



Master  
Biologie cellulaire et physiopathologie  
UE initiation à la communication scientifique



## ***Quelques notions de base de statistiques appliquées à la biologie***

Etienne Roux

*Laboratoire de Physiologie Cellulaire Respiratoire INSERM U 885  
UFR des Sciences de la Vie Université Victor Segalen Bordeaux 2*

**contact: [etienne.roux@u-bordeaux2.fr](mailto:etienne.roux@u-bordeaux2.fr)  
support de cours : [e-fisio.net](http://e-fisio.net)**

### Introduction

*« There are three kinds of lies:  
lies, damned lies, and statistics. »*

attribué par Mark Twain à  
Benjamin Disraeli

Qu'est-ce que les statistiques ?

Pourquoi utiliser les statistiques ?

## Introduction

Qu'est-ce que les statistiques ?  
Pourquoi utiliser les statistiques ?

? **Caractéristiques en biologie de la mesure d'un certain nombre de facteurs**

- variabilité des réponses en biologie
- mesure à partir d'échantillons

? **Types de questions que l'on se pose en recherche :**

- qu'est-ce qui produit un effet ?
- à quoi est due l'amplitude de cet effet ?
- qu'est-ce qui produit l'effet le plus important ?

## Introduction

? caractériser la relation entre variables  
variables qualitatives ou quantitatives

? **Statistiques descriptives**

? à partir des données obtenus sur l'échantillon, inférer les  
caractéristiques de la population d'origine

? estimer si plusieurs échantillons proviennent ou non d'une même  
population

? **Statistiques inférentielles**

## Notion de variables

### **Exemples**

1. cellules musculaires lisses isolées de trachée de rats  
application d'un inhibiteur des SERCA  
réponse calcique à la caféine
2. étudiants inscrits en master BCPP en 2005-2006  
gaucher / droitier
3. étudiants inscrits en maîtrise de BCP  
sexe  
repos/effort  
rythme cardiaque
6. genres fossiles  
durée de vie
7. réponse à un agoniste contractant  
concentration de l'agoniste  
amplitude de la contraction

## Notion de variables

### **Variables qualitatives et quantitatives**

1. cellules musculaires lisses isolées de trachée de rats  
application d'un inhibiteur des SERCA  
réponse calcique à la caféine
2. étudiants inscrits en master BCPP en 2005-2006  
gaucher / droitier
3. étudiants inscrits en maîtrise de BCP  
sexe  
repos/effort  
rythme cardiaque
6. genres fossiles  
durée de vie
7. réponse à un agoniste contractant  
concentration de l'agoniste  
amplitude de la contraction

## Notion de variables

### ***Variables indépendantes et variables dépendantes***

1. cellules musculaires lisses isolées de trachée de rats  
application d'un inhibiteur des SERCA  
réponse calcique à la caféine
2. étudiants inscrits en master BCPP en 2005-2006  
gaucher / droitier
3. étudiants inscrits en maîtrise de BCP  
sexe  
repos/effort  
rythme cardiaque
6. genres fossiles  
durée de vie
7. réponse à un agoniste contractant  
concentration de l'agoniste  
amplitude de la contraction

## Notion de variables

### ***Variables indépendantes et variables dépendantes***

1. liens entre variables qualitatives  
variable indépendante qualitative  
variable dépendante qualitative  
ex : présence d'un inhibiteur des SERCA  
types de réponses (oscillantes vs non oscillantes)
2. liens entre variables qualitatives et quantitatives  
variable indépendante qualitative  
variable dépendante quantitative  
ex : présence d'un inhibiteur des SERCA  
amplitude de la variation de  $[Ca^{2+}]_i$
3. liens entre variables quantitatives  
variable indépendante quantitative  
variable dépendante quantitative  
ex : concentration d'inhibiteur des SERCA  
amplitude de la variation de  $[Ca^{2+}]_i$

## Notion de variables

### **Variables contrôlées et non contrôlées**

1. cellules musculaires lisses isolées de trachée de rats  
application d'un inhibiteur des SERCA  
réponse calcique à la caféine
2. étudiants inscrits en master BCPP en 2005-2006  
gaucher / droitier
3. étudiants inscrits en maîtrise de BCP  
sexe  
repos/effort  
rythme cardiaque
6. genres fossiles  
durée de vie
7. réponse à un agoniste contractant  
concentration de l'agoniste  
amplitude de la contraction

## Notion de variables

### **Variables contrôlées et non contrôlées**

#### ***observation (survey)***

Dans les études d'observations, les variables indépendantes ne sont pas contrôlées.

exemples :

- fréquence des cancers de la thyroïde après l'accident de Tchernobyl, dans une zone géographique donnée.
- durée de vie des genres fossiles
- admission aux urgences pour problèmes respiratoires en fonction de l'intensité de la pollution atmosphérique

## Notion de variables

### *Variables contrôlées et non contrôlées*

#### *expérimentation (experiment)*

Dans les études d'expérimentation, les variables indépendantes sont contrôlées

exemples :

- effet de l'adrénaline sur la fréquence cardiaque.
- détermination sur la souris de la quantité minimale contaminante de cerveau de bovin atteint d'ESB.

## Notion de variables

### *Variables contrôlées et non contrôlées*

#### *observation (survey)*

#### *expérimentation (experiment)*

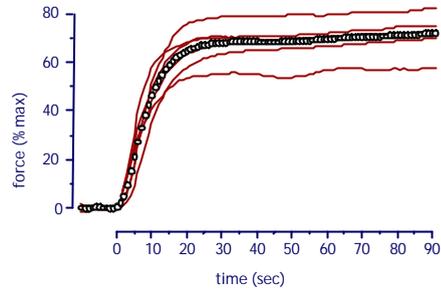
? analyse prospective : observation, expérimentation

? analyse rétrospective : observation

## Variabilité des processus biologiques

### Caractéristiques

ex : tension développée par un anneau de trachée de rat en réponse à une stimulation cholinergique



? la variabilité est la règle

? la variabilité est non prévisible

? la variabilité des résultats est différente de l'erreur instrumentale

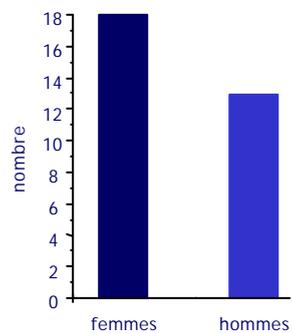
? la « marge d'imprécision » d'estimation de la tendance centrale est un intervalle de probabilité

## Types de distributions

### distribution binomiale

La variable peut prendre deux valeurs - pas forcément numériques.

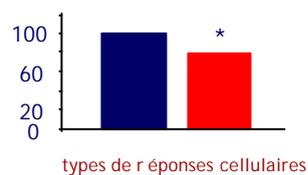
r épartition  
hommes/femmes dans  
une population



proportion de gauchers dans une population

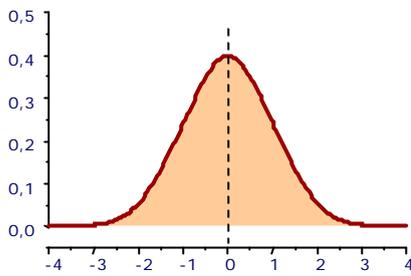
droitiers	gauchers

% of responding cells



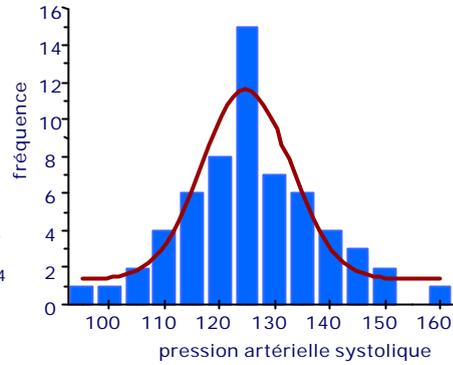
## Types de distributions

### distribution normale



loi de distribution de probabilité, définie par une fonction de densité de probabilité de la forme :

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

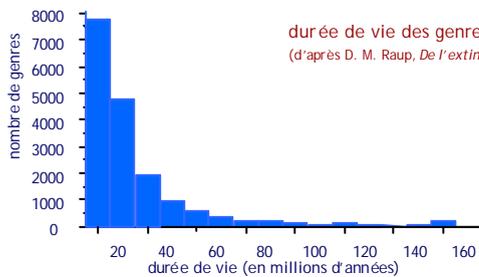
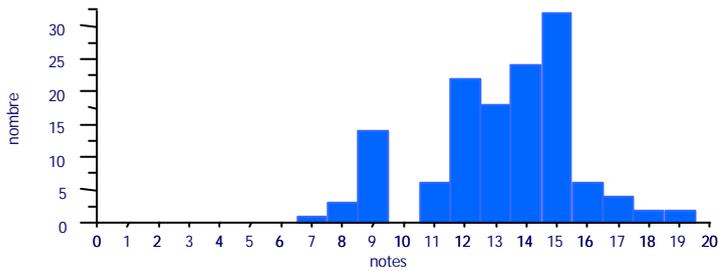


ex : valeur de la pression artérielle systolique dans une population

## Types de distributions

### autres distributions

répartition des notes de TP à un examen de licence BCP



durée de vie des genres fossiles

(d'après D. M. Raup, *De l'extinction des espèces*, Gallimard, Paris, 1993)

## Statistiques descriptives

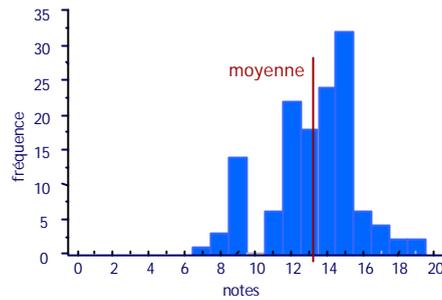
### mesure de la tendance centrale

#### **moyenne arithmétique**

moyenne arithmétique : somme des valeurs de la variable divisée par le nombre de valeurs

= centre de gravité de la distribution

*(pour éviter les biais par simplification, faire le calcul avec une décimale supplémentaire par rapport au nombre de décimales de la valeur exprimée de la moyenne)*



#### *intérêts et limites :*

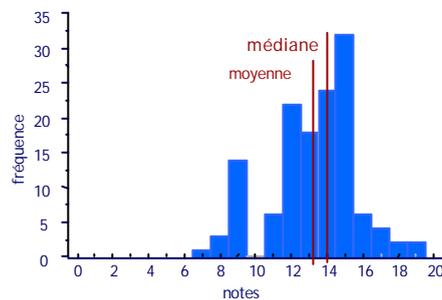
- ? très utilisée en statistiques descriptive et inférentielle
- ? souvent, pas toujours, la mesure la plus pertinente de la tendance centrale

## Statistiques descriptives

### mesure de la tendance centrale

#### **médiane**

valeur de part et d'autre de laquelle se distribue par moitié les valeurs de la variable (50 % des valeurs sont inférieures à la médiane, et 50 % sont supérieures).



#### *intérêts et limites :*

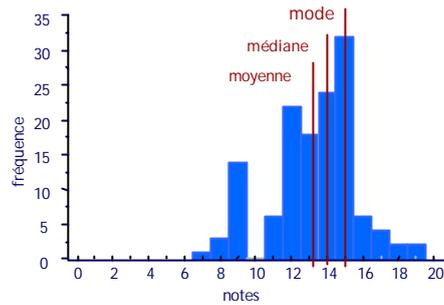
- ? intérêt théorique : dans certains cas, « bonne » manière de décrire la tendance centrale
- ? peu utilisée pour les calculs de signification statistique

## Statistiques descriptives

### mesure de la tendance centrale

#### mode

valeur de la variable qui survient avec la plus grande fréquence  
variables discontinues : valeur exacte  
variables continues : dépend du mode de calcul



#### intérêts et limites :

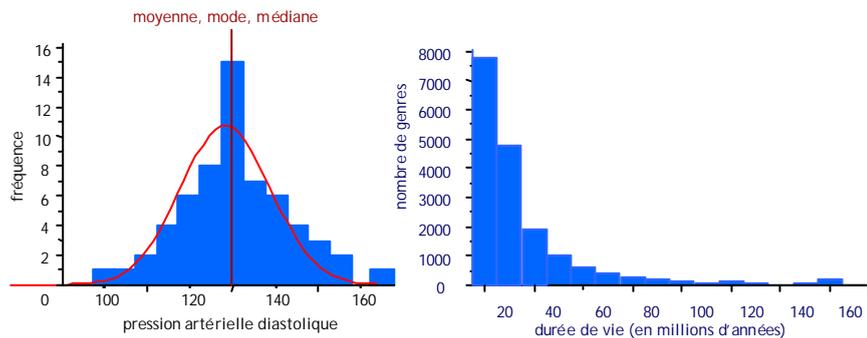
- ? facile à déterminer
- ? difficile à manipuler mathématiquement (pour tester statistiquement des hypothèses)
- ? intérêt théorique : dans certains cas, « bonne » manière de décrire la tendance centrale

## Statistiques descriptives

### mesure de la tendance centrale

#### choix de la mesure

- ? dépend de la loi de distribution
- ? dépend de la question posée



- ? dans la plupart des cas : moyenne
- ? médiane et mode intéressants dans certains cas

## Statistiques descriptives

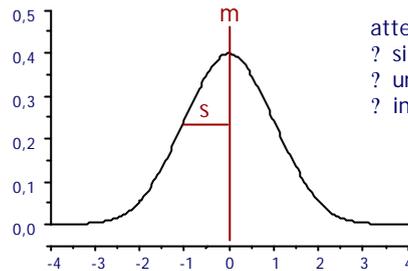
### mesure de la dispersion individuelle

#### écart-type (standard deviation)

racine carrée de la variance

L'écart-type est donc donné par la formule :  $\sqrt{\sum x_i^2 / n - \bar{x}^2}$

loi normale



attention :

- ? signification de l'écart-type
- ? unité de l'écart-type
- ? influence de changement de variable

intérêts et limites :

- ? Après standardisation, permet de comparer la position de plusieurs variables entre elles, même si les unités de mesure de ces variables sont différentes.
- ? quasiment la seule mesure de la dispersion utilisée

## Statistiques descriptives

### mesure de la dispersion individuelle

#### écart (range)

différence entre la plus petite et la plus grande valeur.

intérêt et limites :

- ? facile à calculer
  - ? très instable (une seule valeur extrême modifie fortement la valeur de l'écart)
- La moyenne arithmétique de l'écart déterminé sur des échantillonnages répétés et l'écart moyen (mean range), utilisé comme estimation de l'écart-type de la population.

#### déviatoin moyenne (mean deviation)

moyenne arithmétique de la différence, en valeur absolue, entre chaque valeur et la moyenne arithmétique.

intérêt et limites :

- ? mesure très rarement utilisée

## Statistiques inférentielles : estimations

### questions

#### **estimation des caractéristiques d'une population à partir d'un échantillon**

- ? fréquence de distribution
- ? moyenne et écart-type de la population

*précision de l'estimation*  
*intervalle de confiance*

#### **estimation des différences entre plusieurs populations, à partir d'échantillons**

- ? comparaison à une population théorique
- ? comparaison de plusieurs (2 ou plus) échantillons entre eux

tests statistiques  
*estimation des erreurs*  
*risque de première espèce (a)*  
*risque de deuxième espèce (b)*

## Statistiques inférentielles : estimations

### **estimation des caractéristiques d'une population à partir d'un échantillon**

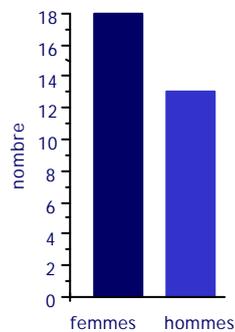
#### fréquence de distribution

r épartition  
hommes/femmes  
dans un échantillon  
d'une population

estimation = valeur de répartition de l'échantillon

échantillon (mesure) :

n = 31  
femmes = 18  
homme = 13



population (estimation) :

femmes = 58,06 %  
homme = 41,94 %

## Statistiques inférentielles : estimations

**estimation des caractéristiques d'une population à partir d'un échantillon**  
moyenne et écart-type de la population

? estimation de la moyenne  
= moyenne de l'échantillon

? estimation de l'écart-type  
écart-type estimé de la population : S  
écart-type calculé de l'échantillon : s

$$s = \sqrt{\frac{S^2}{n-1}}$$

n-1 : degré de liberté

exemple : mesure de la fréquence cardiaque sur un échantillon de 31 personnes  
moyenne de l'échantillon (mesurée) : 86 battements/min  
écart-type de l'échantillon (mesuré) : 13,04 battements/min

moyenne de la population (estimée) : 86 battements/min  
écart-type de la population (estimé) : 13,25 battements/min

## Statistiques inférentielles : estimations

**précision de estimation**

**fluctuation / taille de l'échantillon**

? précision de l'estimation de la moyenne de la population :

dépend de la fluctuation de la moyenne de l'échantillon.

Moins, d'un échantillon à un autre, la valeur moyenne fluctue, plus grande est la précision de l'estimation de la moyenne de la population.

la fluctuation de la moyenne dépend :

- ? de la fluctuation individuelle dans la population
- ? de la taille de l'échantillon

estimation de la fluctuation de la moyenne :  
loi de probabilité de la moyenne ?

## Statistiques inférentielles : estimations

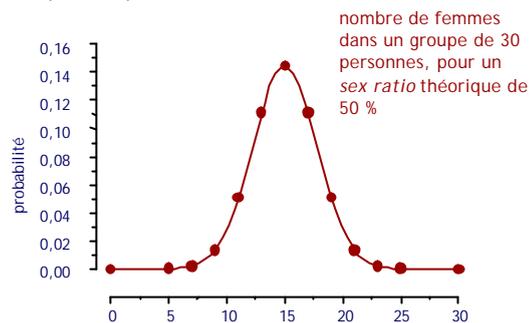
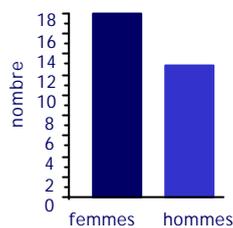
### précision de estimation

#### loi de probabilité de la moyenne

? si la loi de probabilité des variables des individus suit une loi normale, alors la loi de probabilité de la moyenne est également une loi normale

? si la loi de probabilité des variables des individus n'est pas une loi normale, la loi de probabilité de la moyenne est une loi normale, si la taille de l'échantillon est assez grande ( $n > 30$ )

ex : loi binomiale



## Statistiques inférentielles : estimations

### précision de estimation

#### écart-type de la moyenne

estimation de la fluctuation de la moyenne :  
écart-type de la moyenne = standard error of the mean (SEM)

exemple : fréquence cardiaque

moyenne de la population (estimée) : 86 battements/min  
écart-type de la population (estimé) : 13,25 battements/min

SEM = 3,38 battements/min

$$SEM = \frac{s}{\sqrt{n}}$$

exemple : sex ratio

population (estimation) : femmes = 58,06 %    homme = 41,94 %

SEM = 8,86 %

$$SEM = \sqrt{\frac{pq}{n}}$$

NB : la précision dépend de la taille de l'échantillon, pas de la taille de la population

## Statistiques inférentielles : estimations

### **précision de estimation**

#### **intervalle de confiance (confidence interval)**

intervalle autour de la moyenne calculée de l'échantillon dans lequel la moyenne de la population a une probabilité donnée de se trouver.

*exemple : intervalle de confiance à 95 % : la valeur moyenne de la population dont est issu l'échantillon a 95 chances sur 100 de se trouver dans l'intervalle.*

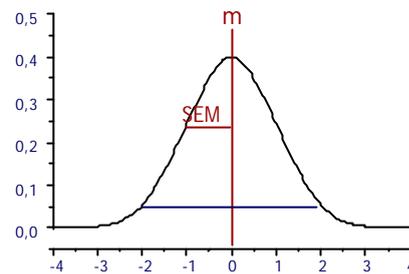
loi normale

dépend :

- ? de la SEM
- ? du % de confiance voulu
- ? du degré de liberté (ddl)

ex :

ddl > 30  
intervalle de confiance à 95 % =  
 $m \pm 1,96 \text{ SEM}$



## Statistiques inférentielles : estimations

### **précision de estimation**

#### **intervalle de confiance (confidence interval)**

exemple : sex ratio mesuré sur un échantillon de 31 personnes

sex ratio de l'échantillon (mesure) :

femmes = 18 (58,06 %)      homme = 13 (41,94 %)

sex ratio de la population (estimation) :

femmes = 58,06 %      homme = 41,94 %

SEM = 8,86 %

(n = 31 ; ddl = 30)

intervalle de confiance à 95 % (IC 95 %) =

femmes = 58,06 %  $\pm$  17.37 %      hommes = 41,94  $\pm$  17.37 %

## Statistiques inférentielles : estimations

### *précision de estimation*

### *intervalle de confiance (confidence interval)*

exemple : fréquence cardiaque mesurée sur 31 personnes (n = 31)

moyenne de l'échantillon (mesurée) : 86 battements/min  
écart-type de l'échantillon (mesuré) : 13,04 battements/min

moyenne de la population (estimée) : 86 battements/min  
écart-type de la population (estimé) : 13,25 battements/min  
SEM = 3,38 battements/min  
(n = 31 ; ddl = 30)

intervalle de confiance à 95 % (IC 95 %) =  
 $86 \pm 6,7$  battements/min

## Statistiques inférentielles : comparaisons statistiques

### *principe des tests : hypothèse nulle*

#### *principe des tests*

Les statistiques inférentielles permettent d'assigner une probabilité à l'obtention d'un résultat pour une hypothèse donnée.

Si cette probabilité est trop faible, on rejette l'hypothèse.

#### *hypothèse nulle (null hypothesis)*

Le principe des tests statistiques est de postuler l'hypothèse nulle : on fait l'hypothèse que les différences observées - entre des valeurs observées ou entre une valeur observées et une valeur théorique - est due aux fluctuations d'échantillonnage.

#### *conditions de rejet de l'hypothèse nulle*

Si la probabilité de l'hypothèse nulle est trop faible, on la rejette, et on accepte l'hypothèse non nulle : les échantillons comparés proviennent de populations différentes.

On dit alors qu'il existe une différence statistiquement significative.

## Statistiques inférentielles : comparaisons statistiques

### **principe des tests : hypothèse nulle**

*exemples :*

a) comparaison de la répartition H/F observée et de la valeur théorique du sex ratio 50 %

b) comparaison des fréquences cardiaques de groupes d'étudiants à la fréquence théorique « normale » de 70 battements/minute

c) comparaison des fréquences cardiaques des hommes et des femmes dans un groupe d'étudiants.

a) & b) hypothèse nulle : la population théorique dont le groupe d'étudiants est un échantillon représentatif n'est pas différente de la population « générale » dont on connaît les valeurs théoriques : sex ratio de 50 % et fréquence cardiaque de 70 batt/min.

c) hypothèse nulle : que la population théorique dont les étudiants masculins sont un échantillon représentatif est identique à la population théorique dans les étudiants féminins sont un échantillon représentatif.

## Statistiques inférentielles : comparaisons statistiques

### **conditions de rejet de l'hypothèse nulle**

Si la probabilité de l'hypothèse nulle est trop faible, on la rejette, et on accepte l'hypothèse non nulle : les échantillons comparés proviennent de populations différentes.

Il existe une différence statistiquement significative.

Par convention, on fixe en général le seuil de signification à 5 %

$p < 0,05$  : différences statistiquement significatives

$p < 0,01$  : différences statistiquement hautement significatives

$p < 0,001$  : différences statistiquement très hautement significatives

le seuil de signification est déterminé avant d'effectuer le test ; le degré de signification est déterminé par le test (= probabilité de rejeter l'hypothèse nulle si elle est vraie).

La différence est significative si le degré de signification est inférieur au seuil de signification.

## Statistiques inférentielles : comparaisons statistiques

### *risques d'erreur*

#### ***risque a (risque de 1<sup>re</sup> espèce) (type 1 error)***

risque de rejeter l'hypothèse nulle si elle est vraie.

Il est connu : seuil (à priori) ou degré (à postériori) de signification du test

#### ***risque b (risque de 2<sup>e</sup> espèce) (type 2 error)***

risque d'accepter l'hypothèse nulle alors qu'elle est fautive.

Le risque de 2<sup>e</sup> espèce correspond au défaut de puissance d'un test

Il est en général indéterminé (on ne connaît pas les caractéristiques des populations théoriques).

## Statistiques inférentielles : comparaisons statistiques

### *risques d'erreur*

#### ***risque a (risque de 1<sup>re</sup> espèce) (type 1 error)***

risque de rejeter l'hypothèse nulle si elle est vraie.

#### ***risque b (risque de 2<sup>e</sup> espèce) (type 2 error)***

risque d'accepter l'hypothèse nulle alors qu'elle est fautive.

Les deux types de risques sont antagonistes.

Si on diminue le risque de 1<sup>re</sup> espèce, on augmente le risque de 2<sup>e</sup> espèce.

Étant donné que le risque de 2<sup>e</sup> espèce n'est pas connu - à la différence du risque de 1<sup>re</sup> espèce - en absence de différence significative, on ne peut pas conclure à l'absence de différence, car on ne contrôle pas le risque d'erreur attaché à cette conclusion.

*Il y a une différence souvent oubliée entre ne pas conclure qu'il existe une différence, et conclure qu'il n'existe pas de différence.*

## Statistiques inférentielles : comparaisons statistiques

### *risques d'erreur*

#### **risque a (risque de 1<sup>e</sup> espèce) (type 1 error)**

risque de rejeter l'hypothèse nulle si elle est vraie.

#### **risque b (risque de 2<sup>e</sup> espèce) (type 2 error)**

risque d'accepter l'hypothèse nulle alors qu'elle est fautive.

#### **risque de se tromper**

Le risque de conclure à tort à une différence, c'est-à-dire le risque de conclure à une différence significative alors qu'il n'y en a pas, n'est pas le risque de 1<sup>re</sup> espèce.

En effet, il s'agit de la probabilité de rejeter l'hypothèse nulle et que l'hypothèse nulle soit vraie.

Donc, le risque de conclure à tort dépend du seuil de signification choisi mais également de la probabilité a priori que l'hypothèse nulle soit vraie.

## Procédures expérimentales

### *méthodologie*

- ? poser une question
- ? émettre une hypothèse
- ? élaborer une procédure expérimentale de test de l'hypothèse

(NB : test ? confirmation)

« Experimental design is not a means of proving that the chosen explanation is correct, but rather a system by which alternatives are eliminated »

Lumley & Benjamin. *Research: some grounds rules*

= critère de réfutation

(Karl Popper. *La logique de la découverte scientifique, Conjectures et réfutations*)

## Procédures expérimentales

### **procédure expérimentale**

#### ? **constitution d'un ou de plusieurs échantillons**

*« statistical techniques depend on the random selection of subjects (sample) from a defined population »*

*Lumley & Benjamin*

#### ? **choix des procédures expérimentales**

#### ? **choix des procédures statistiques**

définition de l'hypothèse nulle  
choix du seuil de signification  
choix du test

*« The identification of an appropriate statistical procedure is an important part of the design and should never be left after the data collection ».*

*Lumley & Benjamin*

## Procédures expérimentales

### **choix du test**

#### ? **types de variables** qualitatives / quantitatives

#### ? **nombre de variables**

#### ? **taille de l'échantillon**

#### ? **loi de répartition** normale ou non

#### ? **mesures répétées ou non / nombre de facteurs**

## Statistiques inférentielles : choix du test

### *liens entre variables qualitatives*

- ? échantillon de taille normale (effectifs calculés > 5) :
- ? test du  $\chi^2$  ( $\chi^2$ ).
  
- ? échantillon de taille réduite (effectifs calculés > 3) :
- ?  $\chi^2$  corrigé (correction de Yates)
  
- ? échantillon de taille très réduite (effectifs calculés < 3) :
- ? « test exact »

## Statistiques inférentielles : choix du test

### *liens entre une variable qualitative et une variable quantitative*

#### **comparaison de deux moyennes**

comparaison de deux moyennes observées  
comparaison d'un moyenne observée à une moyenne théorique

- ? les effectifs sont suffisamment grands ( $n > 30$ ) ou la loi de répartition est normale (faire éventuellement un test de normalité)
- ? test  $t$  de Student

#### **options du test :**

- comparaison d'un moyenne observée et d'une moyenne théorique (one population) ou de deux moyennes observées (two populations)
- mesures appariées (paired) ou non appariées (unpaired)
- comparaison unilatérale (one-tailed) ou bilatérale (two-tailed)

## Statistiques inférentielles : choix du test

### *liens entre une variable qualitative et une variable quantitative*

#### **comparaison de deux moyennes**

comparaison de deux moyennes observées

comparaison d'une moyenne observée à une moyenne théorique

? Les effectifs sont faibles et la répartition n'est pas normale  
(faire éventuellement un test de normalité)

? tests non paramétriques

#### **options :**

##### *séries non appariées :*

test *W* de Wilcoxon

test *U* de Mann et Whitney (équivalent du test précédent)

test *C1* de Fisher-Yates-Terry

##### *séries appariées :*

test *T* de Wilcoxon

## Statistiques inférentielles : choix du test

### *liens entre une variable qualitative et une variable quantitative*

#### **comparaison de plusieurs (= 2) moyennes**

? la loi de répartition de probabilité est normale pour la variable mesurée  
(faire éventuellement un test de normalité)

? Analyse de la variance (ANOVA)

#### **options :**

ANOVA à plusieurs facteurs

#### tests « post-hoc » :

*Méthode de Bonferroni* (test *t*) : recommandé pour un usage général, si les comparaisons ne sont pas trop nombreuses

*Méthode de Tukey* (test *t*) : à utiliser lorsque toutes les comparaisons paire par paire sont intéressantes

*Méthode de Dunnett* : à utiliser lorsque l'on compare le groupe contrôle avec les autres groupes, mais pas les autres groupes en eux.

*Méthode de Sheffé* (test *F*) : à utiliser pour les comparaisons compliquées

## Statistiques inférentielles : choix du test

### liens entre une variable qualitative et une variable quantitative

#### comparaison de plusieurs ( $\geq 2$ ) moyennes

? la loi de répartition de probabilité est normale pour la variable mesurée (faire éventuellement un test de normalité)

? Analyse de la variance (ANOVA)

? la loi de répartition de probabilité n'est pas normale pour la variable mesurée

(faire éventuellement un test de normalité)

? tests non paramétriques

#### tests de normalité

Martinez-Iglewicz \*

(Kolmogorov-Smirnov)

Agostino-Skewness (si  $n > 8$ )

Agostino-Kurtosis (si  $n > 20$ )

Agostino Omnibus\*

\* meilleurs tests

## Statistiques inférentielles : choix du test

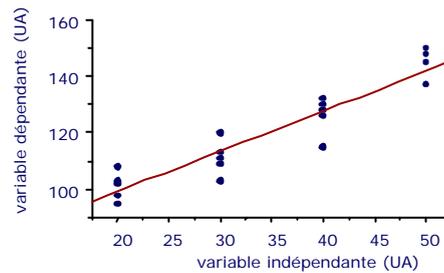
### lien entre deux caractères quantitatifs

#### exemple

Linear Regression

$Y = A + B * X$

Parameter	Value	Error
A	70,94	4,39668
B	1,416	0,11966
R	SD	N
0,94133	5,98312	20



? une des deux distributions liées au moins est normale avec une variance constante

? test de corrélation (ou de régression)

détermine si la pente est statistiquement significative de 0

? si aucune des variables liées n'est normale et de variance constante (petits échantillons)

? test non paramétrique de corrélation des rangs (test de Spearman)

## Statistiques inférentielles : choix du test

***Que faire quand on ne sait pas quoi faire ?***

demander à quelqu'un qui sait

faire appel à un statisticien  
(au moment de concevoir les protocoles)

## Statistiques : éléments de bibliographie

**P. Lazar & D. Schwartz. Éléments de probabilités et statistiques, Flammarion, Paris, 1987.**

*petit livre de base, avec exercices, pour s'initier de manière pratique aux probabilités et statistiques (BU)*

**R. Salamon. Statistique médicale, Masson, Paris, 1988.**

*Petit livre de base contenant l'essentiel des notions en statistiques, et une introduction au calcul des probabilités (BU)*

**D. Schwartz. Méthodes statistiques à l'usage des médecins et des biologistes, 4<sup>e</sup> édition, Flammarion, Paris, 1994.**

*ouvrage français de référence (BU)*

**T. H. Wonnacot & R. J. Wonnacot. Statistique, 4<sup>e</sup> ed, Economica, Paris, 1991.**

*Ouvrage détaillée (900 p) sur la statistique en économie, gestion, sciences et médecine, avec exercices d'applications (BU)*

**J. S. P. Lumley & W. Benjamin. Resarch: some ground rules, Oxford University Press, Oxford, 1994.**

*guide pour savoir comment mener un travail de recherche. N'est pas consacré particulièrement aux statistiques, mais une section est consacrée à l'analyse des résultats, avec une approche utilitaire des statistiques. (BU)*

## Statistiques : éléments de bibliographie

---

**J. Fowler, L. Cohen & P. Jarvis. *Practical statistics for field biology*, Wiley, Chichester, 1998.**

*Bonne introduction aux statistiques en général, bien qu'axé plutôt sur les statistiques de biologie d'observation.*

**S. J. Gould, *L'éventail du vivant*, Seuil, Paris, 1997. (titre original : *Full House*)**

*ouvrage de vulgarisation sur l'analyse des tendances de l'évolution biologique, présente de manière claire les biais possibles et les pièges à éviter dans l'analyse des répartitions asymétriques (en annexe, une introduction au jeu de base-ball).*

**D. M. Raup. *De l'extinction des espèces*, Gallimard, Paris, 1993 (titre original : *Extinction. Bad genes or bad luck?*)**

*Par un spécialiste de paléontologie statistique, l'analyse de la part du hasard dans les extinctions. Contient une présentation claire de quelques questions d'ordre statistique.*