

1

UE 4 : Evaluation des méthodes d'analyse appliquées aux sciences de la vie et de la santé

BIOSTATISTIQUE

G. MAIGNANT

Chercheur CNRS – Laboratoire RETINES – Faculté de Médecine de NICE




2

Date	Horaire	Enseignant	Intitulé du cours
Vendredi 6 septembre	10h15-12h15	PS	Mesure des phénomènes biologiques : types de données, unités, échelles, ordres de grandeur, précision
Vendredi 13 septembre	10h15-12h15	PS	Evénements, probabilités élémentaires et probabilités conditionnelles
Vendredi 20 septembre	10h15-12h15	PS	Indépendance en probabilité et théorème de Bayes
Vendredi 27 septembre	10h15-12h15	PS	Variables aléatoires et lois de probabilité discrète et continue
Vendredi 4 octobre	10h15-12h15	GM	Population, échantillon. Estimation ponctuelle d'un paramètre et estimation par intervalle
Vendredi 11 octobre	10h15-12h15	GM	Les principes des tests d'hypothèse. Tests paramétriques et non paramétriques
Vendredi 18 octobre	10h15-12h15	GM	Etude de la liaison entre variable quantitative et qualitative (comparaison de deux moyennes, échantillons indépendants et appariés)
Lundi 21 octobre	10h15-12h15	GM	Etude de la liaison entre deux variables qualitatives (test du Chi2 : indépendance, homogénéité, adéquation à un modèle théorique)
Vendredi 25 octobre	10h15-12h15	GM	Etude de la liaison entre deux variables quantitatives (corrélation, régression)

■ Probabilités

■ BioStatistiques

Lundi 28 octobre	10h15-12h15	LLP	Statistiques descriptives et indicateurs en épidémiologie : estimation, intervalles
Lundi 4 novembre	10h15-12h15	LLP	Valeur informationnelle d'un signe : sensibilité, spécificité et valeurs prédictives
Vendredi 8 novembre	10h15-12h15	PS	Raisonnement médical, arbre de décision
Vendredi 15 novembre	10h15-12h15	PS	Courbes de survie
Lundi 18 novembre	10h15-12h15	PS	Applications de l'informatique à la décision médicale
Vendredi 22 novembre	10h15-12h15	CP	Statistiques inférentielles et épidémiologie analytique : mesure des risques et puissance en épidémiologie
Lundi 25 novembre	10h15-12h15	LLP	Principes méthodologiques pour les essais cliniques

-  Probabilités
-  Statistiques
-  Applications

•1- **Cours en ligne**

Après les cours en amphi

•2- **Diapositives**

Cours (théorie, formules)

Exercices didactiques avec détails et rappels théoriques. QCM avec explications détaillées. **Concours = QRU**

•3- **Conseils pédagogiques**

Suivre tous les cours (**bien sûr !**)

Tutorat (**indispensable !**)

Noter les commentaires

Réfléchir sur les exercices en amphi



En fin de cours (fréquence à définir..) vos tutrices et moi-même, vous proposons une petite séance de QRU et questions/réponses peut être...

•4- **Cours** Méthodologie statistique. Logique, raisonnement : pas de formules « par cœur ». Pas de calculatrices, pas de calculs !

5

-1- Méthodes statistiques à l'usage des médecins et des biologistes

D. Schwartz.

Flammarion, Médecine Sciences.

-2- Biostatistique.

Beuscart, R. (Dir.) (2009).

Paris : Omniscience.

-3- Probabilités et statistiques

Alain-Jacques Valleron

Masson



Benjamin Disraeli, premier ministre britannique (1804 – 1881) :

« Il y a trois sortes de mensonges : les petits mensonges, les gros mensonges et les statistiques! »

Statistiques, enquêtes, sondages, moyennes, indices... sont diffusés dans les journaux écrits et télévisés.. Le vocabulaire des probabilités et des statistiques est couramment employé dans la vie quotidienne : **espérance de vie**, salaire **moyen**, **fréquence** des bus...

Mal comprise, ou mal utilisée, cette pratique peut conduire à des conclusions surprenantes, voire absurdes.

L'usage de la statistique devient abusif. Le grand public reste perplexe et pense "on fait dire ce que l'on veut aux chiffres".

Il y a 1,5 milliards de personnes dans le monde qui ne connaissent pas ou très peu de problèmes d'hémorroïde, ou de diverticulite (inflammation des parois de l'intestin).

Une étude statistique a prouvé que ces gens, majoritairement, faisaient leurs besoins accroupis.. Et non pas assis sur une cuvette comme par exemple les Européens, lesquels présentent très fréquemment des problèmes d'hémorroïde, ou de diverticulite..

Conclusion « évidente » : la position assise « à l'européenne », favorise l'apparition de problèmes d'hémorroïde, ou de diverticulite.

En fait, **une attitude critique** doit nous amener à nous poser d'autres questions : au-delà des constatations précédentes tous les autres paramètres sont ils identiques par ailleurs ? Alimentation, bien sûr, hygiène, pollution, stress, ...

**Attitude critique, indispensable dans le monde d'aujourd'hui :
Conclusions statistiques, info, réseaux sociaux, ...**

BIOSTATISTIQUE PLAN GÉNÉRAL DU COURS

8

1 - Biostatistique

2- Statistique Descriptive

3 - Statistique Déductive

- ***Liaisons entre caractères qualitatifs***
- ***Liaisons entre caractères quantitatifs***
- ***Liaisons entre caractères qualitatifs et quantitatifs***
- ***Tests non paramétriques***

9

1 - Biostatistique - Méthode

2 - Statistique Descriptive

3 - Statistique Déductive

- ***Liaisons entre caractères qualitatifs***
- ***Liaisons entre caractères quantitatifs***
- ***Liaisons entre caractères qualitatifs et quantitatifs***
- ***Tests non paramétriques***

Application des théories statistiques au domaine du vivant

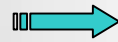


Le domaine de la Santé Publique

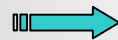
Décrire l'état de santé de populations

Evaluer des traitements, des techniques, des coûts

Mettre en place des observations épidémiologiques, en tirer des conclusions



Analyse descriptive



Analyse déductive

Objectifs pédagogiques :

Notion de variabilité

Définitions essentielles

Les types de variables

Statistique = art de collecter, d'analyser et d'interpréter des « données ».

Données = résultat de l'observation d'un individu, par l'utilisation d'un instrument de mesure, ou par les sens de l'observateur (signes cliniques, biologiques,..)

Cette donnée n'est intéressante que si on peut l'observer/la comparer sur plusieurs individus. Elle ne sera pas strictement équivalente d'un individu à l'autre. On parle donc de **variable**. On dira que la **variable** prend une valeur pour un individu, une autre valeur pour un autre individu, etc ..

On observe une grande **variabilité** des données dans le domaine biologique (taille, poids, groupe sanguin, température corporelle,..)

VARIABILITE



due au hasard



physiologique (intra ou inter sujets)

Paramètre = grandeur apportant une information résumée (ou synthétisée) sur la variable étudiée.

Exemple : moyenne d'une série de valeurs (notion détaillée plus loin..)

Les 2 domaines de la Statistique

12

- **Statistique descriptive** : Description d'une situation à l'aide de paramètres.
- **Statistique déductive (explicative ou inductive)** : Conclusions à partir d'observations et de mesures : **hasard ou autre explication** ?

Exemple : **2 traitements anti cancéreux donnent à 5 ans une survie de 42 % pour l'un et de 48 % pour l'autre. Hasard ou efficacité plus grande pour l'un des deux ?**

La statistique = méthode scientifique

Les statistiques = collections de données, dénombrements (au 17ème siècle, les premiers dénombrements sont demandés par l'état ==> mot "statistiques")

- **Série statistique** Collection d'objets de même nature, avec des caractéristiques différentes d'un objet à l'autre (**variables**).

- **Variables quantitatives** >> mesurables (instr de mesure)
- **Variables qualitatives** >> non mesurables (binaires, nominales...)

- **Population** Série exhaustive de **TOUS** les individus étudiés, sur lesquels on veut appliquer (inférer) des décisions.

- Exemple :** *Population de la France, des étudiants en médecine de Nice, des patients opérés d'une certaine pathologie entre 2 dates précises dans un service donné,...*

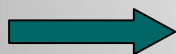
- **Echantillon** S/ensemble fini et d'effectif limité, extrait de la population.

Pourquoi échantillonner ?

- Etude sur l'échantillon et « pari » sur l'application des résultats à la population.
- Population inaccessible dans son entier pour des raisons d'organisation et de moyens limités

Comment échantillonner ?

- L'échantillon doit être *représentatif de la population*.



Tirage au sort (randomisation)

- **Echantillon connu, population inconnue**

15

- 1) Age
- 2) Sexe
- 3) Consommation tabac (nb cig/jour : 0 - 9 ; 10 – 19 ; >19)
- 4) Nb otites dans la dernière année
- 5) Déficit auditif moyen
- 6) Douleur articulaire (absente, modérée, intense)

Classer ces informations dans les différents types de variables définies :

Qualitative binaire, qualitative nominale, qualitative ordinale,

Quantitative discrète, quantitative continue



16

A) Qualitative binaire
D) Quantitative discrète

B) Qualitative nominale
E) Quantitative continue

C) Qualitative ordinale

Age

D : âges = **quantitative discrète**

Sexe

A : masculin ou féminin = **qualitative binaire**

Conso tabac

C : 3 niveaux de consommation
(faible 0 - 9, moyenne 10 - 19, forte >19) =
qualitative ordinale

Nb otites dans la
dernière année

D : nb entier = **quantitative discrète**

Déficit auditif moyen :

E : précision 1/10^{ème} (**11,5 dB**) = **quantitative continue**

Douleur articulaire (absente,
modérée, intense)

C : 3 niveaux = **qualitative ordinale**



17

Le nb d'élèves inscrits en classe de 5ème au Parc Impérial a Nice, est 112.
Un certain nombre de renseignements a été collecté auprès de ces élèves :
sexe, âge, temps du trajet domicile - collège, ...

Répartition des élèves

29 élèves en classe de 5e A (17 filles, 12 garçons)

30 élèves en classe de 5e B (18 filles, 12 garçons)

26 élèves en classe de 5e C (15 filles, 11 garçons)

27 élèves en classe de 5e D (16 filles, 11 garçons)

- 1) Population de référence : Elèves inscrits en 5ème # Population exhaustive
- 2) Quel est l'effectif total de cette population ? 112
- 3) Citer une variable qualitative : sexe
- 4) Citer une variable quantitative discrète : Nb garçons
- 5) Citer une variable quantitative continue : Temps parcours
- 6) Quel est l'effectif de la classe de 5e D ? 27
- 7) Fréquence des filles en 5ème ? $(17+18+15+16)/112 = 58,9\%$

Note Biostat	Rang Classement
12,4	210
4,9	555
18,1	6
5,4	445
19,4	5
16	14

Nature de ces variables ?

Quantitative ?

Qualitative?

Quantitative continue

Qualitative ordinale



variable qualitative ordinale → variable pseudo quantitative

Variation ordinale

Ex : variation de la douleur sur une échelle de numérique, taux de satisfaction de 1 à 5, rang de classement ...

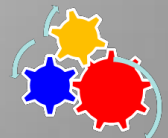
5 = Très bon, 4 = Bon, 3 = Moyen, 2 = Mauvais, 1 = Très mauvais

Ou bien ... 1 = Aucune douleur, 2 = Supportable, 3 = Moyen, 4 = Douloureux, 5 = Très douloureux

Peuvent être considérées comme variables « pseudos quantitatives » pour certains tests

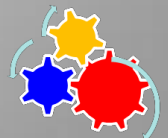
Attention : les nombres affectés aux modalités qualitatives n'ont pas de signification et ne peuvent faire l'objet d'opérations arithmétiques (calcul d'une somme ou d'une moyenne).

Etudié plus loin dans le cours



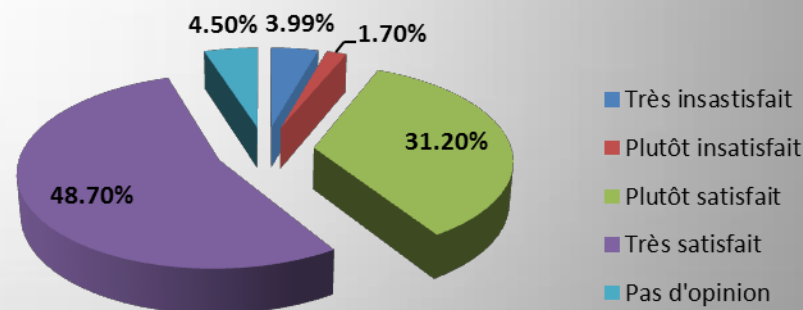
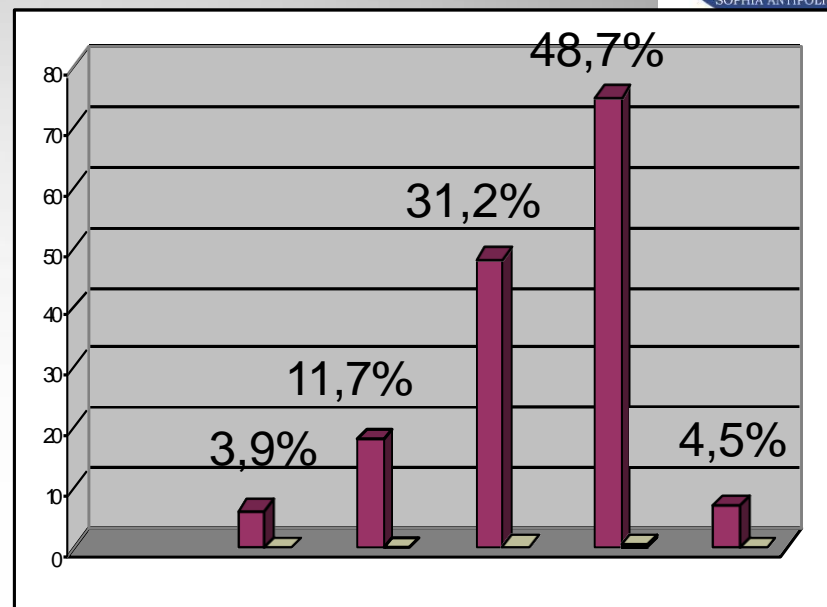
Exemple : Quel est le degré de satisfaction des mères accouchant dans une certaine maternité ?

- a) **Echantillon** : T.A.S mères ayant accouché dans cette maternité sur une période donnée (effectif $n=154$)
- b) **Variable étudiée** : degré de satisfaction (très insatisfait, plutôt insatisfait, plutôt satisfait, très satisfait, pas d'opinion) : variables qualitatives ordinales
- c) **Modalités de réponse = 5**



21

Degré de satisfaction	Nb mères	%
Très insatisfait	6	3,9%
Plutôt insatisfait	18	11,7%
Plutôt satisfait	48	31,2%
Très satisfait	75	48,7%
Pas d'opinion	7	4,5%



Variables qualitatives : tableau de %
histogramme, secteurs ...

Exemple : on s'intéresse aux poids des nouveaux nés dans cette maternité

- a) **Echantillon** : T.A.S mères ayant accouché dans cette maternité pendant une période donnée (effectif $n=1165$)
- b) **Variable étudiée** : poids du nouveau né : variables quantitatives
- c) **Données brutes** : n valeurs

23

Poids (g)	Nb bébés
2200	2
2300	2
2400	4
2500	5
...	
3100	121
3200	150
3300	162
3400	170

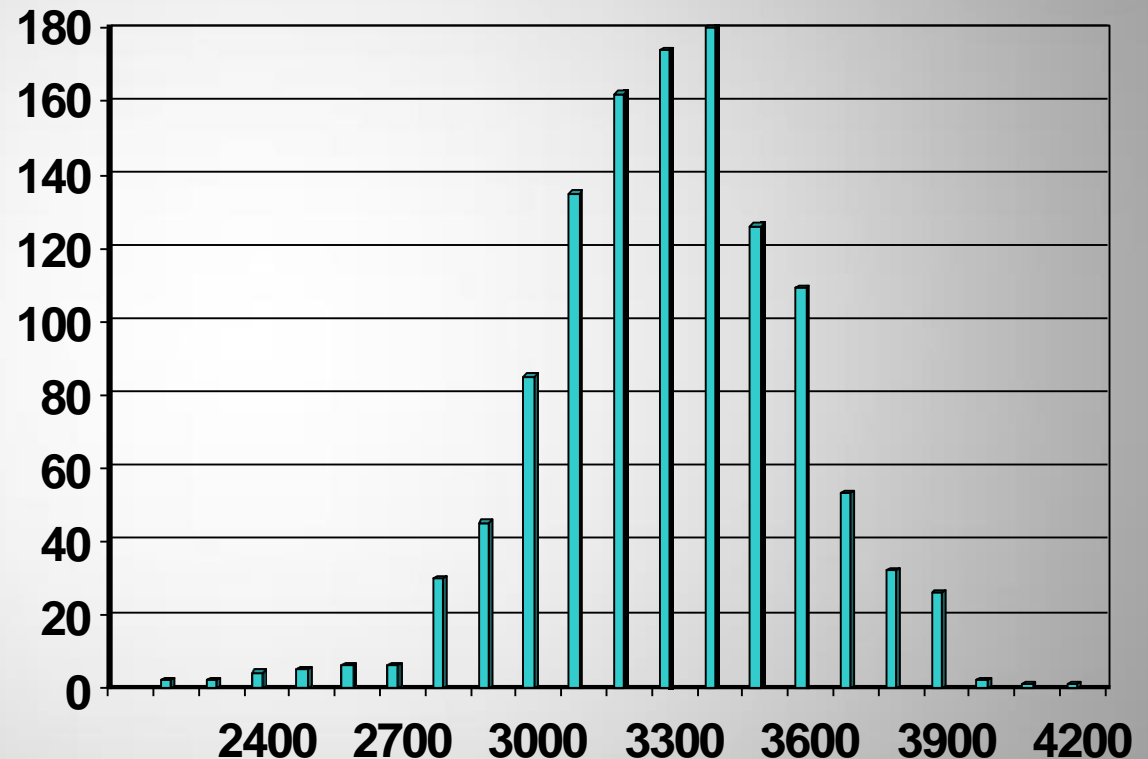


Tableau ou histogramme des effectifs
Mais pas seulement

24

On peut « résumer » en quelques paramètres les caractéristiques de la série de données quantitatives.

Moyenne : cas d'une variable quantitative **discrète**

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

cas d'une variable quantitative **continue**

$$m = \frac{\sum_{i=1}^n n_i x_i}{n}$$

Variance : paramètre indiquant la dispersion des données autour de la moyenne.

Médiane : valeur centrale si rangées par ordre croissant,

50% des valeurs < médiane et 50% des valeurs > médiane

(Exemple : {3,4,**6**,8,10})

Quartiles : Les quartiles partagent la série ordonnée en 4 groupes de même effectif (Exemple $Q_{25} = 1^{\text{er}}$ quartile : 25% de la série est < à cette valeur.)

Exemple : Soit une série de poids de bébés (n = 15 valeurs)

25

1	3400
2	2570
3	3210
4	4070
5	3840
6	4180
7	3480
8	3990
9	2640
10	3000
11	3830
12	1890
13	2350
14	2140
15	2470

Valeurs
rangées
par ordre
croissant

1	1890
2	2140
3	2350
4	2470
5	2570
6	2640
7	3000
8	3210
9	3400
10	3480
11	3830
12	3840
13	3990
14	4070
15	4180

Médiane : $n=15$ donc $(n+1)/2 = 8$
8^{ème} valeur = 3210

Si n pair, médiane située
entre les n° $n/2$ et $n/2+1$. **Moyenne
des 2 valeurs correspondantes**

3^{ème} quartile Q_{75} = $0,75 \times 15 = 11,25$

Le 3^{ème} quartile est situé entre le n°11 et
le n°12 soit $(3830+3840)/2 = \mathbf{3835}$

Moyenne de la série = 3137,3

Comparaison des caractéristiques de la moyenne et de la médiane

26

Moyenne

Avantages

- Facile à calculer, se manipule facilement dans les tests statistiques (sera très utilisée)
- Très significative si la répartition des données est assez symétrique et la dispersion faible

Inconvénients

- Sensible aux valeurs anormales (mini ou maxi)

Médiane

Avantages

- Calcul facile, peu sensible aux valeurs anormales
- Utilisable pour les valeurs ordinales, les classes, etc..

Inconvénients

- Se prête moins aux calculs statistiques.

Ex : Salaire moyen = 2500€, salaire médian = 1800€....10% des salariés > 4000€
Durée moy grippe = 8j durée médiane = 6j
En fait moyenne et médiane sont complémentaires !

27

1 - Biostatistique

2 - Statistique Descriptive

3 - Statistique Dédutive

- ***Liaisons entre caractères qualitatifs***
- ***Liaisons entre caractères quantitatifs***
- ***Liaisons entre caractères qualitatifs et quantitatifs***
- ***Tests non paramétriques***

28

- **Description de populations**
- **Estimation**
- **Notion de sondages**



Objectifs pédagogiques :
Variabilité
Echantillonnage
Intervalle de confiance
Risque α

Exemple :

Glycémie d'une population de sujets normaux = 1g/L.

En **moyenne**, taux de sucre dans le sang chez les sujets normaux = 1g/L

A quoi sert cette information ?

Patient avec une glycémie de 1,2 g/L : glycémie normale ou anormale?

Quelle est la variabilité normale de la glycémie, liée aux variabilités individuelles ?

Variabilité non maîtrisée



Biais

Variabilité maîtrisée



Estimation

RÈGLES DE BASE

Variabilité en Biologie

.....

Toujours

Individus

.....

Tous différents

Groupes similaires

.....

Résultats comparables mais non identiques.

30

Problème : déterminer un paramètre au niveau d'une population à partir d'observations réalisées sur un échantillon de cette population.

Exemples : - durée de séjour moyenne des patients hospitalisés en France, pour une pathologie donnée?
- durée moyenne d'attente aux urgences d'un hôpital ?

- **Estimation PONCTUELLE** Valeur, jugée la meilleure à un instant t (peu fiable)
- **Estimation par INTERVALLE** Intervalle de valeurs contenant la valeur recherchée

Intervalle de confiance autour de la valeur inconnue du paramètre

Méthodologie :

- **Détermination précise de la population étudiée = population cible**
- **Tirage au sort n sujets → Echantillon représentatif**
- **Calcul de l'intervalle de confiance**

Nous allons étudier dans l'ordre...

a) Données quantitatives : Estimation de la moyenne

Diapo 32 à 46 inclus

b) Données qualitatives : Estimation d'un pourcentage

Diapo 47 à 55 inclus

c) Synthèse Estimation

Diapo 56 à 60 inclus

a) Données quantitatives : Estimation de la moyenne

32

Echantillon

effectif = n

moyenne = m

écart type = s



ESTIMATION



Population cible

effectif = N

moyenne = μ

écart type = σ

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

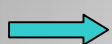
Estimateur de la moyenne vraie μ

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

Estimateur de l'écart type vrai σ

= racine carrée de la variance

Données quantitatives : Estimation de la moyenne



NOTION D'INTERVALLE DE CONFIANCE

$$\mu \in \left[m \pm \frac{\varepsilon S}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

α : Probabilité de se tromper dans l'estimation de μ

Plus α est petit, plus l'intervalle de confiance est grand : on réussit plus souvent. On s'expose aussi au risque de rater la "bonne" estimation. **Compromis universel** $\alpha = 5\%$

Intervalle de confiance à

$$\alpha = 5\% \quad \varepsilon = 1,96$$

$$\alpha = 1\% \quad \varepsilon = 2,6$$

Remarques

Différents échantillons

→ Différentes estimations

Taille échantillon augmente

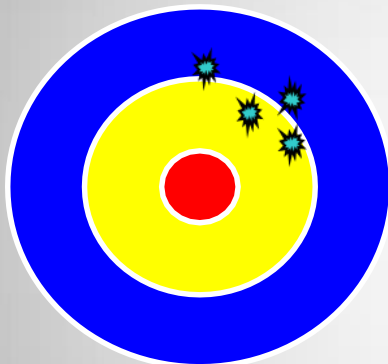
→ Estimation tend vers la moyenne vraie m .

C'est la méthode de calcul des normes des dosages biologiques

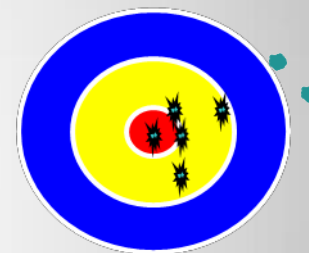
Précision de l'estimation

Intervalle de confiance peut être vu comme une cible

Large = plus de chances de l'atteindre, mauvaise précision de l'estimation



Resserré = risque de rater, meilleure précision de l'estimation



- a) **Indice** qui permet de calculer **la précision** de l'estimation de μ : $i = \varepsilon \frac{s}{\sqrt{n}}$
 Cette valeur est la largeur de l'intervalle de confiance.

Indice petit = meilleure précision

- b) Nombre de sujets nécessaires pour une précision donnée $n = \varepsilon^2 \frac{s^2}{i^2}$

35

Exercice : Précision estimation

s est connu = 0,1 g/l

Cas 1 : 2 mesures = 0,90 ; 0,96 g/l $m = 0,93$

$$IC_{95\%}(\mu) = \left[0,93 - 1,96 \frac{0,1}{\sqrt{2}} ; 0,93 + 1,96 \frac{0,1}{\sqrt{2}} \right]$$

$$\Rightarrow IC_{95\%}(\mu) = [0,7 ; 1,1] \quad \text{Précision de l'estimation } i = 1,96 \frac{s}{\sqrt{n}} = 0,2$$

Cas 2 : 3 mesures = 0,90 ; 0,96 ; 0,92 g/l $m = 0,93$

$$IC_{95\%}(\mu) = \left[0,93 - 1,96 \frac{0,1}{\sqrt{3}} ; 0,93 + 1,96 \frac{0,1}{\sqrt{3}} \right]$$

$$\Rightarrow IC_{95\%}(\mu) = [0,79 ; 1,07] \quad \text{Précision de l'estimation } i = 0,14$$

Cas 3 : 10 mesures $m = 0,93$

$$\Rightarrow IC_{95\%}(\mu) = [0,87 ; 0,99] \quad \text{Précision de l'estimation } i = 0,062$$

Effectif



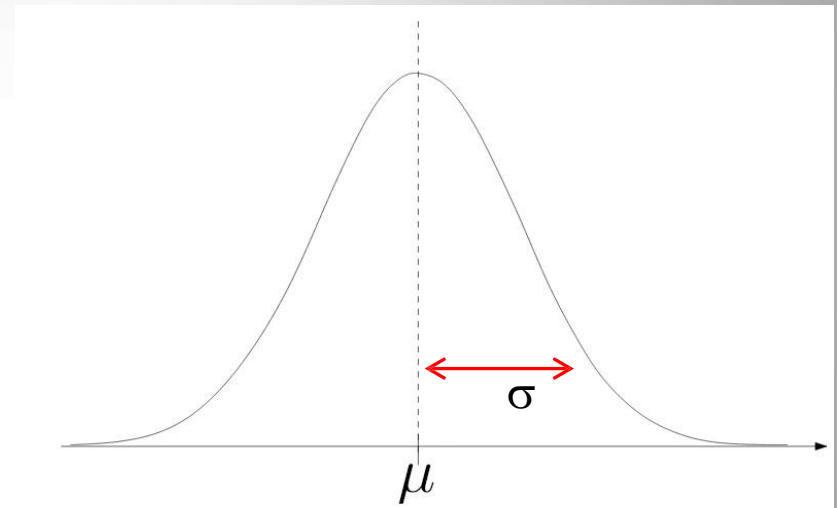
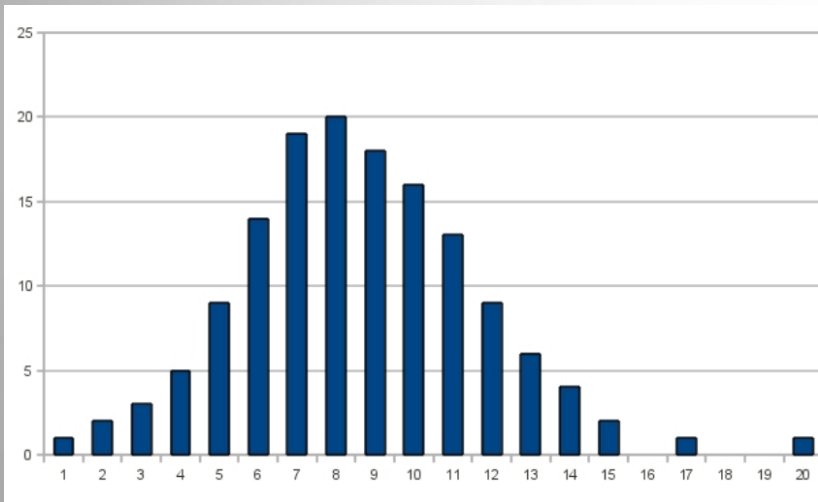
IC resserré, indice

et précision



En sciences humaines on observe souvent des distributions (X) plutôt symétriques autour de la moyenne avec une forme de cloche

Pour pouvoir faire des calculs, on va supposer que X suit une distribution « modèle », pour des **variables quantitatives continues** :
la Loi Normale

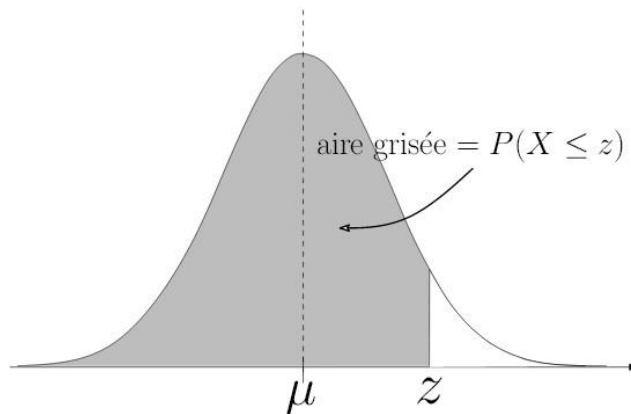


L'écart type caractérise la dispersion des données autour de la moyenne

Par exemple : Répartition des notes à un examen

A quoi ça sert ?

Pour chaque couple : moyenne, écart type (μ, σ)
il existe une loi normale de moyenne μ et d'écart type σ : $\mathcal{N}(\mu, \sigma)$



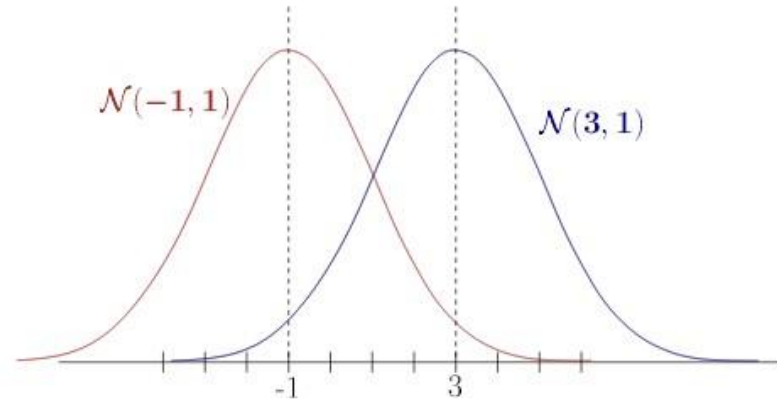
$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

L'aire grisée représente une proportion cumulée : la probabilité que $x \leq z$ donné.

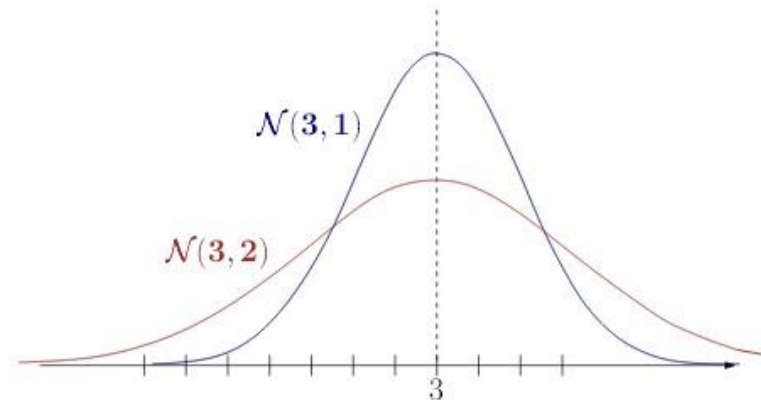
**Cadeau pour les matheux !
Formule inutile pour ce cours !!**

Echantillon effectif >30

Exemples de lois normales avec **moyennes différentes**, même écart-type :



Exemples de lois normales avec même moyenne, **écart-types différents** :



39

Quelques précisions :

Ecart type mesure la dispersion des données autour de la moyenne.

Exemple : répartition des notes d'un examen : plus le niveau des étudiants est homogène, plus les notes sont regroupées autour de la moyenne : **l'écart type est faible**.

Ecart type mesure la variabilité des données entre elles et par rapport à m
 → Si $(x_i - m)$ diminue alors s diminue.

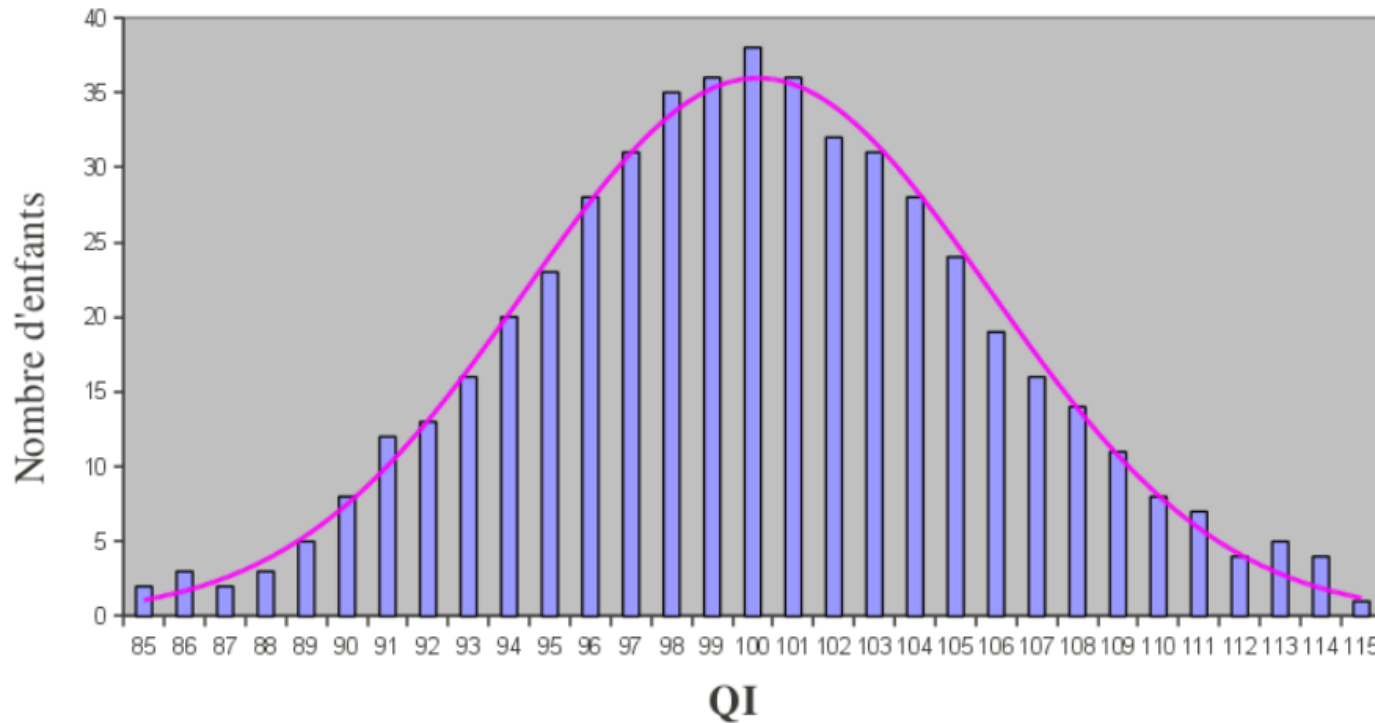
∈

Nb de degrés de liberté = nb des écarts **indépendants** $(x_i - m)$

Effectif = n : n écarts $(x_i - m)$, et leur somme = 0 $\sum (x_i - m) = \sum x_i - n.m = 0$

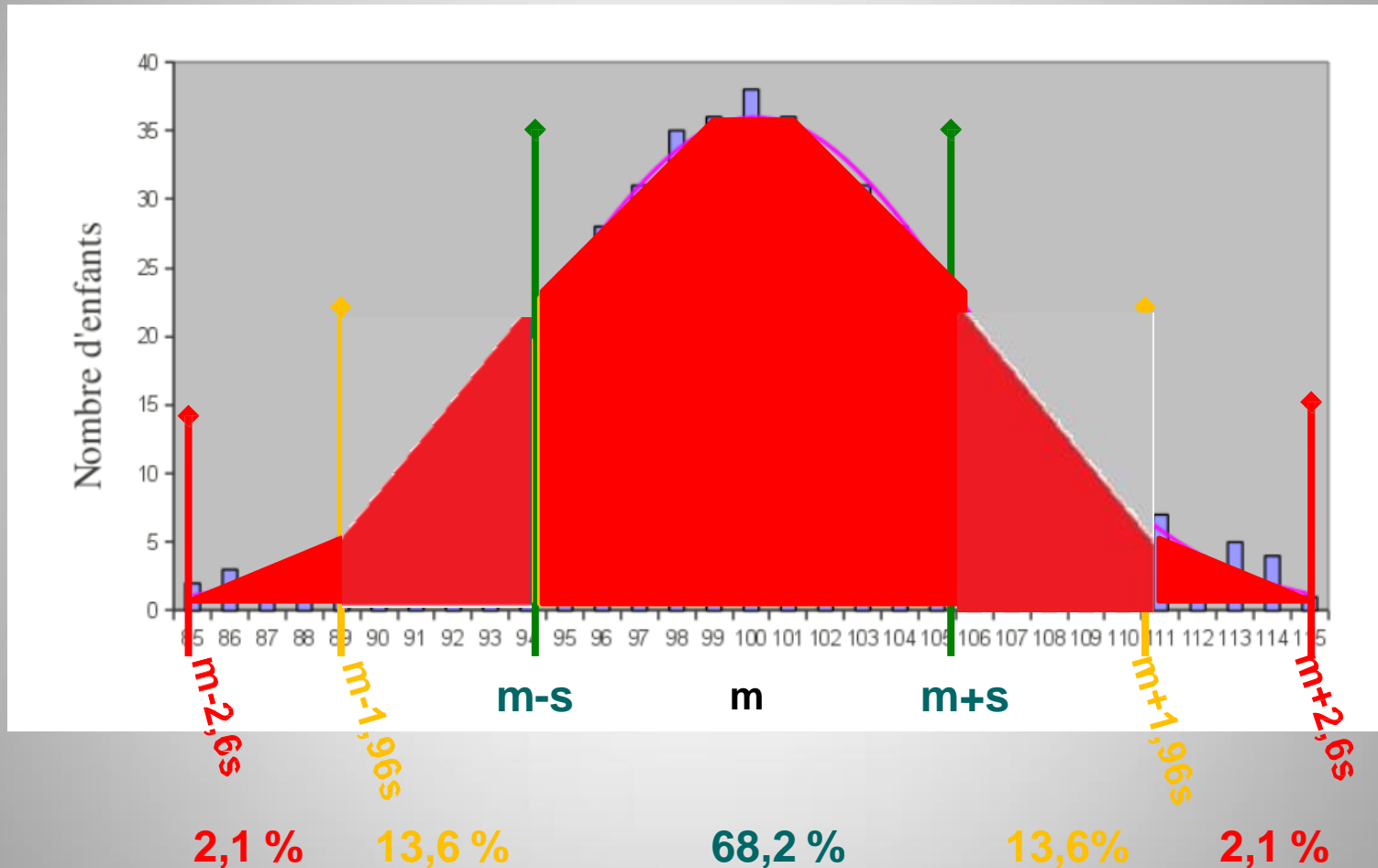
Il suffit de connaître $(n-1)$ valeurs pour les connaître toutes :

$n-1$ degrés de liberté.



Etude du QI de 515 enfants du même âge
 Courbe rose = courbe de la loi normale $\mathcal{N}(\mu, \sigma)$
 $\mu = 100, \sigma = 5,7$

Loi Normale ou de GAUSS
Comment l'utiliser ?



A partir de la Loi Normale ou de GAUSS, on précise :

Intervalles de confiance

- ★ $[m - 1 s \quad ; m + 1 s]$ contient 68,2% de la population
- ★ $[m - 1,96 s \quad ; m + 1,96 s]$ contient 95,4% de la population
- ★ $[m - 2,6 s \quad ; m + 2,6 s]$ contient 99,6% de la population

**Patient avec une glycémie de 1,2 g/l.
Quel est l'intervalle de confiance?**

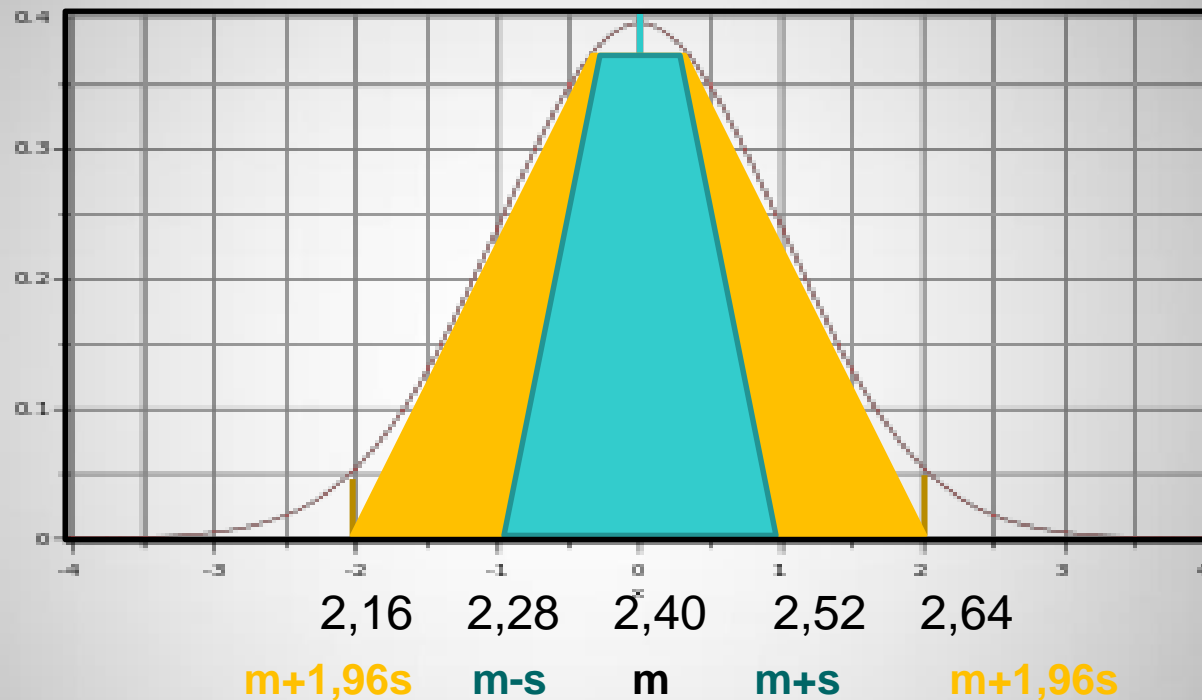
Intervalle de confiance au risque 5% = $[0,90 ; 1,22]$

**On ne peut répondre au patient qu'en connaissant l'intervalle de confiance
du dosage : NORMES**

Renseignements cliniques :		A jeun.		Normes
BIOCHIMIE - SANG				
Indice d'hémolyse		0		<2
<i>Analyseur Roche Cobas 8000.</i>				
Sodium		141	mmol/l	136-146
Potassium		3.97	mmol/l	3.50-5.00
Chlorures		103	mmol/l	98-107
<i>Potentiométrie indirecte. Analyseur Roche Cobas 8000.</i>				
Bicarbonates mesurés		29	mmol/l	22-29
<i>Technique photométrique UV. Analyseur Roche Cobas 8000.</i>				
Trou anionique		9	mmol/l	5-15
Protéines totales		68	g/l	64-83
<i>Test colorimétrique - Biuret. Analyseur Roche Cobas 8000.</i>				
Calcium total		2.33	mmol/l	2.15-2.55
<i>Spectrophotométrie NM-BAPTA. Analyseur Roche Cobas 8000.</i>				
Glucose		5.00	mmol/l	3.90-5.80
<i>Test enzymatique UV - Hexokinase. Analyseur Roche Cobas 8000.</i>				
	soit :	0.90	g/l	0.70-1.05
Urée		5.8	mmol/l	2.9-8.2
<i>Test UV enzymatique. Analyseur Roche Cobas 8000.</i>				
Créatinine		74	μmol/l	45-106
<i>Jaffe-Straucholide NMS-Corrigée. Analyseur Roche Cobas 8000.</i>				
Magnésium		0.86	mmol/l	0.70-1.05
<i>Spectrophotométrie Bleu de Xylylène. Analyseur Roche Cobas 8000.</i>				
Bilirubine totale		8	μmol/l	<21
Bilirubine conjuguée		3	μmol/l	0-4
<i>Photométrie - Diphényldiazonium. Analyseur Roche Cobas 8000.</i>				
Bilirubine non conjuguée		5	μmol/l	<17
LDH		414	U/l	200-480
<i>Test UV-DGKC-Substrat pyruvate. Analyseur Roche Cobas 8000.</i>				
		Résultat vérifié.		
ASAT (Transa.TGO)		32	U/l	10-50
ALAT (Transa.TGP)		32	U/l	10-50
<i>Technique IFCC avec PLP. Analyseur Roche Cobas 8000.</i>				

44

Dans un laboratoire pharmaceutique, un automate remplit des sachets de pilules. Il dépose en moyenne 2,40 g/sachets, avec un écart type de 0,12 g.



Quel % de sachets entre

2,16 g et 2,64 g **95%**

2,28 g et 2,52 g **68,2%**



45

Sur un échantillon de 114 personnes tirées au sort, on a calculé le taux moyen de cholestérol et l'écart type au niveau de cet échantillon

$$m = 195,4 \text{ cg}$$

$$s = 45,6 \text{ cg}$$

Comment s'écrit l'intervalle de confiance à 95% de la moyenne ?

A) $[195,4 \pm (2,56 \times 45,6 / \sqrt{114})]$

B) $[195,4 \pm (2,56 \times 45,6)]$

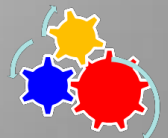
C) $[195,4 \pm (1,96 \times 45,6 / \sqrt{114})]$

D) $[195,4 \pm (1,96 \times 45,6)]$

E) Aucune réponse juste

Réponse C.

En effet IC_{95%} s'écrit avec $\varepsilon = 1,96$. On peut affirmer que cet IC donne bien une estimation de la moyenne μ inconnue au niveau de la population.



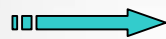
b) Données qualitatives : Fluctuation d'un pourcentage

46

Dans une population, quel % d'individus présentent un caractère donné ?

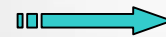
- Echantillon représentatif par tirage au sort (n sujets)
- Calcul d'un % qui tend vers la proportion cherchée, mais s'en écarte suivant une variabilité liée au hasard
- Autre échantillon → autre %

p_{obs}



Estimateur du pourcentage inconnu p

$$s = \sqrt{\frac{p_0 q_0}{n}}$$



Estimateur de l'écart type inconnu σ

avec $q_0 = 1 - p_0$

INTERVALLE DE CONFIANCE

$$p \in [p_{obs} - \varepsilon ; p_{obs} + \varepsilon]$$

$$\alpha = 5\% \quad \varepsilon = 1,96$$

$$\alpha = 1\% \quad \varepsilon = 2,6$$

PACES - UFR Médecine – Université Nice-Sophia Antipolis

Année universitaire 2019 - 2020

47

Exemple : précision d'un sondage

900 personnes ont été interrogées sur leur intention de vote à une élection présidentielle qui oppose 2 candidats A et B.

52% ($p=0,52$) ont déclaré qu'elles **voteraient A**.

Les journaux annoncent que le candidat A arrive en tête.

Vérification statistique de cette affirmation

Pour A $IC_{0,95} = [0,52 \pm 1,96 \sqrt{\frac{0,52 \times 0,48}{900}}] = [0,49 ; 0,55]$

Pour B $IC_{0,95} = [0,48 \pm 1,96 \sqrt{\frac{0,52 \times 0,48}{900}}] = [0,45 ; 0,51]$

52% et 48 % possèdent des IC contenant 50%

Les 2 candidats peuvent être considérés comme à égalité !

