Introduction to data analysis

Part 6. Introduction to statistics (part 1/2)

Clean the data

Explore the data

Store the data

Analyze in depth

A lot is expected from a data scientist

Find the data

Visualize the results

Understand the question

Tell the story

Let's get started.

the master labs • academy

NCOI LEARNING

Introduction to statistics

Explore the data

Analyze in depth

This afternoon you will learn:

- Basic concepts in statistics
- Intro to probability
- Overview of test types
- Intro to hypothesis testing
- Interpreting your results

Statistics in a nutshell

Data

Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized.

Information

When data is processed, organized, structured or presented in a given context so as to make it useful.

Some examples

Visitors to our website	Shoppers in the main street	Client segmentation
Pattern of site traffic from specific country	Segment breakdown of visitors for a particular store	Buying pattern for specific season

Putting it al into context

Your pension fund grew by 5% this year.
Sounds great, no?
Unless you compare it with the general stock market, which rocketed by 12%.

Basic concepts

What
is statistics about?

Process

generates

Population

What
is statistics about?

Characteristics

Variable

Characteristic possibly varies for each element of the population.

What
is statistics about?

Descriptive statistics

Representation of sample values

Inferential statistics

Analysis and interpretation of data

Execution of hypothesis tests

Extrapolation to the entire population

Representative sample = representatieve steekproef
Descriptive statistics = beschrijvende statistiek
Inferential statistics = inferentiële statistiek

What
is statistics about?

Sample

Selection of the population

Using a sample, one can never make a statement about the population with 100% certainty.

Confidence level is expressed by means of a percentage / a chance

Sample = steekproef
Confidence level = betrouwbaarheid

Probability

Theory of probability

Basic principles

Basics

Number of opportunities to pick a 'heart ' from the card game:

Total number of cards = 52
Total number of hearts = 13

P(heart) = 13/52 = 1/4

Sum

Number of opportunities to pick a 'heart' OR a 'clover' from the card game:

Total number of cards = 52
Total number of hearts = 13
Total number of clovers = 13

P(heart) = 13/52 = 1/4
P(clover) = 13/52 = 1/4

P(heart OR clover) = 13/52 + 13/52
= 26/52
= 1/2

Multiplication

Number of opportunities to pick a 'heart' AND NEXT a 'clover' from the card game assuming no cards are removed from the card deck:

Total number of cards = 52
Total number of hearts = 13
Total number of clovers= 13

P(heart) = 13/52 = 1/4
P(clover) = 13/52 = 1/4

P(heart AND clover) = 13/52 * 13/52
= ¼ * ¼
= 1/16

Theory of probability

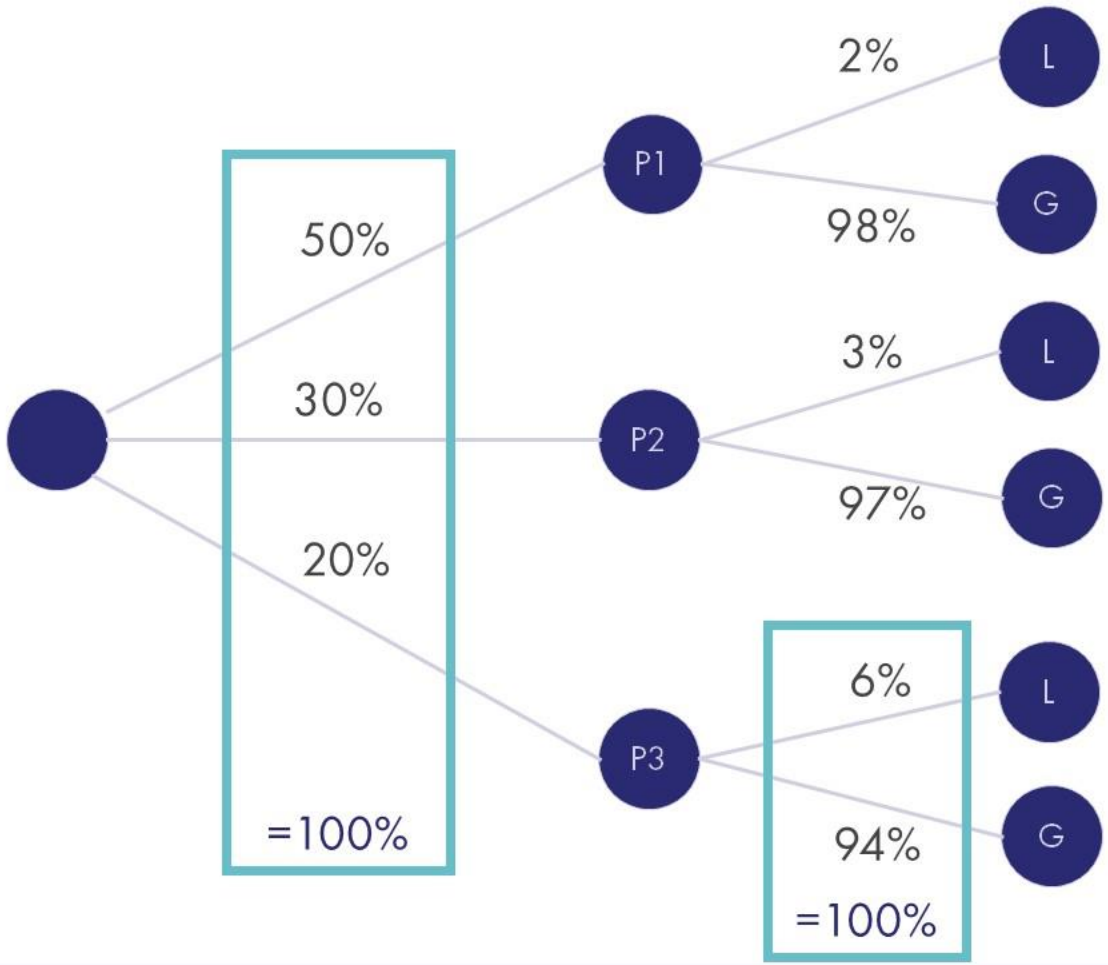
Exercise Coffee

To be able to supply all the demand for coffee a company needs 3 production lines.
Production line P1 produces half the amount of all the packages of coffee. On average 2% of these packages show leakages.
Production line P2 accounts for 30% of the production and supplies 3% of bad output.
Lastly, production line P3 supplies the rest, or 20% of the production. Production Line 3 is the worst line, resulting in 6% of leakages.

Calculate the propability that a randomly picked package of coffee will show leakage.

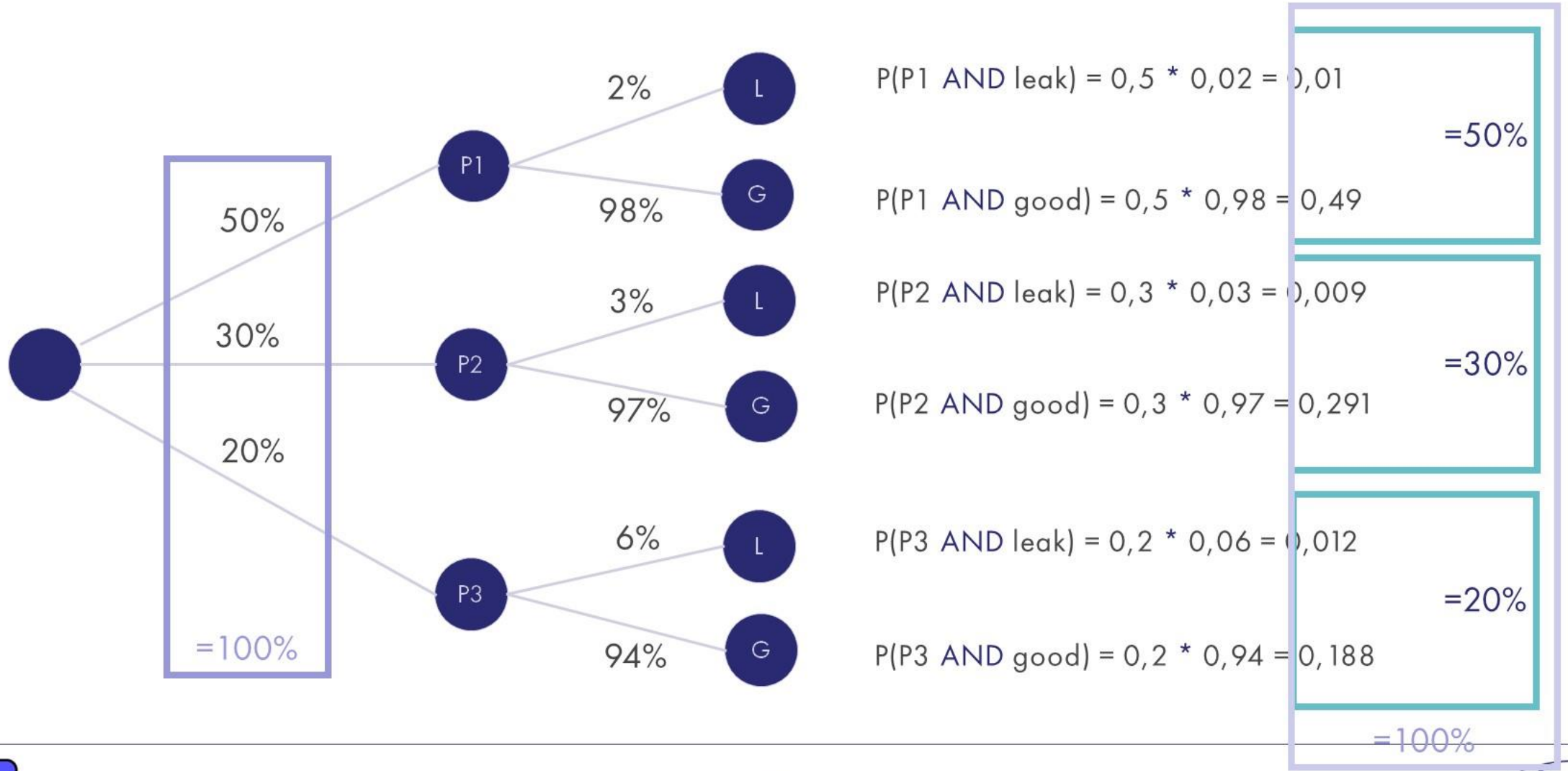
Theory of probability

Exercise Coffee



Theory of probability

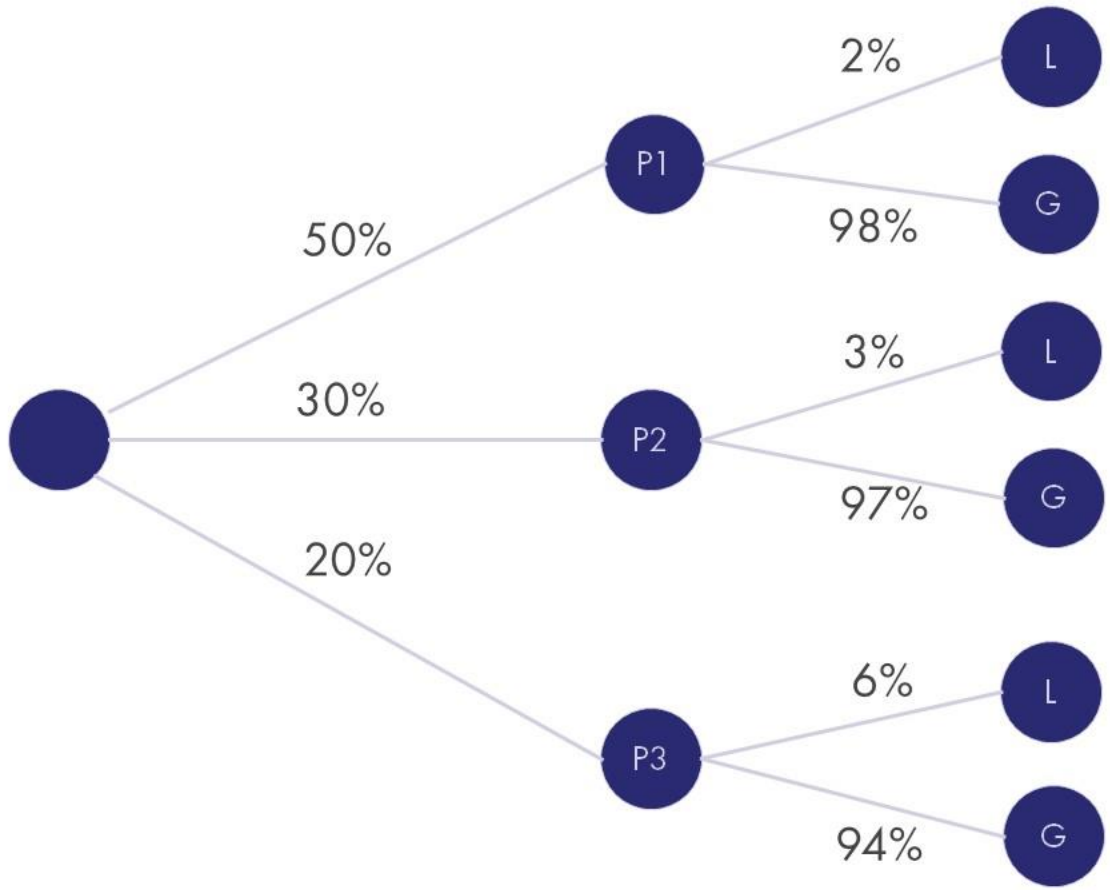
Exercise Coffee



Theory of probability

Exercise Coffee

The propability that a randomly picked package of coffee will show leakage?



P(P1 AND leak) = 0,5 * 0,02 = 0,01

OR +

P(P2 AND leak) = 0,3 * 0,03 = 0,009

OR +

P(P3 AND leak) = 0,2 * 0,06 = 0,012

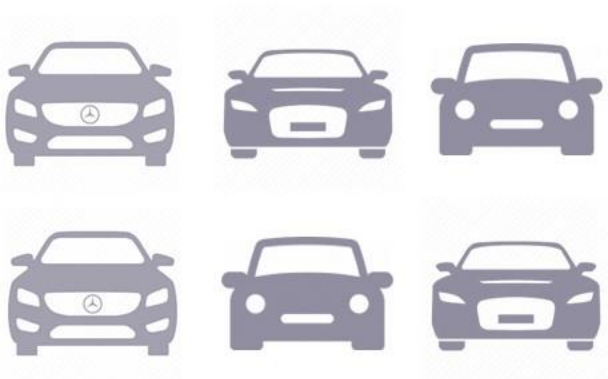
0,031

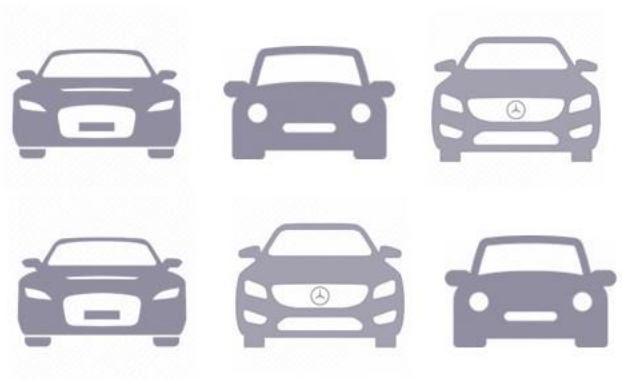
Theory of probability

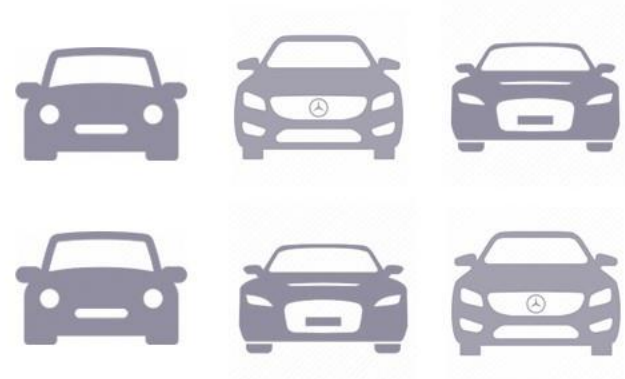
Factorals

Determine the number of ways in which you can arrange a group of objects

$3!$







Theory of probability

Combinations

Counting how many choices you have

$$\frac{n!}{x!\,(n-x)!} = \frac{3!}{1!\,2!} = \frac{3 * \cancel{2!}}{1! * \cancel{2}!} = 3$$

Theory of probability

Combinations

In how many ways can I assign 6 people to 4 chairs?

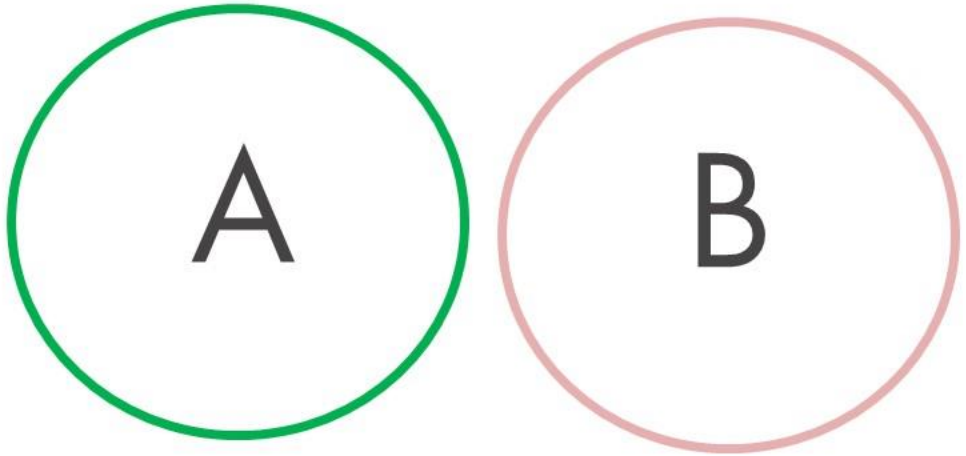
In how many ways can we choose 2 representatives in the group?

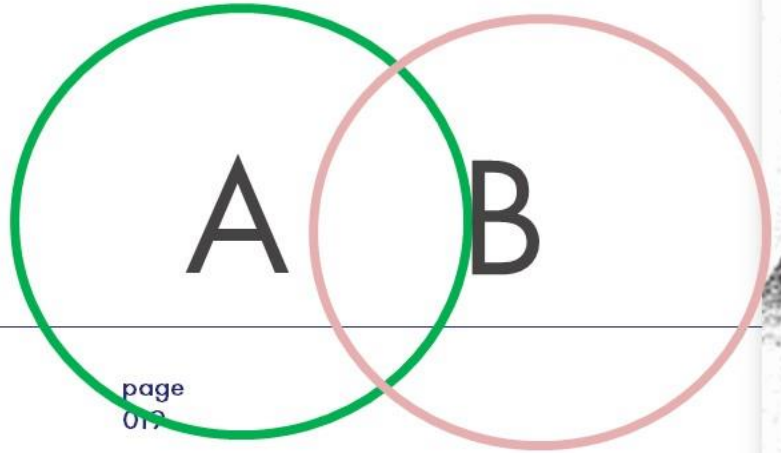
In how many ways can you rearrange the word STATISTICS?

Theory of probability

Bayes' rules

To investigate whether a person is suffering from a particular disease, a medical test can be applied. However, this test does not provide complete certainty. The test can predict with 99 procent certainty if someone has the disease. If someone does not have the disease, the chance is 3 percent that the test still gives a positive result. The disease occurs in 1 percent of the population. In a population survey, the test is performed on you and the test result is positive. How big is the chance that you have the disease?





Theory of probability

Bayes' rules

Let's start with dice. Given the following occurrences.

A = we get an even number
B = we get a number smaller or equal to 4

Are these events independent? How do we decide?

$P(A) =$

$P(B) =$

$P(A \cap B) =$

$P(A|B) =$

Theory of probability

Bayes' rules

Let's continue with coins.

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

Theory of probability

Bayes' rules

To investigate whether a person is suffering from a particular disease, a medical test can be applied. However, this test does not provide complete certainty. The test can predict with 99 procent certainty if someone has the disease. If someone does not have the disease, the chance is 3 percent that the test still gives a positive result. The disease occurs in 1 percent of the population. In a population survey, the test is performed on you and the test result is positive. How big is the chance that you have the disease?

$P(p|z) =$ $P(n|z) =$

$P(p|g) =$ $P(n|g) =$

$P(z) =$ $P(g) =$

Theory of probability

Bayes' rules

To investigate whether a person is suffering from a particular disease, a medical test can be applied. However, this test does not provide complete certainty. The test can predict with 99 procent certainty if someone has the disease. If someone does not have the disease, the chance is 3 percent that the test still gives a positive result. The disease occurs in 1 percent of the population. In a population survey, the test is performed on you and the test result is positive. How big is the chance that you have the disease?

$$P(z|p) = \frac{\text{chance of being sick given a positive results}}{\text{all possible situations where the result is positive}}$$

$$= \frac{P(z \cap p)}{P(p)}$$

Statistical measures

Descriptive statistics for each measurement scale

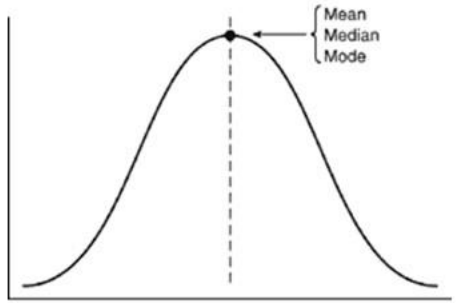
	Discrete		Continuous
	Nominal	Ordinal	Interval/ratio
Central tendency	Mode	Mode Median	Mode Median Arithmetic mean Geometric mean
Variability/dispersion		Range Quartiles Interquartile range	Range Quartiles Interquartile range Mean absolute deviation Variance Standard deviation Coefficient of variance
Central tendency, dispersion and skewness		Boxplot	Boxplot
Correlation			Correlation coefficient Covariance

Descriptive statistics of central tendency

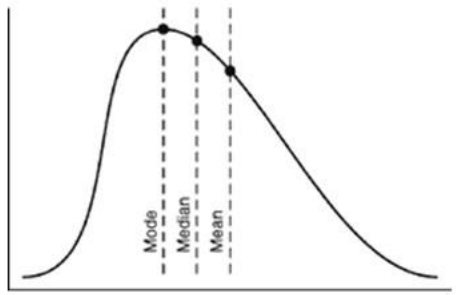
Describe the 'center' of the sample

The 'center' is where we expect to find most data points

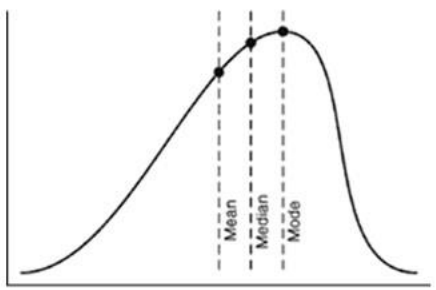
Mode
Median
Arithmetic mean
Geometric mean



Normal Distribution



Poitively Skewed Distribution



Negatively Skewed Distribution

Descriptive statistics of central tendency – Arithmatic Mean

Interval

Mean is the average value of the sample data set

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

Some characteristics

The sum of all values equals the arithmatic mean multiplied by the sample size:

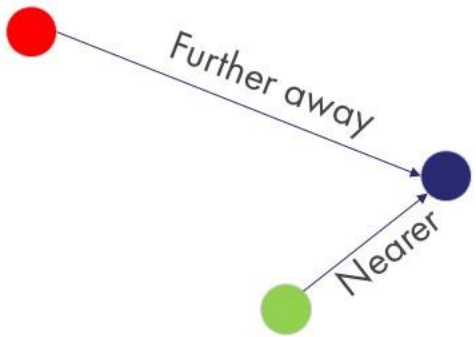
$$\sum_{i=1}^{n} x_i = \text{n} * \bar{x}$$

The sum of all deviations from the mean equals zero:

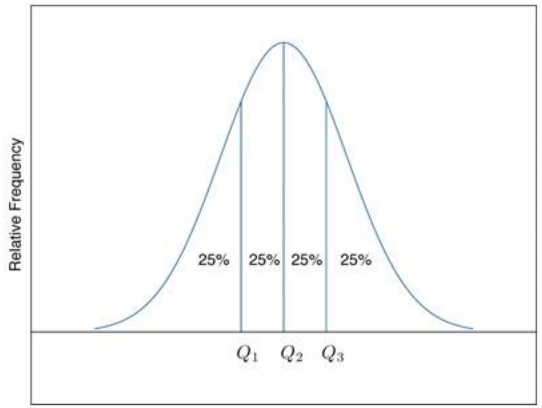
$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

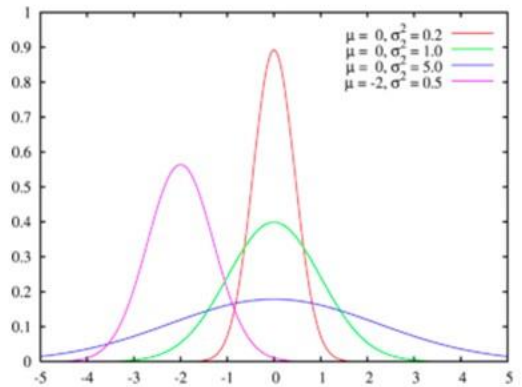
Descriptive statistics of variability or dispersion

Describe the relative position of an observation in comparison to other observations



Range
Quartiles
Interquartile range
Mean absolute deviation
Variance
Standard deviation
Coefficient of variance

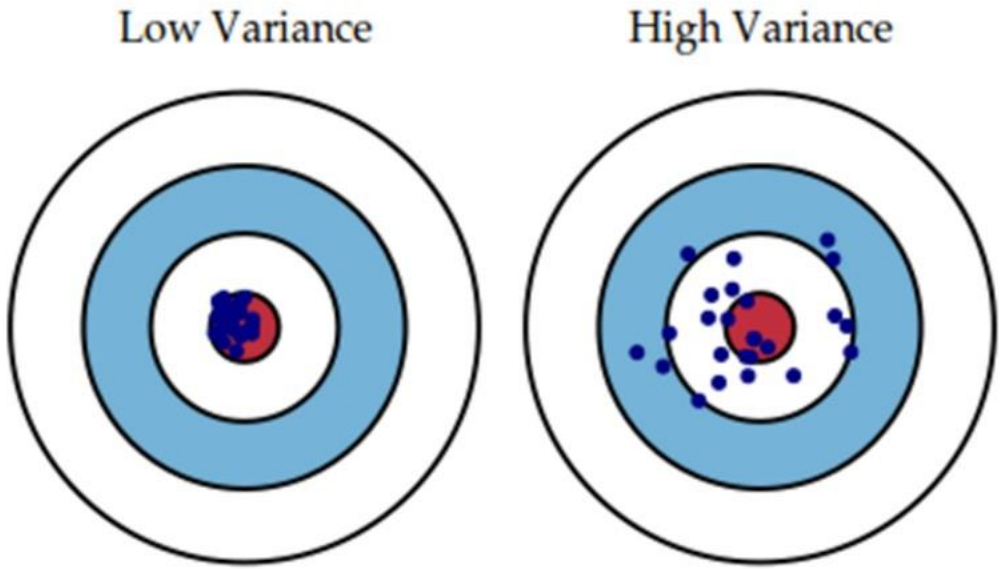




Descriptive statistics of dispersion – Variance

Interval

Variance indicates how much observations differ from the mean and hereby attaches greater importance to observations that are relatively further away



Descriptive statistics of dispersion – Variance

Interval

Variance indicates how much observations differ from the mean and hereby attaches greater importance to observations that are relatively further away

Variance is the mean of the squared deviations from the mean

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

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

Some characteristics

Variance cannot be negative

Variance can only be zero if each observation equals the mean

Descriptive statistics of dispersion – Variance

Interval

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

Degrees of freedom

Given three variables x_1, x_2 and x_3.

$x_1 = 10$

$x_2 = 15$

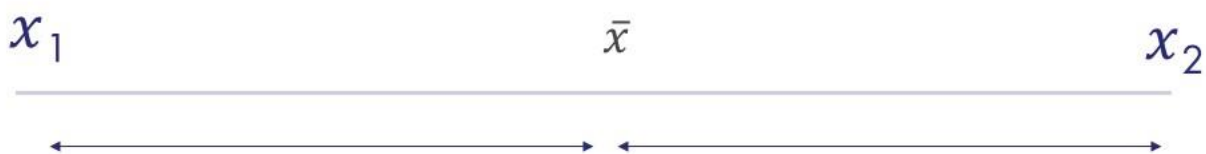
If we want to get an average of 15, what is the value of x_3?

= no freedom to decide on x_3 (n-1 variables) (= *Bessel's correction)*

Descriptive statistics of dispersion – Variance

Variance of a sample with n=2

Denominator = n-1



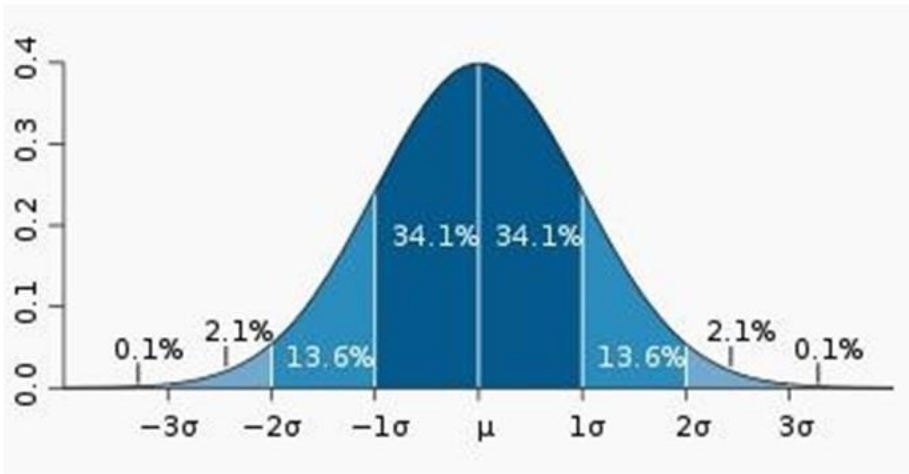
But what does it really mean?
E.g weight variance of 9805

Descriptive statistics of dispersion – Standard deviation

Interval

Standard deviation is the positive square of the variance

$$\sigma = \sqrt{s^2}$$



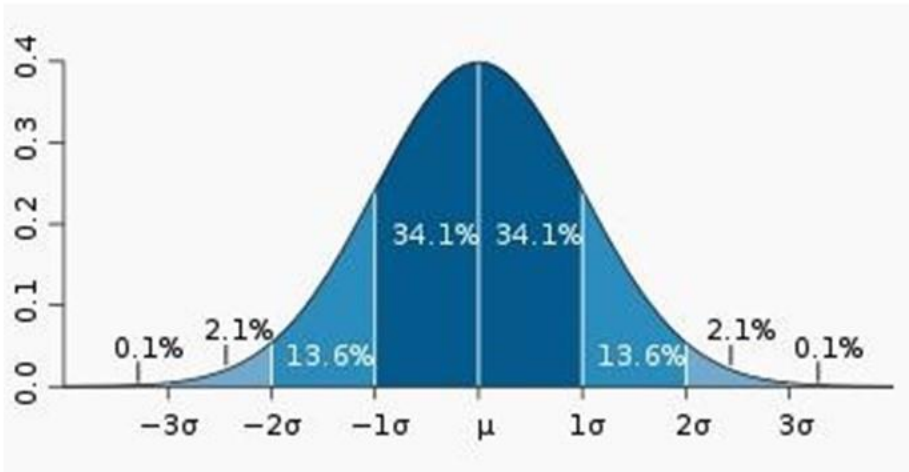
Some characteristics

Standard deviation is expressed in the same unit as the sample data

Descriptive statistics of dispersion – Standard deviation

Interval

Standard deviation tells how far away numbers on a list are from their average



68%

Most entries on the list will be somewhere around one standard deviation away from the average

95%

99,7%

Very few entries will be more than 2 or 3 standard deviations away

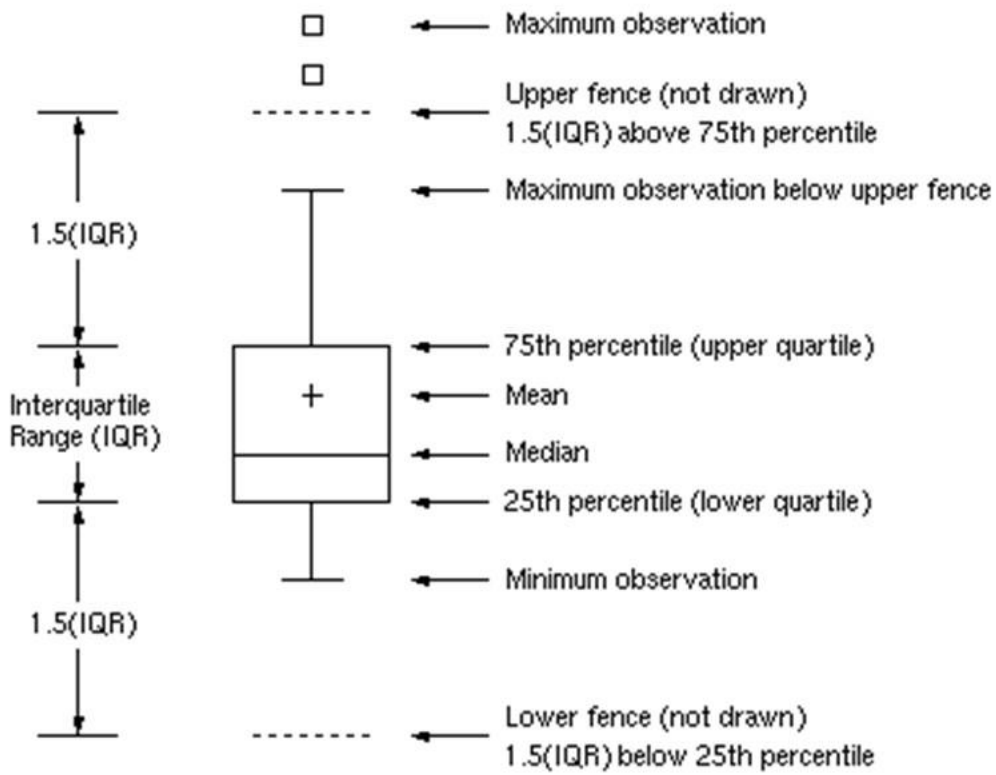
Exploratory Data Analysis

Questions answered by **Exploratory Data Analysis**

- Center
  - Mean
  - Median
  - Mode
- Spread
  - Variance
  - Standard deviation
  - Quartiles
  - Interquartile range
- Skewness
  - Symmetrical
  - Positively skewed
  - Negatively skewed
- Outliers

Questions answered by
Exploratory Data Analysis

Central tendency, dispersion and skewness are respresented in a boxplot

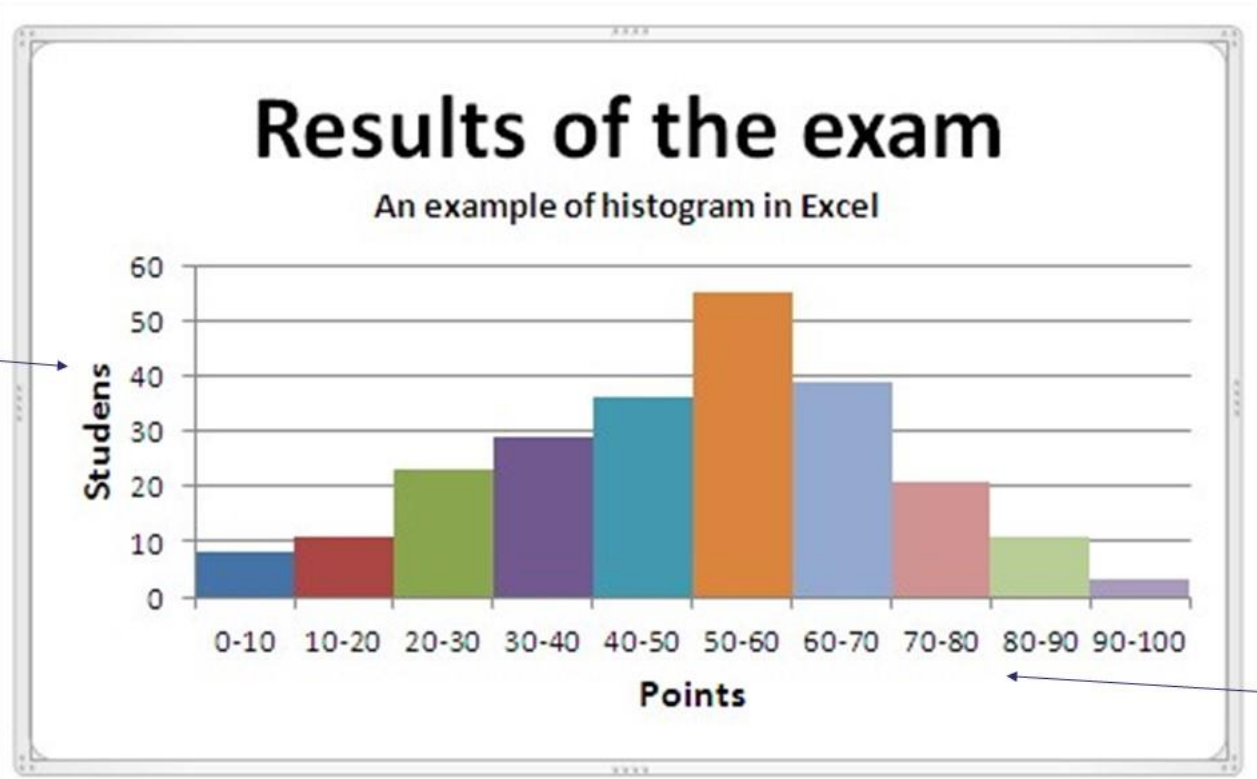


Questions answered by
Exploratory Data Analysis

Graphical representation of the distribution of numerical data

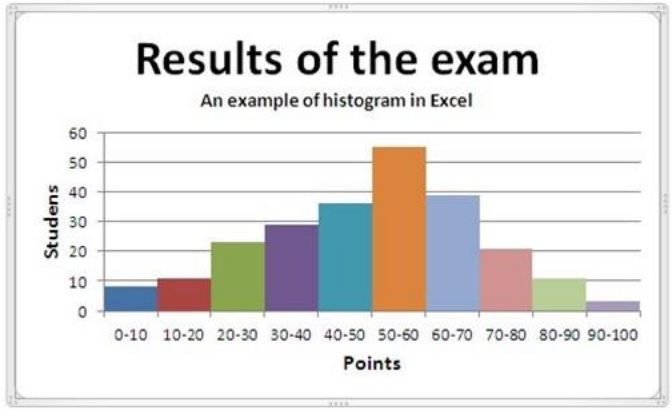
Summarizes a large amount of data

The amount of values that fall into an interval



Intervals

Questions answered by
Exploratory Data Analysis



- Used to represent continuous variables

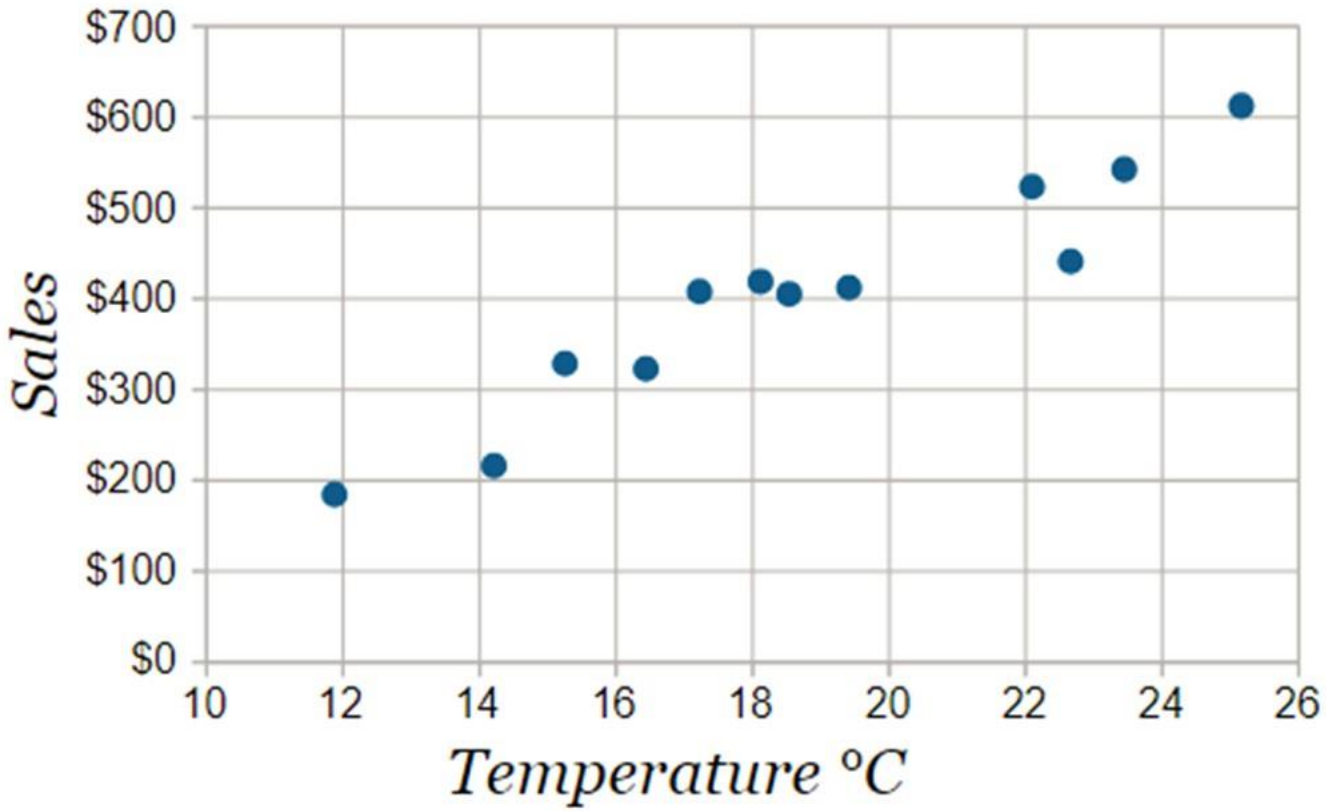
Notice that all the intervals touch each other, there is no space in between

- It is not required that each interval has the same width

However, equal width is easier to interpret

- If the y-axis represents the relative frequency, the total surface of all the rectangles equals 1

Questions answered by
Exploratory Data Analysis



Ice Cream Sales vs Temperature	
Temperature °C	**Ice Cream Sales**
14,2°	$215
16,4°	$325
11,9°	$185
15,2°	$332
18,5°	$406
22,1°	$522
19,4°	$412
25,1°	$614
23,4°	$544
18,1°	$421
22,6°	$445
17,2°	$408

Important distributions

Normal distribution

Gaussian function

Continuous probability distribution

Most important distribution

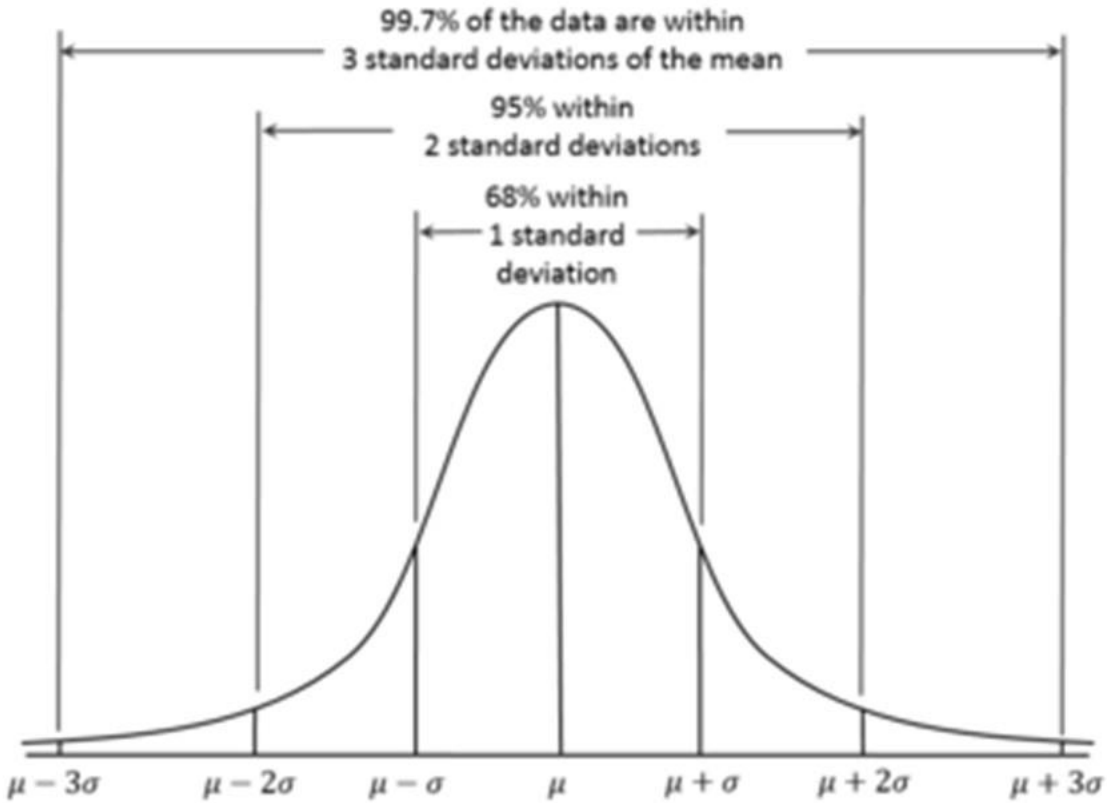
Most used distribution

A large set of observations tends towards a normal distribution
Mostly n ≥ 30

Normal distribution

Gaussian function

- Bell shaped
- Symmetrical with center = μ
- Mode = Median = Mean
- Inflection points for x = μ - **σ** and x = μ + **σ**



Normal distribution

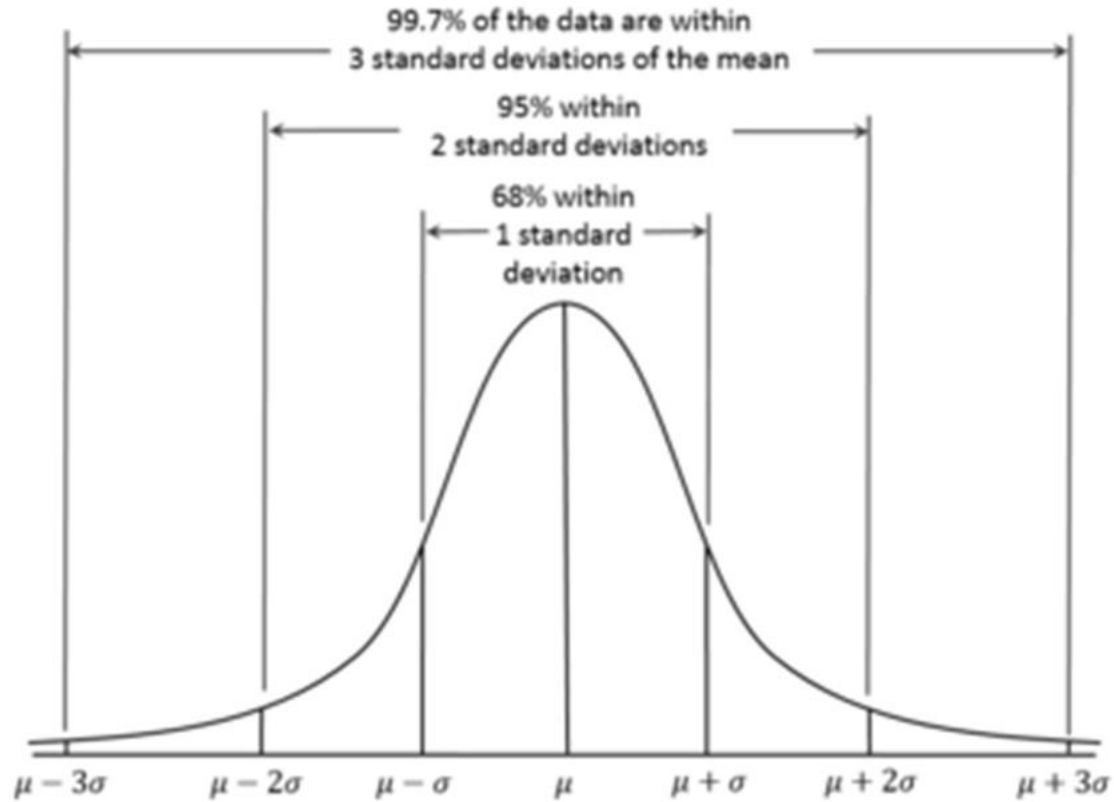
Properties

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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

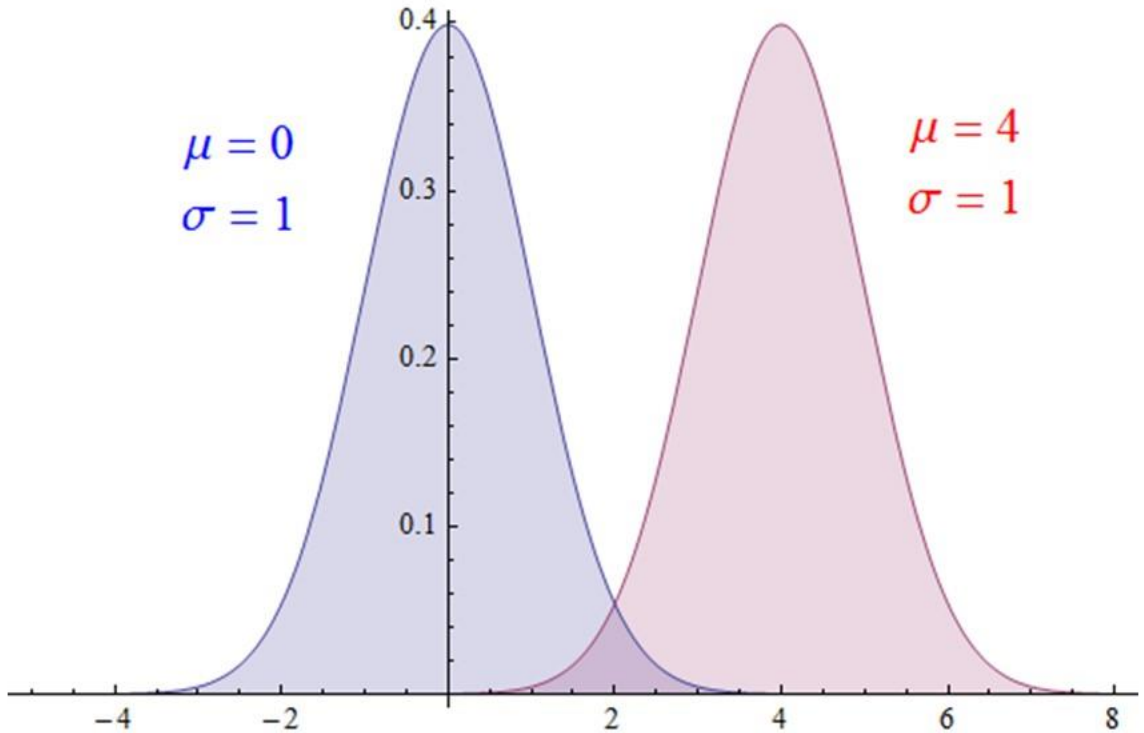
Mean

Variance



Normal distribution

Impact of μ



A change in μ

Changes the position of the curve

Does not change the shape

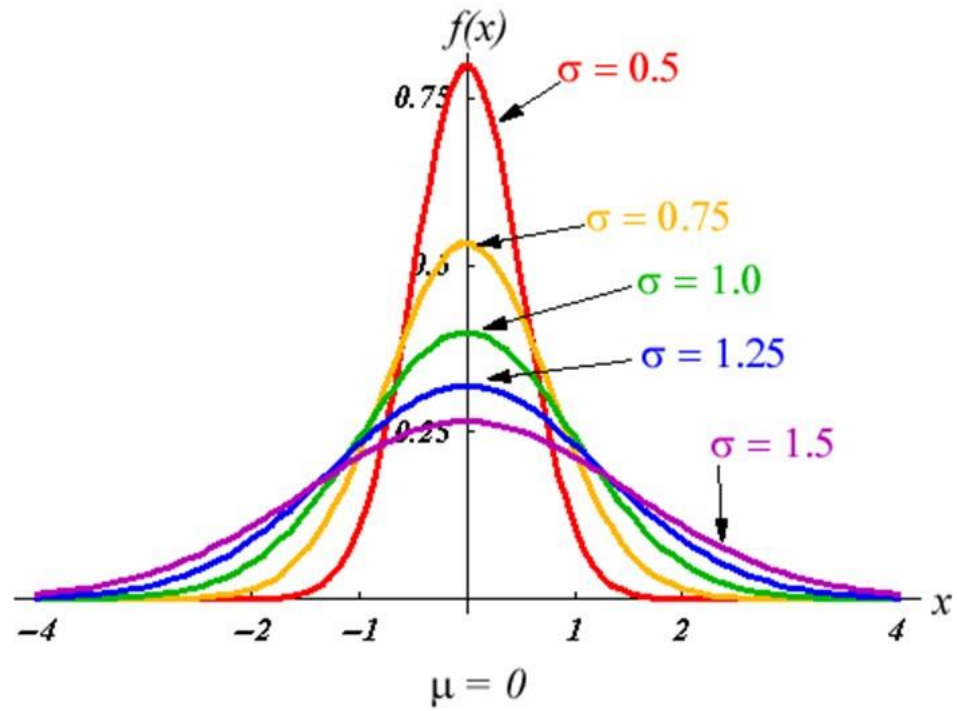
Normal distribution

Impact of σ

A change in **σ**

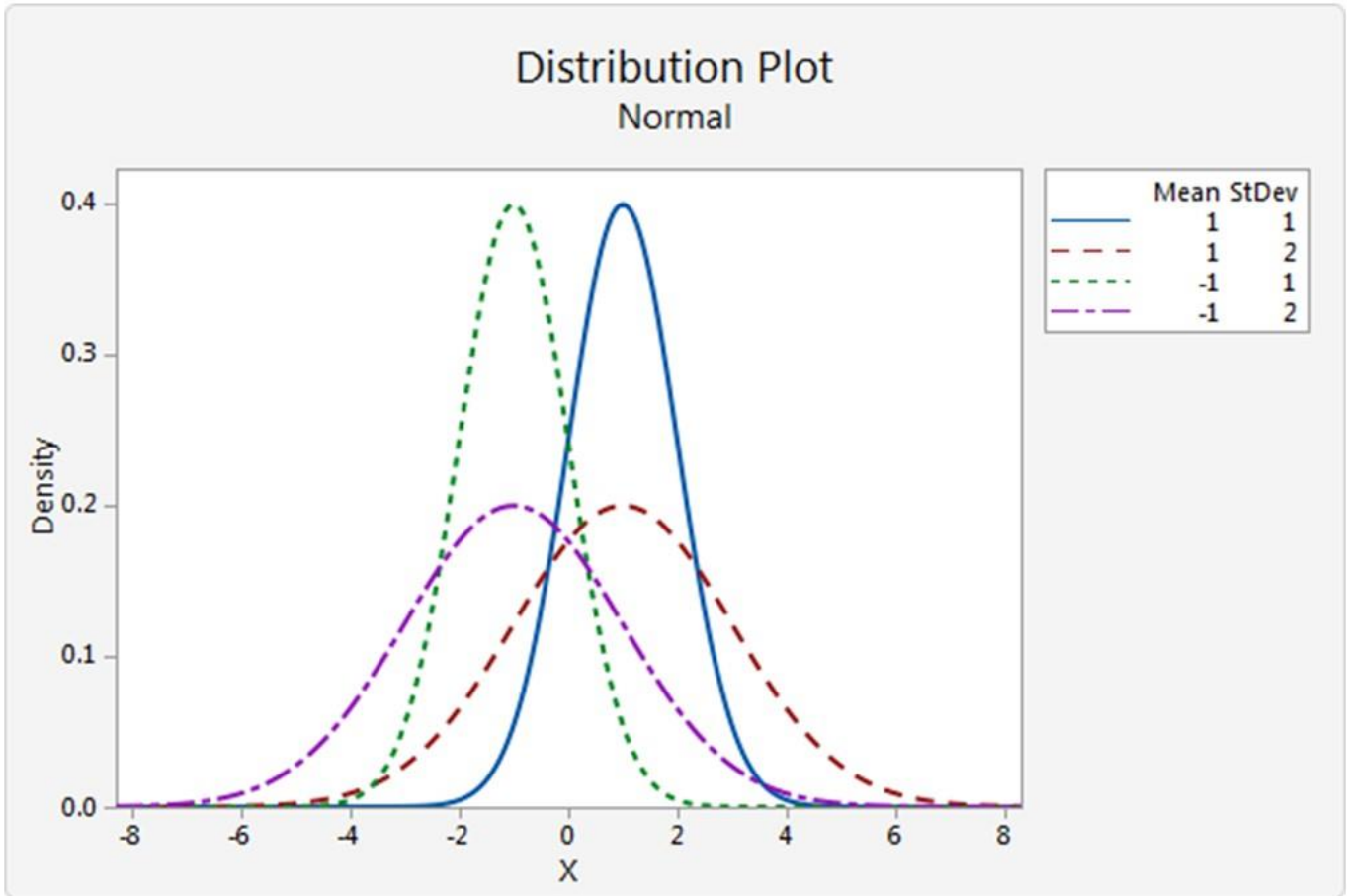
Changes the shape

Does not change the position



Normal distribution

Overview impact of μ and σ



μ

position

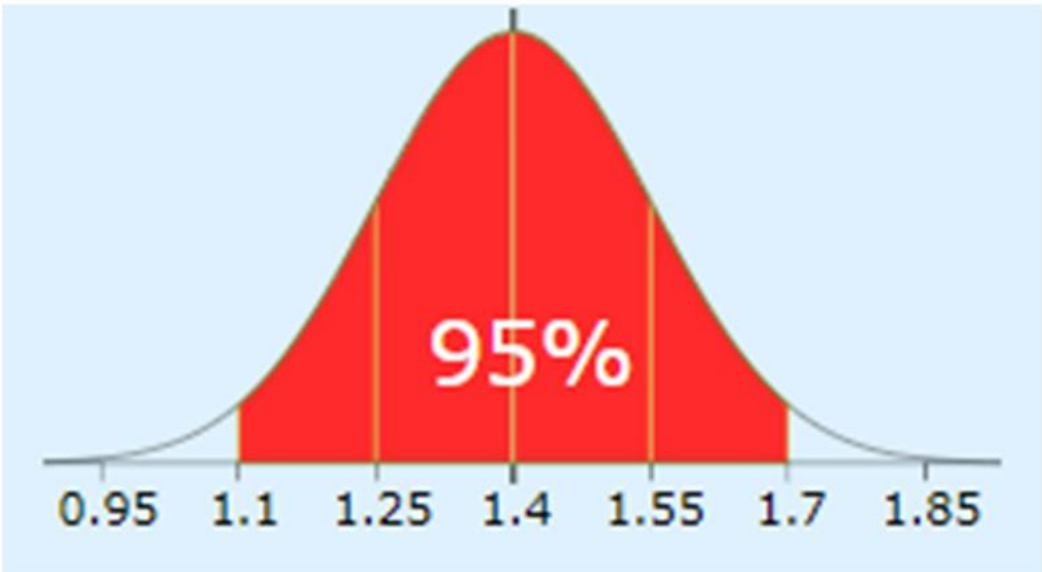
σ

shape

Normal distribution

Standard scores

The number of standard deviations from the mean = Z-score



Normal distribution

Standardize

Normal distribution

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

Mean | Variance

Standard normal distribution

$$Z \sim N(0, 1)$$

Mean | Variance

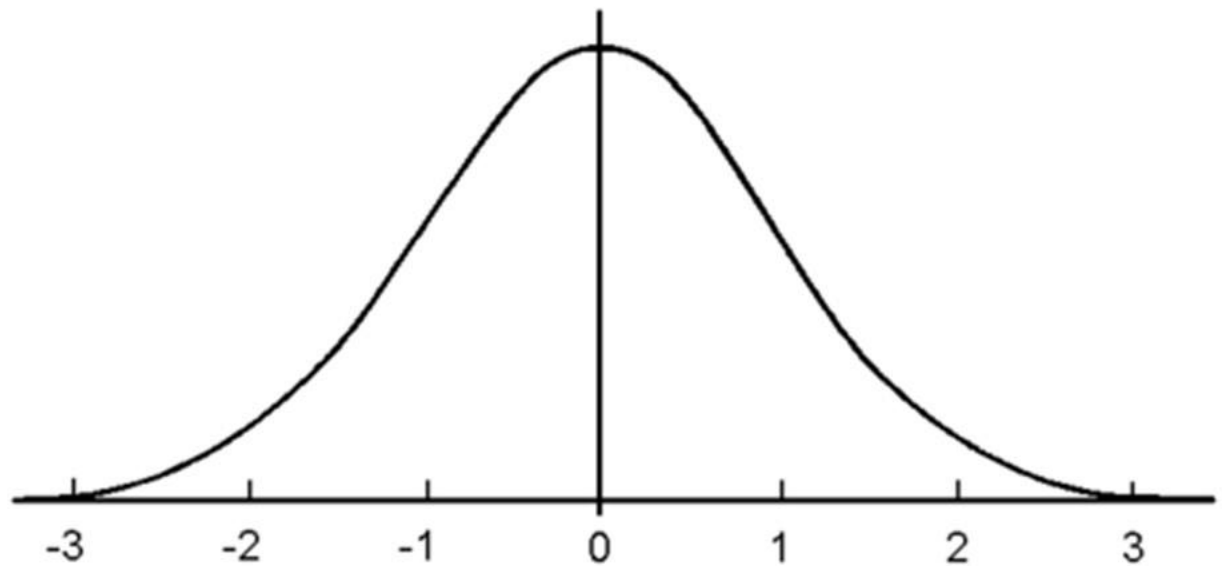
Normalization

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

Standard normal distribution

Deeper interpretation of the bell curve

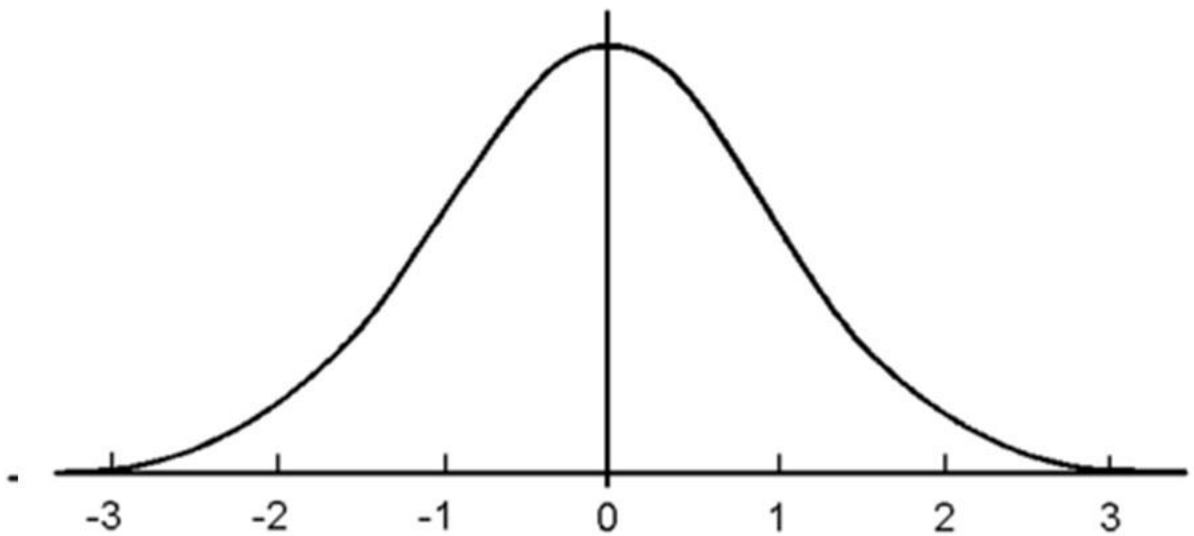
Surface below the curve
represents 100%



Standard normal distribution

Deeper interpretation of the bell curve

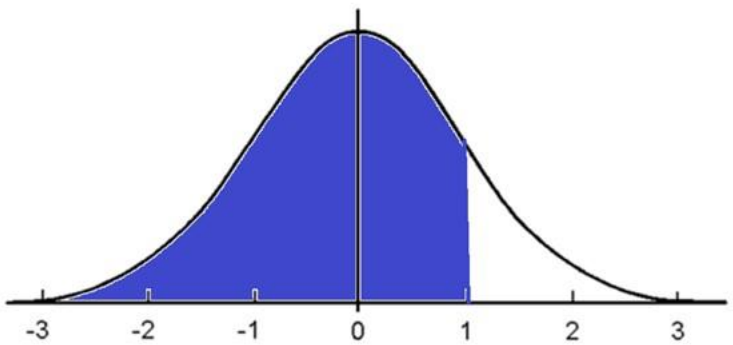
$P(Z \leq x)$ with $x \geq 0$



$P(Z \leq 0) = 0,5$ or 50%

Standard normal distribution

Deeper interpretation of the bell curve



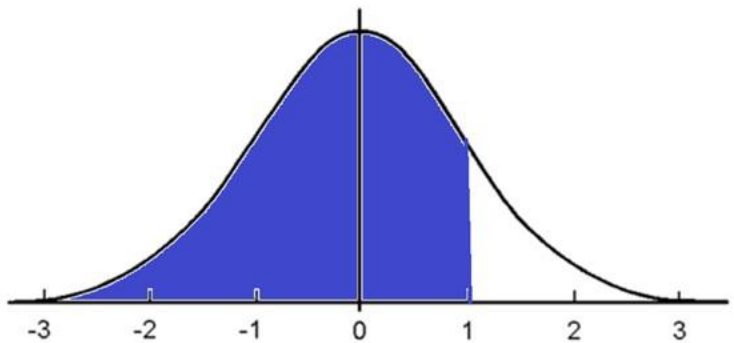
$P(Z \leq 1) = 0,8413$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Standard normal distribution

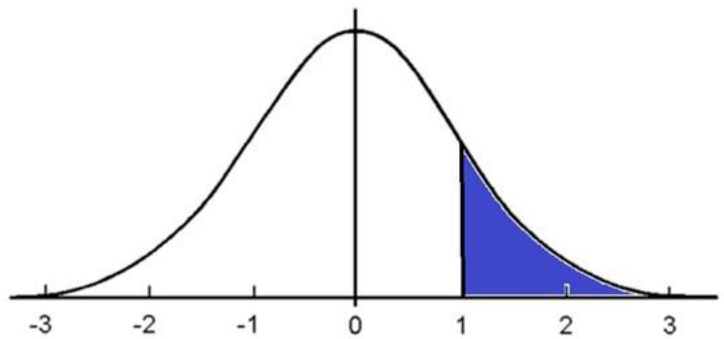
Deeper interpretation of the bell curve

$P(Z \leq 1)$



$P(Z \leq 1) = 0,8413$

$P(Z \geq 1)$



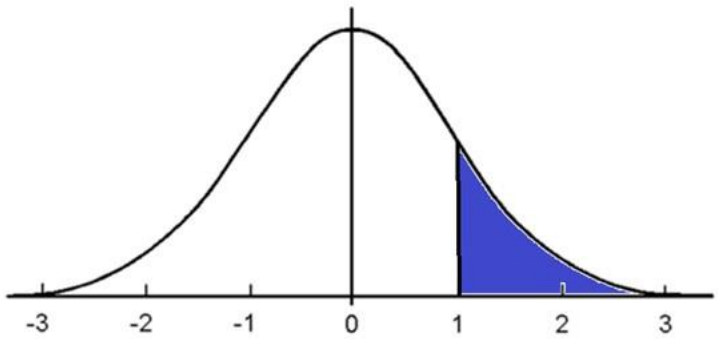
Surface below the curve represents 100%

$$P(Z \geq 1) = 1 - P(Z \leq 1)$$
$$= 1 - 0,8413$$
$$= 0,1587$$

Standard normal distribution

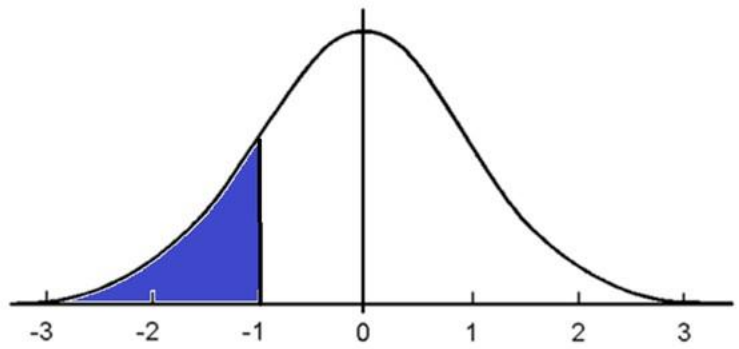
Deeper interpretation of the bell curve

$P(Z \geq 1)$



$P(Z \geq 1) = 0,1587$

$P(Z \leq -1)$



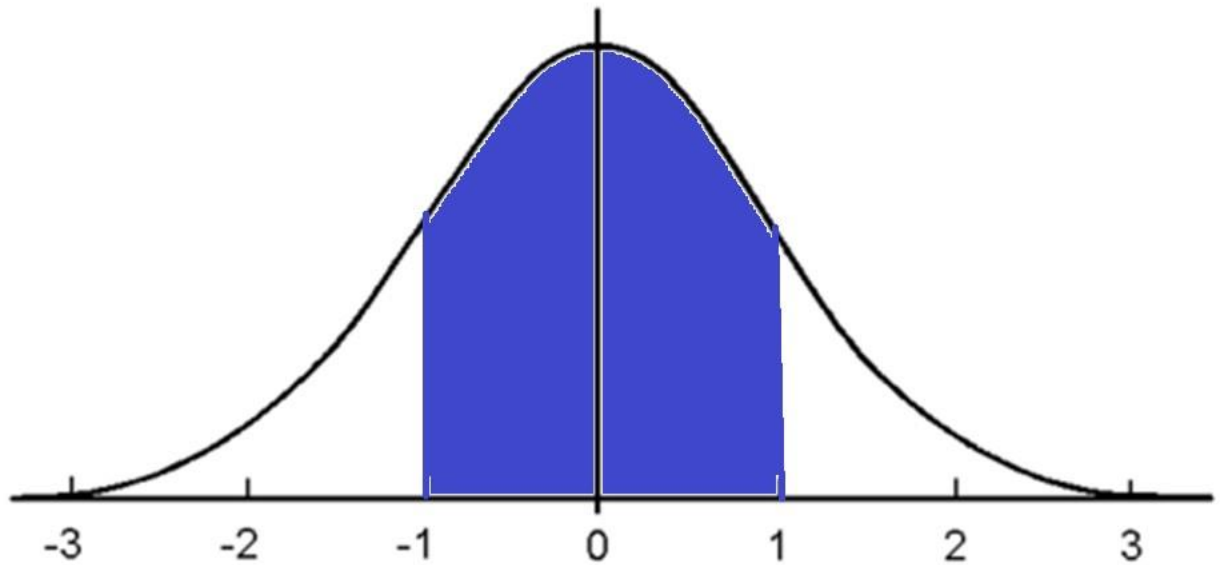
Use symmetry of the curve

$P(Z \leq -1) = 0,1587$

Standard normal distribution

Deeper interpretation of the bell curve

$$P(-1 \leq Z \leq 1)$$



$$P(Z \leq 1) - P(Z \leq -1)$$

$$= 0,8413 - 0,1587$$

$$= 0,6826$$

Student's t-distribution

William Gosset.

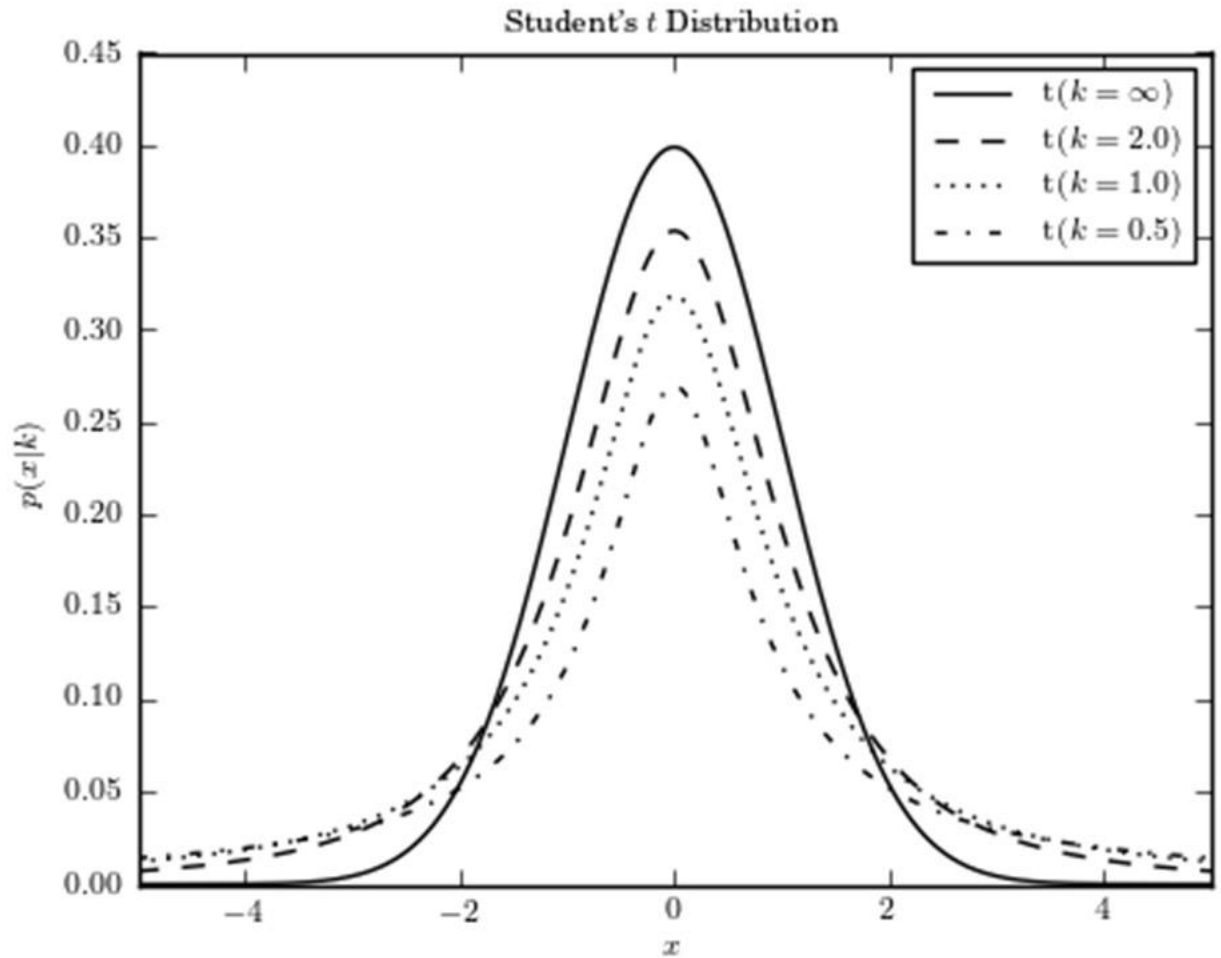
Continuous probability distribution

Important distribution

A small set of observations
Mostly n < 30

Population standard deviation is unknown

Student's t-distribution



μ= 0

Symmetric

Looks like the Standard Normal distribution
= Bell shaped

When to use:

σ is unknown

Student's t-distribution

Degrees of freedom

The freedom to vary

$$\sigma^2 = \frac{k}{k-1}$$

Student's t-distribution

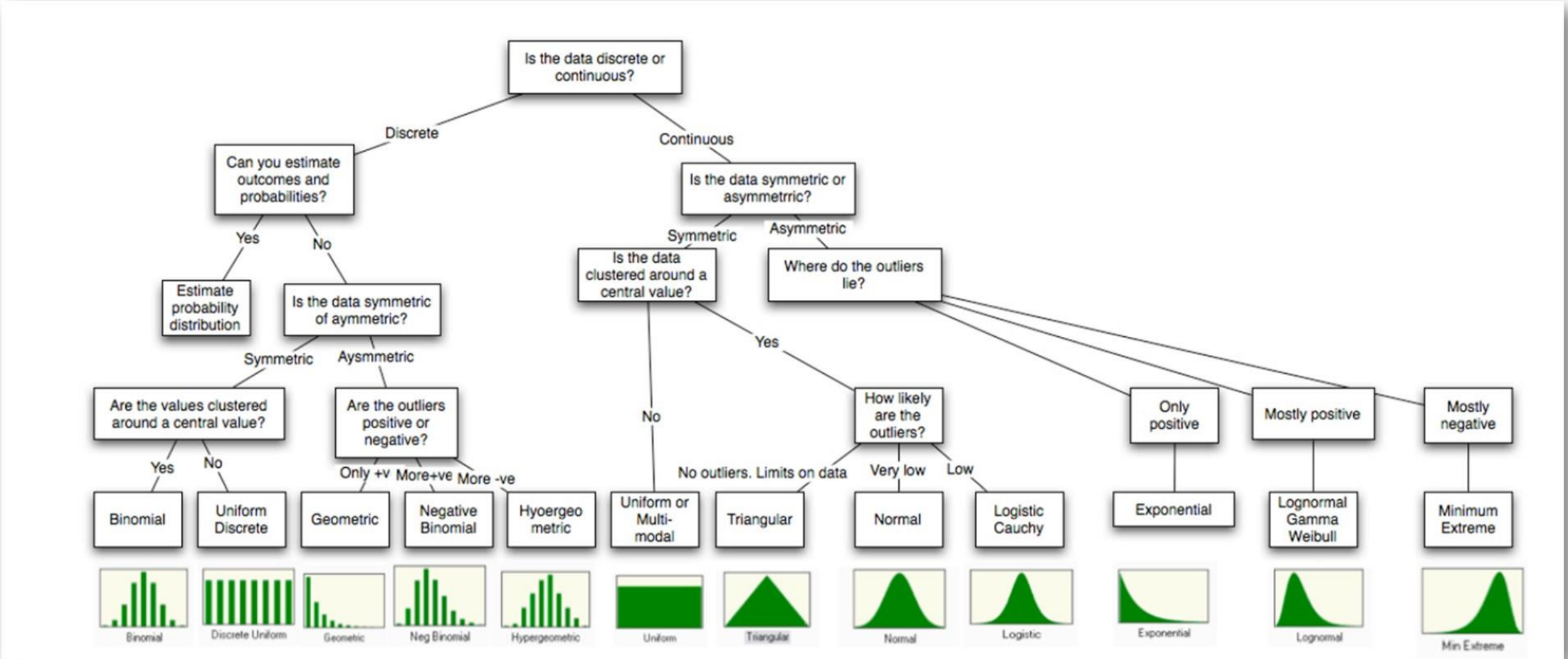
Degrees of freedom

- The t-distribution is of the utmost importance in statistics
- Degrees of freedom indicated with parameter k

$$t = \frac{x - \mu}{s/\sqrt{n}} \sim tk - 1$$

Summary

Which distribution to use?



Hypothesis testing

Hypothesis testing

Is our statement about a population correct or not?

Which statement is best supported by our sample data?

Hypothesis testing
6 steps

1. Null hypothesis
2. Alternative hypothesis
3. Level of significance
4. Test statistic
5. Critical value(s)
6. Decision

Statistical hypothesis testing

Assumptions:

Last year, the monthly energy costs for a family was €260.

In order to check the monthly energy cost for this year, a data scientist takes a sample of 25 families.

The results are the following:

n = 25

$\bar{x}$ = 330,6

s = 154,2

Sample mean = 330,6

Population mean = 260

Does this imply that the population mean is not €260 anymore?

Or is our sample just a bad representation of the population?

Statistical hypothesis testing

Step 1: Nulhypothesis

H_0

A statement that is assumed to be true unless there is strong evidence against it.

- The mean value of a population
- The variance of a population
- The mean difference between two different populations
- The probability distribution followed by a population.

$$H_0: \mu = \mu_0 \qquad H_0: \sigma^2 = {\sigma_0}^2$$

Statistical hypothesis testing
Step 2: Alternative hypothesis

H_a

A statement that is accepted, when the nulhypothesis is rejected

Right-tailed test	Left-tailed test	Two-tailed test
$H_a: \mu > \mu_0$	$H_a: \mu < \mu_0$	$H_a: \mu \neq \mu_0$

Statistical hypothesis testing
Hypotheses

Null hypothesis

H_0

The null hypothesis mostly assumes there is no relationship, or no difference.

The null hypothesis does normally not reflect the desired situation.

H_0 always tests the equality

H_0: =

KEEP CALM AND TEST YOUR HYPOTHESIS

Alternative hypothesis

H_a

The alternative hypothesis mostly holds the statement the researcher wants to prove.

H_a always tests the inequality:

H_a: ≠, >, <

Statistical hypothesis testing
Nondirectional versus directional hypotheses

H_0: parameter = value

E.g. H_0: $\mu = 20$ or H_0: $\sigma = 0,25$
H_0: The data have a normal distribution

H_a: always tests the inequality

Nondirectional

E.g. H_a: $\mu \neq 20$ or H_a: $\sigma \neq 0,25$
H_a: The data do not have a normal distribution

Directional

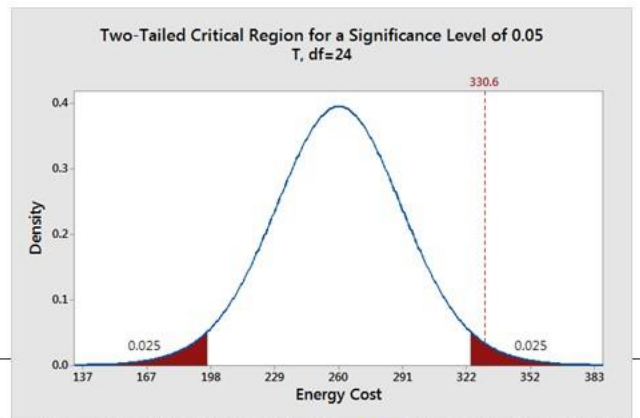
E.g. H_a: $\mu > 20$ or H_a: $\sigma < 0,25$

Statistical hypothesis testing
Nondirectional versus directional hypotheses

H_a: always tests the inequality

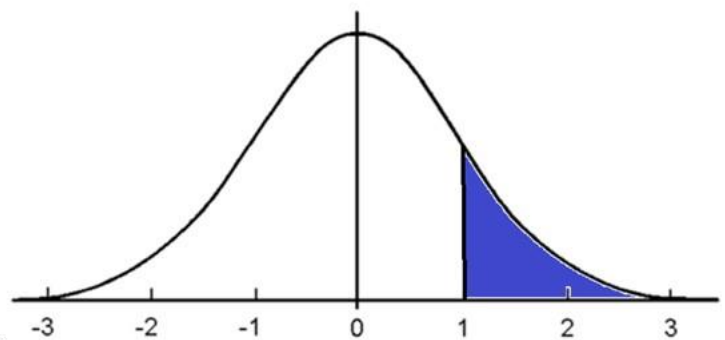
Nondirectional

2-tailed



Directional

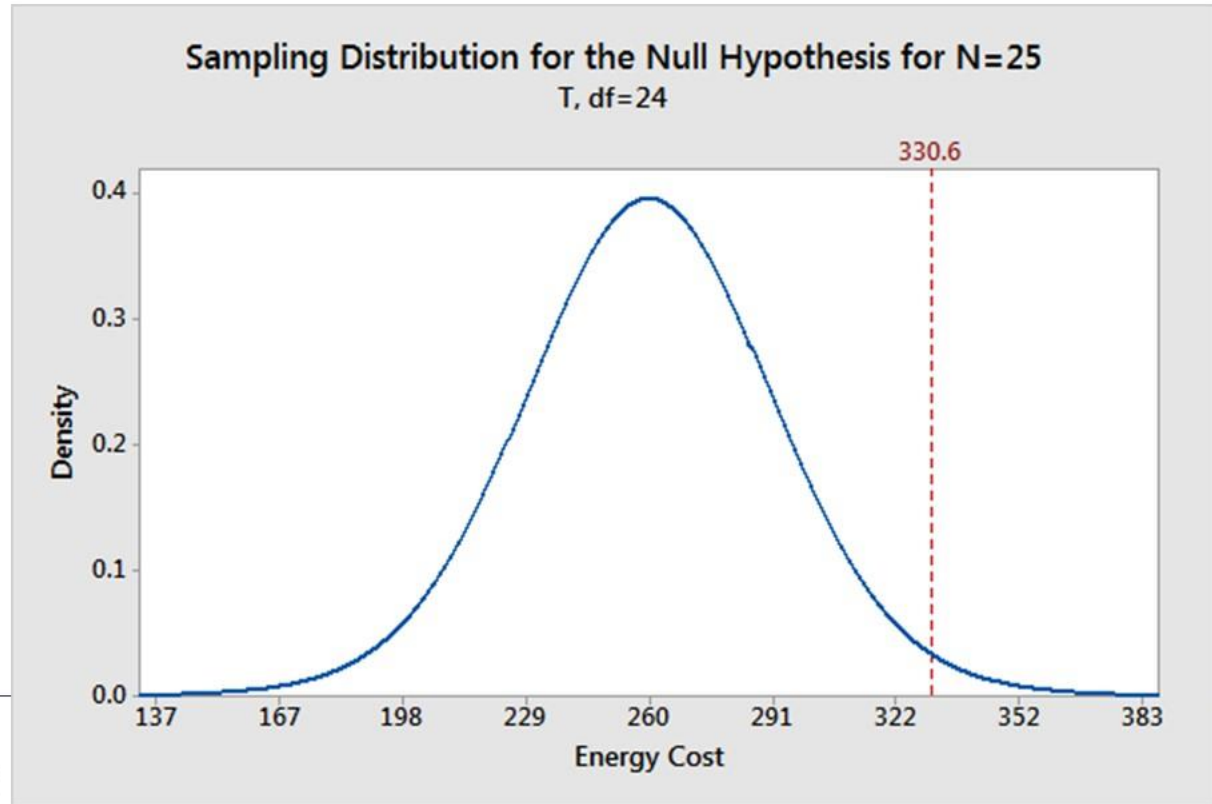
1-tailed



Statistical hypothesis testing

We can use the central limit theorem to take multiple random samples of the same size from the same population

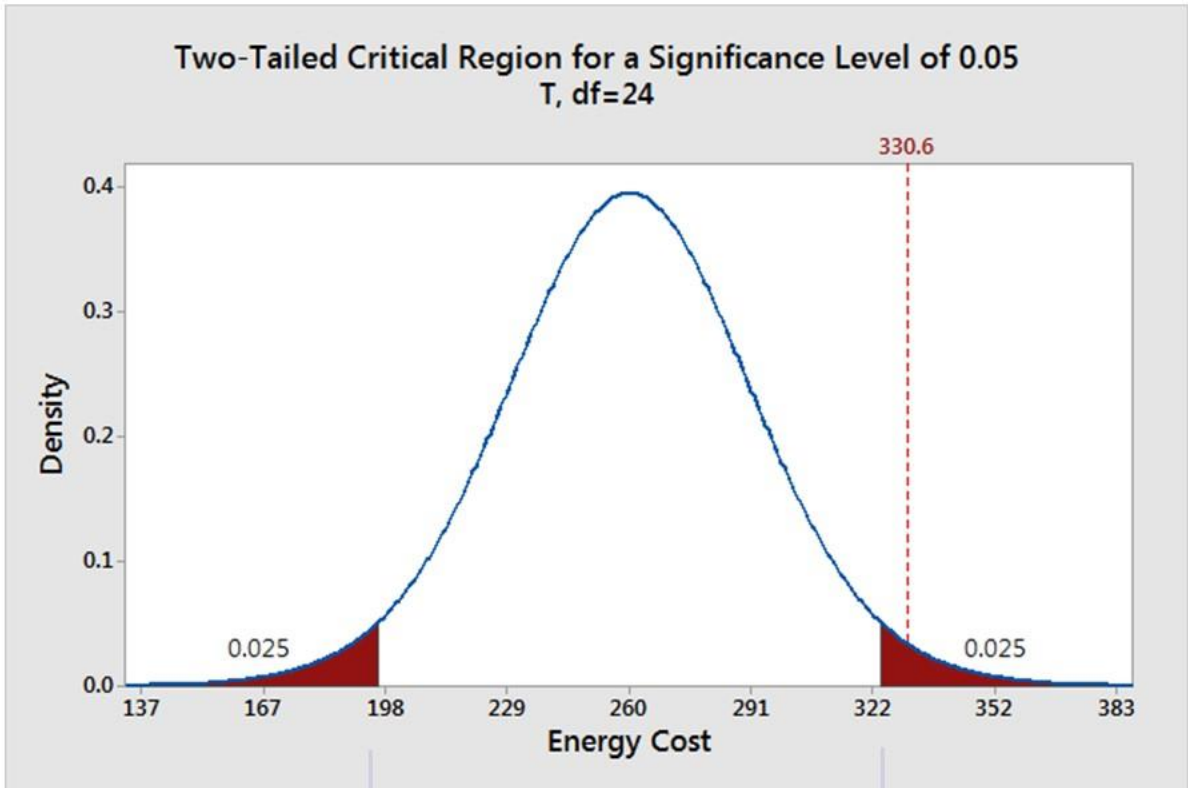
And plot the distribution of the population means



Statistical hypothesis testing
Step 3: Level of significance α

The probability of rejecting the null hypothesis when it is true

$H_0: \mu = 260$



Statistical hypothesis testing
Step 4: test statistic

Two options

1. Calculate confidence interval

   → Determine if value is within acceptable or critical range

2. Calculating p-value

   i.e the chance of finding a value more extreme than the one observed

   → Compare with chosen significance level

Statistical hypothesis testing
Step 4: test statistic

The numerical measure which is computed from sample data to determine whether or not the null hypothesis should be rejected.

N

Assumptions

Test for population mean: μ
Population variance σ^2 is known

Normal distribution

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

Standard normal distribution

$$Z \sim N(0, 1)$$

T

Assumptions

Test for population mean: μ
Population variance σ^2 is unknown
The population variance will be estimated using s^2
A normal distribution is followed!

Student's t-distribution

$$X \sim t_k$$

$$t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Statistical hypothesis testing
Step 5: critical values

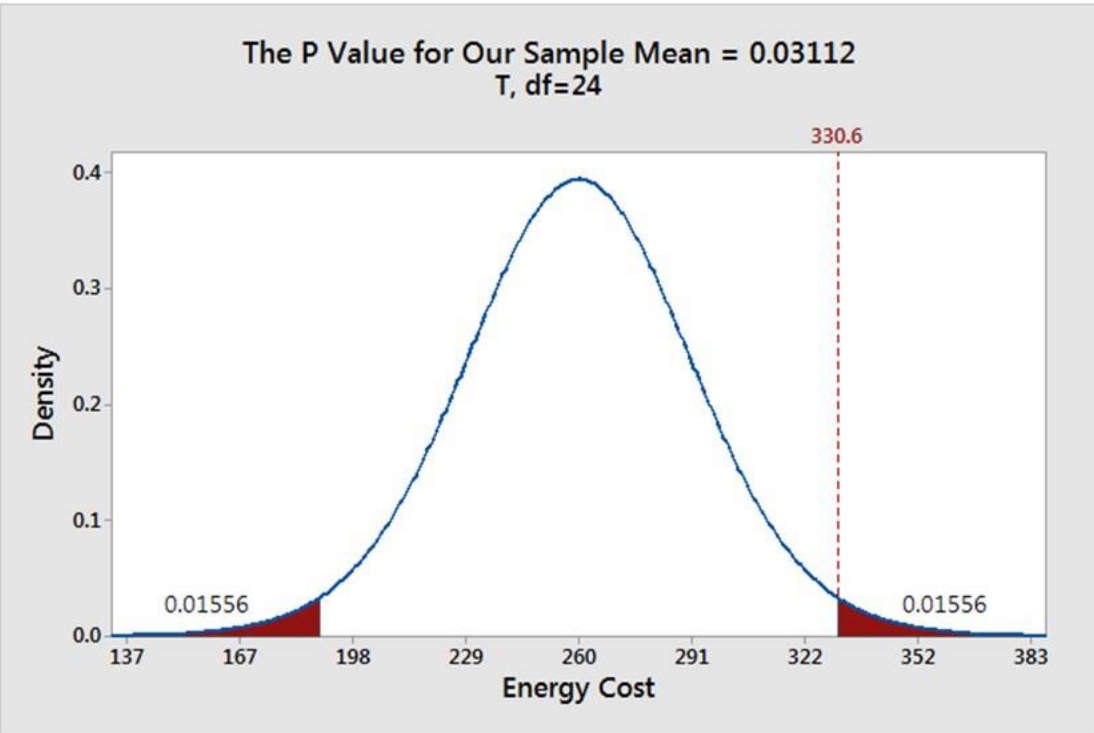
Two-tailed test	Right-tailed test	Left-tailed test
$-t_{\alpha/2;n-1}$ $t_{\alpha/2;n-1}$	$t_{\alpha;n-1}$	$-t_{\alpha;n-1}$

Statistical hypothesis testing
Step 6: Reject or accept

Two-tailed test	Right-tailed test	Left-tailed test
$t < -t_{\alpha/2;n-1}$ $t > t_{\alpha/2;n-1}$	$t > t_{\alpha;n-1}$	$t < -t_{\alpha;n-1}$

Statistical hypothesis testing
Alternative p-value

The chance of finding an effect which is at least as extreme as the one of your sample data, assuming that H_0 is true.



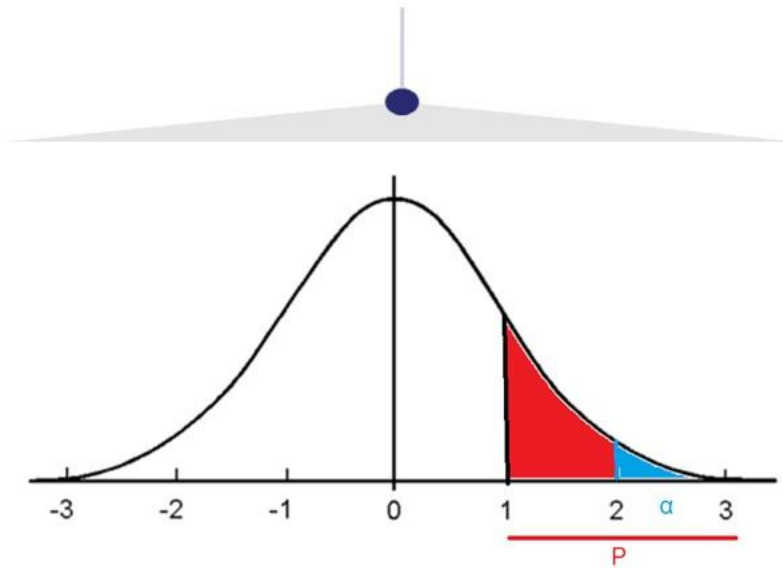
2-tailed: Compare $\frac{p}{2}$ ***to*** $\frac{\alpha}{2}$

In case $\frac{p}{2} \leq \frac{\alpha}{2}$
Reject H_0

Statistical hypothesis testing
Acceptance or rejection of H_0 for 1-tailed test

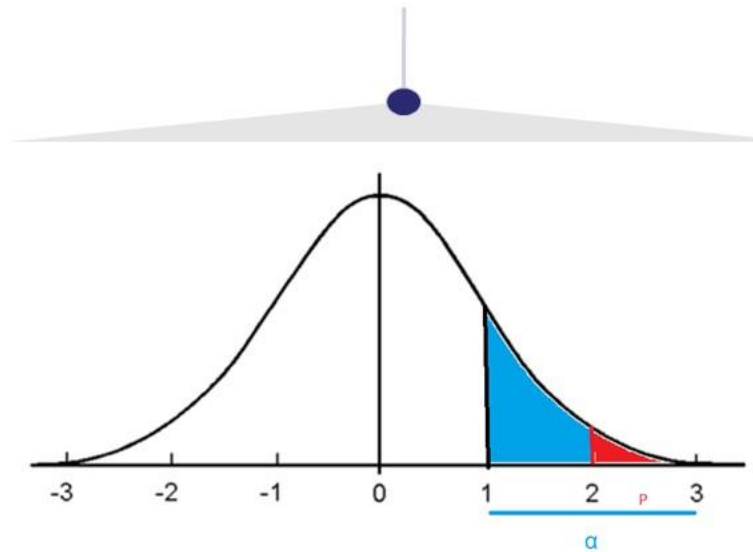
P- value

Assuming the null hypothesis were true, what is the probability of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed.



$P \geq \alpha$

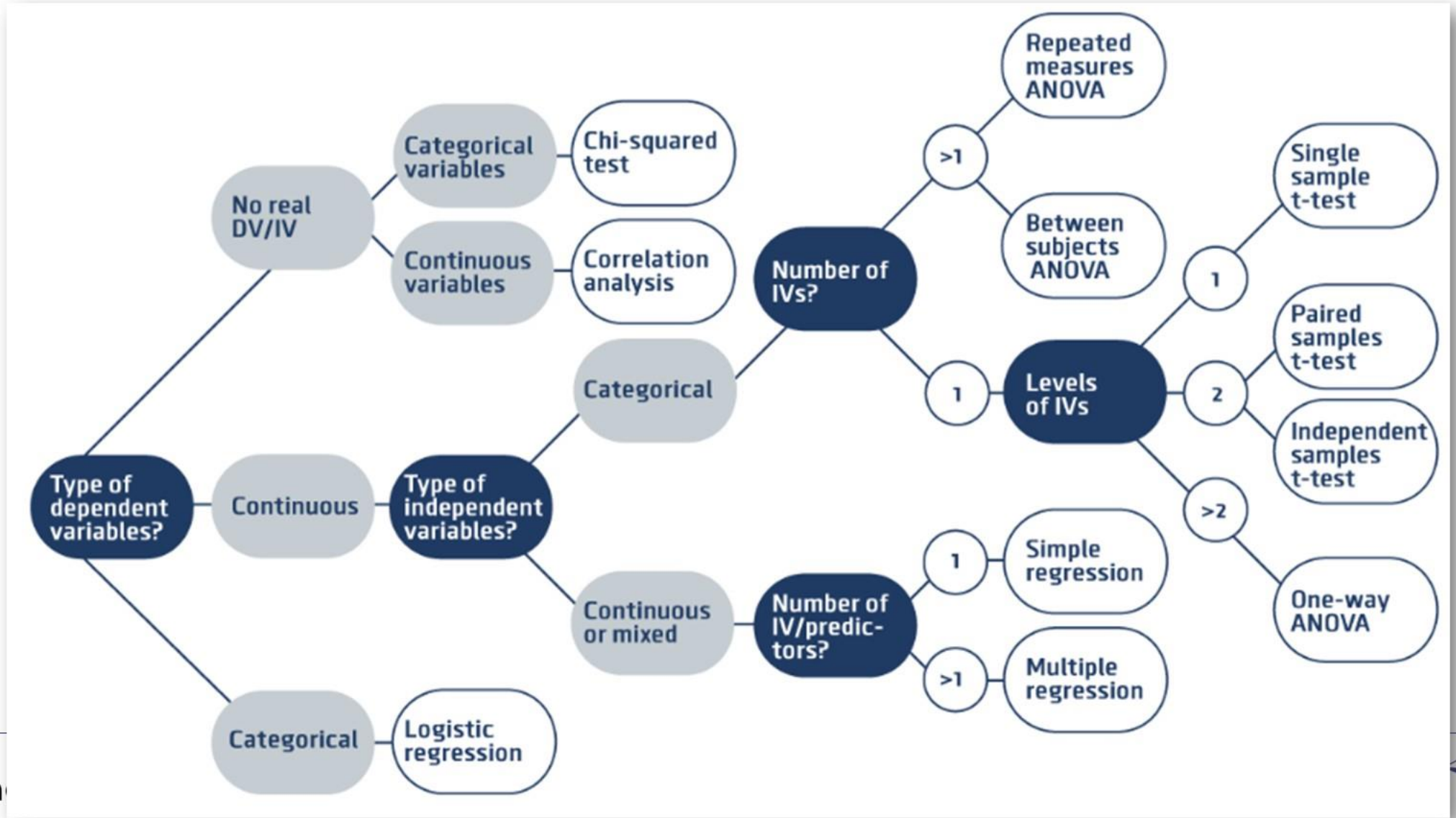
Accept H_0



$P < \alpha$

Reject H_0

Statistical hypothesis testing
Which test do I need? Questions to ask



T-Distribution

Exercise

A survey indicates that for each trip to the supermarket, a shopper spends an average of 45 minutes with a standard deviation of 12 minutes in the store. The length of time spent in the store is represented by the variable x. A shopper enters the store.

Find the probability that the shopper will be in the store for between 24 and 54 minutes

Statistical hypothesis testing
Possible tests

Parameter	1 **Population**	Type of **critical value**
Mean	$H_0: \mu = \mu_0$	Student's t-distribution
Variance	$H_0: \sigma = \sigma_0$	χ^2 - distribution
Proportion	$H_0: \pi = \pi_0$	Standard normal distribution
Median	$H_0: Me = Me_0$	Binomial distribution

When $parameter_0$ is a specific value

Distribution	1 **Population**	Type of test
Poisson	e.g. H_0: The number of visitors follow a poisson distribution	χ^2-test
Normal	H_0: The data have a normal distribution	Kolmogorov – Smirnov Lilliefors test χ^2-test

Statistical hypothesis testing
Possible tests

Parameter	1 **Population**	2 **Populations**	3 **Populations**
Mean	$H_0: \mu = \mu_0$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2 = \mu_3$
Variance	$H_0: \sigma = \sigma_0$	$H_0: \sigma_1 = \sigma_2$	$H_0: \sigma_1 = \sigma_2 = \sigma_3$
Proportion	$H_0: \pi = \pi_0$	$H_0: \pi_1 = \pi_2$	$H_0: \pi_1 = \pi_2 = \pi_3$
Median	$H_0: Me = Me_0$	$H_0: Me_1 = Me_2$	$H_0: Me_1 = Me_2 = Me_3$