ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Instituts thématiques Inserm

Institut national de la santé et de la recherche médicale

Les statistiques : pour quoi faire ?

Pr Gabriel Thabut

Service de Pneumologie et transplantation pulmonaire

& INSERM U1152

Hôpital Bichat, Université Paris 7

Plan

- Statistiques descriptives
- Fluctuations d'échantillonage
  - Intervalle de confiance
  - Principes des test
- Notion de risque
  - Différence p et risque α
- L'analyse multivariée

Les statistiques descriptives

Statistiques descriptives

N=1000

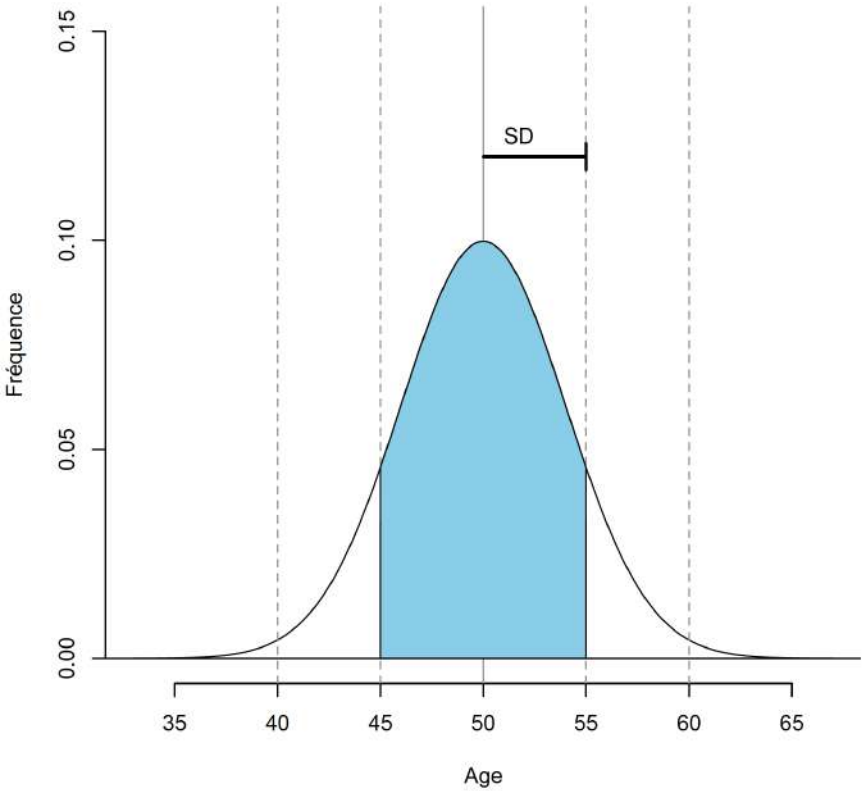
- **Femmes : 530 (53%)**
- **Age : 50 ans (DS:5)**

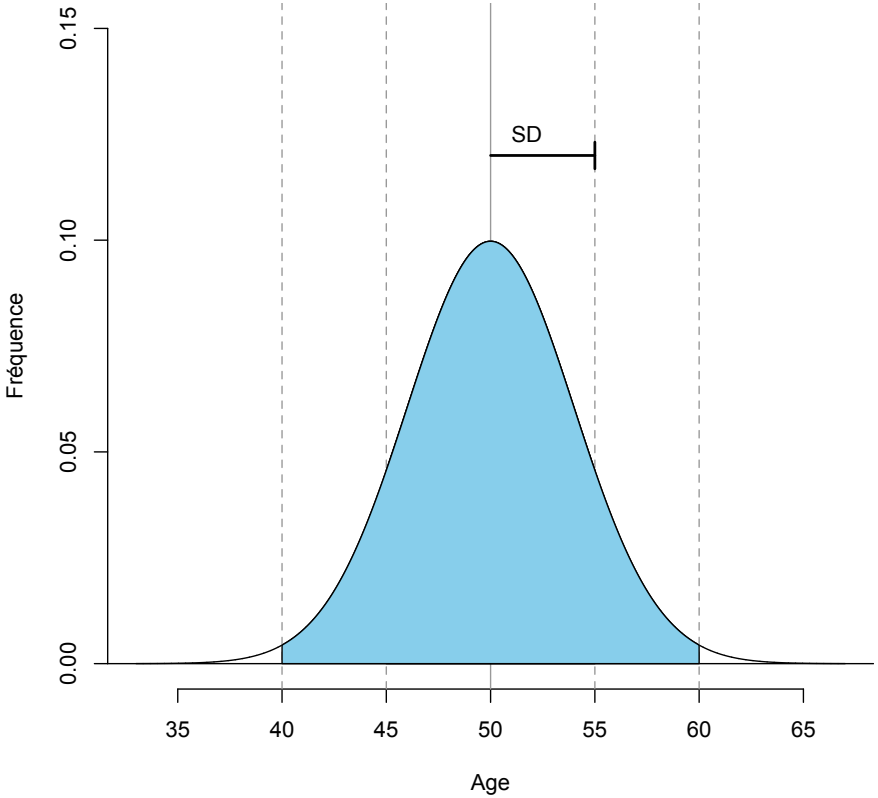
Statistiques descriptives

Interprétation de la DS si distribution normale

Age : 50 (DS=5)

- 68% des individus ont un âge compris entre moyenne ± 1 DS
- 96% des individus ont un âge compris entre moyenne ± 2 DS

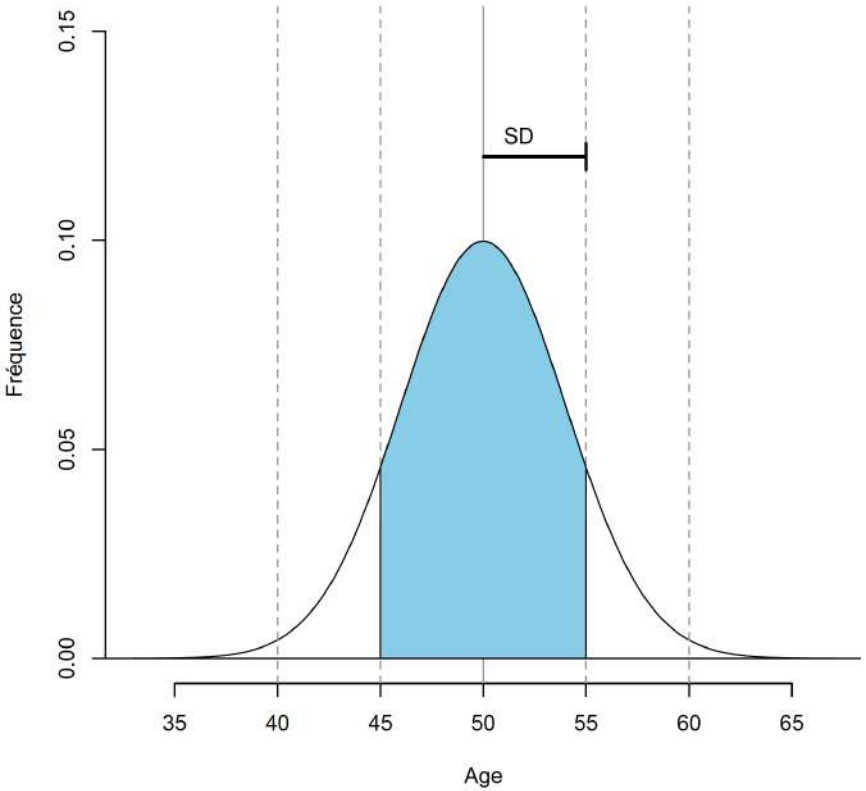




Statistiques descriptives

Interprétation de la DS si distribution normale

Age : 50 (DS=5)



La plupart des caractéristiques démographiques : âge, taille poids suivent une distribution normale

Calcul de la variance et de la déviation standard

Age
50
55
45
52
48

Moyenne : (50 + 55 + 45 + 52 + 48)/5 = 50

$$\sigma^2 = \frac{1}{N}\sum_i^n (x_i - \bar{x})^2$$

Variance = $[(50-50)^2 + (55-50)^2 + (45-50)^2 + (52-50)^2 + (48-50)^2]/5$
= 14,5

Déviation standard = racine carrée de la variance = 3,81

Parameter	Treatment	Baseline	
		n	Mean (SD)
TC (mg/dL)	DTG	356	163.8 (34)
	RAL	366	160.3 (38)
LDL-C (mg/dL)	DTG	354	96.8 (30)
	RAL	363	93.4 (32)
HDL-C (mg/dL)	DTG	356	44.4 (12)
	RAL	366	44.3 (13)
TC:HDL ratio	DTG	356	3.9 (1)
	RAL	366	3.8 (1)
Triglycerides (mg/dL)	DTG	357	113.8 (64)
	RAL	366	115.9 (82)

Interprétation

- Distribution manifestement non normale

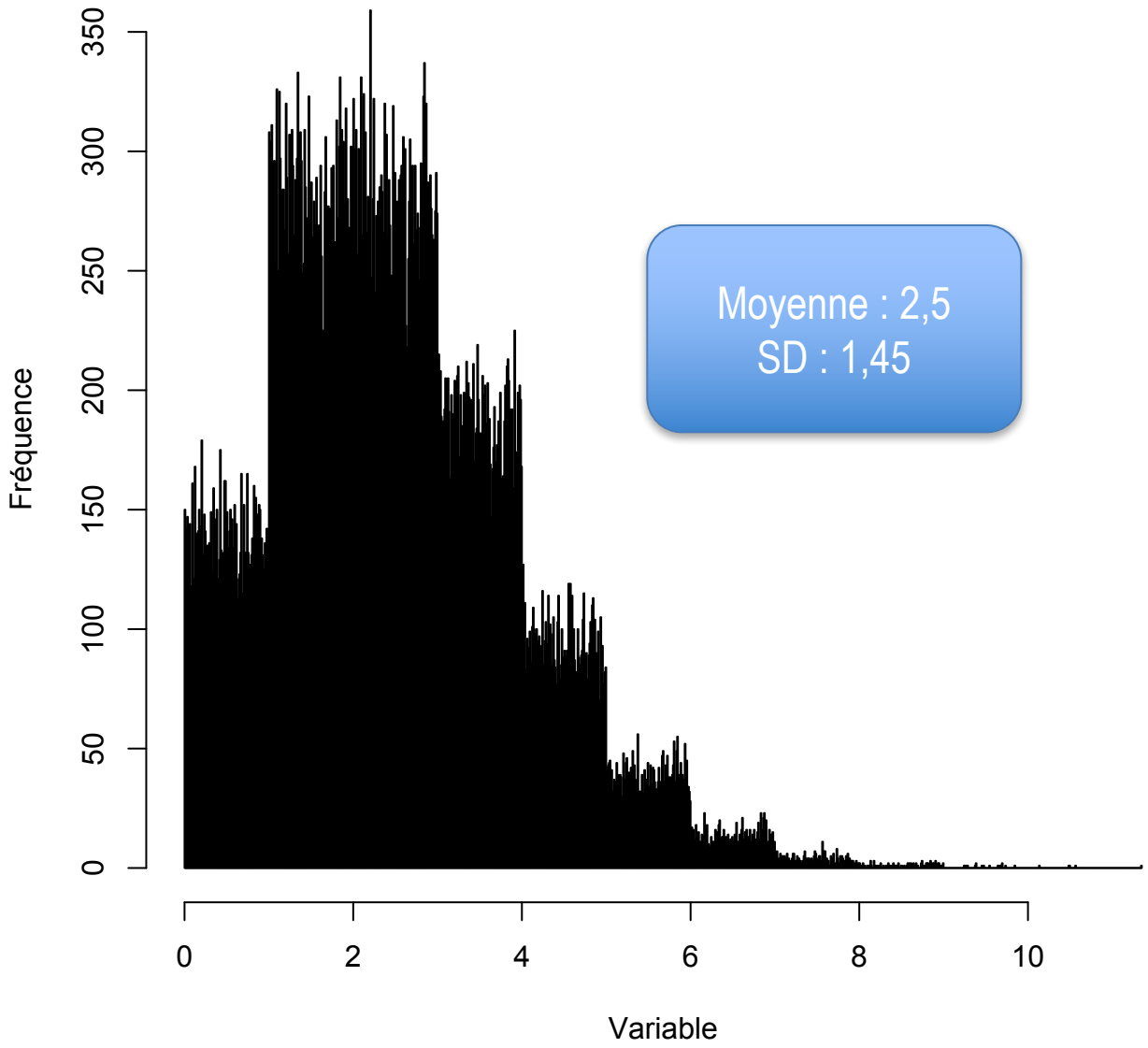
Quercia R, *Clin Drug Investig*, 2015

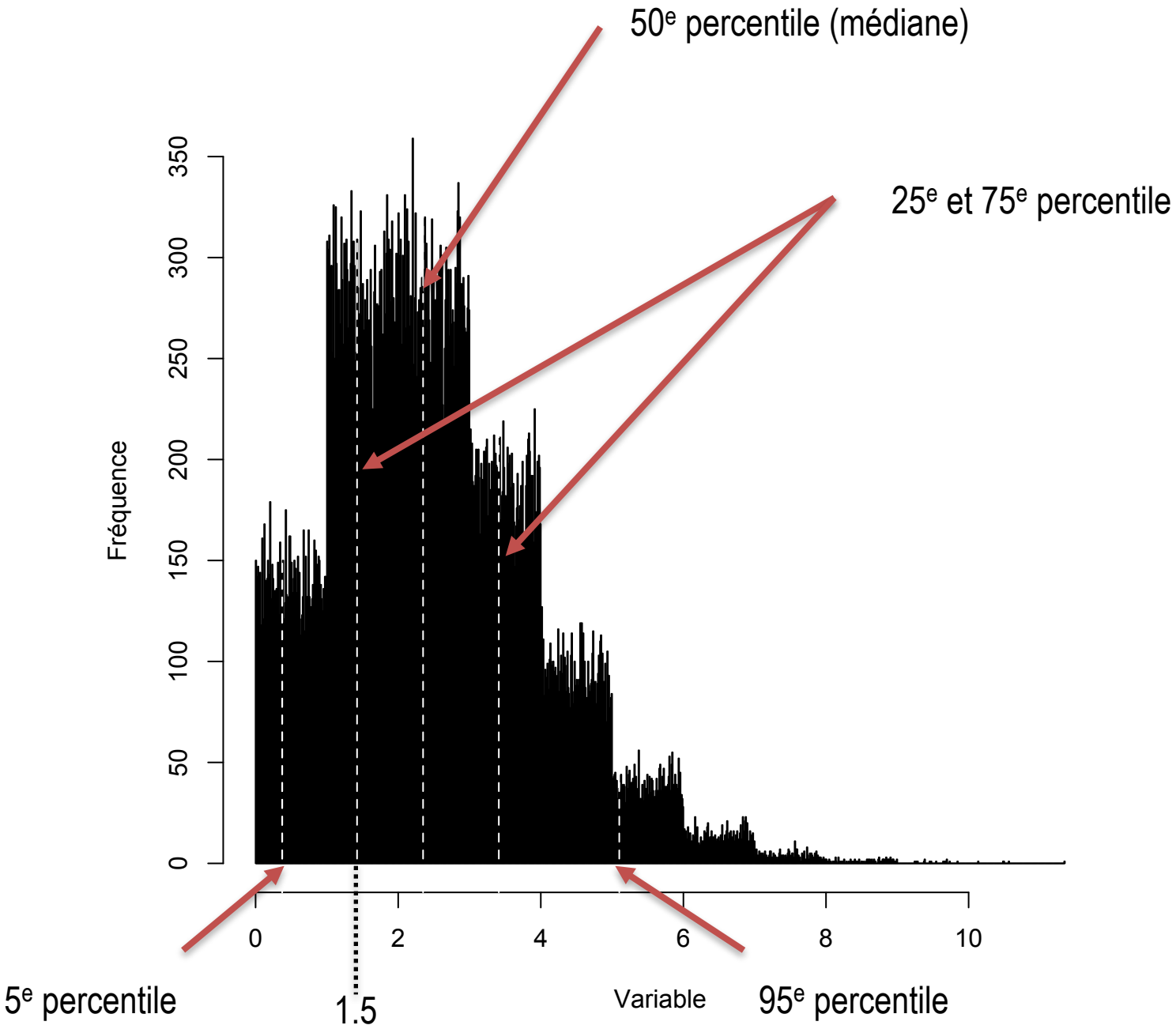
Parameter	Treatment	Baseline	
		n	Mean (SD)
TC (mg/dL)	DTG	356	163.8 (34)
	RAL	366	160.3 (38)
LDL-C (mg/dL)	DTG	354	96.8 (30)
	RAL	363	93.4 (32)
HDL-C (mg/dL)	DTG	356	44.4 (12)
	RAL	366	44.3 (13)
TC:HDL ratio	DTG	356	3.9 (1)
	RAL	366	3.8 (1)
Triglycerides (mg/dL)	DTG	357	113.8 (64)
	RAL	366	115.9 (82)

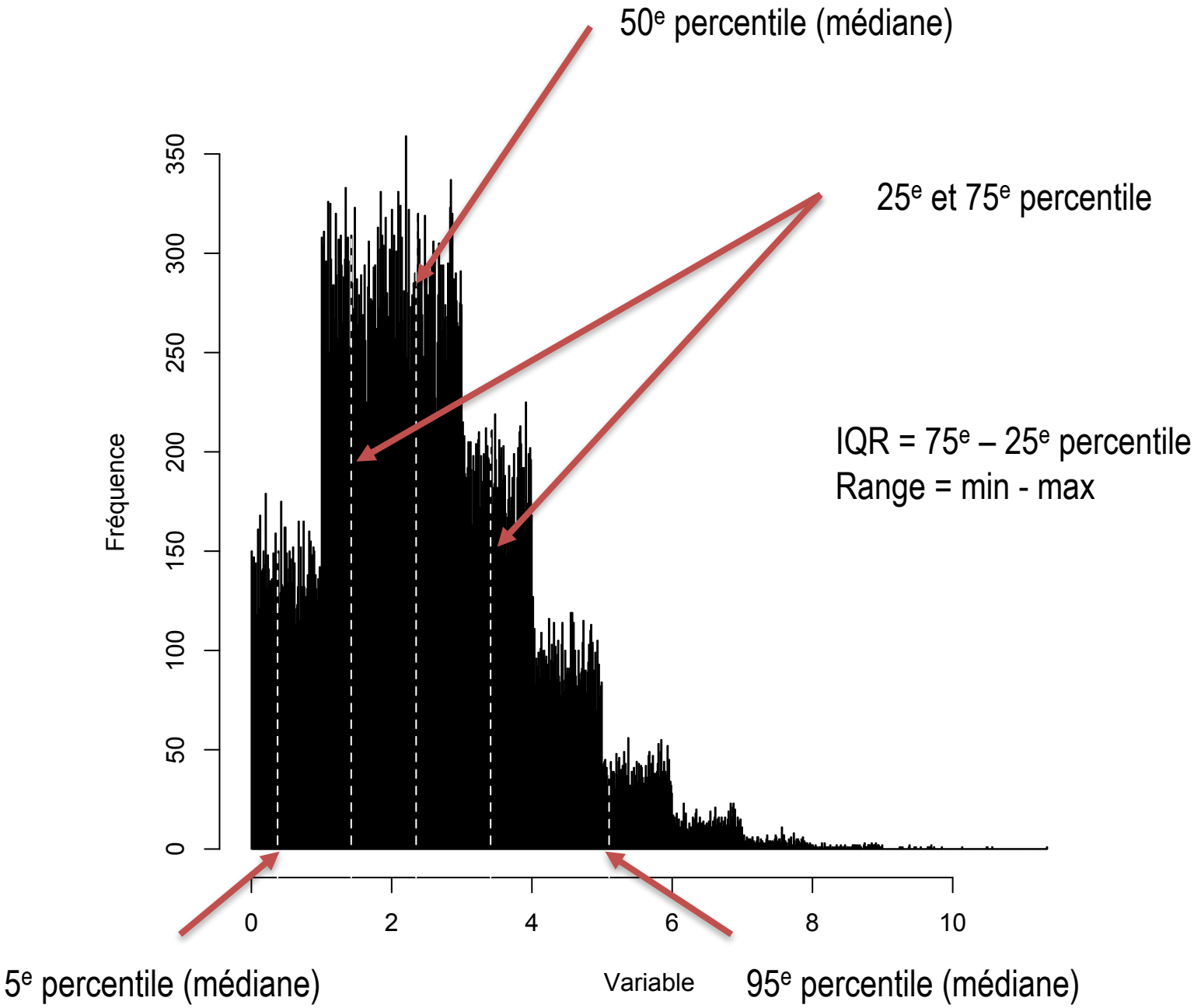
La plupart des paramètres biologiques **ne suivent pas** une distribution normale

Quercia R, *Clin Drug Investig*, 2015

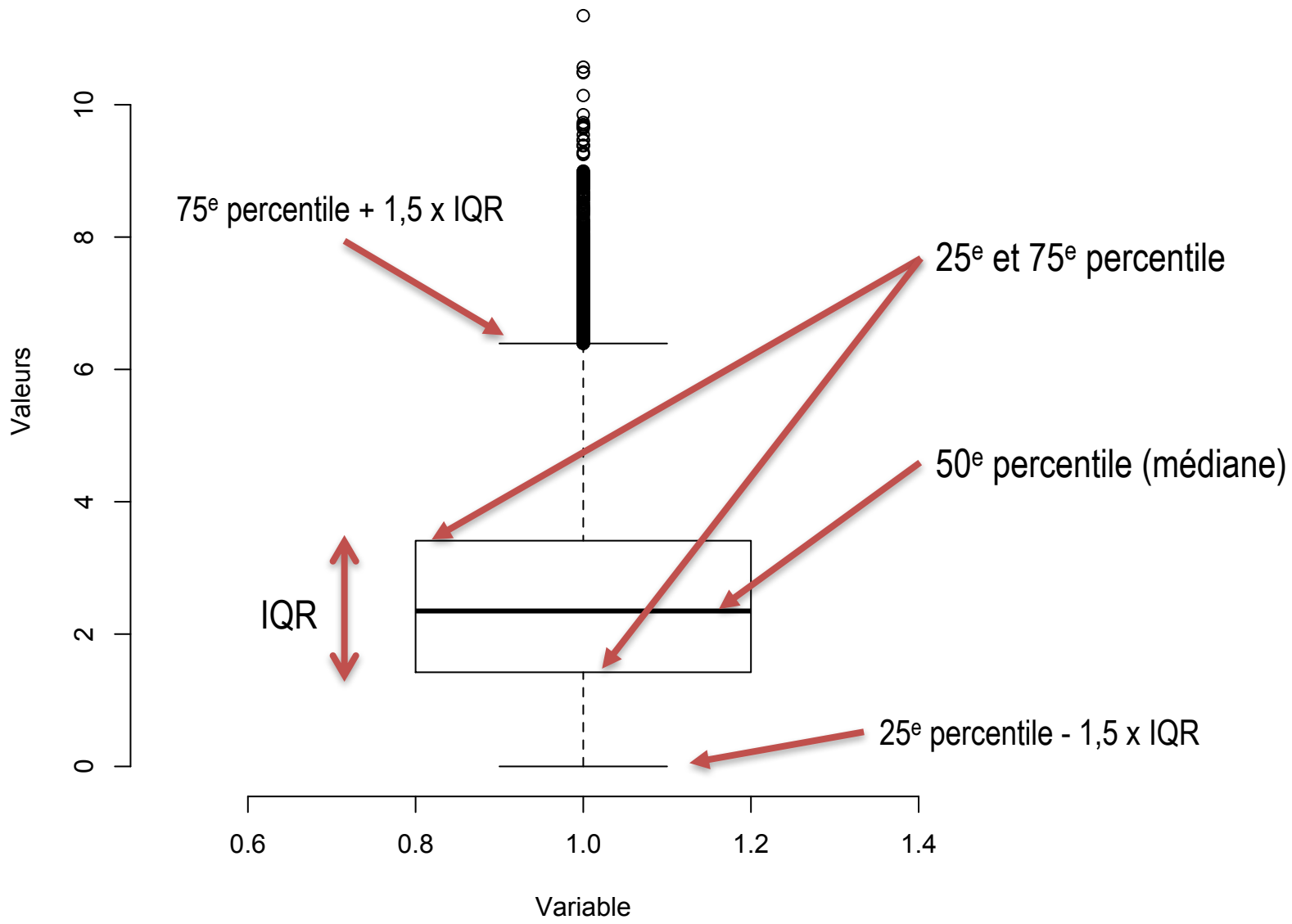
Statistiques descriptives : distribution non normale




50e percentile (médiane)
25e et 75e percentile
Fréquence
0
50
100
150
200
250
300
350
0
2
4
6
8
10
Variable
5e percentile
1.5
95e percentile


50e percentile (médiane)
25e et 75e percentile
IQR = 75e – 25e percentile
Range = min - max
Fréquence
0
50
100
150
200
250
300
350
0
2
4
6
8
10
Variable
5e percentile (médiane)
95e percentile (médiane)

Statistiques descriptives : autres paramètres



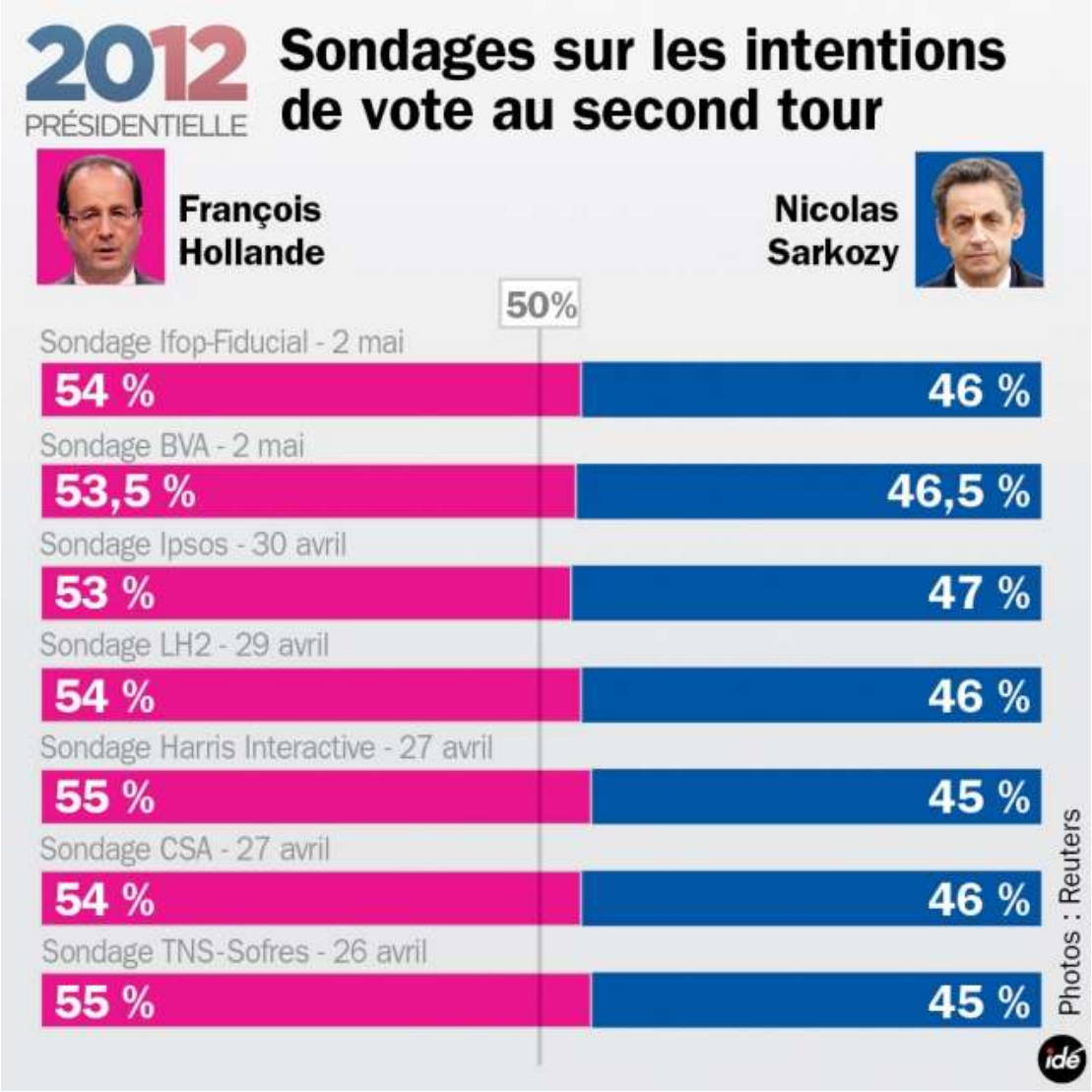
Rapporte médiane, 25e – 75e percentile et range

	OT (n=291)	PI-mono (n=296)	Total (n=587)
Demographic and clinical characteristics			
Age (years)	43 (37–49, 23–75)	45 (39–50, 23–67)	44 (38–49, 23–75)
Route of infection			
Homosexual	175 (60%)	176 (59%)	351 (60%)
Heterosexual	108 (37%)	108 (36%)	216 (37%)
Other	8 (3%)	12 (4%)	20 (3%)
Female	64 (22%)	73 (25%)	137 (23%)
Ethnic origin			
White	206 (71%)	195 (66%)	401 (68%)
Black	73 (25%)	90 (30%)	163 (28%)
Other	12 (4%)	11 (4%)	23 (4%)
Hepatitis C virus antibody positive	7 (2%)	14 (5%)	21 (4%)

Paton NI, *Lancet*, 2015

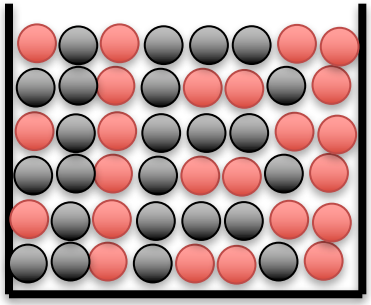
L’inférence statistique

De l'échantillon….



….a la population d'intérêt

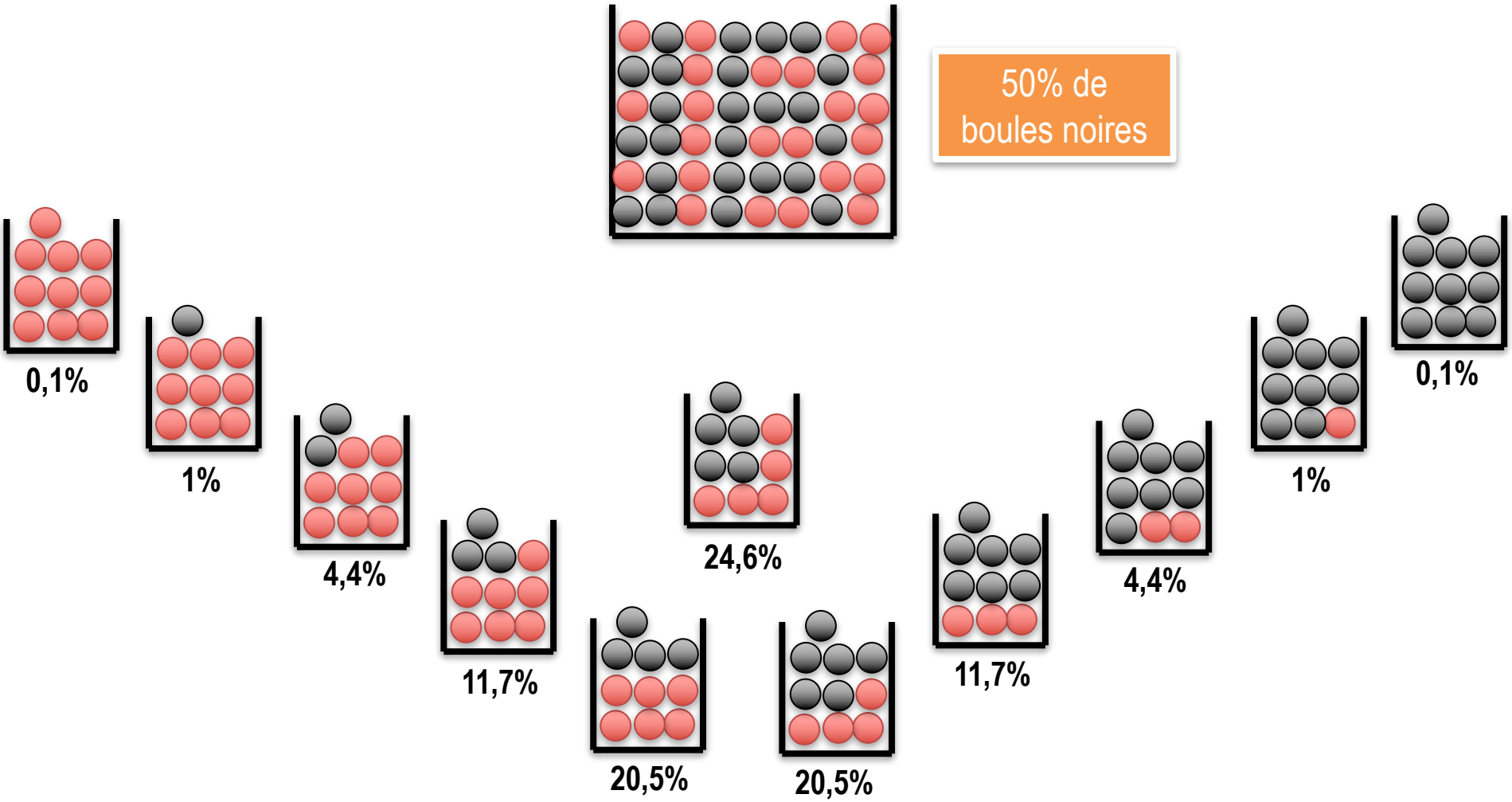
Les fluctuations d'échantillonage



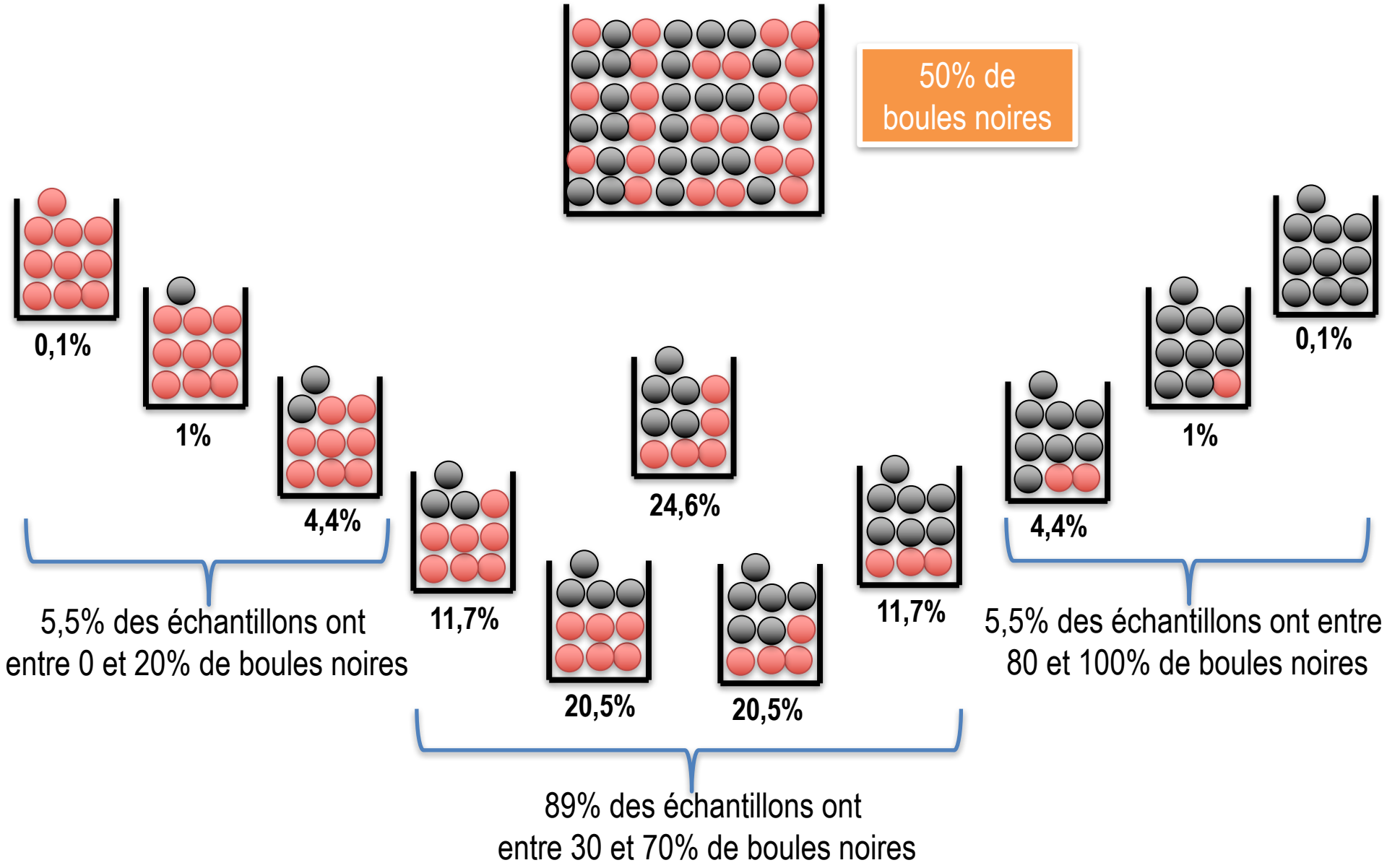
50% de boules noires

Tirage au sort d'échantillons de 10 boules noires

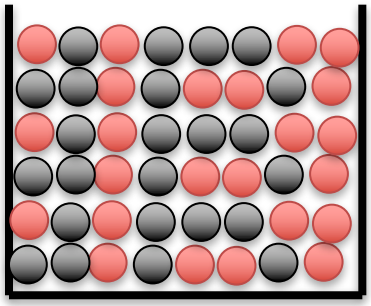
Les fluctuations d'échantillonage



Les fluctuations d'échantillonage

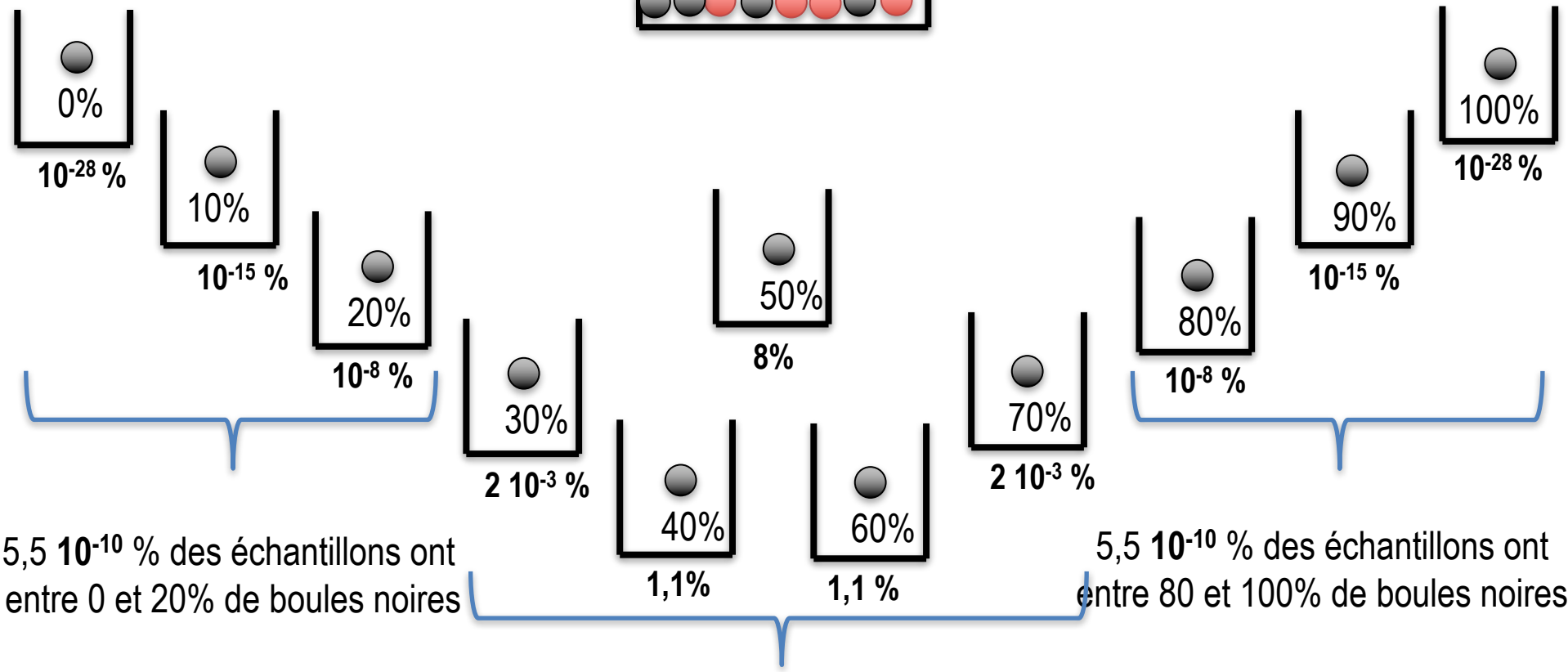


Les fluctuations d'échantillonage : influence de la taille de l'échantillon



50% de boules noires

N=100

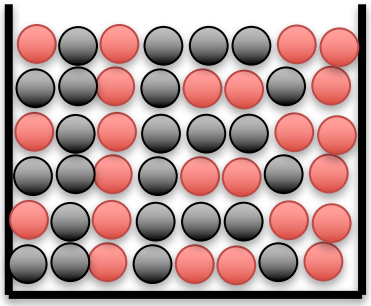


5,5 10^{-10} % des échantillons ont entre 0 et 20% de boules noires

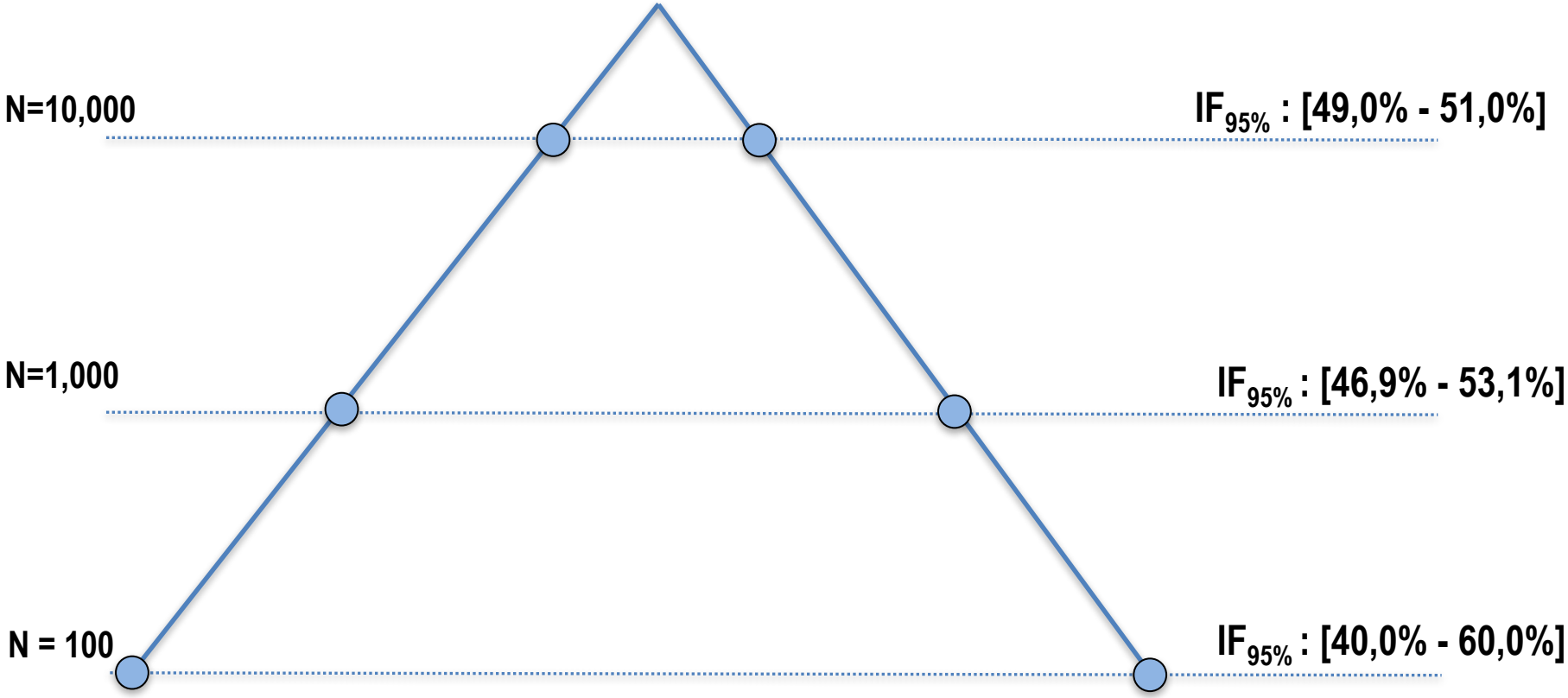
5,5 10^{-10} % des échantillons ont entre 80 et 100% de boules noires

~ 99,999 % des échantillons ont entre 30 et 70% de boules noires

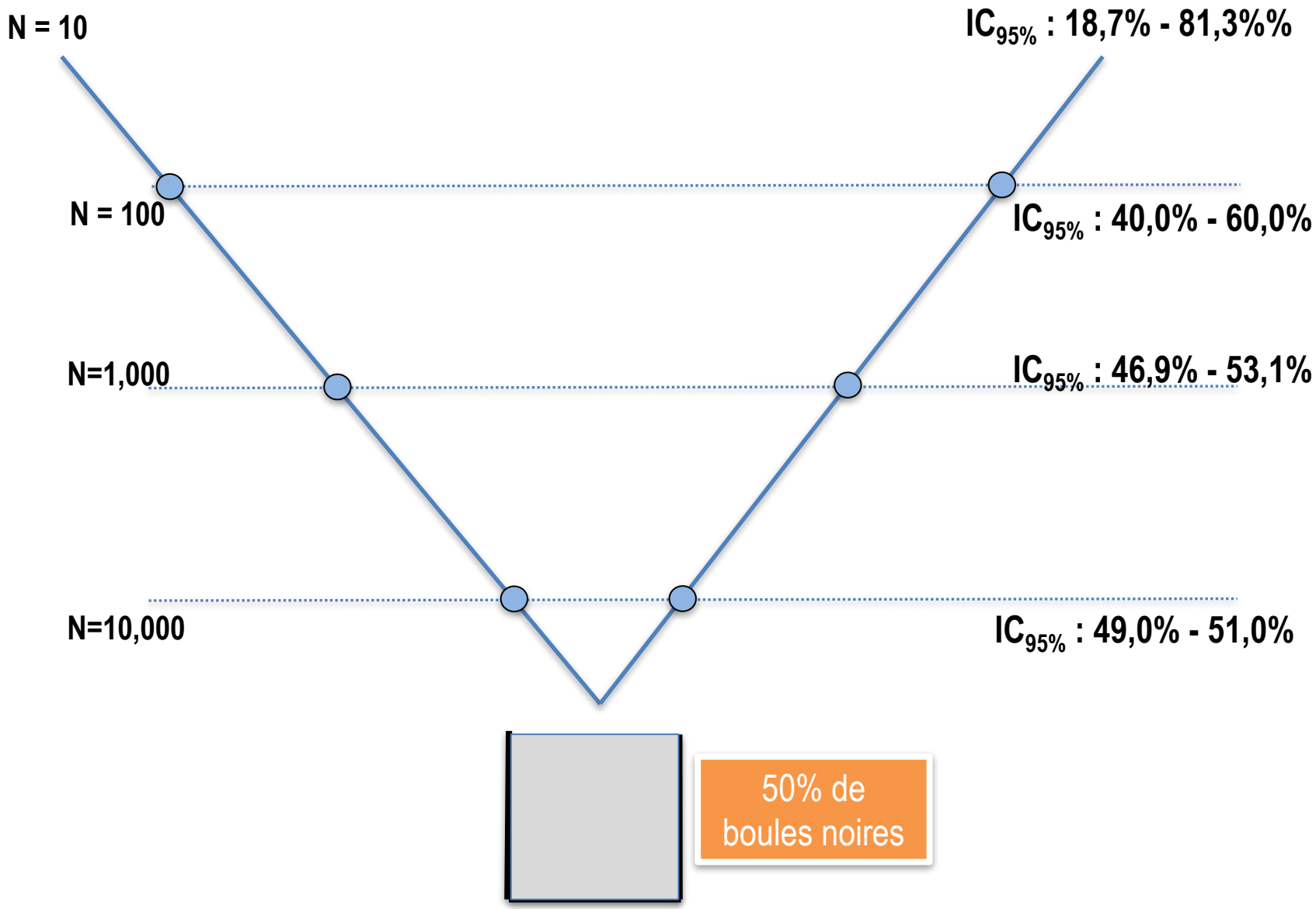
Les fluctuations d'échantillonage : influence de la taille de l'échantillon



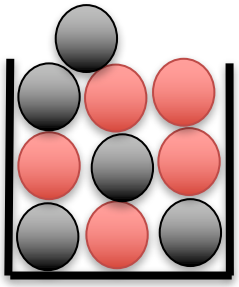
50% de boules noires



L'intervalle de confiance

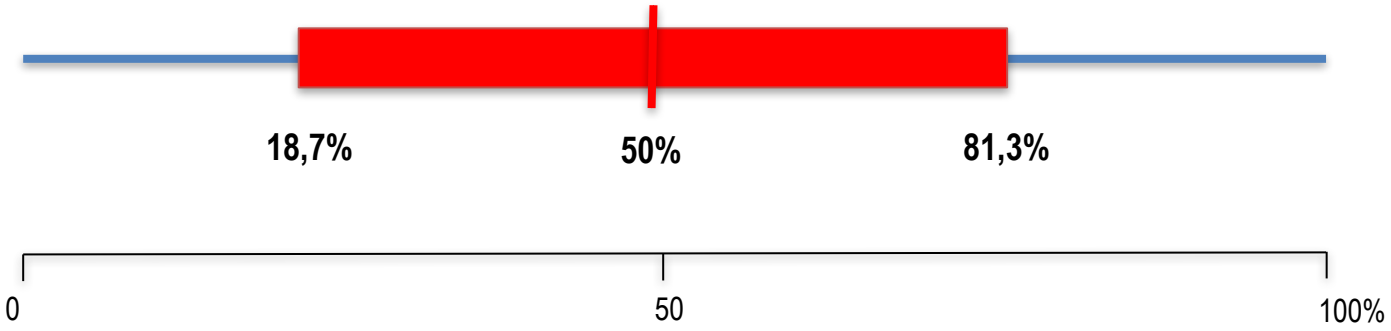


La notion d'intervalle de confiance



Interprétation

- 50% des boules de l'échantillon (n=10) sont noires
- Il y a 95% de chances pour que la proportion de boules noires dans la population dont est issue l'échantillon soit comprise entre 18,7% et 81,3%



Calculer l'intervalle de confiance

Proportion

- Calcul la variance de l'échantillon : $p*q$
- Calcul de l'intervalle de confiance = $p \pm 1{,}96*\sqrt{(p*q/n)}$
- Exemple : $p = 40\%$, $n=10$: IC = $0{,}5 \pm 1{,}96* \sqrt{(0{,}6*0{,}4)/10} = 0{,}1 – 0{,}7$

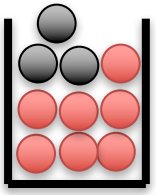
Moyenne

- Calcul la variance de l'échantillon
- Calcul de l'intervalle de confiance = $m \pm 1{,}96*\sqrt{(\sigma^2/n)}$
- Exemple : $m = 3$, $\sigma^2=4$, $n=10$: IC = $3 \pm 1{,}96*\sqrt{(4/10)} = 1{,}8 – 4{,}2$

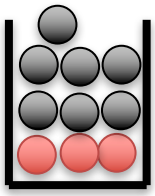
Les tests statistiques

Principes des tests statistiques

N=10



30% de boules noires

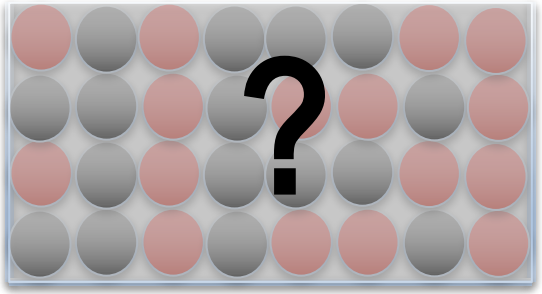


N=10

70% de boules noires

Principes des tests statistiques

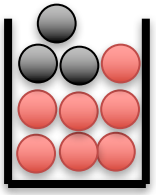
50% de boules noires



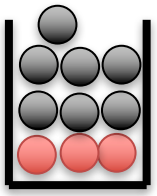
Question posée au test :

- Quelle est la probabilité d'obtenir 30% de boules noires et 70% de boules noires dans 2 échantillons de 10 boules issues de la même urne
- Réponse : 18%

N=10



30% de boules noires



N=10

70% de boules noires

Principes des tests statistiques

Variable binaire

Hommes = ?

p=0,26

Hommes = 51%
n=100

Hommes = 59%
n=100

Principes des tests

p=0,26 signifie : la probabilité d'observer 2 échantillons aléatoires de 100 patients issus de la même population avec une différence de pourcentage d'au moins 8 points est de 26%

p=0,26

Hommes = 51%
n=100

Hommes = 59%
n=100

Principes des tests

Comparaison de pourcentages : test du χ^2

p=0,26

Hommes = 51%
n=100

Hommes = 59%
n=100

Principes des tests

Variable continue

Poids ?

p=0,005

Poids : 78 (10)
n=100

Poids : 82 (10)
n=100

Principes des tests

p=0,005 signifie : la probabilité d'observer 2 échantillons aléatoires de 100 patients issus de la même population avec une différence de poids de 4 kilos est de 0,005%

p=0,005

Hommes = 51%
n=100

Hommes = 59%
n=100

Principes des tests

Comparaison de moyennes:
t-test (Student's t-test)

p=0,005

Hommes = 51%
n=100

Hommes = 59%
n=100

Principes des tests

Proportions

Dans le cas de 2 échantillons de 100 patients avec une proportion d'hommes de 46% et 42% respectivement, le test répond à la question : quelle est la probabilité d'observer une différence d'au moins 4% entre 2 échantillons issus d'une même population composée de 44% d'hommes

Variables continues

Dans le cas de 2 échantillons de 100 patients avec des poids moyen de 78 et 82 kilos (SD =10), le test répond à la question : quelle est la probabilité d'observer une différence d'au moins 4 kgs entre 2 échantillons issus d'une même population dont le poids moyen est de 80 kgs (SD=10)

Principes des tests

- Les tests statistiques ont des conditions de validité (notamment la taille des échantillons)

- Quand les échantillons comparés sont de petite taille (n=30), des tests non paramétriques sont utilisés (test exact de Fisher, test de Mann-Whitney).

Points clés

Les tests statistiques permettent de quantifier la probabilité d'observer les différences entre 2 échantillons liées aux fluctuations d'échantillonage

Rappel : Le hasard peut tout expliquer

Les principaux tests

Type de variable	Nombre de groupes	N > 30	N<30
Continue	2 groupes	t-test (test de Student)	Mann-Withney U test (Wilcoxon)
	> 2 groupes	ANOVA	Kruskal-Wallis
Binaire	2 groupes ou +	Chi-2	McNemmar
Catégorielle	2 groupes ou +	Chi-2	McNemmar

Principe de l'essai randomisé

- Définir une population d'intérêt (sélection)
- Tirer au sort 2 groupes qui ne diffèrent que par le traitement reçu (randomisation) et les fluctuations d'échantillonage
- Mesurer le critère de jugement au terme de l'étude
- Quantifier la probabilité que la différence observée sur le critère de jugement entre les 2 échantillons soit liée aux fluctuations d'échantillonage
- Prendre un risque pour prendre une décision

Principe de l’essai randomisé

Pour que la probabilité observée (p) soit interprétable, il faut conserver la comparabilité des groupes jusqu’au terme de l’étude

- Insu : Ne pas informer les différents participants (malade : simple aveugle) l’équipe soignante et l’ évaluateur (double aveugle) du traitement reçu

  => Assure la meilleure égalité possible du suivi

- Analyser les patients dans leur bras de randomisation (ITT)

ITT, modified ITT, per protocol

- ITT (Full analysis set) : tous les patients inclus dans leur groupe d'assignation
- Modified-ITT : retire les patients
  - Critère d'éligibilité violé
  - Aucune prise du traitement (ou placebo) à l'essai
  - Aucune donnée après randomisation
- Per-protocol

Guidelines ICH-E9

- L'analyse des essais de supériorité doit se faire en ITT ou modified-ITT
- L'analyse des essais de non-infériorité doit se faire en per-protocol et en ITT

ITT et données manquantes

- Arrêt du traitement : suivre le patient jusqu'à la fin de l'étude et recueillir le critère de jugement principal
- Perdu de vue : essayer de retrouver le patient
- Si décès ou autre : méthodes d'imputation

Les fluctuations d'échantillonages ne sont quantifiables que si les groupes tirés au sort sont respectés

Exemple

Critères de sélection

EP avec collapsus

Randomisation

n= 300
Fibrinolytique

n = 300
Placebo

Double-aveugle
Placebo

Décès à J30
15%

Décès à J30
25%

RR = 0,6
[95% IC: 0,4 – 0,8)

p=0,002

Interprétation – risque **α**

- En tirant 2 échantillons de 300 patients d'une population dont la mortalité est de 20%, la probabilité d'observer une différence de 10% est de 0,002
- Cette différence peut-être due
  - Au hasard (fluctuations d'échantillonnage)
  - A l'effet du traitement

<u>Il n'y a aucun moyen de différencier les deux</u>

Notion de risque : je décide que si le p est inférieur à 5% la différence observée est due au traitement

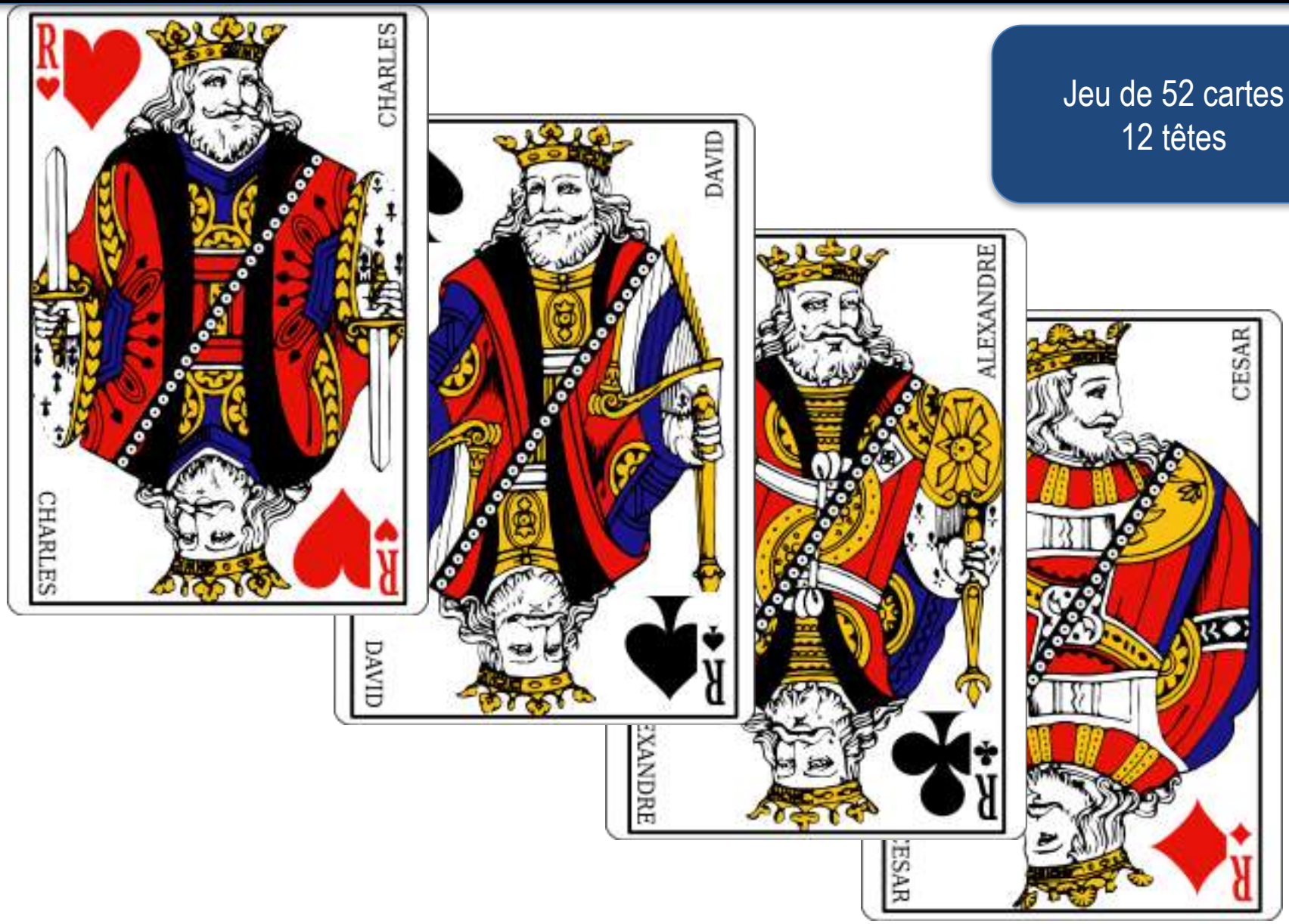
- C'est le **risque de première espèce** ou **risque α**
- Quand je conclus à l'efficacité d'un traitement car $p < 0,05$ j'ai effectivement 5% de conclure à tort à l'effet du traitement

Interprétation – risque **α**

- Par convention le risque α est fixé à 5%
- Ce risque est non négligeable
- La FDA demande 2 essais positifs au risque de 5% pour enregistrer un médicament => risque de 2,5 pour 1000

Risque **α** et petit p sont 2 notions totalement différentes

Risque α et petit p



Jeu de 52 cartes
12 têtes

Risque α et petit p

Quelle est la probabilité de tirer le roi de cœur ?

Risque α et petit p

Quelle est la probabilité de tirer le roi de cœur ?

$p = 1/52 \approx 2\%$

Risque α et petit p

Quelle est la probabilité de gagner au jeu ?

$p = 1/52 \approx 2\%$

Risque α et petit p

Quelle est la probabilité de gagner au jeu ?

p = 1/52 ≈ 2%

Dépend de la règle

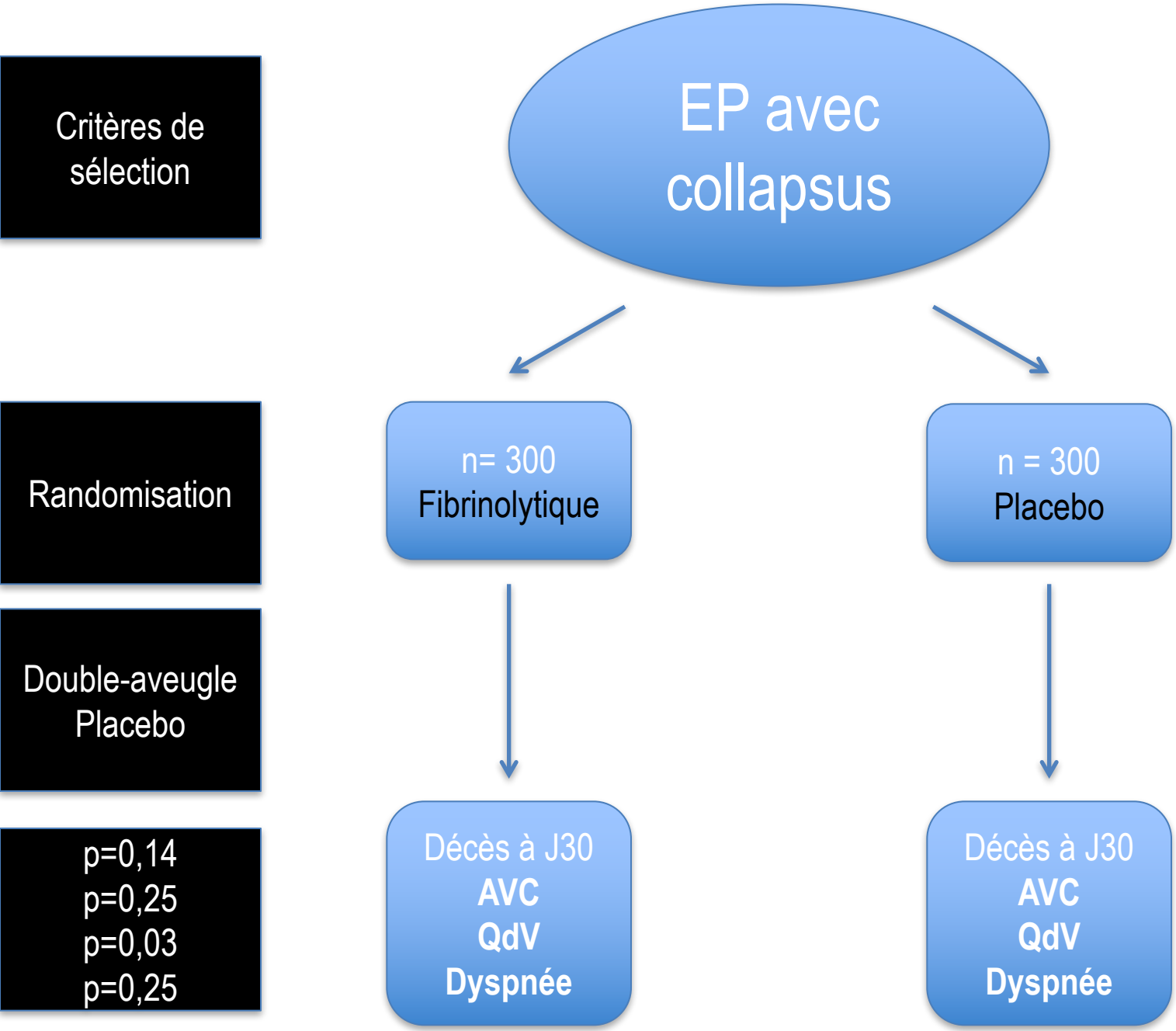
- **Tirer le roi de cœur = 1/52** ≈ 2%
- **Tirer un roi = 4/52** ≈ 7,7%
- **Tirer une tête = 12/52** ≈ 23,1%

Risque α et petit p – En pratique

La règle doit être fixée avant de commencer…

…. et ne doit pas changer en cours de route

=> Définition d'un critère de jugement *a priori*
(clinicaltrial.gov)


Critères de sélection
EP avec collapsus
Randomisation
n= 300
Fibrinolytique
n = 300
Placebo
Double-aveugle
Placebo
Décès à J30
AVC
QdV
Dyspnée
Décès à J30
AVC
QdV
Dyspnée
p=0,14
p=0,25
p=0,03
p=0,25

- La probabilité d’observer une différence de QdV que j’ai observée entre les 2 groupes est de 3%
- En revanche, la probabilité d’observer sur 4 tests au moins un p inférieur à 5% est >> 5% (pas loin de 20%)

  => Inflation du risque alpha

Coprimary end points were the rate of decline in the mean FEV_1 before and after bronchodilation beginning on day 30. Secondary end points included measures of FVC, changes in response on St. George's Respiratory Questionnaire (SGRQ), exacerbations of COPD, and mortality.

Table 2. **Annual Rates of Decline in FEV_1 and FVC before and after Bronchodilation and Scores on Health-Related Quality of Life.***

Variable	Tiotropium		Placebo		Difference between Tiotropium and Placebo (95% CI)	P Value†
	Patients	Mean Decline	Patients	Mean Decline		
	no.	*ml/yr*	*no.*	*ml/yr*		
FEV_1						
Before bronchodilation	2557	30±1	2413	30±1	0±2 (–4 to 4)	0.95
After bronchodilation	2554	40±1	2410	42±1	–2±2 (–6 to 2)	0.21
FVC						
Before bronchodilation	2557	43±3	2413	39±3	4±4 (–4 to 12)	0.30
After bronchodilation	2554	61±3	2410	61±3	1±4 (–7 to 9)	0.84
Total SGRQ score‡	2505	1.25±0.09	2362	1.21±0.09	0.04±0.13 (–0.2 to 0.3)	0.78

Tashkin D, *N Engl J Med*, 2008

Table 3. Exacerbations of COPD and Related Hospitalizations.*

Variable	Tiotropium	Placebo	Relative Risk for Tiotropium vs. Placebo (95% CI)	P Value
Exacerbation†				
Per patient-year — no.	0.73±0.02	0.85±0.02	0.86 (0.81–0.91)	<0.001
Leading to hospitalization — no. per patient-year	0.15±0.01	0.16±0.01	0.94 (0.82–1.07)	0.34
Days per patient-year	12.11±0.32	13.64±0.35	0.89 (0.83–0.95)	0.001
Hospitalization days per patient-year	3.17±0.17	3.13±0.17	1.01 (0.87–1.18)	0.86
Patients with exacerbation — no. (%)‡				
Total	2001 (67.0)	2049 (68.2)	NA	0.35
Leading to hospitalization	759 (25.4)	811 (27.0)	NA	0.18

CONCLUSIONS

In patients with COPD, therapy with tiotropium was associated with improvements in lung function, quality of life, and exacerbations during a 4-year period but did not significantly reduce the rate of decline in FEV_1. (ClinicalTrials.gov number, NCT00144339.)

Interprétation du p

Table 1. Baseline Characteristics of the Patients.*

Variable	Endobronchial-Valve Therapy (N=220)	Control (N=101)	P Value
Demographic characteristic			
Age — yr	65.34±6.83	64.91±5.84	0.56
Male sex — no. (%)	133 (60.4)	49 (48.5)	0.05
White race — no. (%)†	213 (96.8)	99 (98.0)	1.00
Body-mass index‡	25.09±3.96	24.82±3.39	0.51
Smoking history			
Ever smoked — no. (%)	219 (99.6)	99 (98.0)	0.23
Years of smoking	37.74±8.95	36.12±8.62	0.17
No. of pack-yr	63.29±29.58	61.67±30.33	0.64
Receiving continuous oxygen — no./total no. (%)	65/148 (43.9)	30/72 (41.7)	0.77
Medication use			
Short-acting bronchodilator only — no. (%)	17 (7.7)	9 (8.9)	0.83
Long-acting bronchodilator only — no. (%)§	42 (19.1)	21 (20.8)	0.77
Long-acting bronchodilator and inhaled glucocorticoid only — no. (%)§	60 (27.3)	29 (28.7)	0.90
Long-acting beta agonist, inhaled glucocorticoid, and long-acting anticholinergic agent only — no. (%)	101 (45.9)	42 (41.6)	0.67
Any oral glucocorticoid — no. (%)	28 (12.7)	13 (12.9)	0.97
Arterial blood gas¶			
PaO_2 — mm Hg	69.14±10.26	68.44±8.09	0.51
$PaCO_2$ — mm Hg	40.53±4.25	41.61±4.82	0.04

Risque α et *significativité*

Indique quelle est la probabilité que la différence observée soit liée au hasard ……..

…….. Mais elle est bien due au hasard puisque c'est le résultat du tirage au sort

En faisant 20 tests au seuil de 5%, on s'attend à au moins 1 test *significatif* par hasard

Comment maitriser le risque α

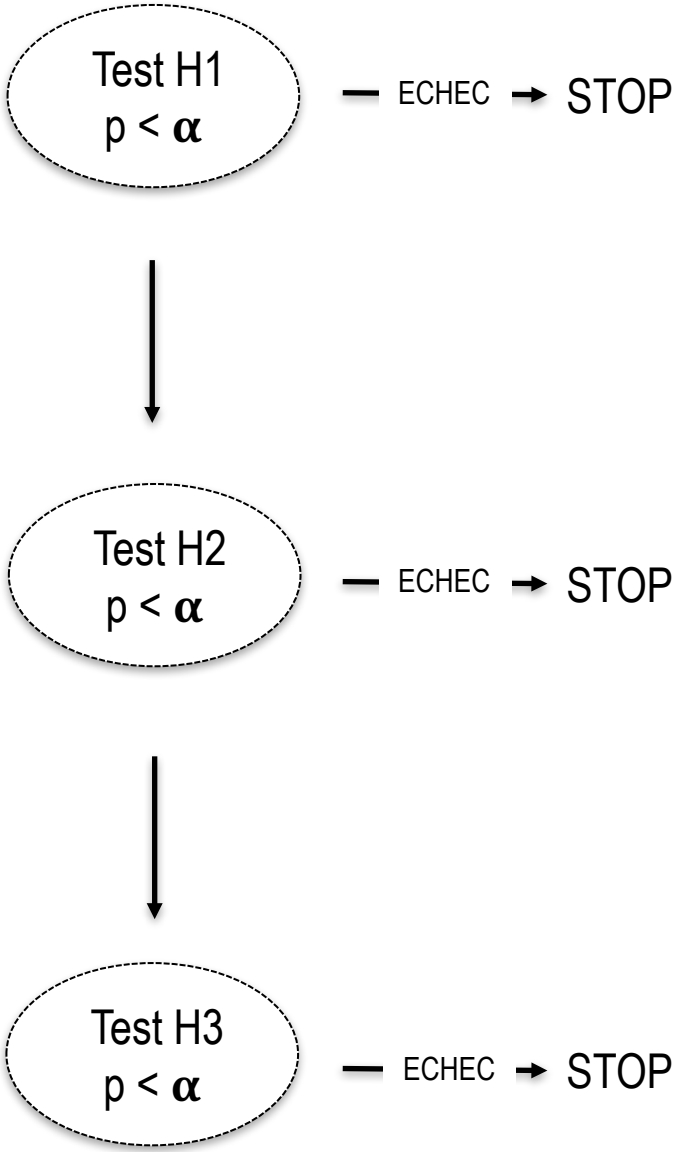
- Définir le critère de jugement *a priori* et **ne pas en changer**

  le rôle des registres d'essais (clinicaltrial.gov) est de s'en assurer

- Si *n* critères de jugement, alors diviser le seuil par *n* (méthode de Bonferroni)
  - 2 critères au seuil de significativité de 2,5% = risque alpha à 5%

- Si analyse intermédiaires, diviser le seuil par le nombre de tests (alpha spending function)

- Plusieurs hypothèses à tester (par exemple : non infériorité puis supériorité)



Fixed-sequence procedure

Hierarchical test procedure

Closed test procedure

Les analyses de sous-groupes

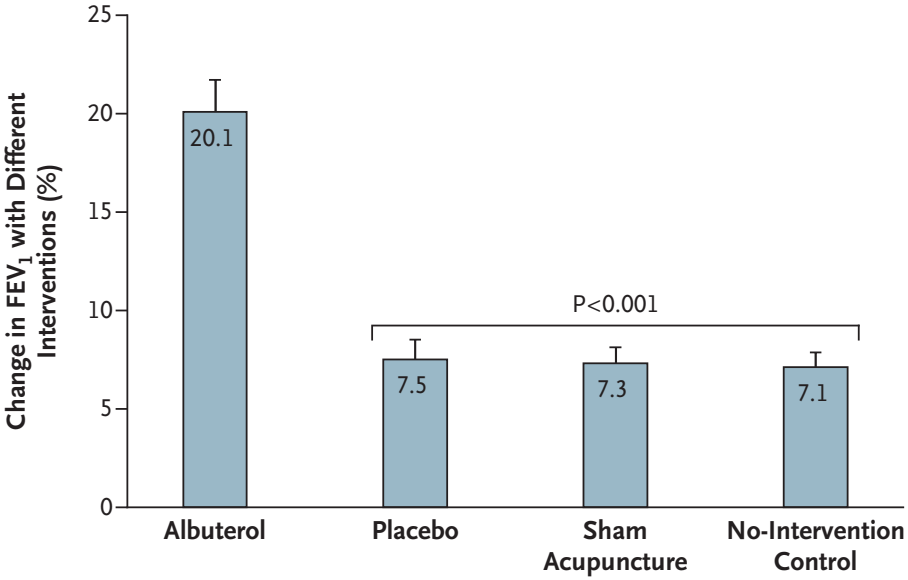
- Si l'auteur rapporte 5 sous-groupes c'est qu'il en a testé 20 !!
- Intérêt des sous-groupes pré-spécifiés
- Les analyses en sous-groupes ne peuvent-être qu'exploratoires

Quelques règles d'interprétation

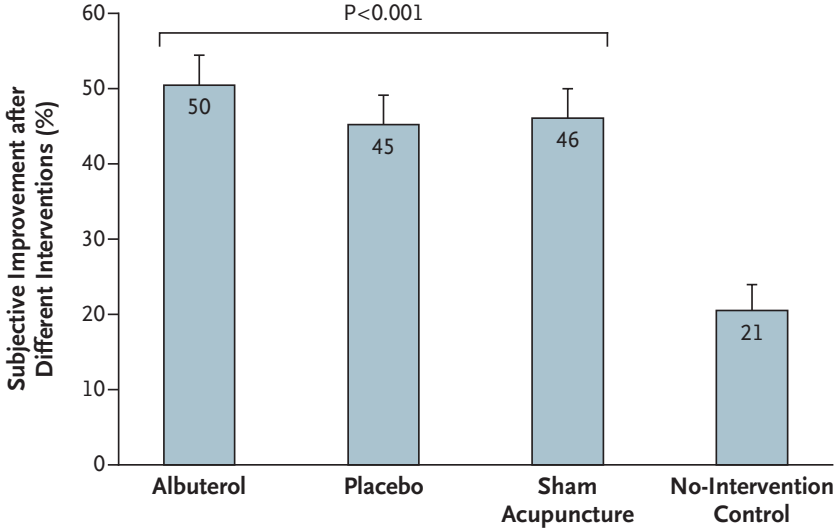
- Seule le p mesuré sur un critère de jugement pré-spécifié est 'confirmatoire'
- Les autres analyses (sous-groupes etc…) sont exploratoires
- Hypothèses pré-spécifiées >> analyses post-hoc
- Proscrire les 'tendances'
- Proscrire : 's'il y avait eu d'avantage de patients le résultat aurait été significatif'

L'effet placebo mythe ou réalité ?

Mesure objective (VEMS)



Mesure subjective (symptômes)



Wechsler ME, *N Engl J Med*, 2011

L’analyse multivariée

- Modèle (ou équation) permettant d’expliquer une variable (dépendante) par un ensemble de variables (indépendantes)
- Suivant le type de variable dépendante,
  - Continue : régression linéaire
  - Binaire : régression logistique
  - Temps avant survenue d’un événement : modèle de survie (Cox)

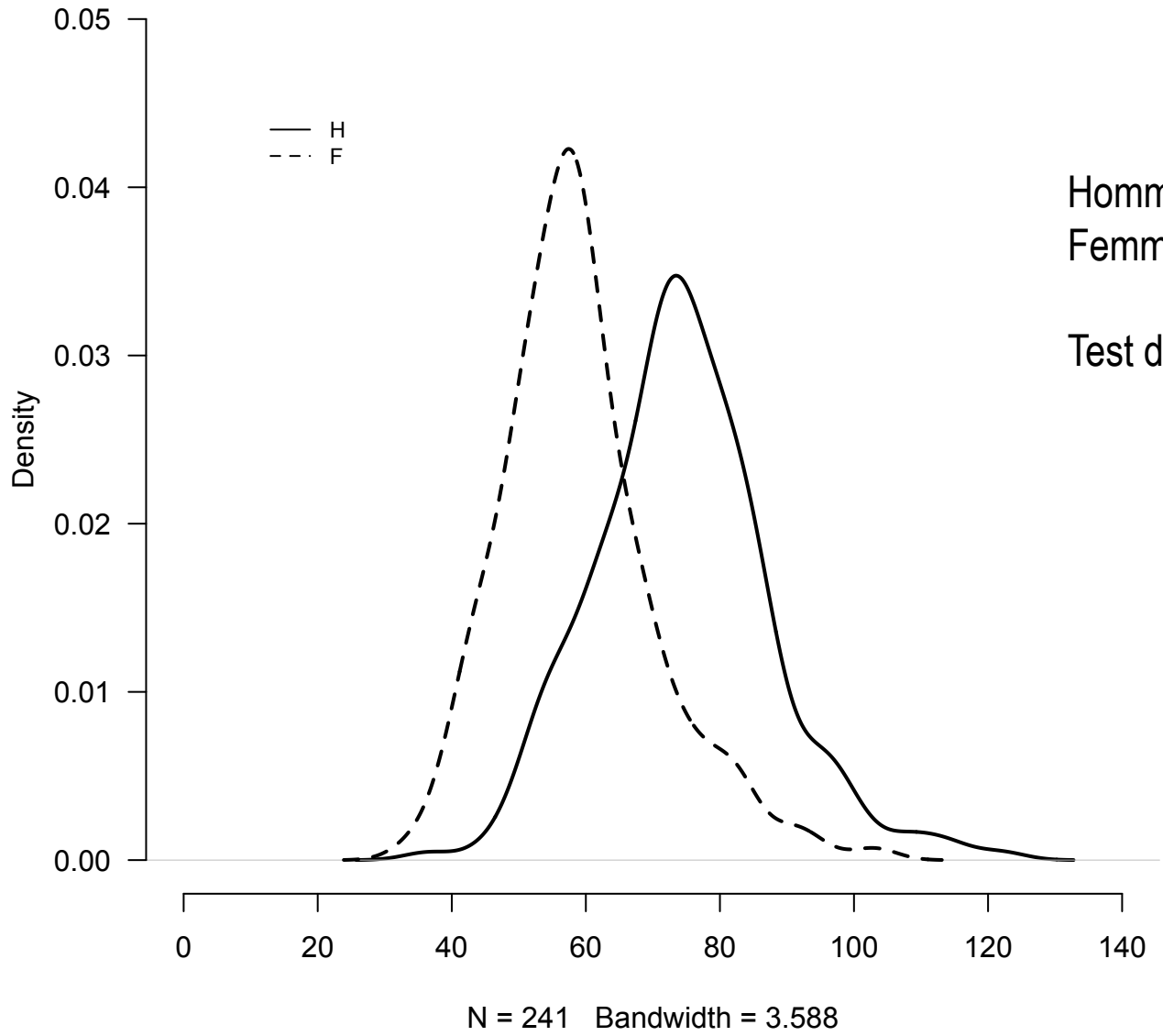
L’analyse multivariée

Variable	Modèle	Paramètre
Continue	Régression linéaire	β
Binaire	Régression logistique	OR
Temps avant évènement	Modèle de survie (Cox)	HR

Exemple d’analyse multivariée

- 410 patients
- Poids en fonction du genre ?

Exemple d'analyse multivariée



Hommes : 74,3 (13,0)
Femmes : 59,0 (11,5)

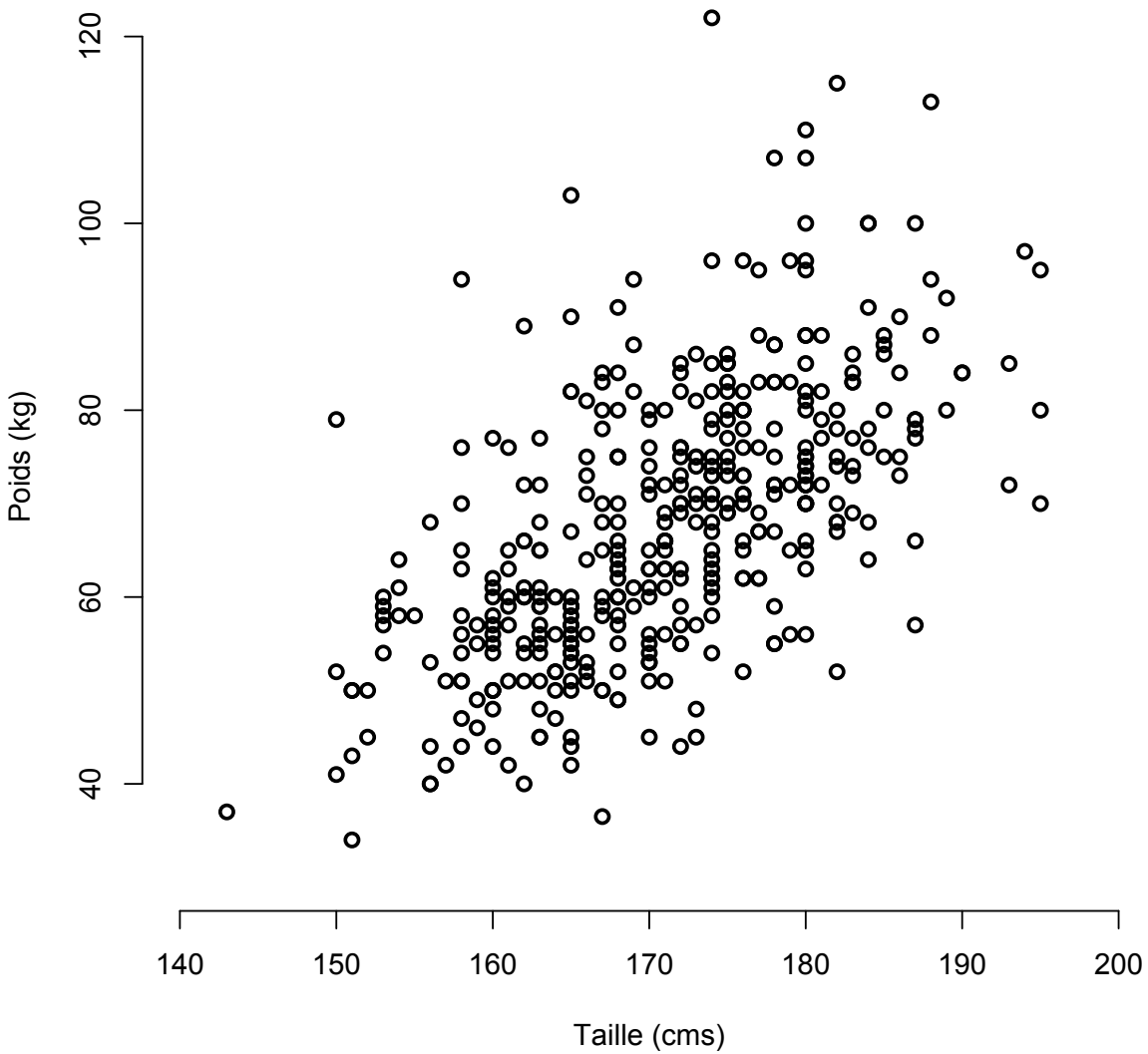
Test de Student : p<0,001

L’analyse multivariée

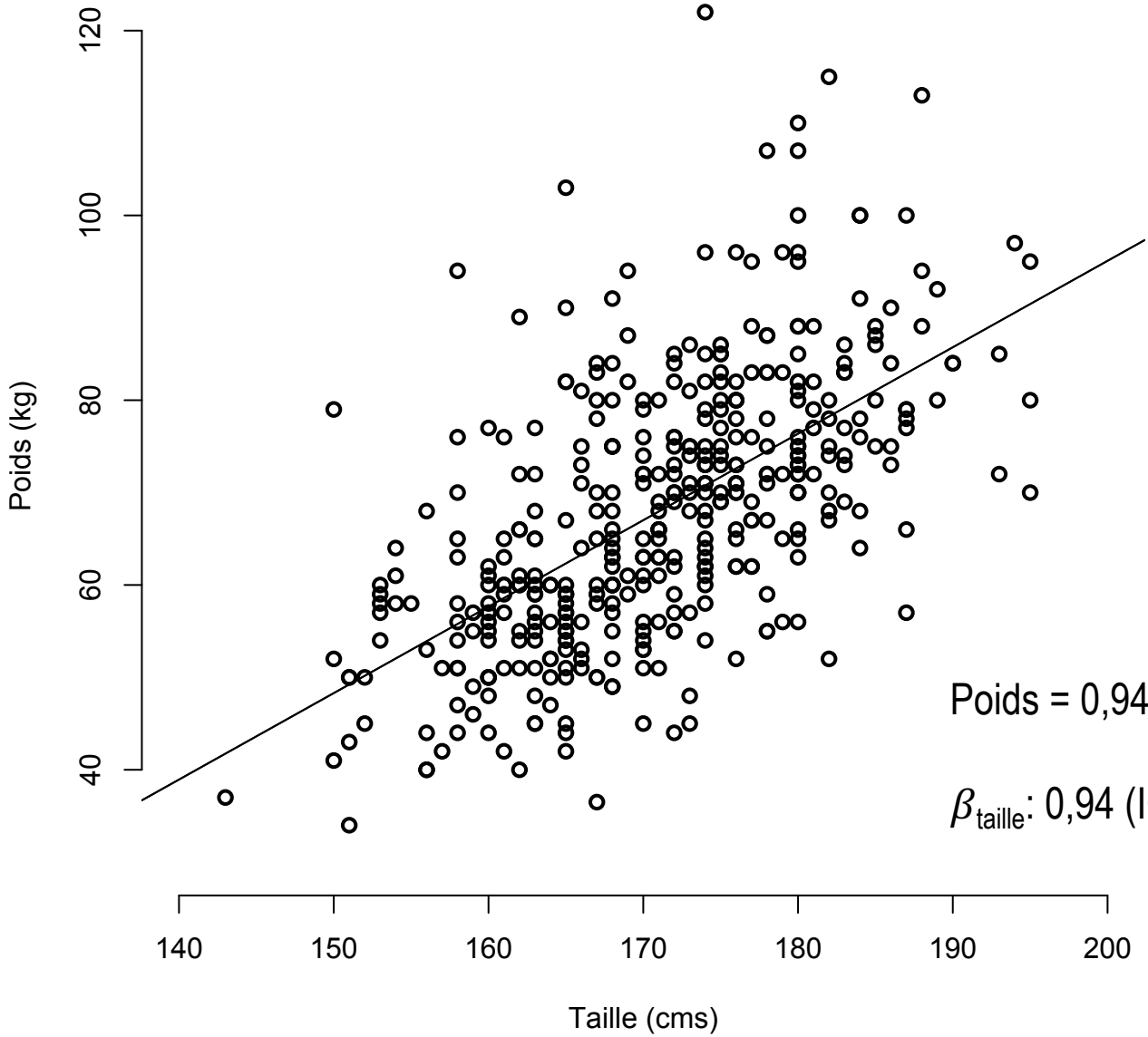
- La différence de poids entre les hommes et les femmes s’explique-t-elle par une différence de taille ?

- Taille des hommes : 176,3 (7,4)
- Taille des femmes : 163,6 (6,4)
- $P < 0,001$

Exemple d'analyse multivariée



Régression linéaire multiple

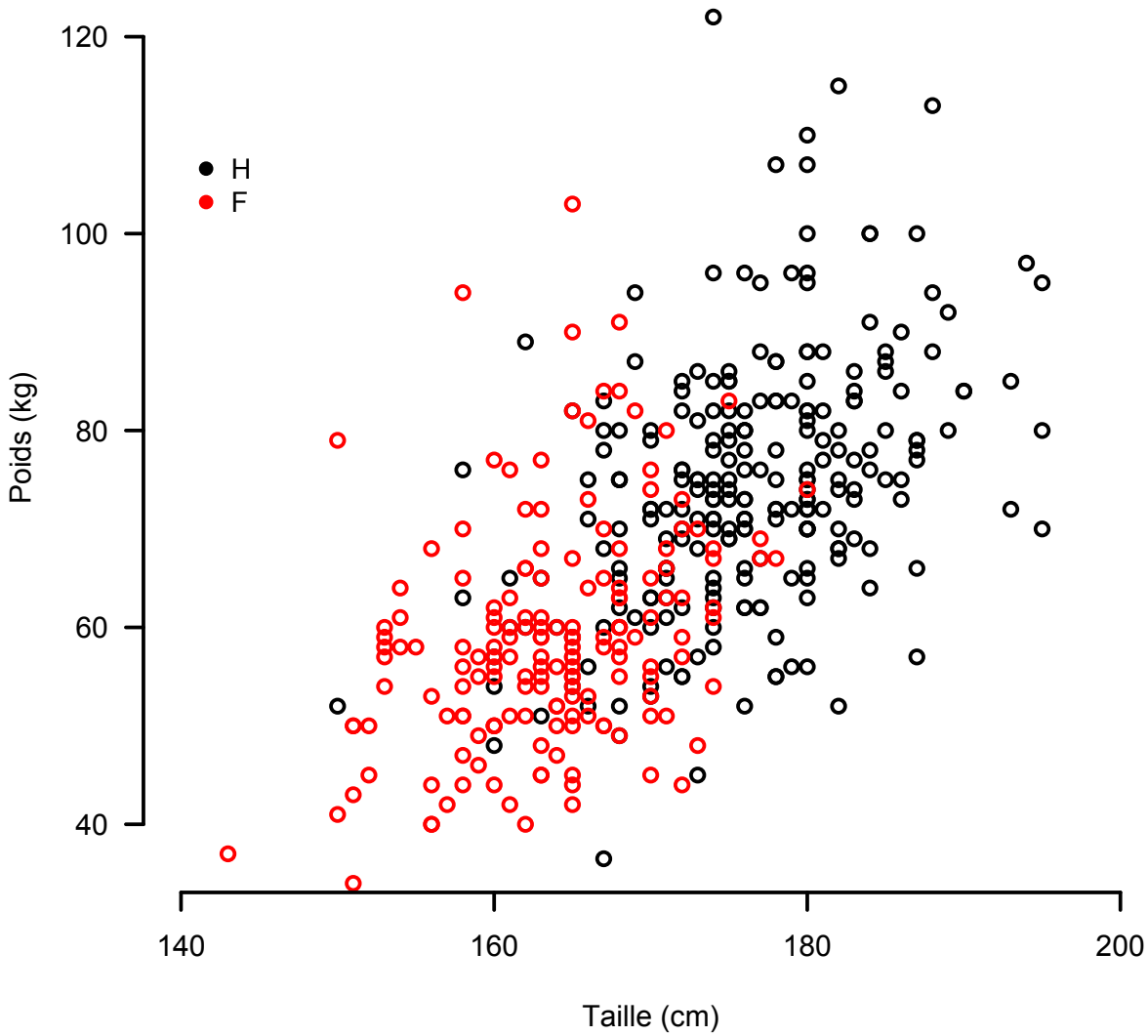


Poids = 0,94*taille – 92,1

β_{taille}: 0,94 (IC95% : 0,82 – 1,05) ; p<0,001

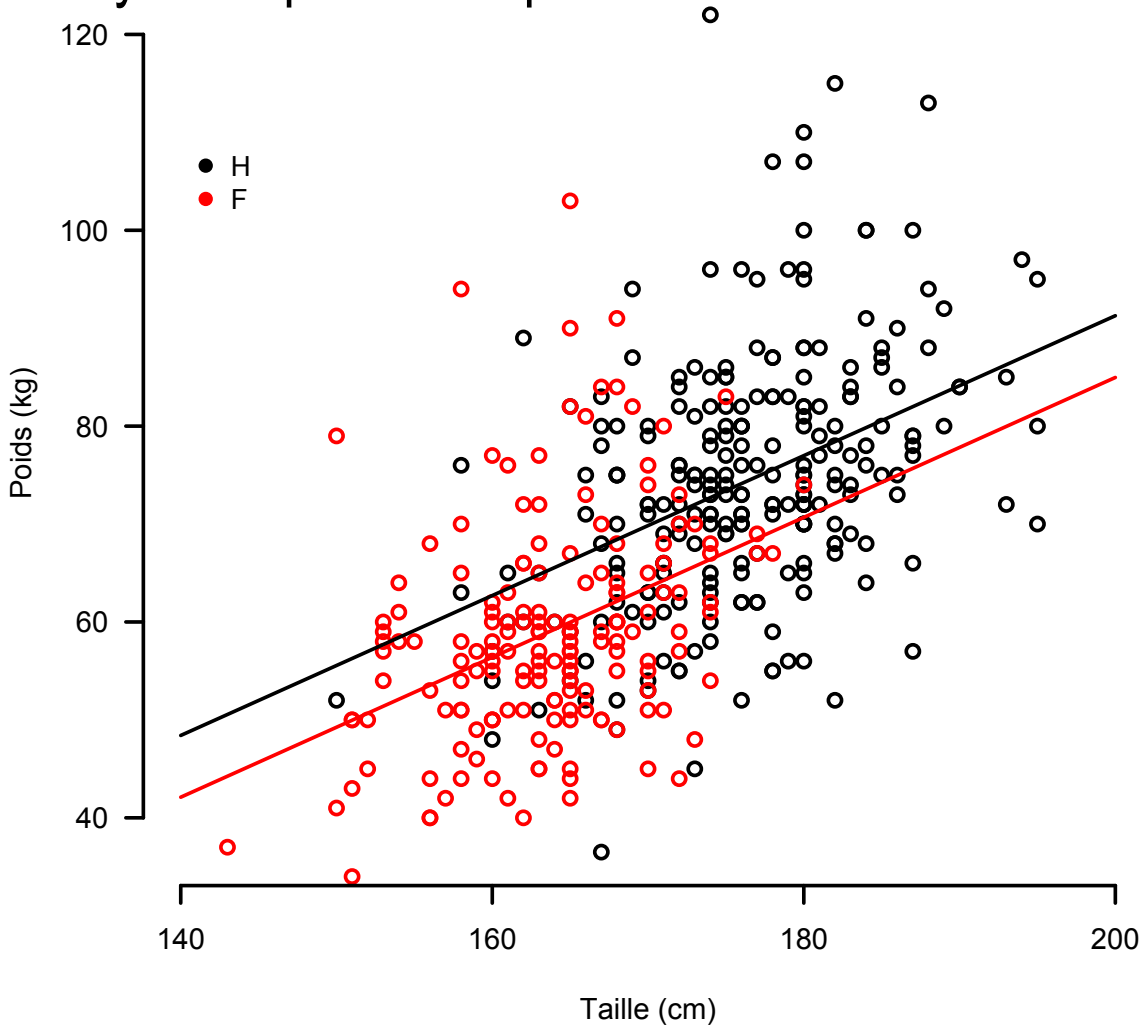
Exemple d'analyse multivariée

Analyse du poids des patients en fonction de leur taille et de leur sexe



Exemple d'analyse multivariée

Analyse du poids des patients en fonction de leur taille et de leur sexe



Poids = 0,71*Taille – 6,3*F – 51,6

Exemple d'analyse multivariée

Paramètre	β	p
Taille	0,71 (0,08)	<0,001
Sexe (F)	-6,3 (1,52)	<0,001
Intercept	-51,6 (15,3)	<0,001

Interprétation

1 – permet d'analyser indépendamment chaque facteur inclus dans l'analyse multivariée
répond à la question : quelles est l'influence du genre, toutes choses égales par ailleurs

- A taille égale, les femmes pèsent 6,3 kilos de moins que les hommes. Si on regardait simplement le poids des hommes vs celui des femmes : 74,3 vs 59,0
- Chaque augmentation de 1 cm de la taille est associée à une augmentation de 0,7 kilos quelque soit le sexe

Régression logistique

Table 2. Factors Independently Associated with Exacerbations at Enrollment and Follow-up.*

Time Period and Factor	Odds Ratio (95% CI)	P Value
History of severe exacerbations at enrollment		
FEV_1, per percentage-point decrease	1.02 (1.01–1.03)	0.001
SGRQ, per 1-point increase	1.03 (1.02–1.04)	<0.001
Age, per 1-year increase	0.97 (0.95–0.99)	0.002
PA:A ratio >1	4.78 (3.43–6.65)	<0.001

OR ≈ RR

Wells JM, *N Engl J Med*, 2012

Prédiction

- **Permet de faire des prédictions (scores pronostiques etc….)**

⇒ Si taille = 1,60 et Genre = F : Poids = 0,71x160 – 6,3 – 51,6 = 55,7 kgs

⇒ Scores GAP, BODE etc....

- **Fait des hypothèses**

=> Relation Taille / Poids linéaire etc..

Pour aller plus loin

- Ecole d’été de santé publique et d’épidémiologie de Bicêtre => http://www.eespe.u-psud.fr

Merci de votre

attention