta T_2$.
Determine α and β so that T is an unbiased estimator of θ:

a) When $b_1 \neq b_2$.
b) When $b_1 = b_2$.

Exercise 7. The elements of a population possess a character X that follows a density distribution

$$f_\theta(x) = \frac{2}{\sqrt{\pi}\theta^{3/2}} x^2 e^{-x^2/\theta}.$$

Where $\theta > 0$. A series of n independent experiments yielded the values $x_1, x_2, \dots, x_n$.
1. Determine an estimator $\hat{\theta}$ of the parameter θ using the maximum likelihood method.
2. Is this estimator unbiased? Convergent? Efficient?

Exercise 8. In order to improve customer satisfaction, a major ISP compiles statistics on the number of hotline calls received, so as to evaluate the waiting time for the customer and the number of employees needed to staff the switchboard. The results of the survey cover 200 consecutive one-minute sequences, during which the average number of calls was 3 per minute. Calls are assumed to be evenly distributed over time: a time interval is divided into one-second units, so in each time unit there is at most one call.

1. What is the probability distribution of the number of calls received in 4 minutes?
2. Show that this distribution can be approximated by a Poisson distribution.
3. Give a confidence interval for the average number of calls in 4 minutes.

Exercise 9. The medical staff of a large company compile statistics on the cholesterol levels of their employees; the observations on 100 employees selected at random are as follows.

Cholestérol level in cg : (center of classes)	Number of employes
120	9
160	22
200	25
240	21
280	16
320	7

1. Calculate the mean m_e and standard deviation σ_e on the sample.
2. Estimate the mean and standard deviation for cholesterol levels throughout the company.
3. Determine a confidence interval for the mean.
4. Determine the minimum sample size so that the amplitude of the confidence interval is less than 10.

Exercise 10. The average height of a random sample of 40 people taken from a population of 780 is 1.70m. The standard deviation for the whole population is 24cm. Find the 95% confidence interval for the mean height of the population.

Exercise 11. 500 students sit an exam. A random sample of 38 marks gives a mean equal to 8.65 and a standard deviation equal to 2.82. Find the confidence interval for the population mean scores at 90%, 95% and 99%.

Exercise 12. It is assumed that the weights of 3000 students at a university follow a normal distribution with mean 68kg and standard deviation σ=3kg.

1. What is the sampling mean and standard deviation of the means if we extract 80 samples of 25 students each.
   a) In the case of a non-exhaustive draw.
   b) In the case of exhaustive sampling.
2. For how many samples can we expect to find an average:
   a) Between 68.3kg and 68.8kg.
   b) Less than 68.4kg.

Exercise 13. The human resources manager of a company has established that the results of a test measuring the manual dexterity of the workforce assigned to tasks involving the assembly of complex parts are distributed according to the normal distribution of mean m = 72 and variance σ^2= 36.

1. What is the probability that a randomly selected employee will score less than 63 on the manual dexterity test?
2. A random sample of 25 employees took the manual dexterity test.

a) What is the distribution of the sample mean?
b) What are the mean and standard deviation of the distribution of the mean?

3. What is the probability that the sample mean lies between 69 and 75?
4. What is the probability that the standard deviation between the sample mean and the population mean is greater than 3?

Exercise 14. It is assumed that students in a statistics lecture have normally distributed grades with mean m = 72 and standard deviation σ= 9.

1. Find the probability that a single randomly selected student has a grade above 80.
2. Find the probability that a random sample of 10 students will have an average grade above 80.
3. Answer the previous question (2.) assuming that the population does not follow a normal distribution.

Exercise 15. Two samples of size n_1 = 120 and n_2 = 150 were taken independently from a population of economics students.
We find that 48 students in the first sample and 66 students in the second sample have a scientific secondary education. Let p be the proportion of students in the population with a scientific background.
Calculate three-point estimates of p.

Exercise 16. A company uses insulating material in the assembly of a certain type of electric motor. It is important not only that the average thickness of the components meets the company's requirements, but also that the variability in thickness does not fluctuate too widely.
A random sample of 20 insulating components taken from a batch gives the following thicknesses:

Thickness in mm									
5,6	5,9	6,2	6,1	6,6	5,9	5,9	5,6	6,2	5,8
5,5	5,6	6	6,3	6,2	5,9	6,2	6	6,2	6,3

1. Assuming that the thickness of this insulating material is distributed according to a normal distribution, estimate by confidence interval the standard deviation of the thickness for the entire production run. Use a confidence level of 95%.
2. Estimate also the mean thickness of the insulating material for the whole production with a confidence level of 95%.