Statistics1

Representation of Data

- Qualitative variables (Frequency table) → Bar chart, Pie chart
- Quantitative variables (Frequency Distribution table) → Histogramme, Polygone, Cumulative frequency dist.

→ Frequency table

Classes	Frequency
Class 1	5
Class 2	32
Class 3	40
Total	77

Frequency of each class

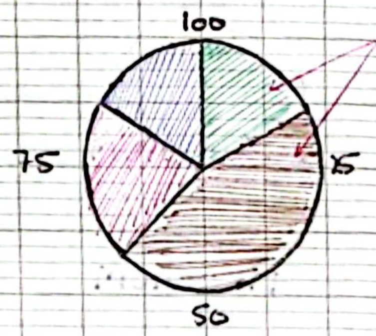
Total nb of frequency (Sum of all frenquency)

→ Bar chart

x'ox : classes
y'oy : frenquency

with a gap between classes

→ Pie Chart



Pourcentage of each class or relative frequency : $\frac{\text{Frequency}}{\text{total}}$

N.B !!

Sum of relative frequency equal 1, or 100.

→ Frequency Distribution

Step 1 : Number of classes : $2^K > n$

K: nb of classes

n: nb total of observation

Step 2 : Class interval : $i \geqslant \frac{(H - L)}{K}$

H : Hightest value

L : Lowest value

Step 3 : Set the individual class limits.

Classes
[200 to 600[
[600 to 1000[
[1000 to 1400[

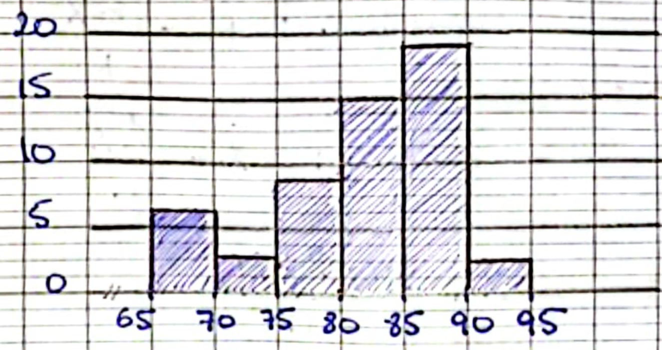
N.B !!

. Nb of classes should be correct.

. Midpoint = $\frac{200 + 600}{2}$

Step 4 : Count the nb of items in each class (Frequency)

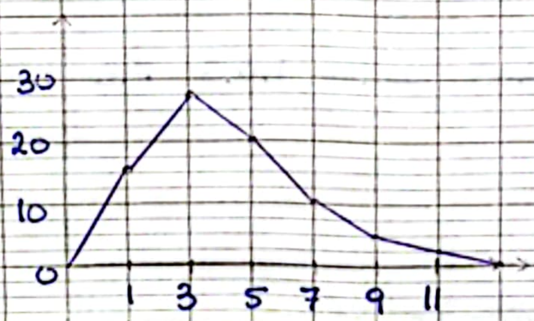
→ Histograms



x'ox: classes
y'oy: Frequency

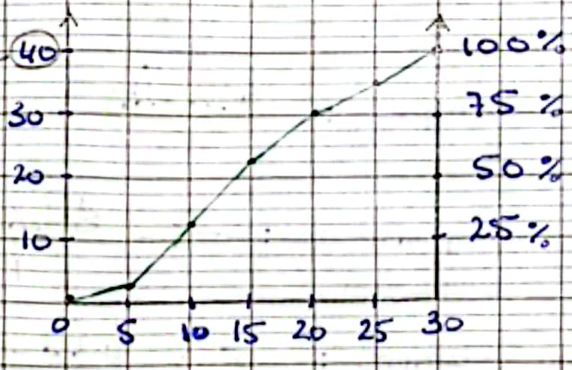
without gap between classes.

→ Polygone



x'ox: Midpoint of class
y'oy: Frequecy

→ Cumulative frequency distribution



x'x: classes
y'y1: Cumulative Frequency
y'y2: Pourcentage

N.B !!
Cumulative frequency:

	Frequency	Cumulative fr
	8	8
8 + 11 → 19	11	19
19 + 23 → 42	23	42
42 + 38 → 80	38	80

The last one is the total frequency

* Central tendency

→ Arithmetic mean :
(without Frequency)

Population : $\mu = \frac{\Sigma X}{N}$

Sample : $\bar{X} = \frac{\Sigma \bar{X}}{n}$

→ Weighted mean : $\bar{X}_w = \frac{\Sigma fX}{n}$

N.B !!
For interval classes : $\bar{X} = \frac{\Sigma fM}{n}$

→ Mode : The class who has the bigger frequency

* Dispersion

→ Rang : rang = Largest value - Smallest value

→ Mean Deviation :

Population : $\frac{\Sigma|X - \mu|}{N}$

Sample : $\frac{\Sigma|X - \bar{X}|}{n}$

→ Variance :

Population : $\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$

Sample : $s^2 = \frac{\Sigma (X - \bar{X})^2}{n-1}$

→ Standard deviation :

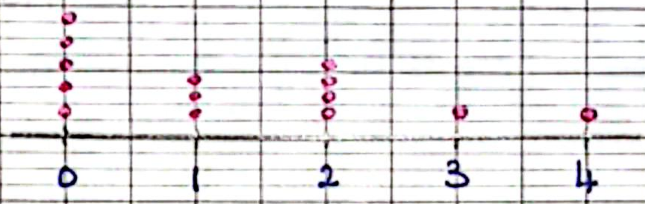
Population : $\sigma = \sqrt{\sigma^2}$

Sample : $s = \sqrt{s^2}$

N.B !! For interval class : $s^2 = \frac{\Sigma F (M - \bar{X})^2}{n-1}$

* Displaying and Exploring Data

→ Dot Plot



. every observation is represented by a (•)

. The total nb of observation is represented by the total nb of dots

→ Percentile

Location of a percentile : $L_p = \frac{P}{100}(n+1)$

→ Quartiles

. First quartile (25%) : $L_{25} = \frac{25}{100}(n+1)$

. Second quartile (50%) : $L_{50} = \frac{50}{100}(n+1)$
Median

. Thir quartile (75%) : $L_{75} = \frac{75}{100}(n+1)$

. Forth quartile (100%)

N.B

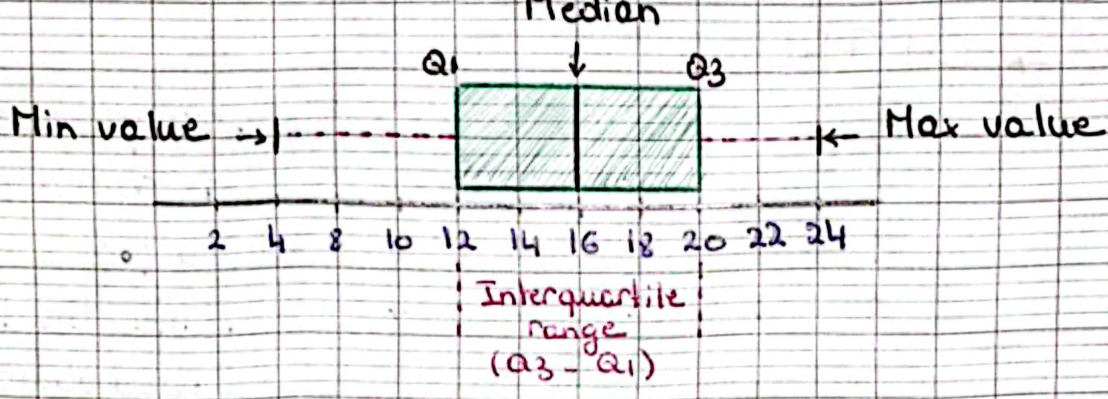
For exemple if Q_1 is between 43 and 61 and $L_{25} = 1{,}75$

=> Q_1 = Smallest val + 0,x (big val - Small val)

x: nb after (,) in the location

$Q_1 = 43 + 0{,}75\,(61 - 43)$

→ Box Plot



. outlier > $Q_3 + 1,5\,(Q_3 - Q_1)$
. outlier < $Q_1 - 1,5\,(Q_3 - Q_1)$

* Probability

→ Classical Probability

. Probability of an event $= \dfrac{\text{Nb of favorables outcomes}}{\text{Total nb of possible outcomes}}$

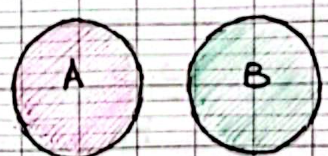
→ Empirical Probability

. Probability of Successful flight $= \dfrac{\text{Nb of successful flight}}{\text{Total nb of flights.}}$

→ Addition Rules

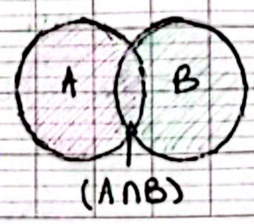
. if A and B are mutually exclusive

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$



. if A and B are not mutually exclusive

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$



N.B!!

Contingency Table: table used to classify sample observation according to two or more characteristics

→ Multiplication Rules

. if A and B are independents

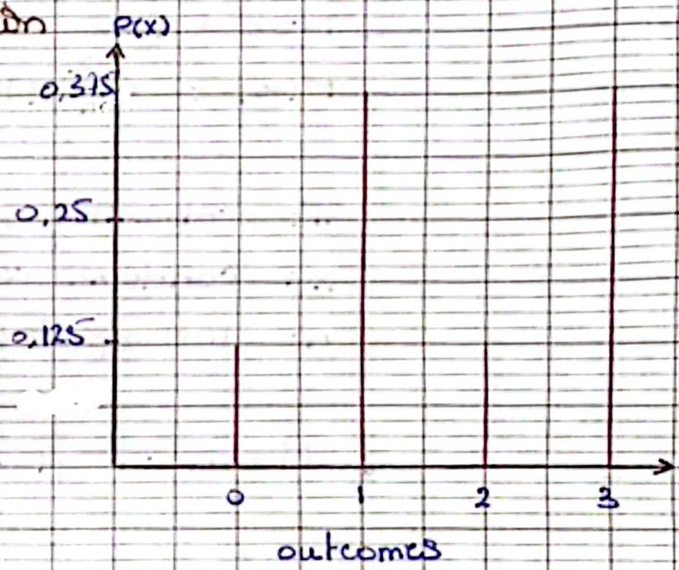
$p(A \text{ and } B) = p(A \cap B) = p(A)\,p(B)$.

. if A and B are not independents

$p(A \text{ and } B) = p(A \cap B) = p(A)\,p(B/A)$

→ Probability Distribution

Outcomes (X)	Probability of outcomes P(x)
0	0,125
1	0,375
2	0,125
3	0,375
Total	1



→ Mean of a probability Distribution

$$\mu = \sum [XP(x)]$$

→ Variance and standard deviation of a probability distribution

$$\sigma^2 = \sum [(X-\mu)^2 P(x)] \quad ; \quad \sqrt{\sigma^2}$$

Discrete Probability Distribution

Binomial Probability Distribution

1. There are only 2 possible outcomes
2. The outcomes are mutually exclusives
3. The randome variable is the result of counts.
4. Each trial is independent of any other trial

- Probability

$$P(x) = C_n^x (\pi)^x (1-\pi)^{n-x}$$

π: probability of success (given)

x: randome variable

n: nb of trial (given)

- Mean : $\mu = n\pi$

- Variance: $\sigma^2 = n\pi(1-\pi)$

Poisson Probability Distribution

Describe the nb of times some event occurs during a specified interval (time, distance, area)

- Probability

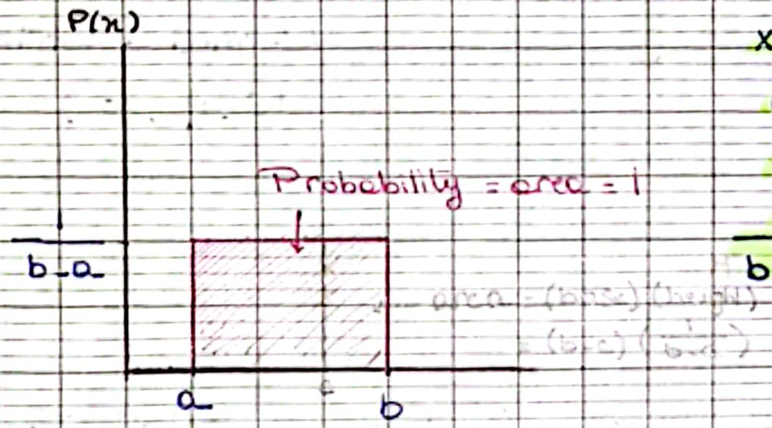
$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

- Mean : $\mu = n\pi$

- Variance : $\sigma^2 = n\pi$

* Continous Probability Distribution

→ The Uniform Distribution



x: outcomes from a to b

a: lower limit of the interv

b: upper limit of the interval

$\frac{1}{b-a}$: Probability of any particular outcome from a to b

Some of all probabilities from a to b equal 1 (area of the rectangle)

$P(x>b)$ or $P(x<a) = 0$

↳ Mean , $\mu = \frac{a+b}{2}$

↳ Standard deviation : $\sigma = \sqrt{\frac{(b-a)^2}{12}}$

N.B !!

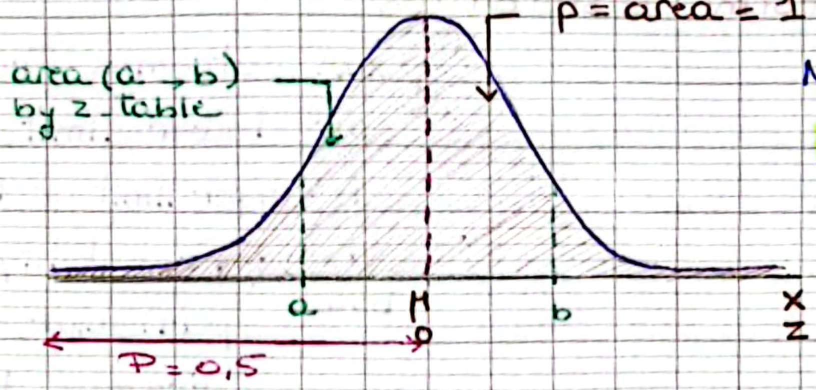
. The prob of $X = C$ => $\frac{1}{b-a}$

. The prob of x between c and b => area of the small rectangle

area = (base) (height)

$(b-c)\left(\frac{1}{b-a}\right)$

→ The Normal Distribution



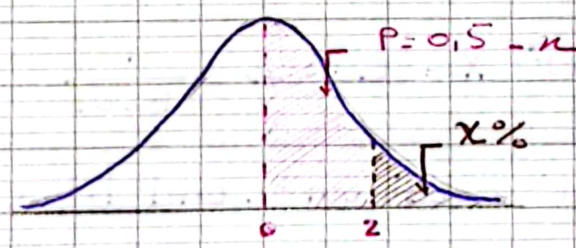
Normal Distribution:
$N(\mu, \sigma)$ given.

. Convert normal distribution into z_distribution ($N(0,1)$)
by finding z_value:

$$z = \frac{X - \mu}{\sigma}$$

. Finding the Probability:
- Find z_value
- Find P using z table.

. Finding X using Z
what is the value of X for a x %



. Find the value of Z using P from the table

. $\Rightarrow z = \frac{X - \mu}{\sigma} \Rightarrow X = z\sigma + \mu$

→ The Empirical Rule

- About 68% of X are between $(\mu - 1\sigma)$ and $(\mu + 1\sigma)$
- About 95% of X are between $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$
- About 99,7% (partially all) of X are between $(\mu - 3\sigma)$ and $(\mu + 3\sigma)$

* Sampling Error

Sampling Error is the difference between a sample statistic and population parameter:

$\bar{X} - \mu$

$S - \sigma$

$S^2 - \sigma^2$

$p - \pi$

* Central limit Theoreme

- given: population mean (μ)
  population standard deviation (σ)
  nb of observation in the sample (n)

- $\mu_{\bar{X}} = \mu_{pop}$

- $\sigma^2_{\bar{X}} = \dfrac{\sigma^2_{pop}}{n} \Rightarrow \sigma_{\bar{X}} = \dfrac{\sigma_{pop}}{\sqrt{n}}$

- $\sigma > \sigma_{\bar{X}}$

- To find the probability a sample mean falls:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

* Confidence interval Estimates

confidence interval is a range of values within wich the population parameter is expected to occure

↳ Confidence interval for population mean with σ Known (z_dist)

given: Sample mean ($\bar{X}$)
population standard deviation (σ)
nb of observation in the sample (n)
confidence level

Formula : $\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$

z : from the table for a particular level of confidance

↳ Confidence interval for population mean with σ unknown (t_dist)

given: Sample mean($\bar{X}$)
Sample standard deviation (s)
nb of obs in the sample (n)
confidence level

Formula: $\bar{X} \pm t \frac{\sigma}{\sqrt{n}}$

t : from the table for a particular level of confidence and sample size ($n-1$)

* Hypothesis Testing

1. State Ho and H1

- Ho : . = , ⩽ , or ⩾
  - . no more than, at least , has not changed

- H_1 : . Opposite of H_0
  - . ≠ , > , <
  - . Larger (more) than, smaller/less , has increased , different from.

2. Level of significance (α)

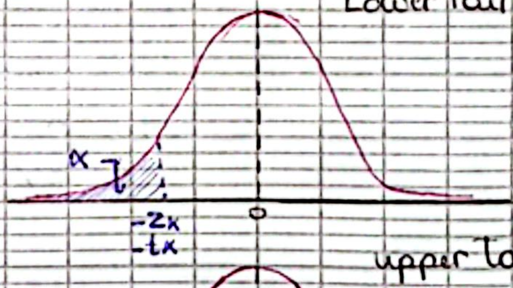
given: 0.01, 0.05, 0.10

3. Test statistic

Find the critical value:

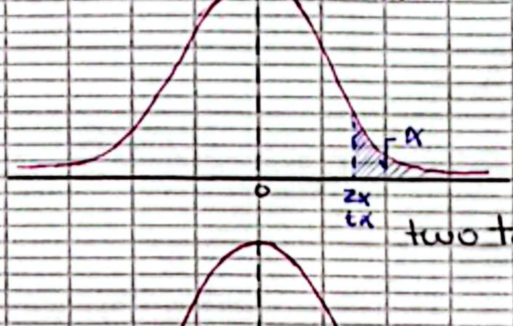
Lower tail

a. $H_0 : \mu \geq n$
$H_1 : \mu < n$



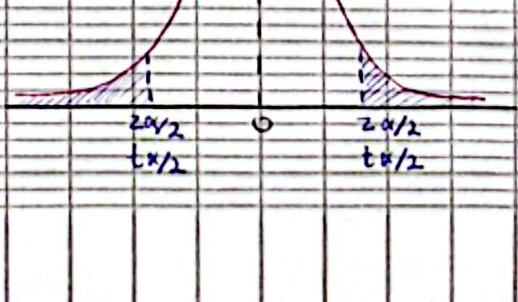
upper Tail

b. $H_0 : \mu \leq n$
$H_1 : \mu > n$



two tail

c. $H_0 : \mu = n$
$H_1 : \mu \neq n$



4. Decision Rule

from the sample

a. $H_0 : \mu \geq n$ => Reject Ho if $z < -z\alpha$

$H_1 : \mu < n$ $\quad t < -t_{\alpha, n-1}$

b. $H_0 : \mu \leq n$ => Reject Ho if $z > z\alpha$

$H_1 : \mu > n$ $\quad t > t_{\alpha, n-1}$

c. $H_0 : \mu = n$ => Reject Ho if $|z| > z_{\alpha/2}$

$H_1 : \mu \neq n$ $\quad |t| > t_{\alpha/2, n-1}$

5. Take a sample and make the decision

1. If σ is known (z)

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

2. If σ is unknown (t)

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$