Mathematics for Supply Chain

Msc Supply Chain & Purchasing

Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

1 Course summary

We first provide some recall on the sampling estimation and the main difference between a pointwise estimation. Some properties, in terms of mean and variance are also stated.

In a second part, we focus on the estimation by confidence intervals in three different cases:

- estimation of the mean of a population when the its variance is known,
- estimation of the mean of a population when the its variance is unknown,
- estimation of the proportion of a population.

1.1 Sampling Estimation and Central Limit Theorem

When we have access to sample of a population, it gives the opportunity to have an estimation of several statistics about this population, such as the mean.

More concretely, imagine that you have sample $(x_1, x_2, x_3, \ldots, x_n)$ where $x_i \in \mathbb{R}$ then you are able to compute the mean value of this sample:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

You hope that this value is representative of the population. But what happen if I am changing my sample and that I have the values $(x'_1, x'_2, x'_3, \ldots, x'_n)$. Do you think that will necessarily have the estimation of the mean? **The answer is clearly no!**

Exemple 1.1. *Imagine that you roll a dice where the possible output is from* 1 *to* 6 *and you want to estimation the mean value of the dice. To estimate the mean value, you roll the dice twice to have an estimation of the mean value that you can reach.*

In the first experiment you will have 3 *and* 5 *for instance, then the mean value is equal to* 4*. If I am doing it a second time, I will get the values* 5 *and* 6 *and you will have a mean equal to* 5.5

In which estimation may I trust? How can I know which one is the most reliable and may I have an idea of the possible value that I can have and far I am from the real mean value?

All these answers are given by the Central Limit Theorem, which is stated below

Theorem 1.1: Central Limit Theorem

Let $X_1, X_2, ..., X_n$ of real random variables that are **independent** and **identically distributed** with the same unknown mean parameter μ and the same variance σ^2.
Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then:

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{loi} Z \sim \mathcal{N}(0,1),$$

i.e. the random variable Z_n defined as above can be approximated, for sufficiently large n, by a random variable Z which is normally distributed with a mean equal to 0 and a variance equal to 1.

In practice, sufficiently large means that $n \geq 30$. In other words, whatever the distribution the random variable (gaussian or not), the estimator of the mean $\bar{X}_n$ is normally distributed when is large enough.

Remark There is no condition on the size of n in the case where the data are normally distributed (see the previous notes on the normalization process to understand why).

Properties of the estimator of the mean $\bar{X}_n$ The estimator of the mean has the following properties:

(i) $\mathbb{E}[\bar{X}_n] = \mu$, *i.e.* the expectation of this random variable is no more than the value of the unknown parameter μ,

(ii) $\sigma_{\bar{X}_n} = \dfrac{\sigma}{\sqrt{n}}$, *i.e.* the standard deviation of the mean estimator is equal to standard deviation of a random variable X_i divided the square root of the sample size n.

First step to confidence interval Let us suppose that our sample size n is larger than 30 and consider the random variable

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

This random variable follows a $\mathcal{N}(0, 1)$. For this reason, we know that a proportion $1 - \alpha$ of the values of this random variable will fall in the range $[-z_{1-\alpha/2},\ z_{1-\alpha/2}]$ where $z_{1-\alpha/2}$ denotes the quantile of order $1 - \alpha/2$, *i.e.*, with probability $1 - \alpha$, we are sure that

$$-z_{1-\alpha/2} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}.$$

This is the first step to build the different confidence intervals with a rate of confidence $1 - \alpha$ that we are going to draw in the next sections.

1.2 Confidence Interval on the mean: σ is known

In this section, we consider that **we know** the variance σ at the population scale. According to the previous section, we have, with probability equal to $1 - \alpha$ that the random variable Z_n is bounded as:

$$-z_{1-\alpha/2} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}.$$

Using this expression, we can say that, with the same probability, the unknown parameter μ we are looking for, is also located in the range:

$$\left[\bar{x}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

where $z_{1-\alpha/2}$ denotes the quantile of order $1 - \alpha/2$ of a gaussian distribution with a mean equal to 0 and a variance equal to 1.

Exemple 1.2. *We have drawn* $10,000$ *samples of perfume in order to measure the quantity of liquid present in the vials. We want to check that the average fill is always equal to* 50.2 *ml. The average volume of perfume in our sample vials is* 50.5 *ml. It is further assumed that the filling process follows a normal distribution with standard deviation* $\sigma = 2$.

Can we say that the machine is correctly adjusted with an error rate of 10%?

According to the statement, we have $n = 10,000, \bar{x}_n = 51, \sigma = 2$ *and a margin of error* $\alpha = 0.1$ *because we want a confidence interval of* $1-\alpha = 90\%$. *Remember that our confidence interval is symmetrical around* $\bar{x}_n$ *and verifies*

$$\mathbb{P}\left[\bar{x}_{inf} \leq \mu \bar{x}_{sup}\right] = \mathbb{P}\left[\bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

The aim here is to determine the upper bound $\bar{x}_{sup}$. *By definition,* $z_{1-\alpha/2}$ *is the value, for a reduced centered variable* Z, *for which we have :*

$$\mathbb{P}\left[Z \leq z_{1-\alpha/2}\right] = 1 - \frac{\alpha}{2}.$$

In our case $\alpha = 0.1$, *so* $1 - \frac{\alpha}{2} = 0.95$. *By searching the Z table, we find that the value of* $z_{1-\alpha/2}$ *is equal to* 1.64

According to what we saw earlier, we therefore have

$$\bar{x}_{sup} = \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 50.5 + 1.64 \times \frac{2}{\sqrt{10,000}} = 50.5 + 0.0328 \simeq 50.53.$$

In fine, our confidence interval is as follows:

$$[50.47, 50.53].$$

Or $50.2 \notin [50.47, 50.53]$, *so we can't say that our machine is well-tuned with a rate of error of* 10%.

Important to remember

Theory

Let us consider a sample of size n denoted $x_1, \ldots, x_n$,then, the estimator of the mean$\bar{x}_n$ est donné par

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_{n-1} + x_n}{n}.$$

This estimator of the mean $\bar{X}_n$ is a random variable whose distribution depends on the context. In the case where the standard deviation σ of the distribution is known and the data are from a normal distribution or our sample size is greater than 30, then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \simeq Z \sim \mathcal{N}(0,1),$$

where μ is the mean parameter we aim to estimate.

Confidence interval (symmetrical!, but non-symmetrical confidence intervals are also possible) of level $1-\alpha$ for the mean μ.

$$\left[\bar{x}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[\bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

is the quantile of order α of the normal distribution, *i.e.* it is the value for which a random variable Z following a normal distribution verifies :

$$P[Z \leq z_\alpha] = \alpha.$$

We can also say that a proportion $1-\alpha$ of estimates of the mean $\bar{x}_n$ fall within the interval

$$\left[\mu - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \mu + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Pratice

To give a confidence interval of level $1-\alpha$ on an unknown parameter such as the mean μ **in the case where the standard deviation of the distribution σ is known**, we must

1. estimate the mean value $\bar{x}_n$ from the data

2. check the size n of our sample
3. determine the value of $z_{1-\alpha/2}$
4. calculate the bounds of the confidence interval from the above information

$$\left[\bar{x}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

If you want to check whether a machine is up to standard (you know the reference value μ), you can check whether $\bar{x}_n$ lies in the interval

$$\left[\mu - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \mu + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

We'll proceed in the same way for the construction of this interval.

1.3 Confidence Interval on the mean: σ is unknown

In this section, we consider that we **do not know** the variance σ at the population scale.
According to the previous section, we have, with probability equal to $1-\alpha$ that the random variable Z_n is bounded as:

$$-z_{1-\alpha/2} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}.$$

However, we do not have access to σ to compute our confidence interval. Tee only thing that we can do is to estimate this value on the sample that we have, thus, we will make the following approximation:

$$\sigma^2 \simeq s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2,$$

where $\bar{x}_n$ is the mean value on the sample.
If we replace the value σ by estimated one s on the sample, it is no more true to say that the random variable $\frac{\bar{X}_n - \mu}{\frac{s}{\sqrt{n}}}$ follows a normal distribution. It will follows a **Student distribution** with $n-1$ degree of freedom where n is the sample size.

Student Distribution It is another distribution which is close the normal distribution and which depends on only one parameter that is the number of degree of freedom p. We denote this distribution $\mathcal{T}_p$.

With Excel

With Excel, we can compute the probabilities and the quantiles associated to the Student distribution with p degree of freedom using the functions below: (i) the probabilities and (ii) the quantiles

Version	Function
ENGLISH	T.DIST(t, p, CUMULATIVE)
FRENCH	LOI.STUDENT.N(t; p, CUMULATIVE)

where p is the number of degree of freedom and $t \in \mathbb{R}$. The last parameter is called Cumulative, if:

- **TRUE**: it computes $\mathbb{P}[T \leq t] = F(t)$, otherwise
- **FALSE**: it computes $f(t)$ the value of the density for the given t.

Version	Function
ENGLISH	T.INV(α, p)
FRENCH	LOI.STUDENT.INVERSE.N(α; p)

where p is the number of degree of freedom and $\alpha \in [0, 1]$.

Despite the distribution is not the same, the way our confidence interval is defined remains the same. In this specific case, where σ is unknown, we now that, with probability $1 - \alpha$, the unknown parameter μ is in the range:

$$\left[\bar{x}_n - t_{n-1,1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + t_{n-1,1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

where $t_{n-1,1-\alpha/2}$ denotes the quantile of order $1-\alpha/2$ of a Student distribution with $n-1$ degree of freedom.

Exemple 1.3. *A pharmaceutical laboratory wishes to study the reliability of an automatic filling machine which is responsible for filling boxes containing a medical preparation. We wish to check that each box contains a mass $\mu = 86.65$ grams of this preparation. It is assumed that the filling of the cans is distributed according to a normal with unknown parameters μ and σ^2. To check the correct setting of its machine with a risk of error of 0.1, the laboratory took a sample of 1000 cans; from this sample, we obtained a mean mass of $\bar{x}_n = 86g$ and a standard deviation $s = 0.12g$.*

We are in the case where the standard deviation of our population, here the variability of mass due to the automaton, our estimator of the mean therefore follows a Student's law whose degree of freedom is equal to the sample size minus one, i.e. 999.
The confidence interval $I_{1-\alpha}$ of level $1-\alpha = 0.9$ is defined by :

$$
\begin{aligned}
& \quad \downarrow \textit{ definition of a confidence intervale with a rate of confidence } 1-\alpha \\
I_{1-\alpha=0.9} &= \left[\bar{x}_n + t_{\alpha/2}\frac{s}{\sqrt{n}}; \bar{x}_n + t_{1-\alpha/2}\frac{s}{\sqrt{n}}\right], \\
& \quad \downarrow \textit{ If } 1-\alpha = 0.9 \textit{ then } \alpha = 0.1, \textit{ so } \alpha/2 = 0.05 \textit{ and } 1-\alpha/2 = 0.95. \\
& \quad \downarrow \textit{ replace } \bar{x}, s \textit{ and } n \textit{ by their values.} \\
&= \left[86 + t_{0.05}\frac{0.12}{\sqrt{1000}}; 86 + t_{0.95}\frac{0.12}{\sqrt{1000}}\right], \\
& \quad \downarrow \textit{ Look for the quantile of order } 0.95 \textit{ using Excel} \\
&= \left[86 - 1.646 \times \frac{0.12}{\sqrt{1000}}; 86 + t_{0.95}\frac{0.12}{\sqrt{1000}}\right], \\
& \quad \downarrow \textit{ compute the bounds} \\
I_{1-\alpha=0.9} &= [85.99; 86.01]
\end{aligned}
$$

The pharmaceutical company can therefore conclude that the automaton is incorrectly set up, and should therefore increase the average filling value of their automaton.

Important to remember

Theory

Considering a sample of n measurements denoted $x_1, \dots, x_n$, then the estimators of the mean $\bar{x}_n$ and variance s^2 are given by

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_{n-1} + x_n}{n}.$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

This estimator of the mean $\bar{X}_n$ is a **random variable** whose distribution depends on the context. In the case where we don't know the standard deviation σ of the distribution and the data come from a normal distribution or our sample size n is greater than 30, then

$$\frac{\bar{X}_n - \mu}{s/\sqrt{n}} \simeq T \sim \mathcal{T}_{n-1},$$

where μ is the unknown parameter to be estimated.
Confidence interval (symmetrical!, but non-symmetrical confidence intervals are also possible) of level $1-\alpha$ for the mean μ.

$$\left[\bar{x}_n - t_{1-\alpha/2}\sqrt{s^2/n}; \bar{x}_n + t_{1-\alpha/2}\sqrt{s^2/n}\right] = \left[\bar{x}_n + t_{\alpha/2}\sqrt{s^2/n}; \bar{x}_n + t_{1-\alpha/2}\sqrt{s^2/n}\right],$$

where t_{alpha} is the quantile of order α of the Student's law with $n-1$ degrees of freedom, *i.e.* is the value for which a random variable T following a Student's law with $n-1$ degrees of freedom verifies :

$$P[T \leq t_\alpha] = \alpha.$$

We can also say that a proportion $1-\alpha$ of estimates of the mean $\bar{x}$ fall within the interval

$$\left[\mu - t_{1-\alpha/2}\sqrt{s^2/n}; \mu + t_{1-\alpha/2}\sqrt{s^2/n}\right].$$

To give a confidence interval of level $1-\alpha$ on an unknown parameter such as the mean μ **in the case where the standard deviation of the distribution σ is unknown**, we must

1. give an estimate of the mean value $\bar{x}$ from the data
2. estimate the standard deviation s from the data
3. check the size n of our sample
4. determine the value of $t_{1-\alpha/2}$
5. calculate the bounds of the confidence interval from the above information

$$\left[\bar{x}_n - t_{1-\alpha/2}\sqrt{s^2/n}; \bar{x}_n + t_{1-\alpha/2}\sqrt{s^2/n}\right]$$

If you want to check whether a machine is up to standard (you know the reference value μ), you can check whether $\bar{x}$ lies in the interval

$$\left[\mu - t_{1-\alpha/2}\sqrt{s^2/n}; \mu + t_{1-\alpha/2}\sqrt{s^2/n}\right].$$

Proceed in the same way for the construction of this interval.

1.4 Confidence Interval on the proportion

Whether or not an individual has a particular characteristic (binary side) can be modelled by a random variable $X \sim \mathcal{B}(p)$ or $\mathcal{B}(1,p)$ where the parameter p is unknown and represents the probability of an individual having the characteristic in question.

Now consider an independent sample $x_1, x_2, ..., x_n$ with the same distribution as $X \sim \mathcal{B}(p)$, then the proportion estimator defined by

$$\bar{P} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

follows a distribution with expectation and variance defined by

$$\mathbb{E}\left[\bar{P} = p\right] \quad \text{et} \quad Var(X) = \frac{p(1-p)}{n}.$$

In addition, we have $n\bar{P} \sim \mathcal{B}(n,p)$ as the sum of n independent Bernoulli random variables with parameter p.

We have seen that if our sample was large enough, we could approximate the Binomial distribution by a Normal distribution, *i.e.* when n is large enough

$$n\bar{P} \sim \mathcal{N}\left(np, np(1-p)\right),$$

where np and $np(1-p)$ denote the expectation and variance, respectively, of the binomial distribution $\mathcal{B}(n,p)$. We can therefore approximate our random variable $\bar{P}$ as follows

$$\frac{\bar{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1),$$

where we have simply subtracted the mean and divided by the standard deviation to obtain a centered and reduced Normal distribution.

Based on this approximation, we are once again able to construct intervals at a $1-\alpha$ confidence level on our unknown parameter p. More precisely, we are able to estimate with what probability our sampled value will fall within a given confidence interval, i.e. we have:

$$\mathbb{P}\left[p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \bar{p} \leq p + z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha.$$

From this result, we can also state that we have a probability of $1-\alpha$ that our unknown parameter belongs to the interval

$$\left[\bar{p} + z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right].$$

The deviation from the true population value is called the margin of error and still has the same meaning as in the previous sections, only its expression changes.

Exemple 1.4. *The percentage obtained for a question in a survey is* 20%*. We wish to determine a margin of error of* 2.5% *(remember that the margin of error is half the length of our confidence interval) in* 95% *of cases. What size should our sample be? What would our margin of error be if our sample size were equal to* 100*?*

For the first question in this example, remember that the margin of error is equal to

$$z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}.$$

We want a confidence interval of 0.95*, so we have* $\alpha = 0.05$*, so we look up the value of* $z_{1-\alpha/2=0.975}$ *in the table of the normal distribution and find (or remember) that this value is* 1.96*. Now we want :*

$$\begin{aligned} z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} &= 0.025, \\ 1.96 \times \sqrt{\frac{0.2(1-0.2)}{n}} &= 0.025. \end{aligned}$$

We thus have:

$$n = \frac{01.96\sqrt{0.2 \times 0.8}}{0.025} = 983.3,$$

which means that n *must be at least equal to* 984.

For the second question, we start again from the definition of the margin of error and simply perform the calculation assuming that the proportions remain unchanged, which gives us

$$1.96 \times \sqrt{\frac{0.2 \times 0.8}{100}} = 0.0784.$$

We therefore have a margin of error of 0.0784*, i.e. a confidence interval for the estimate of the proportion* p *of the form:*

$$[0.2 - 0.0784; 0.2 + 0.0784] = [0.1216; 0.2784].$$

Important to remember

Theory

Considering a sample of n measurements denoted $x_1, \cdots, x_n$, an estimator of the proportion $\bar{p}$ (or f) is given by

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_{n-1} + x_n}{n},$$

where x_i take the values 0 or 1.
This estimator of the proportion $\bar{p}$ is a **random variable** again. Asymptotically, when the sample size is large enough to verify $nbarp \geq 5$ and $n(1-\bar{p}) \geq 5$

$$\frac{\bar{p} - p}{\sqrt{\bar{p}(1-\bar{p})/n}} \simeq Z \sim \mathcal{N}(0,1),$$

where p is the unknown parameter to be estimated.
Confidence interval (symmetrical!, but non-symmetrical confidence intervals are also possible) of level $1-\alpha$ for proportion p.

$$\left[\bar{p} - z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right].$$

Which is the same as:

$$\left[\bar{p} + z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right],$$

where z_{alpha} is the quantile of order α of the centered-reduced Normal distribution, $i.e.$ is the value for which a random variable Z following a centered-reduced Normal distribution verifies :

$$P[Z \leq z_\alpha] = \alpha.$$

We can also say that a proportion $1-\alpha$ of the estimates of the proportion $\bar{p}$ fall within the interval

$$\left[p - t_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}; p + t_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}\right].$$

Practice

To give a confidence interval of level $1-\alpha$ on an unknown parameter such as the proportion p, we must

1. estimate the proportion *barp* from the data
2. check the size n of our sample
3. determine the value of $z_{1-\alpha/2}$
4. calculate the bounds of the confidence interval from the above information

$$\left[\bar{p} - z_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}; \bar{p} + z_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}\right]$$

If you want to check whether a machine is up to standard (you know the reference value μ), you can check whether $\bar{x}$ lies in the interval

$$\left[p - z_{1-\alpha/2}\sqrt{p(1-p)/n}; p + z_{1-\alpha/2}\sqrt{p(1-p)/n}\right].$$

We will proceed in the same way for the construction of this interval.

2 Exercises

Exercice 2.1 (Study of a Lifetime)**.**

The lifetime of a light bulb, given in hours, is represented by a random variable X whose distribution is assumed to be Normal with standard deviation $\sigma = 400$, while the parameter of the mean μ is unknown.
Measurements of the lifetime of a batch of 9 bulbs gave the following results:

$$2000; 1890; 3180; 1990; 2563; 2876; 3098; 2413; 2596.$$

a) Determine a confidence interval for the average life of a light bulb at the 90% level.

b) Can we assert, with a risk of error of 10%, that the average life of a light bulb is equal to 2500 hours?

Exercice 2.2 (Study of a factory)**.**

A factory specializing in cable construction wants to check the reliability of its products by assessing the maximum mass its cables can support.
To do this, we model the maximum mass, in tonnes, supported by a cable as a random variable X following a Normal distribution with unknown mean μ and standard deviation $\sigma = 0.5$. A study was carried out on a sample of 50 cables. On average, the maximum load supported by a cable was 12.2 tons.

a) Determine a confidence interval for μ at the 0.99 level.

b) Can we state that the machine, with a risk of error of 1%, produces cables capable of supporting a mass of at least 11.5 tonnes?

c) Determine the minimum size of the sample studied so that the length of the confidence interval at the 99% level is less than 0.2.

Exercice 2.3 (Biology)**.**

A biologist is studying a type of algae that attacks marine plants. The toxin contained in this alga is obtained in the form of an organic solution. The biologist measures the quantity of toxin per gram of solution, modeled by a random variable X following a normal distribution whose expectation μ and variance σ^2 are unknown. He obtained the following nine measurements, expressed in milligrams:

$$1.2; 0.8; 0.6; 1.1; 1.2; 0.9; 1.5; 0.9; 1.0$$

The standard deviation associated with this sample is $s = 0.26$.

a) *Calculate the mean* $\bar{x}_n$ *associated with this sample.*

b) *Give a confidence interval of level* 0.90 *of toxin quantity.*

c) *Can we say, with a confidence level of* 0.80, *that the quantity* μ *of toxin per gram of solution is equal to* 1.3 *mg.*

Exercice 2.4 (IT Firm)**.**

An information technology (IT) consulting firm specializing in health care solutions wants to study communication deficiencies in the health care industry. A random sample of 70 *health care clinicians reveals the following:*

- *Time wasted in a day due to outdated communication technologies:*

$$\bar{x} = 45 \textit{ minutes} \quad \textit{and} \quad s = 10 \textit{ minutes.}$$

- *Thirty-six health care clinicians cite inefficiency of pagers as the reason for the wasted time.*

Answer to the following questions:

1. *Construct a* 99% *confidence interval estimate for the population mean time wasted in a day due to outdated communication technologies.*
2. *Construct a* 95% *confidence interval estimate for the population proportion of health care clinicians who cite inefficiency of pagers as the reason for the wasted time.*

Exercice 2.5 (Decision and Costs)**.**

The cost of certain decisions made for a corporate strategy committee (in thousands of dollars) is studied in order to conduct a statistical study of these decisions. These values can be found in the file ***CostDecision*** *This would allow, among other things, to help in the decision making process for future decisions, risky or not.*

Use the data to:

- *Estimate the mean cost for each decision that have been made.*
- *Build a confidence interval with a a level of significance of:*
  1. $1 - \alpha = 99\%$
  2. $1 - \alpha = 95\%$
  3. $1 - \alpha = 90\%$
- *Can we say that, with probability equal to* 95%*, the mean average cost is equal to* 20*?*

Exercice 2.6 (Elections)**.**

In a survey preceding the presidential elections, 500 *people were questioned. Although this is not the case in practice, to simplify the calculations, we assume that the* 500 *people constitute an independent and identically distributed sample (often referred to as i.i.d.) of the French population. Of the* 500 *people, 150 said they would vote for candidate* C_1 *and* 140 *for candidate* C_2*.*

a) Give a point estimate of voting intentions for each candidate, as a percentage.

b) Give a 95% *confidence interval for each of the two voting intentions.*

c) Can we predict which of C_1 *or* C_2 *will be elected?*

Exercice 2.7 (A Clinical Study)**.**

A pharmaceutical company is setting up a test to assess the efficacy of a new migraine drug. Two groups of 125 *migraine sufferers, treated as random samples, take part in the test. Patients in group A are given the new drug, while patients in group B receive a placebo (no active ingredient). After four days of treatment,* 75 *patients in group A and* 65 *patients in group B report a reduction in migraine intensity.*

a) Determine the confidence intervals at the 0.95 *confidence level for the proportions of patients reporting a decrease in migraine intensity in each sample.*

b) Do these confidence intervals allow, at the 0.95 *confidence level, to consider that the drug is more effective than the placebo?*

c) What should be the minimum size of each sample so that, with proportions identical to those observed previously, the results confirm the efficacy of the drug, at the 0.95 *confidence level?*

Exercice 2.8 (Online Sells).

A major online retailer wants to check the efficiency of its site, and in particular its accessibility.
To this end, it has recorded the connection times to its web page of several users and wishes to check that the latter is well below 1s. The data relating to this study can be found in the file ***Speed***.

Can we say that the specifications have been met with a confidence level of 95%*?*
Same question with a confidence level of 80%.