v.25-1

MSE213
Probability and Statistics for Materials Science - Preliminary version

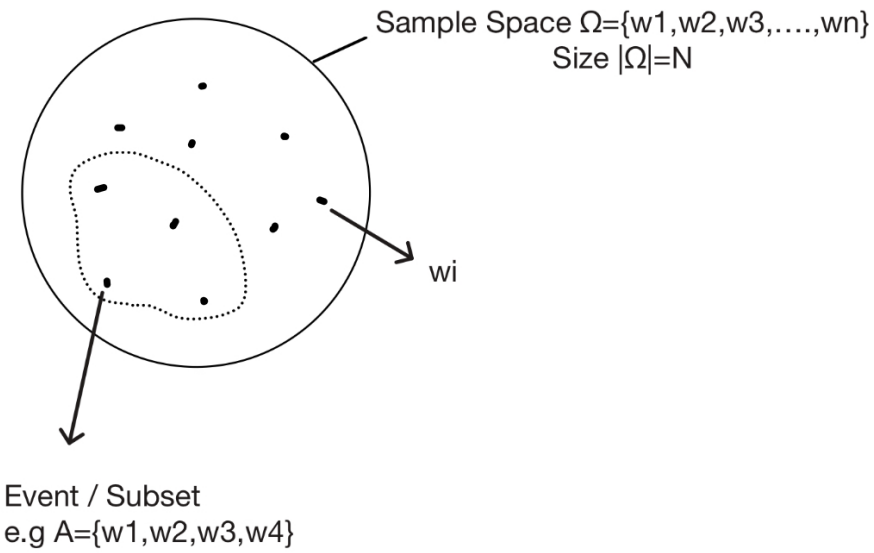
Note that these lecture notes are meant to *complement* the lectures, they are not complete and not a replacement.
This is a preliminary version, if you find mistakes, please contact gregor.jotzu@epfl.ch

2025

Contents

1 Basics of Probability

1.1 Set Theory - Definitions



Examples:

Sample spaces:
→ Discrete:
$\Omega = \{heads, tails\}$ (one coin toss)
$\Omega = \{hh, tt, ht, th\}$ (2 **ordered** coin tosses)
$\Omega = \{hh, tt, ht\}$ (2 **unordered** coin tosses)
$\Omega = \{1, 2, 3, 4, 5, 6\}$ (die throw)
$\Omega = \{Emilie, Jean, Sam\}$ (pick a person)

→ Continuous:
$\Omega = \{0 \leq t \leq \infty\}$ (decay time of a nucleus)

Events:
$E = \{1, 2, 3\}$ ("odd" die result)
$A = \{hh, ht\}$ ("first one heads" coin result)
$B = \{hh, tt\}$ ("two of the same face")
$C = \{tt\}$ ("two tails") → *C is elementary, it contains only one element*

Elementary events: $\{heads\}, \{hh\}$

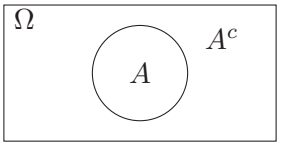
1.2 Operations on sets

Complement: all elements except A:

$$A^c = \overline{A}$$

Example:

$$A = \{th,\, tt\}$$

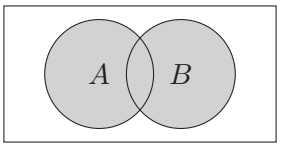


Union / OR:

$$A \cup B$$

Example:

$$A \cup B = \{tt,\, hh,\, th\}$$



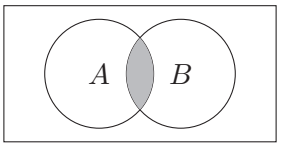
(Venn diagram)

Intersection / AND:

$$A \cap B$$

If $A \cap C = \varnothing$ then A and C are **mutually exclusive**.
Example:

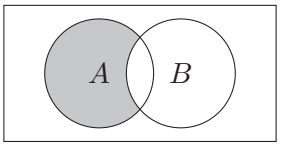
$$A \cap B = \{\text{something}\}$$



Exclusion / A without B:

$$A \setminus B$$

Example:



Containment:

$$C \subseteq B \quad \text{if} \quad B \cap C = C$$

Example:

$$B \cap C = \{tt\} = C$$

Note:

$$A \cap \bar{A} = \varnothing \quad \text{always.}$$

1.3 Axioms of probability

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- If $A \cap B = \{\}$ then $P(A \cup B) = P(A) + P(B)$ and if $A \cap B = A \cap C = B \cap C = \{\}$ then $P(A \cup B \cup C) = P(A) + P(B) + P(C)$ etc...

1.4 Uniform ("Laplace") probabilities

Generally: for independent experiments, Laplace says that **all elementary events have equal probabilities**.

$P(\{\omega_i\}) = P(\{\omega_j\})$ for all i,j

$\sum_{i=1}^{n} P(\{\omega_i\}) = P(\{\Omega\}) = 1 \rightarrow P(\{\omega_i\}) = \frac{1}{|\Omega|} = \frac{1}{N}$

Example: $P(\{heads\}) = P(\{tails\}) = \frac{1}{2}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of elements in A}}{\text{total number of elements}}$$

Example: $P(\text{“}odd\text{”}) = \frac{3}{6} = \frac{1}{2}$

1.5 General Results

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example: For a die
P("odd" ∪ "bigger than 4") $= P(1,3,5,6) = P(1,3,5) + P(5,6) - P(5) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3}$

1.6 Conditional probabilities/ Independent events

Probability of A if B is known to be true ("A given B")" $\rightarrow B$ becomes the new sample space

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

For uniform probabilities ("Laplace"):

$$P(A \mid B) = \frac{|A \cap B|}{|B|}$$

$$P(A \mid B) = 0 \quad \text{if } A \cap B = \varnothing$$

$$P(A \mid B) = 1 \quad \text{if } B \subseteq A \quad (\text{i.e. } A \cap B = B)$$

Independence

$$P(A \mid B) = P(A) \quad \Rightarrow \quad P(A) \text{ is independent of } P(B)$$

Then, for A and B independent:

$$\frac{P(A \cap B)}{P(B)} = P(A) \quad \Leftrightarrow \quad P(A \cap B) = P(A) \cdot P(B)$$

To prove that A and B are independent events you need to show showing that either of the equalities above is true (when one is true, both are true).

Example: Rolling a fair die

$$P(\text{“odd”} \cap \text{“}>4\text{”}) = P(\{5\}) = \frac{1}{6}$$

$$= P(\{1,3,5\}) \cdot P(\{5,6\}) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

$\Rightarrow$ These two events are **independent**.

$$P(\text{“odd”} \cap \text{“}<4\text{”}) = P(\{5\}) = \frac{1}{6}$$

$$\neq P(\{1,3,5\}) \cdot P(\{4,5,6\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$\Rightarrow$ These two events are **not independent**.

Note that “<4” is NOT the dual of “>4” (that would be “≥ 4”), otherwise the independence the first pair would imply the independence of the second pair.

2 Random variables and Bernoulli distribution

2.1 Probability Trees

2.1.1 Chain rule

$$P(A \cap B) = P(B \mid A) \cdot P(A) = P(A \mid B) \cdot P(B)$$

$$P(A \cap B \cap C) = P(C \mid A \cap B) \cdot P(A \cap B) = P(C \mid A \cap B) \cdot P(B \mid A) \cdot P(A) = ...$$

(From the definition of conditional probability)

2.1.2 Total Probability

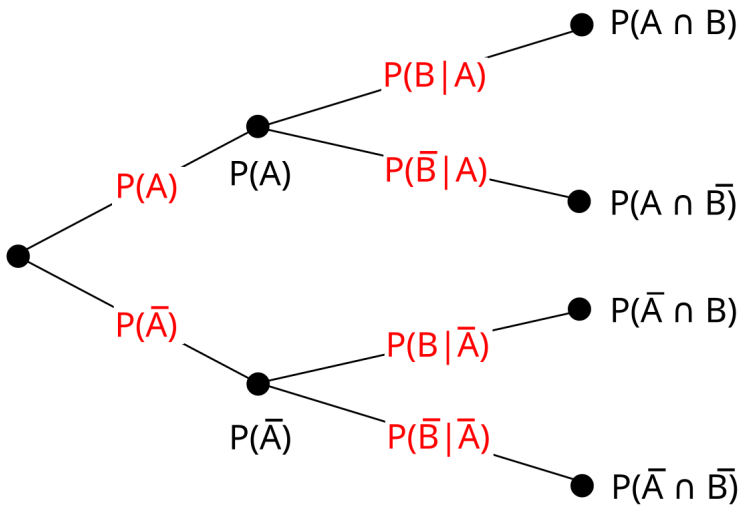
$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = P(B \mid A) \cdot P(A) + P(B \mid \bar{A}) \cdot P(\bar{A})$$

Generally:

$$P(B) = \sum_{i=1} P(C_i \cap B)$$

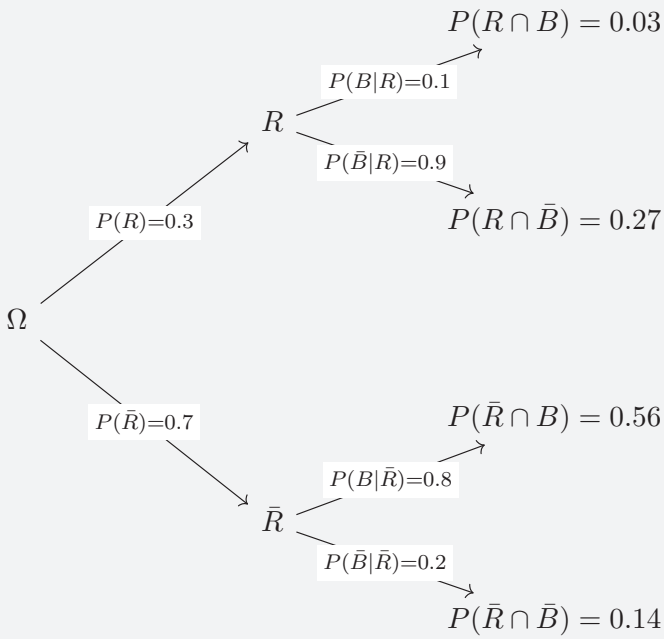
if $C_1 \cup C_2 \cup C_3 \cup ... \cup C_n = \Omega$ (complete)
and $C_i \cap C_j = \{\}$ (mutually exclusive)

2.1.3 Tree Diagram



Example:

There is a 30% probability that it rains in the morning (R). In this case, Prof. J goes to EPFL by bike with 10% probability. If it does not rain, he goes by bike with 80% probability. What is the total probability for going by bike (B) ?



Compute $P(B)$:

$$P(B) = P(B \mid R) \cdot P(R) + P(B \mid \bar{R}) \cdot P(\bar{R}) = 0.1 \cdot 0.3 + 0.7 \cdot 0.8 = 0.59$$

2.2 Bayes' Rule

Bayes' rule permits the inversion of the probability tree.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Example:
You see Prof. J come by bike (B is true). What is the probability that it was raining? (Compute $P(R \mid B)$)

$$P(R \mid B) = \frac{P(B \mid R) \cdot P(R)}{P(B)} = \frac{0.1 \cdot 0.3}{0.3 \cdot 0.1 + (1 - 0.3) \cdot 0.8} \approx 0.051 \approx 5.1\%$$

2.3 Random Variables

The function X assigns a numerical value x_i to each element of the sample space ω_i. Usually, this value is a real number, then :

$$X : \Omega \to \mathbb{R}$$

$$\omega_i \longrightarrow x_i$$

Example:

$$X(\text{“heads”}) = 1 \to \omega_1 = \text{“}heads\text{”} \to x_1 = 1$$

$$X(\text{“tails”}) = 0 \to \omega_2 = \text{“}tails\text{”} \to x_2 = 0$$

The function X is known as a random variable. We write :

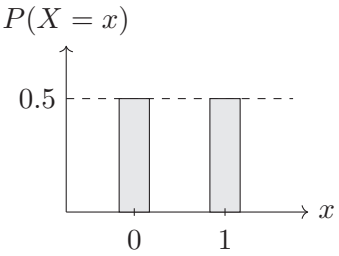
$$P(\omega_i) = p_i = p_{\omega_i}$$

$P(X)$ is the **probability mass function** for discrete state spaces.
$P(X)$ is the **probability density function** for continuous state spaces.

Example:

$P(\text{"heads"}) = p_1 = \frac{1}{2}$
hence $P(\text{"tails"}) = p_2 = \frac{1}{2}$

We can plot our probability function for a discrete Ω :



The connection between the elements and the assigned value can be very different.

For example we could have for an unordered two coins:
$\omega_1 = \{2 \text{ heads}\} \to x_1 = 100 \to p_1 = \frac{1}{4}$
$\omega_2 = \{2 \text{ tails}\} \to x_2 = -5 \to p_2 = \frac{1}{4}$
$\omega_1 = \{\text{one of each}\} \to x_3 = 0 \to p_3 = \frac{1}{2}$

2.4 Cumulative Distribution Function

What is the probability for $(X \leq x)$?

$$CDF(x) = P(X \leq x)$$

It is called a cumulative distribution function, CDF(x).

For a discrete distribution :

$$CDF(X) = \sum_{i=-\infty}^{X} P(X)$$

For a continuous distribution :

$$F(x) = \int_{-\infty}^{x} P(x)\,dx$$

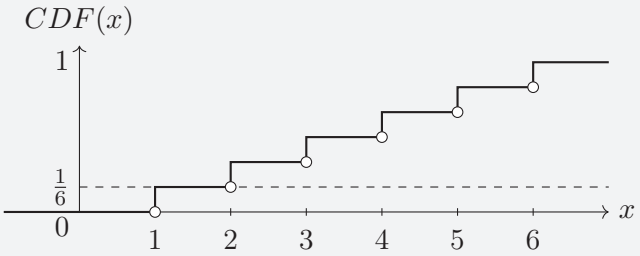
For CDF we also have :

$$\lim_{x\to-\infty} CDF(x) = 0 \qquad \lim_{x\to+\infty} CDF(x) = 1$$

CDF is non-decreasing :

$$CDF(x) \geq CDF(y) \to x \geq y$$

Example: dice throw



2.5 Expectation Value

If you sample a random variable infinitely many times, which arithmetic mean do you expect ?

Definition : the **expectation value**

$$\mathbb{E}(X) = \sum_{i=1}^{n} P(\omega_i) \cdot X(\omega_i = \sum x_i \cdot p_i = \mu \qquad \text{for a discrete distribution}$$

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} X \cdot P(X)\,dX \qquad \text{for a continuous distribution}$$

It is also called the mean gain or the **true mean**.

For a die:

$$\mathbb{E}(X) = \frac{1+2+3+4+5+6}{6} = 3.5$$

For the coin game above, we have:

$$\mathbb{E}(X) = \frac{1}{4}100 + \frac{1}{4}(-5) + \frac{1}{2}0 = 23.75$$

- A game is considered fair if $\mathbb{E}(X) = 0$
- We have $\mathbb{E}(f(x)) = \sum f(x_i) \cdot p(x_i)$

Examples:
$\mathbb{E}(aX+b) = \mathbb{E}(a) + \mathbb{E}(bX) = \sum a \cdot p_i + \sum b \cdot x_i \cdot p_i = a\sum p_i + b\sum x_i \cdot p_i = a + b\mathbb{E}(X)$

For a single coin toss with 0 for tails, 1 for heads:

$$\mathbb{E}(X) = 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = 0.5$$

$$\mathbb{E}(X^2) = \frac{1}{2} \cdot 0^2 + \frac{1}{2} \cdot 1^2 = 0.5 \neq (\mathbb{E}(X))^2$$

$$\mathbb{E}(cos(X)) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0.54 = 0.77 \neq cos(\mathbb{E}(X))$$

2.6 Variance

The mean $\mu = \mathbb{E}(X)$ gives the "central tendency". To measure spread, we define the **variance**:

$$\mathbb{V}(X) = \mathrm{Var}(X) = \mathbb{E}\big[(X-\mu)^2\big]$$

Equivalent formula:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Example:
For a die:

$$\mathbb{V}(X) = (\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)) - (\frac{21}{6})^2 = \frac{91}{6} - \frac{441}{36} \cong 2.9$$

For a coin:

$$\mathbb{V}(X) = (\frac{1}{2}(0^2+1^2)) - (\frac{1}{2})^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

Properties:

$$\mathbb{V}(aX) = a^2\mathbb{V}(X) \qquad \mathbb{V}(aX+b) = a^2\mathbb{V}(X)$$

$$\frac{\mathbb{V}(bX)}{\mathbb{E}(bX)} = b\frac{\mathbb{V}(X)}{\mathbb{E}(X)}$$

2.7 Standard Deviation

Definition:

$$\sigma = SD(X) = \sqrt{\mathbb{V}(X)}$$

Note:

$$\sigma(bX) = |b|\sigma(X)$$

So: $\dfrac{\sigma(bX)}{\mathbb{E}(bX)} = \dfrac{|b|}{b} \cdot \dfrac{\sigma(X)}{\mathbb{E}(X)} = sign(b) \cdot \dfrac{\sigma(X)}{\mathbb{E}(X)}$

2.8 Bernoulli Distributions

2 possible outcomes: $\Omega = \{\omega_0, \omega_1\} = \{\text{fail,win}\}$

- $X(\omega_0) = 0$
- $X(\omega_1) = 1$
- $P(\omega_0) = 1 - p$
- $P(\omega_1) = p$

$\rightarrow \mathbb{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p$
$\rightarrow \mathbb{V}(X) = \sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = (0^2 \cdot (1-p) + 1^2 \cdot p) - p^2 = p(1-p)$

2.9 Binomial distribution

Repeat Bernoulli experiment n times. The probability of k total "wins" is given by : $P(X = k)$

$X = X_1 + X_2 + ... + X_n \rightarrow$ total number of "wins"

$\Omega = \{(n \cdot \omega, 0 \cdot f), (n-1 \cdot \omega, 1 \cdot f), ..., (0 \cdot \omega, n \cdot f)\}$

We then have: $P(X = k) = p^k \cdot (1-p)^{n-k} \cdot \binom{n}{k}$

with $\binom{n}{k} = \dfrac{n!}{n!(n-k)!}$ where $n! = n \cdot (n-1) \cdot ... \cdot 1$

Example: (for n=2)
$P(X = 0) = (1-p)^2$
$P(X = 1) = p^2$
$P(X = 2) = p(1-p) \cdot 2$

Expectation value:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \cdots + X_n) = \sum_{j=1}^{n} \mathbb{E}(X_j) = n\,p$$

where each X_j is one Bernoulli experiment with parameter p.

Variance:

$$\mathbb{V}(X) = \sum_{j=1}^{n} \mathbb{V}(X_j) = n\,p(1-p)$$

because X_i and X_j are independent.

$$\frac{\sigma}{\mu} = \frac{\sqrt{\mathbb{V}(X)}}{\mathbb{E}(X)} = \frac{\sqrt{n\,p(1-p)}}{n\,p} = \frac{1}{\sqrt{n}} \frac{\sqrt{p(1-p)}}{p} \quad \Rightarrow \quad \text{relative width decreases.}$$

3 Normal distribution

3.1 Sum of expectation values

Let X_1 and X_2 be random variables on the same sample space $\{\omega_i\}$.

$$\mathbb{E}(X_1 \pm X_2) = \sum_i P(\omega_i)\big(X_1(\omega_i) \pm X_2(\omega_i)\big) = \sum_i P(\omega_i)X_1(\omega_i) \pm \sum_i P(\omega_i)X_2(\omega_i)$$

$$\mathbb{E}(X_1 \pm X_2) = \mathbb{E}(X_1) \pm \mathbb{E}(X_2) = \mu_1 \pm \mu_2$$

The expectation value of the sum (or the difference) of two random variables is the sum (or difference) of the individual expectation values.

3.2 Sum of variances

We compute the variance of a sum:

$$\mathbb{V}(X_1 + X_2) = \mathbb{E}\big[(X_1 + X_2)^2\big] - \big(\mathbb{E}(X_1 + X_2)\big)^2$$

Expand:

$$(X_1 + X_2)^2 = X_1^2 + 2X_1X_2 + X_2^2$$

So:

$$\mathbb{E}\big[(X_1 + X_2)^2\big] = \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1X_2) + \mathbb{E}(X_2^2)$$

Also

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Hence

$$\begin{aligned}\mathbb{V}(X_1 + X_2) &= \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1X_2) + \mathbb{E}(X_2^2) - \big(\mathbb{E}(X_1) + \mathbb{E}(X_2)\big)^2 \\ &= \mathbb{E}(X_1^2) - [\mathbb{E}(X_1)]^2 + \mathbb{E}(X_2^2) - [\mathbb{E}(X_2)]^2 + 2\big(\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)\big) \\ &= \mathbb{V}(X_1) + \mathbb{V}(X_2) + 2\mathrm{Cov}(X_1, X_2)\end{aligned}$$

Where $\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)$ is the **covariance**.
More generally:

$$\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) \pm 2\,\mathrm{Cov}(X_1, X_2)$$

If X_1 and X_2 are independent, $\mathrm{Cov}(X_1, X_2) = 0$ and

$$\mathbb{V}(X_1 \pm X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2)$$

3.3 More on Binomial Distribution

We have the following:
- n independent trials,
- each trial: 1 (success) with prob. p, and 0 (fail) with prob. $1-p$,
- let X be the *total number of wins*

The probability of having exactly k successes is :

$$P(X=k) = p^k(1-p)^{n-k}\binom{n}{k} \qquad k = 0, 1, \ldots, n$$

where $\binom{n}{k}$ is the binomial coefficient.

Note: For $p = \frac{1}{2} \to P(X=k) = (\frac{1}{2})^n \cdot \binom{n}{k}$

Mean and variance of binomial distribution

$X = X_1 + X_2 + \cdots + X_n$

All X_j are independent and each X_j is a Bernoulli event.

Mean

$$\mathbb{E}(X) = \mathbb{E}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = np$$

Variance:

$$\mathbb{V}(X) = \mathbb{V}\Big(\sum_{i=1}^{n} X_i\Big)$$

Independence implies

$$\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(X_i)$$

For a Bernoulli variable X_i:

$$\mathbb{E}(X_i) = p, \qquad \mathbb{V}(X_i) = p(1-p)$$

Therefore:

$$\mathbb{V}(X) = \sum_{i=1}^{n} p(1-p) = np \cdot (1-p) = \sigma^2$$

Standard deviation:

$$\sigma = \sqrt{\mathbb{V}(X)} = \sqrt{np \cdot (1-p)}$$

Relative width:

$$\frac{\sigma}{\mu} = \frac{\sqrt{\mathbb{V}(X)}}{\mathbb{E}(X)} = \frac{\sqrt{np(1-p)}}{np} = \frac{1}{\sqrt{n}} \cdot \frac{\sqrt{p(1-p)}}{p}$$

3.4 From Binomial to Gaussian/Normal

Can we get an approximate/smooth function for $P(X = k)$?

We use Stirling's/Demoivre's approximation:

(for $n \to \infty$)

$$n! \approx \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot \exp(-n)$$

Then we get:

$$P(X = k) \approx \frac{1}{\sqrt{2\pi\, np(1-p)}} \exp\left(-\frac{(k-np)^2}{2\, np(1-p)}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(k-\mu)^2}{2\sigma^2})$$

This corresponds to the *normal* or *Gaussian* distribution

$$\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(k-\mu)^2}{2\sigma^2}) = \mathcal{N}\big(\mu, \sigma\big)$$

3.5 The *Standard Normal* Distribution

Also called the **scaled and centered** or **reduced and centered** distribution.

We shift and scale all values:

$$z = \frac{X - \mu}{\sigma} \quad \text{[dimensionless]}$$

z means: "How many standard deviations away from the mean is a value?"

$$P(z) = \mathcal{N}(\mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Cumulative Distribution Function

$$P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z'^2/2}\, dz' = \Phi(z)$$

With $\Phi(z)$ the error function given by :

$$\Phi(z) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$$

We have the following:

$$P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) = \Phi(z)$$

$$P(X \leq \mu) = \Phi(0) = 0.5$$

$$P(X > a) = 1 - P(Z \leq a) = P(Z \leq -a) \quad \text{by symmetry of the Gaussian}$$

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

4 Probability intervals and Central Limit Theorem

4.1 From the X range / interval to P

→ We know which probability distribution is followed, as well as its parameters:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

→ We are asked the probability P to find a result between a and b, those being anything from $-\infty$ to $+\infty$

Recipe:

$$P(a \leq X \leq b) \quad \text{or, for } a = -\infty \quad P(X \leq B)$$

1. Go from X to Z:
$$Z = \frac{X - \mu}{\sigma}$$

2. Use the standard normal distribution:
$$P(X \leq b) = P\left(Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) \text{ (to be looked up in the table)}$$
$$= \frac{1}{2}(1 + erf(\frac{Z}{\sqrt{2}})) \text{ (calculator)}$$

4.1.1 Useful identities

$$P(Z \leq b) = 1 - P(Z > b)$$

$$P(Z \leq -a) = P(Z \geq a)$$

$$\Phi(-z) = 1 - \Phi(z)$$

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) = \Phi(Z_{high}) - \Phi(Z_{low})$$

Example: A population where $\mu = 180$ cm and $\sigma = 8$ cm
What's the probability of a randomly chosen person to be 196cm tall or smaller ?

$$P(X \leq 196) = P(Z \leq \frac{196 - 180}{8}) = \Phi(2) = 0.9772 \approx 98\%$$

4.2 Quantiles - From some p to a range/interval $-\infty$ to b

A **quantile** $\tilde{x_a}$ is the value of X under which a fraction a of the population sits.
- a **percentile** has a as a percentage (*e.g.* $\tilde{x_{95\%}}$ *means that 95% of the population is equal or below this value.*)
- a **quartile** is above either $1/4$ or $3/4$ of the values.
- the **median** has 50% of the population below or equal to its value.

$$P(X < \tilde{x_p}) = p$$

→ *Below which value of X do i find the result with probability p ?*

Recipe:
- Know P
- Look for: $\tilde{x_p}$
- Find z via $\Phi(z) = p$
- Compute $X = z \cdot \sigma + \mu$

Example: For the same distribution as before, below which height are 98% of the population?

$$0.98 = \Phi(z) \quad \Rightarrow \quad z \approx 2.06$$

$$\tilde{x_{0.98}} = z\sigma + \mu = 2.06 \cdot 8 + 180 \approx 196 \text{ cm}$$

4.3 Predictive Confidence Interval

"In which interval, symmetric around the mean are k% of the results of my distribution ?"

We have:

$$P(a < X \leq b) = p$$

$$P(\mu - c < X \leq \mu + c)$$

Then:

$$P(-\frac{c}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{c}{\sigma}) = P(-\frac{c}{\sigma} < Z \leq \frac{c}{\sigma}) = \Phi(\frac{c}{\sigma}) - \Phi(-\frac{c}{\sigma})$$

Finally:

$$\Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = k$$

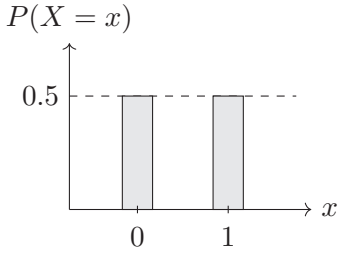
Example: For that same population once again, in which height range (symmetric around the mean) sit 95% of the population ?

$$2\Phi(z) - 1 = 0.95 \Rightarrow \Phi(z) = \frac{1 + 0.95}{2} = 0.975 \Rightarrow z \approx 1.96$$

$$c = z \cdot \sigma = 1.96\sigma = 15.68$$

$$\Rightarrow \text{Interval } = [180 \pm 15.7\text{cm}]$$

4.4 Bernoulli to Binomial review



For one experiment, we have $\mu_{single} = \frac{1}{2}$ and $\sigma_{single} = \frac{1}{2}$

$$X = X_1 + X_2 + X_3 ... + X_n$$

Knowing that all X_i are identical and independently distributed, we have:

$$\mu = \frac{n}{2} \quad \text{and} \quad \sigma = \frac{\sqrt{n}}{2}$$

$$Y = \frac{X}{n} \quad ; \quad E(Y) = \frac{1}{2} \quad ; \quad V(Y) = V(\frac{X}{n}) = \frac{1}{n^2} \cdot \mathbb{V}(X) = \frac{1}{4n} \quad \Rightarrow \quad \sigma_Y = \frac{1}{2\sqrt{n}}$$

4.5 General σ of averages

If:

$$X_G = \frac{X_1 + \cdots + X_n}{n} \qquad X_i \text{ being independent identically distributed events}$$

Then:

$$\mathbb{E}(X_G) = E(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n} \cdot n \cdot \mathbb{E}(X_i) = \mu_i = \mu_G$$

$$\mathbb{V}(X_G) = \frac{1}{n^2}(\mathbb{V}(X_1) + ... + \mathbb{V}(X_n)) = \frac{1}{n}\mathbb{V}(X_i) = \frac{\sigma_i^2}{n}$$

$$\Rightarrow \sigma_{X_G} = \frac{\sigma_i}{\sqrt{n}}$$

Example: Average weight of a grain of wheat
average weight : 50mg ; $\sigma = 10$mg
Looking at the average weight of a collection of 100 grains, in which range do we expect our results to be 95% of the time ?

$$\mu_G = \mu = 50\text{mg} \qquad \sigma_G = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1\text{mg}$$

$$\rightarrow \mathcal{N}(\mu_G, \sigma_G)$$

$$z = 1.96$$

So the range is $\mu_G \pm 1.96 \cdot \sigma_G \Rightarrow \mu_G = 50\text{mg} \pm 2\text{mg}$

4.6 Central Limit Theorem

For any distribution of X_i, the random variable

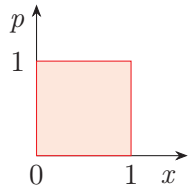
$$X_G = \frac{X_1 + \cdots + X_n}{n}$$

will be approximately Gaussian/Normal for large n, with $\mu_G = \mu_i$, $\sigma_G = \frac{\sigma_i}{\sqrt{n}}$

$$X_G \sim \mathcal{N}\left(\mu, \frac{\sigma_i^2}{n}\right)$$

5 Central Limit Theorem

5.1 CLT in action



$$\int_{-\infty}^{+\infty} p(x) = \int_{0}^{1} 1dx = 1$$

We now consider a random variable Y:

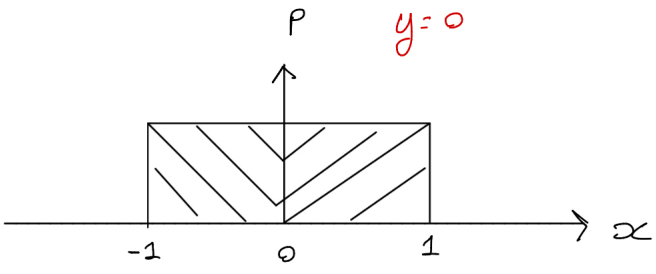
$$Y = \frac{X_1 + X_2}{2}$$

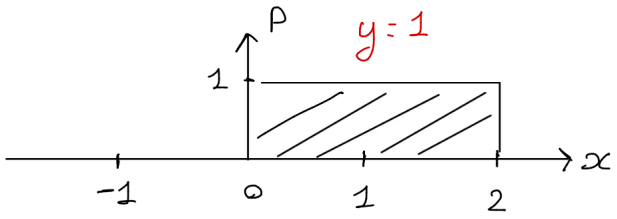
We have now:

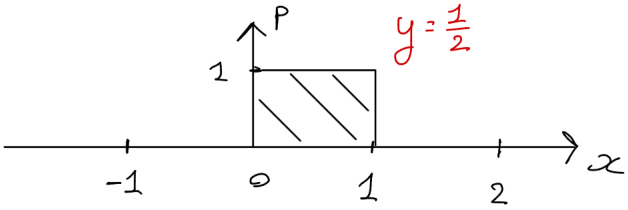
$$X_2 = 2Y - X_1$$

This means:

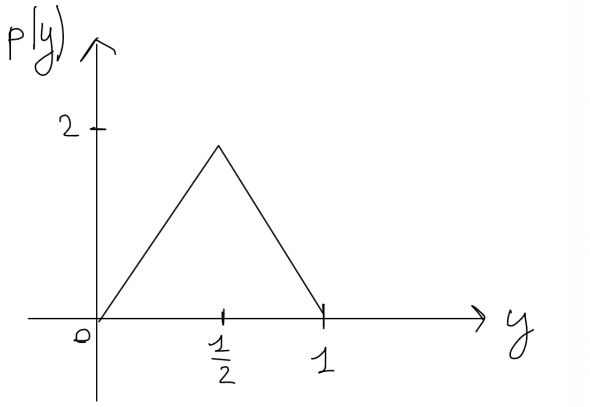
$$P(Y) = 2\int_{-\infty}^{+\infty} p(x) \cdot (2y - x)dx$$



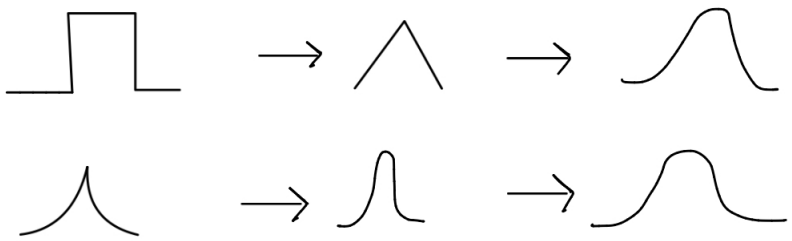




This leads to:



In general we have:



Let X_1 and X_2 two random variable such that :

$$X_1 \sim X_2 \sim \mathcal{N}(\mu = 0, \sigma^2 = \frac{1}{\sqrt{2}})$$

We know: $Y = \frac{X_1+X_2}{2}$ and we are looking for $p(y)$.

Reminder:

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} e^{-(x-a)^2}\, dx = \text{Cst}$$

We start from the expression of $p(y)$:

$$p(y) = C \int_{-\infty}^{\infty} p(x) \cdot p(2y - x)\, dx$$

$$= C \int_{-\infty}^{\infty} e^{-x^2} \cdot e^{-(2y-x)^2}\, dx$$

$$= C \int_{-\infty}^{\infty} e^{-x^2-4y^2-x^2+4yx}\, dx$$

$$= C \int_{-\infty}^{\infty} e^{-2(x^2+y^2-2yx)-2y^2}\, dx$$

$$= C\, e^{-2y^2} \cdot \int_{-\infty}^{\infty} e^{-2(x-y)^2}\, dx$$

We consider :

$$\int_{-\infty}^{\infty} e^{-2(x-y)^2}\, dx = \text{Cst}$$

Then :

$$p(y) = C' e^{-2y^2}$$

This leads to :

$$Y \sim \mathcal{N}\left(\mu_y = 0, \sigma_Y = \frac{1}{2}\right)$$

5.2 Real Data

Statistics is about :
- reducing data
- extracting $P(x)$ or its parameters from a sample $\overrightarrow{x}$

It works because of the **law of large numbers** : for large n (i.e. many data points), the relative frequencies of the values $\overrightarrow{x}$ will tend towards the underlying probability distribution.

Given data $\overrightarrow{x} = (170, 163, 186)$:

- **sample arithmetic mean**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{170 + 163 + 186}{3} = 171$$

- **sample biaised variance**

$$\tilde{S}^2_{\text{biased}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 48,6$$

- **sample biased standard deviation**

$$\tilde{S}_{biased} = \sqrt{\tilde{S}^2_{biased}}$$

Median: $\tilde{x} = \tilde{x}_{0.5}$: Half the data points are below or equals to this value, half above or equal. To find it :

1. sort the data $\rightarrow \overrightarrow{x_S} = (x_{S1}, x_{S2}, \ldots, x_{Sn})$
2. for n odd: middle value $\rightarrow \tilde{x} = x_{S\cdot(\frac{n}{2}+1)}$
3. for n even: average of the two middle values $\rightarrow \tilde{x} = \frac{x_{S\cdot(\frac{n}{2})} + x_{S\cdot(\frac{n}{2}+1)}}{2}$

The median is useful to tell if a distribution is assymetrical, because, contrary to the mean, it doesn't take outliers into account.

Example: Yearly income of 11 citizens

$$\tilde{x} = (90'000 \times (5 \text{ times}), 110'000 \times (5 \text{ times}), 10'000'000)$$

We have:

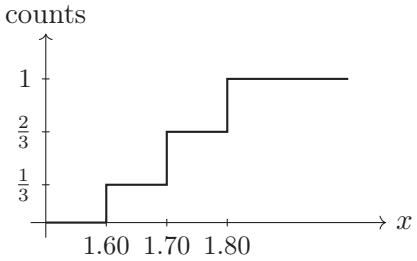
$$\overline{x} = 1'000'000\text{CHF} \qquad \tilde{x} = 110'000\text{CHF}$$

Quantiles

The q-quantile $\tilde{x}_q$ is such that a fraction q is below or equal to $\tilde{x}_q$ and $1 - q$ is above or equal to $\tilde{x}_q$.

Cumulative empirical distribution function

Example of empirical CDF (schematic):



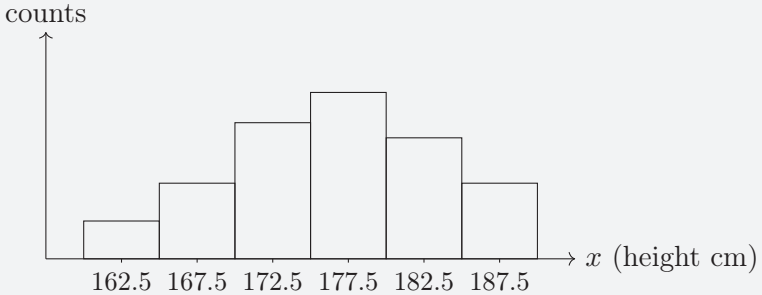
5.3 Histograms

To build a histogram:

- pick a **bin width** w (e.g. 5 cm),
- choose center bins (e.g. $170, 175, 180, \dots$),
- count how many observations fall into each bin.

Example with human heights (cm): let $w = 5\,\text{cm}$ and choose bins

$$(160, 165), (165, 170), (170, 175), (175, 180), (180, 185), (185, 190)$$



We can summarise the histogram in a small table of midpoints z_i and counts c_i:

bin center z_i (cm)	162.5	167.5	172.5	177.5	182.5	187.5
count c_i	2	4	7	9	6	4

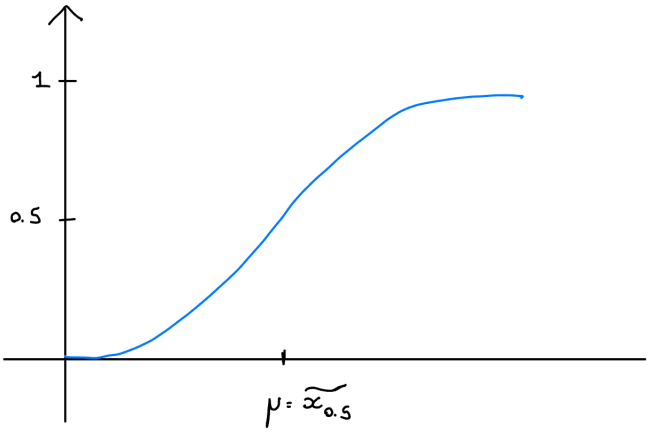
RECAP Ch 5

Continuous distribution

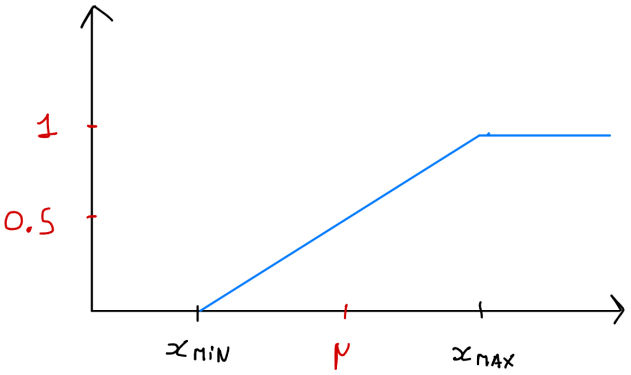
$$F(\bar{x}_q) = q$$

Example: $F(\bar{x}_{0.25}) = 0.25$

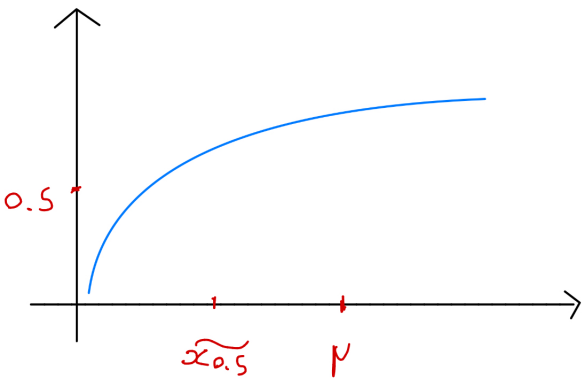
Gaussian (Normal) distribution:



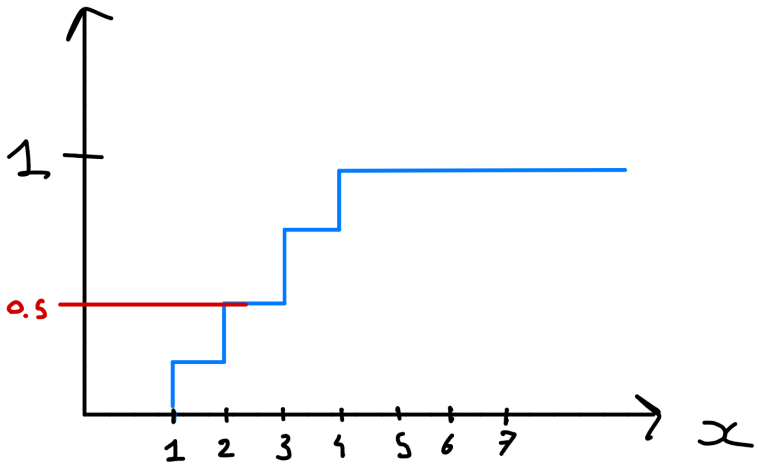
Uniform distribution:



Exponential distribution



Discrete distribution:



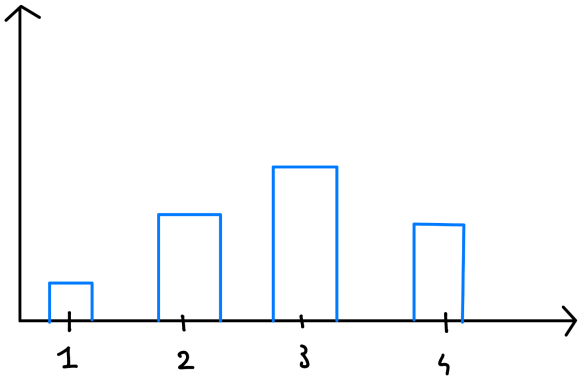
$\tilde{x}_{0.5} \rightarrow$ any value between 2 and 3 (convention $\rightarrow$ take 2.5)

What is the quantil for $\tilde{x}_{0.25}$:

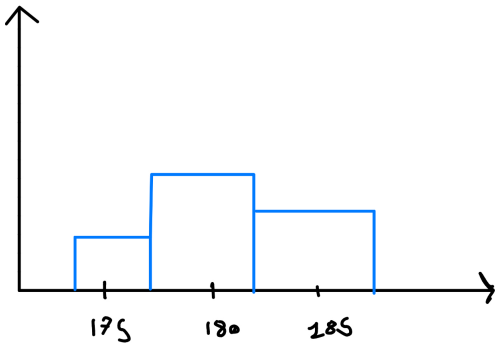
At least 25% of the probability is at or below $\tilde{x}_{0.25}$ and at least 75% of the probability is at or above $\tilde{x}_{0.25}$

Example:
$\tilde{x}_{0.2} = 0.2 \cdot 6 = 1.2 \rightarrow 2^{nd} number$
$\tilde{x}_{0.5} = 0.5 \cdot 6 = 3 \rightarrow$ between 3^{rd} and 4^{th}

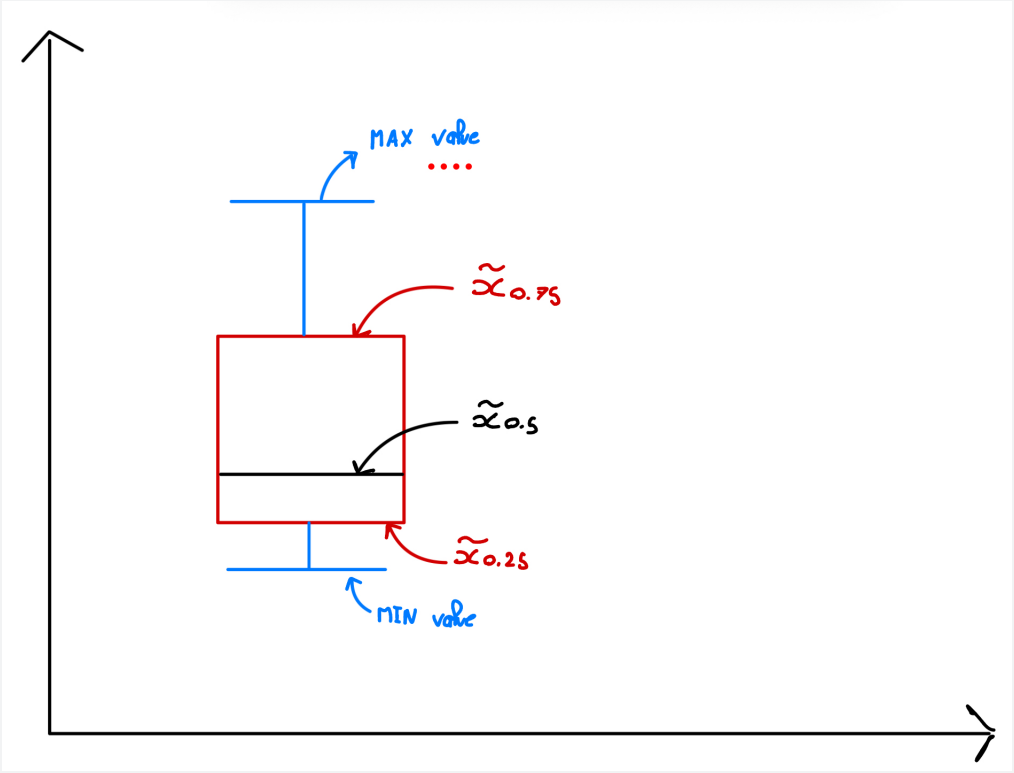
Histogram:
Discrete data



Binned data



Other example:

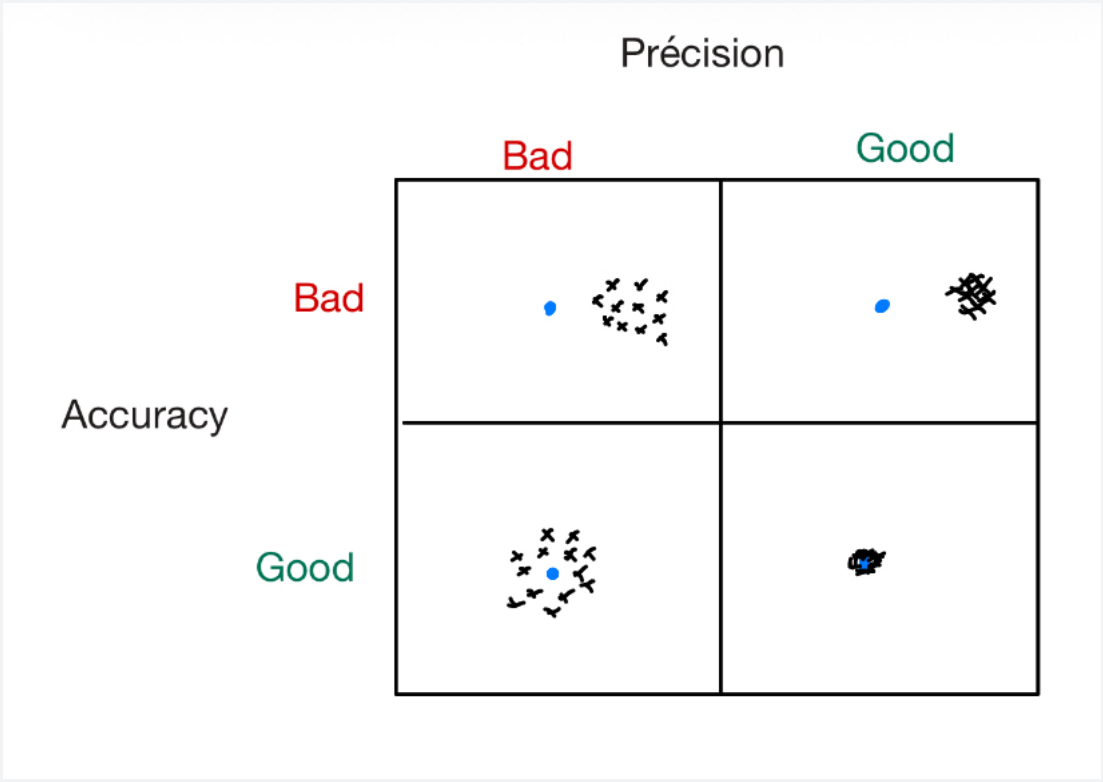


6 Estimators and bias

To go form data (samples) to discreet estimators, it should be.

- Accurate (close to the true value on average)
- Precise (low variance)

Example



6.1 Estimator for the Mean

Given $X_1, \ldots, X_n$ i.i.d :

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

$$\mathbb{E}(\bar{X}) = \mathbb{E}(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \cdot n \cdot \mathbb{E}(X) = \mu$$

$\rightarrow$ The sample mean is an **unbiased estimator** of μ.

6.2 Estimator for the Variance

$$\widetilde{S}^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \qquad \text{note: } \mathbb{V}(\bar{X}) = \frac{\mathbb{V}(X)}{n}$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

We have the following:

$$\mathbb{E}(\widetilde{S}) = \mathbb{E}\left[\frac{1}{n}\sum_i (x_i - \bar{x})^2\right] = \frac{1}{n}\sum_i \mathbb{E}(X_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n}\sum_i \mathbb{E}(X_i^2) + \frac{1}{n}\sum_i \mathbb{E}(\bar{x}^2) - 2\mathbb{E}\left(\bar{x}\frac{1}{n}\sum_i x_i\right)$$

$$= \frac{1}{n}n\mathbb{E}(X^2) + \frac{1}{n}n\mathbb{E}(\bar{X}^2) - 2\mathbb{E}(\bar{X}^2) = \mathbb{E}(X^2) - \mathbb{E}(\bar{X}^2)$$

$\mathbb{E}(\tilde{S})$ is also equal to:

$$\mathbb{E}(\widetilde{S}) = \mathbb{V}(X) + \mu^2 - \mathbb{V}(\bar{X}) - \mu^2 = \mathbb{V}(X) - \frac{1}{n}\mathbb{V}(X)$$

This means:

$$\mathbb{E}(\widetilde{S}^2) = \frac{n-1}{n}\sigma^2$$

$$\frac{n}{n-1}\cdot\widetilde{S}^2 = \frac{n}{n-1}\cdot\frac{1}{n}\sum_i (x_i - \bar{x})^2 = \boxed{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2}$$

$\Downarrow$ unbiased estimator for the variance

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

Example

For $n = 1$ and $\bar{x} = 3$:

$$\hat{S}^2 = \frac{(3-3)^2}{1} = 0$$

$$\mu = 35 \to S_\mu^2 = \frac{(3-\mu)^2}{1} = 0.5^2 = 0.25$$

Same for $n = 2$ and $\bar{x} = 4$:

$$\hat{S}^2 = \frac{(3-4)^2 + (5-4)^2}{2} = 1$$

$$S_\mu^2 = \frac{(3-3.5)^2 + (5-3.5)^2}{2} = 1.25$$

7 Z-test and t-test

7.1 Hypotheses

Yes/No questions

We want to answer questions such as:

- Is the mean weight of cars in the USA bigger than in Europe ?
- Is there a particle at 125 GeV ?
- Is the dice fair?

Example:

$$H_A : \text{“The die is fair”} \Rightarrow p_1 = p_2 = \cdots = \frac{1}{6}$$

Probability of rolling 6 ten times:

$$P(10 \text{ times } 6 \mid H_1) = \left(\frac{1}{6}\right)^{10} \approx 0.0000017\%$$

If this probability is extremely small $\Rightarrow$ we reject H_1 with a significance α:

$$\text{confidence} : 1 - \alpha = 99.999\% \quad \Rightarrow \quad \text{significance} : \alpha = (\frac{1}{6})^{10}$$

—

Alternative case:

$$H_2 : \ p_6 = 0.9, \quad p_2 = 0.01, \quad \text{other}=0$$

$$P(10 \text{ times } 6 \mid H_2) = (0.9)^{10} \approx 35\%$$

7.2 The Null Hypothesis H_0

Steps:

1. Get a yes/no hypothesis (ex: “my friend cheats” = “H_1 : the dice doesn’t follow a uniform distribution")

2. Take the logical opposite of H_1 : $\mathrm{H}_0 = \overline{\mathrm{H}_1}$: “the dice does follow a uniform distribution"

3. Choose a level of statistical significance

$$\text{Often } \alpha = 0.05$$

4. Compute: $P(\text{outcome} \mid H_0) \rightarrow$ if $P < \alpha \Rightarrow$ reject H_0

7.3 The z-test (Gaussian test for the mean)

Assumptions:
- We have measured N_s elements with mean $\bar{x}$
- We know (or assume) the standard deviation of the population σ
- Thanks to CLT, means follow a Gaussian distribution

When is this test useful ?
- fluctuations come from a well-characterized measurement method
- full population is known, but we want to see if one subgroup is representative

⋆ two-sided question → *what is the probability it's in a range ?*
- fixed deviation, search for P
- fixed P, search for deviation, infer σ from your data

⋆ one-sided question → *what is the probability it's in* $]-\infty, a]$ *or* $[b, +\infty[$ *?*
- fixed deviation, search for P

z-test Recipe : Define:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

For a fixed probability P (e.g. 95%) :
under which condition is the true mean of my distribution with probability P ?

- above $\mu_0 \Rightarrow \Phi(+z) = P$
- below $\mu_0 \Rightarrow \Phi(-z) = P$
- in the $\overline{x} \pm |\mu_0 - \overline{x}|$ range $\Rightarrow \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = P$

For a fixed μ_0 or deviation z :
with which probability P is the true mean that represents the population :

- above $\mu_0 \Rightarrow P = \Phi(z)$
- below $\mu_0 \Rightarrow P = \Phi(-z)$
- in the $\overline{x} \pm |\mu_0 - \overline{x}|$ range $\Rightarrow P = 2\Phi(z) - 1$

7.4 The t-test / Student's t-test

Method :
- measure n elements with mean $\overline{x}$ → *the minimum value of n for this is 2*
- standard deviation σ is not known and n is small
- we can assume a Gaussian distribution from the central limit theorem

The t distribution depends on $\bar{x}, S^2$ and the degrees of freedom ($df = \nu$):

$$\nu = n - 1$$

For $\nu \to +\infty$, the t-distribution is a Gaussian.
The distribution is overall wider than a Gaussian.

7.5 The one-sample t-test

We do not know σ, but we estimate the variance as:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

The t-value is:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

Degrees of freedom:

$$\nu = n - 1$$

$\Rightarrow \bar{x}$ follows a t-distribution with $(\bar{x}, S^2, \nu)$

7.6 The two-sample z-test

We compare the mean difference between two independent Gaussian distributions X and Y :

$$\sigma_x \text{ and } \sigma_y \text{ are known}$$

Mean:

$$\bar{x} - \bar{y}$$

Variance:

$$\mathbb{V}(X) + \mathbb{V}(Y) = \mathbb{V}(X - Y)$$

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

z quantifies the difference between two distributions.

How to choose a test

- if the variance σ^2 is known, or n is very large, there is no need for a t-test
- if n is small, prefer a t-test.

8 Two-sample tests

8.1 Two-sample tests

8.1.1 Two-sample z-test

- Two independent samples/datasets X, Y with measured means $\bar{x}, \bar{y}$
- We know σ_x, σ_y (or large n_x, n_y)

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Limits to consider :
1. $n_y \to \infty \Rightarrow z \to \frac{\bar{x}-\bar{y}}{\sigma_x/\sqrt{n_x}}$
2. $\sigma_x = \sigma_y = \sigma$, $n_x = n_y = N/2$

$$z = \frac{\bar{x} - \bar{y}}{2\sigma/\sqrt{N}}$$

8.1.2 Two-sample Student's t-test

We know $\sigma_x = \sigma_y = \sigma$.
We can estimate σ from :

$$S^2 = \frac{1}{(n_x - 1) + (n_y - 1)} \left(\sum_{i=1}^{n_x} (x_i - \bar{x})^2 + \sum_{j=1}^{n_y} (y_j - \bar{y})^2 \right)$$

Degrees of freedom:

$$\nu = N_T - 2 = (n_x - 1) + (n_y - 1)$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S^2}{n_x} + \frac{S^2}{n_y}}}$$

8.1.3 Two sample Welch / Behrens–Fisher test

- Two independent samples X, Y
- σ_x may differ from σ_y (estimate from S_x and S_y separately)

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

Effective degrees of freedom:

$$\nu = \text{round} \left[\frac{(S_x^2/n_x + S_y^2/n_y)^2}{\frac{(S_x^2/n_x)^2}{n_x - 1} - \frac{(S_y^2/n_y)^2}{n_y - 1}} \right]$$

Limits to consider :
1. $n_x \to \infty \Rightarrow \nu \to n_y - 1$
2. $S_x = S_y$, $n_x = n_y \Rightarrow \nu = N_T - 2$

8.2 Paired test

For two samples X, Y, the **pairwise difference** d is :

$$d_i = x_i - y_i$$

We consider a random variable D and we do a 1-sample t-test for d :

$$t_D = \frac{\bar{d} - \mu_{0_D}}{S_D/\sqrt{n_D}}$$

$$n_x = n_y = n_D \text{ by construction}$$

This way of comparing data gives better and cleaner data to analyse evolution.

Example:

x	y	$d = x - y$
101	91	10
98	89	9
122	112	10

Here, d clearly has smaller variance than x or y.

8.3 The χ^2 (chi-square) test

We have data: Can they be described by a certain probability density/mass function ?

Need discrete data $\rightarrow$ bin continuous data if needed.

Measure of deviation from expectation value:

$$\chi^2 = \sum_i \frac{(n_i - Np_i)^2}{Np_i} = N \sum_i \frac{(f_i - p_i)^2}{p_i}$$

Example: Dice throw

H_0 : The die is fair $\rightarrow p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$

We throw $N_T = 60$ times. Count number n_i for each outcome.

outcome x_i	1	2	3	4	5	6
proba p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Abs freq n_i	12	9	11	8	10	10
real freq $\frac{n_i}{N_T}$	$\frac{12}{60}$	$\frac{9}{60}$	$\frac{11}{60}$	$\frac{8}{60}$	$\frac{10}{60}$	$\frac{10}{60}$
expec value $N_T p_i$	10	10	10	10	10	10

$$N_T = 60 \qquad || \qquad \text{numbers of possible outcomes } k = 6$$

$$\chi^2 = \frac{(12-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(10-10)^2}{10} = 1$$

$$\nu = k - 1 = 6 - 1 = 5$$

Degrees of freedom:

$$df = \nu = k - 1 \quad \text{(PDF with no free parameters)}$$

If $\chi^2 \approx 0 \rightarrow$ data fits PDF well, no deviation
If $\chi^2 \gg 0 \rightarrow$ unlikely for this data to originate from this PDF

The χ^2 table shows the critical χ^2 for which the law is follow, meaning that any χ^2 below that indicates no deviation from the PDF, while any χ^2 above that indicates that the PDF is not followed.

χ^2 use criterion:

$$N \cdot \min(p_i) > 5 \qquad \text{(follows } N_T >> k\text{)}$$

If we also estimate m parameters of the PDF from our data:

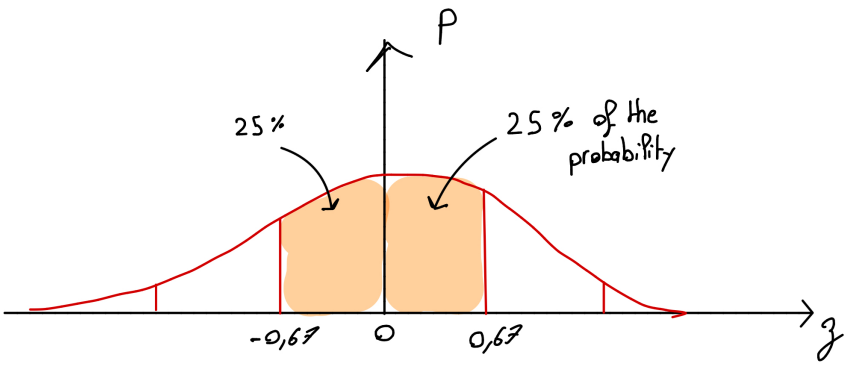
$$\nu = k - 1 - m$$

RECAP Ch 8:

The Chi-square test compares discrete (or binned) data to a probability mass (or binned density) distribution.

H_0 : (Null hypothesis) The proposed distribution can describe the data.

Example:



We consider a standard normal distribution. From the standard normal table, the value corresponding to 75% is:

$$z = 0.67 \rightarrow \textit{ we look in the table}$$

The data are divided into four bins defined by the cut points $-\infty$, -0.67, 0, 0.67 and $+\infty$

It gives us the following:

	$(-\infty, -0.67)$	$(-0.67, 0)$	$(0, 0.67)$	$(0.67, +\infty)$	Total
Expected counts $N \cdot p_i$	25	25	25	25	100
Observed counts	20	25	25	30	100
Difference	-5	0	0	$+5$	0
Chi-square contribution	$\frac{(-5)^2}{25}$	0	0	$\frac{5^2}{25}$	2

Test Statistic:

The chi-square test statistic is defined as:

$$\chi^2 = \sum_i \frac{(n_i - N_T \cdot p_i)^2}{N_T \cdot p_i}$$

Degrees of Freedom:

The degree of freedom is:

$$\nu = 4 - 1 = 3$$

Decision Rule:

At the 5% significance level, we compare the observed statistic with the critical value:

$$\chi^2_{0.95,\,3}$$

Since:

$$\chi^2 = 2 < \chi^2_{0.95,\,3}$$

we do not reject the null hypothesis.

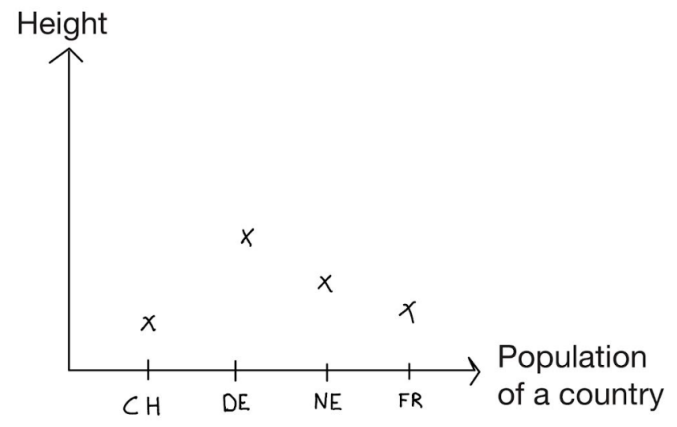
9 One-factor ANOVA test

9.1 Multiple Sample Testing: ANOVA

Goal: test whether the means of several groups are statistically equivalent or not. This is essentially an extension of Student's t-test to more than two groups.

$$H_0 : \mu_1 = \mu_2 = \dots$$

Example: Heights of people by country
H_0 : the mean height in all these countries is the same



Factor: The property of a group (nationality, temperature, material...) / independent variable / parameter

Levels: how many values the factor has

We name the data $x_{i,j}$, with i being the level and j the element within the group i.

Example: $x_{2,7}$ is the seventh German

9.1.1 1-Factor ANOVA

→ 1-factor ANOVA means no distinct separation within groups (e.g. if we examine nationality, we don't consider age, gender, wealth for each nationality).

H_0 : X_i does not depend on i

For each group i with sample size n_i:

$$\bar{x}_i = x_{i,\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j} \quad \text{is the average over } j \text{ for each level}$$

Variance estimator:

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$$

$$\sum_{j=1}^{n_i}(x_{i,j}-\bar{x}_i)^2 = SS_i \rightarrow \text{sum of squares for group } i$$

Total:

$$\bar{x}_T = \bar{x}_{\cdot,\cdot} = \frac{\sum_{i=1}^{I}(\sum_{j=1}^{n_i} x_{i,j})}{\sum_{i=1}^{I}(\sum_{j=1}^{n_i} 1)} = \frac{\sum_i n_i \cdot \bar{x}_i}{\sum_i n_i}$$

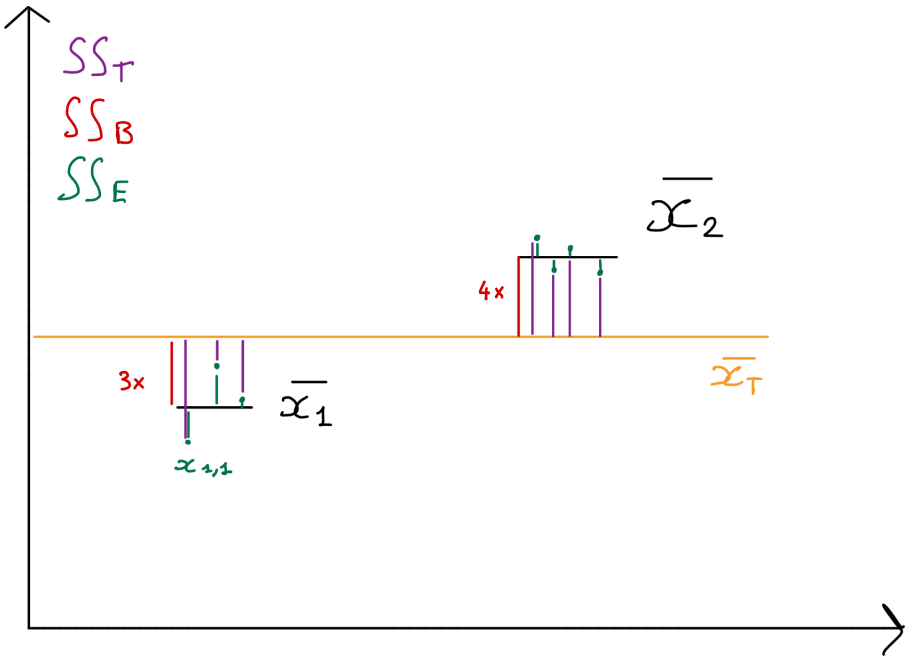
For all $n_i = n$

$$\bar{x}_T = \frac{1}{I}\sum_{i=1}^{I}\bar{x}_i \rightarrow \text{ means of means}$$

$$SS_T = \sum_i\sum_j (x_{i,j}-\bar{x}_T)^2 = \sum_i\sum_j ((x_{i,j}-\bar{x}_i)^2 + (\bar{x}_i-\bar{x}_T)^2)$$
$$= \sum_i SS_i + \sum_i n_i(\bar{x}_i-\bar{x}_T)^2 = SS_E + SS_B$$

where $SS_B \rightarrow$ square sum of between
$SS_E \rightarrow$ square sum of the error

- SS_E: variability **within** groups (Error)
- SS_B: variability **between** groups (Factor)



Expected Values under H_0

Assuming the groups come from the same probability distribution:

$$\mathbb{E}[SS_E] = \mathbb{E}\big[\sum_i S_i^2(n_i-1)\big] = \sum_i \sigma^2(n_i-1) = \sigma^2(N_T - I)$$

For all $n_i = n$:

$$\mathbb{E}(SS_E) = \sigma^2 \cdot I \cdot (n-1) \quad \textit{where } I \textit{ in the number of levels}$$

$$\mathbb{E}[SS_B] = \mathbb{E}\big[\sum_i n_i(\bar{x}_i - \bar{x}_T)^2\big]$$

If all $n_i = n$:

$$\mathbb{E}(SS_B) = n\mathbb{E}\big[\sum_i S^2_{\bar{x}_i} \cdot (I-1)\big] = n(I-1) \cdot \sigma_{\overline{x}^2 = \sigma^2(I-1)}$$

where: $\sigma^2_{\bar{x}_i} = \dfrac{\sigma^2}{n} = \sigma_B$ and $S^2_{\overline{x}} = \dfrac{\sum(\overline{x}_i - \overline{x}_T)^2}{I-1}$

F-statistic

We define MS_E the mean square error and MS_B the mean square between. They both have different degrees of freedom.

$$MS_E = \frac{SS_E}{N_T - I} \qquad \textit{and} \qquad MS_B = \frac{SS_B}{I-1}$$

$$\nu_E = N_T - I \qquad \textit{and} \qquad \nu_B = I - 1$$

for $n_i = n \quad \forall i$:

$$\mathbb{E}(MS_E) = \mathbb{E}(MS_B) = \sigma^2$$

and so we define the F-statistic as :

$$F_{measured} = \frac{MS_B}{MS_E}$$

Decision rule (95% confidence):

$$F_{\text{measured}} > F_{\nu_B, \nu_E}\textit{(95\%)} \Rightarrow \text{reject } H_0$$

ANOVA Summary Table

Source of Variation	df	SS	MS	$F_{Measured}$
Between groups (Factor)	$I-1$	SS_B	$SS_B/(I-1)$	MS_B/MS_E
Within groups (Error)	$N_T - I$	SS_E	$SS_E/(N_T - I)$	
Total	$N_T - 1$	SS_T		

Example:

i	copper	gold	silver
$\bar{x}_i$	11	12	13
S_i	1	2	1
n_i	10	20	10

- $I = 3$
- $N_T = 40$
- $\bar{x}_T = 12$
- $SS_B = 10(-1)^2 + 20(0)^2 + 10(-1)^2 = 20$
- $MS_B = \frac{20}{2} = 10$
- $SS_E = 1^2(10-1) + 2^2(20-1) + 1^2(10-1) = 94$
- $MS_E = \frac{94}{37} = 2,54$

$$F_M = \frac{10}{2,54} = 3,94 > F_{2,37} = 3,25$$

$\rightarrow$ we can reject the hypothesis $\rightarrow$ There is a difference between each of these metals' studied property.

10 Linear regression

We have data $\{(x_i, y_i)\}$.

10.1 Constant function

- ⋆ **Model:** $y = a$ (a is some constant)
- ⋆ **Estimator:** $\hat{y} = \hat{a}$
- ⋆ **Residual:** $r_i = y_i - \hat{y}_i(x) = y_i - \hat{a}$

We want to minimize the squared residual. Thus, we need to find $\hat{a}$ so that

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{a})^2$$

has the smallest possible value.

Take the derivative:

$$\frac{\partial SS_E}{\partial a} = \sum_{i=1}^{n} 2(y_i - a)(-1) = -2 \sum_{i=1}^{n} (y_i - a) \stackrel{!}{=} 0$$

Hence

$$-2n\overline{y} + 2na \stackrel{!}{=} 0 \quad \Rightarrow \quad \hat{a} = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$$

Also

$$\frac{\partial^2 SS_E}{\partial a^2} = 2n > 0 \qquad \hat{a} = \overline{y} \text{ is the only minimum of } \mathrm{SS}_E$$

⇒ **The mean is then the best estimator for a constant function.**

10.2 Linear function

Model:

$$y = a + bx$$

where a is the intercept and b is the slope.

We want to find $\hat{a}, \hat{b}$ that minimize the residual.

Residual:

$$r_i = y_i - (\hat{a} + \hat{b} x_i)$$

The sum of squared errors is

$$SS_E = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Derivative with respect to $\hat{a}$

$$\frac{\partial SS_E}{\partial a} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = -2\sum_{i=1}^{n}(y_i - a - bx_i) \stackrel{!}{=} 0$$

Therefore

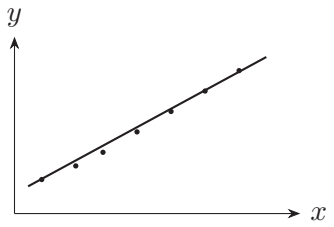
$$\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i \stackrel{!}{=} 0$$

so

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

We check:

$$\frac{\partial^2 SS_E}{\partial a^2} = 2n > 0$$



Derivative with respect to $\hat{b}$

$$\frac{\partial SS_E}{\partial b} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i) = -2\sum_{i=1}^{n} x_i(y_i - a - bx_i) \stackrel{!}{=} 0$$

Substitute $a = \bar{y} - b\bar{x}$ and simplify, which gives

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Check:

$$\frac{\partial^2 SS_E}{\partial b^2} = 2\sum_{i=1}^{n} x_i^2 \geq 0$$

$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ measures how x and y vary together around their means (negative or positive slope).

10.3 ANOVA for linear regression

We consider the null hypothesis:

$$H_0 : y \text{ does not depend on } x \quad \Longleftrightarrow \quad b = 0 \quad \text{(hypothesis)}.$$

ANOVA table for linear regression

Source	df	SS	MS	F_M
Model	1	$SS_M = \sum(\hat{y}(x_i) - \bar{y})^2$	$MS_M = \dfrac{SS_M}{\nu_M}$	$F_M = \dfrac{MS_M}{MS_E}$
Error (residual)	$n-2$	$SS_E = \sum(y_i - \hat{y}(x_i))^2$	$MS_E = \dfrac{SS_E}{\nu_E}$	
Total	$n-1$	$SS_T = \sum(y_i - \bar{y})^2$		

$$\nu_M = n - 1 \qquad | \qquad \nu_E = n - v_M - 1 = n - 2$$

Here, n is the number of slopes.

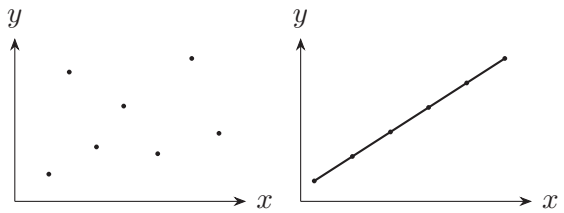
If $F_M > F_{\nu_M,\nu_E}$ (e.g. at $P = 95\%$), then we **reject** H_0, so there is a correlation between y and x, which can mean that using the model is justified. It is however always useful to superimpose the model function over the data to check for visible correlation.

Goodness of fit for regression

$$R^2 = \frac{SS_M}{SS_T} = 1 - \frac{SS_E}{SS_T} \qquad | \qquad 0 < R^2 < 1$$

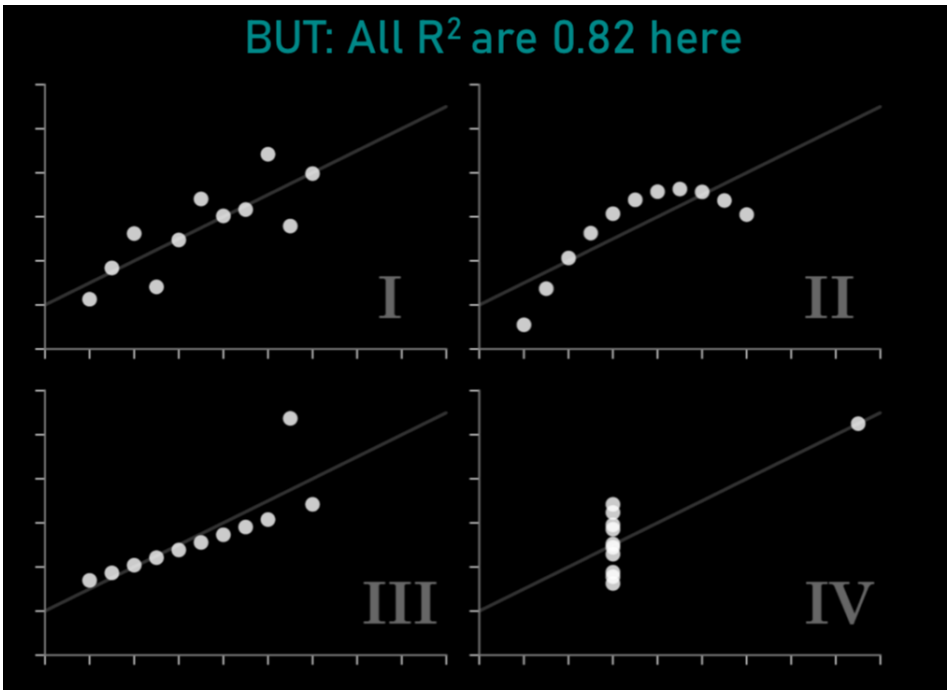
The closer R^2 is to 1, the better the fit is.

Terrible fit vs perfect fit (schematic):



$R^2 \approx 0$ $\qquad$ $R^2 \approx 1$

However, R^2 is not in itself a guarantee of the fit ; it is always important to plot the data to see.



10.4 What is "linear" in linear regression?

We are interested in linearity in the parameters a and b. Could the following functions be used to do a linear regression?

$y = a + bx$	Yes
$y = a + bx^2$	Yes
$y = a + b \sin x$	Yes
$y = a + \sin(bx)$	No

The first three are linear in the parameters a, b (the parameters appear only to the first power, not inside non-linear functions). The last one is **non-linear** in b.

Shape of the SS_E as a function of b

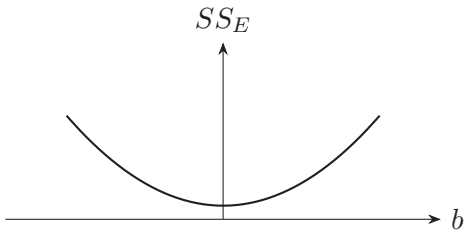
For a linear model in the parameters, say

$$y = a + bx$$

the function

$$SS_E(b)$$

as a function of b has a single minimum (a "nice" convex shape).

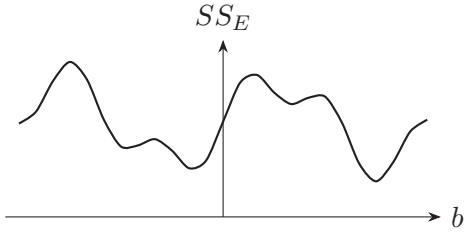


linear case $y = a + bx$

For a nonlinear model like

$$y = a + \sin(bx)$$

the corresponding $SS_E(b)$ can have many local minima and a complicated shape. It is very difficult to guarantee that an algorithm finds the *global* optimum.



nonlinear case $y = a + \sin(bx)$

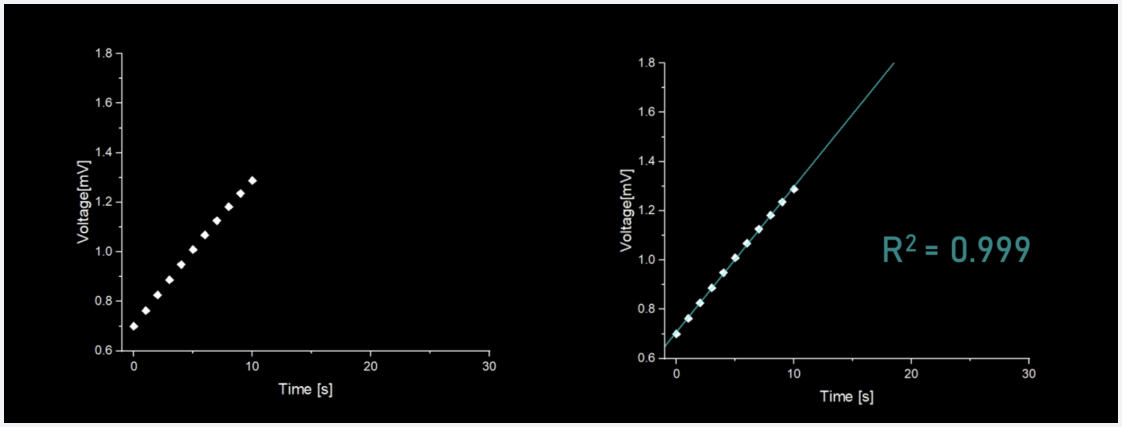
10.5 Assumptions for Linear Regression

We assume that :

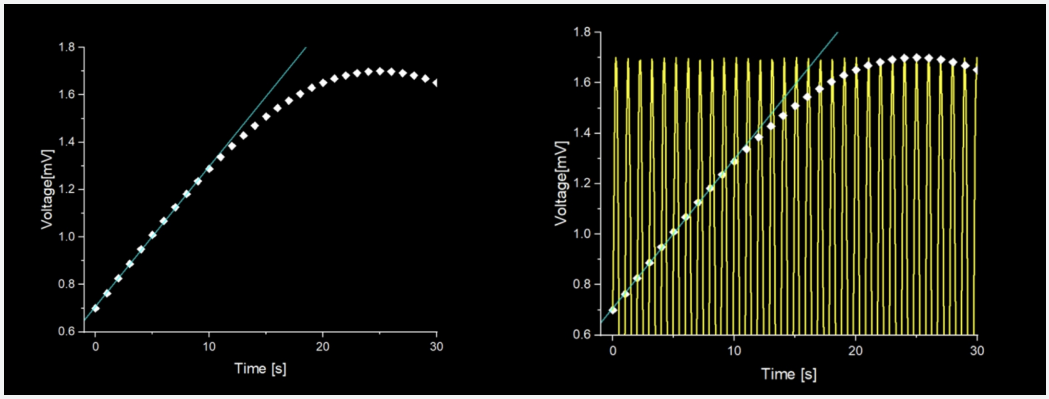
- the residual is a random variable : $R \sim \mathcal{N}(0, \sigma^2)$
- the residual is independent of x and y
- there is no error in x

Fitting commandment: Only trust the model for the values where the data is.

Example: We take some data and approximate the linear regression



In reality, if we look at more data, this linear regression can become totally false !



On that last graph, we can also observe that the data is taken at a certain frequency, and so we can't trust what is going on between the data points.

10.6 Parameter Uncertainty

$$R \sim \mathcal{N}(0, \sigma^2)$$

$$\rightarrow S_E^2 = \frac{1}{n-2} \sum_i^n (r_i^2) \qquad || \qquad r_i = y_i - \hat{y}(x_i)$$

Variances of estimators:

$$\mathbb{V}(\hat{a} + \hat{b}x_j) = \frac{\sigma^2}{n} + \frac{\sigma^2 (x_j - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

Where $\frac{\sigma^2}{n}$ is the uncertainty about the height at which the curve lies, and the last term the uncertainty about the slope.

$$\mathbb{V}(\hat{b}) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

11 Linear regression with more than 1 slope

Multi-factor :

$$y = a + b_1 x_1 + b_2 x_2 + \dots$$

$$y_i = a + b_1 x_{1,i} + b_2 x_{2,i} + r_i$$

Multi-function :

$$y = a + b_1 x + b_2 x^2 + \dots$$

$$y_i = a + b_1 x_i + b_2 x_i^2 + r_i$$

A model with n parameters can fit n data points perfectly.

11.1 Model Comparison

$n = 10$ for all the models.

Model	ν_M	ν_E	SS_E
M1: $y = a$	0	9	something
M2: $y = a + bx$	1	8	less
M3: $y = a + b_1 x + b_2 x^2$	2	7	even less

Nested model:

M1 is "nested" in M2 if you can get M1 by setting one or more slopes in M2 to 0. In this case,
- M2 $\rightarrow$ full model
- M1 $\rightarrow$ reduced model

$$\Delta SS_E = SSE_{\text{reduced}} - SSE_{\text{full}}$$

$\Delta p =$ Difference in free parameters // number of b's set to 0 from M2 to M1

11.2 Model Comparison: F-Test

H_0: The reduced model describes the data as well as the full model

$$F_m = \frac{\Delta SS_E / \Delta p}{SS_{E_{\text{full}}} / \nu_{m,Full}} \quad \overset{H_0}{>} \quad F_{\Delta p, \nu_{E,Full}}$$

Example:
Compare a cubic polynomial ($a + b_1x + b_2x^2 + b_3x^3$) to a quintic one ($a + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5$) for 10 data points.

$$SS_{E_{\text{full}}} = 180 \qquad SS_{E_{\text{reduced}}} = 400$$

$$\Delta SSE = 400 - 180 = 220$$

$$F_m = \frac{220/2}{180/(10-5-1)} = \frac{110}{45} = 2.4$$

$$F_{2,4} \approx 6.9$$

$\Rightarrow H_0$ is NOT rejected $\rightarrow$ Cubic model is enough $\rightarrow$ No need for the quintic one

11.3 Matrix Form of Linear Regression

$$y_1 = a + b_n x_{1,1} + b_2 x_{2,1} + \cdots + r_1$$
$$y_2 = a + b_n x_{1,2} + b_2 x_{2,2} + \cdots + r_2$$
$$\vdots$$

We write:

$$\overrightarrow{y} = X\overrightarrow{\beta} + \overrightarrow{r}$$

With:

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \cdots \\ 1 & x_{1,2} & x_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$\overrightarrow{\beta} = (a, b_1, b_2, \dots)$$

$$\hat{\beta} = (X^\top X)^{-1} X^\top \overrightarrow{y}$$

12 Multi-factor statistics

We have a normal distribution for a population (e.g. people's heights)

$$X \sim N(\mu, \sigma^2)$$

The data is characterized by the following table:

	Taille moyenne en cm		Echantillon	Population
	Moyenne	IC +/-	n	N
Total	**171.1**	**0.2**	**21 873**	**7 182 252**
Sexe				
Hommes	177.6	0.2	10 122	3 548 077
Femmes	164.7	0.2	11 751	3 634 175
Âge				
15-24 ans	172.7	0.5	2 115	804 408
25-34 ans	172.9	0.5	2 173	1 110 406
35-44 ans	172.3	0.4	3 115	1 218 144
45-54 ans	171.6	0.4	3 865	1 223 931
55-64 ans	171.1	0.3	4 202	1 190 409
65-74 ans	168.8	0.4	3 418	847 919
75+ ans	166.8	0.4	2 985	787 035

To have 95% probability to have the mean inside the range (-0.2;+0.2):

$$\Phi(z) - \Phi(-z) = 0.95$$

$$\Phi(z) - (1 - \Phi(z)) = 2\Phi(z) - 1 = 0.95$$

$$\Phi(z) = 0.975 \rightarrow z = 1.96$$

We have : $z \cdot \frac{\sigma}{\sqrt{n}} = 0.2\text{cm} \rightarrow \sigma = \frac{0.2 \cdot \sqrt{10,000}}{2} = 10 \text{ cm}$

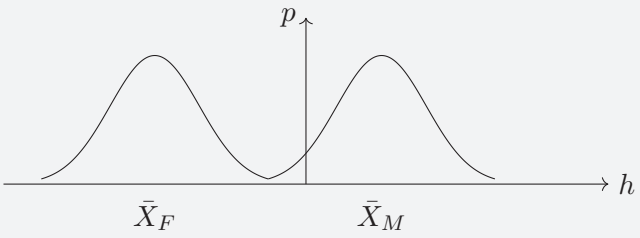
Comparing two population means
Suppose we compare two populations (e.g. average height of men vs. women, or two age groups). Let

$$\bar{X}_M,\ \bar{X}_F$$

Two sample z-test (because n=10,000)

$$z = \frac{\bar{X}_m - \bar{X}_F}{\sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_F^2}{n_F}}} = \frac{12.9}{\sqrt{\frac{10^2+10^2}{10{,}000}}} = 91 \to p = 1 - \frac{1}{10^{180}}$$

Graphically, the population distributions are wide, but the distributions of the sample means are much narrower ("uncertainty of the mean").



Correlation and causation

Correlation does not necessarily mean causation; there can be different causal structures:

1. Direct cause: $A \to B$.
2. Common cause: A and B are both influenced by C.
3. Intermediate cause: $A(\to C) \to B$.

12.1 Multi-factor situations

Example: we look at the height of people, grouped by :
- factor I: age group, $i = 1, \ldots, 7$
- factor J: sex, $j = 1, 2$ (e.g. male/female)

We can then have $x_{i,j,k} = x_{2,2,521} = 163,2\text{cm}$.
In this case, this person is a woman in the 25-34 age group.

The average height of this age group is then: $\bar{x}_{2,2,\cdot} = \frac{1}{n_{2,2}} \sum_{k=1}^{K} x_{2,2,k}$

The average height over all women (summed over age groups) is :

$$\bar{x}_{\cdot,2,\cdot} = \frac{1}{\sum_i \sum_k 1} \sum_i \sum_k x_{i,2,k}$$

and the total mean (all people, all groups) is :

$$\bar{x} = \frac{1}{\sum_i \sum_j \sum_k 1} \sum_i \sum_j \sum_k x_{i,j,k}$$

12.1.1 Table of means

There are several ways of presenting data : either by showing each of them, or grouping them.

Raw Data, element table:

ID	Sex	Age	Height
Jean	M	27	176
Marie	F	38	178

Statistical table:

Height	Male 20y	Female 20y	Male 21y
Mean	176 cm	166 cm	175 cm
Variance	1.7 cm^2	2.0 cm^2	2.6 cm^2
n	98	93	103

We can arrange the group means in a table:

	$i=1$	$i=2$	...	$i=I$	partial means
$j=1$	$\overline{x}_{1,1,\cdot}$	$\overline{x}_{2,1,\cdot}$	...	$\overline{x}_{I,1,\cdot}$	$\overline{x}_{\cdot,1,\cdot}$
$j=2$	$\overline{x}_{1,2,\cdot}$	$\overline{x}_{2,2,\cdot}$	...	$\overline{x}_{I,2,\cdot}$	$\overline{x}_{\cdot,2,\cdot}$
...	...	...	...	...	...
$j=J$	$\overline{x}_{1,J,\cdot}$	$\overline{x}_{2,J,\cdot}$	...	$\overline{x}_{I,J,\cdot}$	$\overline{x}_{\cdot,J,\cdot}$
partial means	$\overline{x}_{1,\cdot,\cdot}$	$\overline{x}_{2,\cdot,\cdot}$	...	$\overline{x}_{I,\cdot,\cdot}$	$\overline{x}_T$

Example: Resistance data (in Ω), $n = 10$ for all i

	Copper	Silver	All (partial mean)
20°C	10	12	11
60°C	11	13	12
100°C	15	17	16
All (partial mean)	12	14	13 (total mean)

We decompose the total sum of squares into components:

$$SS_{i,j} = \sum_k (x_{i,j,k} - \bar{x}_{i,j,\cdot})^2 \text{ (for raw data)} = (n_{i,j}-1)\cdot S_{i,j}^2 \text{ (cumulative)}$$

$$SS_E = \sum_i \sum_j SS_{i,j}$$

$$SS_{B,j} = \sum_i \sum_j n_{i,j}(\bar{x}_{i,j,\cdot} - \bar{x}_T)^2 = In \sum_j (\overline{x}_{i,j,\cdot} - \overline{x}_T)^2 \text{ if all } n \text{ equal}$$

$$SS_T = \sum_i \sum_j \sum_k (x_{ij,k} - \bar{x}_T)^2$$

12.2 Non-interacting 2-factor ANOVA

Source	ν	SS	MS	F
Factor I	$I-1$	$SS_{B,I}$	$\frac{SS_{B,I}}{\nu_I}$	$\frac{MS_{B,I}}{MS_E}$
Factor J	$J-1$	$SS_{B,J}$	$\frac{SS_{B,J}}{\nu_J}$	$\frac{MS_{B,J}}{MS_E}$
Error	$\sum_i \sum_j (n_{i,j}-1) = IJ(n-1) \rightarrow$ if n same	SS_E	$\frac{SS_E}{\nu_E}$	
Total	$N_T - 1 = IJn - 1 \rightarrow$ if all n equal	SS_T		

12.3 Interacting factors

Non-interacting model:

$$X_{i,j} = A_i + B_j$$

Interacting model:

$$X_{i,j} = A_i + A_i \cdot B_j$$

⋆ Alternative interacting forms:

$$X_{i,j} = A_i \cdot B_j, \qquad X_{i,j} = e^{A_i} \cdot \sin B_j$$

⋆ log-transform example (to obtain non-interacting model) :

$$\log(X_{i,j}) = \log(A_i \cdot B_j) = \log A_i + \log B_j = C_i + D_j$$

Example: Weight in grams of weirdly-filled crêpes

j/i	No cheese	Cheese	All
No banana	$100 = (170 - 20 - 50)$	140	$120 = (170 - 50)$
Banana	200	240	$220 = (170 + 50)$
All	$150 = (170 - 20)$	$190 = (170 + 20)$	170

For the non-interacting case, we have :

$$x_{i,j}^{Expected} = \bar{x}_T + (\bar{x}_{i,\cdot,\cdot} - \bar{x}_T) + (\bar{x}_{\cdot,j,\cdot} - \bar{x}_T)$$

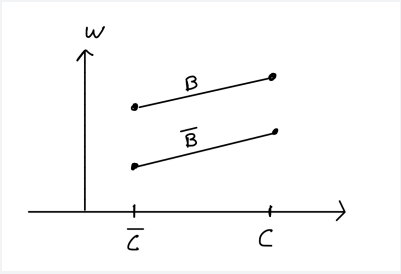
For the interacting case we have :

$$\begin{aligned} x_{i,j}^{Interaction} &= \bar{x}_{i,j,\cdot} - \bar{x}_{i,j}^{Expected} \\ &= \bar{x}_{i,j,\cdot} - \bar{x}_T - (\bar{x}_{i,\cdot,\cdot} - \bar{x}_T) - (\bar{x}_{\cdot,j,\cdot} - \bar{x}_T) \\ &= \overline{x}_{i,j,\cdot} - \overline{x}_{i,\cdot,\cdot} - \overline{x}_{\cdot,j,\cdot} + \overline{x}_T \end{aligned}$$

Detecting Interaction - Graphical View

- **No interaction**: lines are parallel
- **Interaction**: lines cross or diverge

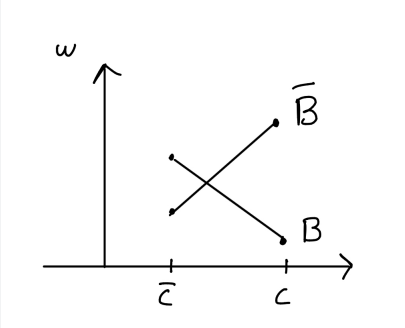
For the previous example, we have:



→ No interaction

Example: taste score of those weirdly-filled crêpes

j/i	$\bar{C}$	C	All
$\bar{B}$	3	7	5
B	5	1	3
All	4	4	4



→ Interacting

Deviation from $\bar{x}_T$:

-1	+3	+1
-1	-3	-1
0	0	0

Expected deviation:

+1	-1
-1	-1

Expected (non-interacting) value:

+5	+5
+3	+3

Interaction:

$$x_{i,j}^{Interaction} = \bar{x}_{i,j,\cdot} - \bar{x}_{i,\cdot,\cdot} - \bar{x}_{\cdot,j,\cdot} + \bar{x}_T$$

-2	+2	0
+2	-2	0
0	0	0

Example: Noise in a Room

j/i	No Train	Train	All
Window open	10	30	20
Window closed	1	3	2
All	5.5	16.5	11

Expected:

14.5	25.5
-3.5	7.5

Interacting:

$$x_{1,1}^{Interaction} = 10 - 20 - 5.5 + 11 = -4.5$$

Sum of squares of interaction:

$$SS_{B,I:J} = n\sum_i\sum_j (x_{i,j}^{Interaction})^2 == n\sum_i\sum_j (\bar{x}_{i,j,\cdot} - \bar{x}_{i,\cdot,\cdot} - \bar{x}_{\cdot,j,\cdot} + \bar{x}_T)^2$$

$$SS_T = SS_E + SS_{B,I} + SS_{B,J} + SS_{B,I:J}$$

ANOVA Table (Two-Factor with Interaction) :

Source	df	SS	MS	F
Factor I	$I-1$	$SS_{B,I}$	$\frac{SS_{B,I}}{\nu_i}$	$\frac{MS_{B;I}}{MS_E}$
Factor J	$J-1$	$SS_{B,J}$	$\frac{SS_{B,J}}{\nu_J}$	$\frac{MS_{B,J}}{MS_E}$
Interaction $I:J$	$(I-1)(J-1)$	$SS_{B,I,J}$	$\frac{SS_{B,I,J}}{\nu_{I,J}}$	$\frac{MS_{IJ}}{MS_E}$
Error	$IJ(n-1)$	SS_E	MS_E	
Total	$IJn-1$	SS_T		

$$\nu_I = I - 1$$

$$\nu_J = J - 1$$

$$\nu_{IJ} = (I-1)(J-1) = \nu_i \cdot \nu_j$$

$$\nu_E = I\,J\,(n-1)$$

$$\begin{aligned}\nu_T &= (I-1) + (J-1) + (I-1)(J-1) + IJ(n-1)\\ &= I + J - 2 + IJ - I - J + 1 + IJn - IJ\\ &= IJn - 1\\ &= N_T - 1\end{aligned}$$

Example: degrees of freedom

j \ i	a	b	c	All
α	1	2	*None*	1
β	3	4	*None*	2
γ	5	6	*None*	3
δ	*None*	*None*	*None*	*None*
All	1	2	*None*	$\bar{x}_T$

- To be determined
- Fixed (Given)

13 Propagation of error/uncertainty

13.1 Expectation Value

Let X be a random variable.

$$\mathbb{E}[X] = \mu \qquad \mathrm{Var}(X) = \mathbb{E}\big[(X-\mu)^2\big] = \mathbb{E}[X^2] - \mu^2$$

In general,

$$\mathbb{E}[f(X)] \neq f(\mathbb{E}[X]) = f(\mu)$$

Linear case

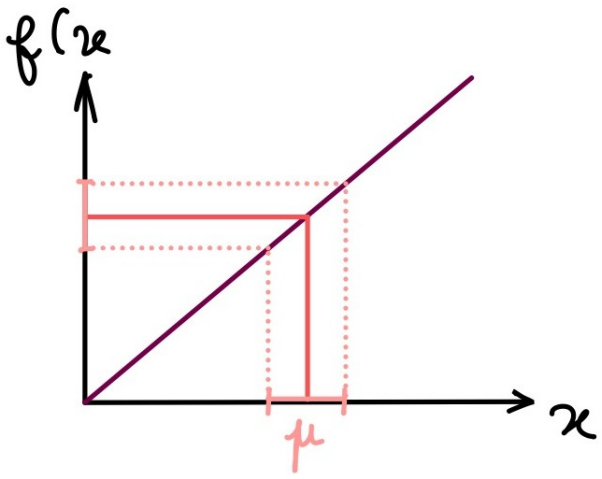
If f is linear, $f(x) = a + bx$, then:

$$\mathbb{E}[f(X)] = a + b\,\mathbb{E}[X] = a + b\mu$$

Reminder: computing expectations

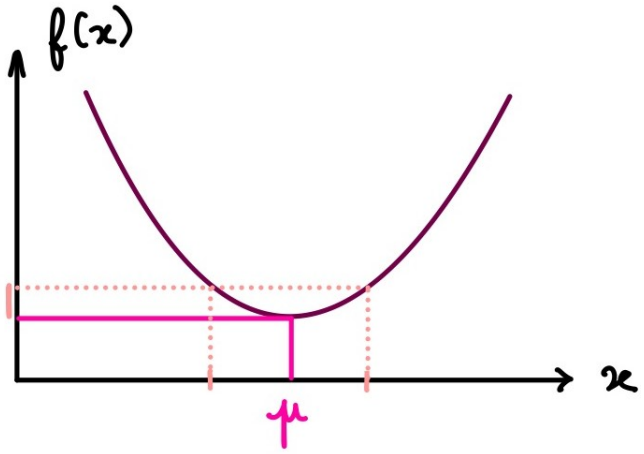
$$\mathbb{E}[X] = \sum_i x_i\, p_i \quad \text{(discrete)} \qquad \text{or} \qquad \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot p(x)\, dx \quad \text{(continuous)}$$

Illustration (linear mapping of uncertainty)



← Here, we see that any uncertainty on μ translates into more or less that same uncertainty in f in a linear case.

Here, we see that any uncertainty on μ is shifted up on f. The propagation of the uncertainty is not linear →



Example:

$$f = a + bX + cX^2$$

Rewrite around the mean μ (idea: separate mean and fluctuation):

$$f = a + b(X - \mu) + b\mu + c(X - \mu)^2 - c\mu^2 + 2c\mu X$$

Expectation value:

$$\begin{aligned}\mathbb{E}(f) &= a + b(E(X) - \mu) + b\mu + c(E(X) - \mu)^2 - c\mu^2 + 2c\mu\mathbb{E}(X) \\ &= a + b\mu + c\mu^2 + c \cdot \mathbb{V}(X) \\ &= f(\mu) + c \cdot \sigma_x^2\end{aligned}$$

For a Taylor expansion:

$$c = \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}|_\mu$$

13.2 Effects on variance

Basic rules

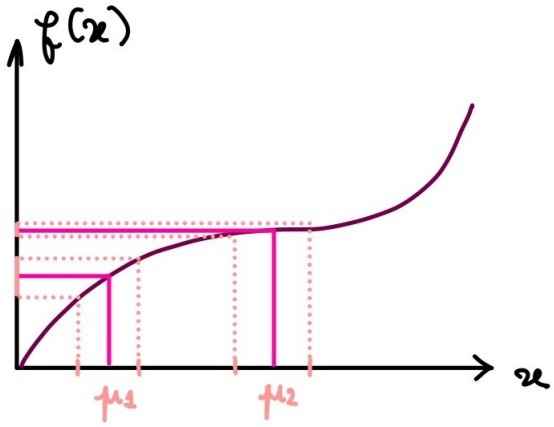
$$\mathbb{V}(a + bX) = b^2\,\mathbb{V}(X)$$

First-order propagation (linearization around μ):

$$\mathbb{V}\big(f(X)\big) \approx \left(\frac{\partial f}{\partial x}\bigg|_\mu\right)^2 \cdot \mathbb{V}(X) \qquad \text{for small variations}$$

Equivalently (using standard deviations):

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial x}\right)_\mu^2 \sigma_X^2 \qquad \sigma_f \approx \left|\frac{\partial f}{\partial x}\bigg|_\mu\right| \cdot \sigma_X$$



← Here, we see that, since the function is not linear, uncertainty on μ doesn't translate to something linear on f, which shows the non-linear propagation of uncertainty for non-linear functions.

Sum of two variables

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathrm{Cov}(X,Y)$$

If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$ and:

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

Example: Height of some people in centimetres, two distributions defined by X and Y

$$\mu_X = 175 \quad \mu_Y = 165 \quad \sigma_X = 10 \quad \sigma_Y = 9$$

Then:

$$\mu_{X+Y} = 340\text{cm} \qquad \sigma_{X+Y} = \sqrt{10^2 + 9^2} = \sqrt{181} \approx 13.5\text{cm}$$

Special Case: Product of Two Quantities

Let

$$f = X \cdot Y$$

Using logarithms:

$$\ln f = \ln(X \cdot Y) = \ln X + \ln Y$$

For small relative uncertainties, the relative uncertainty propagates as:

$$\frac{\sigma_f}{f} \approx \sqrt{\left(\frac{\sigma_X}{\mu_X}\right)^2 + \left(\frac{\sigma_Y}{\mu_Y}\right)^2}$$

The mean of f:

$$\mu_f \approx \mu_X \cdot \mu_Y$$