# Hypothesis Testing as Bivariate Analyses

j3

– 01.04.26 –


Institut Méditerranéen d'Océanologie

Institut de Recherche pour le Développement FRANCE

Plateforme OMICS - MIO

**Bio-informatique & Sciences de l'Environnement : Exploration de la Diversité Taxonomique des Ecosystèmes par Metabarcoding**

cnrs

Variability in species richness between the 2 stations, depths

**Is there a real significative difference or just a coincidence ?**

Using **statistics** to answer your question !!

# Population VS samples

Population: set of individuals or objects of the same kind (very large or infinite)

→ We can't study an entire population: in statistics, we study a limited number of individuals, a part of the population: **a sample**

→ We try to **deduce properties** of the population from the sample

→ If we want to **study the variability** of a variable of interest in the population, we need a **representative sample** (drawn at random)

In a population, we can measure a characteristic: **a variable** that is the result of a random phenomenon.

- Qualitative
- Quantitative (continuous)

A **probability law** describes the random behavior of a phenomenon that depends on chance.

In a population, we can measure a characteristic: **a variable** that is the result of a random phenomenon.

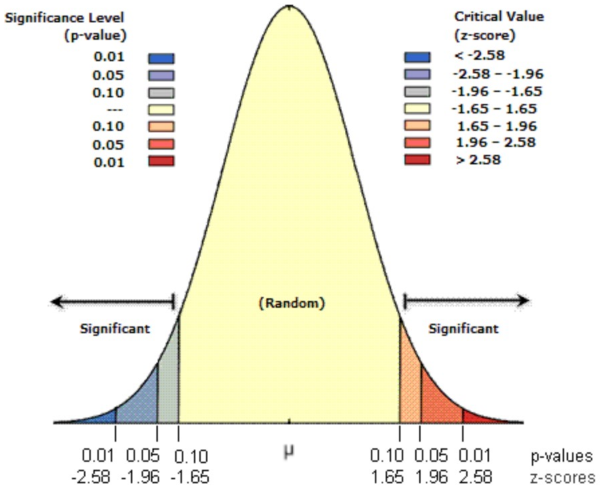
- Qualitative
- Quantitative (continuous)

A **probability law** describes the random behavior of a phenomenon that depends on chance.

## THE NORMAL LAW

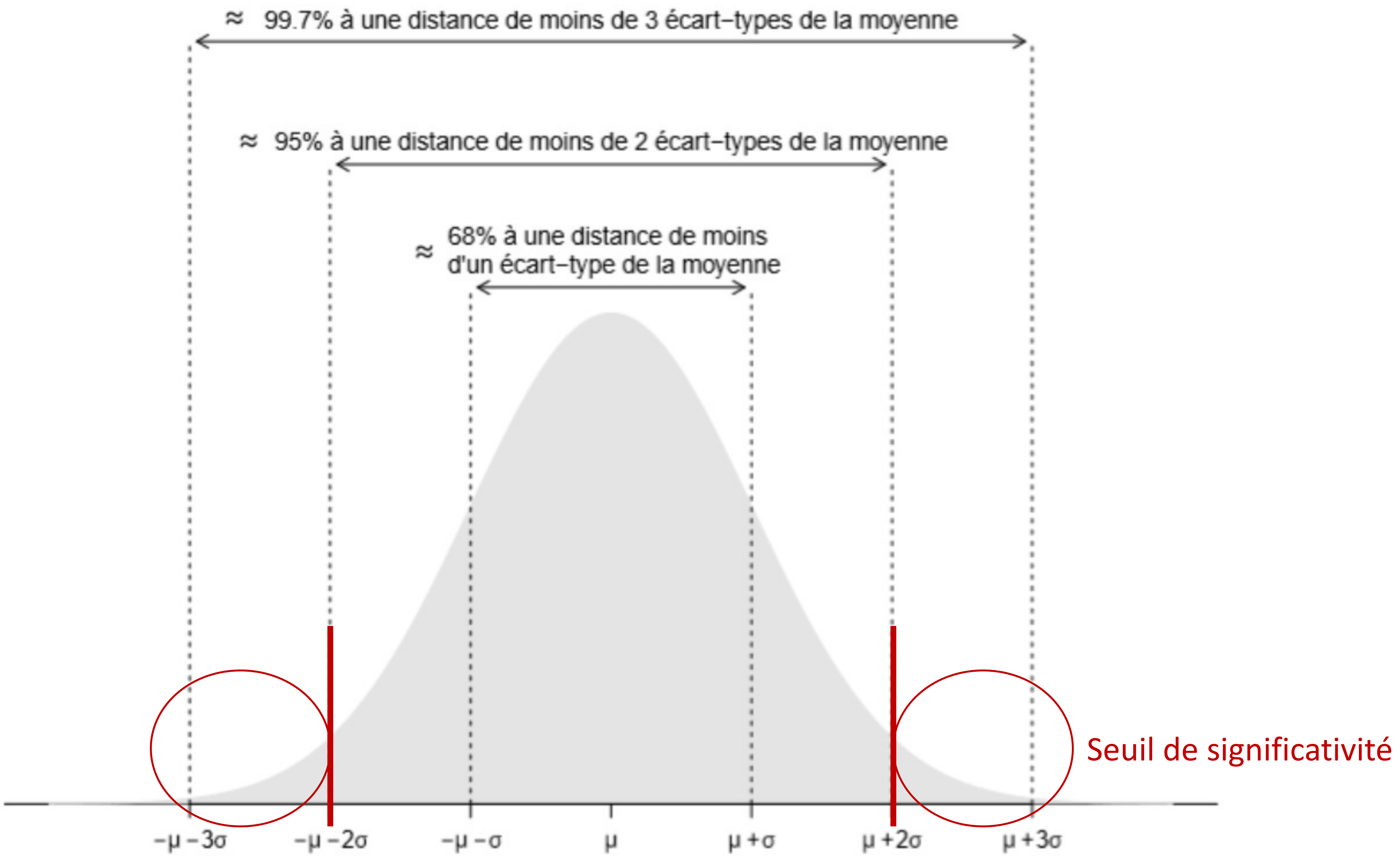
If we have 1000 samples of a variable following a normal distribution, and plot the number of samples equal to each value, we obtain a "bell" curve / gaussian distribution

$X \sim N(\mu,\sigma 2)$ with $\mu$ and $\sigma 2$ the parameters of the distribution:
- $\mu$: expectation of X
-$\sigma$ = dispertion around the mean = Standard deviation (SD) of X




- Symetric, unimodal
- Center around the mean

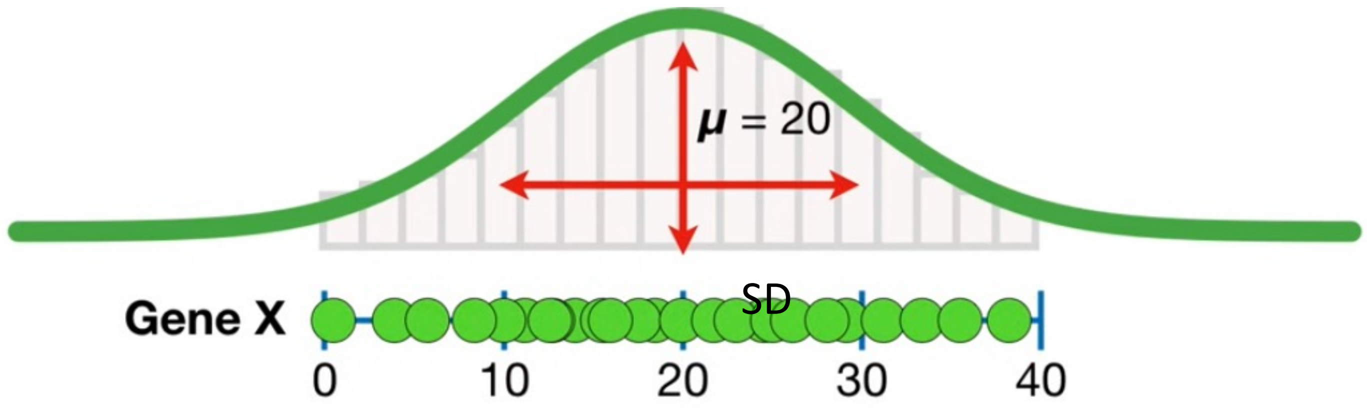
# Répartition des valeurs autour de la moyenne



# Remember : Descriptive statistics (Univariate analysis)


**Merely describe, show and summarize collected data**

- **Central tendency** (mean, mediane…)
- **Dispersion** (variance, standard deviation)
- **Frequency distribution** (count, relative, cumulative)

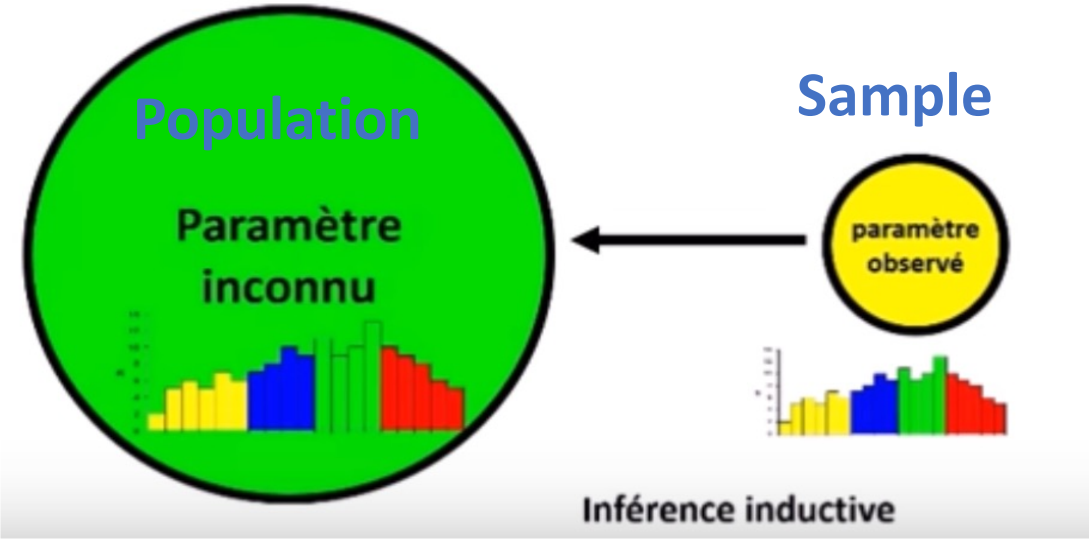



Identify the characteritics of data for each variable(s)

→ **Allows you to formulate hypotheses and guide statistical analyzes**

# Inferential Statistics

## Predictions - Generalizations


Make inferences about the population

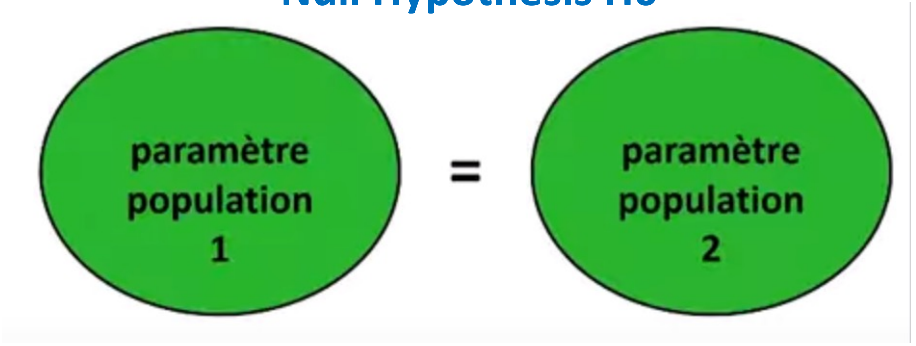
- How can I use my sample to make predictions about the population = **Estimation**
- How do I prove a theory about my data's behaviour (comparison) = **Hypothesis Testing**

# Hypothesis testing approach

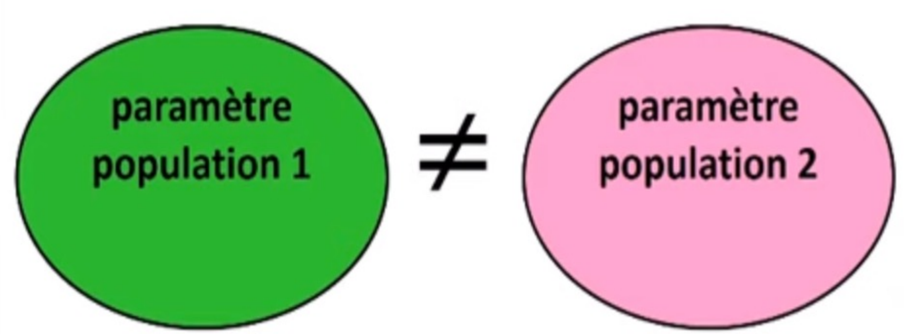
Trying to validate a hypothesis relating to a population parameter from a sample comparisons

**Is there a real difference or just a coincidence (chance)**

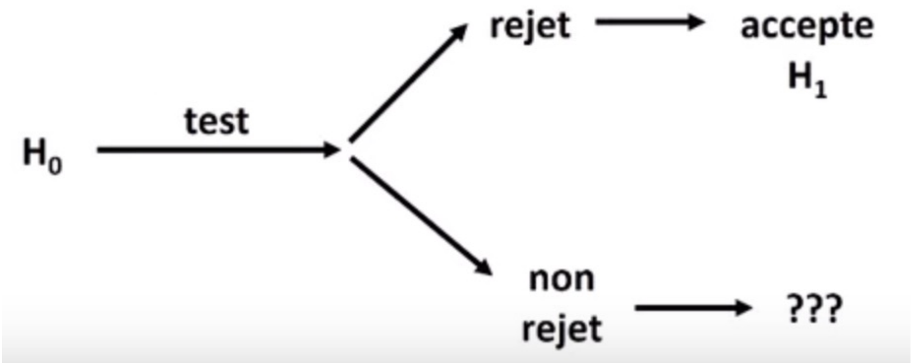
**Null Hypothesis H0**




**Alternative hypothesis H1**




**We are testing the null hypothesis!**

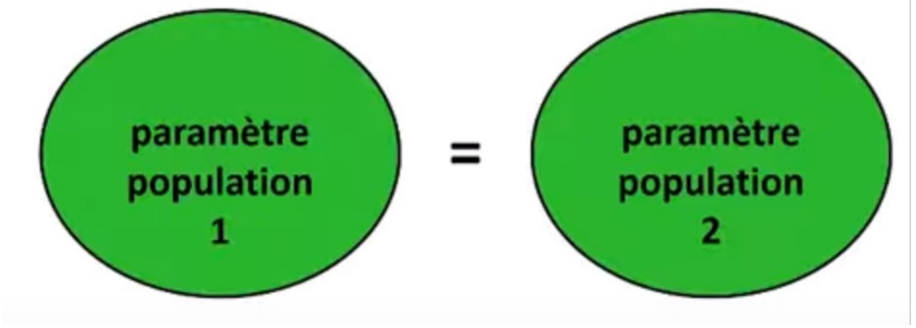


# Hypothesis testing approach

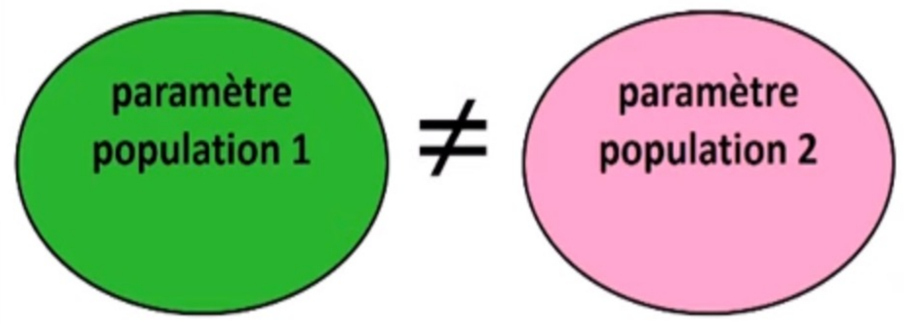
Trying to validate a hypothesis relating to a population parameter from a sample comparisons

**Is there a real difference or just a coincidence (chance)**

**Null Hypothesis H0**




**Alternative hypothesis H1**

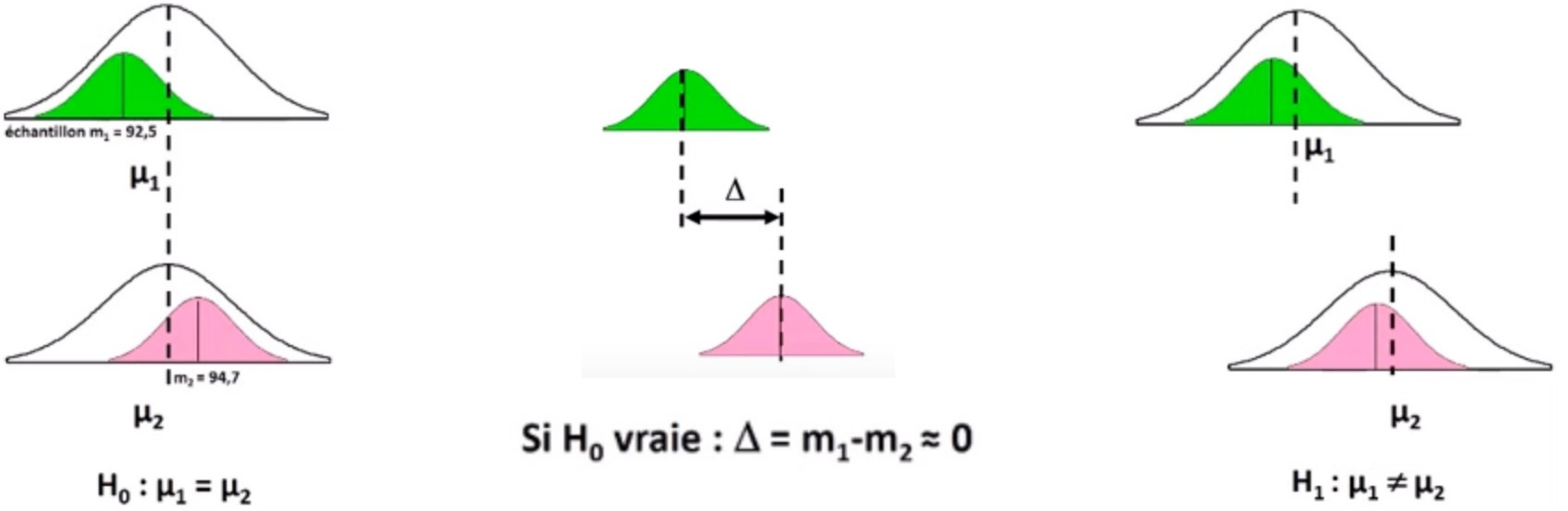



**"Absence of Evidence is not Evidence of Absence"**

# Hypothesis testing & mean comparison

**If H0 true… no difference**

**If H0 rejected, H1 accepted**



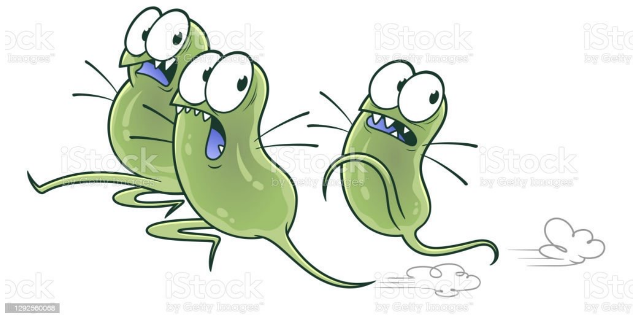

**SAME distribution**
→ **Sampling fluctuation**

**Two different distributions**

Inference Issue : Subjected to errors!!

The risk is linked to the result of hypothesis testing

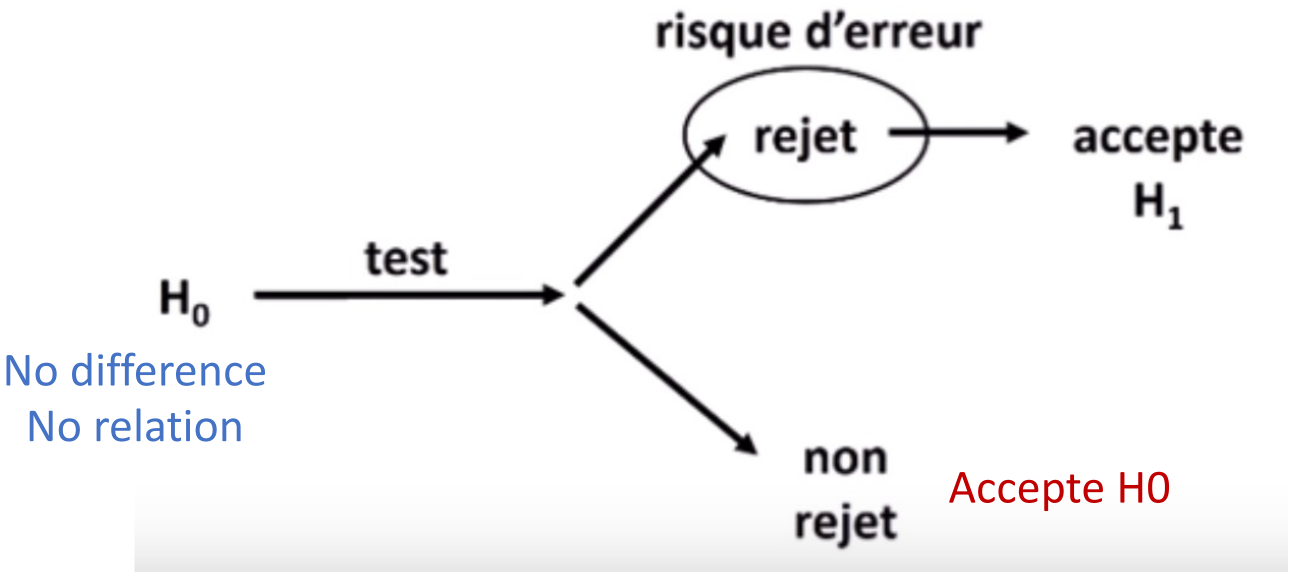
Because of your sampling!



# The risk of Type I error $\alpha$

- **A probability between 0 and 1, or 0 and 100%**

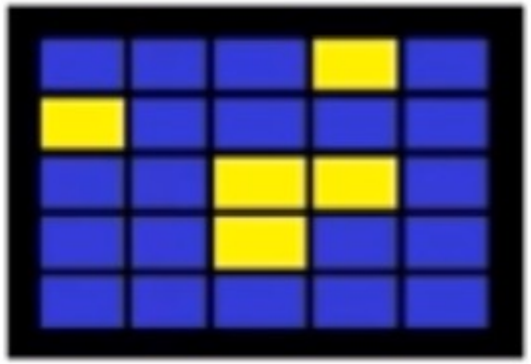
- **Is when a difference is affirmed but there is none (=False positive)!!**



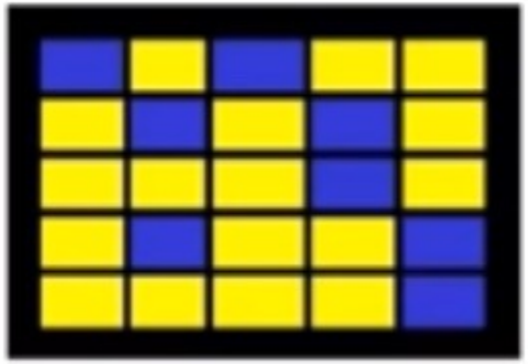

**$\alpha$= Risk to reject H0 if H0 is true**

# Sampling

25 tiles
→ 80% blue



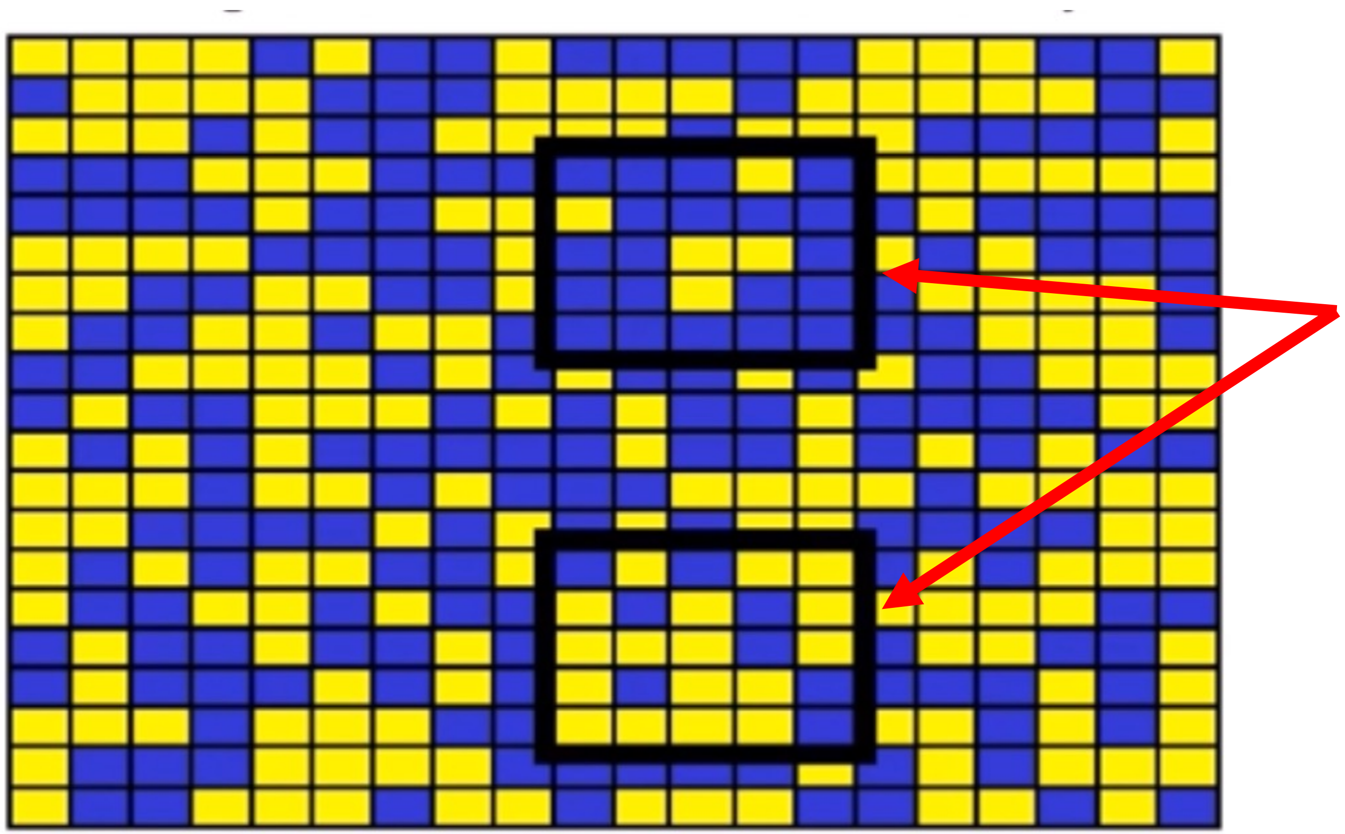
25 tiles
→ 32% blue



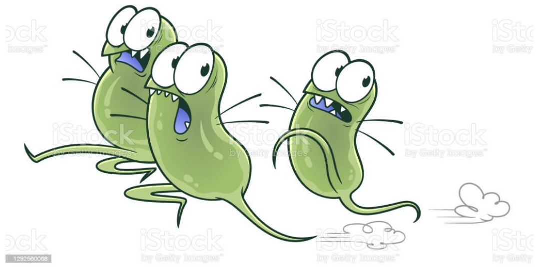
Do the two samples come from the same population? (same distribution)?

- H0 is rejected
- but let's go to the store...see the population

# Come from the same population (50% blue, 50 % yellow)!!

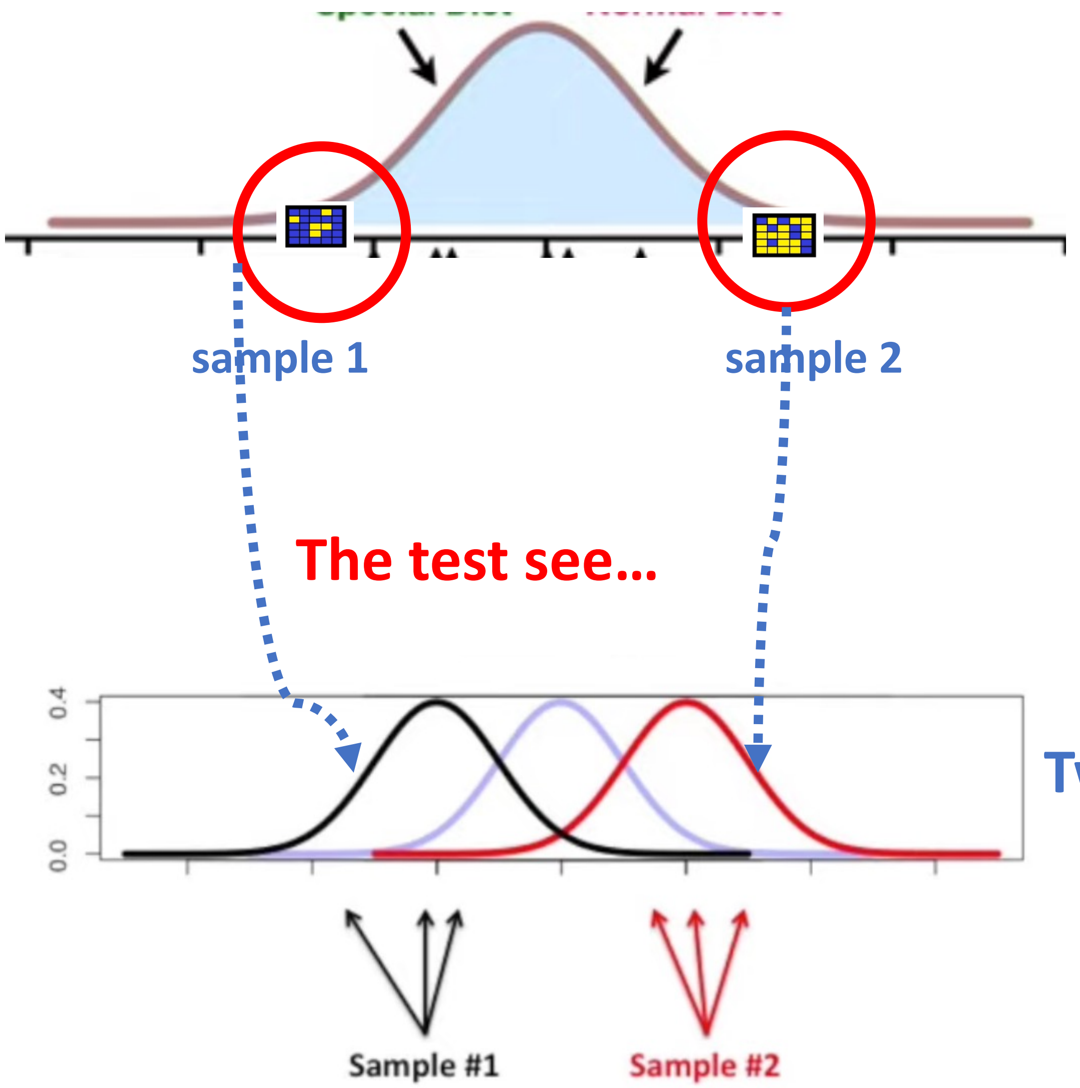


Rare sample type



Conclude on the basis of our samples that they came from two different distributions
= Type I error

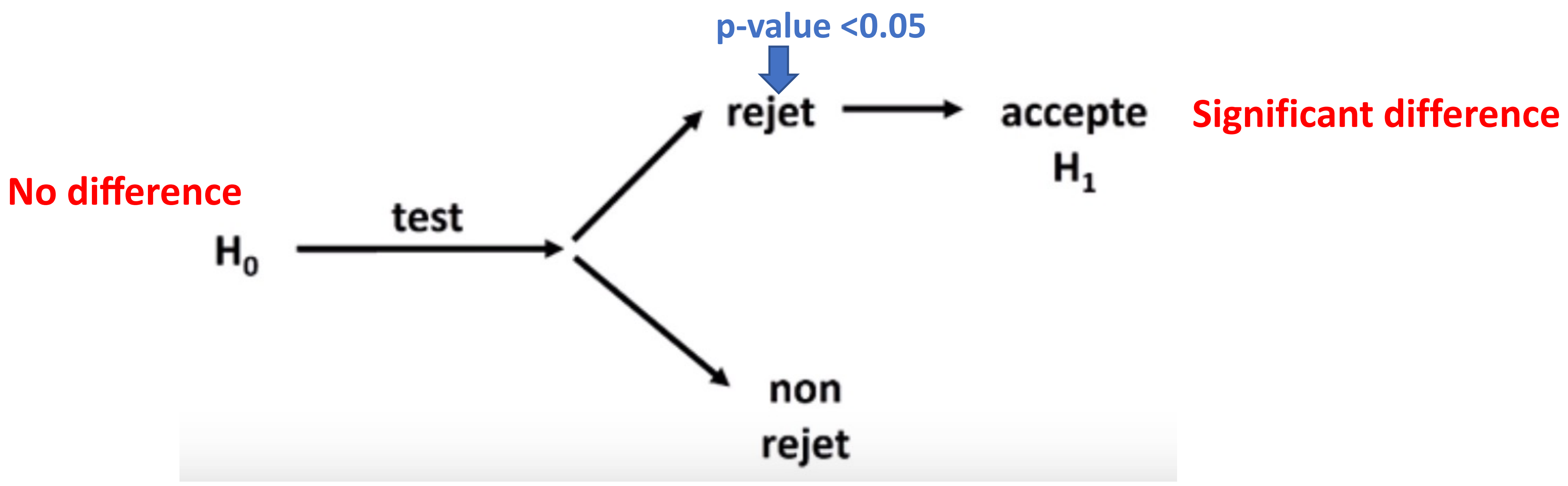
# Data come from the same distribution but …




Two different distributions

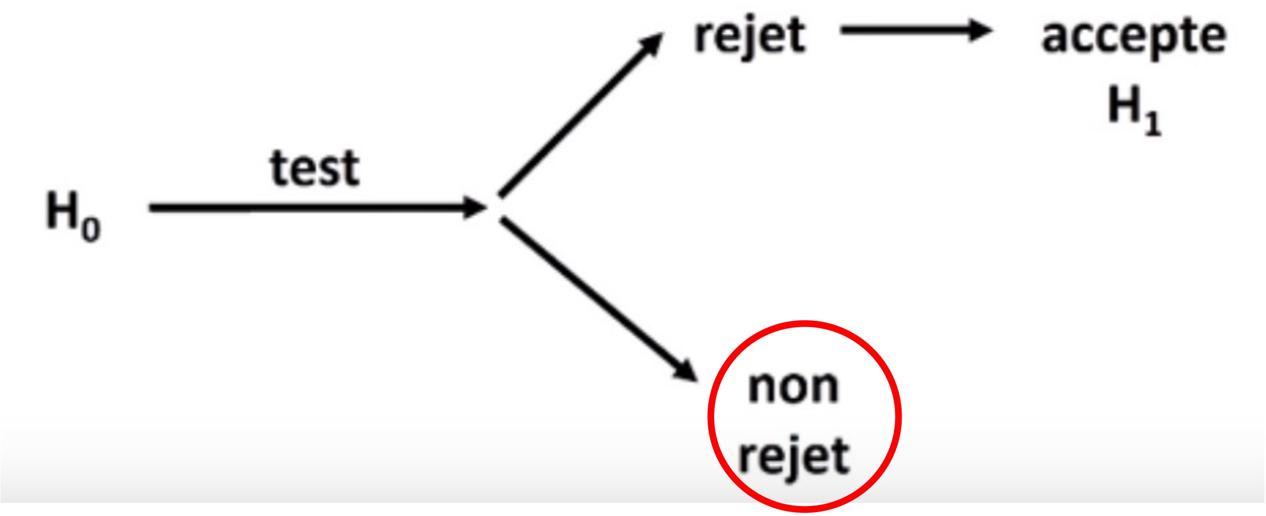
- $\alpha$ is choosen before the test : Significance threshold

- $\alpha$ often set 5% (H0 wrongly rejected)
- In science the "almost no chance" translates to in less than 5% of cases where H0 is true = $p\text{-value} < 0.05$

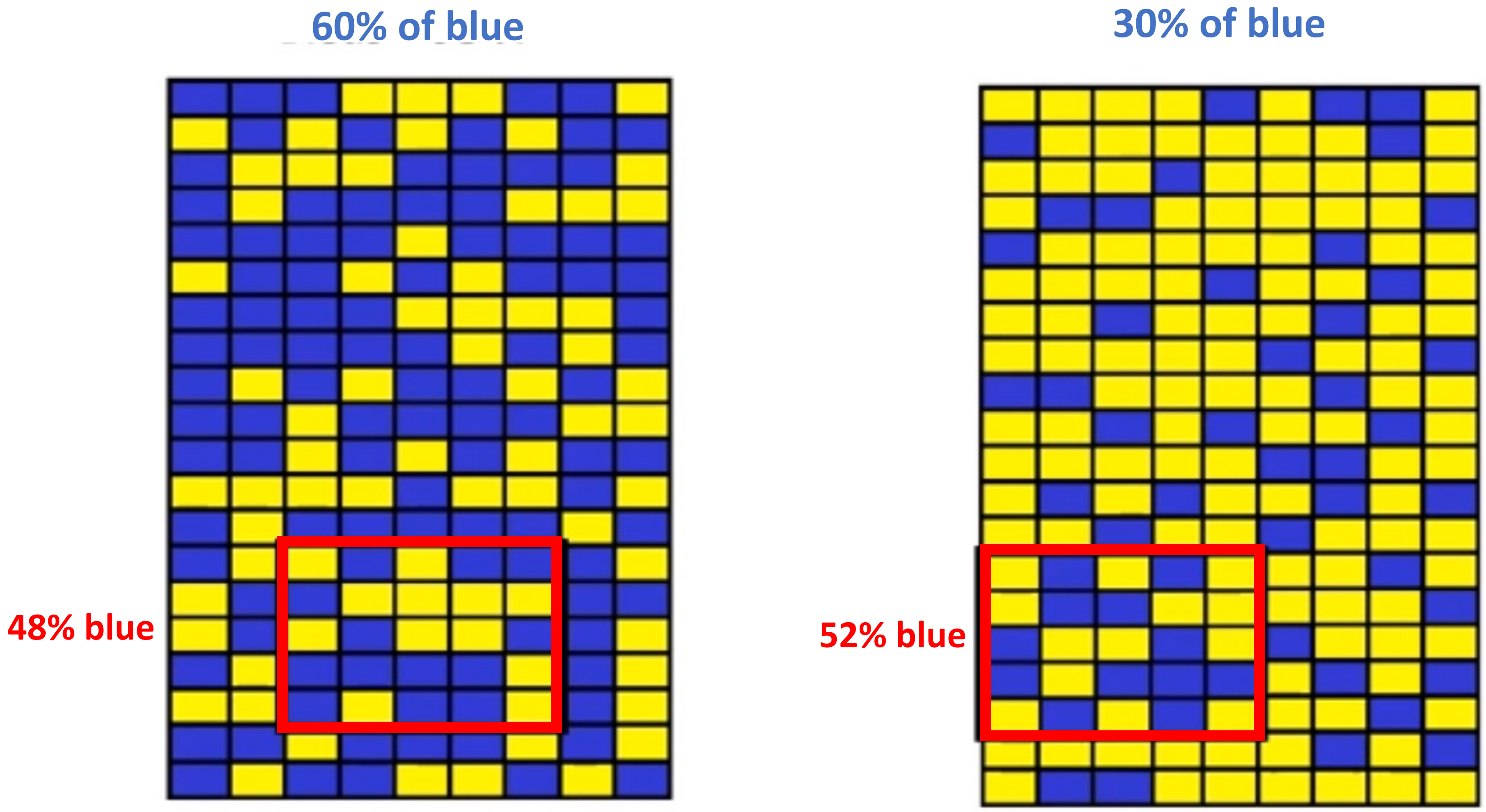


# Risk of Type II Error : $\beta$

Failing to conclude **a difference when there is a true one** ("False Negative")
Probability of not rejecting H0, if H1 is true

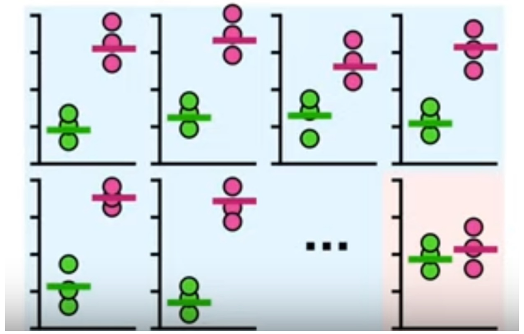


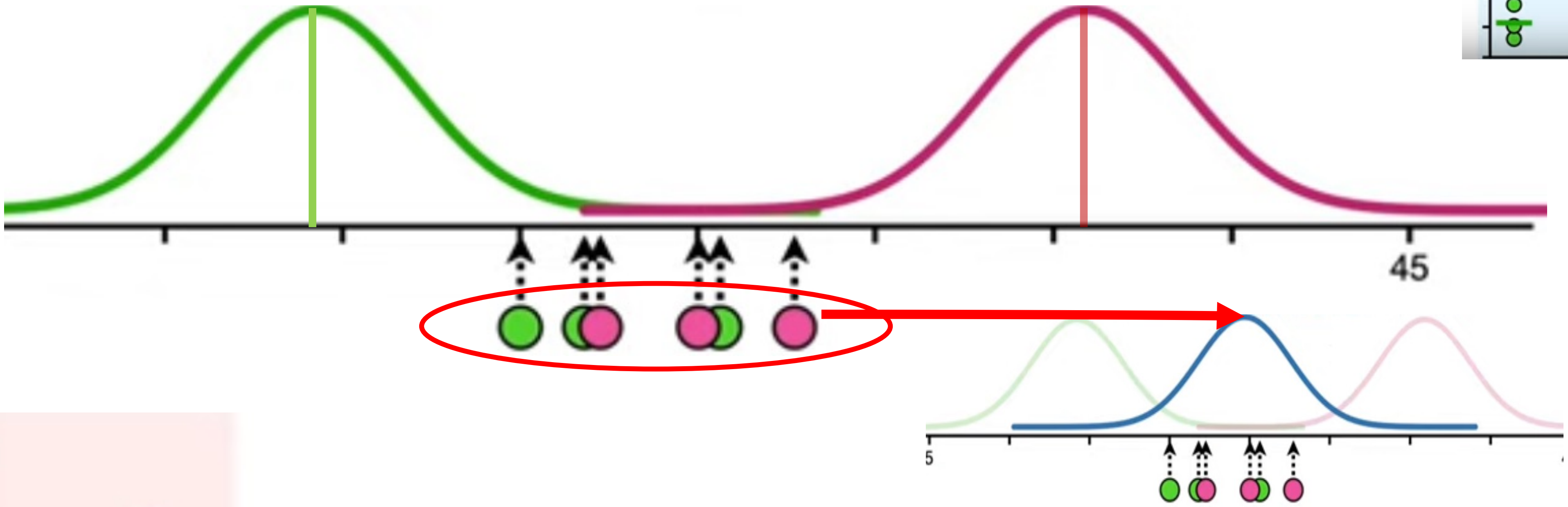

$\beta$ is not calculable

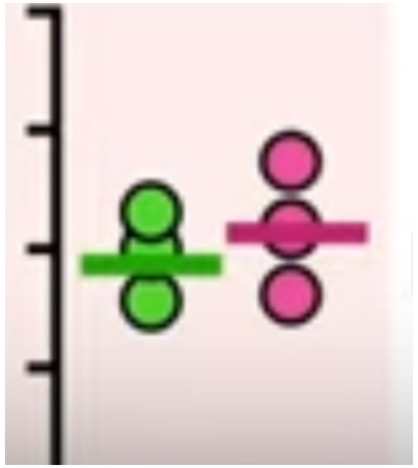



- 2 different tiles = 2 different populations, H0 should be rejected But that would not have been the case during the test with our sampling…

# But sometimes…





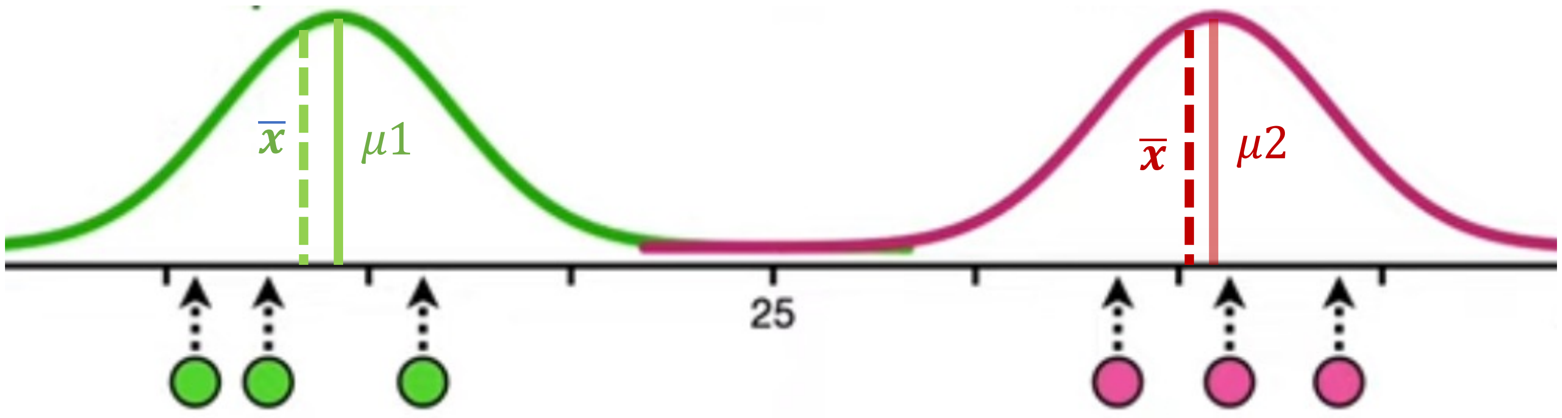



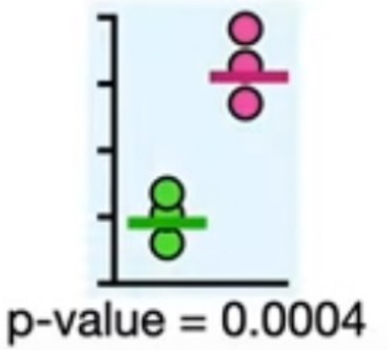
p=0.23!!!

Even if two different distributions (pop)…the test (your data) thinks they come from the SAME distribution!

**Unable to correctly reject H0…**

# Scientifically … representative sampling of population



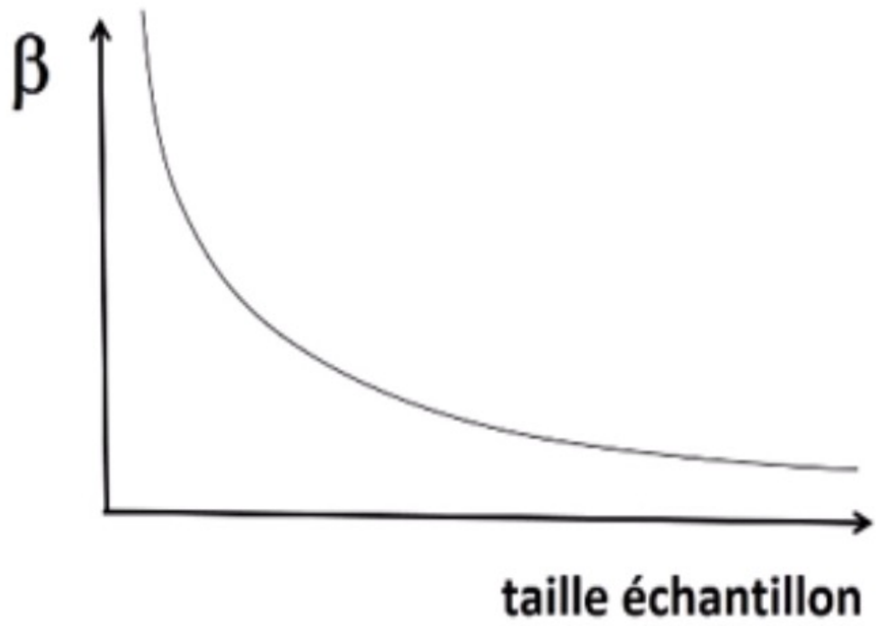

→ H0 correctly rejected

→ = Data do not belong to same distribution

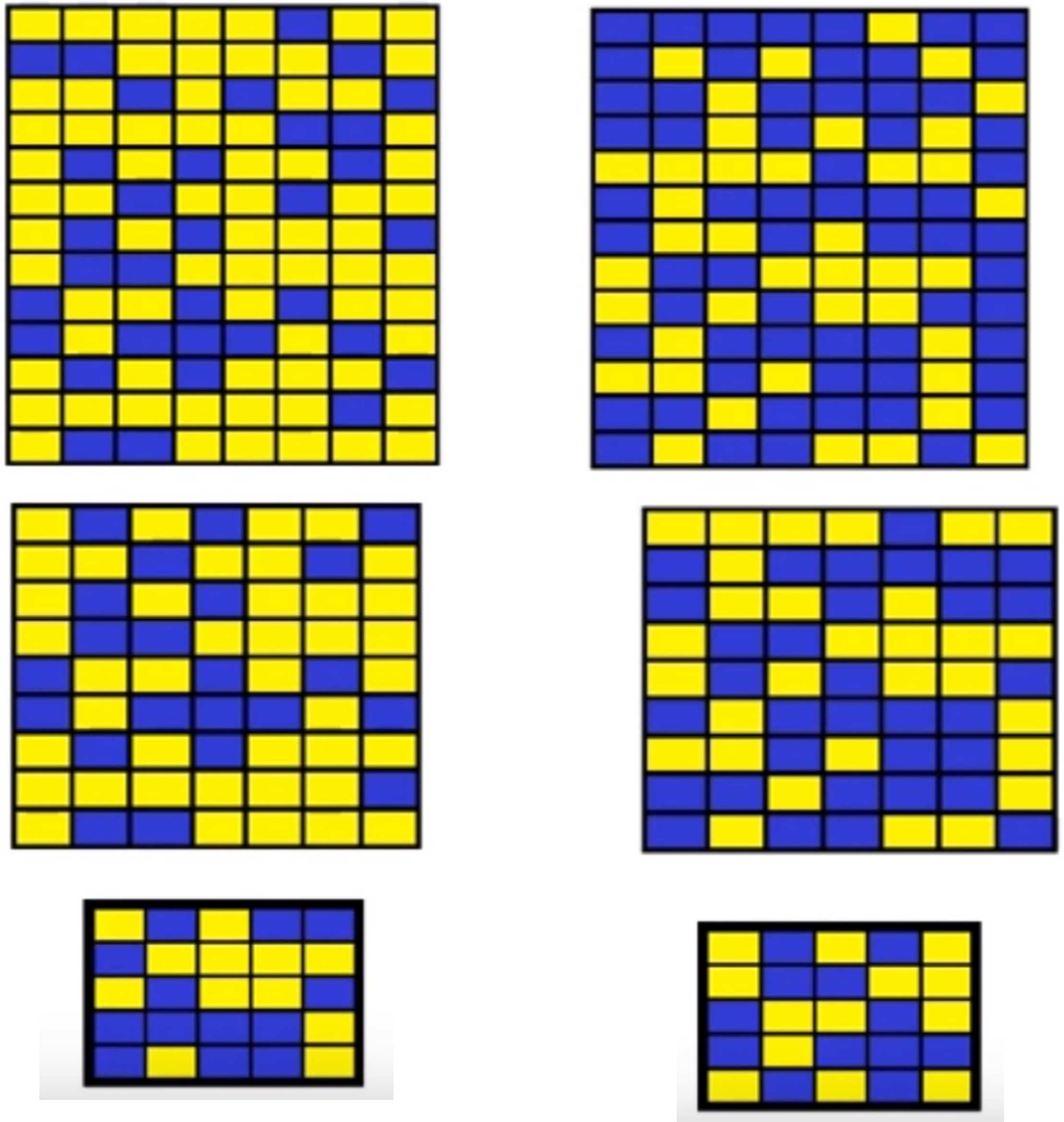
→ Two different populations

# Fundamental relationship

Power $=1-\beta$




The more the size increases, the more the differences appear! The power of the test increases!



**Power:** Probability of correctly reject the H0 hypothesis

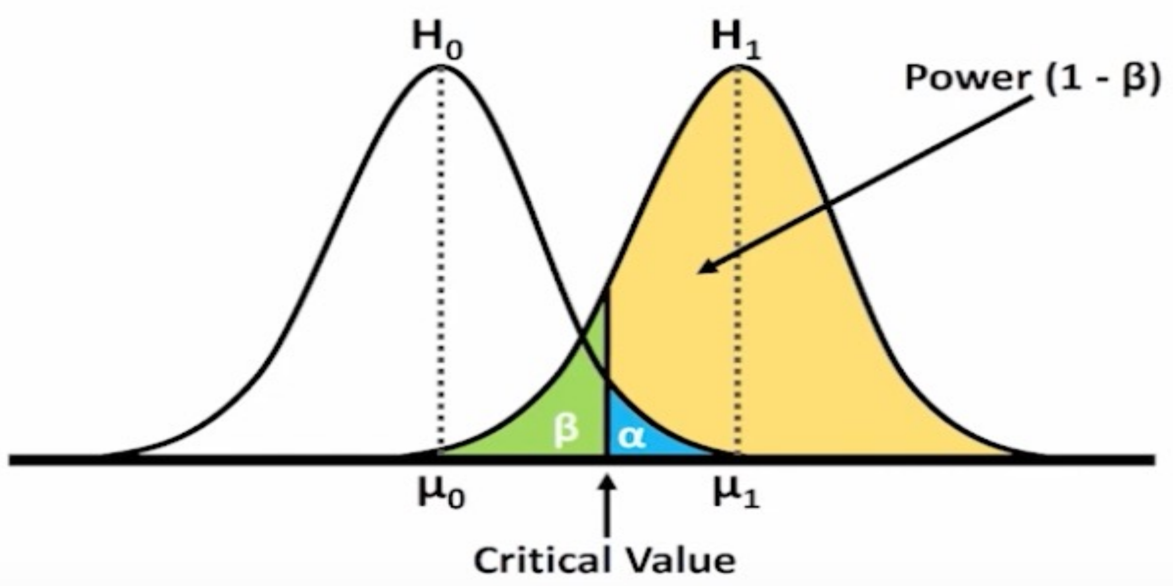
Ability of a test to detect differences

# Summary

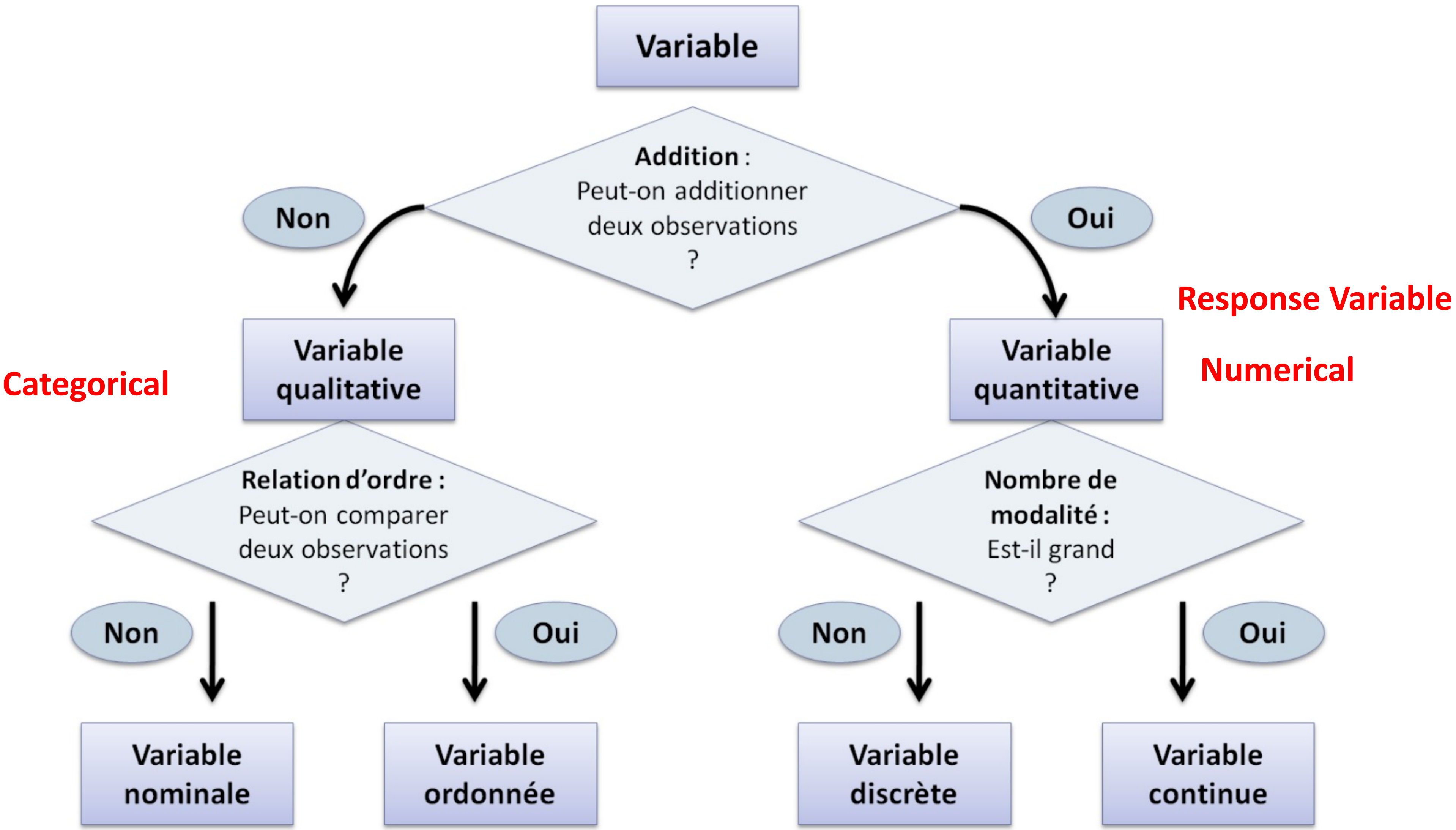
Population

TEST échantillons

| | $H_0$ vraie | $H_1$ vraie |
|---|---|---|
| accepter $H_0$ | OK | *erreur type 2* $\beta$ Faux Négatif |
| rejeter $H_0$ | *erreur type 1* $\alpha$ Faux positif | OK |



# Reminder on variables… important for statistical tests




- Maried, single…

→ No relation order

- Behaviour
- good, excellent…

Child in family (1,2,3..)
finite number of real values

Size, weight : infinite

# Bivariate Hypothesis Testing

- Seek to **quantify the association** between a **variable to be explained** (response/Quantitative) and an **explanatory variable** (factor/categorical)

- **Make statistical inferences about the relationship between two variables,**
  **One quantitative variable (response) & one qualitative (explicative)!**

→ Can variations in species richness (response variable) be explained by the explanatory variable (factor) Treatment

→ **Comparison of mean between groups**

- **Parametric or non parametric test??**
- **which test?? significance ? (p-value)**
- **How many groups??**
- **Post hoc test required ??**

Which test for independent samples?
ONE  categorical variable (H/F) & ONE continuous  variable (numerical)

**Normalité des données?**
**Shapiro, Q-Q plots**

## Which test for independent samples?
## ONE categorical variable (H/F) & ONE continuous variable (numerical)

**Normalité des données?**

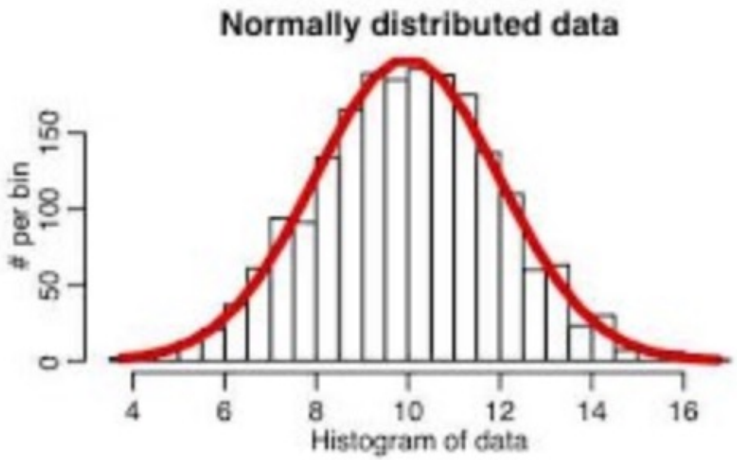
**Shapiro, Q-Q plots**

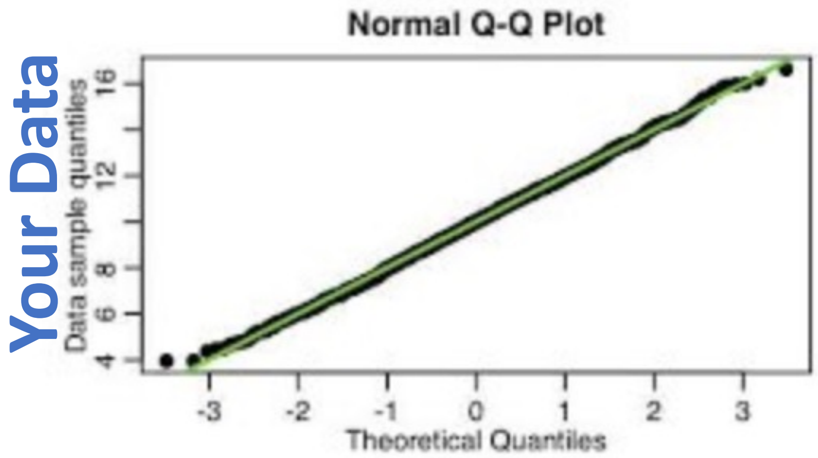
Yes

**Parametric test**

# Q-Q plot normale: Compare your distribution with a normal distribution

## Do my data follow a normal distribution ?



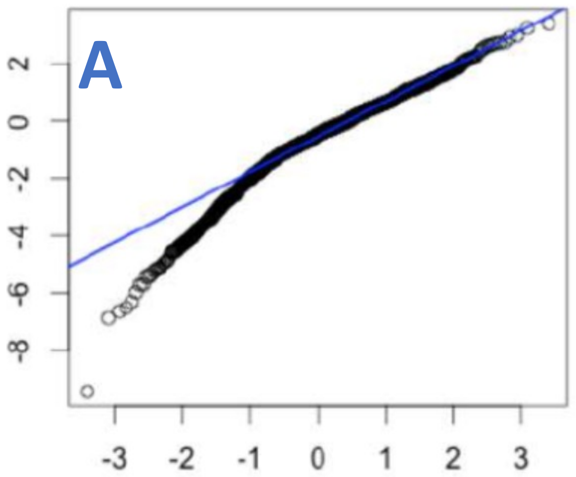



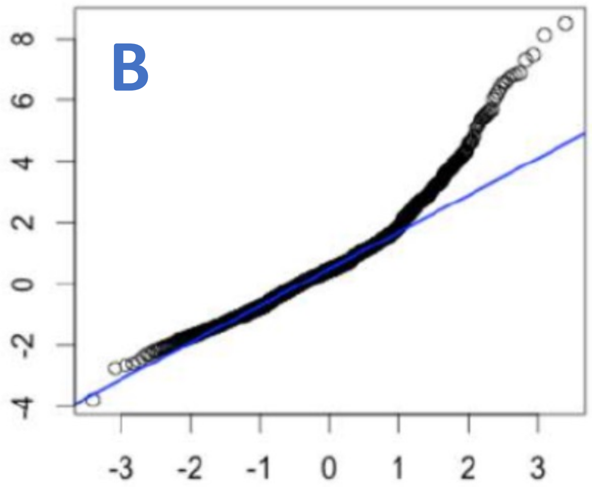

Normal Data (u= 0, SD=1)

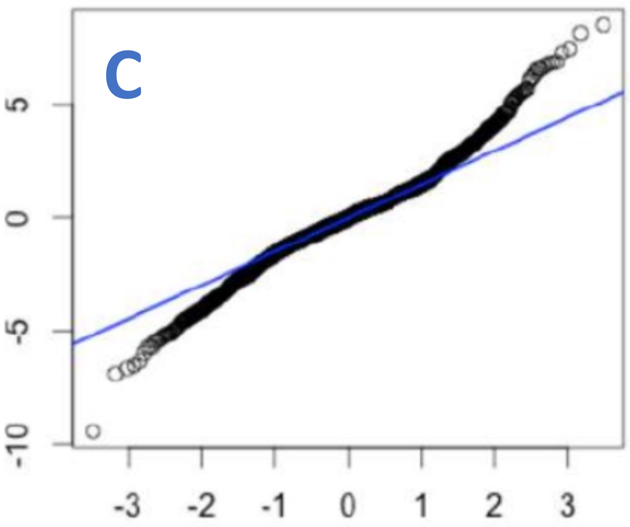
Conclusion?

**The line draws by QQ-Plot indicates the position that the points must have to follow a normal distribution**

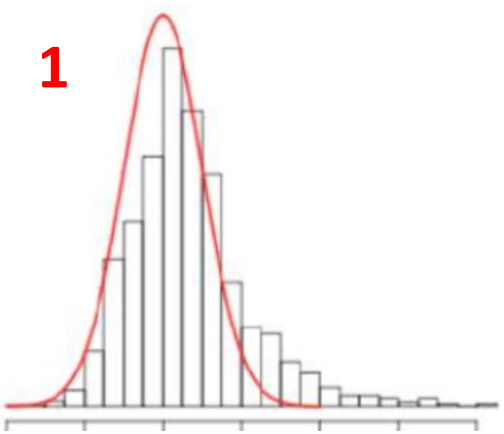
# What are the distributions (bottom) corresponding to these QQ-plots?

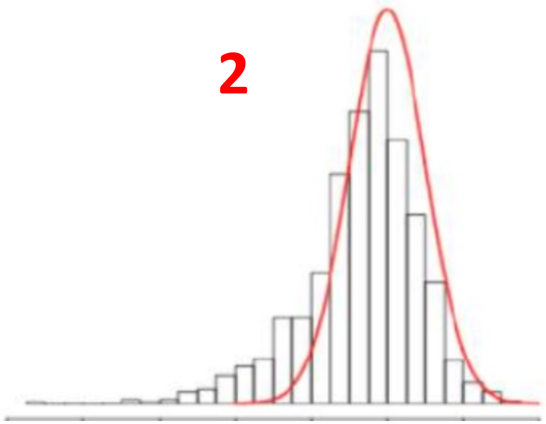


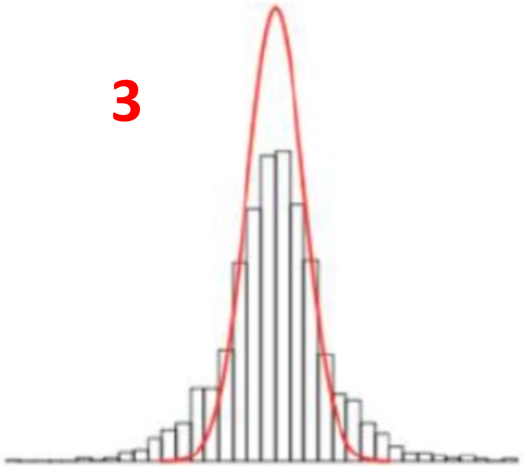



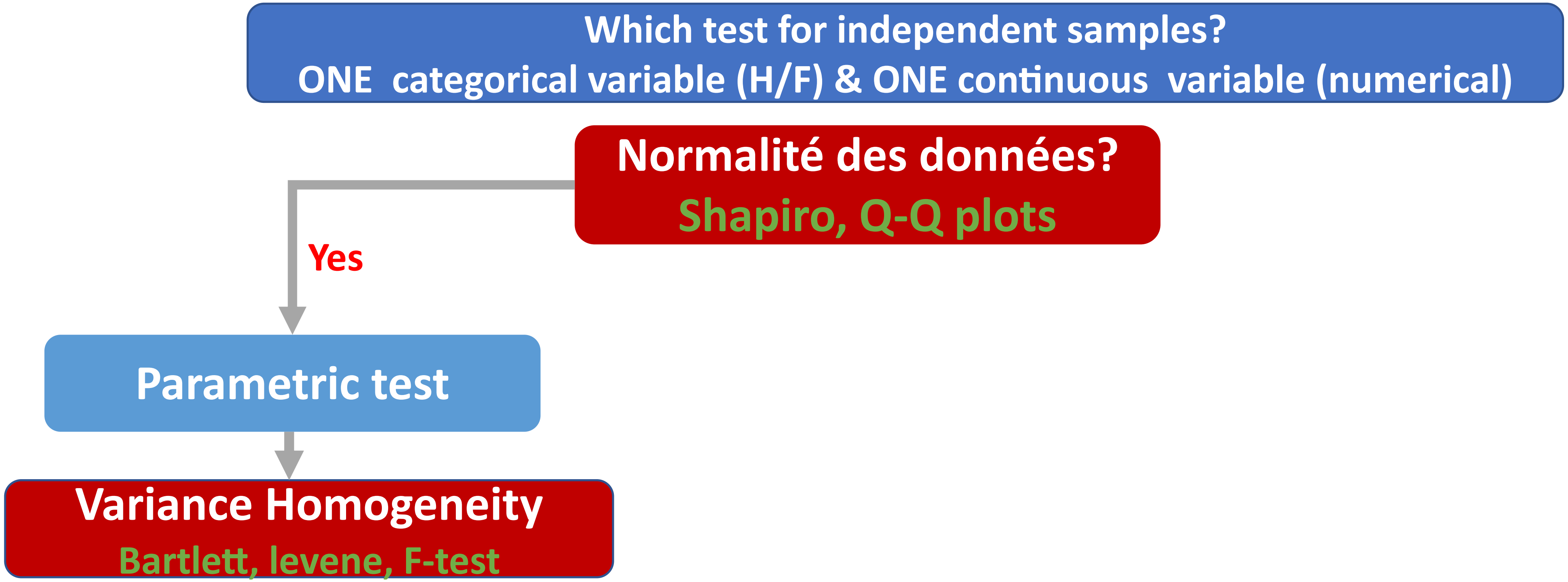




???????????



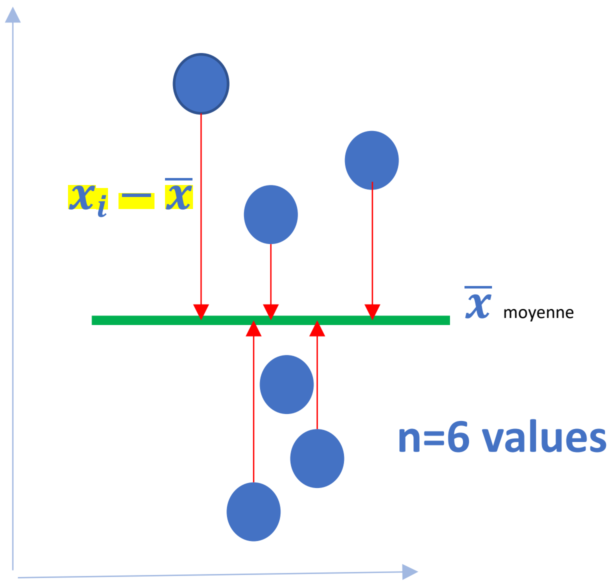

Which test for independent samples?
ONE categorical variable (H/F) & ONE continuous variable (numerical)
Normalité des données?
Shapiro, Q-Q plots
Yes
Parametric test
Variance Homogeneity
Bartlett, levene, F-test

# Variance= $S^2/\sigma^2$

- Variance measures the degree of dispersion of a data set around the mean
- Arithmetic mean of squared deviations from the mean! ☹

→ square unit

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




## Standard Deviation=S/ $\sigma$

$$S = \sqrt{S^2}$$

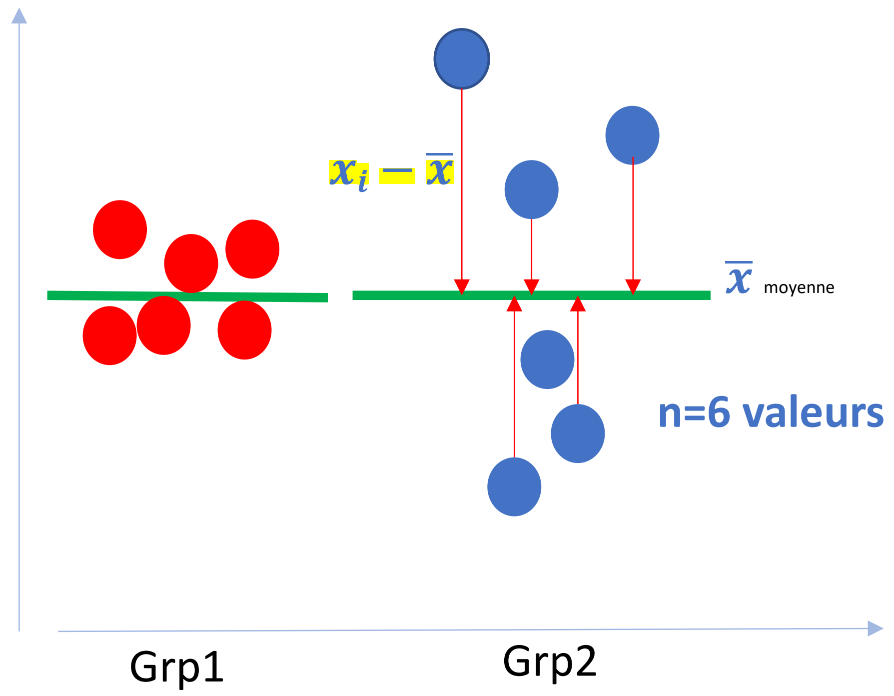
The advantage of the standard deviation : expressed in the same unit as the data series

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

SS will be greater in the sample….??

Results of test using variance :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| groupe | 3 | 13.03 | 4.343 | 0.211 | 0.887 |
| Residuals | 14 | 288.75 | 20.625 | | |



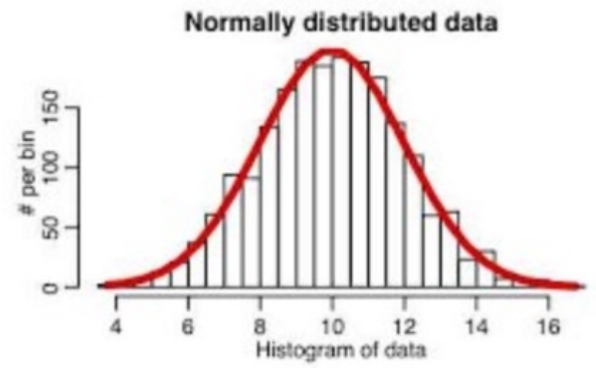

- ***Sum of Squares* (= SS, Sum Sq) in your results!**

→ Numerator of variance!!

- ***Mean Square (= Mean Sq= VARIANCE formula!!!)***

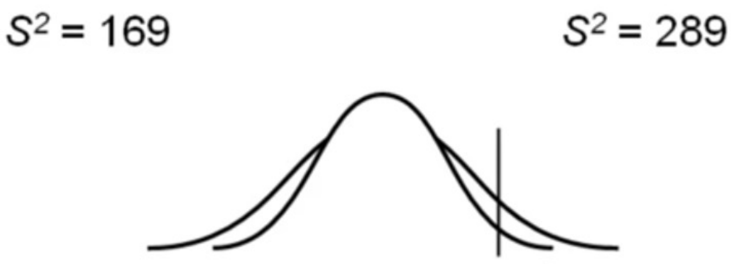
# Requirement for parametric test… check-list!

- Check normality of data:  Shapiro Test & QQ-plots!!
- Shapiro: H0 is «data follow normal distribution»




- Check variance Homogeneity : F-test (2 groups), Bartlett's & Levene's tests

H0: « No difference »

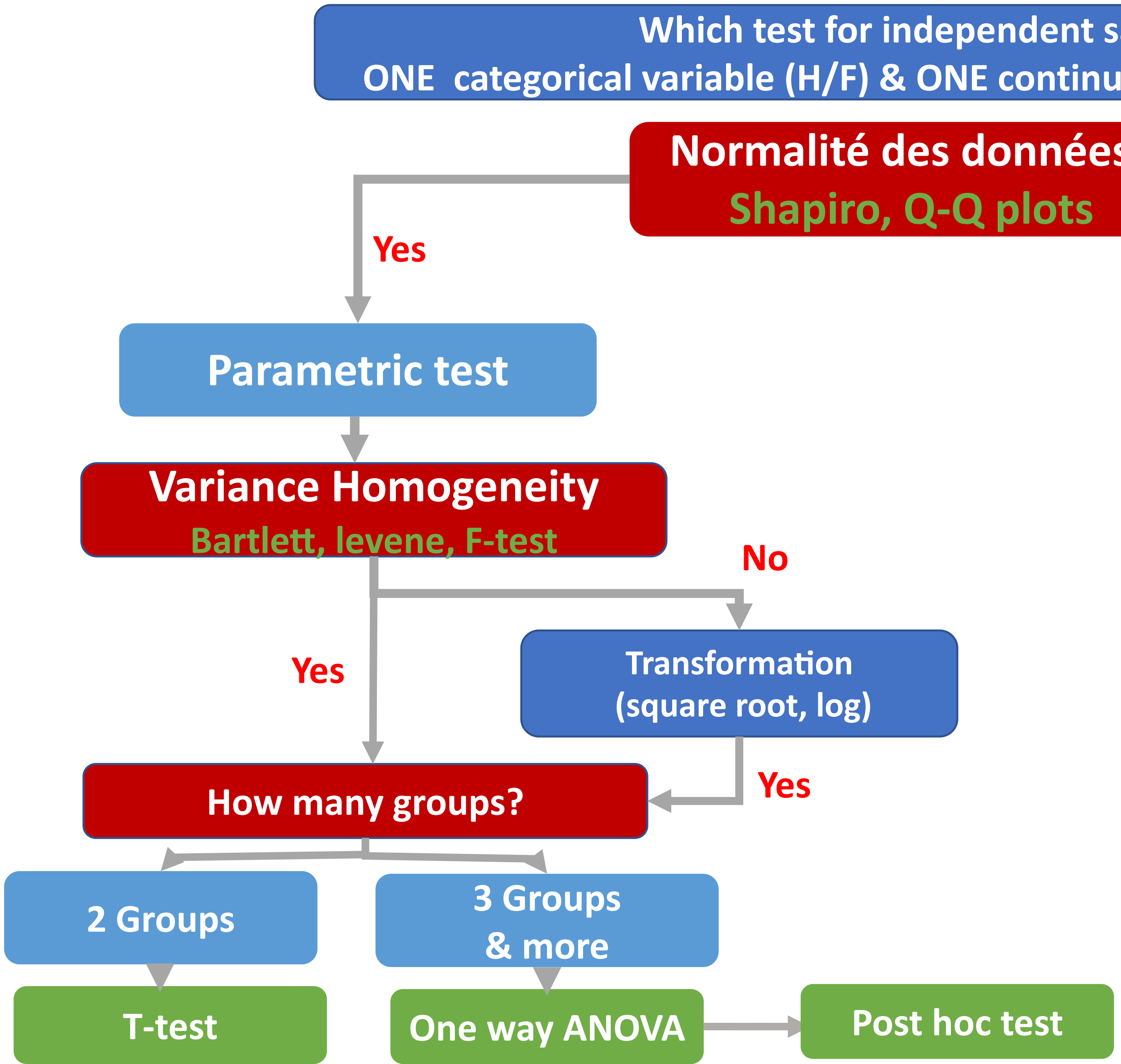


# Parametric Tests

Follow a known distribution (Normal distribution) → **Position** parameters
**Dispersion** parameters

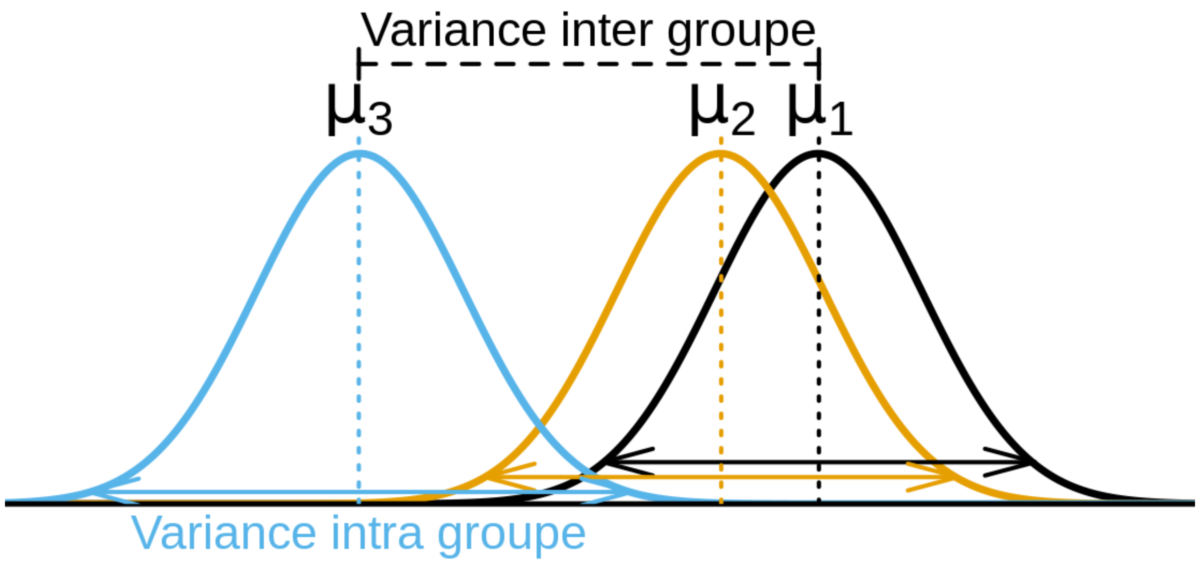
Conditions are required (variance homogeneity)

- **T-test (paired or unpaired):** Compare of the means from 2 sample groups for one variable
- **One way Anova** (variance analysis) : compare the means of three or more sample groups for one variable


Which test for independent samples?
ONE categorical variable (H/F) & ONE continuous variable (numerical)
Normalité des données?
Shapiro, Q-Q plots
Yes
Parametric test
Variance Homogeneity
Bartlett, levene, F-test
No
Yes
Transformation
(square root, log)
Yes
How many groups?
2 Groups
3 Groups
& more
T-test
One way ANOVA
Post hoc test

# ANOVA: ANalysis Of VAriance (One way Anova= Univariate)
# (3 groups at least)

- **Compare the variance of the group means to that within groups (i.e. intra-group variance) for a single explanatory variable (qualitative)**



## ANOVA: ANalysis Of VAriance (One way Anova= Univariate)

- ***Postulate*** = The **VARIATIONS** observed between the **MEANS** of the different groups (AT LEAST 3) are so small that they are easily explained by chance!!!

- **Evaluation : Compare the variance of the group means to that within groups (i.e. intra-group variance)**

- **ANOVA → variations through the Variance quantity**

**Variance inter-groupes** + **Variance intra-groupes**

**attribuable au facteur**

**attribuable à l'expérimentale**

**(fluctuation de l'échantillonnage, hasard)**

Factor effect!

- **Statistic F** = $\frac{Inter-group\ Variance}{Intra-group\ Variance}$

Chance /fluctuation

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| groupe | 3 | 13.03 | 4.343 | 0.211 | 0.887 |
| Residuals | 14 | 288.75 | 20.625 | | |

**Idea :**

**if the factor really has an effect, the part of the variations that can be attributed to it = Inter-group variance will be significantly higher than the part of the variations that cannot be attributed to it = Intra-group variance!**

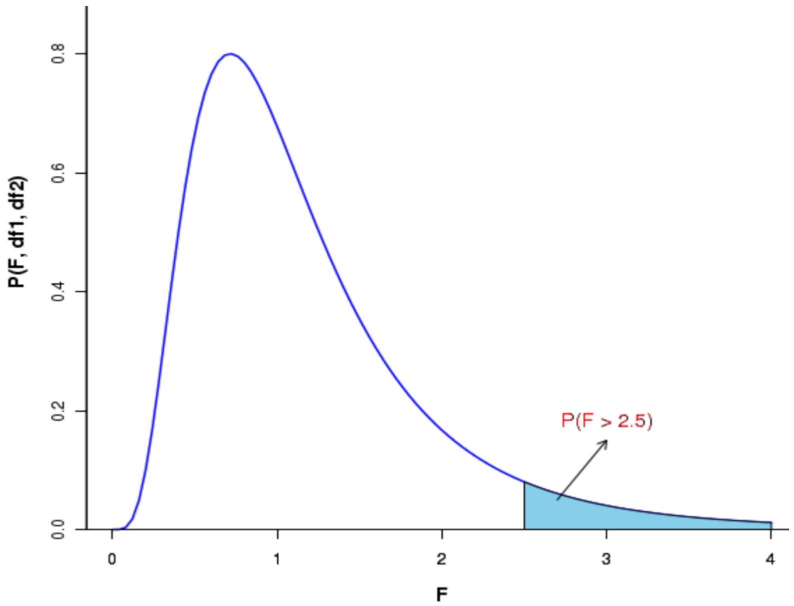
**Statistic F** Follows a so-called **Fisher-Snedecor** law:

= **Distribution F used for test of variances,** distribution of variances not being normal

- Relation of an observed value of F with the a priori probability of encountering such a value (> or =) by chance!
- → **probability given by the law = p-value!**

Denominator $S^2$ Numerator $S^2$ $S^2$

```
           Df  Sum Sq  Mean Sq  F value  Pr(>F)
groupe      3   13.03    4.343    0.211   0.887
Residuals  14  288.75   20.625
```




| variances | | ddl | F |
|---|---|---|---|
| entre k groupes | $v_k$ | k-1 | $v_k / v_r$ |
| résiduelle | $v_r$ | N - k | |

Degré de liberté

- **Two-ways ANOVA : Influences of TWO qualitative variables on ONE quantitative variable**

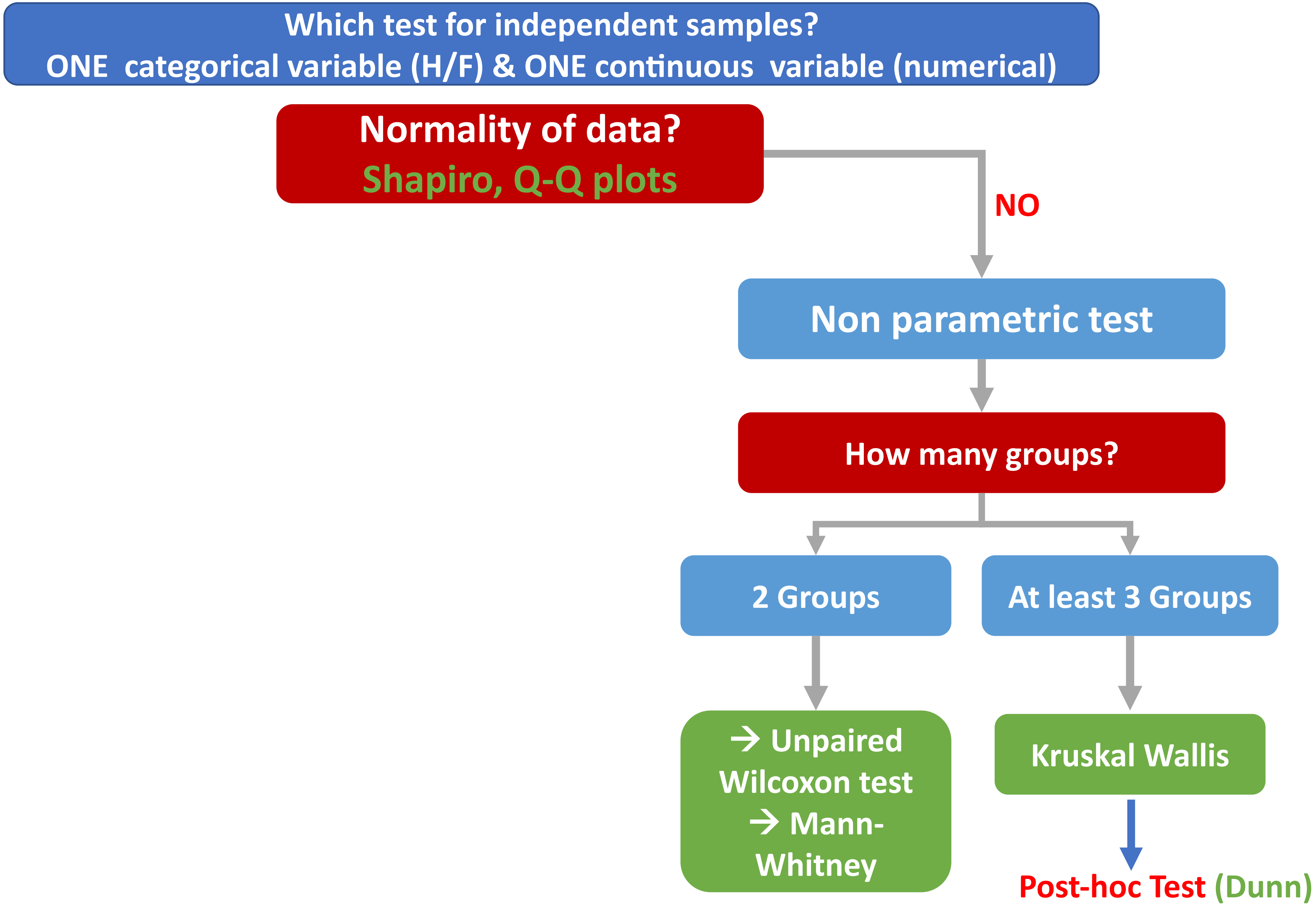
  **Exple: Influence of soil type and degree of humidity (ordinal variable) on plant yield**

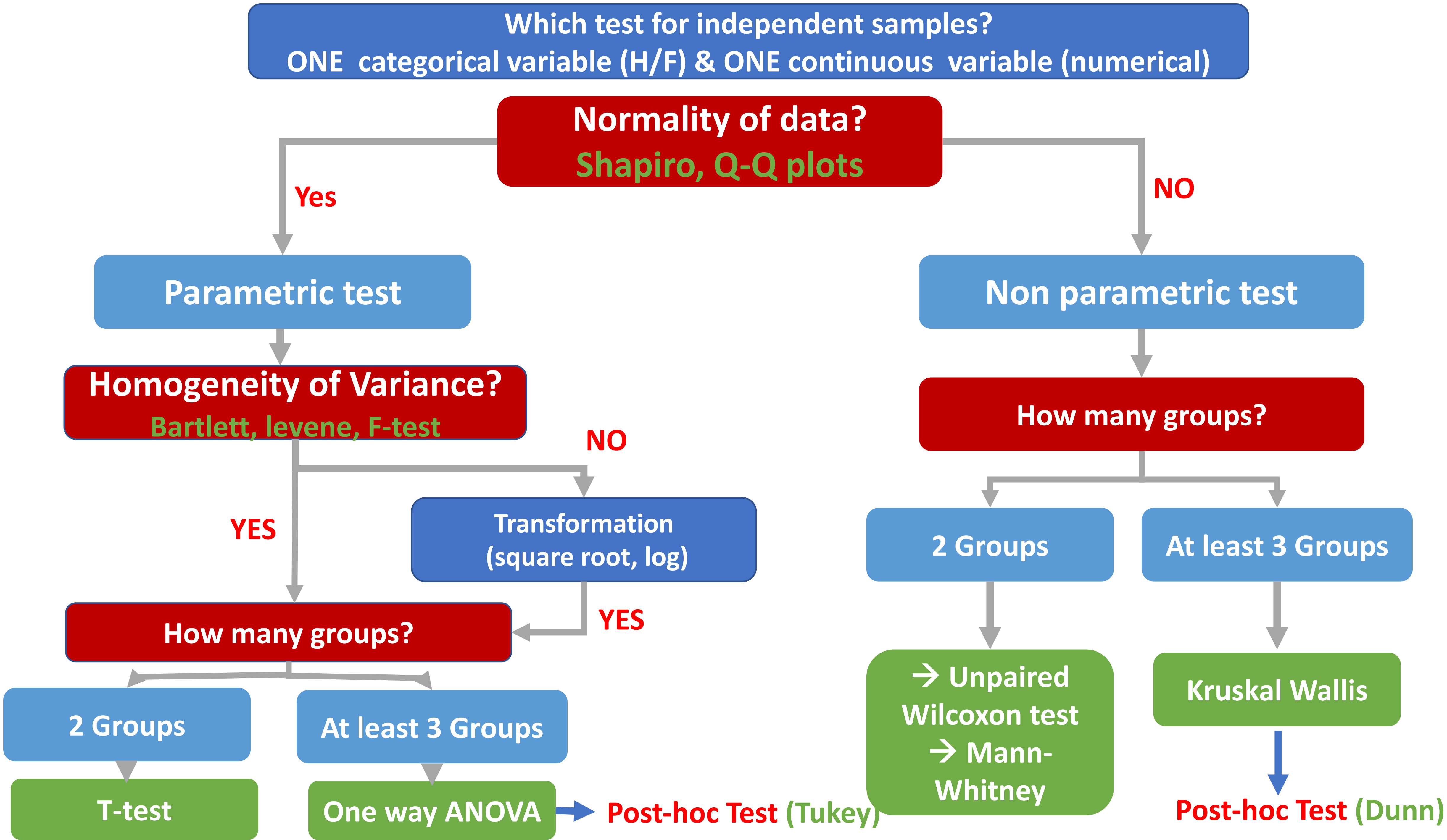
# Non-parametric tests

**No assumptions are made for the distribution of data:**
**Distribution-free tests, they are alternative to parametric tests**

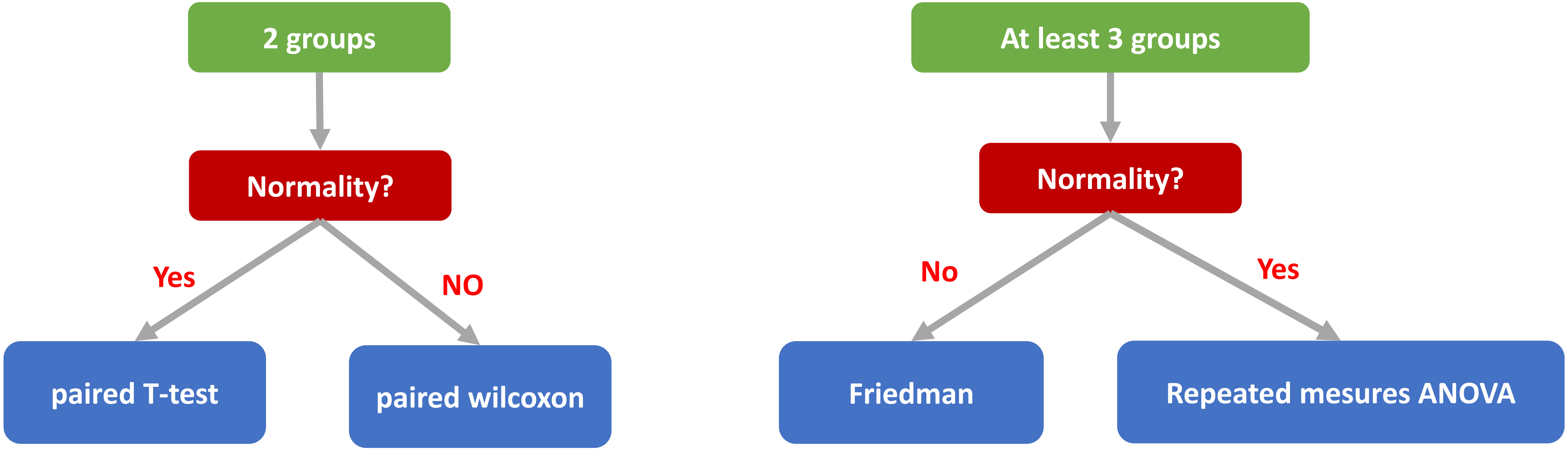
- **Wilcoxon Rank test** **: samples are paired/unpaired, 2 sample groups**
- **Mann-Withney test: Independent samples, 2 sample groups**
- **Kruskal wallis test : Independant samples, Three or more groups**

→ Based on the average ranks: we classify the values, we replace by a position (1,2 etc), Compares the average of the ranks between the groups


Which test for independent samples?
ONE categorical variable (H/F) & ONE continuous variable (numerical)
Normality of data?
Shapiro, Q-Q plots
NO
Non parametric test
How many groups?
2 Groups
At least 3 Groups
→ Unpaired Wilcoxon test
→ Mann-Whitney
Kruskal Wallis
Post-hoc Test (Dunn)


Which test for independent samples?
ONE categorical variable (H/F) & ONE continuous variable (numerical)
Normality of data?
Shapiro, Q-Q plots
Yes
NO
Parametric test
Non parametric test
Homogeneity of Variance?
Bartlett, levene, F-test
NO
YES
Transformation
(square root, log)
YES
How many groups?
2 Groups
At least 3 Groups
T-test
One way ANOVA
Post-hoc Test (Tukey)
How many groups?
2 Groups
At least 3 Groups
→ Unpaired Wilcoxon test
→ Mann-Whitney
Kruskal Wallis
Post-hoc Test (Dunn)

# Repeated measurements – paired samples
# Exple= time series, Before-After Treatment…



# Post-hoc Test

**Statistical tests with at least 3 groups!**
**After ANOVA, Kruskal-wallis**

**→ The result of an ANOVA test is an Overall p-value**

**Exple: You are comparing the effect of 3 soil types (A,B,C) on plant growth**
**ANOVA returns a p-value of 0.03**
**It does not tell you which pair of groups are significantly differents!!!!**

**→ Post-hoc Test! Multiple comparisons (eg: Gp A vs. Grp. B; GrpB vs. Grp C; Grp C vs. Grp A!)**

- Parametric Post-hoc test (ANOVA) → **Tukey Test**
- Non-parametric Post-hoc test (Krukal wallis) → **Dunn Test**

# Multiple Testing Issue: increasing the risk…

Test is based on probabilities, so there is always a risk of drawing the wrong conclusion!

→ **No hypothesis test is 100% reliable**


Performing hypothesis testing:

- You have two hypotheses :
- H0: Null hypothesis = the reference hypothesis : No difference
- H1: Alternative hypothesis: There is a difference

- You encounter: Type I error : $\alpha$ = Risk alpha

$\alpha$ = 0.05 Is the probability (significance threshold) to incorrectly reject H0!
In other words, an acceptable chance of a false positive!!

# Differential abundance : Multiple testing!!

**ONE TEST :** $P_{\text{False Positive}} = P_{\text{error}} = \underline{\alpha} = 0.05$

**Complementary Prob** $P_{\text{no\_error}} = 1 - \underline{\alpha} = 0.95$

**TWO TEST without making error :** $P_{\text{no\_error in two tests}} = (1 - \underline{\alpha}) * (1 - \underline{\alpha}) = (1 - \underline{\alpha})^2$

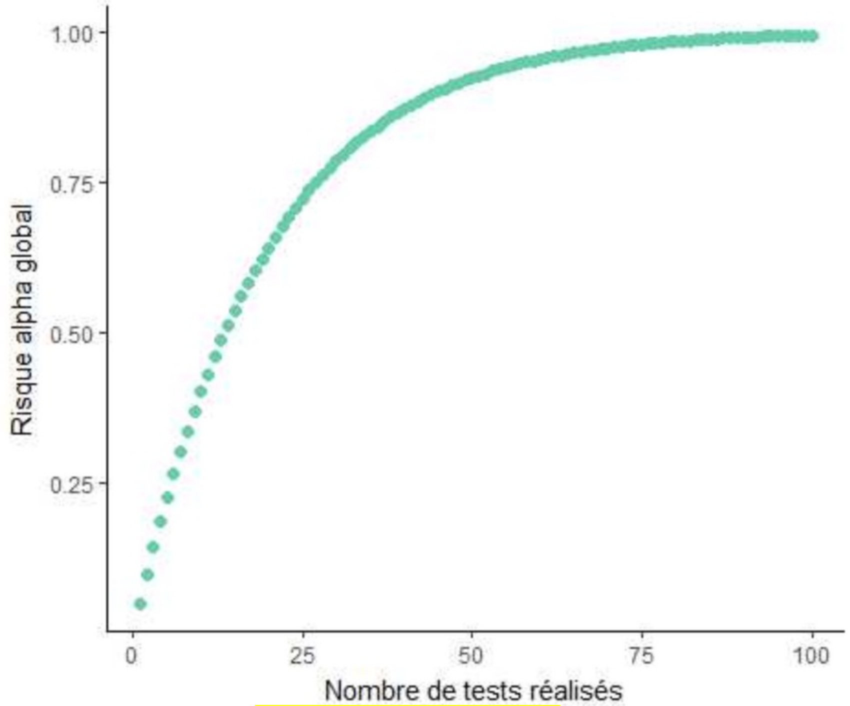
**Complementary Prob** $P_{\text{at\_least\_ONE\_error in two tests}} = 1 - (1 - \underline{\alpha})^2$

**Generalization to n TESTS** $P_{\text{at\_least\_ONE\_error in\_n\_tests}} = 1 - (1 - \underline{\alpha})^n$

**It's called the global $\underline{\alpha}$ risk**

# What does it means…

- You test ONE ASVs (n=1) for differential abundance: $1-(1-\alpha)^n = 1-(1-0.5)^1 = 0.05$
- You test 3 ASVs (n=3): $1-(1-0.05)^3 = 0.14$
- You test 100 ASVs (n=100): $1-(1-0.05)^{100} = 0.9941$




**The global risk α reach 0.9941=99.41%!!!!**

**→ 99% to wrongly reject the H0 at least One times**

**Need to ajusted this phenomen by using p-value adjusted!**

# FDR : False Discovery Rate : Benjamini-Hochker

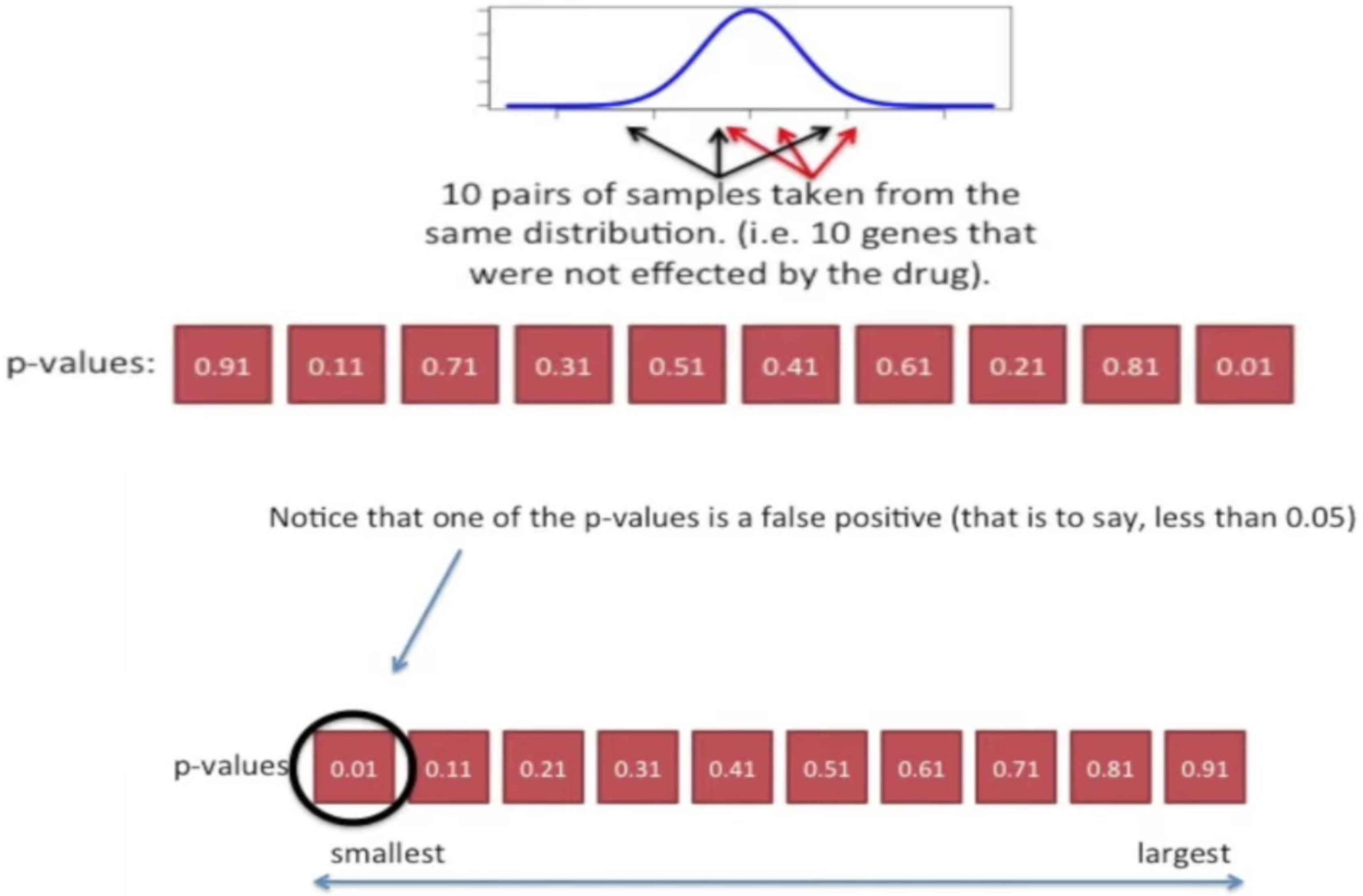
The idea : Discard bad data that looks good!!!

Benjamini-hocherk **adjusts p-values**
to limit the number of **false positives**
that are reported as **significant** (pvalue < 0.05)

**Adjusts p-values**
means that it makes them **larger**!

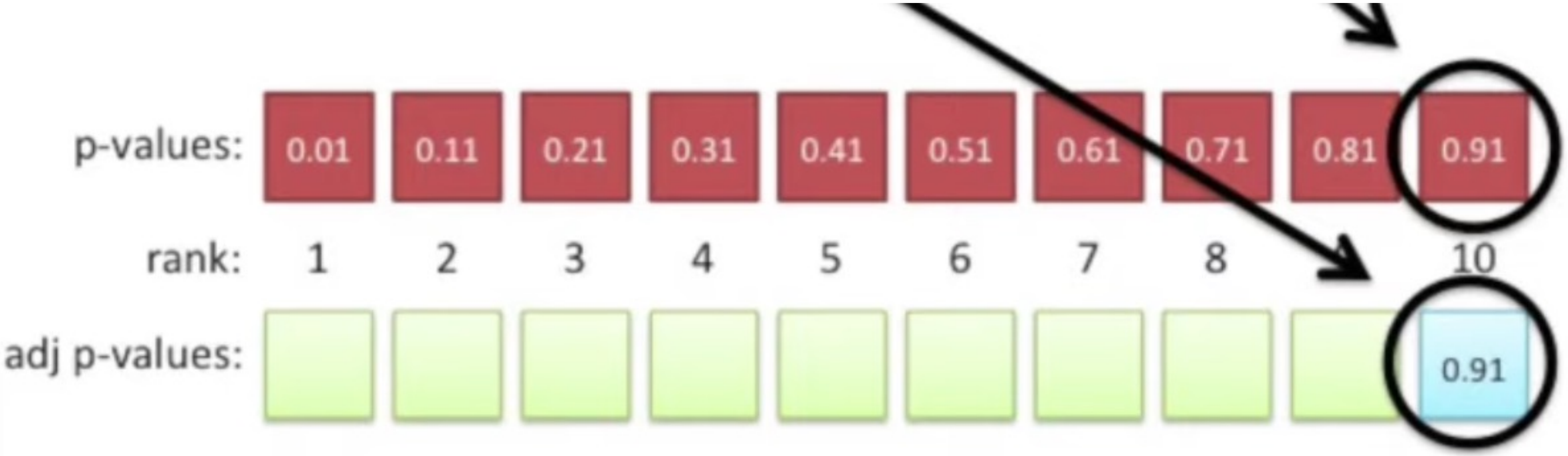
**Using FDR cutoff < 0.05**
means less than 5% of the significant results will be false positives

# Mathematical approach FDR-Benjamini-Hochker



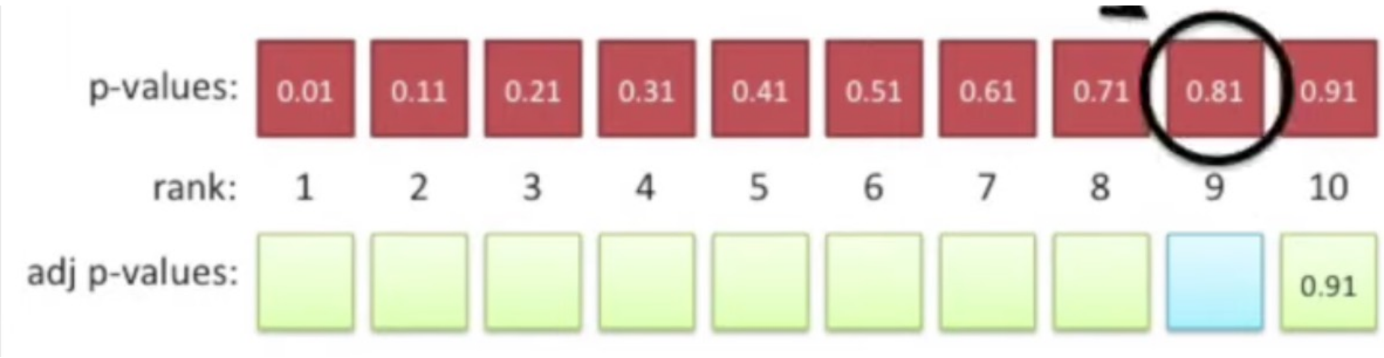

**1- Ranking pvalue**

# Prepare space for adjusted p-value


**2- Largest adjusted pvalue and larger pvalue are same**

## Next adjusted pvalue ….

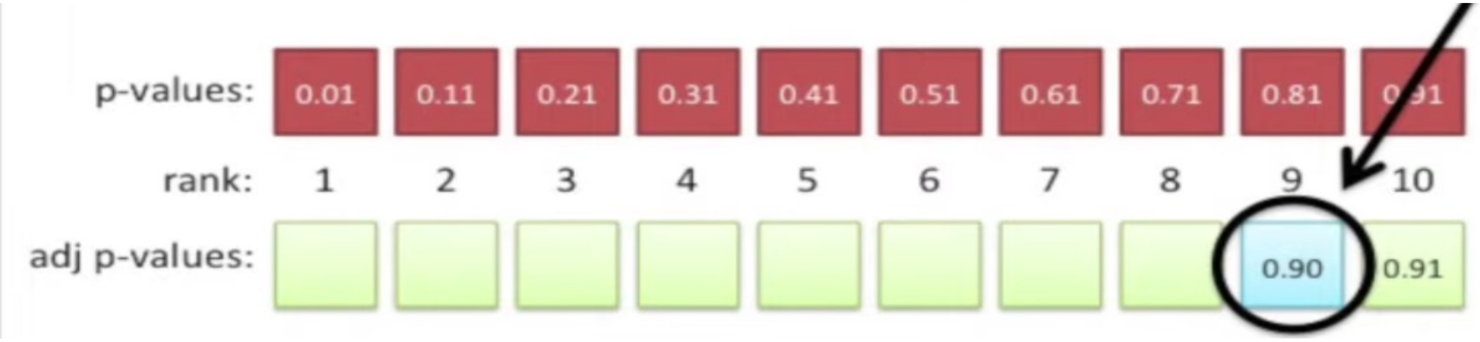



### The smallest of the two options

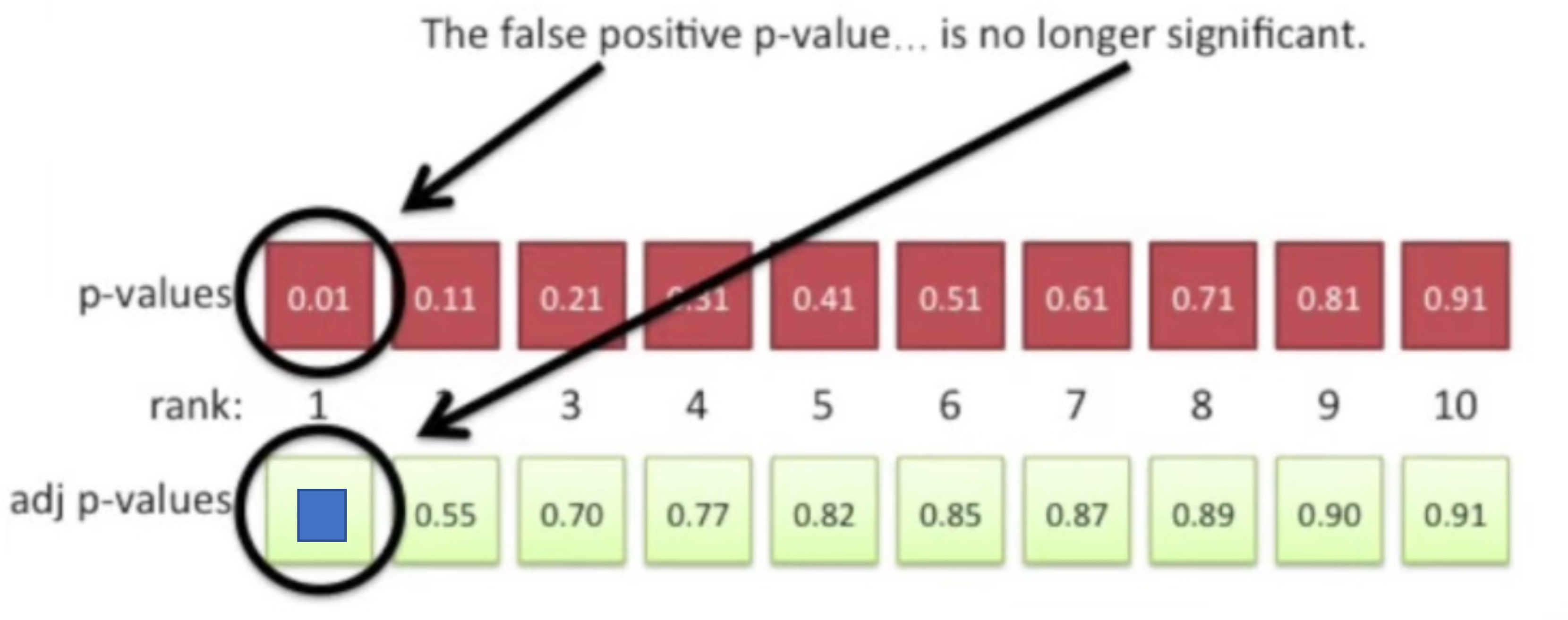
b: the current p-value $* \left( \frac{\text{total \# of p-values}}{\text{p-value rank}} \right)$

a: The previous adjusted p-value = 0.91

b: $0.81 * \left( \frac{10}{9} \right) = 0.90$



# Finally…


The false positive p-value… is no longer significant.
p-values
0.01
0.11
0.21
0.41
0.51
0.61
0.71
0.81
0.91
rank:
1
3
4
5
6
7
8
9
10
adj p-values
0.55
0.70
0.77
0.82
0.85
0.87
0.89
0.90
0.91

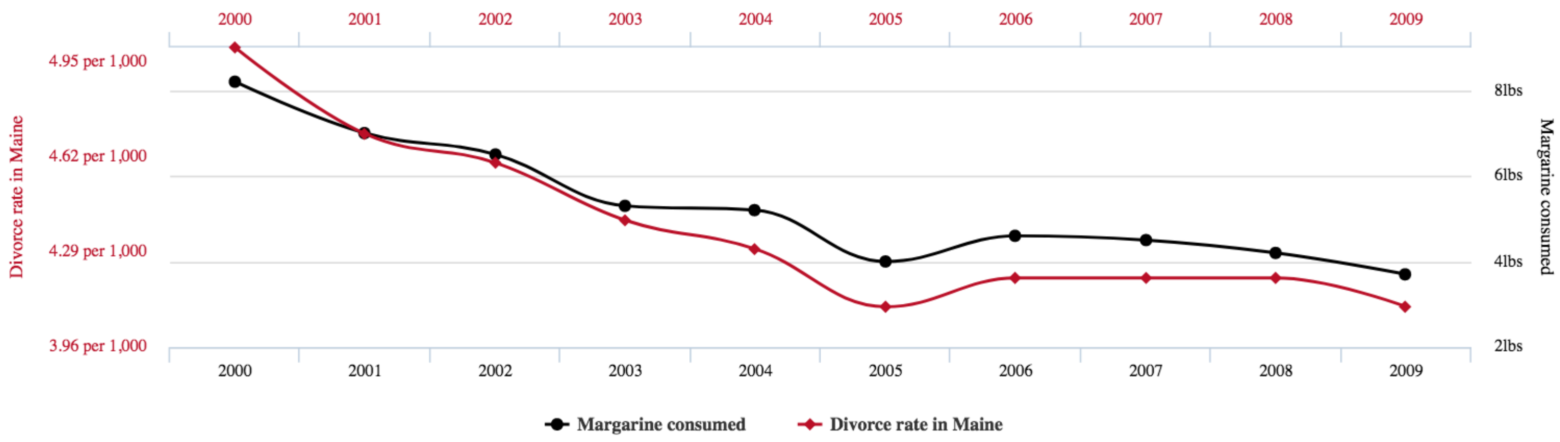
# Correlation & linear Regression as Bivariate Analyses

**Objective : Analyze the link that may exist between two variables (here: quantitatives)**
**(Two qualitative variables -> Khi2 test)**

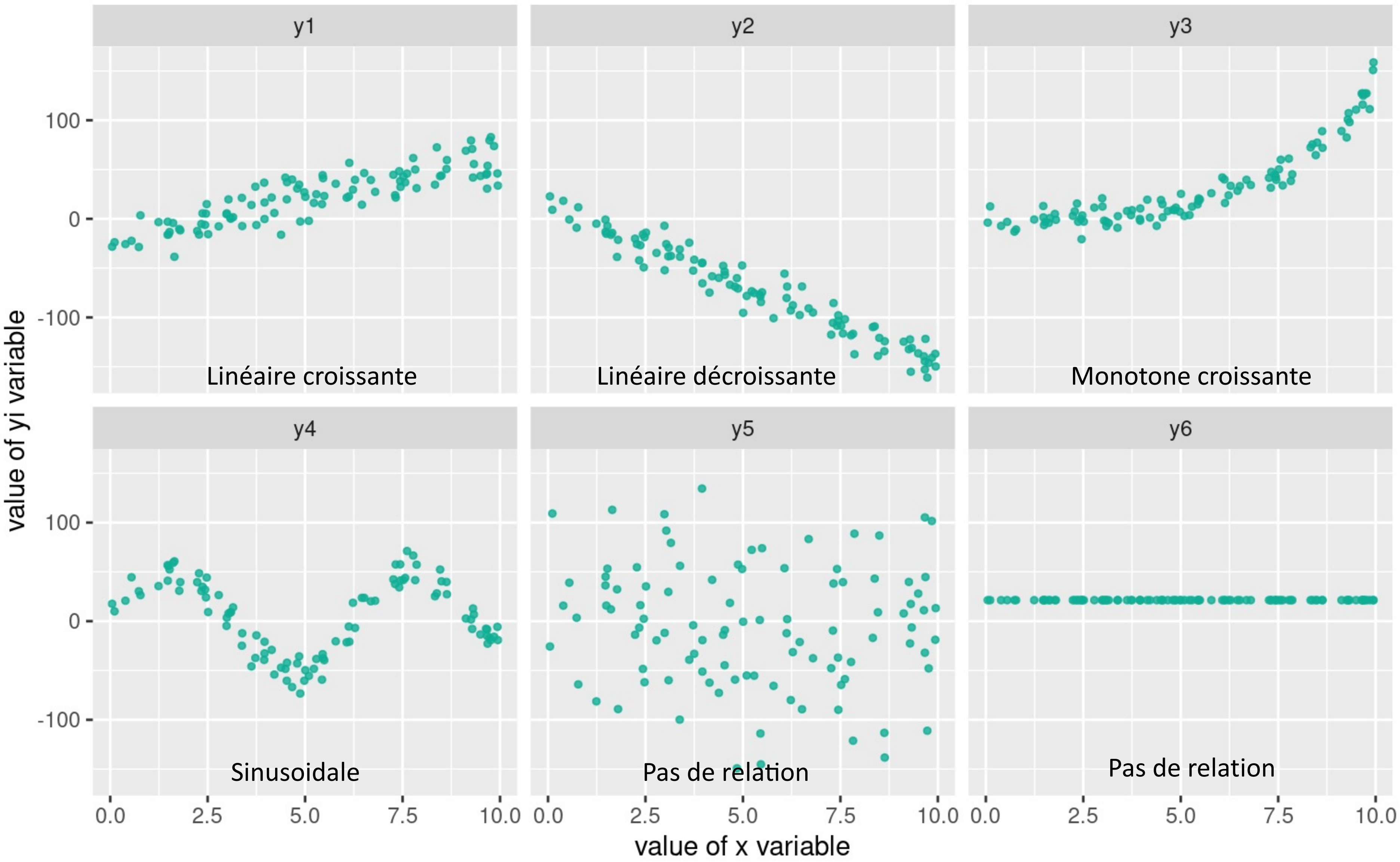
**Link/relationship/dependence** between the variables

→ The values of two variables **do not evolve independently** but on the contrary, present a certain form, a certain regularity

→ Intensity of the association does not indicate causality …



# What are the relationship between the variables in each graph?

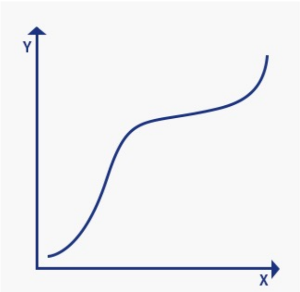


# About Correlation Coefficient r

## Intensity & Direction of the association between two variables

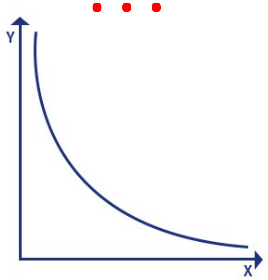
**Methods available:**

- **Strict Linear Relationship : Pearson** (**r**, parametric)
- **Monotonous relationship,** non-parametric : **Spearman** (**Rho**, rank-based)
  **Kendall** (**Tau**), Alternative to Spearman (small sampling)

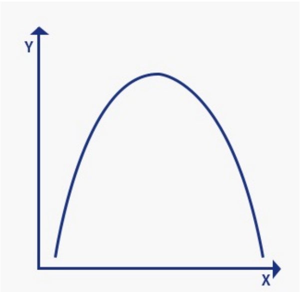



Monotone croissante

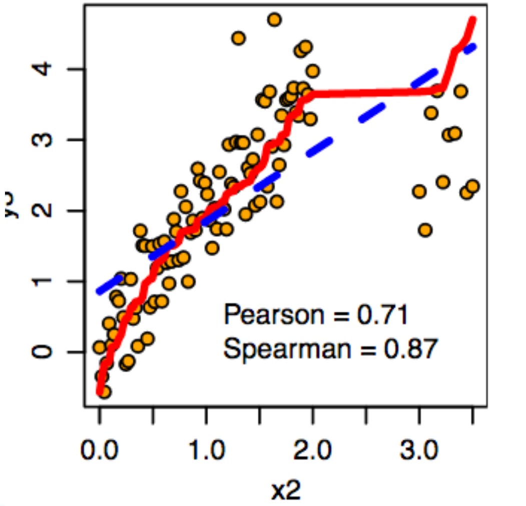
**???**




Monotone décroissante




point d'inflection : Unimodale

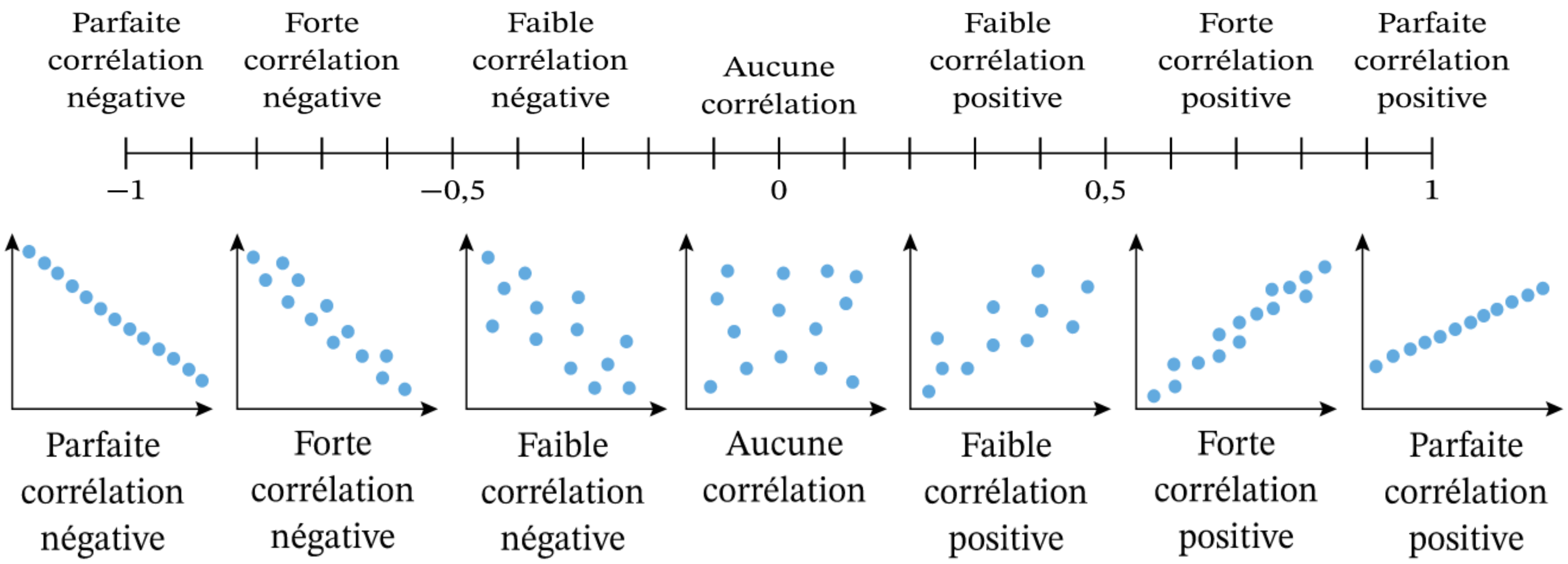



**Coefficient r range between -1 et 1**

- **Positive correlation** : The values of both variables tend to increase together
- **Negative correlation** : The values of one variable tend to increase and the values of the other variable decrease
- **Zero** : no **LINEAR** association (**Pearson**)

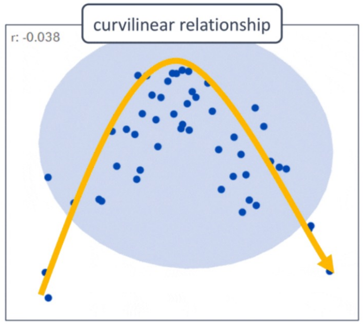
Distribution law of r under the H0 hypothesis (No statistical link between X and Y) → Access to p-values to check signifiance

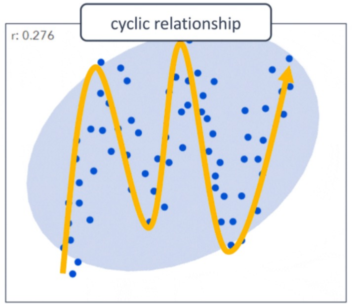
# For information!!!



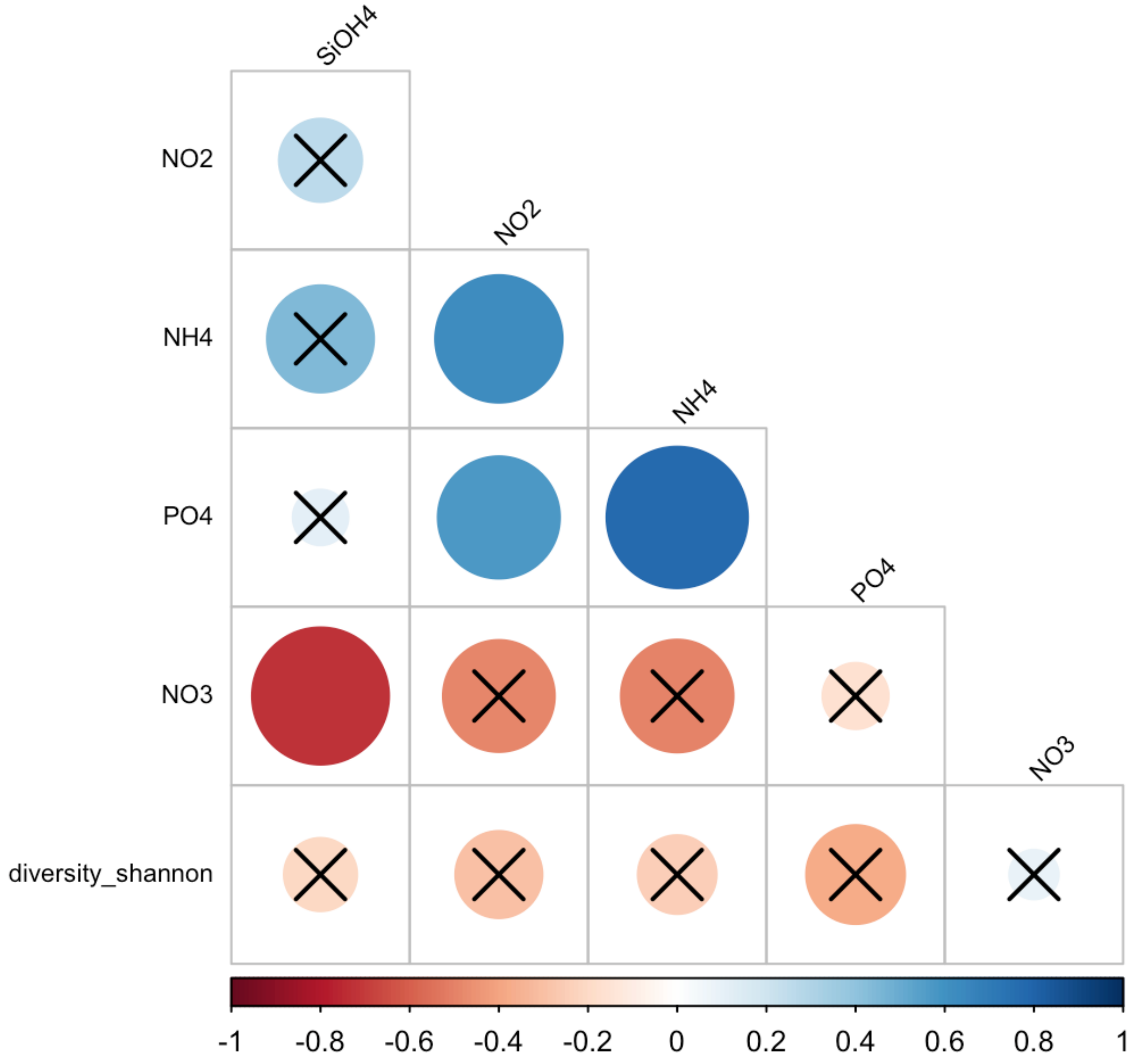

# Because inspecting your results graphically is never useless...

- r close to Zero: no association??



Example of bivariate correlation plot (with significant test)



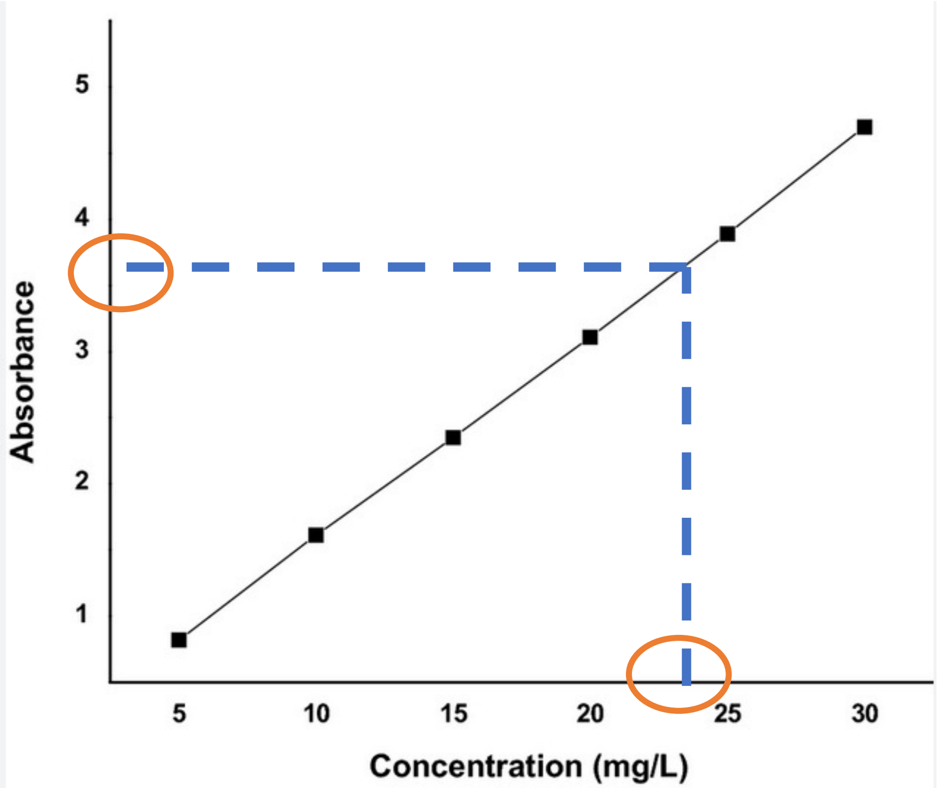
# About simple Linear Regression

- Only for quantitative variables

- Plot the scatter plot Is there a relationship?
- Is it linear?
- What orientation (positive, negative)?
- If the association is **linear** → Make a **regression**

**Requierement**
- Normal distribution
- Variance homogeneity

# Your favorite linear regression… calibration curve!!!

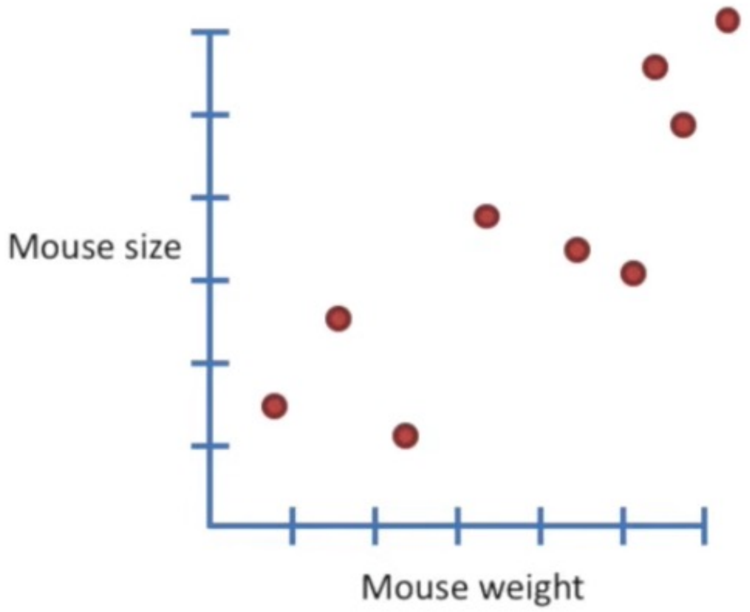



Explain and predict!

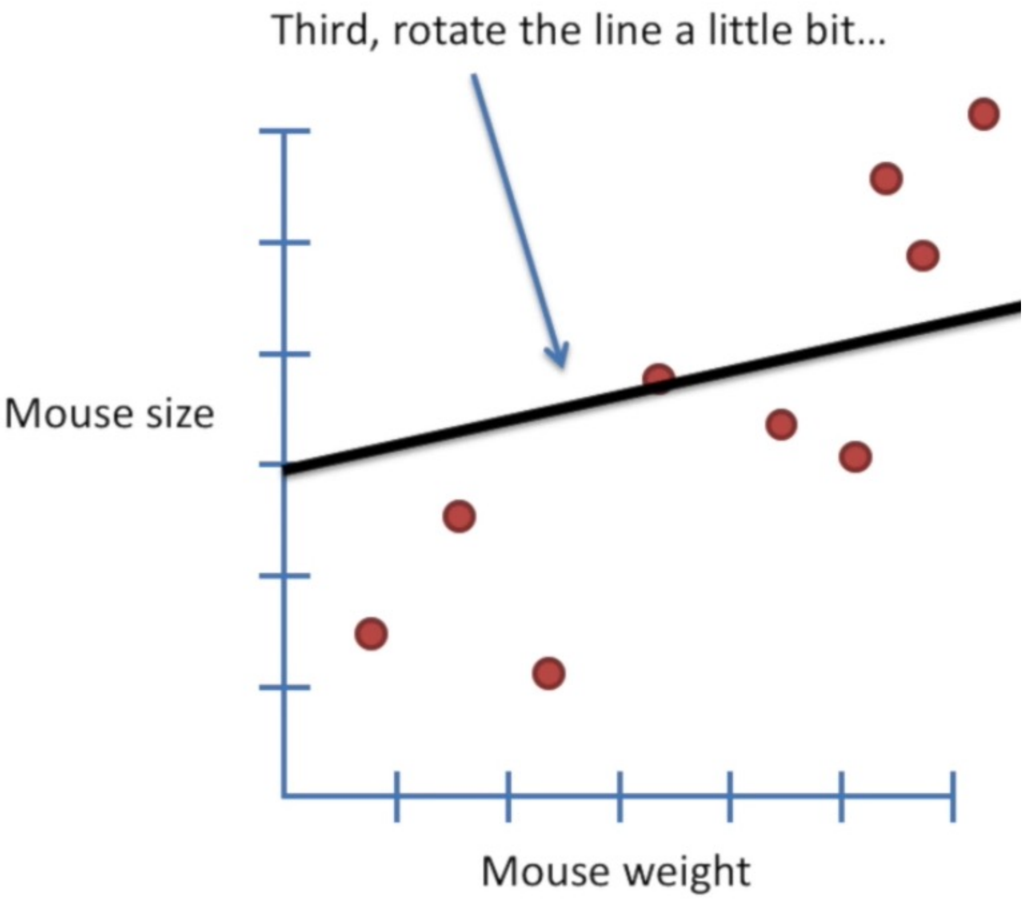
Models a linear type relationship (Y=aX+b)

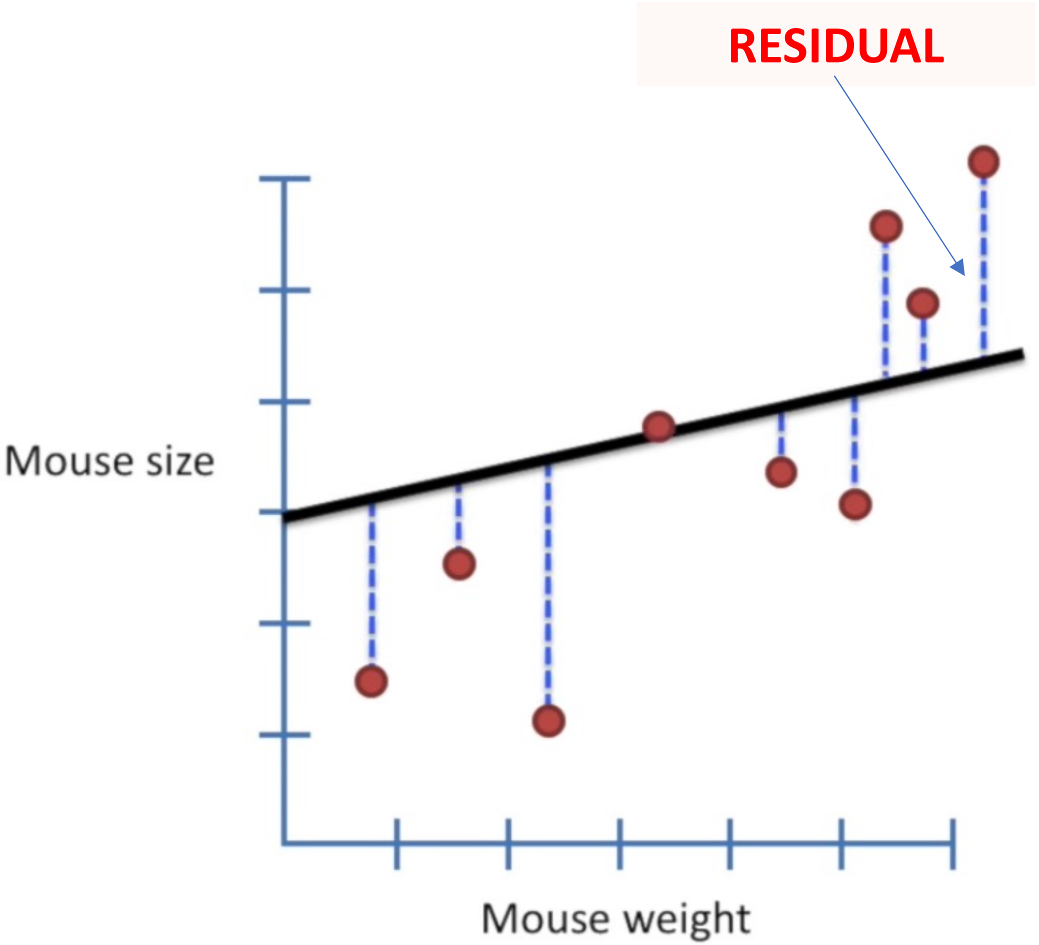
Model seeking to establish a linear relationship between a variable, called explained/dependent (Y), and another called explanatory/independent (X)

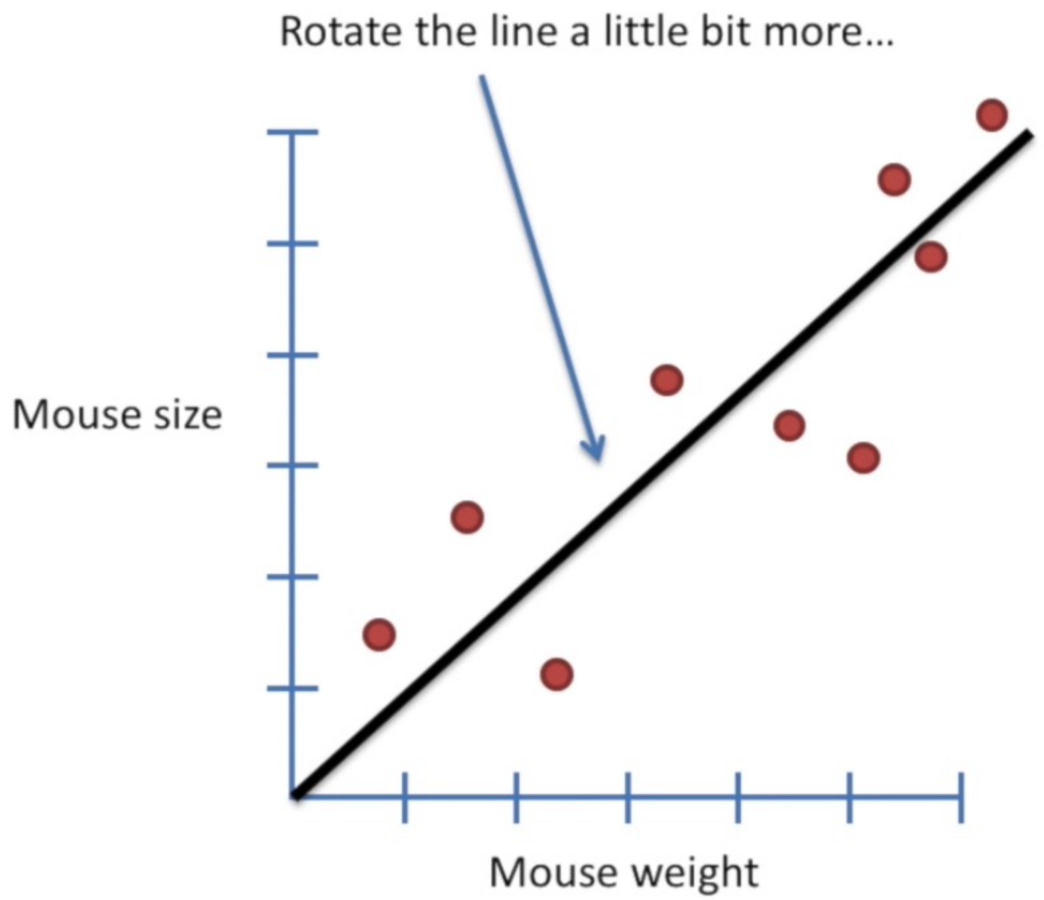
**Can mouse Weight predict Size correctly? ($R^2$)**
**Relationship is due to chance? (p-value)**

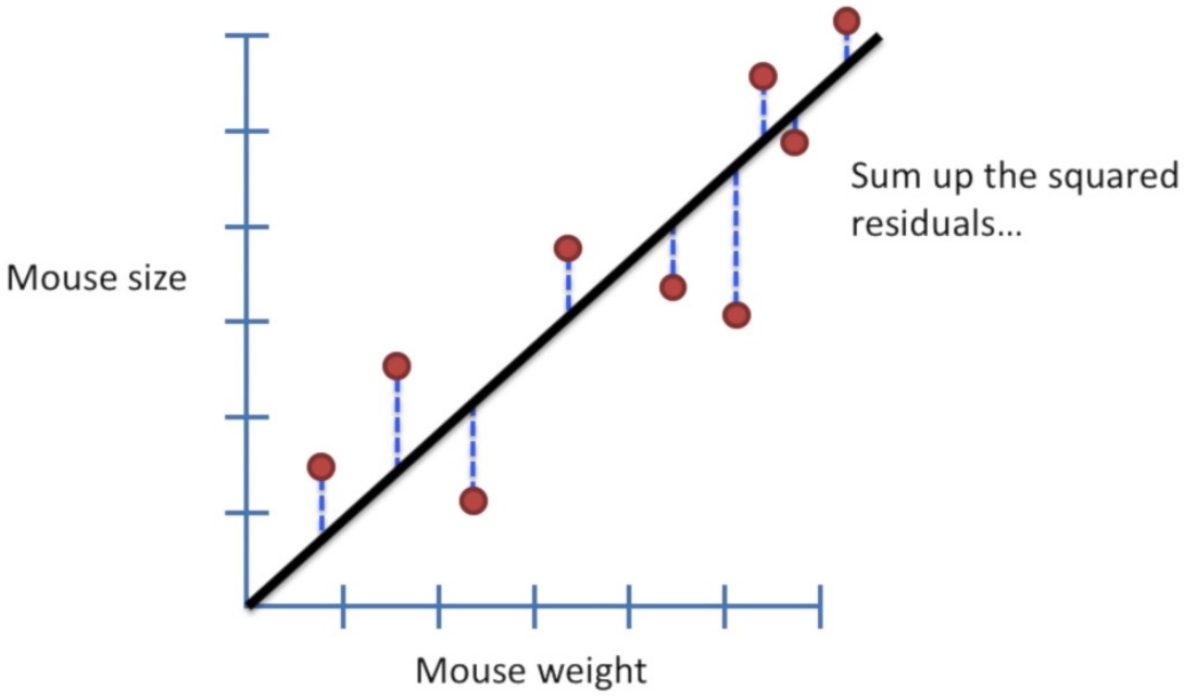


# Least square method



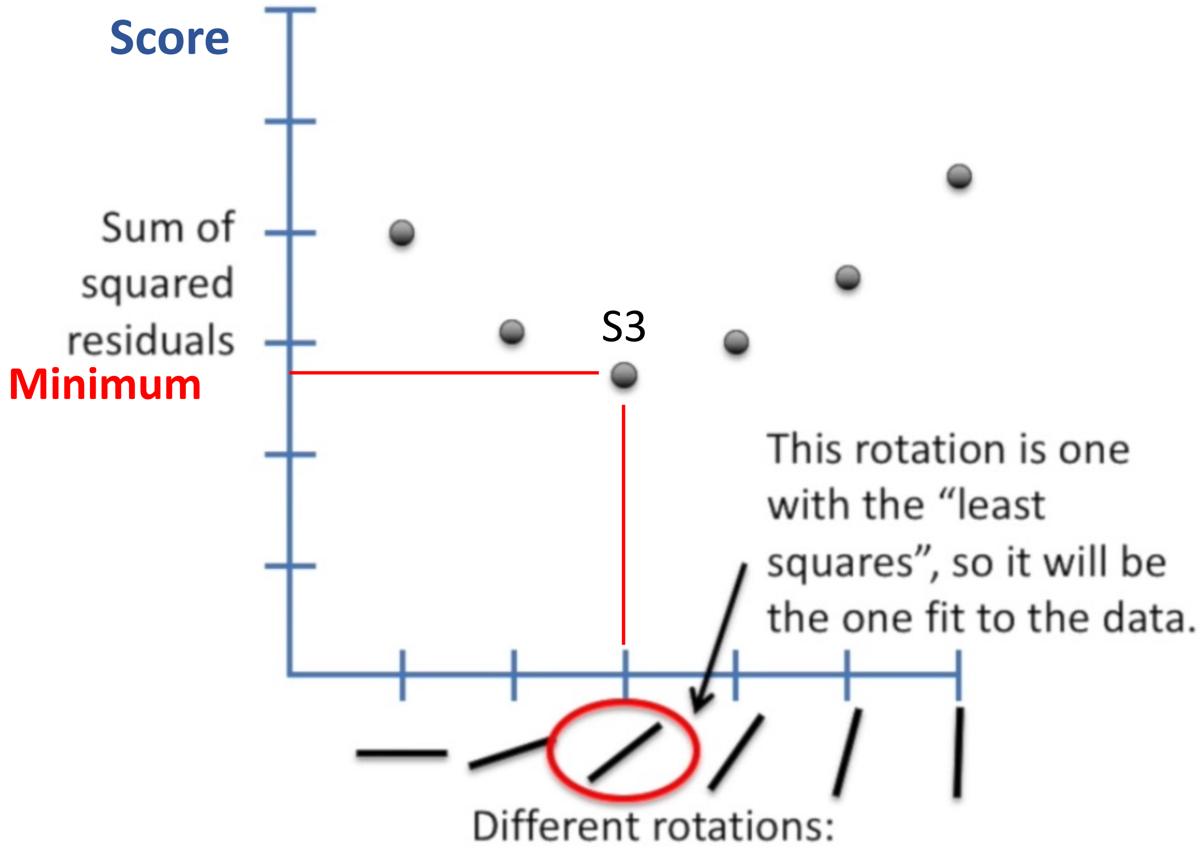





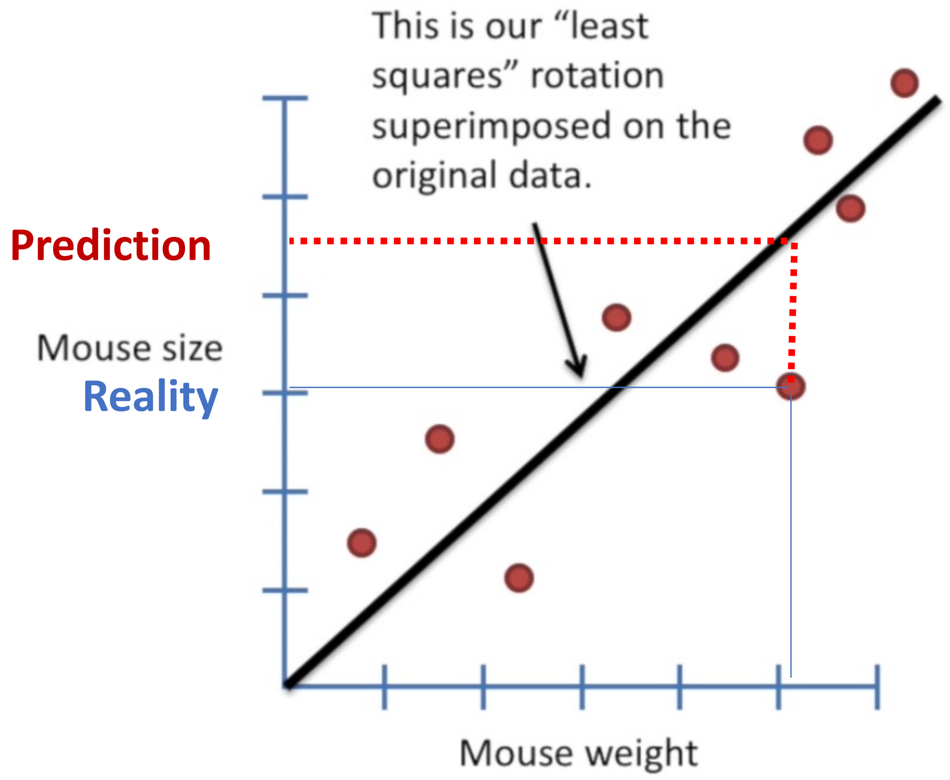




**Again & again, recalculate**

# Resume : Sums of squared residuals for each rotation




**Best rotation (=line position), the one which minimize the score of Sums of squared residuals !!!!**




$$y = 0.1 + \boxed{0.78x}$$

Dependence to « Mouse weight »

**Coefficient $R^2$ = prediction quality**

**how good is the linear regression model to predict Mouse size taking into account Mouse weight!!**

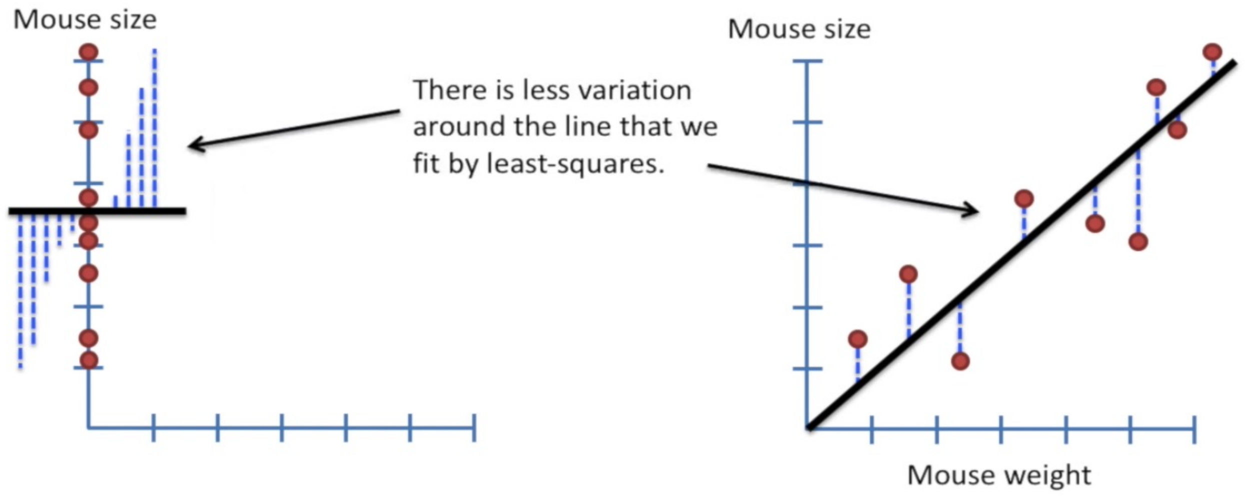
# $R^2$ : Determination Coefficient

$$R^2 = \frac{Var(mean) - Var(fit)}{Var(mean)}$$

$$R^2 = \frac{\textit{Variation expliquée}}{\textit{Variation totale}}$$

$$Var(mean) = \frac{SS(mean)}{n}$$

$Var(mean) = 11.1$




$$Var(fit) = \frac{(data - line)^2}{n}$$

$Var(fit) = 4.4$

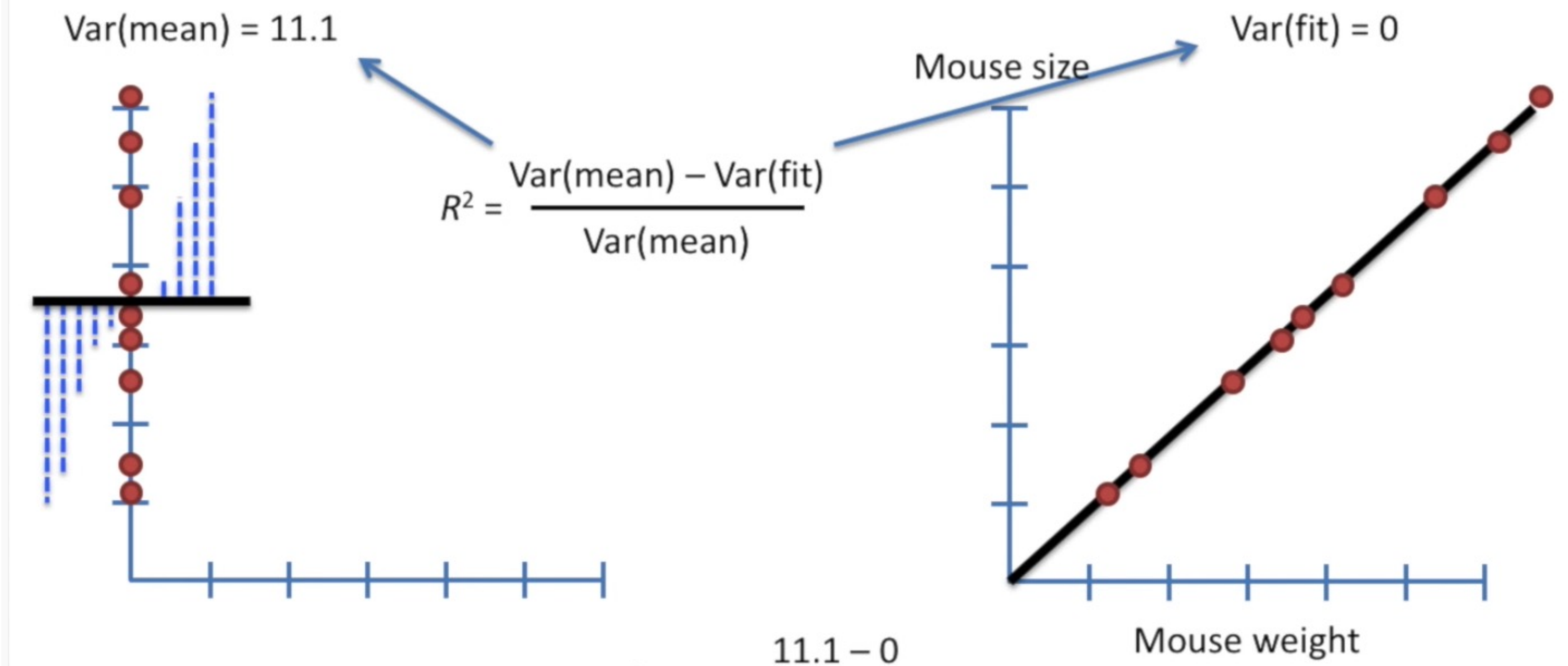
$$R^2 = \frac{11.1 - 4.4}{11.1} = 0.6 = 60\%$$

→ $R^2$ provides percentage variation in Y which is explained by all th X together **= % variation of the response variable explained by a linear model (weight variable)**

→ $R^2$ between 0 and 1

→ Here the established model explains 60% of the variability/variance of the "Mouse size »

# TO be sure …




$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

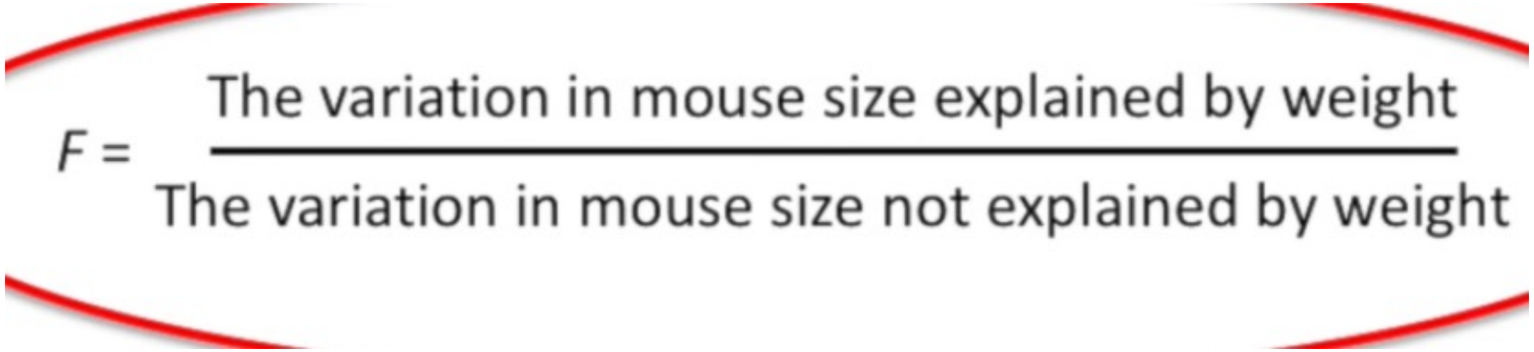
$$R^2 = \frac{11.1 - 0}{11.1}$$

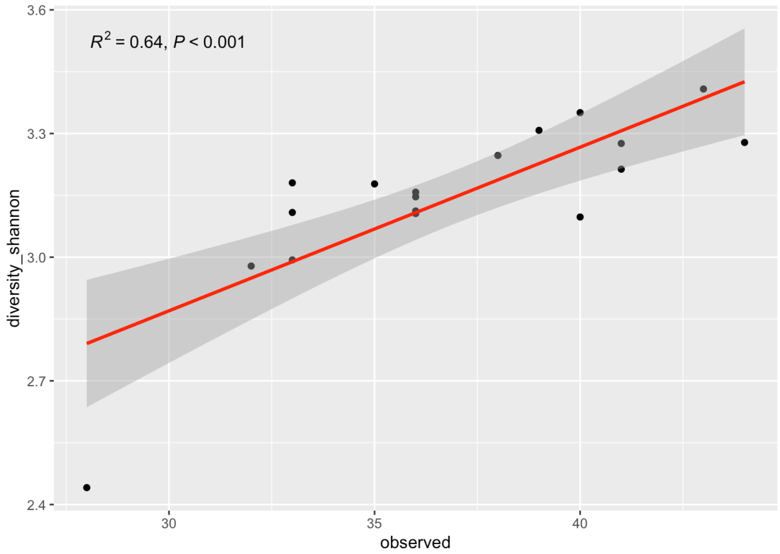
$$R^2 = 1 = 100\%$$

**$R^2$ = 1 = max strength of linear regression model**
**(Lesser $R^2$ the value, the more scattered are the dara points)**

# $R^2$ & significance?

**$R^2$ based on variance comparison … so p-value is given by the ratio F & distribution F**



# Relation between r & $R^2$

**Correlation coefficent of Pearson r** can be linked to linear regression $R^2$
**Its square is the explained variance by the regression** (**$R^2$**)

r =0.5 -> $R^2$ = 0.25 -> 25% of the Y variance explained by X variable… ☹