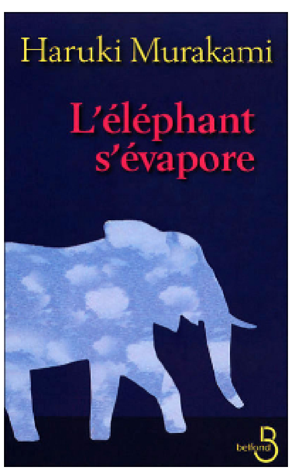
Statistique inférentielle

1) Notions de probabilités

2) Tests statistiques



J'ai 26 ans, je travaille dans le département du contrôle des marchandises […]. Il serait impossible de les contrôler soigneusement une à une […]. Par conséquent, on se borne à tirer sur quelques boucles de chaussures, à grignoter quelques gâteaux à titre d'échantillon.

Le communiqué du kangourou
nouvelle tirée du recueil *L'éléphant s'évapore*
Haruki Murakami

Points-clés

- Statistique inférentielle
- Test statistique
  - Risques d'erreur
  - P-value
  - En pratique (un petit quizz!)
- On n'oublie pas les représentations graphiques

Statistique inférentielle

- Tirer des conclusions à l'échelle d'une **population** à partir d'informations recueillies sur un **échantillon**.
- Sondage, recensement, échantillon représentatif...
- Lorsque l'on avance des informations quantitatives à l'échelle de la population, on ne parle plus de mesure mais d'**estimation**.
- Les mesures effectuées sur l'échantillon sont des **observations** de la variable aléatoire traduisant le phénomène à l'échelle de la population.

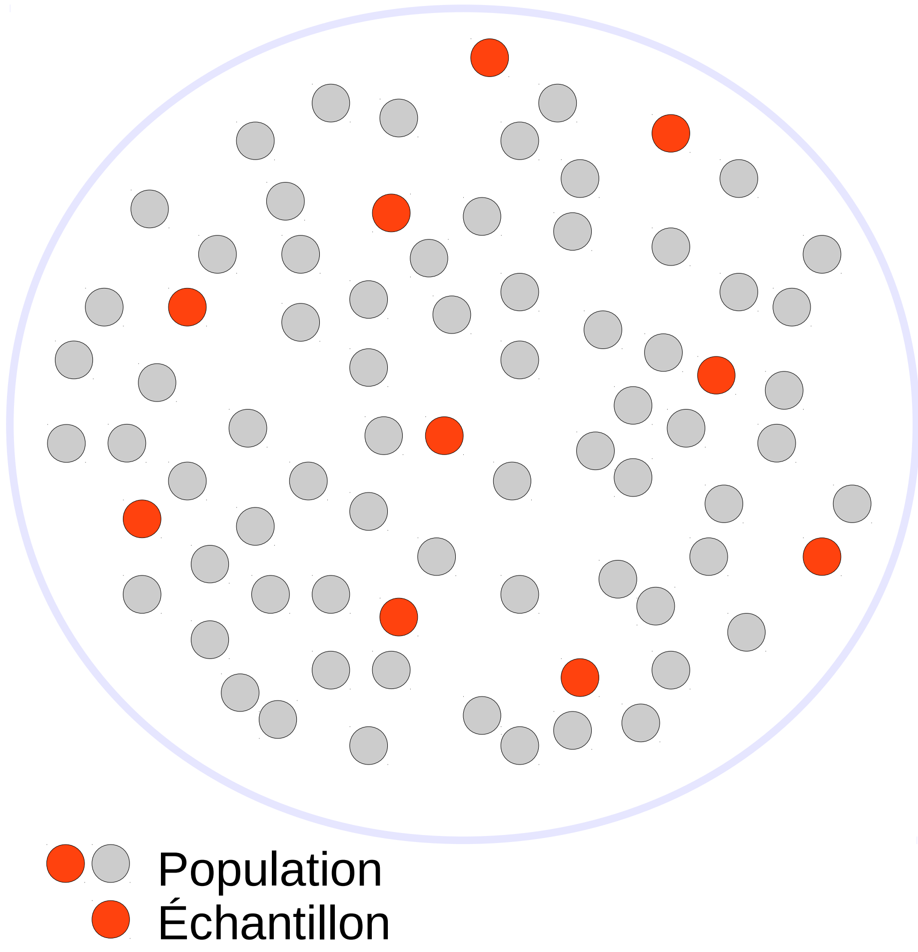
Statistique inférentielle

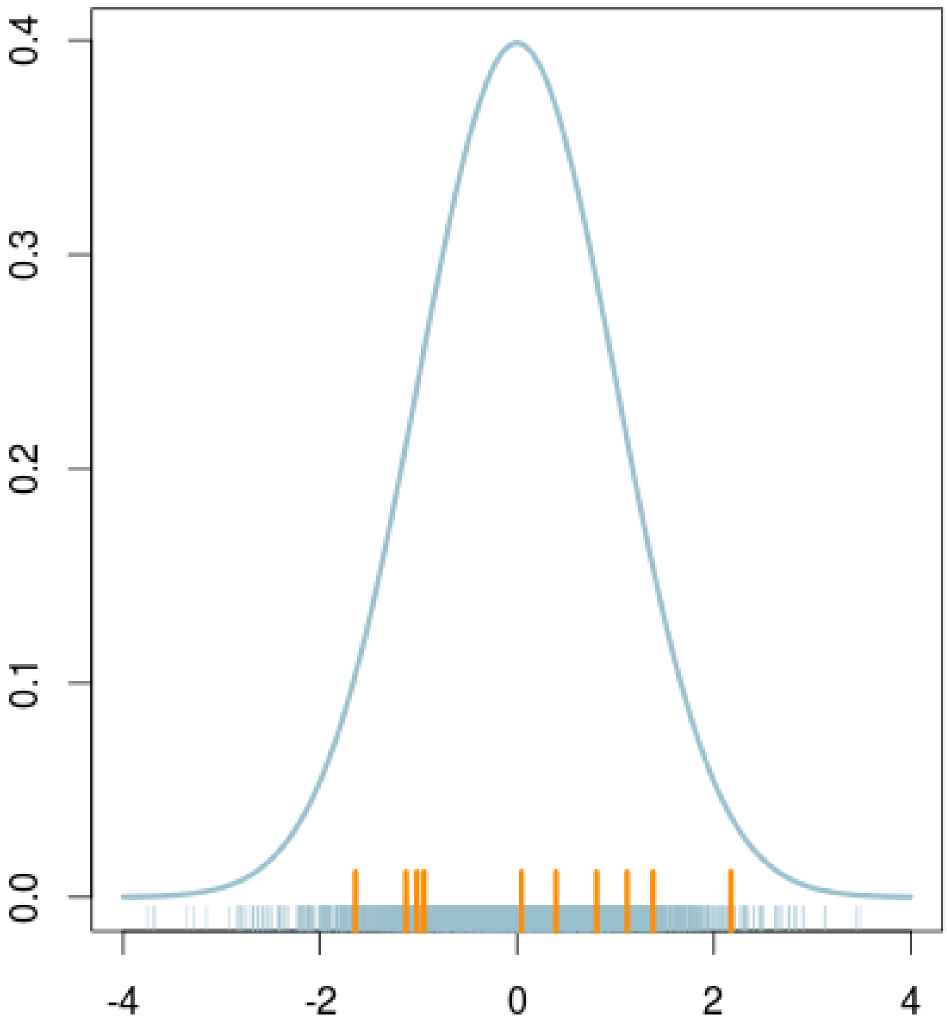
Dans un fabrique de pain d'épice, le procédé mis en œuvre pour vérifier l’aspect moelleux du produit fini consiste à plier une tranche et à mesurer l’angle d’inclinaison nécessaire pour la casser (un tel test est dit destructif). La règle étant qu’un bon pain d'épice doit avoir un angle de rupture de 50° (valeur fictive) : si l'angle est inférieur, le pain est trop sec, s'il est supérieur, le pain est trop moelleux. Tout lot doit être validé avant d’être commercialisé.

Il va de soi qu’une tranche cassée n’est pas commercialisable ainsi qu’un pain n’étant pas convenablement moelleux (angle de rupture $\neq$ 50°).

Dans de telles conditions, **il est impossible de tester l’ensemble des produits** (test destructif). Il est donc nécessaire d’effectuer les mesures sur un **échantillon représentatif** de la population (éviter par exemple de prendre les n premiers ou les n derniers pains fabriqués dans une journée ou, sur une même ligne de production si plusieurs fonctionnent en parallèle). L’angle moyen de rupture calculé sur l’échantillon est un **estimateur** de cet angle chez les pains d'épice du même lot (aux conditions de fabrication analogues).

Statistique inférentielle





• **Intervalle de confiance** : calculer, à partir de l'échantillon, un intervalle dans lequel la moyenne (par exemple) de la population devrait se trouver.

• **Test statistique** : ce qui est observé sur un échantillon permet-il d'invalider une hypothèse faite sur la population ?

D'où sort la courbe en cloche ?

Par exemple d'une planche avec des clous et des billes !

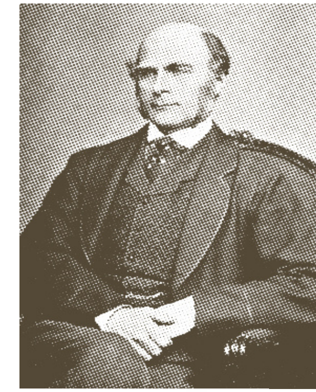
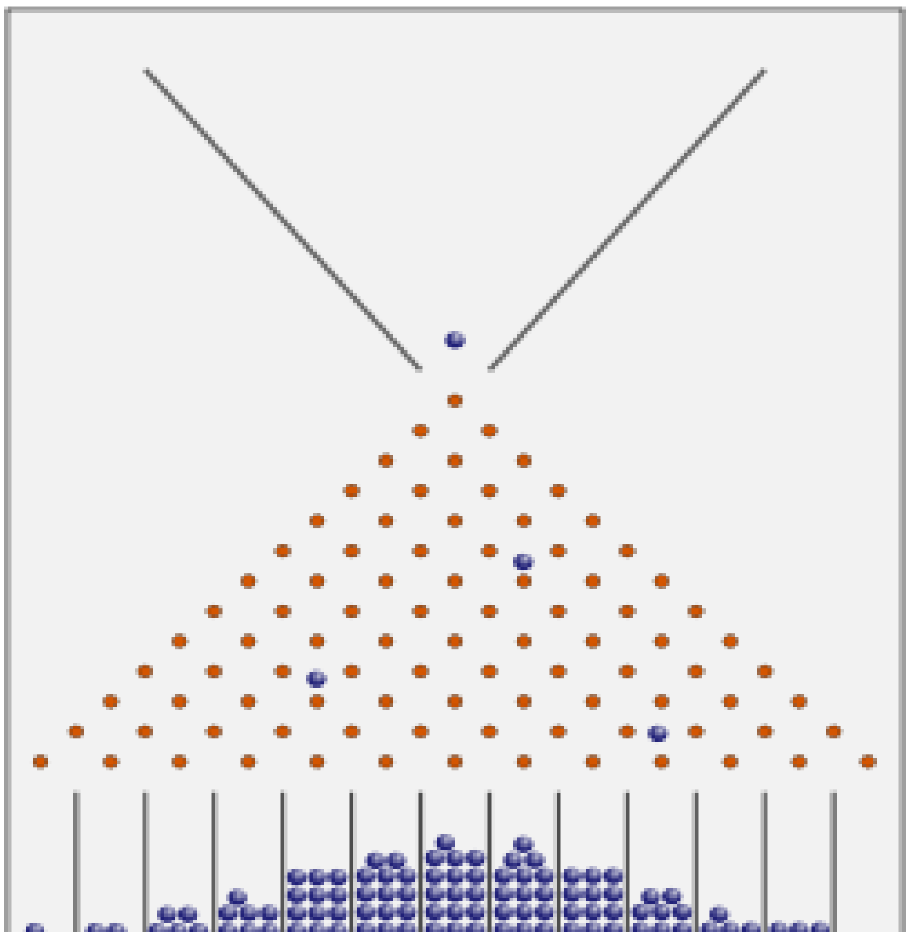
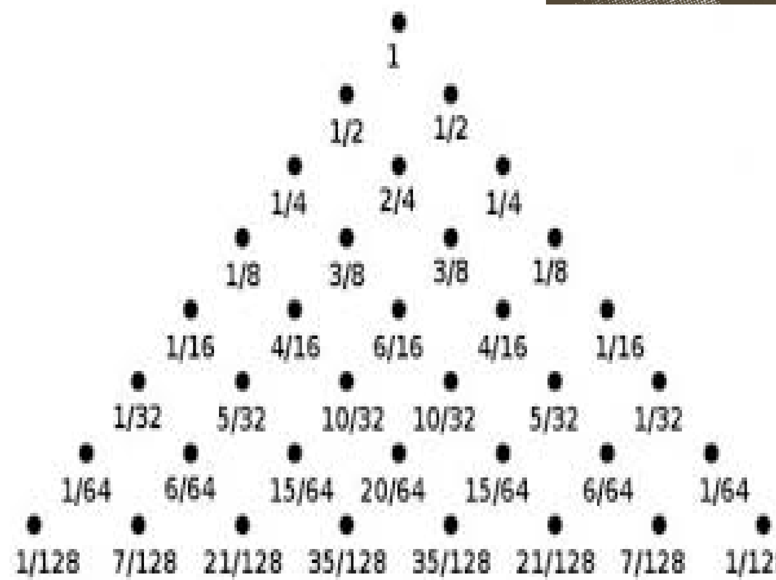


Planche de Galton

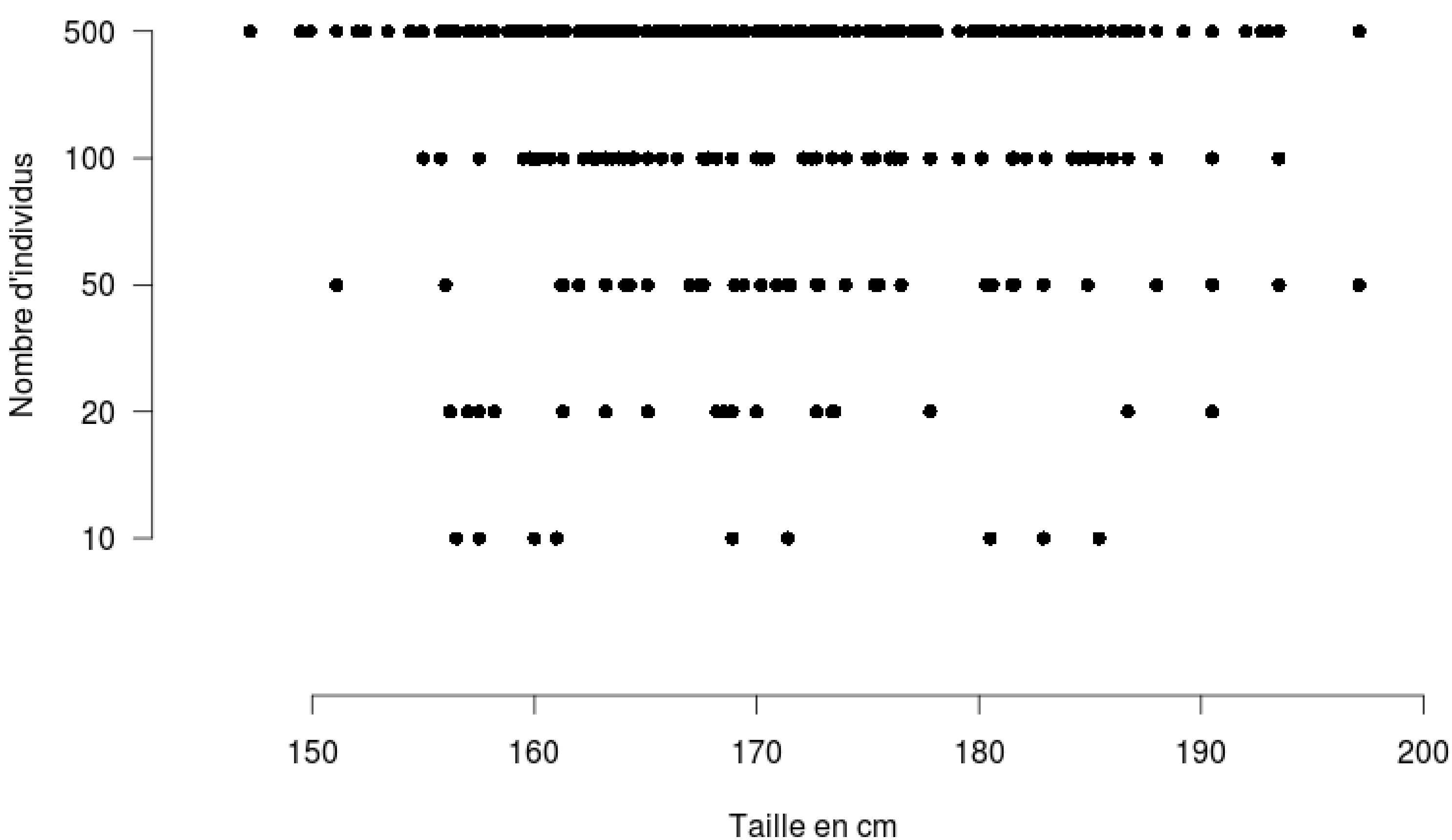




Sébastien Déjean - math.univ-toulouse.fr/~sdejean

D'où sort la courbe en cloche ?

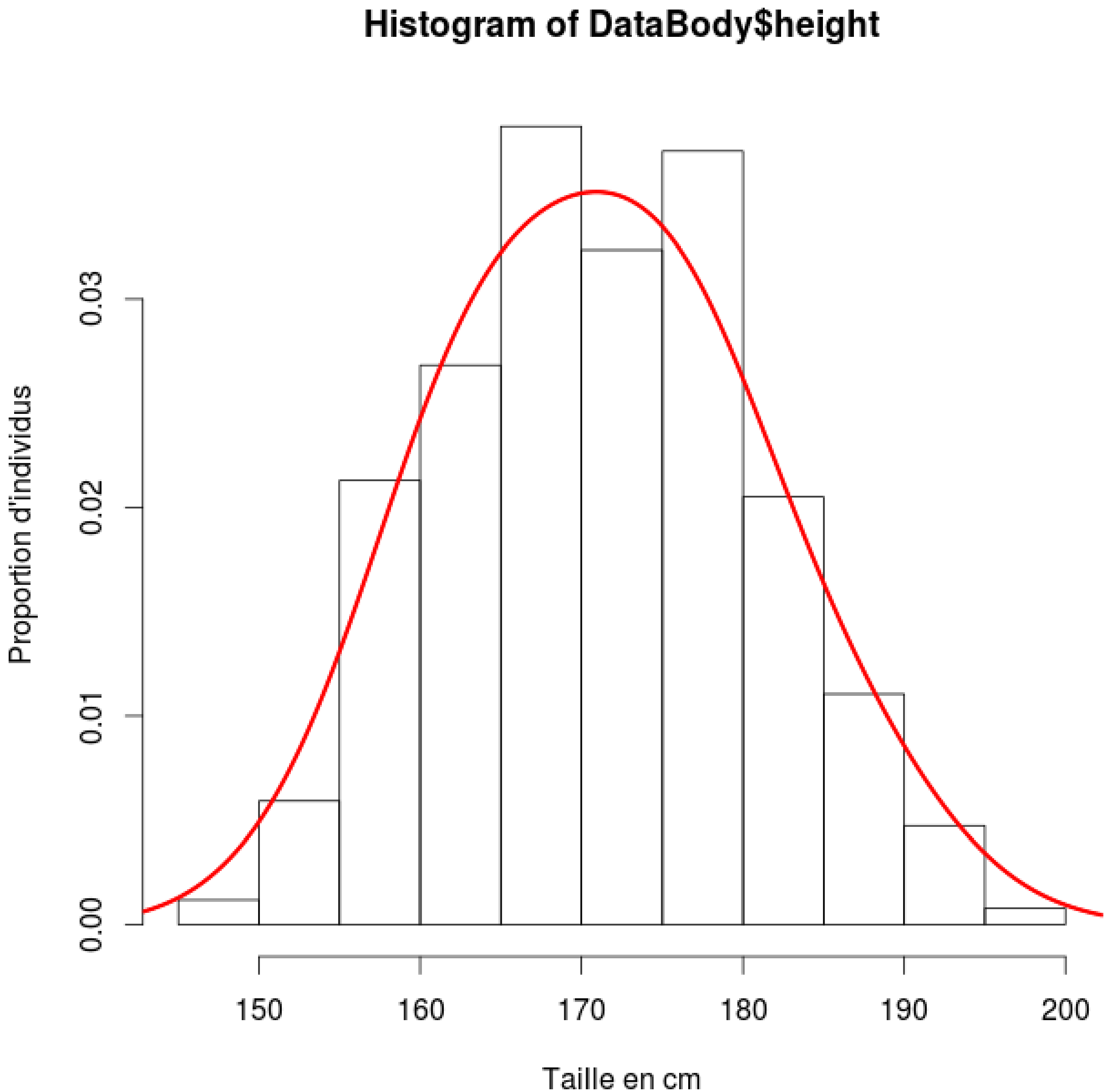
De la taille (de beaucoup) d'individus



Sébastien Déjean - math.univ-toulouse.fr/~sdejean

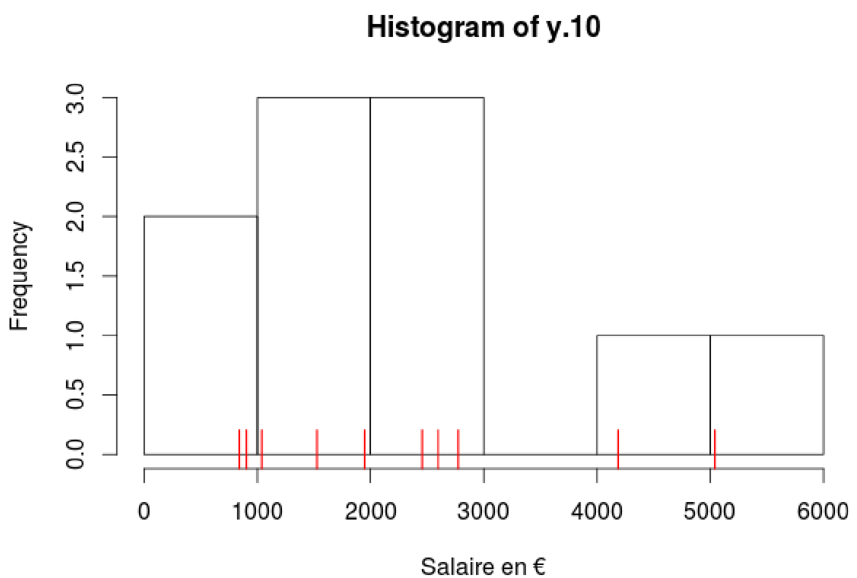
D'où sort la courbe en cloche ?

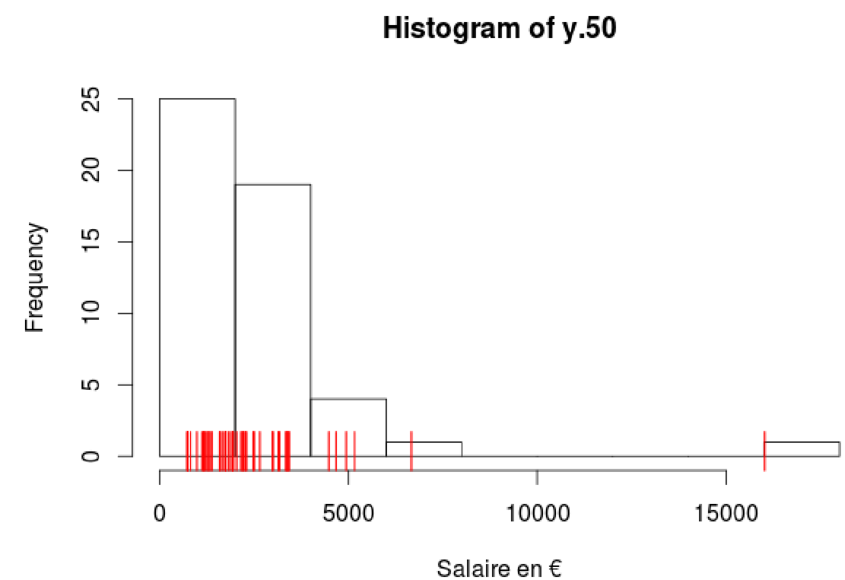
De la taille (de beaucoup) d'individus

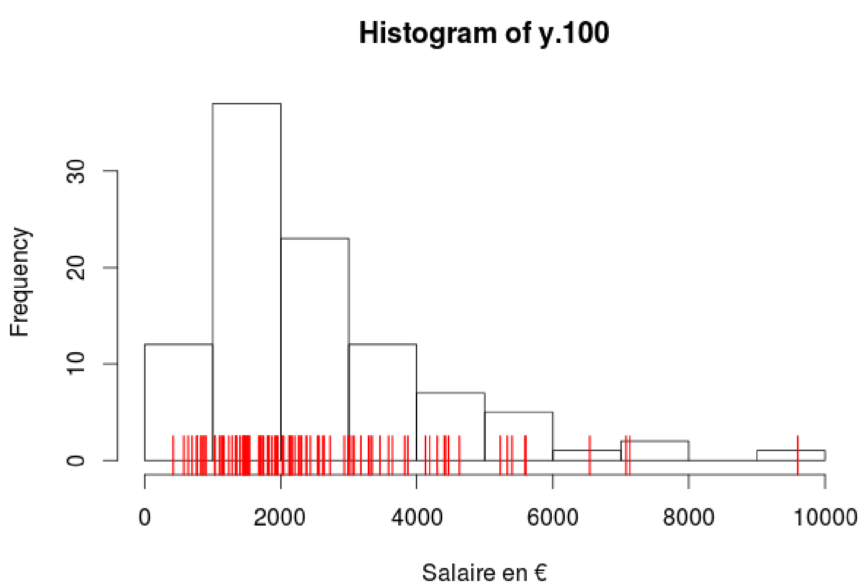


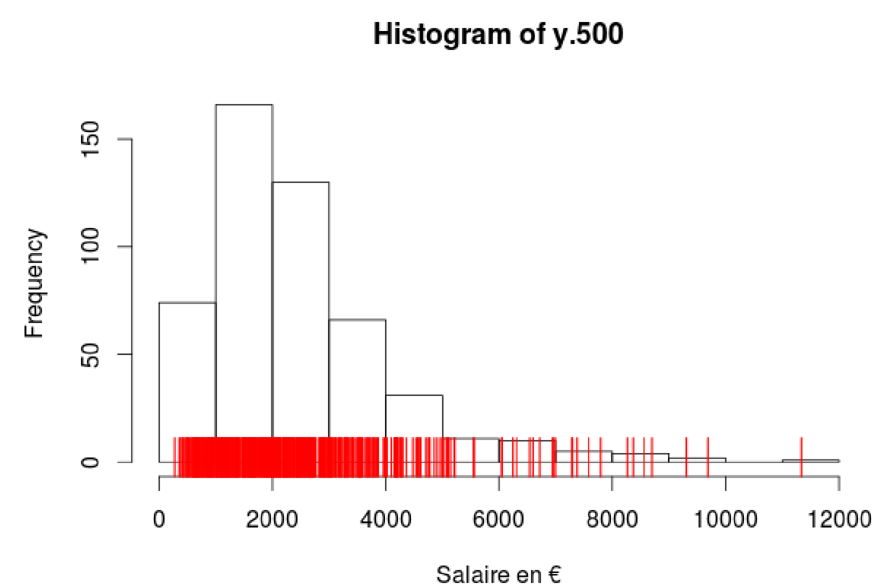
Il n'y a pas que la courbe en cloche

Exemple : distribution de salaires



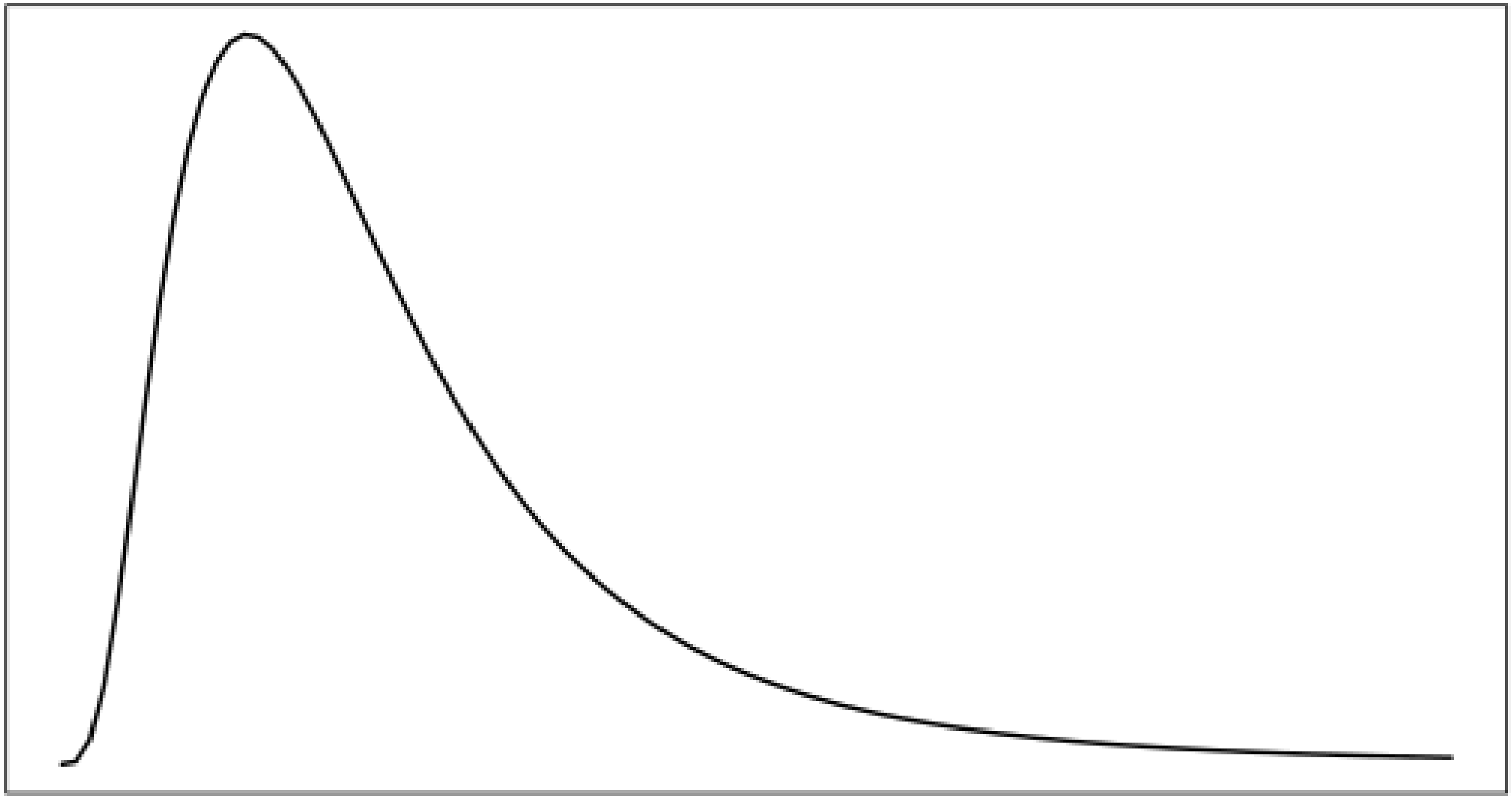






D'autres formes sont possibles

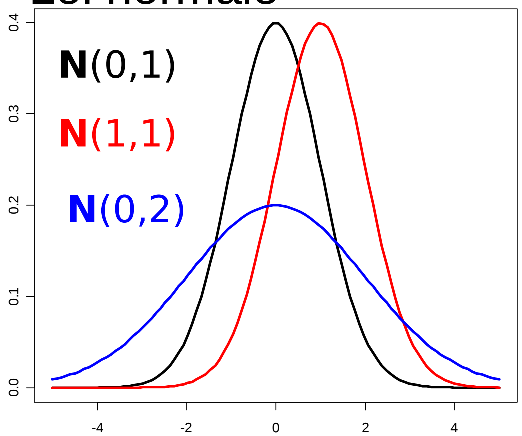
Distribution de χ^2, distribution de Fisher



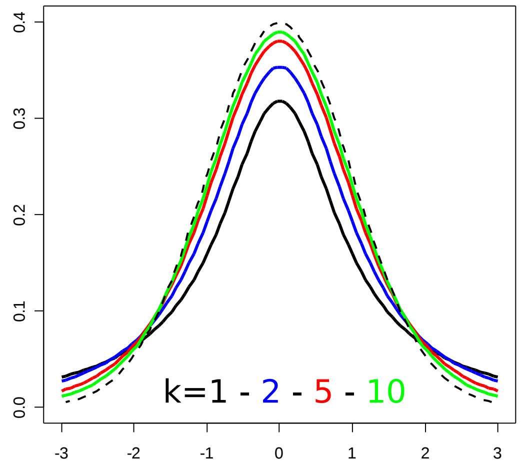
D'autres formes sont possibles

En agissant sur les paramètres des lois de probabilité

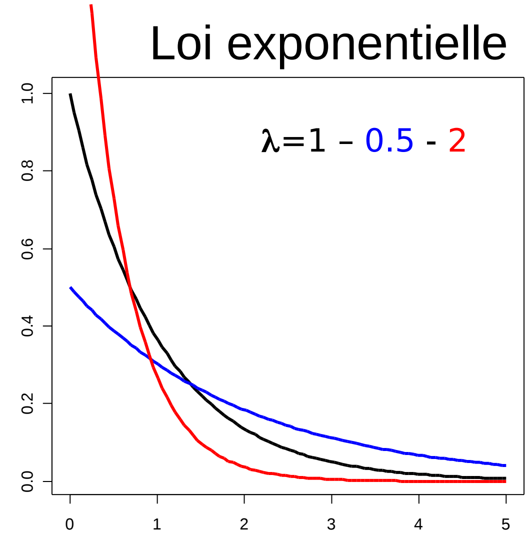
Loi normale



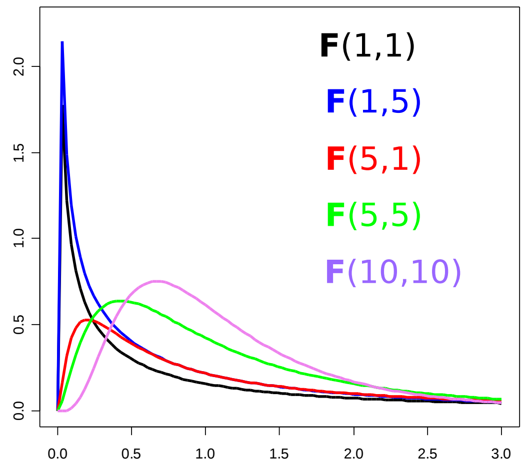
Loi de Student



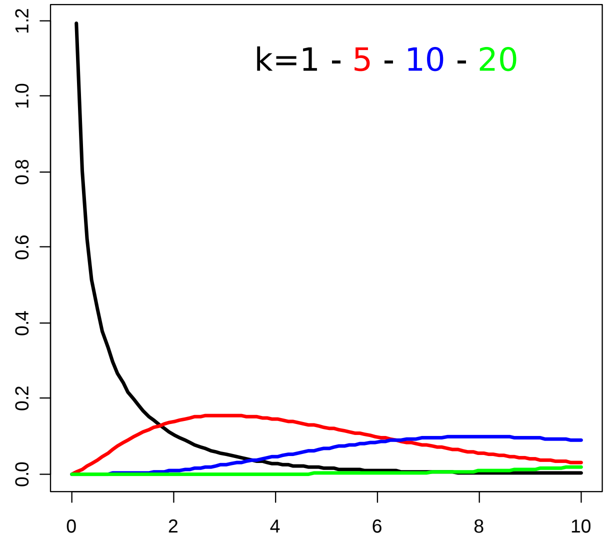
Loi exponentielle



Loi de Fisher

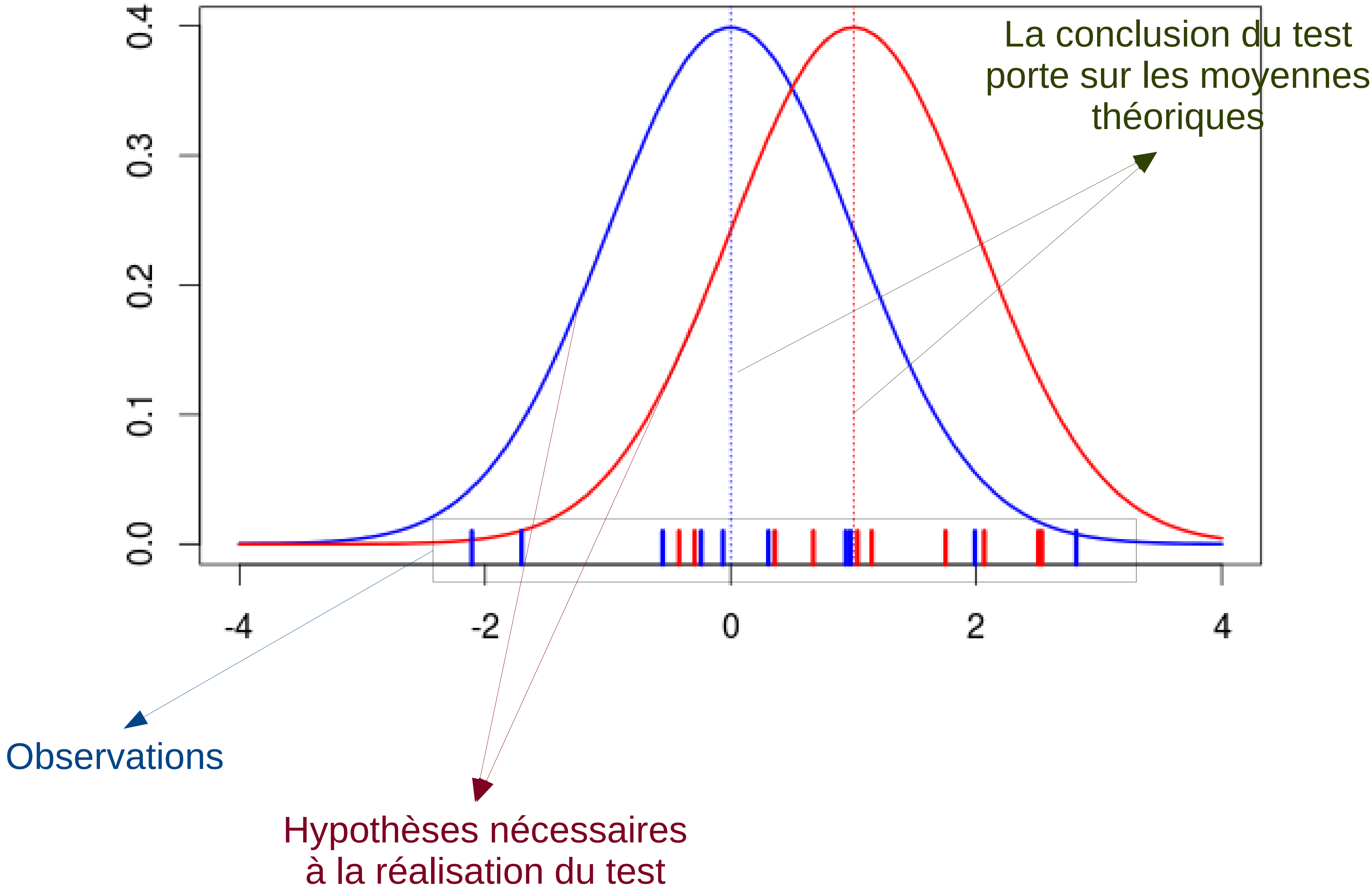


Loi du chi²



Test statistique ?

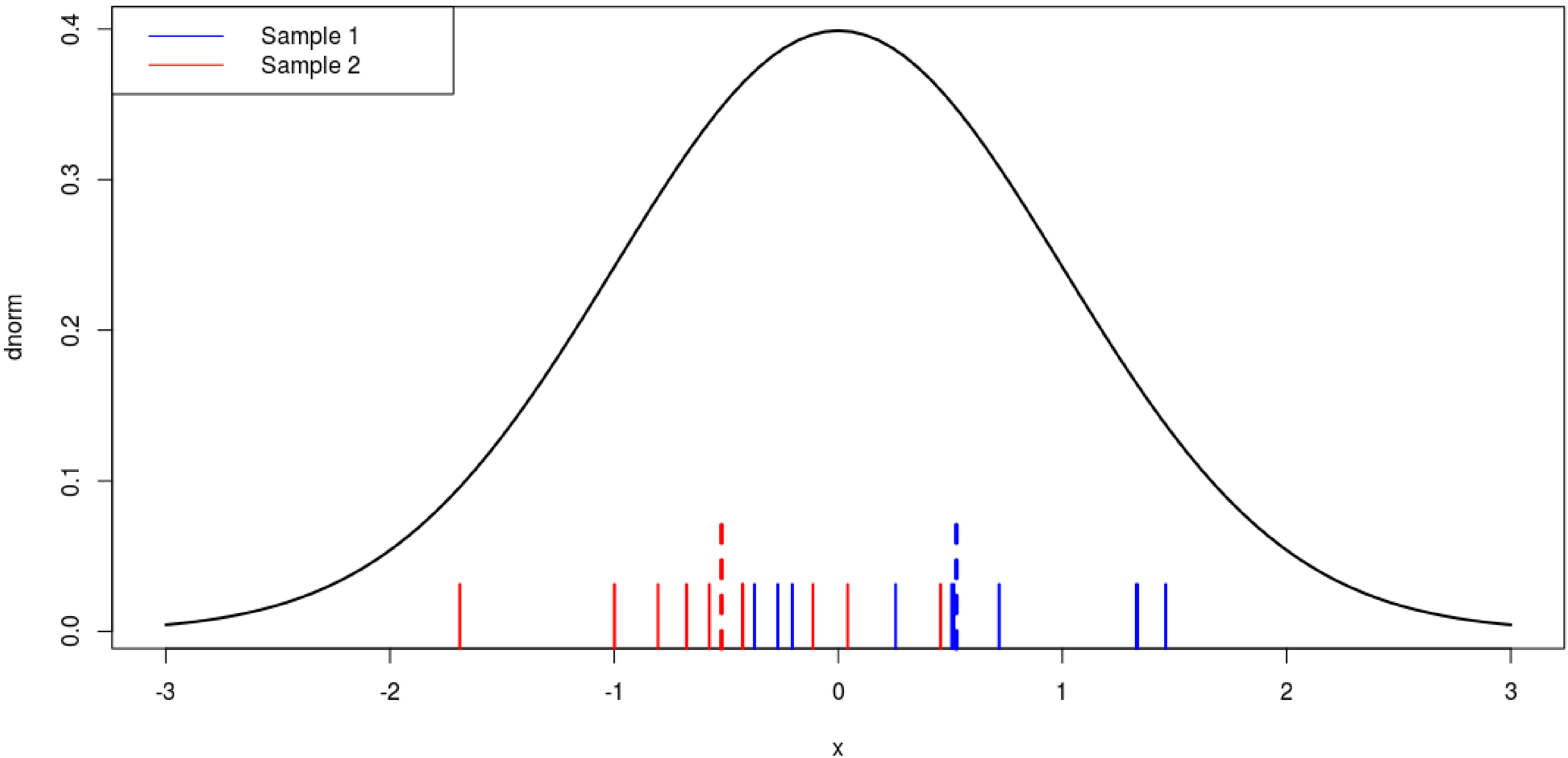
Exemple : le test de Student de comparaison de 2 moyennes



Faux positif ?

2 échantillons indépendants tirés de **la même population** peuvent conduire à une conclusion erronée

```
                Welch Two Sample t-test
data:  matrice[indice, 1:10] and matrice[indice, 11:20]
t = 3.6523, df = 17.61, p-value = 0.001878
alternative hypothesis: true difference in means is not equal to 0
```



Test statistique

Exemple : Fabrication industrielle de pain d'épice dont l'angle de rupture d'une tranche doit être de 50°. Des facteurs incontrôlés font que cet angle est aléatoire.

Question : comment décider qu'un lot est conforme ?

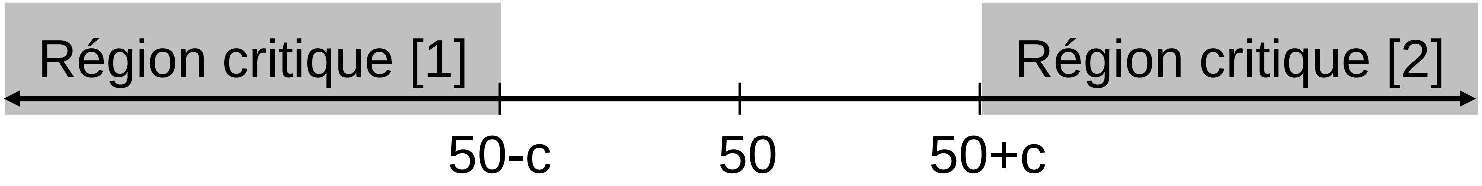
Hypothèses :

H0 : le lot est conforme ($\mu=50$)
H1 : le lot n'est pas conforme ($\mu\neq50$)

Pour trancher entre les 2 hypothèses, on tire au hasard un échantillon de *n* tranches et on en mesure l'angle de rupture $(X_i)_{i=1,\ldots n}$.

On suppose que chaque **X_i suit une loi $N(\mu,\sigma^2)$**.

Règle de décision (principe): Rejet de H0 $\Leftrightarrow \overline{X} \notin [50\text{-}c\ ;\ 50\text{+}c]$



Test statistique : risques d'erreur

Rappel :
H0 : lot conforme
H1 : lot non conforme

		Décision	
		H1 (rejet de H0)	H0 (accept. H0)
Réalité	H0	α	Bonne décision
	H1	Bonne décision	β

Interprétation des risques (en termes de pain d'épice) :

- α : rejeter le lot de biscuits alors qu'il est conforme (gaspillage !)
  ➔ *Le patron ne va pas être content. En fait non, il ne le saura pas.*

- β : déclarer conforme, et donc vendre, des biscuits « défectueux »
  ➔ *Dans ce cas, c'est le client qui n'est pas content. Lui, le saura.*

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Test statistique : risques d'erreur

- Gène différentiellement exprimé entre WT et KO ?
  - H0 : le gène n'est pas différentiellement exprimé ($\mu_{WT}=\mu_{KO}$)
  - H1 : le gène est différentiellement exprimé ($\mu_{WT}\neq\mu_{KO}$)
- α : décider qu'un gène est différentiellement exprimé alors qu'il ne l'est pas. Faux-positif, perte de temps et d'argent pour le confirmer (RT-qPCR...)
- β : décider qu'un gène n'est pas différentiellement exprimé alors qu'il l'est. Faux négatif, on passe peut-être à côté d'une découverte importante.
- Une nouvelle molécule pour augmenter...
  - H0 : l'effet de la nouvelle molécule est similaire à celui d'un placebo ($\mu_{drug}=\mu_{placebo}$)
  - H1 : l'effet de la nouvelle molécule est supérieur à celui d'un placebo ($\mu_{drug}>\mu_{placebo}$)
- α : décider qu'une molécule est meilleur qu'un placebo alors qu'elle ne l'est pas. Commercialiser une nouvelle molécule sans effet (assurance maladie...)
- β **:** décider qu'une molécule a le même effet qu'un placebo alors qu'elle est plus efficace. Des malades ne vont pas bénéficier de ce traitement pourtant efficace.

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Région critique et risque α

Règle de décision : Rejet de H0 $\Leftrightarrow \bar{X} \notin [50-c \,;\, 50+c]$

α = P[Rejeter H0 // H0 vraie] = $P[\bar{X} \notin [50-c \,;\, 50+c] \,//\, \mu = 50]$

Sous H0 ($\mu = 50$)

$$X_i \sim N(\mu,\sigma^2) \Rightarrow \bar{X} \sim N(\mu,\sigma^2/n) \quad ?$$

Pour l'application numérique : $n=16$ et $\sigma^2=9$

$$\alpha = P[\text{« } N(50,9/16) \text{ »} \notin [50-c \,;\, 50+c]\,]$$

Le risque α est la probabilité qu'une variable aléatoire suivant une loi normale de moyenne 50 et de variance 9/16 n'appartienne pas à l'intervalle [50-c ; 50+c].

Calculs

$$X_i \sim N(\mu,\sigma^2) \Rightarrow X \sim \overline{N}(\mu,\sigma^2/n)$$

Données simulées avec $\sigma = 5$

Données pour les échantillons de longueur 3

				Moyennes
Ech_1	172	171	166	170
Ech_2	172	166	179	173
Ech_3	164	169	165	166
Ech_4	160	177	173	170
Ech_5	172	172	171	171
Ech_6	163	168	171	167
Ech_7	164	172	170	169
Ech_8	163	166	174	168
Ech_9	164	181	160	168
Ech_10	164	170	167	167
Ech_11	173	160	163	165
Ech_12	173	177	165	172
Ech_13	172	166	176	171
Ech_14	171	179	167	172
Ech_15	171	167	172	170
Ech_16	175	176	172	175
Ech_17	170	170	173	171
Ech_18	166	163	177	168
Ech_19	171	183	179	178
Ech_20	168	172	178	173

Moyenne des moyennes **170**

Ecart-type des moy. **3.0** $.\simeq 5/\sqrt{3}$

Données pour les échantillons de longueur 10

											Moyennes
Ech_1	167	162	170	176	167	170	167	169	168	169	169
Ech_2	168	171	170	173	155	171	166	168	166	172	168
Ech_3	163	175	169	182	170	169	168	173	172	176	172
Ech_4	173	168	169	166	170	166	176	171	173	177	171
Ech_5	169	171	168	172	171	165	172	163	168	171	169
Ech_6	167	165	167	169	177	167	169	162	166	176	169
Ech_7	163	179	169	175	173	165	165	171	170	171	170
Ech_8	165	165	173	172	177	174	163	164	174	170	170
Ech_9	167	170	167	172	172	170	180	169	170	167	170
Ech_10	171	165	168	169	171	167	159	167	159	167	166
Ech_11	167	176	170	167	170	171	173	164	165	167	169
Ech_12	170	177	168	165	162	172	173	170	168	158	168
Ech_13	172	171	171	171	169	170	166	165	168	168	169
Ech_14	173	177	179	172	164	173	174	174	174	166	173
Ech_15	165	173	170	173	164	172	169	167	175	172	170
Ech_16	167	165	175	164	168	158	168	171	172	173	168
Ech_17	168	175	178	167	174	167	168	167	171	175	171
Ech_18	165	171	167	175	174	163	176	167	165	167	169
Ech_19	162	168	160	172	170	166	174	174	175	166	169
Ech_20	179	167	173	172	173	170	165	166	172	176	171

Moyenne des moyennes **170**

Ecart-type des moy. **1.5**

$.\simeq 5/\sqrt{10}$

Pour les échantillons de longueur 100

Moyenne des moyennes **170**

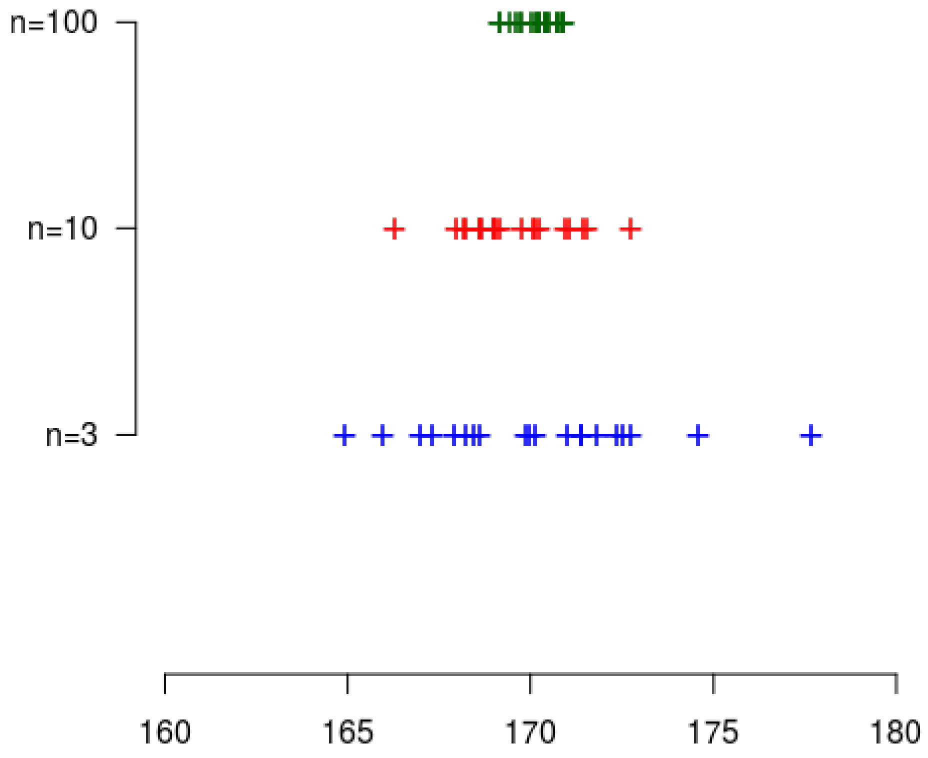
Ecart-type des moy. **0.5** $.\simeq 5/\sqrt{100}$

Précision d'une moyenne

$$X_i \sim N(\mu,\sigma^2) \;\Rightarrow\; \bar{X} \sim N(\mu,\sigma^2/n) \quad ???$$

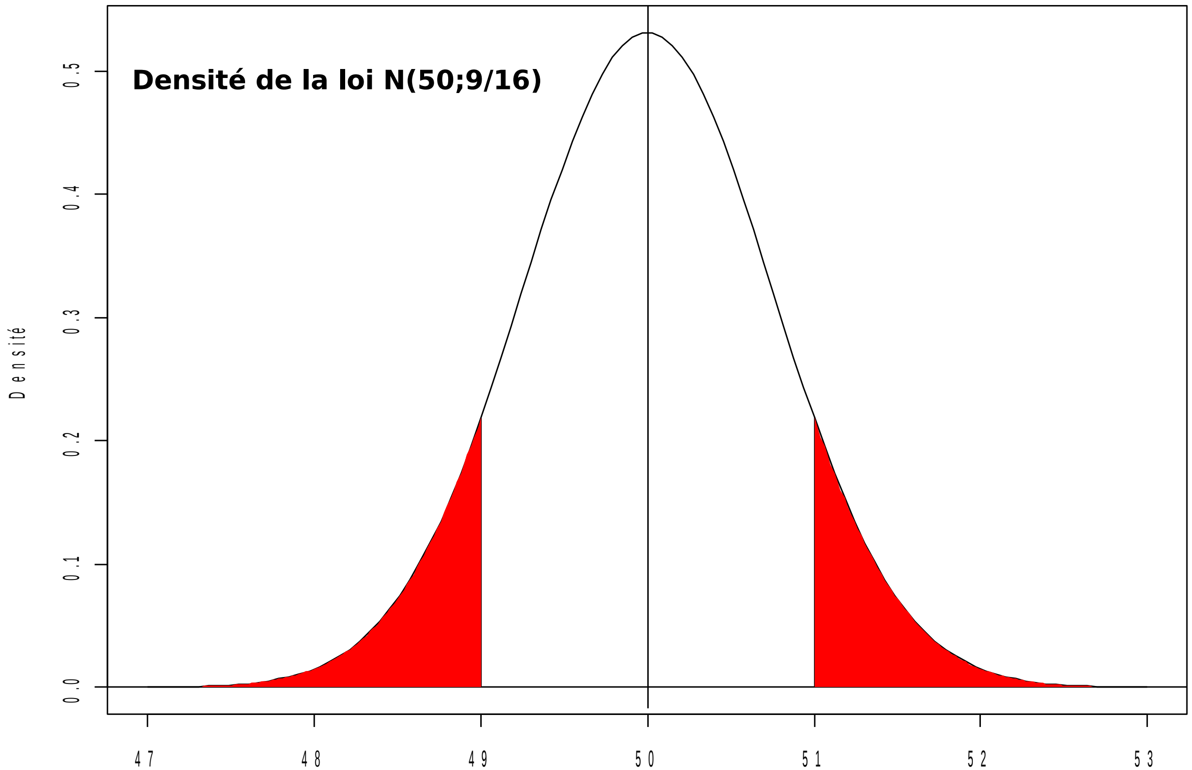
Une moyenne est plus précise quand elle est calculée à partir d'un plus grand nombre d'observations.

20 moyennes de tailles d'individus calculées à partir d'échantillons de longueur 3 (bleu), 10 (rouge) et 100 (vert).



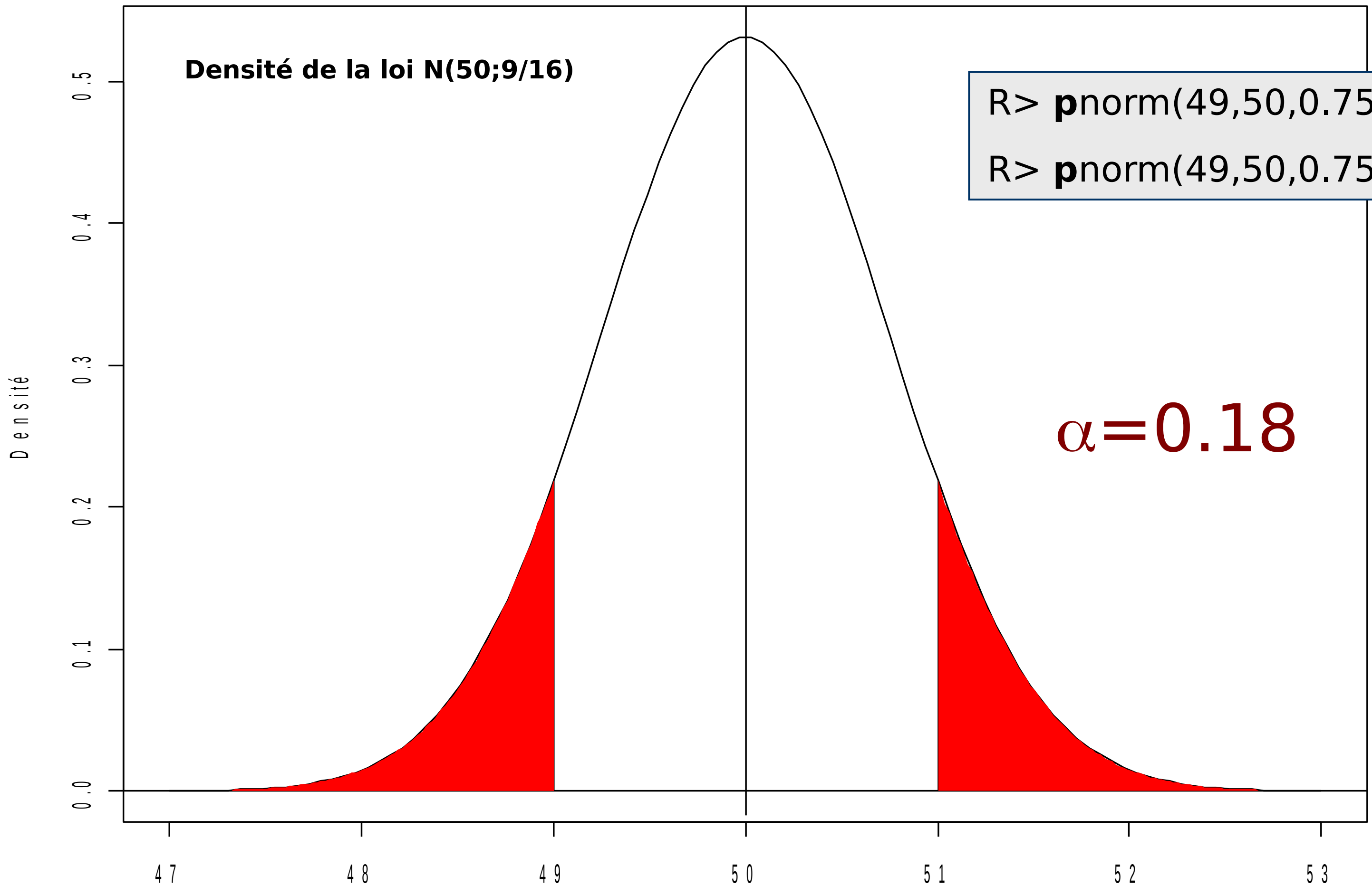
Région critique et risque α

Calculer *la probabilité qu'une variable aléatoire suivant une loi normale de moyenne 50 et de variance 9/16 n'appartienne pas à l'intervalle [50-c ; 50+c]*, revient à calculer l'aire située sous la courbe représentant la distribution de probabilités N(50,9/16) et hors de l'intervalle [50-c ; 50+c].



Région critique et risque α

➔ Exemple : prenons **c=1**, la région critique est **![49 ; 51]**. Calculons le risque α associé. α=P[« N(50,9/16) » $\notin$ [49 ; 51]

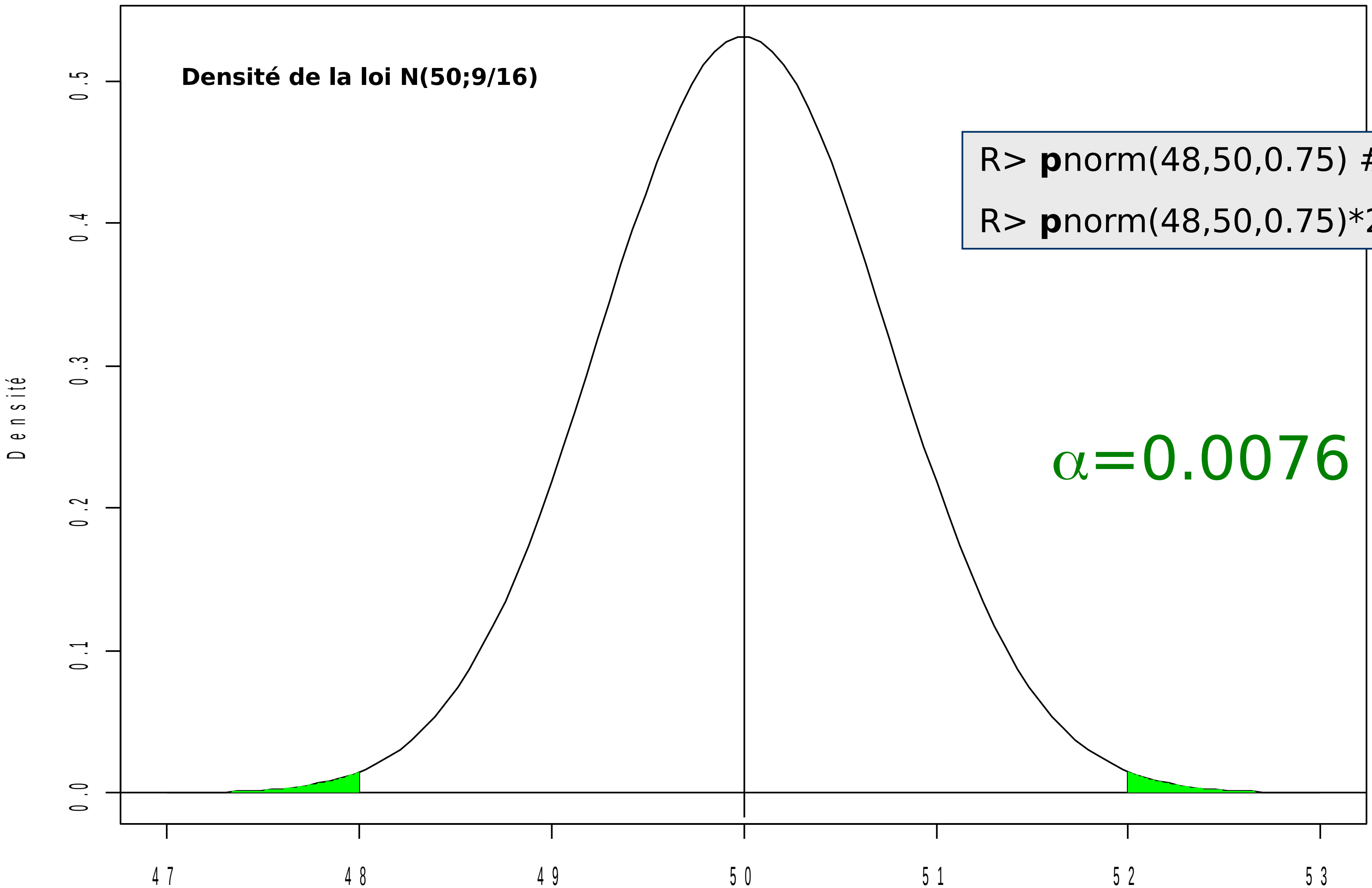



```
R> pnorm(49,50,0.75) # 0.09
R> pnorm(49,50,0.75)*2 # 0.18
```

![49 ; 51] =]-∞ ; 49] ∪ **[51 ; +∞** [

Région critique et risque α

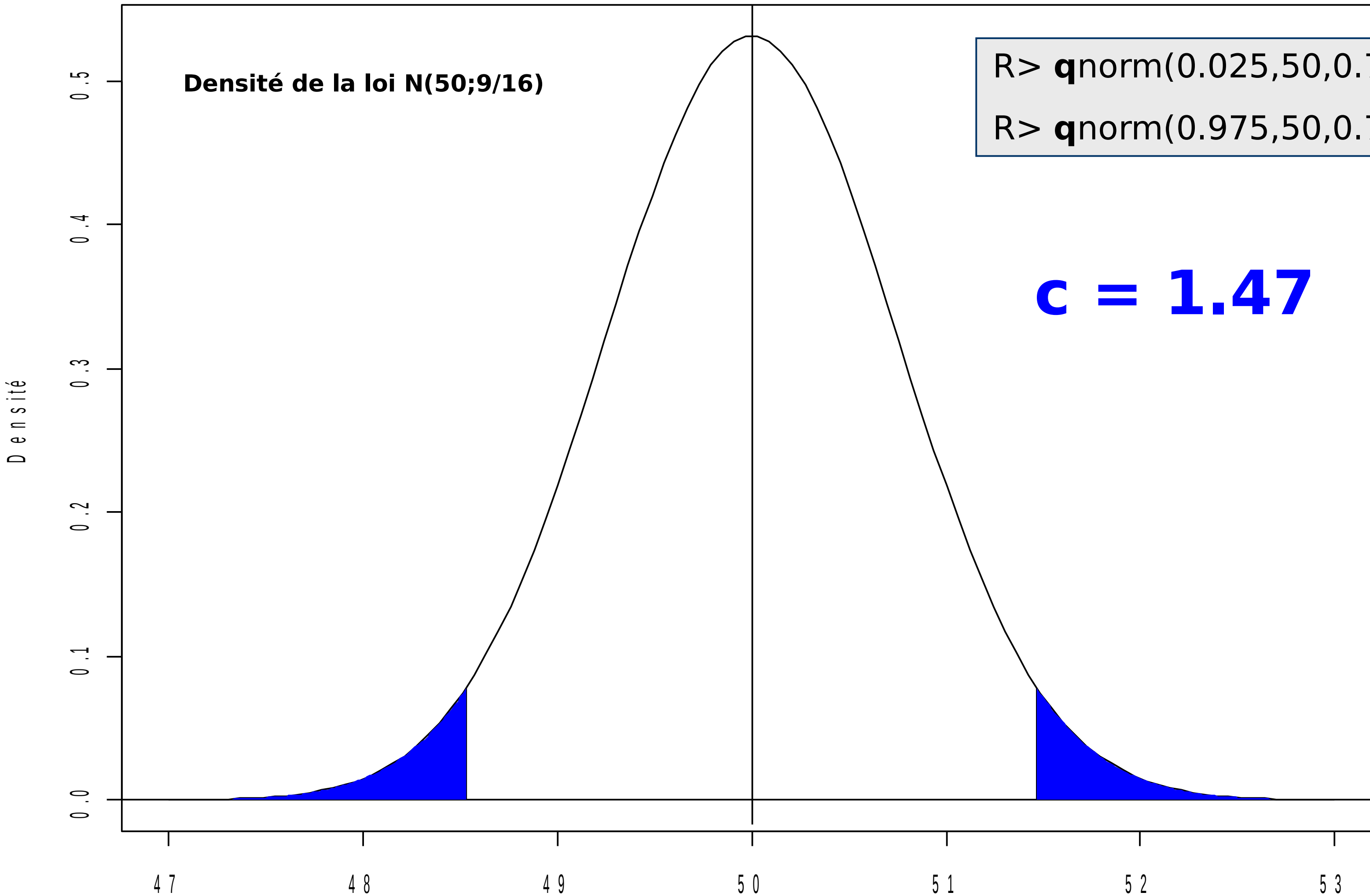
➔ Exemple : prenons **c=2**, la région critique est **![48 ; 52]**. Calculons le risque α associé. α=P[« N(50,9/16) » $\notin$ [48 ; 52]




```
R> pnorm(48,50,0.75) # 0.0038
R> pnorm(48,50,0.75)*2 # 0.0076
```

Région critique et risque α

➔ Trouver c tel que : $\alpha = P[\text{« } N(50,9/16) \text{ »} \notin [50-c\ ;\ 50+c]\ ] = \mathbf{0.05}$



R> **q**norm(0.025,50,0.75) # 48.53

R> **q**norm(0.975,50,0.75) # 51.47

c = 1.47

P-value

*

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

P-value

“We teach it because it’s what we do;
we do it because it’s what we teach.”

Q: Why do so many colleges and grad schools teach p = 0.05?

A: Because that’s still what the scientific community and journal editors use.

Q: Why do so many people still use p = 0.05?

A: Because that’s what they were taught in college or grad school.

George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College

P-value

An unhealthy obsession with p-values is ruining science
http://www.vox.com/2016/3/15/11225162/p-value-simple-definition-hacking

"The proportion of papers that use p-values is going up over time, and the most significant results have become even more significant over time."

John Ioannidis

Though statisticians have long been pointing out problems with "significance doping" and "P-dolatory" (the "worship of false significance") ***journals have increasingly relied on p-values to determine whether a study should be published.***

"It's this number that looks like you could use it to make a decision that might otherwise be difficult to make or require a whole lot more effort to make,"

"The p-value was never intended to be a substitute for scientific reasoning,"

Ron Wasserstein, Executive director of the American Statistical Association

P-value

Good luck trying to find a really clear definition of a p-value.

Not Even Scientists Can Easily Explain P-values
`http://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/`

The ASA's Statement on p-Values: Context, Process, and Purpose
`http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108`

"Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (for example, the sample mean difference between two compared groups) would be equal to or more extreme than its observed value."

*I called Rebecca Goldin, the director for Stats.org and a professor at George Mason University, for help parsing that still **perplexing definition**.*

Digression

p-value ⟷ sondage

recherche scientifique ⟷ Journalisme politique

Le Parisien abandonne les sondages politiques pendant la campagne
http://www.leparisien.fr/flash-actualite-politique/le-parisien-abandonne-les-sondages-politiques-pendant-la-campagne-03-01-2017-6520437.php

Article qui pourrait être ré-intitulé d'après le premier article cité en "*Une obsession malsaine pour les sondages ruine le journalisme (politique)*" ou d'après le second "*Même les sondeurs (ceux que l'on voit à la télé, pas les "vrais") ne savent pas expliquer facilement les résultats d'un sondage*"

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Digression

Le Parisien abandonne les sondages politiques pendant la campagne

*Le directeur des rédactions du Parisien/Aujourd'hui en France Stéphane Albouy a annoncé mardi sur France Inter que le quotidien ne commanderait plus de sondages politiques, une "pause" pendant la campagne pour "**se concentrer sur le journalisme de terrain**".*
"C'est une réflexion qu'on a mené depuis quelques temps déjà, notamment après le Brexit et l'élection de Donald Trump", explique-t-il à l'AFP, ajoutant que le journal ne commandait plus de sondages depuis plusieurs semaines déjà.
"Ce n'est pas une question de défiance envers les sondeurs mais une façon de travailler différemment que nous voulons tester pour la suite de la campagne", poursuit-il.
*Il souhaite notamment éviter "ce côté course de petits chevaux où on se focalise sur qui prend la première position" afin de "**se concentrer sur le fond, sur les programmes**".*
Il ne s'interdit pas toutefois de commenter les sondages commandés par d'autres médias.
Consommateur de sondages, le titre y consacre "quelques dizaines de milliers d'euros par an", selon Stéphane Albouy, qui insiste sur le fait qu'il ne s'agit pas avec cette "pause" de réaliser des économies.
*"On peut entendre les critiques qui nous sont faites, à nous, médias, d'être coupés d'une forme de réalité. **Nous allons privilégier le terrain**", explique-t-il, rappelant que le journal s'appuie sur un réseau de 140 journalistes déployés en Ile-de-France.*
*"**Déployer ces journalistes sur le terrain, cela coûte plus cher que les sondages, et nous oblige aussi à être plus exigeants**", estime-t-il.*

P-value

- La p-value n’est pas facile à définir simplement :
  - *Not Even Scientists Can Easily Explain P-values*, C. Aschwanden 24/11/2015, http://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/
  - *An unhealthy obsession with p-values is ruining science*, J. Belluz, 15/03/2016, http://www.vox.com/2016/3/15/11225162/p-value-simple-definition-hacking
- « Degré de significativité »
- C'est la plus petite des valeurs de α pour lesquelles les observations conduisent au rejet de H0.
- C'est donc la probabilité, sous H0, d’observer les données ou des données « plus extrêmes ».
- Comparer la p-value et α : rejet de H0 ⇔ p-value < α
- Plus une p-value est petite, plus le risque de se tromper en rejetant l’hypothèse H0 est faible.

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

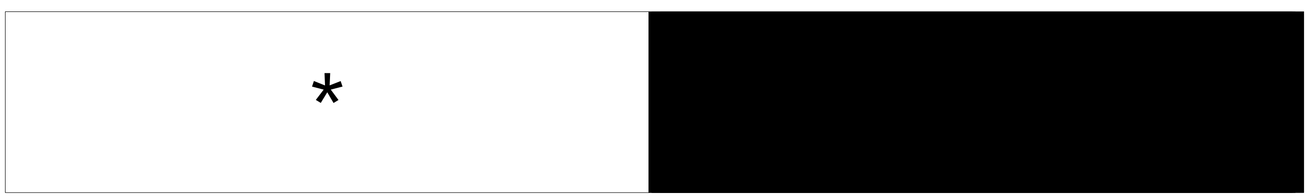
P-value et *

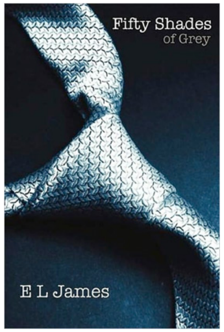
Une histoire vraie, retour d’un referee

Second, the statistical methods performed are confusing and interpretation of significance is improper. Details about the stat’s need to be moved to the methods section. Commenting on the level of statistical significance based on the p-value is incorrect. ***A p-value is either less than alpha value (rejecting null hypothesis) or it is not (retaining null hypothesis); a smaller p-value does not indicate that something has greater or stronger significance. Please delete adjectives (i.e. slightly, strongly, etc.) accordingly.***

5%

p-value < 5%





Roman de E.L. JAMES sorti en 2011

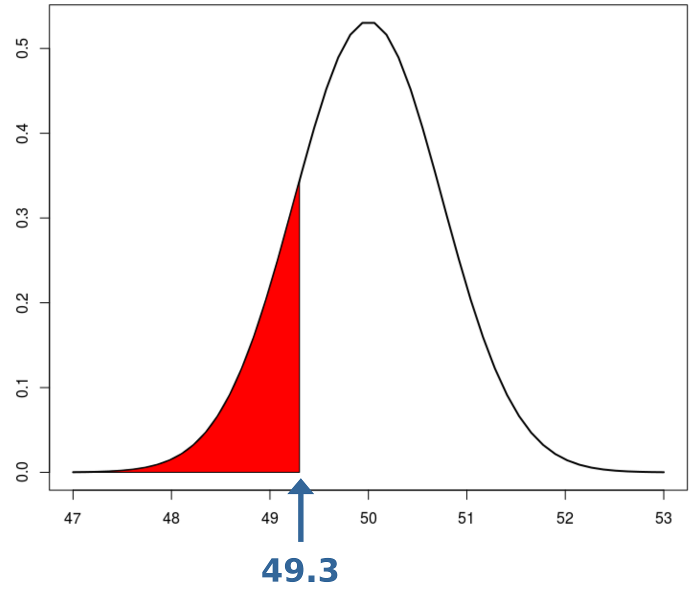
There is a reason that the speedometer in your car doesn't just read "slow" and "fast" -- Frank Harrell (warning about the use of cutoffs after logistic regression) R-help (February 2011)

Album de Jean-Jacques Goldman sorti en 1987

P-value (exemple)

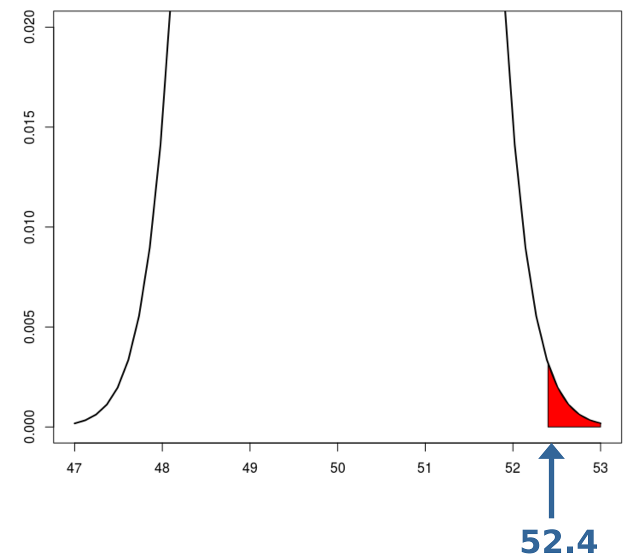
Cas 1) : Angle moyen pour 16 tranches : **49.3**

Cette valeur **n'est pas dans la région critique** (![48.53 ; 51.47]), on ne peut pas rejeter H0, la production du jour est probablement conforme. La p-value associée à la valeur 49.3 est environ 0.17 ce qui est supérieur au seuil de 5%. Elle indique qu'en supposant que l'hypothèse nulle est vraie (angle moyen de la population = 50), la probabilité d'observer un angle moyen de 49.3 pour un échantillon de taille 16 est de 17 %.



Cas 2) : Angle moyen pour 16 tranches : **52.4**

Cette valeur **est dans la région critique**, on rejette H0, la production du jour n'est pas conforme (au seuil de 5%). La p-value associée à la valeur 52.4 est de l'ordre de 0.0007 ce qui est inférieur au seuil de 5%. Elle indique qu'en supposant que l'hypothèse nulle est vraie (angle moyen de la population = 50), la probabilité d'observer un angle moyen de 52.4 pour un échantillon de taille 16 est de 0.007 %. Ce qui nous incite à rejeter cette hypothèse au niveau de la population.



Et le risque β ?

β = P[Accepter H0 // H1 vraie] = P[$\bar{X} \in [50-c\ ;\ 50+c]$ // μ =**???**]

Le calcul explicite du risque β nécessite des valeurs de μ.

$\beta(\mu)=P[«N(\mu,9/16)»\in[50-c;50+c]\ ]$

Pour **c=1 :** RC=!**[49 ; 51]** - α=0.18

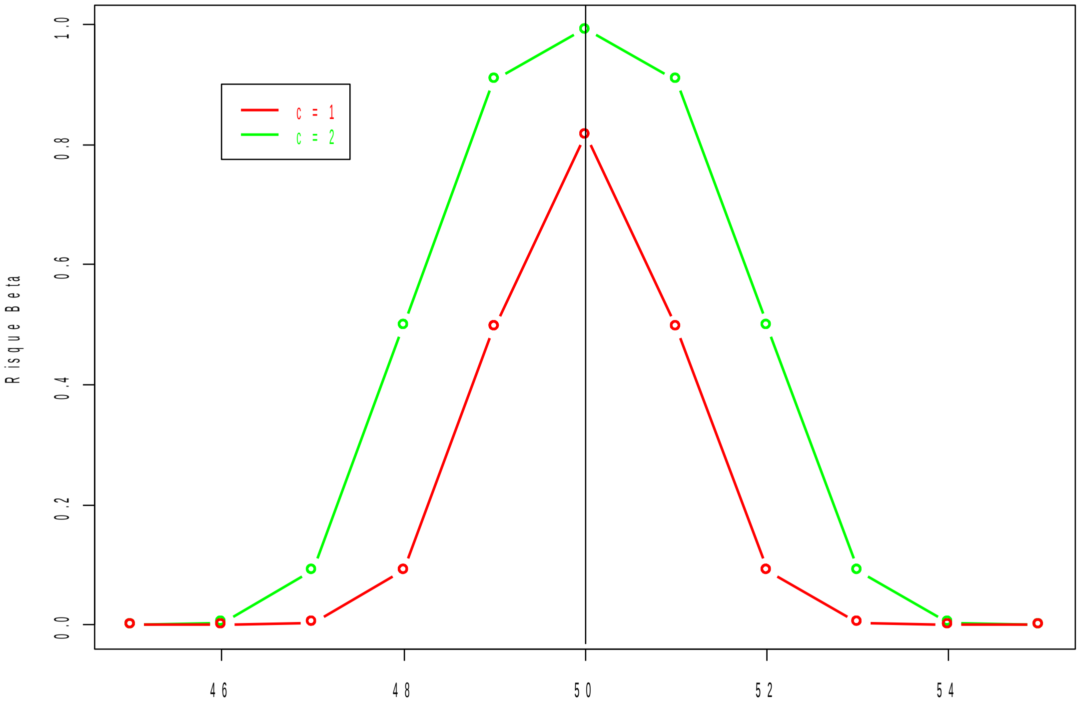
```
R> pnorm(51,45,0.75)-pnorm(49,45,0.75)
```

β(**45**)=4.10-8
β(**46**)=3.10-5
β(**47**)=0.0038
β(**48**)=0.091
β(**49**)=0.496
β(**50**)=0.818
β(**51**)=0.496
β(**52**)=0.091
β(**53**)=0.0038
β(**54**)=3.10-5
β(**55**)=4.10-8

Pour **c=2 :** RC=!**[48 ; 52]** - α=0.0076

```
R> pnorm(52,45,0.75)-pnorm(48,45,0.75)
```

β(**45**)=3.10-5
β(**46**)=0.0038
β(**47**)=0.091
β(**48**)=0.5
β(**49**)=0.91
β(**50**)=0.99
β(**51**)=0.91
β(**52**)=0.5
β(**53**)=0.091
β(**54**)=0.0038
β(**55**)=3.10-5



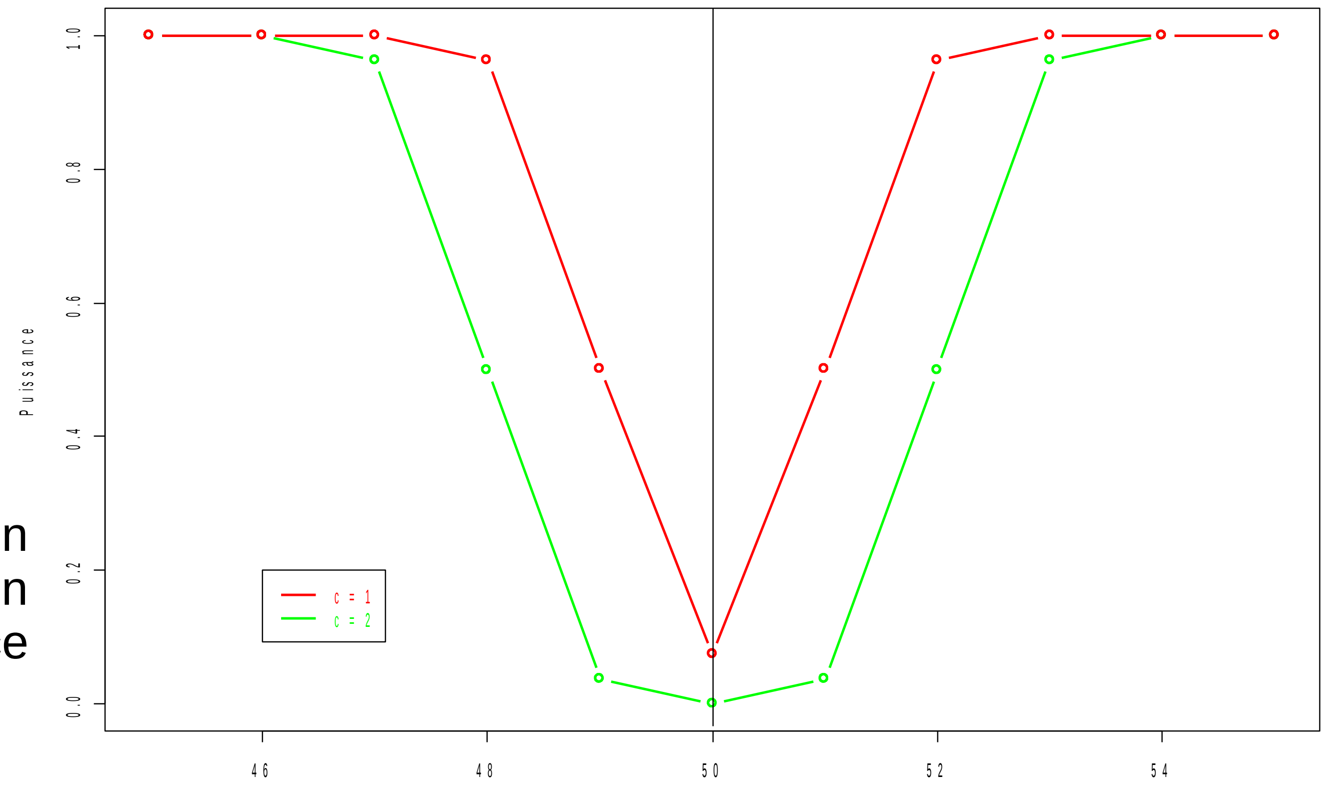
Puissance d'un test

La puissance d'un test est la probabilité de détecter une différence (rejeter H0) lorsqu'elle existe.

$$\mathbf{P} = 1\text{-}\beta$$

$$= 1\text{-P[Accepter H0 // H1 vraie]}$$

$$= \text{P[Rejeter H0 // H1 vraie]}$$



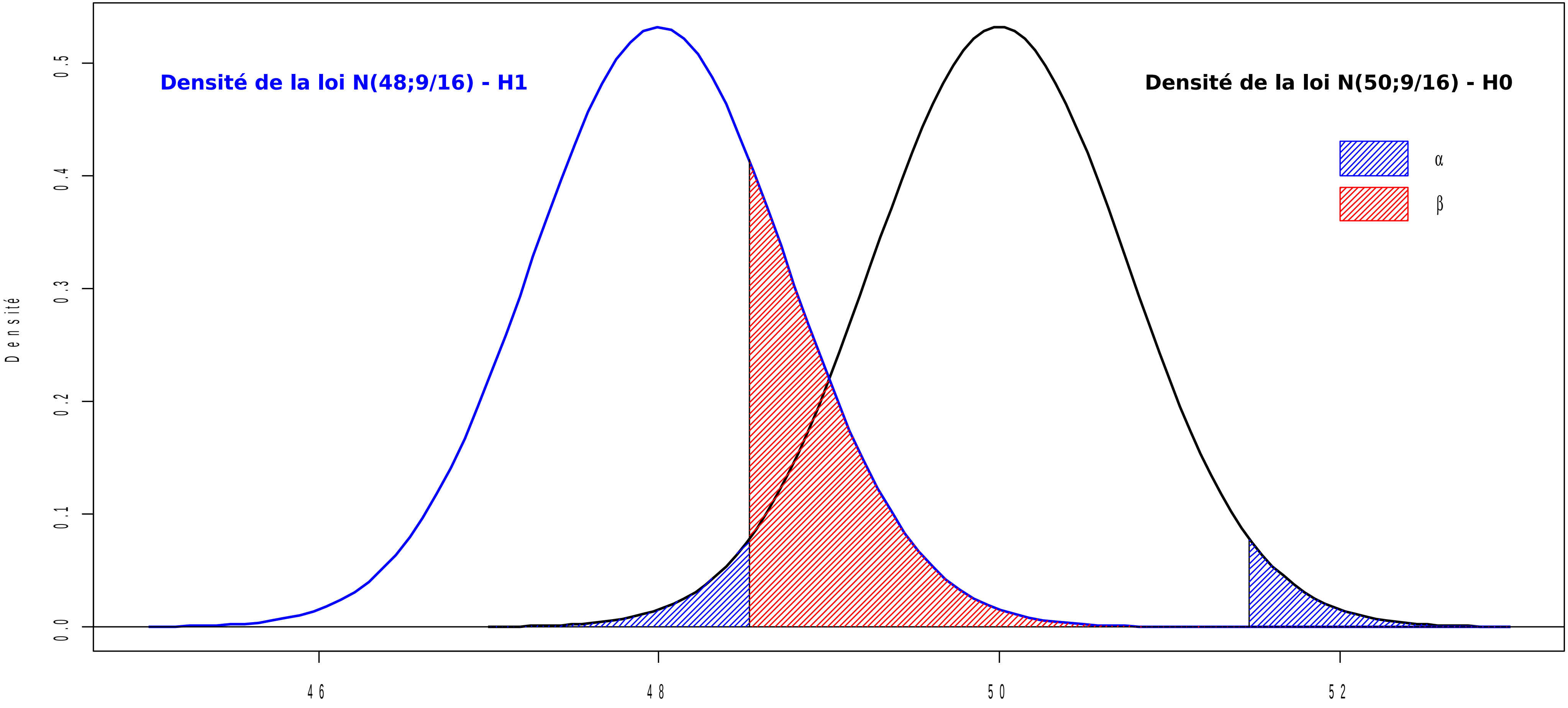
Représentation de la fonction puissance

Représentation graphique de α et β

H0: μ=50
H1: μ=48

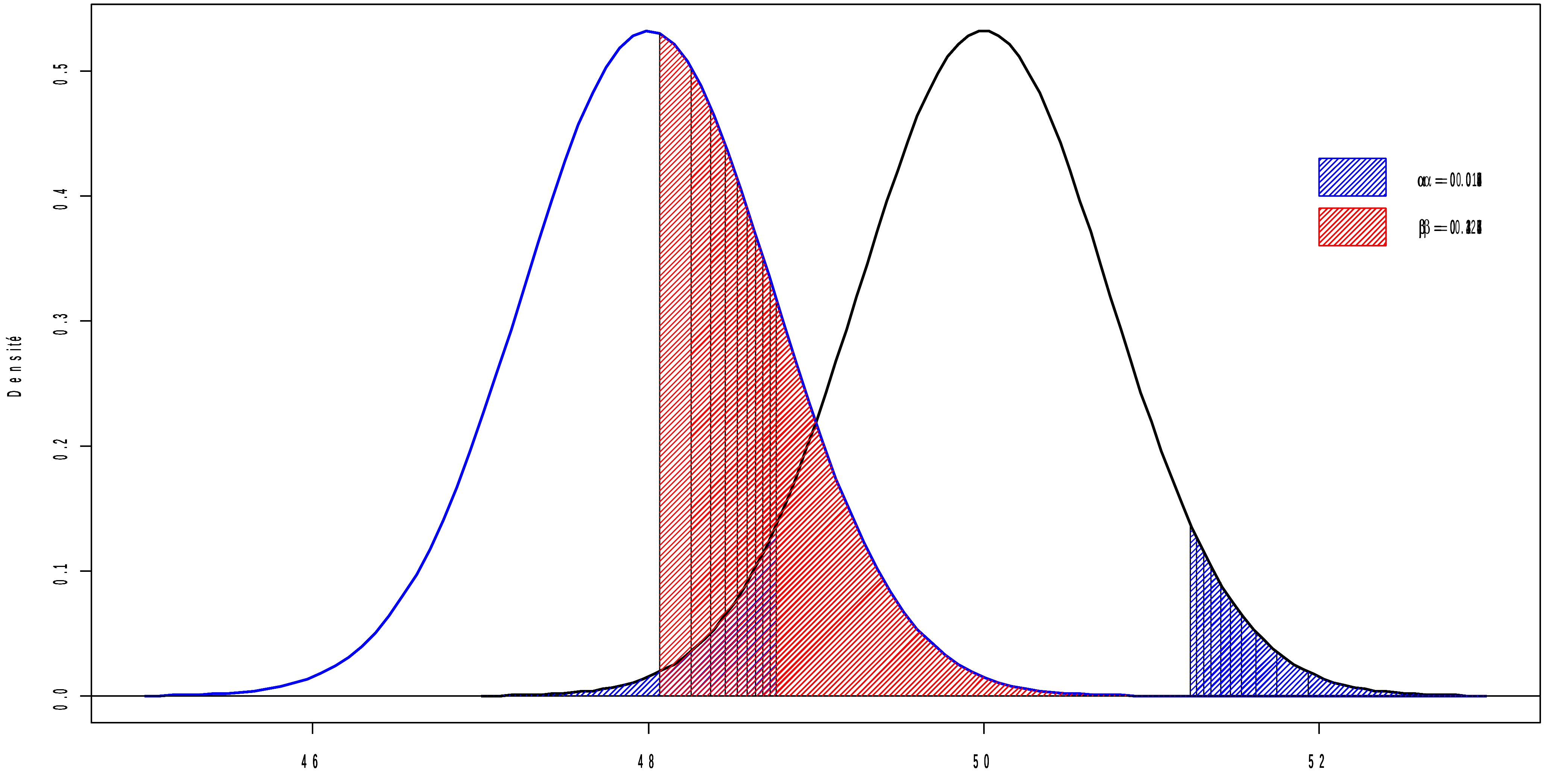
Rappel : pour α=5%, la région critique est]-∞ ; 48.53] ∪ [51.47 ; +∞ [
Dans ces conditions, β = 0.24

R> 1- **p**norm(48.53,48,0.75)



Représentation graphique de α et β

Variations de α de 0.01 à 0.1



Test « significatif »

- Si le test conduit à rejeter H0, le risque de se tromper (α) est faible. La conclusion en faveur de H1 est solide. Le test est dit **significatif**.
- Si le test conduit à accepter H0,le risque de se tromper (β) peut être grand (selon l'hypothèse alternative). Cette conclusion est moins solide. Dans ce cas, le test est dit **non significatif**. D'où l'habitude d'affirmer « on ne peut pas rejeter H0 » plutôt que « on accepte H0 ».
- D'où la nécessaire réflexion du choix des hypothèses H0 et H1 ; H1 étant celle que l'on souhaite voir satisfaite avec un faible risque de se tromper.

Statistical significance

en.wikipedia.org/wiki/Statistical_hypothesis_testing

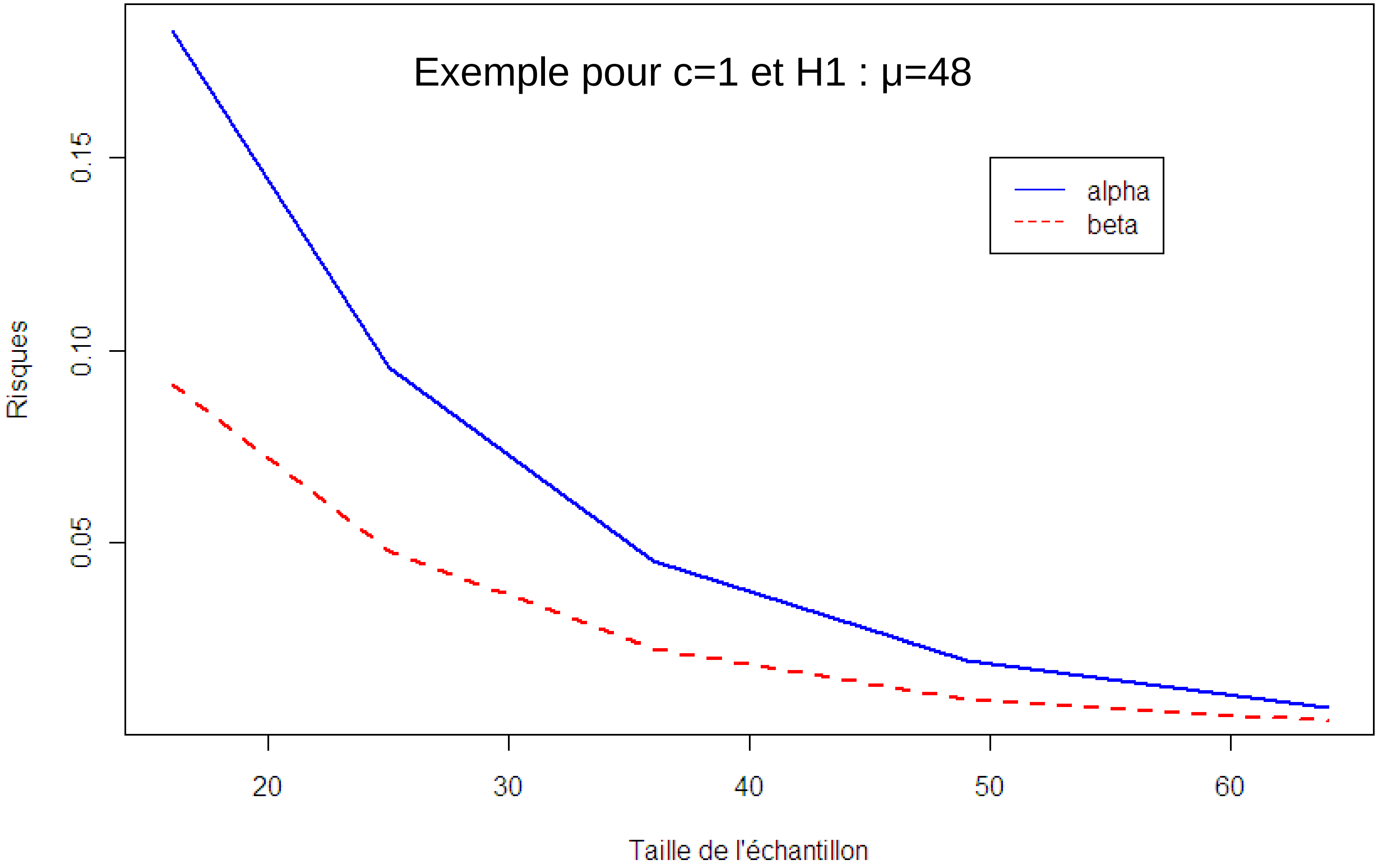
It is important to note the difference between ***accepting the null hypothesis*** *and simply* ***failing to reject it****. The* ***"fail to reject" terminology*** *highlights the fact that the null hypothesis is assumed to be true from the start of the test; if there is a lack of evidence against it, it simply continues to be assumed true. The phrase "accept the null hypothesis" may suggest it has been proved simply because it has not been disproved, a logical fallacy known as the argument from ignorance. Unless a test with particularly high power is used,* ***the idea of "accepting" the null hypothesis may be dangerous****. Nonetheless the terminology is prevalent throughout statistics, where the meaning actually intended is well understood.*

en.wikipedia.org/wiki/Statistical_significance

… a result has ***statistical significance*** *when* ***it is very unlikely to have occurred given the null hypothesis****. More precisely, the significance level defined for a study, α, is the probability of the study rejecting the null hypothesis, given that it were true; and the p-value of a result, p, is the probability of obtaining a result at least as extreme, given that the null hypothesis were true. The result is* ***statistically significant****, by the standards of the study, when* ***$p < \alpha$****.*

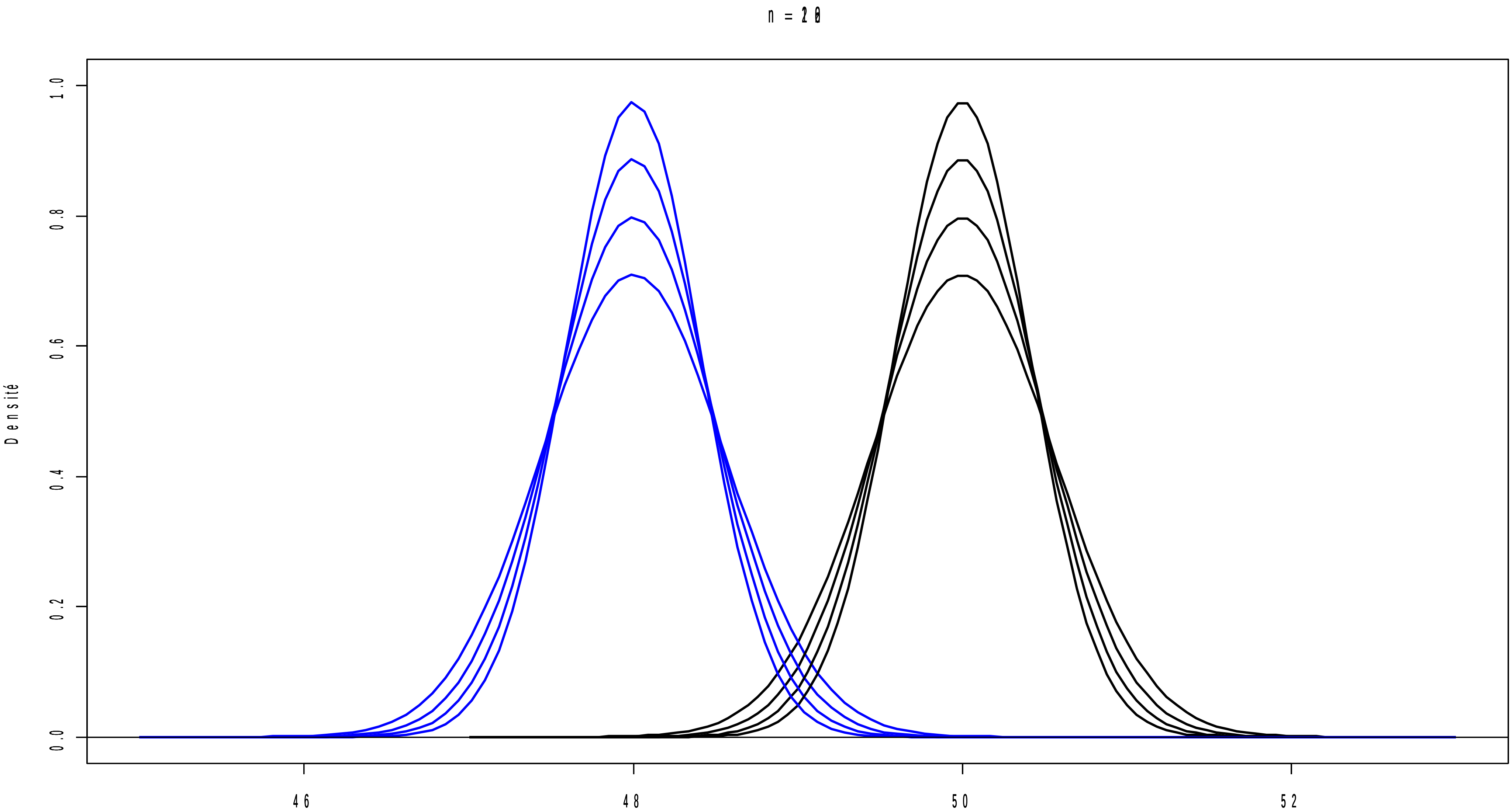
Diminuer α et β

Le seul moyen de diminuer simultanément les risques α et β consiste à augmenter la taille de l'échantillon (ce qui implique une diminution de la variance de $\overline{X}$ et donc diminue le recouvrement des 2 courbes). σ^2/n



Représentation graphique de α et β

Variations de la taille de l'échantillon



Calcul d'effectif : principe

- La puissance diminue (= le risque β augmente) quand :
  - Le risque α diminue et / ou
  - La taille de l'échantillon diminue et / ou
  - La taille de l'effet recherché diminue. En d'autres termes, l'hypothèse alternative H1 « se rapproche » de l'hypothèse nulle H0.

- Une fois fixés :
  - Le risque α
  - Le risque β ou la puissance du test (1-β)
  - La taille de l'effet à mettre en évidence

seule la taille de l'échantillon reste comme inconnue et on peut donc la définir **a priori** compte tenu des autres informations.

Calcul d'effectif : exemple

Extrait de l'aide en ligne de la fonction `power.t.test()` de R

```
Power calculations for one and two sample t tests

Description
Compute power of test, or determine parameters to obtain target power.

Usage
power.t.test(n = NULL, delta = NULL, sd = 1, sig.level = 0.05,
             power = NULL,
             type = c("two.sample", "one.sample", "paired"),
             alternative = c("two.sided", "one.sided"),
             strict = FALSE)

Arguments
n           Number of observations (per group)
delta       True difference in means
sd          Standard deviation
sig.level   Significance level (Type I error probability)
power       Power of test (1 minus Type II error probability)
type        Type of t test alternative One- or two-sided test
strict      Use strict interpretation in two-sided case

Details
Exactly one of the parameters n, delta, power, sd, and sig.level must be
passed as NULL, and that parameter is determined from the others.

...
```

Calcul d'effectif : exemple

R> power.t.test(n = 25, delta = 1, sd = 1, sig.level = 0.05, power = **NULL**)

```
     Two-sample t test power calculation

              n = 25
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.9337076
    alternative = two.sided

 NOTE: n is number in *each* group
```

R> power.t.test(n = 30, delta = 1, sig.level = **NULL** ,power= 0.9)

```
              n = 30
          delta = 1
             sd = 1
      sig.level = 0.01286591
          power = 0.90
```

R> power.t.test(n = **NULL**, delta = 1, sd = 1, sig.level = 0.05, power = 0,9)

```
     Two-sample t test power calculation

              n = 22.02110
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

 NOTE: n is number in *each* group
```

R> power.t.test(n = 30, delta = **NULL**, sig = 0.05 ,power= 0.9)

```
              n = 30
          delta = 0.8511743
             sd = 1
      sig.level = 0.05
          power = 0.9
```

En pratique

Panorama de quelques tests statistiques

Problèmes à 1 échantillon

Source : Wikipedia

Type de test	Test paramétrique	Test non paramétrique
Conformité à 1 standard	Test de comparaison de moyenne (Student), d'écart-type, d'une proportion à une valeur de référence	
Adéquation à une loi		• Kolmogorov-Smirnov • χ^2 d'adéquation • Shapiro-Wilk

Association entre variables

Type de test	Test paramétrique	Test non paramétrique
2 variables quantitatives	Coefficient de corrélation de Pearson	
2 variables qualitatives		χ^2 d'indépendance

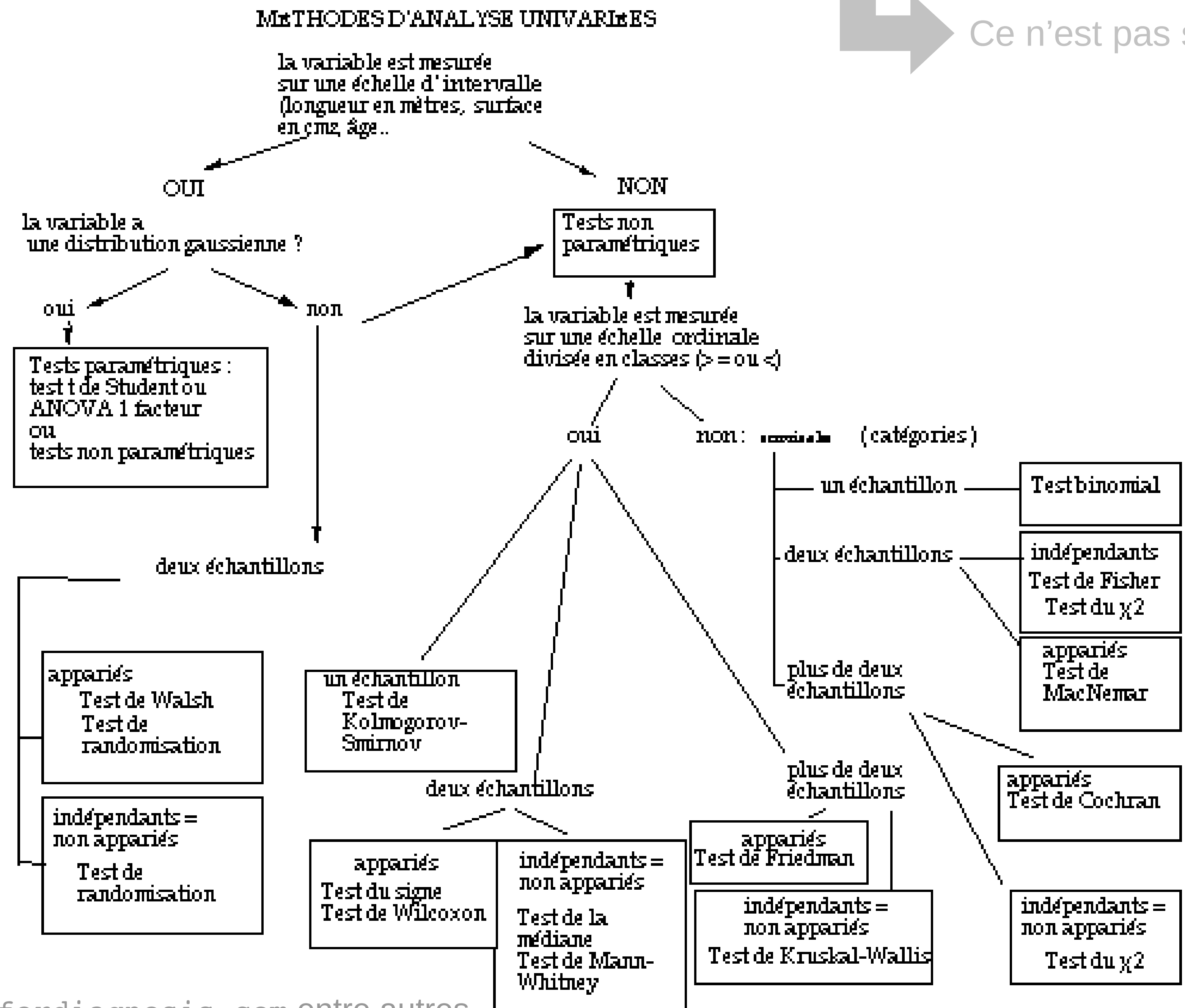
Panorama de quelques tests statistiques

Problèmes à K échantillons : comparaison de population

Type de test	Test paramétrique	Test non paramétrique
Comparaison de populations, les fonctions de répartition sont les mêmes dans les groupes		• **Kolmogorov-Smirnov** • Cramer – von Mises
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de tendance centrale)	• **Test de comparaison de moyennes** (K=2) • **ANOVA** (analyse de variance) à 1 facteur	• **somme des rangs de Wilcoxon** (K=2) • **Mann - Whitney** (K=2) • Kruskal - Wallis • Test des médianes
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de dispersion)	• **Fisher** (K=2) • Bartlett • Cochran • F-max de Hartley	• Ansari - Bradley • Siegel-Tukey • Test des différences extrêmes de Moses
Tests pour K échantillons appariés (mesures répétées ou blocs aléatoires complets)	• **Test de Student de comparaison de moyennes pour échantillons appariés** (K=2) • Test de comparaison de variances pour échantillons appariés (K=2) • ANOVA pour blocs aléatoires complets	• Test des signes (K=2) • **Rangs signés de Wilcoxon** (K=2) • Friedman • Test de McNemar (K=2, variables binaires) • Test Q de Cochran (variables binaires)
Tests multivariés pour K échantillons indépendants	• T^2 de Hotelling, comparaison de K=2 barycentres (vecteur des moyennes) • **MANOVA** (analyse de variance mutlivariée), comparaison de K barycentres : Lambda de Wilks, Trace de Pillai, Trace de Hotelling-Lawley, La plus grande valeur propre de Roy	Source : Wikipedia

Comment s'y retrouver ?

Ce n'est pas simple...



Source : `lookfordiagnosis.com` entre autres

Données indépendantes ou appariées ?

→ *Attention de ne pas se tromper dans ce choix !*

- Données **indépendantes** : les observations sont indépendantes à l'intérieur de chaque échantillon et d'un échantillon à l'autre

Ex: résultats scolaires filles et garçons, dosage d'un produit chez 2 groupes de patients ayant reçu une molécule ou un placebo...

- Données **appariées** : les mêmes individus sont soumis à 2 mesures successives d'une même variable

Ex: notes de copies soumises à une double correction, dosage d'un produit avant et après un traitement chez les mêmes individus...

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

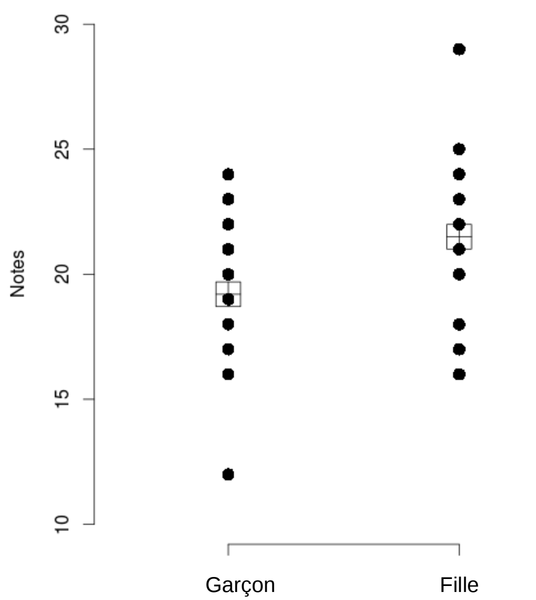
Données indépendantes ou appariées

Garçon	Fille
18	22
21	25
16	17
22	24
19	18
24	29
17	20
20	23
23	21
12	16

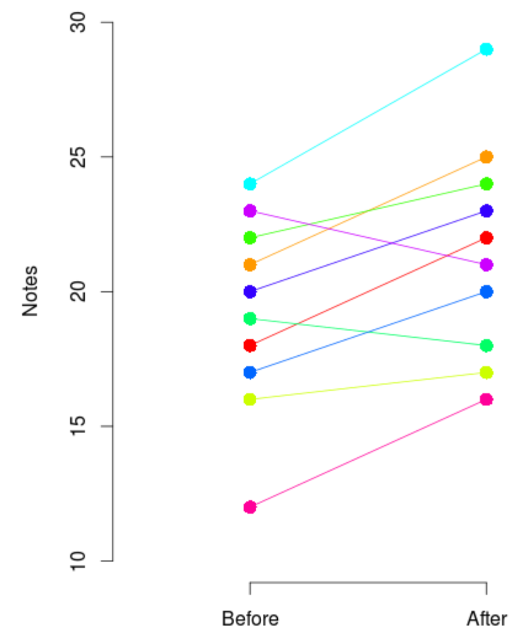
=

Garçon	Fille
18	16
21	21
16	23
22	20
19	29
24	18
17	24
20	17
23	25
12	22

	Avant	Après
Louise	18	22
Léo	21	25
Emma	16	17
Gabriel	22	24
Chloé	19	18
Adam	24	29
Lola	17	20
Timéo	20	23
Inès	23	21
Raphaël	12	16



t.test(x,y, **paired=FALSE**)
p-value = **0,1928**



t.test(x,y, **paired=TRUE**)
p-value = **0,0118** **

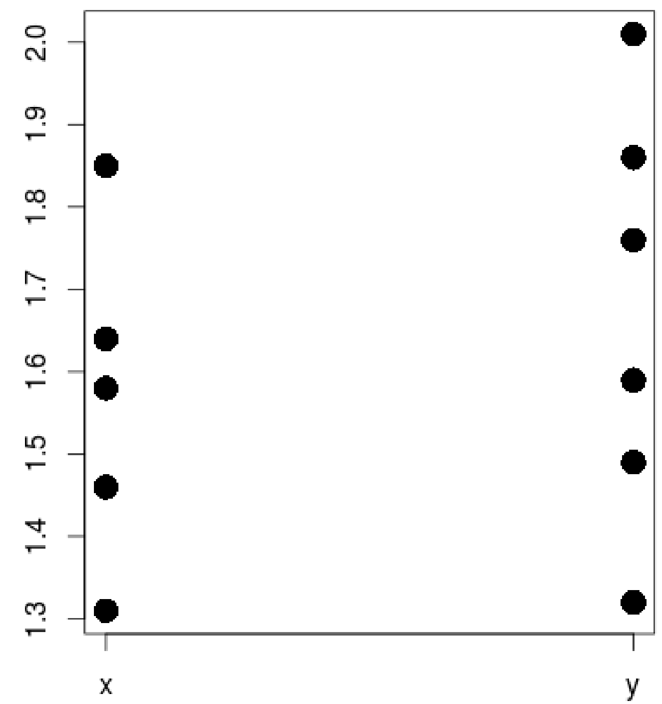
Test paramétrique ou non paramétrique ?

→ *Se tromper dans ce choix n'est pas forcément gênant*

- Test **paramétrique** : les hypothèses nulle et alternative du test portent sur un paramètre statistique (moyenne ou variance par exemple). Ces tests nécessitent généralement des conditions de validité (distribution normale des données par exemple).

- Test **non paramétrique** : un test non paramétrique porte globalement sur la répartition des données sans hypothèse sur leur distribution.

Quizz



- Données

```
X  1.31 1.46 1.85 1.58 1.64
Y  1.49 1.32 2.01 1.59 1.76 1.86
```

- Question : y a-t-il une différence entre x et y ?

☞ Dois-je utiliser un test pour données appariées ?

NON, les données ne sont pas appariées car les 2 échantillons sont de taille différente.

Sébastien Déjean - math.univ-t

R est d’accord !

```
x <- c(1.31,1.46,1.85,1.58,1.64)
y <- c(1.49,1.32,2.01,1.59,1.76,1.86)

wilcox.test(x,y, paired=TRUE)
Erreur dans wilcox.test.default(x, y,
paired = TRUE) :
  'x' et 'y' doivent avoir la même longueur
t.test(x,y, paired=TRUE)

t.test(x,y, paired=TRUE)
Erreur dans complete.cases(x, y) :
  les arguments n'ont pas tous la même
taille
```

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Données indépendantes !

☞ OK, mais quel test utiliser ?

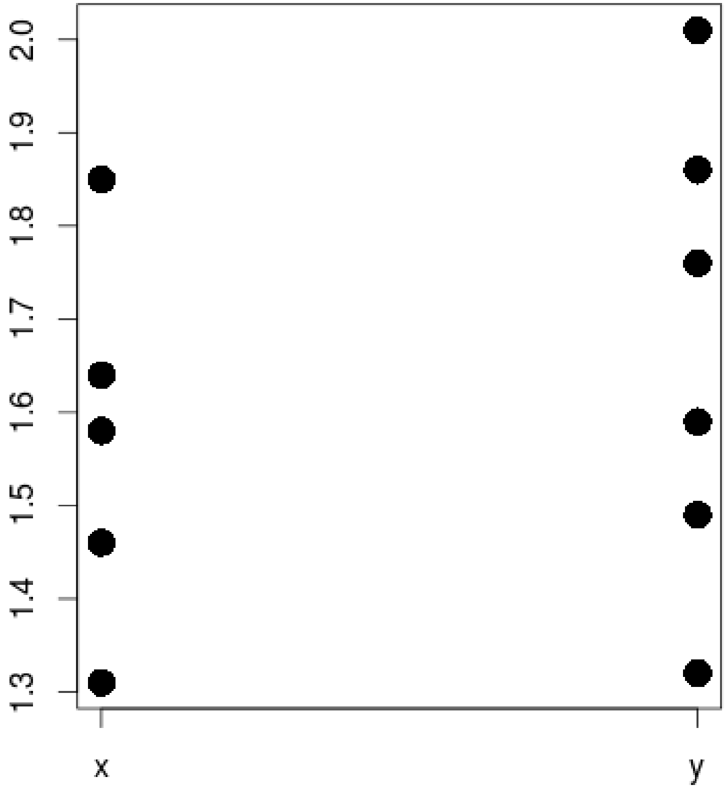
Le t-test ?

YES NO

Ne fais pas de test,
tu n’as pas assez de données !

Si tu insistes, fais plutôt un test de Wilcoxon

Une idée du résultat ?



☞ Combien vaut la p-value du test de Wilcoxon qui teste un éventuel décalage dans les 2 distributions dont sont issus les 2 échantillons ?

Et le résultat est...

```
> wilcox.test(x,y)
Wilcoxon rank sum test
data:  x and y
W = 10, p-value = 0.4286
alternative hypothesis: true location shift is not equal to 0
```

- Les 2 tests sont d'accord... ouf !

Et si ce n'était pas le cas ?

- Trouver pourquoi...

```
> t.test(x,y, var.equal=TRUE)
Two Sample t-test
data:  x and y
t = -0.7381, df = 9, p-value = 0.4792
alternative hypothesis: true difference in means is not equal to 0 95 percent
confidence interval:
 -0.4213783  0.2140450
sample estimates:
mean of x mean of y
 1.568000  1.671667
```

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Autre exemple

```
X  18 21 16 22 19 24 17 20 23 12

Y  22 25 17 24 18 29 20 23 21 16
```

☞ Dois-je faire un test pour données appariées ?

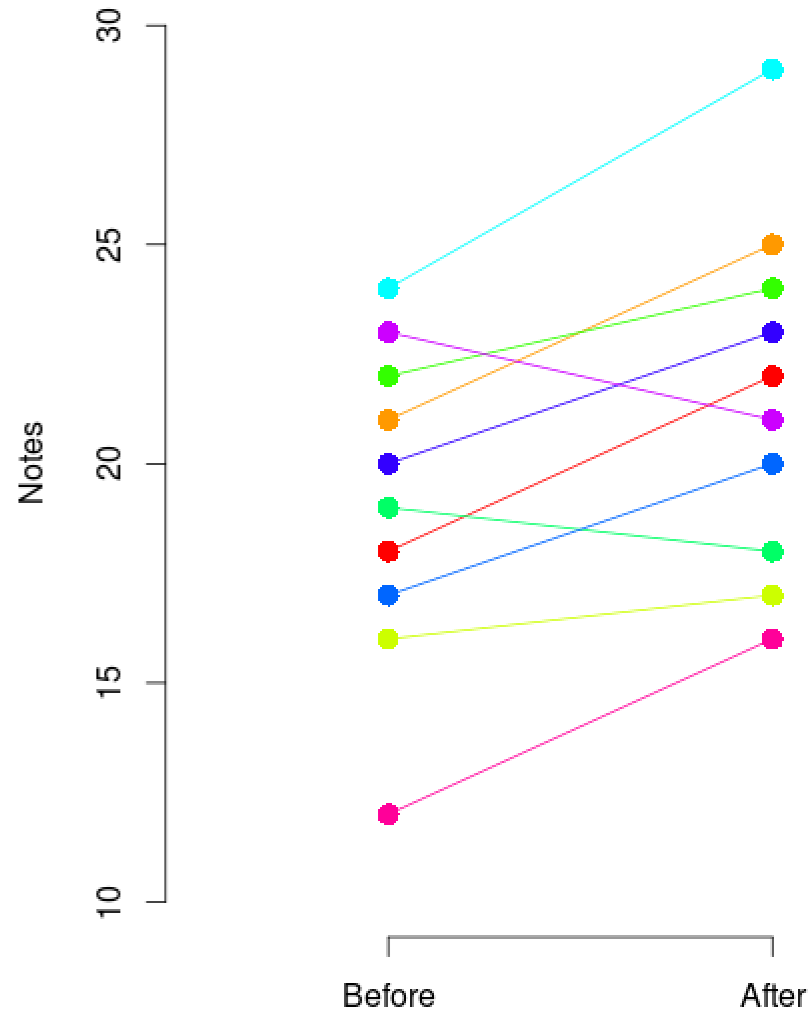
?

Le fait d'avoir 2 échantillons de tailles égales est une condition nécessaire mais pas [...]

[...] =   suffisante

Ce sont des données appariées

	Avant	Après
Louise	18	22
Léo	21	25
Emma	16	17
Gabriel	22	24
Chloé	19	18
Adam	24	29
Lola	17	20
Timéo	20	23
Inès	23	21
Raphaël	12	16



Resultats

> wilcox.test(x,y, **paired=TRUE**)
Wilcoxon signed rank test with continuity correction

data: x and y
V = 5, p-value = **0.02428**
alternative hypothesis: true location shift is not equal to 0

Les 2 tests sont d'accord (re-ouf !) pour décider que le décalage est significatif ou que la différence des moyennes n'est pas nulle.

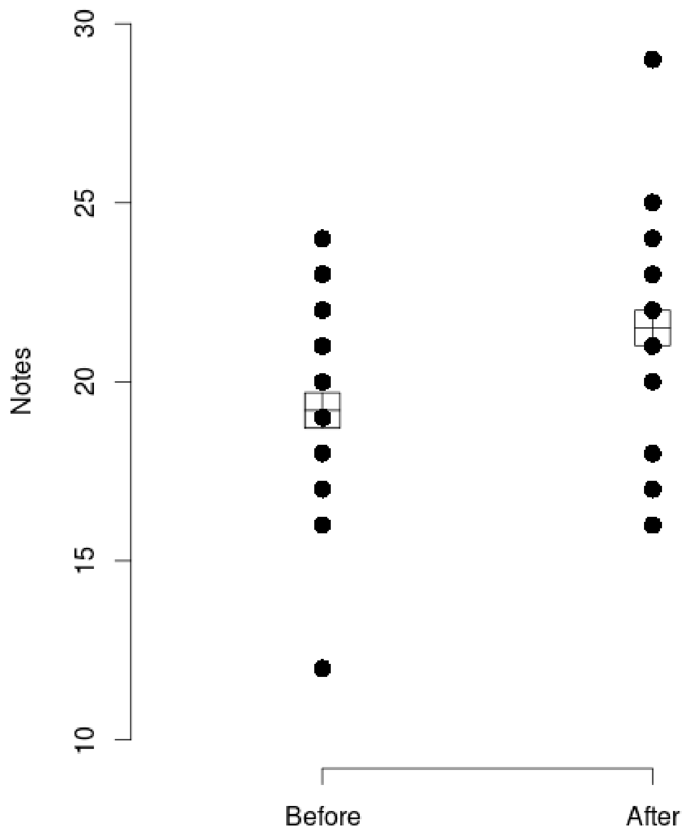
> t.test(x,y, **paired=TRUE**)
Paired t-test

data: x and y
t = -3.1461, df = 9, p-value = **0.01181**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.953766 -0.646234
sample estimates:
mean of the differences
-2.3

Et si je m'étais trompé

```
> wilcox.test(x,y, paired=FALSE)
Wilcoxon rank sum test with continuity correction

data:  x and y
W = 35, p-value = 0.2716
alternative hypothesis: true location shift is
not equal to 0
```

```
> t.test(x,y, paired=FALSE)
Two Sample t-test

data:  x and y
t = -1.3529, df = 18, p-value = 0.1928
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 -5.871567  1.271567
sample estimates:
mean of x mean of y
     19.2      21.5
```



Les 2 tests seraient d'accord pour décider que le décalage n'est pas significatif ou que la différence des moyennes est nulle. Ce qui serait une conclusion fausse si j'ai des données effectivement appariées (mais ce n'est pas de la faute des tests...)

Cas de 2 échantillons

« Comparaison de moyennes »

Type de test Type de données	Test paramétrique	Test non paramétrique
Données indépendantes	Test de Student pour 2 échantillons	Test de Wilcoxon-Mann-Whitney *Rank-sum test*
Données appariées	Test de Student pour 1 échantillon (sur la différence)	Test de Wilcoxon *Signed-rank test*

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Le test de Wilcoxon-Mann-Whitney

Exemple : la concentration d'un produit est mesurée sur 2 échantillons indépendants de taille respective n_1=5 et n_2=6. Voici les mesures :

```
Ech 1 : 1.31 1.46 1.85 1.58 1.64
Ech 2 : 1.49 1.32 2.01 1.59 1.76 1.86
```

Les distributions des données sont-elle significativement différentes dans les 2 populations dont sont issues les 2 échantillons?

Procédure du test de W-M-W

1) Classer toutes les observations par ordre croissant
2) Affecter son rang à chaque observation
3) Calculer la somme des rangs d'un échantillon

L'hypothèse d'absence de décalage entre les 2 distributions est rejetée si cette valeur W s'éloigne « trop » d'une valeur « moyenne ».

1)	1.31	1.32	1.46	1.49	1.58	1.59	1.64	1.76	1.85	1.86	2.01
2)	1	2	3	4	5	6	7	8	9	10	11

3) Somme des rangs en bleu : W = **25**

```
> x<-c(1.31,1.46,1.85,1.58,1.64)
> y<-c(1.49,1.32,2.01,1.59,1.76,1.86)
> wilcox.test(x,y)
        Wilcoxon rank sum test
data:  x and y
W = 10, p-value = 0.4286
alternative hypothesis: true location
shift is not equal to 0
```

La p-value obtenue ici (`0.4286`) indique qu'il n'y a pas de décalage (*shift*) entre les positions des 2 séries d'observations.

Le test de Student

On considère le même problème que précédemment et on applique un test de Student pour comparer la moyenne des 2 échantillons même si les conditions d'application sont plus que discutables.

Calculs

	1.31	1.49
	1.46	1.32
	1.85	2.01
	1.58	1.59
	1.64	1.76
		1.86
Moyenne	1.658	1.672
Variance	0.041	0.064
Var. Commune	0.054	

t = -0.738

Formules

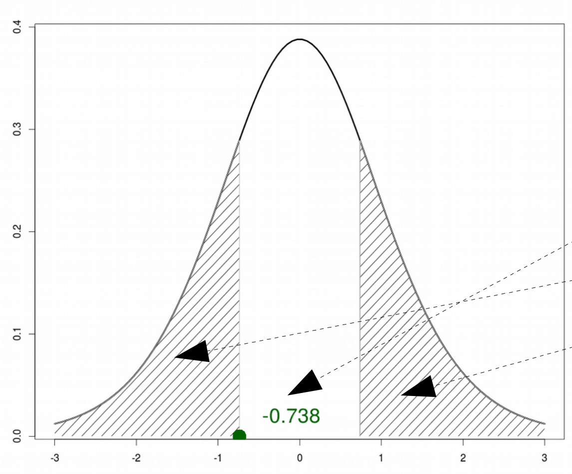
Sous H0, hypothèse d'égalité des moyennes, on a :

$$\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim Student(n_1+n_2-2)$$

Avec s^2 la variance commune aux 2 échantillons

$$s^2=\frac{(n_1-1)V_1+(n_2-1)V_2}{n_1+n_2-2}$$

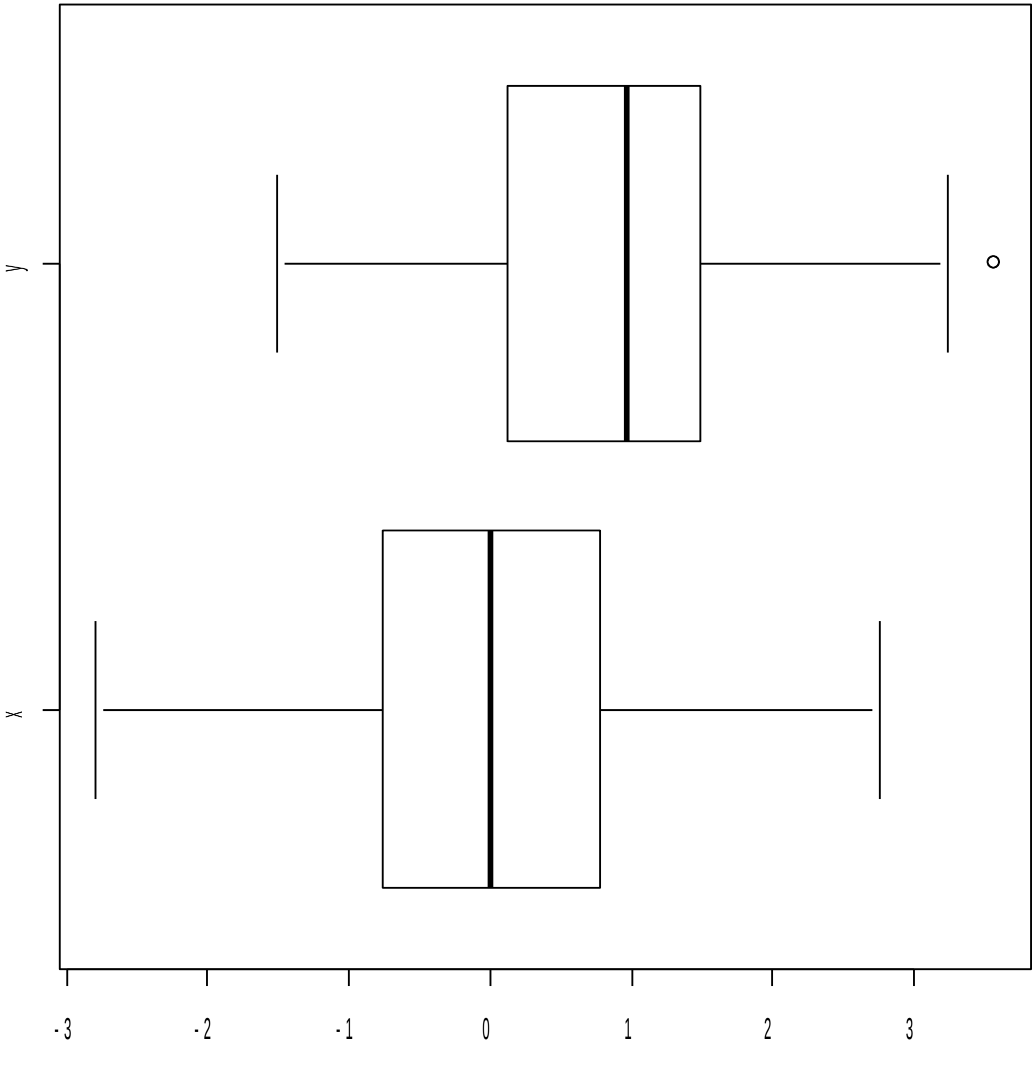
Densité de la loi de Student à 9 ddl

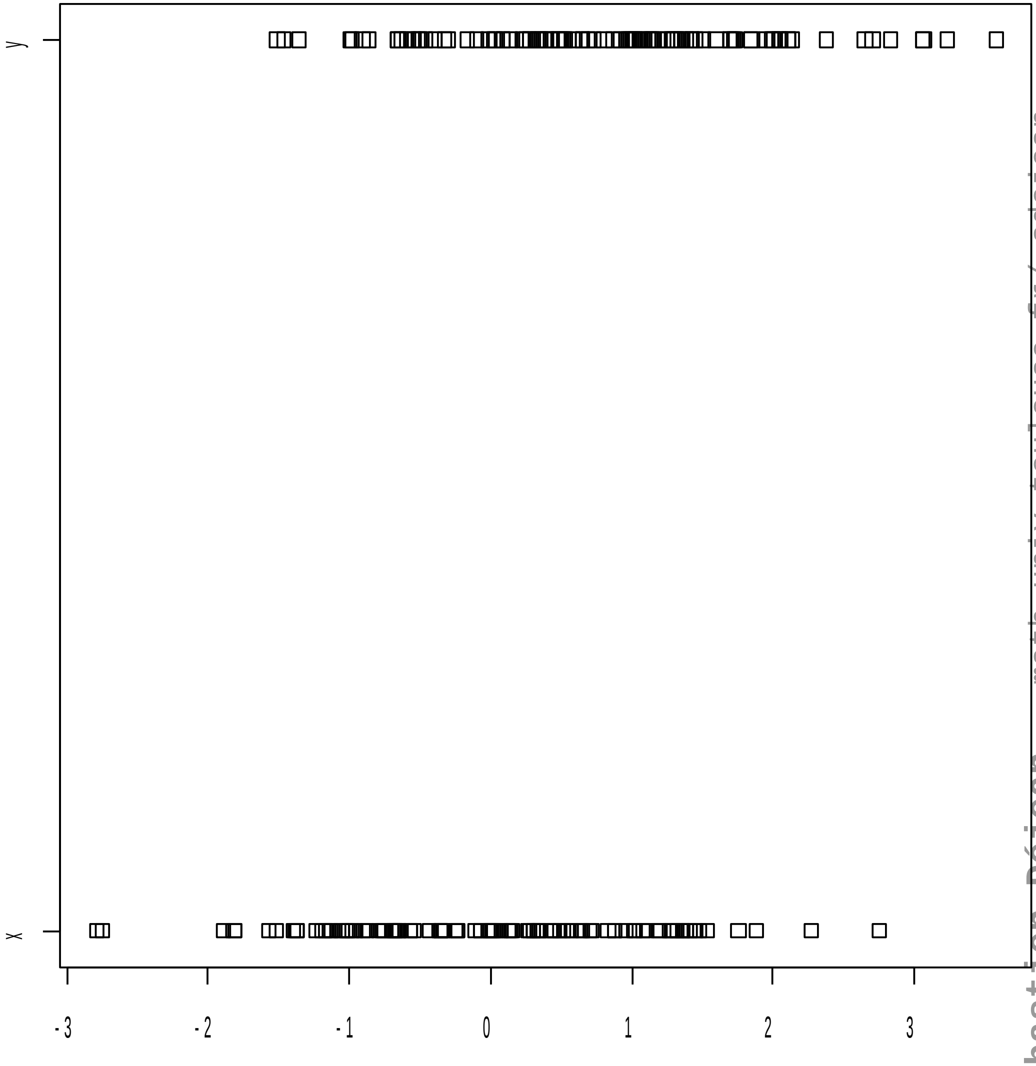



```
> t.test(x,y,var.equal=T)
        Two Sample t-test
data:  x and y
t = -0.7381, df = 9, p-value = 0.4792
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
 -0.4213783  0.2140450
sample estimates:
mean of x mean of y
 1.568000  1.671667
```

Mise en œuvre de quelques tests

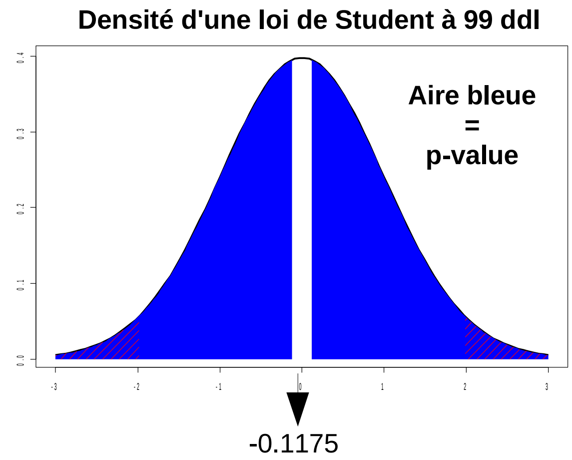
Données simulées : génération aléatoire selon une loi normale de 2 échantillons de longueur 100 : x ~ N(0,1) et y ~ N(1,1)





Mise en œuvre de quelques tests

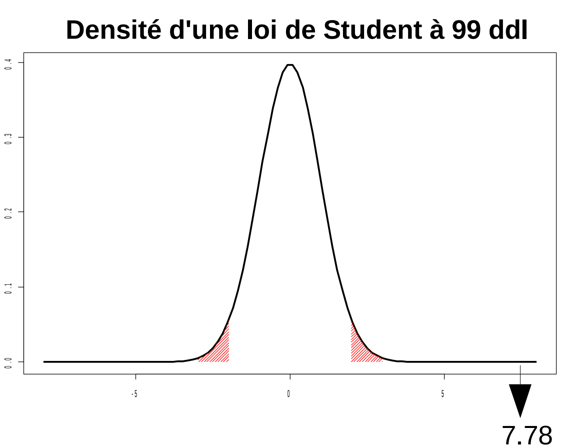
Test de Student pour un échantillon



On ne peut pas rejeter H0, la moyenne est probablement nulle.

```
        One Sample t-test

data:  x
t = -0.1175, df = 99, p-value = 0.9067
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.2239679  0.1989233
sample estimates:
  mean of x
-0.01252230
```

Rejet de H0 avec une très faible probabilité de se tromper.




```
        One Sample t-test

data:  y
t = 7.78, df = 99, p-value = 7.082e-12
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.6291375 1.0599157
sample estimates:
mean of x
0.8445266
```

Mise en œuvre de quelques tests

Test de Fisher d'égalité des variances

```
        F test to compare two variances

data:  x and y
F = 0.9637, num df = 99, denom df = 99, p-value = 0.8545
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6484291 1.4323091
sample estimates:
ratio of variances
         0.9637173
```

On ne peut pas rejeter H0, les 2 variances sont très probablement égales.

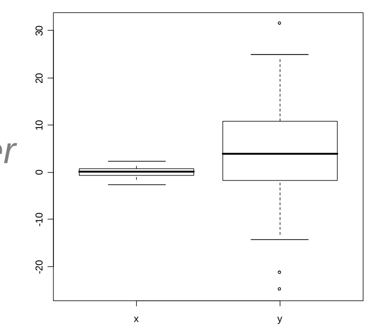
Mise en œuvre de quelques tests

Test de Student pour 2 échantillons

Two Sample t-test (*variances supposées égales*)

```
data:  x and y
t = -5.6342, df = 198, p-value = 5.982e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.1570238 -0.5570741
sample estimates:
  mean of x   mean of y
-0.01252230  0.84452662
```

Pour effectuer ce test, on suppose les 2 variances égales. Cela peut être contrôlé par un test de Fisher d'égalité des variances. Dans le cas ci-contre, la comparaison des moyennes n'a pas vraiment de sens.



On rejette H0, les 2 moyennes sont très probablement différentes.

Welch Two Sample t-test

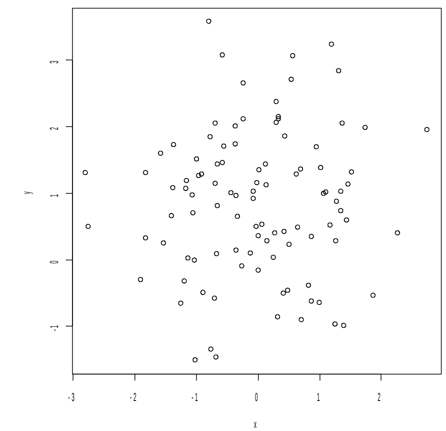
adaptation du test de Student sans l'hypothèse de variances égales

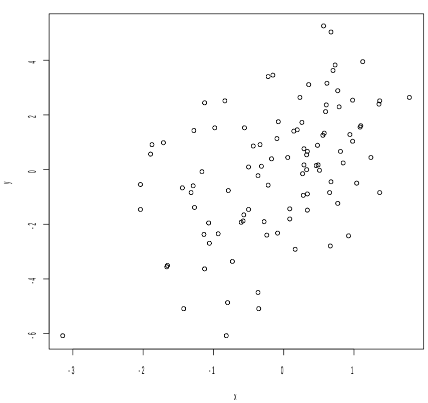
```
data:  x and y
t = -5.6342, df = 197.932, p-value = 5.985e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.1570244 -0.5570734
sample estimates:
  mean of x   mean of y
-0.01252230  0.84452662
```

Mise en œuvre de quelques tests

Test sur le coefficient de corrélation

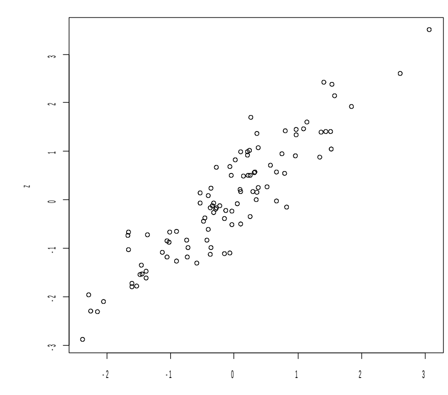
```
     Pearson's product-moment correlation
data:  x and y
t = 0.5464, df = 98, p-value = 0.586
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1428544  0.2488346
sample estimates: cor = 0.05511005
```





```
     Pearson's product-moment correlation
data:  x and z1
t = 5.5115, df = 98, p-value = 2.88e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3206572 0.6233025
sample estimates: cor = 0.486438
```

```
     Pearson's product-moment correlation
data:  x and z2
t = 22.3231, df = 98, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8748002 0.9415099
sample estimates: cor = 0.914144
```



Mise en œuvre de quelques tests

Test de normalité Kolmogorov-Smirnov

```
> ks.test(x,y) # x et Y sont-ils des échantillons d'une même distribution ?
        Two-sample Kolmogorov-Smirnov test
data:  x and y
D = 0.33, p-value = 3.729e-05
alternative hypothesis: two-sided
```

Probablement pas, avec une faible chance de se tromper

```
> ks.test(x,"pnorm") # x est-il un échantillon d'une loi normale N(0,1) ?
        One-sample Kolmogorov-Smirnov test
data:  x
D = 0.0718, p-value = 0.6803
alternative hypothesis: two-sided
```

Les données ne permettent pas de dire le contraire.

```
> ks.test(y,"pnorm") # y est-il un échantillon d'une loi normale N(0,1) ?
        One-sample Kolmogorov-Smirnov test
data:  y
D = 0.3408, p-value = 1.641e-10
alternative hypothesis: two-sided
```

Probablement pas, avec une très faible chance de se tromper.

```
> ks.test(y,"pnorm",1) y est-il un échantillon d'une loi normale N(1,1) ?
        One-sample Kolmogorov-Smirnov test
data:  y
D = 0.0923, p-value = 0.3614
alternative hypothesis: two-sided
```

Les données ne permettent pas de dire le contraire.

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

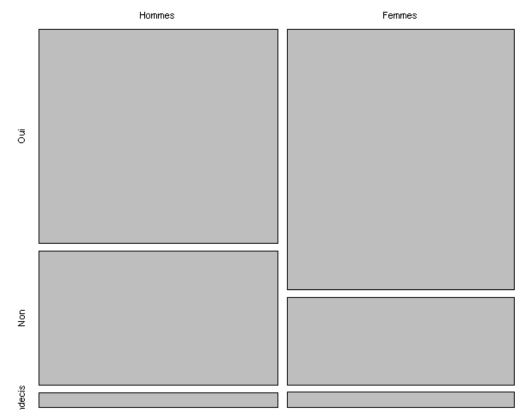
Quelques tests classiques

Test du χ^2 d'indépendance

- **Données** : effectifs recueillies dans une table de contingence (tableau croisé pour 2 variables qualitatives)
- **Question** : les 2 variables qualitatives sont-elles indépendantes ?

- **Exemple** : 1250 personnes ont répondu à la question « Êtes-vous satisfaits des programmes TV ? ». On souhaite savoir si la satisfaction dépend du sexe.

	OUI	NON	Indécis	**Somme**
Hommes	378	237	26	**641**
Femmes	438	146	25	**609**
Somme	**816**	**383**	**51**	**1250**



Hypothèse H0 : Satisfaction et sexe sont indépendants

Effectifs théoriques sous l'hypothèse d'indépendance
*(effectif d'une case = effectif de la ligne * effectif de la colonne / effectif total)*

	OUI	NON	Indécis
Hommes	418	196	26
Femmes	398	187	25

Statistique de test : $\chi^2_{obs} = \Sigma\,(\text{Obs-Théo})^2 / \text{Théo}$

Reflète l'écart entre les données observées et les effectifs théoriques en cas d'indépendance

```
Pearson's Chi-squared test
data:  data.chisq
X-squared = 25.2501, df = 2, p-value = 3.289e-06
```

Les 2 caractères ne semblent pas indépendants.

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

Test et visualisation des données

Y a-t-il un effet du facteur WT/mut sur la variable Vx ?

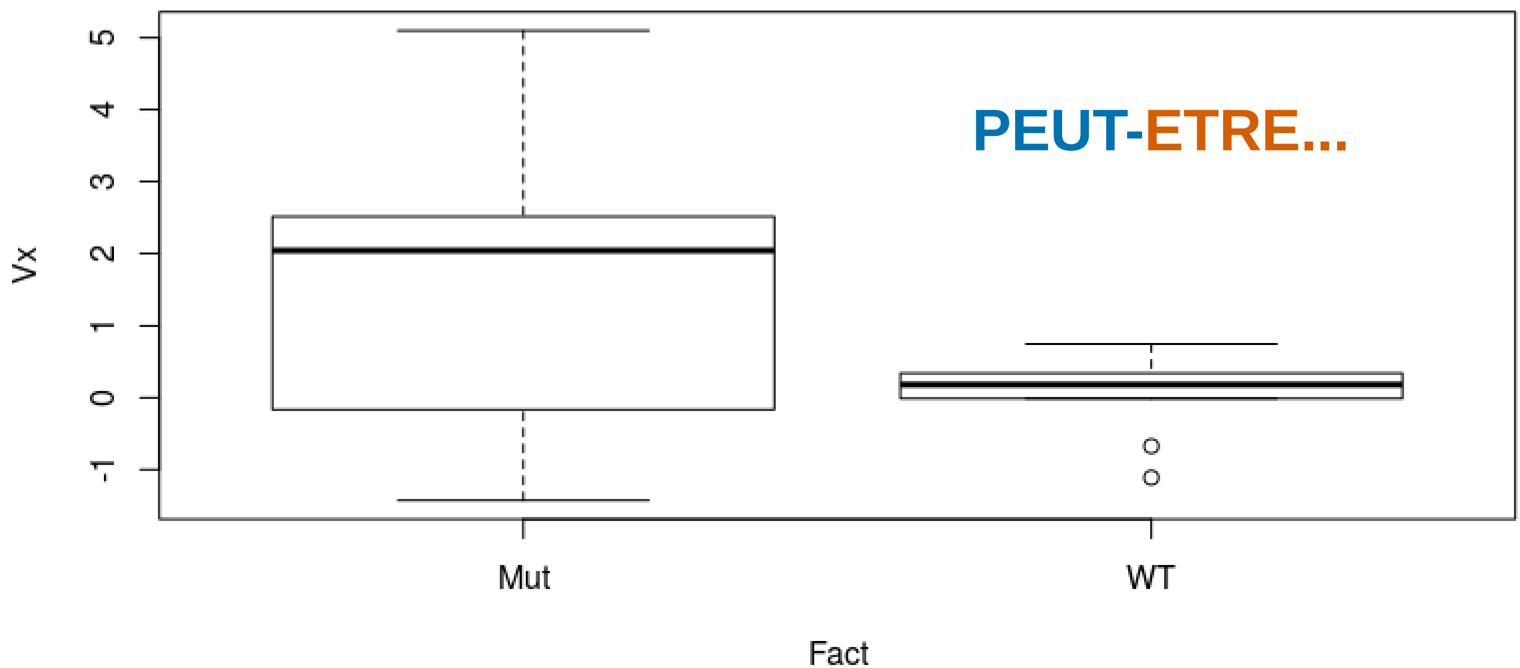
	Vx	Fact
1	-1.11	WT
2	-0.01	WT
3	0.20	WT
4	-0.67	WT
5	0.05	WT
6	0.17	WT
7	0.34	WT
8	0.24	WT
9	0.54	WT
10	0.75	WT
11	2.51	Mut
12	-0.43	Mut
13	2.09	Mut
14	2.21	Mut
15	4.36	Mut
16	-0.17	Mut
17	-1.43	Mut
18	1.99	Mut
19	0.50	Mut
20	5.10	Mut

```
> t.test(Vx~Fact)
        Welch Two Sample t-test
data:  Vx by Fact
t = 2.3854, df = 10.269, p-value = 0.03765
```

OUI !

```
> wilcox.test(Vx~Fact)
    Wilcoxon rank sum test
data:  Vx by Fact
W = 72, p-value = 0.1051
```

NON !?



Test et visualisation des données

Données: 1 facteur (WT/Mut), 2 variables quantitatives

	factor	Vx	Vy
1	WT	2.0	2.00
2	Mut	3.0	2.50
3	WT	4.5	3.50
4	Mut	5.0	3.25
5	Mut	5.5	3.30
6	WT	6.0	4.30
7	Mut	7.0	4.20
8	WT	8.0	5.10
9	Mut	8.5	4.80
10	Mut	9.0	5.00
11	WT	10.0	6.00
12	WT	11.0	6.50

Le facteur influence-t-il Vx et Vy?

Vx

```
> t.test(Vx~fact)
        Welch Two Sample t-test
data:  Vx by fact
t = -0.34852, df = 8.7078, p-value = 0.7357
```

NO

```
> wilcox.test(Vx~fact)

     Wilcoxon rank sum test

data:  Vx by fact
W = 16, p-value = 0.8182
```

NO

Vy

```
> t.test(Vy~fact)

     Welch Two Sample t-test

data:  Vy by fact
t = -0.91815, df = 8.1062, p-value = 0.385
```

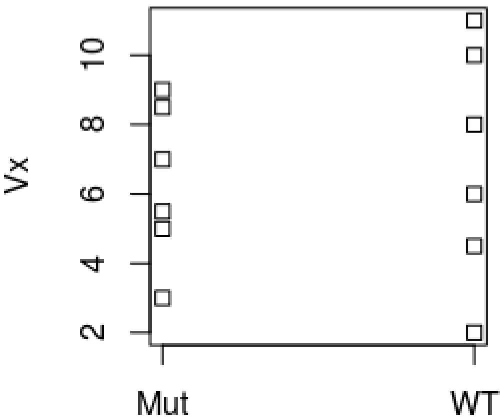
NO

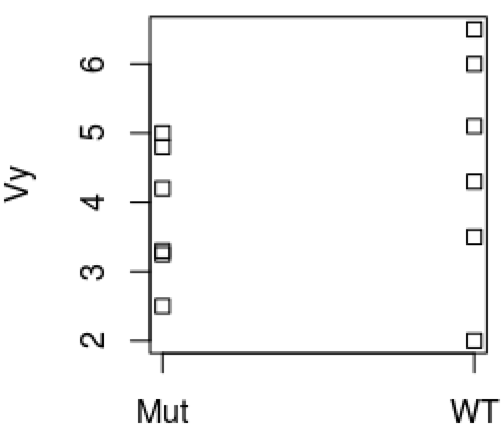
```
> wilcox.test(Vy~fact)

     Wilcoxon rank sum test

data:  Vy by fact
W = 11, p-value = 0.3095
```

NO

Test et visualisation des données

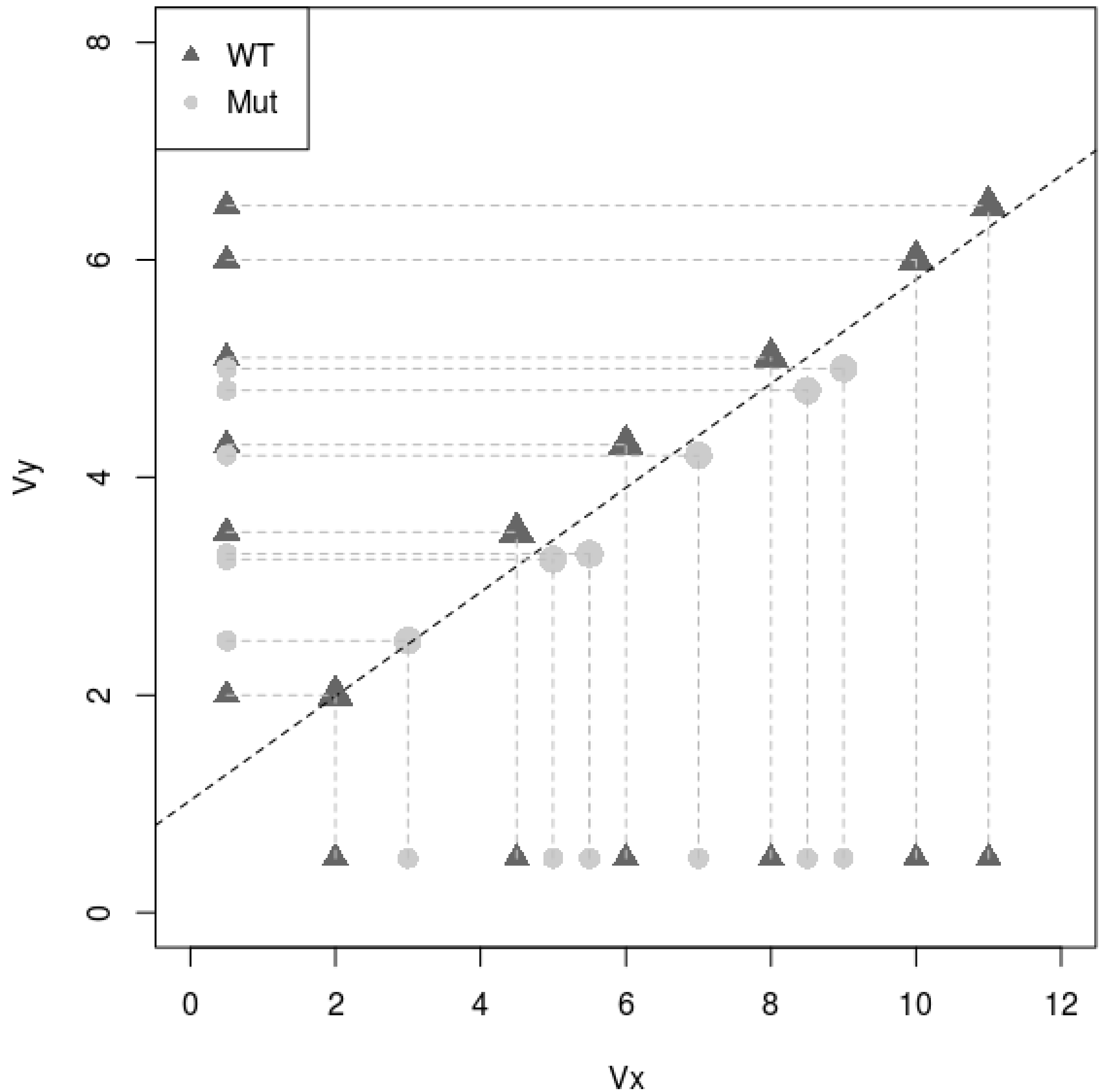
Les *stripcharts* semblent d'accord





	factor	Vx	Vy
1	WT	2.0	2.00
2	Mut	3.0	2.50
3	WT	4.5	3.50
4	Mut	5.0	3.25
5	Mut	5.5	3.30
6	WT	6.0	4.30
7	Mut	7.0	4.20
8	WT	8.0	5.10
9	Mut	8.5	4.80
10	Mut	9.0	5.00
11	WT	10.0	6.00
12	WT	11.0	6.50

Et que dit un nuage de points 2D ?



ANOVA 2 facteurs

Id	genotype	treatment	X1	X2	X3	X4
1	WT	CTRL	10.4	10.4	10.1	10.1
2	WT	CTRL	10.5	10.5	10.2	10.2
3	WT	CTRL	9.6	9.6	9.8	9.8
4	WT	CTRL	9.5	9.5	9.9	9.9
5	WT	CTRL	10.0	10.0	10.0	10.0
6	WT	Treat	6.4	6.4	5.1	8.1
7	WT	Treat	6.5	6.5	5.2	8.2
8	WT	Treat	5.6	5.6	4.8	7.8
9	WT	Treat	5.8	5.8	4.9	8.9
10	WT	Treat	6.0	6.0	5.0	8.0
11	Mut	CTRL	12.1	10.3	5.1	5.1
12	Mut	CTRL	12.2	10.6	5.2	5.2
13	Mut	CTRL	11.8	9.7	4.8	4.8
14	Mut	CTRL	11.9	9.4	4.9	4.9
15	Mut	CTRL	12.0	10.0	5.0	5.0
16	Mut	Treat	8.1	6.3	10.1	10.1
17	Mut	Treat	8.2	6.6	10.2	10.2
18	Mut	Treat	7.8	5.5	9.8	9.8
19	Mut	Treat	7.9	5.9	9.9	9.9
20	Mut	Treat	8.0	6.0	10.0	10.0

Sébastien Déjean - math.univ-toulouse.fr/~sdejean

ANOVA table

X1	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	1	19.40	19.4	192.600	2.44e-10	***
treatment	1	78.80	78.8	782.179	5.17e-15	***
genotype:treatment	1	0.00	0.0	0.045	0.835	

X2	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	1	0.00	0.00	0.0	1	
treatment	1	77.62	77.62	413.4	7.42e-13	***
genotype:treatment	1	0.00	0.00	0.0	1	

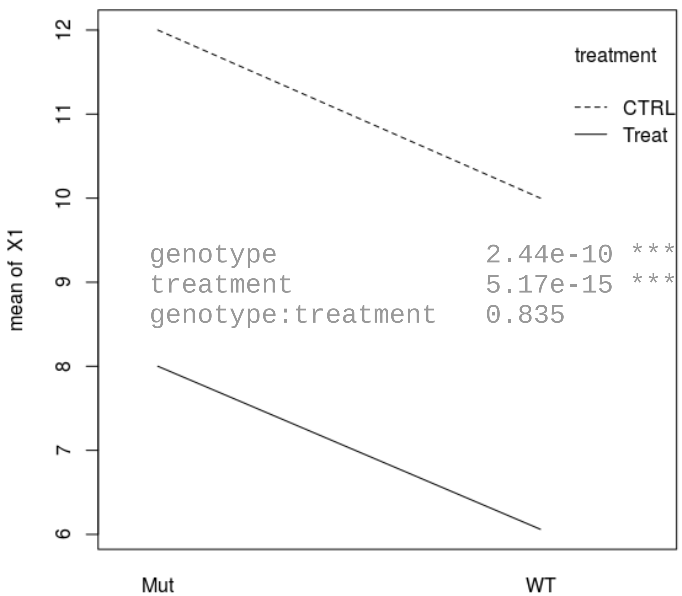
X3	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	1	0.0	0.00	0	1	
treatment	1	0.0	0.00	0	1	
genotype:treatment	1	125.0	125.00	5000	<2e-16	***

X4	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	1	12.8	12.80	204.8	1.54e-10	***
treatment	1	12.8	12.80	204.8	1.54e-10	***
genotype:treatment	1	57.8	57.80	924.8	1.38e-15	***

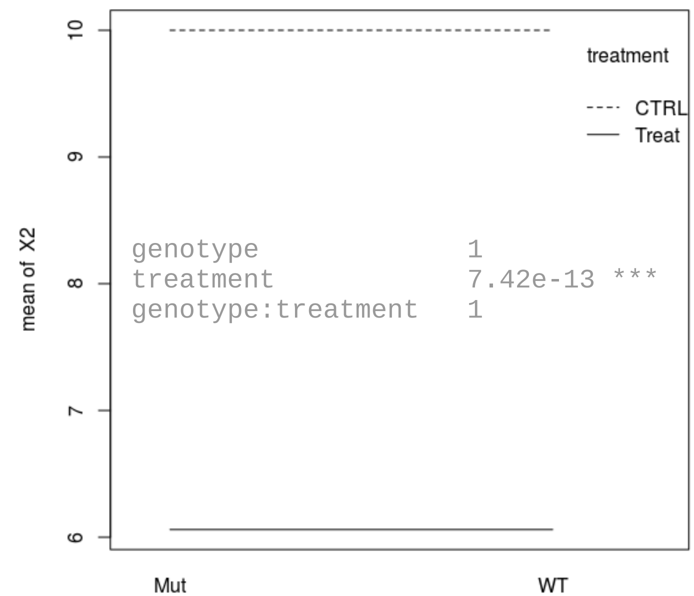
Graphiques d'interaction

psychstat3.missouristate.edu/Documents/MultiBook3/Mlt08.htm

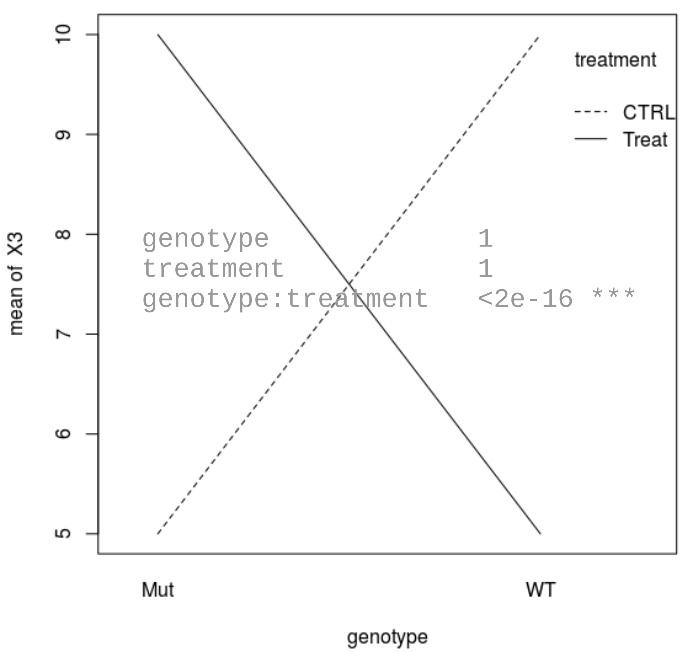
X1



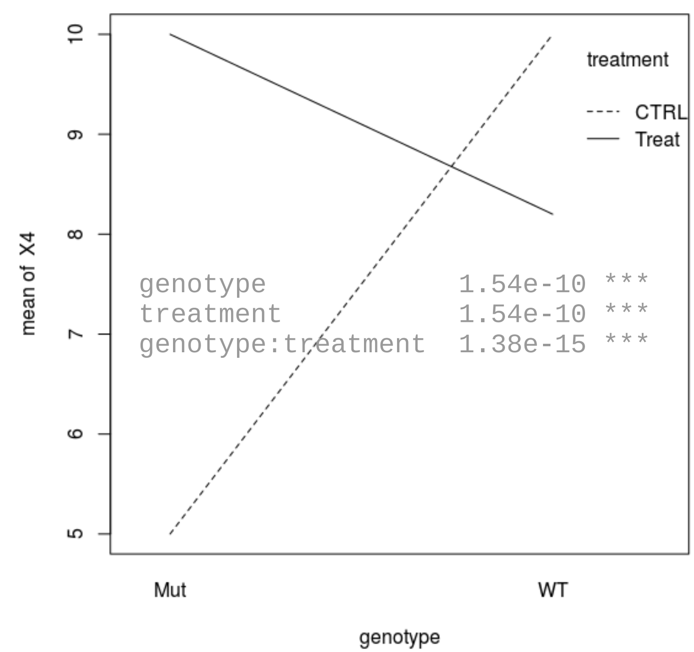
X2



X3



X4

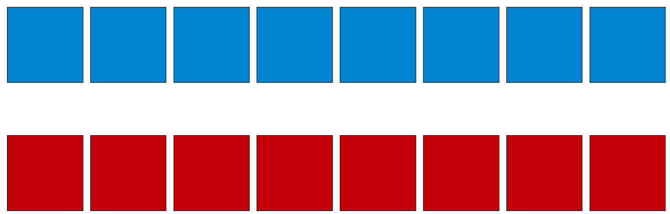


Planification expérimentale

2 conditions à l'étude : Contrôle / Traitement

Jour 1

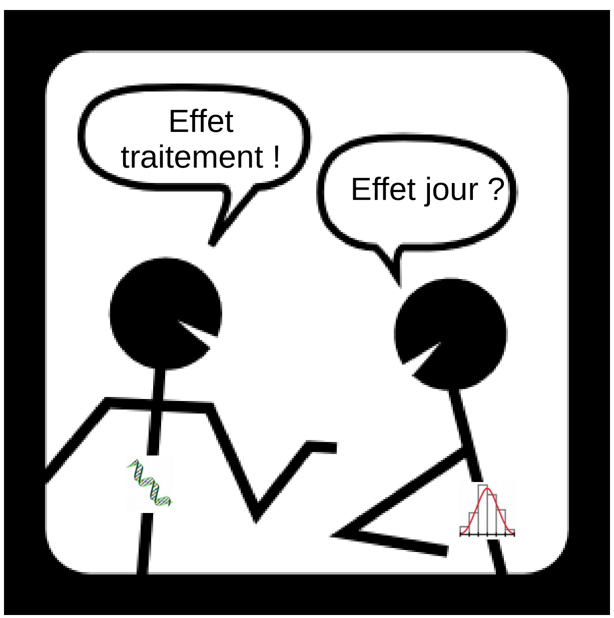
Jour 2



8 échantillons Contrôle

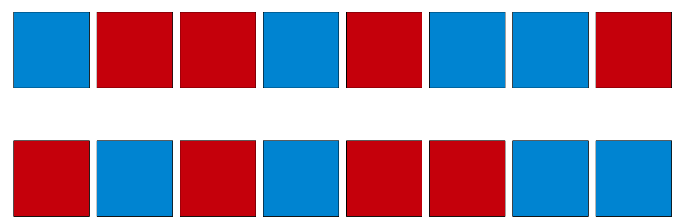
8 échantillons Traitement

Test statistique : les moyennes des 2 séries sont significativement différentes ! *



Jour 1

Jour 2



Randomisation

To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.

R.A. Fisher

Conclusion

Après nous avoir convaincu de leur objectivité fondamentale, il ne reste plus aux chiffres qu'à nous amener doucement à penser qu'ils en déterminent le monopole. Dès lors, une forme de hiérarchie gagne l'argumentation et le raisonnement : contenir quelques chiffres qualifie automatiquement votre discours, même si personne ne prend la peine de comprendre vraiment ce qu'ils signifient, voire même s'ils sont sans rapport avec le sujet traité ! A contrario, de ce fait, toute argumentation purement textuelle semble dépréciée […] comme si le raisonnement et la rigueur ne pouvait exister hors des chiffres.

Lorsqu'on invoque les mathématiques pour garantir des résultats qui ne dépendent que des choix faits au départ, on trompe le lecteur et d'une certaine façon, on contraint cette discipline scientifique à blanchir des hypothèses douteuses. Les mathématiques sont alors prises en otage, ni plus ni moins. […] L'outil mathématique fait son travail, que l'hypothèse soit plausible ou non, qu'elle soit légitime ou non. En aucun cas, il n'assume la garantie des hypothèses sur lesquelles on le fait travailler. Un outil reste un outil.