Mathematics I

Exercise sheet 2 – Sampling, estimation and confidence intervals

EM Lyon Business School – MSc Supply Chain & Operation Management – G. Metzler

Content. Course on sampling, the Central Limit Theorem, confidence intervals (mean with σ known or unknown, proportion) and sample size, then exercises.
Prerequisites. Sheet 1 (normal distribution, standardisation, quantiles). **Tools.** Z and Student tables, or Excel.

1 Course: sampling, estimation and confidence intervals

In practice we never observe a whole population: we audit 400 orders out of 50 000, we time 36 receptions, we weigh 100 boxes. This course explains how to go from a **sample** to a statement about the **population**, with a controlled margin of error. We first study the behaviour of the sample mean (Central Limit Theorem), then we build **confidence intervals** in three situations:

- the mean of a population when its standard deviation σ is known,
- the mean of a population when σ is unknown,
- a proportion (rate of defects, of late deliveries, of returns…).

We finish with the question every manager asks before an audit: *how many observations do I need?*

Why it matters in supply chain

Most supply chain KPIs are computed on samples: OTIF measured on audited orders, inventory accuracy on counted locations, average lead time on recent deliveries. A KPI without its margin of error can lead to wrong decisions: penalising a provider who meets its contract, or ignoring a real drift.

1.1 Population, sample and estimators

Definition 1.1: Vocabulary

- The **population** is the set of all the objects studied (all the orders shipped this year); its characteristics are unknown **parameters**: the mean μ, the standard deviation σ, a proportion p.
- A **sample** is a subset of n objects actually observed, with values $x_1, \dots, x_n$.
- An **estimator** is a quantity computed on the sample to approximate a parameter: $\bar{x}_n$ for μ, s for σ, $\bar{p}$ for p.

Parameter (population)	Estimator (sample)	Example
mean μ	$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$	mean dock-to-stock time
standard deviation σ	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$	variability of a filling machine
proportion p	$\bar{p} = \frac{\text{number of "successes"}}{n}$	rate of late deliveries

A sample is only useful if it is **representative**: the objects must be drawn at random (*simple random sampling*), or in each relevant sub-group (*stratified sampling*: each DC, each shift, each product family). Timing only the receptions of Monday morning, or auditing only the orders of one picker, introduces a **bias** that no sample size can correct.

1.2 Sampling variability and the Central Limit Theorem

When we have a sample $(x_1, x_2, \ldots, x_n)$, we can compute its mean

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

and hope that it is representative of the population. But with another sample $(x'_1, x'_2, \ldots, x'_n)$, would we get the same estimate? **Clearly not!** The estimator is itself a **random variable**, denoted $\bar{X}_n$.

Example 1.1. *Imagine that you roll a dice (values from* 1 *to* 6*, true mean* 3.5*) and that you estimate its mean with two rolls. In a first experiment you get* 3 *and* 5*: the mean is* 4*. In a second experiment you get* 5 *and* 6*: the mean is* 5.5*. With 10 or 50 rolls, the estimates are much closer to* 3.5 *(Figure 1).*

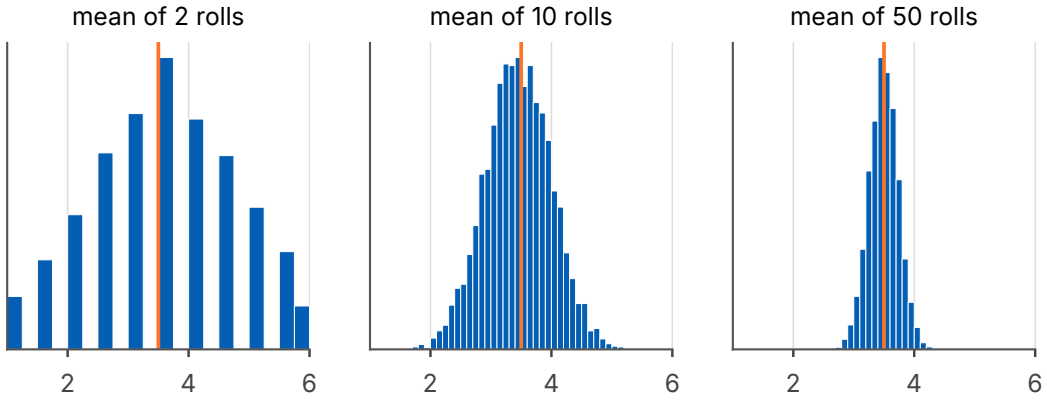


Figure 1: Distribution of the mean of 2, 10 and 50 rolls of a dice (5 000 experiments each). The estimates concentrate around the true mean 3.5 and take a bell shape.

Which estimate can I trust? How far am I from the true mean? The answers are given by two properties and by the Central Limit Theorem.

Properties of the estimator of the mean $\bar{X}_n$. If $X_1, \ldots, X_n$ are independent with mean μ and standard deviation σ:

(i) $\mathbb{E}[\bar{X}_n] = \mu$: on average, the estimator gives the right value (it is **unbiased**);

(ii) $\sigma_{\bar{X}_n} = \dfrac{\sigma}{\sqrt{n}}$: the standard deviation of the estimator, called the **standard error**, decreases with the sample size.

Indeed $\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum \mathbb{E}[X_i] = \mu$ and, the X_i being independent, $Var(\bar{X}_n) = \frac{1}{n^2}\sum Var(X_i) = \frac{\sigma^2}{n}$. Consequence: to divide the error by 2, you need 4 times more data.

Theorem 1.1: Central Limit Theorem

Let $X_1, X_2, ..., X_n$ be real random variables, **independent** and **identically distributed**, with mean μ and variance σ^2, and let $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \underset{n\to\infty}{\longrightarrow} Z \sim \mathcal{N}(0,1) \quad \text{(in distribution)},$$

i.e. for n large enough, $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$ and the total $\sum X_i$ is approximately $\mathcal{N}(n\mu, n\sigma^2)$.

In practice, "large enough" means $n \geq 30$. In other words, **whatever the distribution of the data** (normal or not), the mean of a large sample is normally distributed (Figure 2).

Remark. There is no condition on n when the data are themselves normally distributed: a sum of independent normal variables is normal.

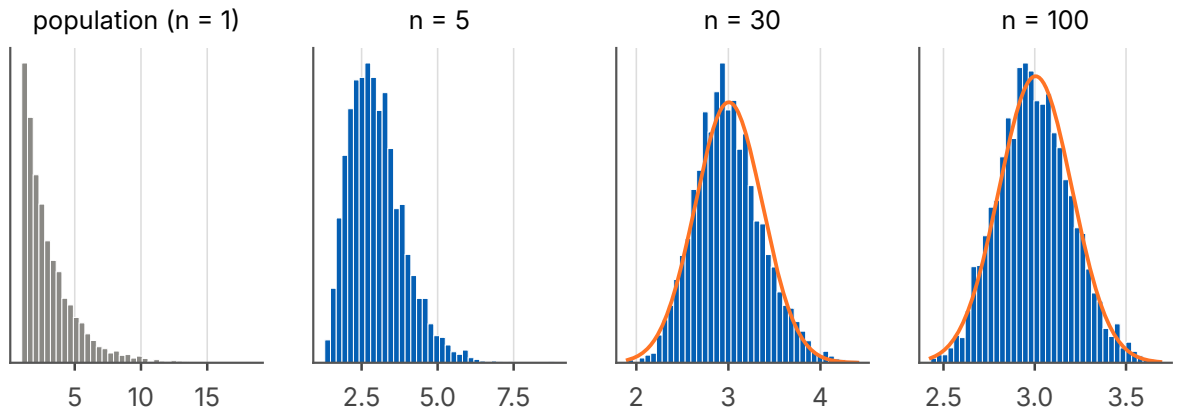


Figure 2: Lead times of a supplier (very skewed population, left) and distribution of the mean of n deliveries. From $n = 30$, the normal approximation (orange) is excellent.

Example 1.2 (Supply chain example: loading roll cages)**.** *At a parcel hub, roll cages are loaded with 40 parcels. A parcel weighs on average 2.4 kg with a standard deviation of 1.1 kg (skewed distribution: many light parcels, a few heavy ones). By the CLT the total weight T of a cage is approximately normal, with mean $40 \times 2.4 = 96$ kg and standard deviation $1.1\sqrt{40} = 6.96$ kg. The probability of exceeding the 100 kg limit of the cage is $\mathbb{P}[T > 100] = 1 - \Phi\big(\frac{100-96}{6.96}\big) = 1 - \Phi(0.57) = 0.28$: more than one cage out of four is overloaded.*

1.3 From the CLT to a confidence interval

Suppose that $n \geq 30$ (or that the data are normal) and consider

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

A proportion $1-\alpha$ of the values of a $\mathcal{N}(0,1)$ variable fall in $[-z_{1-\alpha/2},\ z_{1-\alpha/2}]$, where $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ (Figure 3). Hence, with probability $1-\alpha$:

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

This is the starting point of all the confidence intervals of this sheet. The quantity α is the **risk** (the probability that the interval misses μ) and $1-\alpha$ the **confidence level**.

confidence level $1-\alpha$	90 %	95 %	99 %
risk α	0.10	0.05	0.01
quantile $z_{1-\alpha/2}$	1.645	1.960	2.576

1.4 Confidence interval on the mean: σ is known

In this section, we assume that **we know** the standard deviation σ of the population (for instance from a long history of the process, or from the specifications of a machine).

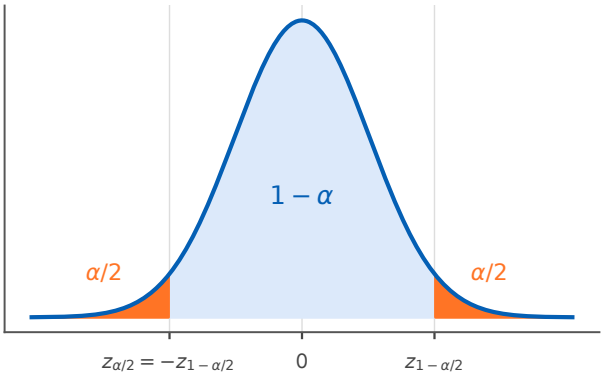


Figure 3: A proportion $1-\alpha$ of the standard normal distribution lies between $-z_{1-\alpha/2}$ and $z_{1-\alpha/2}$; $\alpha/2$ is left in each tail. For $1-\alpha = 95\,\%$, $z_{0.975} = 1.96$.

Definition 1.2: Confidence interval on μ, σ known

A confidence interval of level $1-\alpha$ for the mean μ is

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

The half-length $E = z_{1-\alpha/2}\,\sigma/\sqrt{n}$ is the **margin of error**.

What does "95 % confidence" mean? The interval is random: it changes with the sample. The procedure captures the true mean μ for 95 % of the samples (Figure 4). Once the sample is drawn, the interval either contains μ or not; we are "95 % confident" in the method. It does **not** say that 95 % of the individual values lie in the interval: individual values are $\sqrt{n}$ times more spread than their mean.

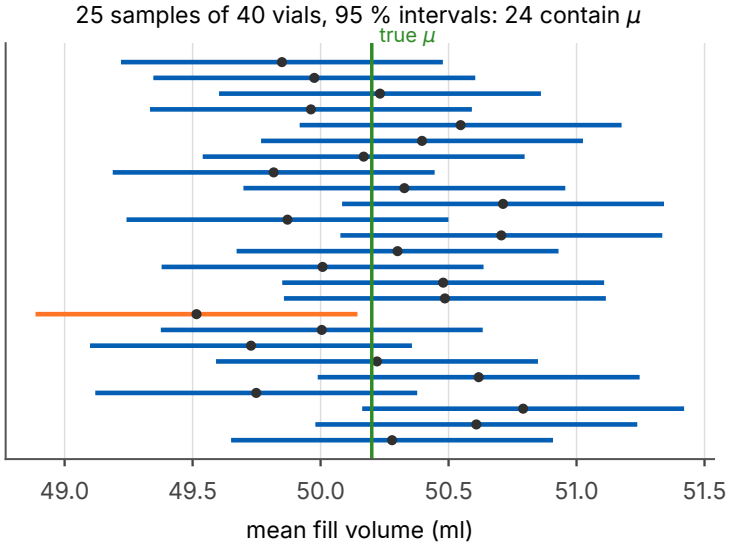


Figure 4: 25 auditors each take 40 vials and compute a 95 % confidence interval. Most intervals contain the true mean (green line); about 1 out of 20 misses it (orange).

What makes an interval narrower? The margin of error $z_{1-\alpha/2}\,\sigma/\sqrt{n}$ decreases when the sample size n increases, when the process is more regular (σ smaller), and when we accept a lower confidence level.

Example 1.3. *We have drawn* $10\,000$ *vials of perfume in order to measure the quantity of liquid present in the vials. We want to check that the average fill is equal to* $50.2\,$*ml. The average volume in our sample is* $\bar{x}_n = 50.5\,$*ml, and the filling process is normal with standard deviation* $\sigma = 2\,$*ml. Can we say that the machine is correctly adjusted, with a risk of error of* $10\,\%$*?*

We have $n = 10\,000$*,* $\bar{x}_n = 50.5$*,* $\sigma = 2$ *and* $\alpha = 0.1$ *(confidence level* $90\,\%$*). By definition,* $z_{1-\alpha/2}$ *is the value such that* $\mathbb{P}[Z \leq z_{1-\alpha/2}] = 1 - \frac{\alpha}{2} = 0.95$*; the* Z *table gives* $z_{0.95} = 1.645$*. The margin of error is*

$$E = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.645 \times \frac{2}{\sqrt{10\,000}} = 0.033,$$

so the confidence interval is $[50.467;\ 50.533]$*. Since* $50.2 \notin [50.467;\ 50.533]$*, the machine is not set at 50.2 ml (at the* $10\,\%$ *risk): it overfills by about 0.3 ml, i.e. 3 litres of perfume given away every* $10\,000$ *vials.*

Important to remember: mean, σ known

Theory. If σ is known and the data are normal or $n \geq 30$: $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \simeq Z \sim \mathcal{N}(0,1)$. Symmetric confidence interval of level $1-\alpha$ for μ:

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

where $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of $\mathcal{N}(0,1)$, *i.e.* $\mathbb{P}[Z \leq z_{1-\alpha/2}] = 1-\alpha/2$ (note that $z_{\alpha/2} = -z_{1-\alpha/2}$).

Practice. (1) compute $\bar{x}_n$; (2) check that $n \geq 30$ or that the data are normal; (3) find $z_{1-\alpha/2}$ (table, or `NORM.S.INV(1-α/2)`); (4) compute the bounds. Excel gives the margin directly: `CONFIDENCE.NORM(α;σ;n)` (French `INTERVALLE.CONFIANCE.NORMAL`).
To check a machine whose reference setting μ_0 is known, check whether $\bar{x}_n$ lies in $\left[\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}};\ \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ (this is a test, see sheet 3).

1.5 Confidence interval on the mean: σ is unknown

Most of the time we **do not know** σ: we estimate it on the sample with

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

We divide by $n-1$ and not by n because the deviations are measured from $\bar{x}_n$, itself estimated on the same data: dividing by n would underestimate σ^2 (Excel: `STDEV.S`, not `STDEV.P`).

Replacing σ by s adds uncertainty: $\dfrac{\bar{X}_n - \mu}{s/\sqrt{n}}$ no longer follows a normal law but a **Student distribution** with $n-1$ degrees of freedom (when the data are normal).

Student distribution. It is a symmetric bell curve, close to the normal distribution but with heavier tails, which depends on one parameter: the number of degrees of freedom p. We denote it $\mathcal{T}_p$. When p grows it gets closer to $\mathcal{N}(0,1)$ (Figure 5); its quantiles are larger than the normal ones for small samples:

n	5	10	20	30	100	∞ (normal)
$t_{0.975}(n-1)$	2.776	2.262	2.093	2.045	1.984	1.960

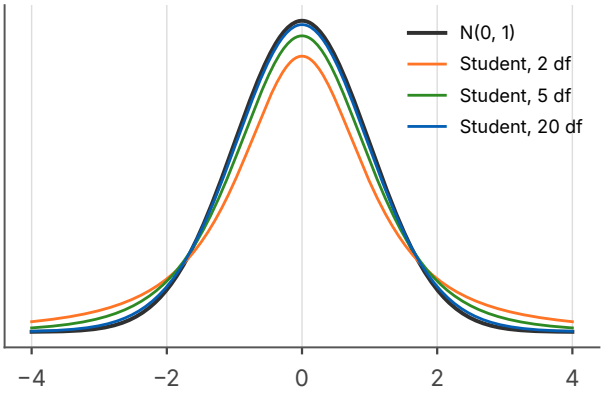


Figure 5: Student distributions with 2, 5 and 20 degrees of freedom, and the standard normal distribution.

With Excel: Student distribution

Version	Function
ENGLISH	`T.DIST`$(t;\ p;\ \texttt{TRUE}) = \mathbb{P}[T \le t]$, `T.INV`$(\alpha;\ p)$ quantile of order α, `T.INV.2T`$(\alpha;\ p) = t_{1-\alpha/2}$, `CONFIDENCE.T`$(\alpha;\ s;\ n)$ margin of error
FRENCH	`LOI.STUDENT.N`, `LOI.STUDENT.INVERSE.N`, `LOI.STUDENT.INVERSE.BILATERALE`, `INTERVALLE.CONFIANCE.STUDENT`

where p is the number of degrees of freedom ($n-1$ for a confidence interval on a mean).

Definition 1.3: Confidence interval on μ, σ unknown

If the data are normal (or $n \ge 30$), a confidence interval of level $1-\alpha$ for μ is

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

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

Example 1.4. *A pharmaceutical laboratory wishes to check an automatic machine filling boxes with a medical preparation: each box should contain $\mu_0 = 86.65\,g$. The filling is assumed normal with unknown μ and σ. On a sample of 1000 boxes, the mean mass is $\bar{x}_n = 86\,g$ and the standard deviation $s = 0.12\,g$. We build a confidence interval of level $1-\alpha = 0.9$.*

↓ *definition of a confidence interval of level $1-\alpha$, σ unknown*

$$I_{0.9} = \left[\bar{x}_n - t_{1-\alpha/2}\frac{s}{\sqrt{n}};\ \bar{x}_n + t_{1-\alpha/2}\frac{s}{\sqrt{n}}\right],$$

↓ *$1-\alpha = 0.9$, so $1-\alpha/2 = 0.95$; we replace $\bar{x}_n$, s and n by their values*

$$= \left[86 - t_{0.95}(999)\frac{0.12}{\sqrt{1000}};\ 86 + t_{0.95}(999)\frac{0.12}{\sqrt{1000}}\right],$$

↓ *quantile of order 0.95 with 999 degrees of freedom:* `=T.INV(0.95;999)`

$$= [86 - 1.646 \times 0.0038;\ 86 + 1.646 \times 0.0038],$$

↓ *compute the bounds*

$$= [85.994;\ 86.006].$$

The value 86.65 is far outside the interval: the machine is incorrectly set and underfills the boxes by about 0.65 g. With $n = 1000$, the Student quantile (1.646) is almost the normal one (1.645).

Important to remember: mean, σ unknown

Theory. With $\bar{x}_n = \frac{1}{n}\sum x_i$ and $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x}_n)^2$: if the data are normal (or $n \geq 30$), $\dfrac{\bar{X}_n - \mu}{s/\sqrt{n}} \sim \mathcal{T}_{n-1}$, and the confidence interval of level $1-\alpha$ for μ is

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

where $t_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of the Student law with $n-1$ degrees of freedom.

Practice. (1) compute $\bar{x}_n$ and s; (2) check n (normality needed if $n < 30$); (3) find $t_{1-\alpha/2}(n-1)$ (`T.INV.2T`$(\alpha; n-1)$); (4) compute the bounds. Using z instead of t for a small sample gives an interval that is too narrow: overconfidence.

1.6 Confidence interval on a proportion

Whether or not an object has a characteristic (an order is late or not, a part is defective or not) is modelled by a Bernoulli random variable $X \sim \mathcal{B}(1,p)$, equal to 1 with probability p and 0 otherwise. The parameter p is the unknown proportion in the population.

For an independent sample $X_1, ..., X_n$ of $\mathcal{B}(1,p)$, the estimator of the proportion

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

is the share of "successes" in the sample. Since $n\bar{P} = \sum X_i \sim \mathcal{B}(n,p)$:

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

For n large enough the binomial distribution is close to a normal one (it is again the CLT), so

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

valid in practice when $n\bar{p} \geq 5$ and $n(1-\bar{p}) \geq 5$. Replacing the unknown p by $\bar{p}$ in the standard error gives the interval below.

Definition 1.4: Confidence interval on a proportion

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

The margin of error is $E = z_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}$.

Example 1.5 (Supply chain example: OTIF of a logistics provider)**.** *A 3PL must deliver* $95\,\%$ *of the orders On Time In Full (OTIF). An audit of 400 random orders finds 368 OTIF orders:* $\bar{p} = 0.92$*. The conditions* $400 \times 0.92 \geq 5$ *and* $400 \times 0.08 \geq 5$ *hold, and*

$$E = 1.96\sqrt{\frac{0.92 \times 0.08}{400}} = 0.027, \qquad p \in [89.3\,\%;\ 94.7\,\%] \text{ at } 95\,\%.$$

The whole interval is below the contractual $95\,\%$*: the target is very likely not met.*

Example 1.6. *In a survey, the percentage obtained for a question is* $20\,\%$*. We want a margin of error of* 2.5 *points (half the length of the interval) in* $95\,\%$ *of the cases. What should the sample size be? What would the margin of error be with a sample of size* 100*?*

We want a $95\,\%$ *confidence interval, so* $\alpha = 0.05$ *and* $z_{0.975} = 1.96$*. We want*

$$1.96 \times \sqrt{\frac{0.2 \times 0.8}{n}} = 0.025 \iff n = \left(\frac{1.96}{0.025}\right)^2 \times 0.2 \times 0.8 = 983.4,$$

which means that n *must be at least equal to* 984*. With* $n = 100$ *and the same proportion, the margin is* $1.96 \times \sqrt{0.16/100} = 0.0784$*, i.e. the interval* $[0.1216;\ 0.2784]$*: ten times fewer respondents, a margin three times larger (Figure 6).*

Important to remember: proportion

Theory. With $\bar{p} = \frac{1}{n}\sum x_i$ ($x_i \in \{0,1\}$) and $n\bar{p} \geq 5$, $n(1-\bar{p}) \geq 5$: $\dfrac{\bar{P} - p}{\sqrt{\bar{p}(1-\bar{p})/n}} \simeq Z \sim \mathcal{N}(0,1)$, and the confidence interval of level $1-\alpha$ for p is

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

Practice. (1) compute $\bar{p}$; (2) check the conditions on n; (3) find $z_{1-\alpha/2}$; (4) compute the bounds. To check a reference proportion p_0 (a contractual rate), check whether $\bar{p}$ lies in $\left[p_0 - z_{1-\alpha/2}\sqrt{p_0(1-p_0)/n};\ p_0 + z_{1-\alpha/2}\sqrt{p_0(1-p_0)/n}\right]$.

1.7 How many observations do I need?

Before an audit or a survey, we choose the margin of error E and the confidence level, and we deduce the sample size by inverting the formula of the margin:

$$\text{mean: } n \geq \left(\frac{z_{1-\alpha/2}\,\sigma}{E}\right)^2, \qquad \text{proportion: } n \geq \left(\frac{z_{1-\alpha/2}}{E}\right)^2 p(1-p).$$

Always round **up**. If p is unknown, take $p = 0.5$, the worst case ($p(1-p) \leq 0.25$). If σ is unknown, use a pilot study or the history.

Example 1.7 (Timing receptions)**.** *A DC wants to estimate its mean dock-to-stock time within* ± 3 *minutes at* $95\,\%$*; the history gives* $\sigma \approx 12$ *minutes. Then* $n \geq (1.96 \times 12/3)^2 = 61.5$*:* ***62*** *receptions must be timed.*

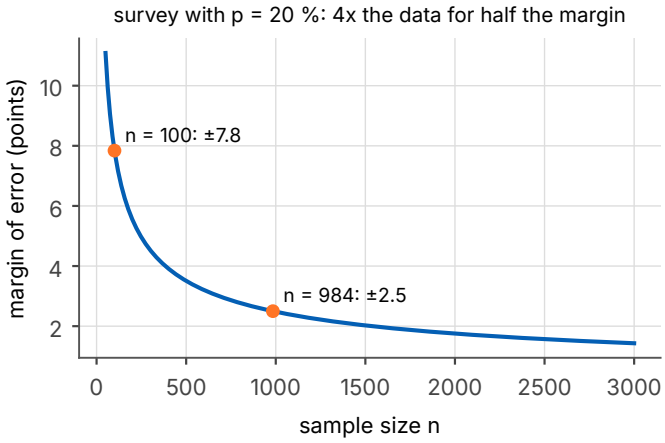
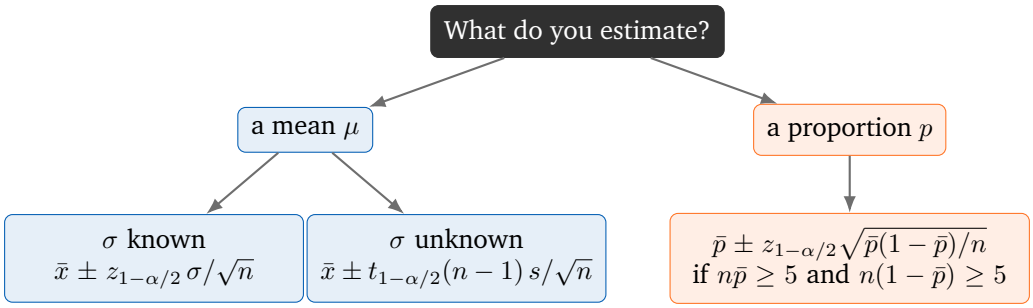


Figure 6: Margin of error of a survey ($p = 20\,\%$, $95\,\%$ confidence) as a function of the sample size: dividing the margin by 2 requires 4 times more observations.

1.8 Summary: which interval?



Common mistakes

- Using σ (population) and s (sample) indifferently, or dividing by n instead of $n-1$.
- Using z instead of t for a small sample with unknown σ.
- Reading the interval as the range of the individual values.
- Forgetting the conditions: $n \geq 30$ or normal data; $n\bar{p} \geq 5$ and $n(1-\bar{p}) \geq 5$.

2 Exercises

Exercise 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?

Exercise 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.

Exercise 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.

Exercise 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:*

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

- *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.*

Exercise 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?*

Exercise 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?

Exercise 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?*

Exercise 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%*.*