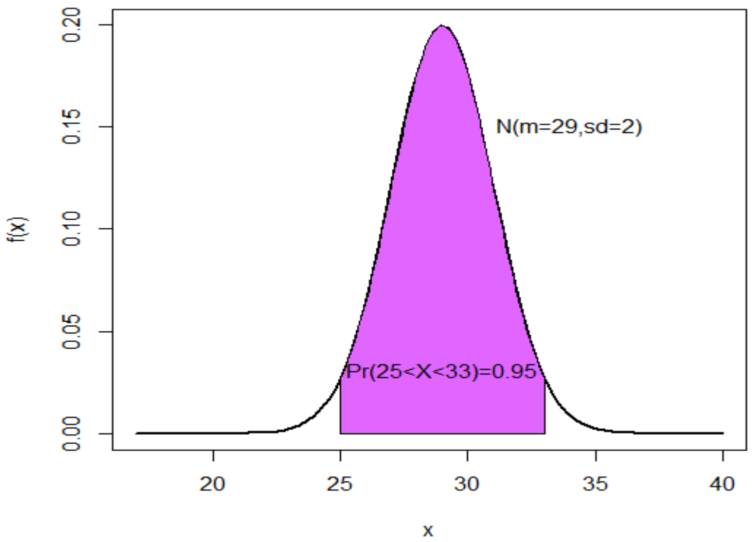


ESTIMATION DE PARAMÈTRES STATISTIQUES

Plan du cours

01 Estimations ponctuels versus estimations par l'intervalle de confiance

02 Intervalle de confiance de la Moyenne d'une population

03 Intervalle de confiance de la Fréquence (relative) d'une population

04 Intervalle de confiance de la difference de moyennes (fréquences)

Variables aléatoires continues: Loi Normale (de GAUSS)

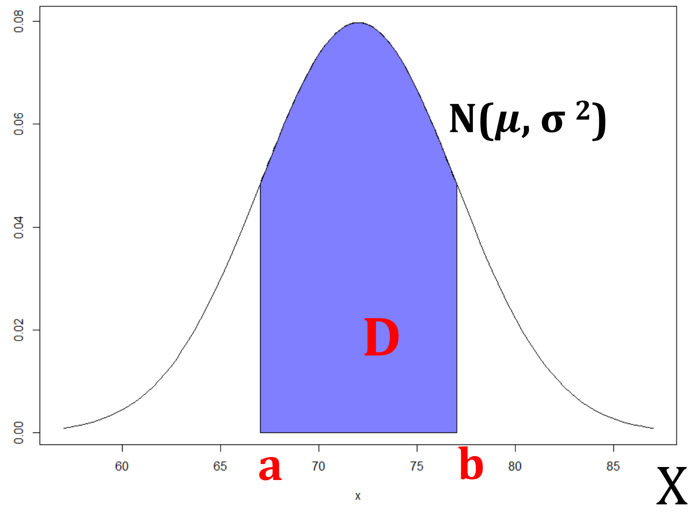
! Loi Normale: **N(μ, σ ²)** dépend de la moyenne et la variation/déviation standard

Fonction densité de probabilité (f):

f:**R** → [0, ∞)

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

- Fonction f est une fonction paire (la courbe est symétrique par rapport a l'axe Oy)
- **l'aire** du domaine compris entre l'axe d' abscisses et la courbe de f est égale a **1.**
- $Pr(a \leq X \leq b) =$ Aire (**D**)



$E(X) = \mu$

$V(X) = \sigma^2$

Applications:

le test Z - pour comparaison des moyennes/fréquences; intervalles de confiance pour les moyennes/fréquences

Variables aléatoires continues: Loi Normale centrée réduite

Loi normale centrée réduite N(0,1)

Pour simplifier l'utilisation de la lois Normale

On fait une transformation: $\boxed{Z = \dfrac{X - \mu}{\sigma}}$

Donc pour la variable aléatoire Z:

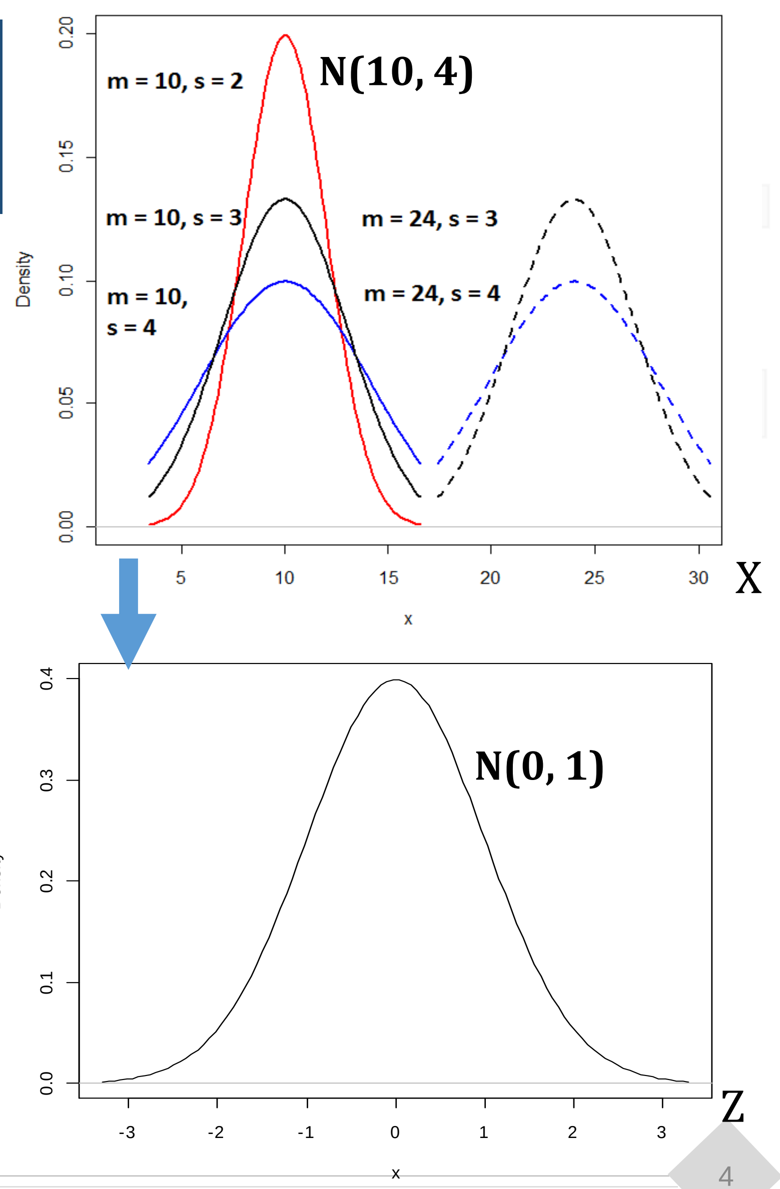
$\mu = 0$

$\sigma = 1$

Z: sans unité de mesure

La fonction densité de probabilité:

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



Loi Normale centrée réduite & Probabilité d'observer un événement donné

Pour le calcul de probabilités sans utiliser la fonction densité de probabilité, il y a des tables qui permet un utilisation pratique de ce lois.

Propriétés:

1. **Pr(Z ≥ u)=1 – Pr(Z<u), ou u>0 et** $Z = \frac{X-\mu}{\sigma}$
2. **Pr(Z≤-u)=Pr(Z≥ u) (**la courbe est symétrique par rapport a l'axe Oy)
3. **Pr(Z ≥-u)=1 – Pr(Z<-u)=1 –(1 – Pr(Z ≤ u))=Pr(Z ≤ u)**
4. **Pr(u≤ Z≤ v)=Pr(Z ≤ v) – Pr(Z ≤ u)**
5. **Pr(-u≤ Z ≤ u)=Pr(Z ≤ u) – Pr(Z ≤-u)=2·Pr(Z ≤ u)−1**

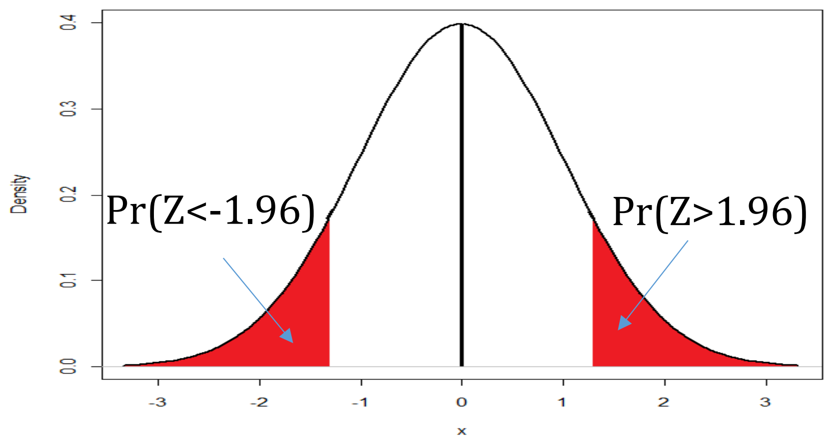
Exemple:

$Pr(\mu - \sigma \leq X \leq \mu + \sigma) = Pr(-1 \leq Z \leq 1) = ??$ ⇔

?? % des valeurs de la variable X sont dans l'intervalle **[μ-σ, μ +σ]**

$Pr(\mu - 1{,}96 \cdot \sigma \leq X \leq \mu + 1{,}96 \cdot \sigma) = Pr(-1{,}96 \leq Z \leq 1{,}96) = ??$

⇔ ?? % des valeurs sont dans l'intervalle **[μ-1,96σ; μ+1,96σ]**



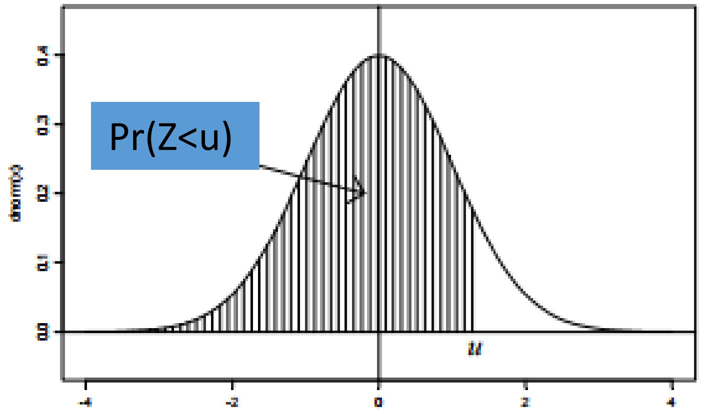
Variables aléatoires continues: Loi Normale centrée réduite

- Ce tableau donne l'aire sous la courbe normale centrée réduite

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,575
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,614
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,651
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,722
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,7823	0,785
0,8	0,7881	0,7910	0,7939	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,8106	0,813
0,9	0,8159	0,8186	0,8212	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,862
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,917
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,944
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,963
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	[illegible]	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,976
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,981
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,985
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,995
2,6	0,9953	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9961	0,9962	0,9963	0,996
2,7	0,9965	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9971	0,9972	0,9973	0,997
2,8	0,9974	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9979	0,9979	0,9980	0,998
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,999
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,999
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

deuxième décimale

Première décimale



Exemple: u=1,96 ⇒ $Pr(Z<1,96)=0,9750$

$Pr(-1,96 \leq Z \leq 1,96) = Pr(Z \leq 1,96) - Pr(Z \leq -1,96) =$

$= 2 \cdot Pr(Z \leq 1,96) - 1 = 2*0,975 - 1 = 0,95$

Exemple

- X: le poids (kg) chez les étudiants MDFR I
- **μ=72 kg, σ=5 kg**
- Questions:

a) Pr (77 ≤ X ≤ 87)=? b) Pr (67 ≤ X ≤ 77)=?

c) Pr (62.2 ≤ X ≤ 81.8)=?

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 72}{5}$$

77 ≤ X ≤ 87⇒ 5 ≤ X-72 ≤ 15⇒ 1 ≤ Z ≤ 3 ⇒

Pr(77 ≤ X ≤ 87) =Pr(1 ≤ Z ≤ 3)=??

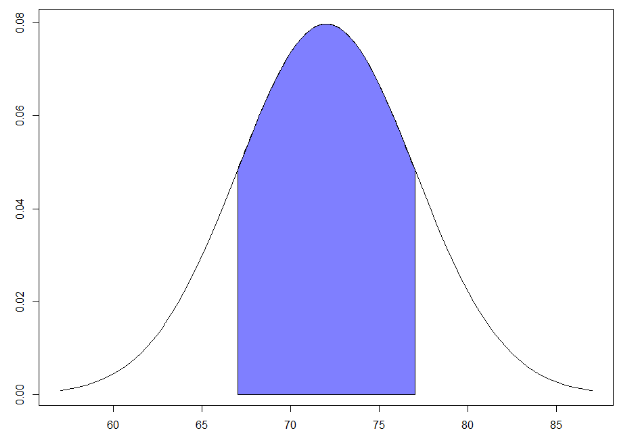
= Pr(Z ≤ 3) − Pr(Z ≤ 1)=0,9987 − 0,8413 = 0,1574

Pr(67 ≤ X ≤ 77)= Pr(−1 ≤ Z ≤ 1)=??

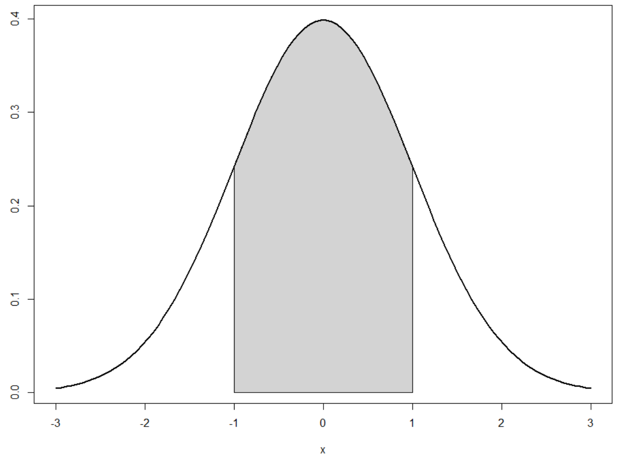
= 2Pr(Z ≤ 1)−1=2*0,8413 − 1 = 0,6826

Pr(62,2 ≤ X ≤ 81,8)=??=0,95

m -1,96*σ m +1,96*σ



X



Z

Statistique inférentielle

Estimation des paramètres (inconnus) d'une population

Tester des hypothèses statistiques

Principes généraux

Pour l’étude sur une population **P** des paramètres d’une certaine variable (quantitative ou qualitative), il est souvent nécessaire de suivre la procédure suivante:

1. Un échantillon représentatif de cette population est extrait.
2. A l’aide de la statistique descriptive, la distribution de la caractéristique sur l'échantillon extrait est décrite:
   - Variable qualitative: **fréquence (relative) observée**
   - Variable quantitative : **moyenne et l’ecart type.**
1. À l'aide de la statistique inférentielle, les résultats observés dans l'échantillon sont étendus à la population.

Intervalle de confiance de 95% (probabilité de 0,95): exemple d'un article scientifique

Randomized Controlled Trial > Clin Breast Cancer. 2022 Jun;22(4):359-366.
doi: 10.1016/j.clbc.2022.01.008. Epub 2022 Jan 31.

The Cardioprotective Effect of Vitamin D in Breast Cancer Patients Receiving Adjuvant Doxorubicin Based Chemotherapy

Noha A El-Bassiouny [1], Maged W Helmy [2], Mostafa Alaa Eldin Hassan [3], Gehan A Khedr [4]

Objectif principal de etude:

- étudier l'effet protecteur potentiel de la vitamine D (Vit D) sur la cardiotoxicité induite par la DOX (CIVD) chez les patientes atteintes d'un cancer du sein précoce recevant une chimiothérapie adjuvante à base de DOX (AC).

Table 2 Serum Levels of Vit D in Control Group and Vit D Group at Baseline and After 4 Cycles of AC

Vit D Baseline (ng/mL)	Control Group (n = 50)	Vit D Group (n = 50)
Min-Max	15.50 - 24.20	15.50 - 31.60
Mean ± SD	20.65 ± 2.22	21.37 ± 3.07
95% CI for mean	20.01 - 21.28	20.50 - 22.25
Vit D after 4x AC (ng/mL)	Control Group (n = 50)	Vit D Group (n = 50)
Min-Max	14.70 - 23.40	20.70 - 36.30
Mean ± SD	18.90 ± 2.27[a]	28.62 ± 4.81[bc]
95% CI for mean	18.25 - 19.54	27.25 - 29.99

Source: El-Bassiouny NA, Helmy MW, Hassan MAE, Khedr GA. The Cardioprotective Effect of Vitamin D in Breast Cancer Patients Receiving Adjuvant Doxorubicin Based Chemotherapy. Clin Breast Cancer. 2022 Jun;22(4):359-366. Lien vers l'article: https://pubmed.ncbi.nlm.nih.gov/35241369/

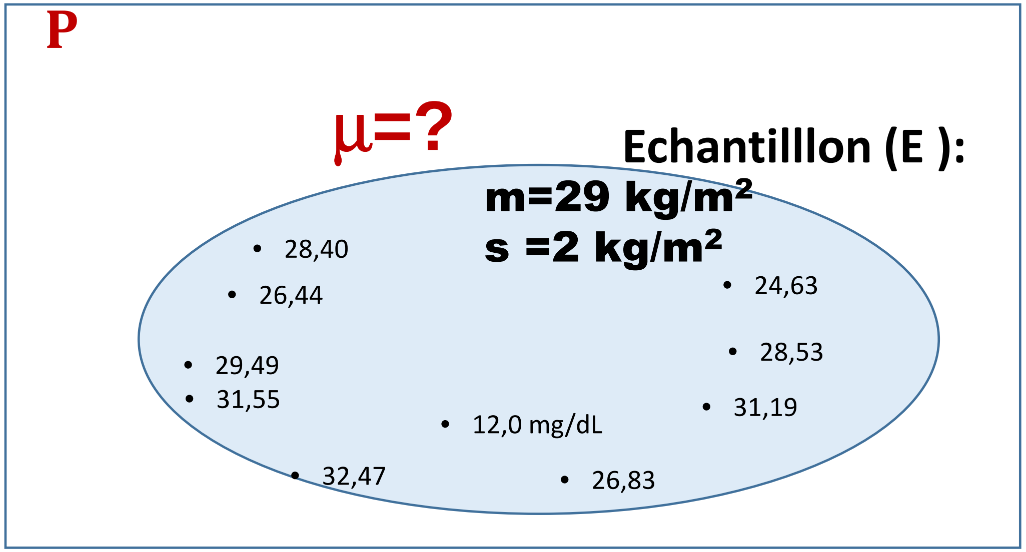
Cas d’une variable quantitative

Moyenne theorique μ de la variable X et la variance theorique σ^2 sur la populationP sont inconnus.

- ✓ un échantillon **E** représentatif de la population P est extrait de manière aléatoire.
- ✓ dans l'échantillon E on observes la moyenne ***m*** et la variance s^2 de la variable X.
- ✓ nous essayons d'estimer les valeurs inconnues de la $\boldsymbol{\mu}$ et $\boldsymbol{\sigma^2}$ à l'aide de **m** et de $\mathbf{s^2}$ observés.

Exemple

Population (**P**): **moyenne μ, variance σ^2 de la variable X.**



Quel est l'indice de masse corporelle (IMC) moyen chez les patients diabétiques?
Sur un échantillon de 10 patients, les résultats ont été obtenus:

28.40; 26.44; 29.49; 31.55; 31.40; 32.47; 24.63; 28.53; 31.19; 26.83

Statistique inférentielle: estimations

estimation d'un paramètre inconnu sur une population à l'aide de statistiques déterminées sur un échantillon représentatif.

Les statistiques déterminées sur l'échantillon:

Le symbole ^ est souvent utilisé pour décrire le paramètre à estimer

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

$$\widehat{\sigma^2} = S^2 = \frac{\sum_{i=1}^{n} (x_i - m)^2}{n-1}$$

Paramètres inconnues de la population:

E

Échantillon (s) de taille n

X: variable(quantitative)

x_i-Valeurs de X mesurées sur l'échantillon

$$\mu = \frac{\sum_{i=1}^{N} x}{N} \quad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Extrapoler le résultat au niveau de la population P

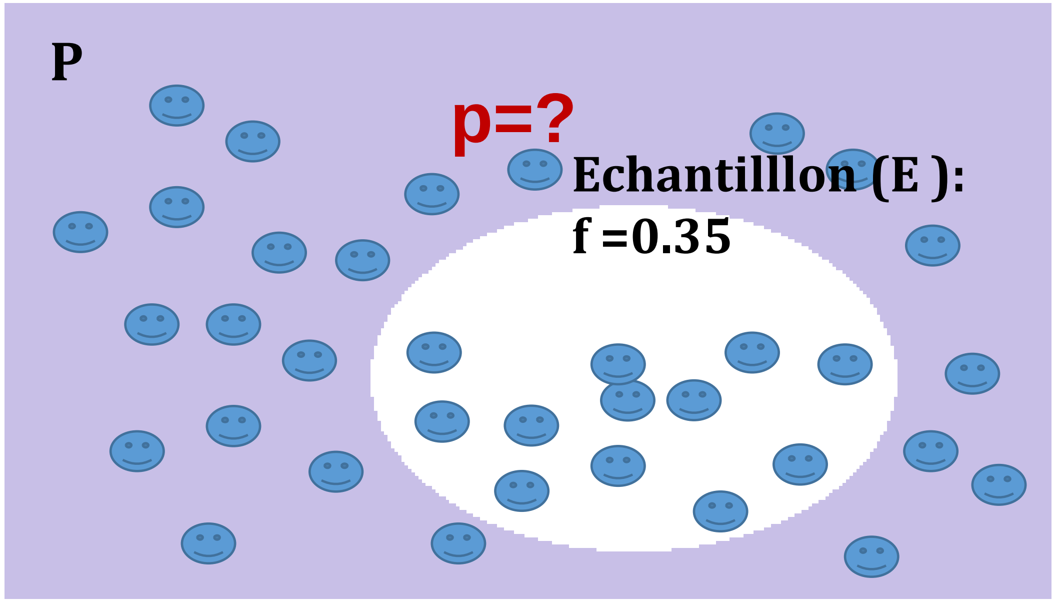
Cas d’une variabile qualitative (X)

La fréquence relative théorique p de la variable X sur la P est inconnue.

- un échantillon E représentatif de la population P est extrait de manière aléatoire.
- dans l'échantillon E pour la variable X, une fréquence **f** est observée.
- nous essayons d'estimer la fréquence théorique **p** en utilisant sa fréquence observée (**f**).

Exemple

Population(**P**)**:** la fréquence relative **p (inconnue)** de la variable X.



Quelle est la proportion d'obeses parmi les patients diabétiques? Sachant que sur un échantillon de 100 diabétiques, 35% étaient obese.

Types d'Estimations

Population

Echantillon

Proportion observée (**f**)

Proportion theorique **p**
? – estimation ponctuelle
[a, b]– intervalle de confiance

Moyenne observée (**m**)

Moyenne theorique **μ**
? – estimation ponctuelle
[a, b]– intervalle de confiance

Types d'estimations:
estimation ponctuelle et estimation par intervalle de confiance

L'estimateur ponctuel = une statistique de l'échantillon permettant d'estimer **sans biais** la valeur du paramètre sur la population.

- La moyenne de l'échantillon est l'estimateur ponctuel de la moyenne théorique μ.
- La fréquence relative (f) de l' échantillon est l'estimateur ponctuel de la fréquence relative théorique (p) sur la population

Intervalle de confiance (notation: IC) = intervalle de valeurs [a, b] qui contiendra le paramètre de population avec une probabilité élevée.

L'estimateur ponctuel avec biais versus estimateur sans biais

Définition: Un estimateur T d'un paramètre θ sur une population est un estimateur **ponctuel sans biais** si l'espérance mathématique M (T) = θ (l'estimation est correcte),

Exemple:

- moyenne **m de l'echantillon**: **estimateur ponctuel sans biais pour la moyenne** μ

  M(m) = μ

- fréquence relative **f**: **estimateur ponctuel sans biais pour la** fréquence théorique p

  M(f) = p

- Si M(T) ≠ **θ** (estimateur avec biais, l'estimation est incorrecte),

Exemple:

- variance standard descriptive (s^2): **estimateur avec biais** pour la variance σ^2 parce que **$E(s^2) \neq \sigma^2$ ou** $E\left(s^2\right) = \frac{n}{n-1}\sigma^2 \neq \sigma^2$

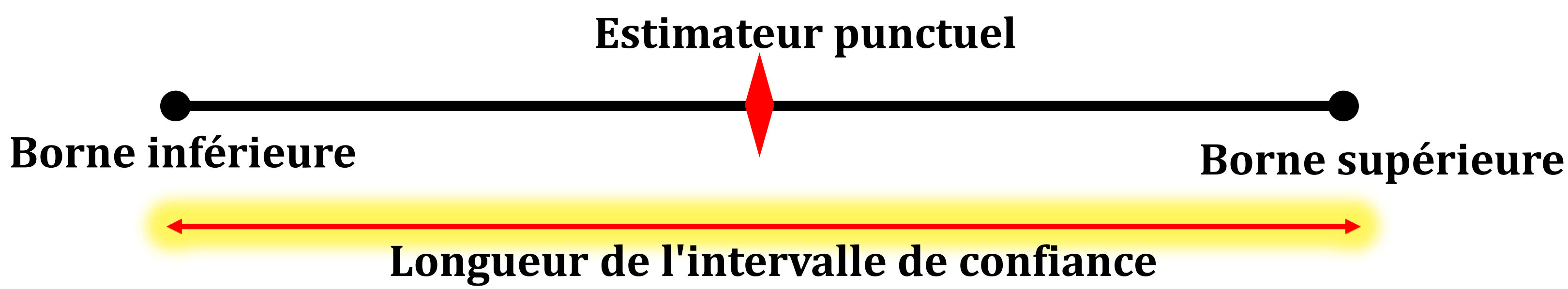
Estimateurs ponctuels

Le paramètre sur la population que l'on souhaite estimer	L'estimateur ponctuel du paramètre
Moyenne (μ)	m
Variance (σ^2)	$S^2 = \frac{n}{n-1} s^2 = \frac{\sum_{i=1}^{n}(x_i - m)^2}{n-1}$
Deviation standard ou $\mathrm{l'ecart - type(\sigma)}$	$S = \sqrt{S^2}$= écart type de l'échantillonnage
Frequence relative (p)	f

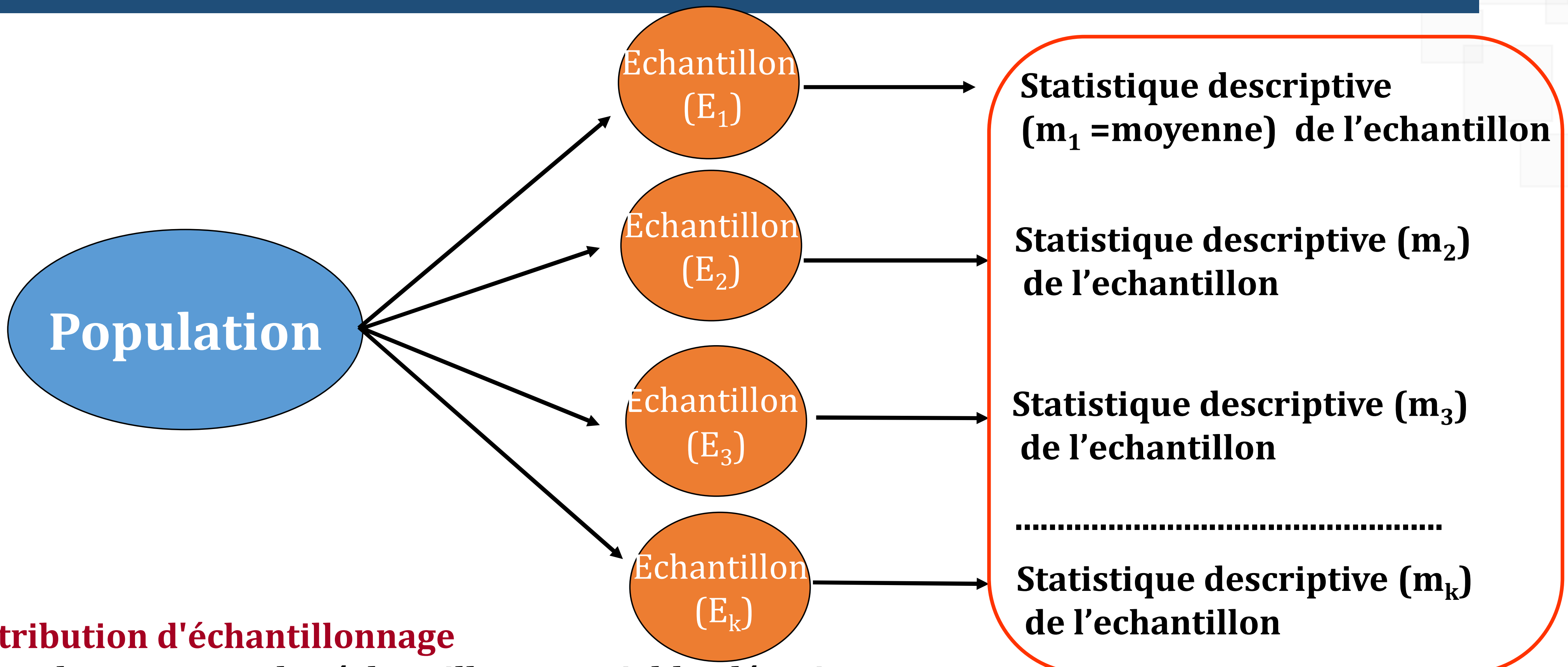
Notations: m = ***moyenne calculée sur l'échantillon****; s^2 = la variance descriptive calculée sur l'échantillon; f = la fréquence relative observée sur l'échantillon*

L’estimateur ponctuel vs. l’intervalle de confiance

- **L'estimateur ponctuel** = une valeur unique qui dépend des fluctuations de l'échantillonnage et peut être plus ou moins éloignée de la valeur du paramètre estimé.
- Il est conseillé d’estimer un paramètre non par une seule valeur mais par un intervalle (pas un intervalle) mais par
- **l'intervalle de confiance** qui peut être défini comme l’intervalle contenant le paramètre estimé avec une probabilité élevée.



Distribution d'échantillonnage



Distribution d'échantillonnage

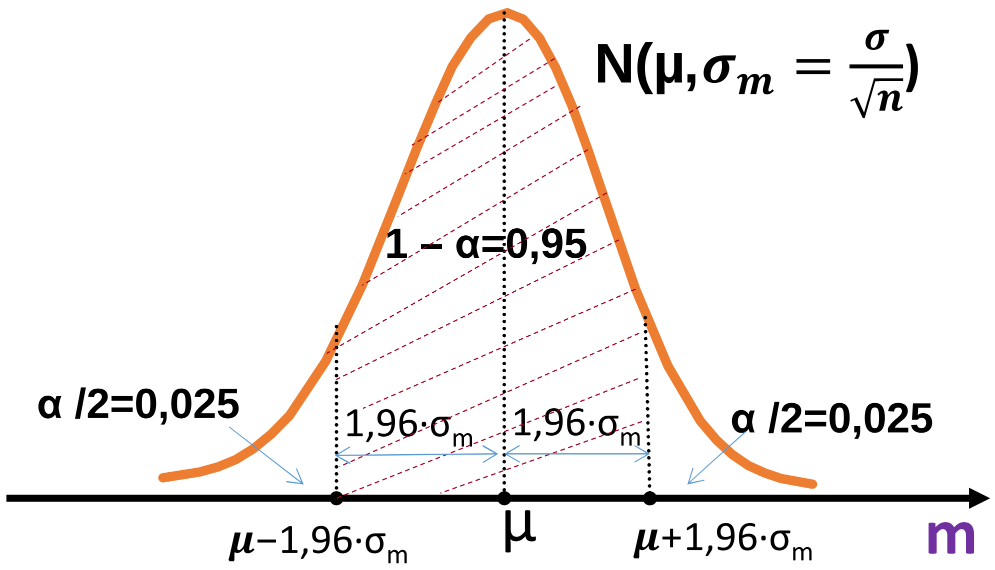
Exemple: moyenne des échantillons≈ variable aléatoire

-> distribution des moyennes des échantillons = distribution d'échantillonnage de la moyenne.

THÉORÈME DE LA LIMITE CENTRALE

- **CAS D'UNE VARIABLE QUANTITATIVE:**

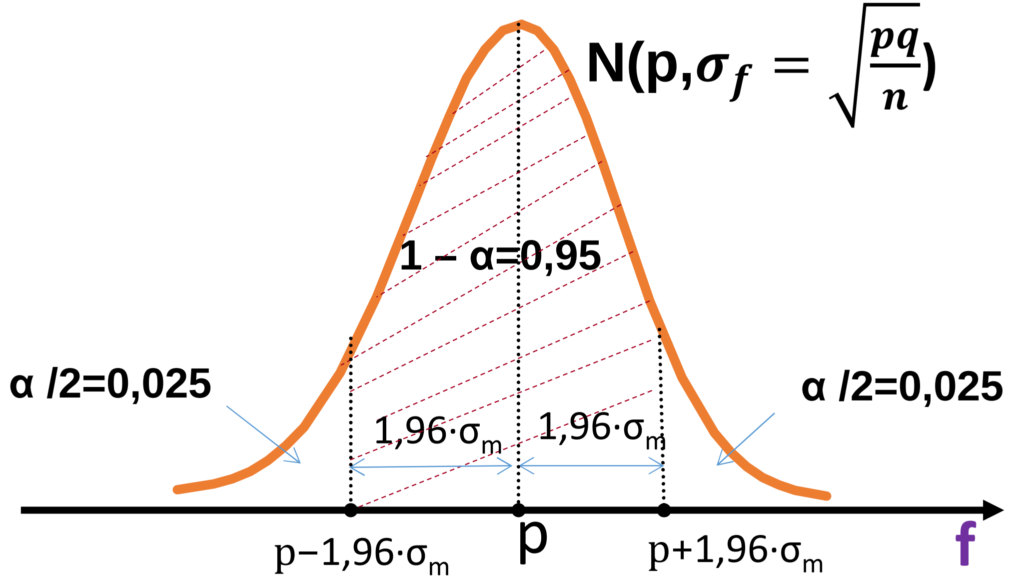
La distribution des moyennes des grands échantillons (n≥30) de taille *n*, extraite d'une distribution de moyenne μ et de variance σ peut être approximée par une distribution normale de moyenne μ et l'écart type $\sigma_m = \frac{\sigma}{\sqrt{n}}$ (appelée *l'erreur standard*).



THÉORÈME DE LA LIMITE CENTRALE

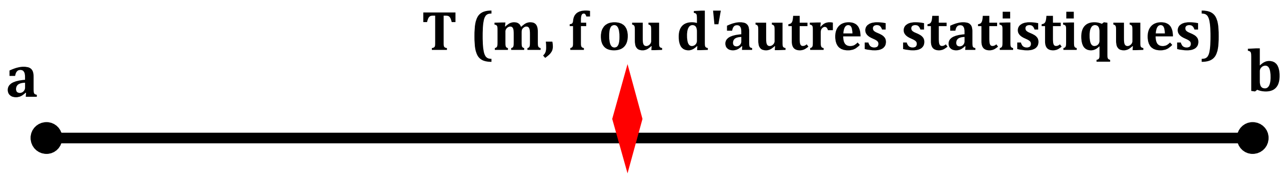
- **CAS D'UNE VARIABLE QUALITATIVE:**

La distribution des fréquences des grands échantillons (*np ou n*(1-p) ≥ 10)* de taille *n*, peut être approximée par une distribution normale de moyenne p et l'écart type $\sigma_f = \sqrt{\frac{pq}{n}}$ (appelée l'erreur standard).



Intervalle de confiance (IC)

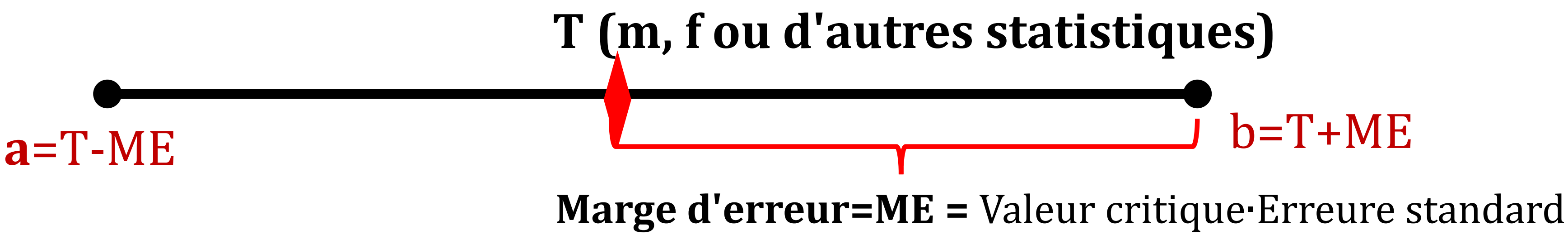
* intervalle aléatoire [a, b] centré sur l'estimateur ponctuel T du paramètre θ.
* Intervalle de valeurs qui contiendra le paramètre θ avec une probabilité de $1 - \alpha$: $\Pr(a \leq \theta \leq b) = 1 - \alpha$
* **Niveau de confiance ($1- \alpha$) choisi *a priori*: généralement égal à 0.95**
* (mais peut être: 0.90, 0.99).
* Niveau de signification: $\alpha=0.05$
* Si $\alpha = 0.05$ => **95% IC: [a; b]**



Intervalle de confiance (IC)

- **Intervalle de confiance**:
  - Considérer la variabilité d'échantillonnage ➔ il a des valeurs différentes pour chaque échantillon
  - Il peut être calculé a l'aide d'un seul échantillon
  - Fournit des informations sur le paramètre inconnu de la population
  - **Formule generale:**

    Estimateur ponctuel T ± (Valeur critique)×(Erreure standard)

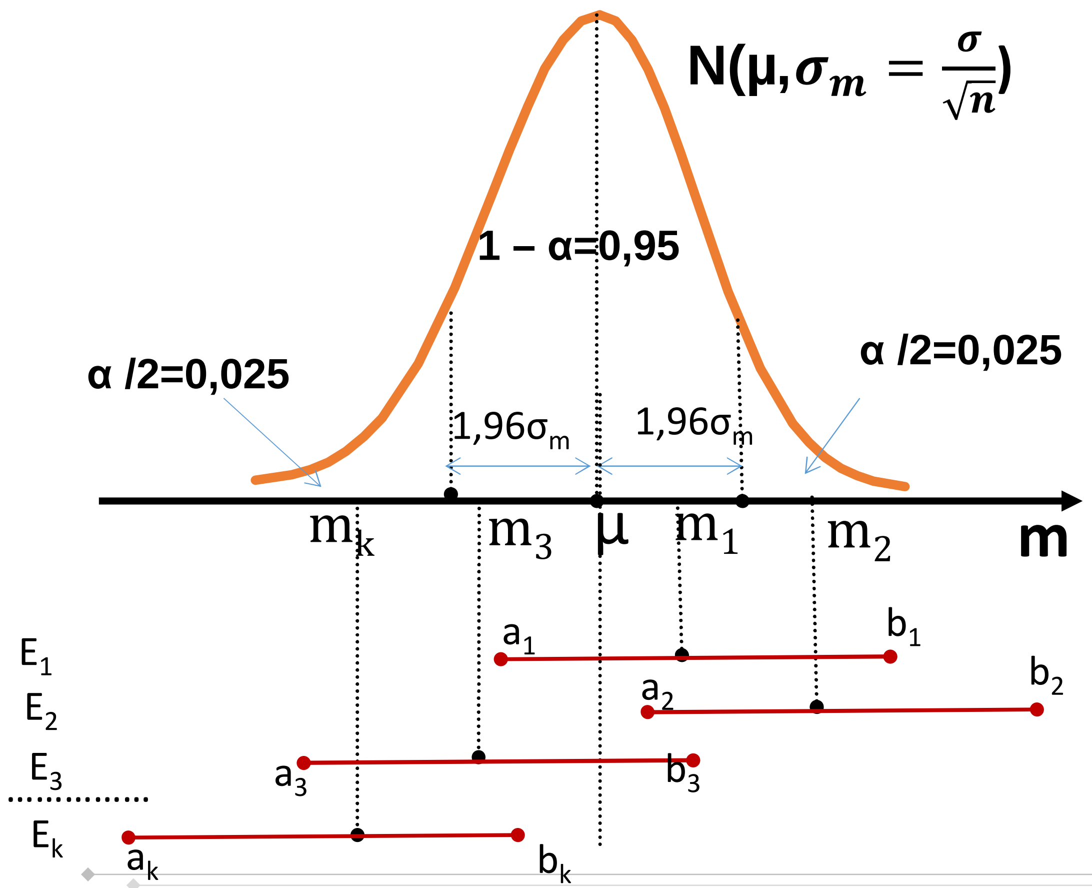


Intervalle de confiance (IC)

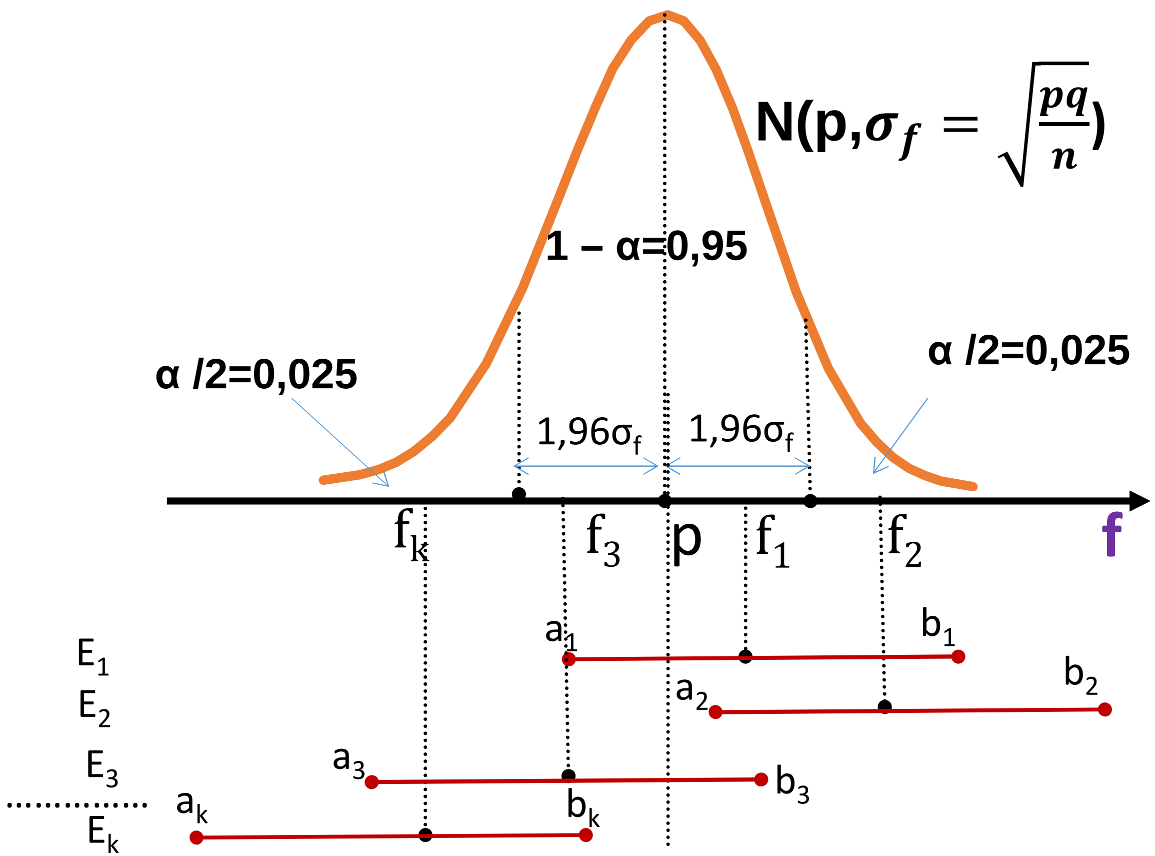
Longueur de l'intervalle de confiance:

- est d'autant plus petit que le volume de l'échantillon est grand.
- varie en fonction du niveau de signification (α)
- **Il ne peut jamais y avoir 100% de confiance.**
- **Interpretation probabilistique:**
  - Si **tous les échantillons possibles de volume = n** ont été extraits de la population et les moyennes et les intervalles de confiance associés ont été calculés, **95% des intervalles de confiance contiennent** le paramètre (moyenne ou fréquence) de population.
  - **Interpretation pratique:** Nous sommes sûrs à 95% que l'intervalle de confiance [a, b] contient le paramètre de population.

Intervalle de confiance pour la moyenne



Intervalle de confiance pour la frequence



ESTIMATION DE LA MOYENNE DE POPULATION CAS DE GRANDS ÉCHANTILLONS (n≥30)

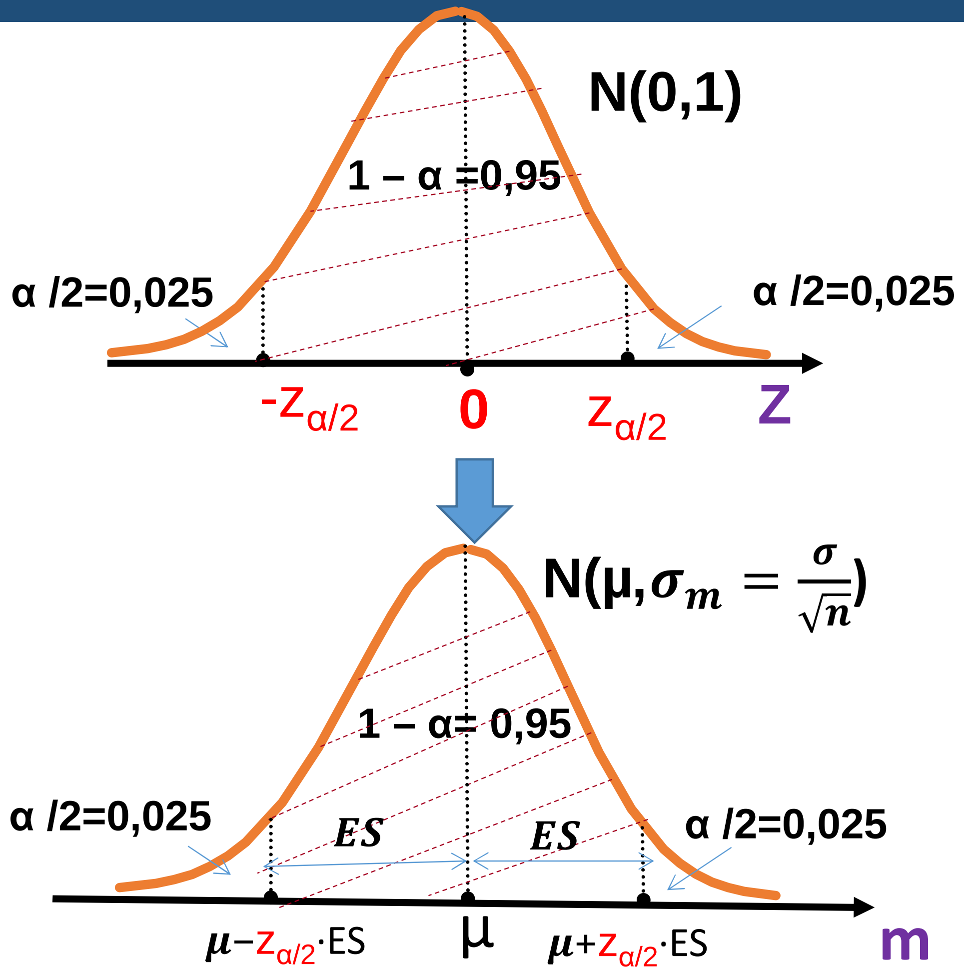
❑Principe:

- population(P): moyenne μ, l' écart type σ.

Théorème de la Limite Centrale:

- moyenne de la distribution d' échantillonnage: $\mathbf{m} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- variable aléatoire : $Z = \frac{\mathbf{m}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$ ou $\frac{\sigma}{\sqrt{n}}$=ES (l'erreur standard)
- [a, b]=? de sorte que $Pr(a \leq \mu \leq b) = 1 - \alpha$

ESTIMATION DE LA MOYENNE DE POPULATION
CAS DE GRANDS ÉCHANTILLONS (n≥30)



$$\Pr(a \leq \mu \leq b) = 1 - \alpha = 0{,}95$$

$$\Pr(? \leq Z \leq ?) = 0{,}95$$

$$-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$$

$$-Z_{\alpha/2} \leq \frac{m - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha/2}$$

$$\underbrace{m - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{a} \leq \mu \leq \underbrace{m + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{b}$$

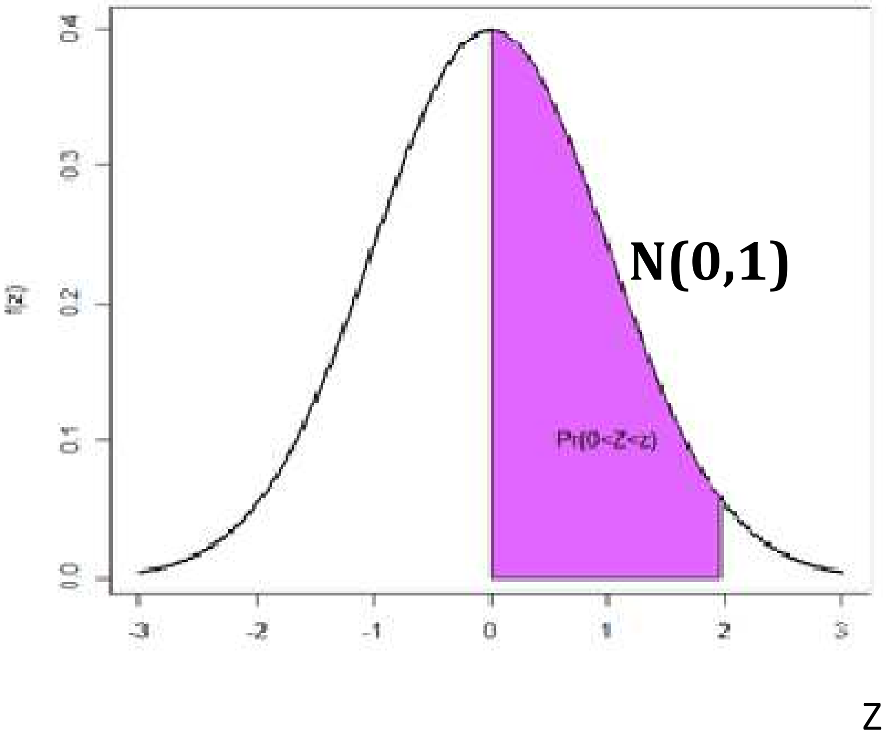
$$\boldsymbol{m - Z_{\alpha/2}ES \leq \mu \leq m + Z_{\alpha/2}ES}$$

$$\text{ES} = \boldsymbol{\sigma_m} = \frac{\sigma}{\sqrt{n}}$$

Table de la distribution normale standard

Scor z	0,06
1,9	0,4750

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981



$1\text{-}\alpha=0.95 \Leftrightarrow (1\text{-}\alpha)/2=0,475 \rightarrow Z_{\alpha/2}=1,96$

$0,475=\Pr(0<Z<1,96)$

Source: http://statcalculators.com/z-score-table-standard-normal-distribution/

ESTIMATION DE LA MOYENNE DE POPULATION
CAS DE GRANDS ÉCHANTILLONS (n≥30)

- L'intervalle de confiance de la moyenne μ avec le seuil α est:

100 (1-α)% IC: $$\left[m - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, m + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- **Lorsque l'ecart type σ est inconnu**, il peut être estimé à l'aide de la déviation standard d'echantillonage (**S**) ou de la déviation standard descriptive de l'échantillon (**s**)

100 (1-α)% IC: $$\left[m - Z_{\alpha/2}\frac{S}{\sqrt{n}}, m + Z_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

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

100 (1-α)% IC: $$\left[m - Z_{\alpha/2}\frac{s}{\sqrt{n-1}}, m + Z_{\alpha/2}\frac{s}{\sqrt{n-1}}\right]$$

ESTIMATION DE LA MOYENNE DE POPULATION
CAS DE GRANDS ÉCHANTILLONS (n≥30)

- Le plus souvent utilisé est le niveau de signification $\alpha = 0,05$.
- Alors $Z_{\alpha/2}$=1,96 et l'intervalle de confiance sera calculé en regardant SI

L'ecart type (σ) sur la population est connu?

OUI

95% IC: $$\left[m - 1,96\frac{\sigma}{\sqrt{n}}, m + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

NON

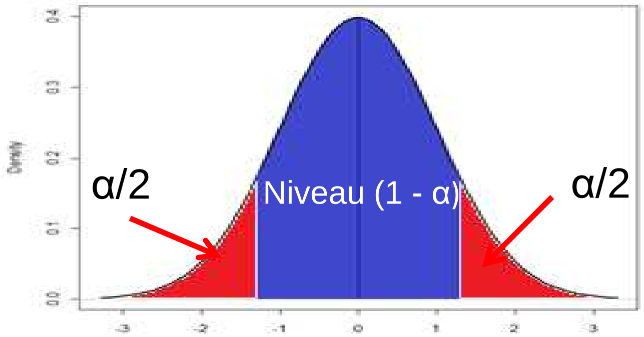
$$\sigma \approx S = \sqrt{\frac{n}{n-1}}\,s$$

95% IC: $$\left[m - 1,96\frac{s}{\sqrt{n-1}}, m + 1,96\frac{s}{\sqrt{n-1}}\right]$$

$$\left[m - 1,96\frac{S}{\sqrt{n}}, m + 1,96\frac{S}{\sqrt{n}}\right]$$

Facteurs qui influencent la largeur/précision d'un intervalle de confiance

- La largeur d'un intervalle de confiance (IC) est déterminée par **n**, **σ** et **α**:
- Si:
  - ÷ **Taille de l' échantillon « n »** augmente, la largeur diminue,
  - ÷ **déviation standard (σ):** variabilité d'observations individuelles - augmente, la largeur augmente (relation de directe proportionnalité),
  - ÷ **le niveau de confiance (1- α)** augmente (donc **α** diminue), la largeur augmente.
  - ÷ **Le risqué α** augmente, la largeur diminue



- La largeur de IC quantifie la **précision** de l'estimation :
  - Si l'IC est large ⇨ l'estimation est imprécise/faible,
  - Si l'IC est étroit ⇨ l'estimation est précise

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE GRANDS ÉCHANTILLONS (n≥30)

EXEMPLE

- Un échantillon aléatoire de taille n=49 a été tiré de la population de femmes de la Ro atteintes d'HTA. Le cholestérol moyen était égal à 220 mg / dl et l'écart type d'echantillonage S = 18 mg / dl.
- Quelle est l'estimation pour la moyenne du cholestérol chez les femmes roumaines a l'aide **l'intervalle de confiance de 95%**? Mais le **l'IC de 99%?**
- **Réponse:** m= 220 mg/dl -estimateur ponctuel du parametre inconnu de population (moyenne du cholesterol)

Longueur = 225-215 = 10

$$95\%\ IC: \left[220 - 1{,}96\frac{18}{\sqrt{49}}; 220 + 1{,}96\frac{18}{\sqrt{49}}\right] = [215; 225]$$

Longueur = 227-213 = 14

$$99\%\ IC: \left[220 - 2{,}58\frac{18}{\sqrt{49}}; 220 + 2{,}58\frac{18}{\sqrt{49}}\right] = [213; 227]$$

Représentativité de l'échantillon a l’aide de l’IC

- Un échantillon est représentatif d'une population si la valeur vraie du paramètre dans la population à partir de laquelle l'échantillon a été extrait est trouvée dans l'intervalle de confiance estimé.
- Si le niveau de confiance est de 0,05, il y a 95% de chances que le paramètre de population soit trouvé dans l'intervalle de confiance.

Exemple

- Sur une population de 10 000 nouveau-nés, nous connaissons le poids moyen à la naissance: $\mu = 3400$ g. Dans un échantillon aléatoire de 10 nouveau-nés:
- moyenne m = 3 200 g, l'écart type s = 300 g.
- **L'échantillon est-il représentatif ou non de la population?**
- Reponse: on calcule 95% IC: [3200-1,96*300/3; 3200+1,96*300/3]
  - 95% IC: [3004; 3396]
  - L'échantillonnage aléatoire implique le "droit" de chaque membre de la population d'être choisi dans l'échantillon mais ne garantit pas la représentativité proportionnelle de toutes les parties d'une population.
  - Un autre échantillon -> autre moyenne, un autre écart

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE PETITS ÉCHANTILLONS (n<30)

❑**Principe:**

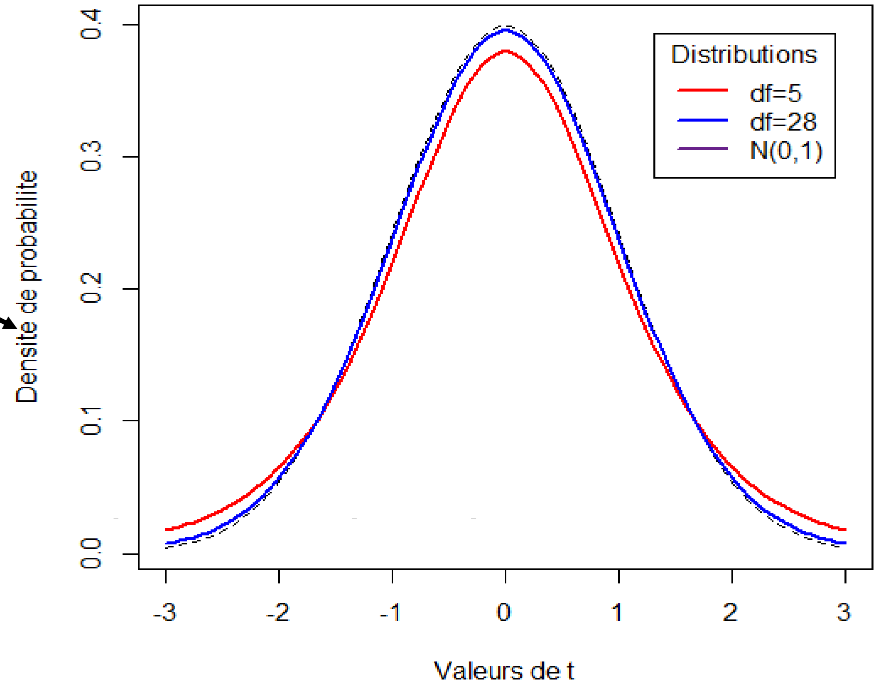
- population(P): moyenne μ, l'ecart type σ, **normalement distribués** par rapport à la variable d'etude.
- moyenne des échantillons (**m**) est une variable aleatoare
- variable : $Z = \frac{\mathbf{m}-\mu}{\frac{s}{\sqrt{n}}} \approx Student(t)$ avec df=n−1
- [a, b]=? tel que $\Pr(a \leq \mu \leq b) = 1-\alpha$

Student vs Normale: differences entre les deux distributions

La distribution de Student ressemble à la distribution normale mais les "queues" sont plus évasées à mesure que la taille de l'échantillon diminue

Distribution de Student:

- df=grade de liberté=n-1
- moyenne= 0; $\sigma^2 \neq 1$;
- symétrique autour de la moyenne



ESTIMATION DE LA MOYENNE DE POPULATION
CAS DE PETITS ÉCHANTILLONS (n<30)

100 (1-α)% IC: $$\left[m - t_{n-1,1-\frac{\alpha}{2}} ES ; m + t_{n-1,1-\frac{\alpha}{2}} ES \right]$$

100 (1-α)% IC: $$\left[m - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} ; m + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

100 (1-α)% IC: $$\left[m - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} ; m + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \right]$$

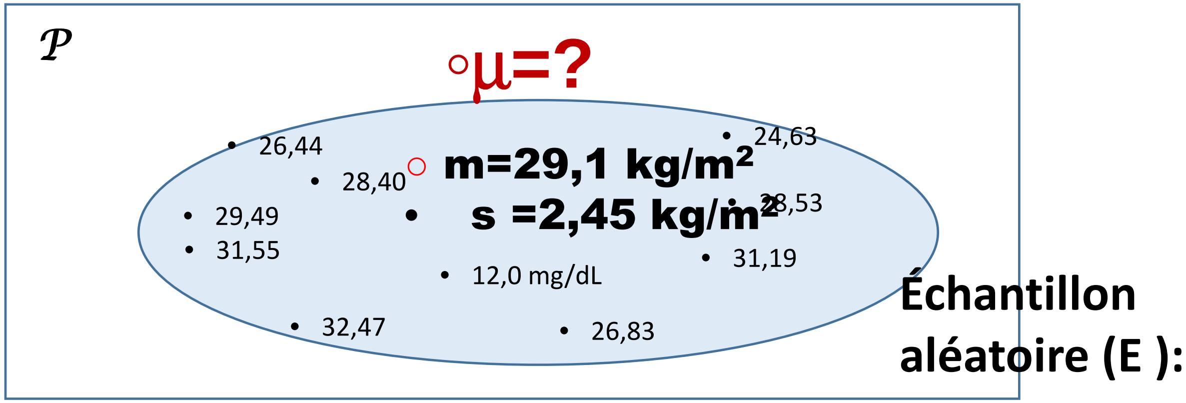
S (majuscule) =déviation standard d'echantillonage

s (minuscule) =déviation standard descriptive (de l'echantillon)

!!!! L'augmentation du *n* va déterminer que la valeur du t se rapproche de 1,96 ⇨ la courbe tend vers la distribution Normale.

Exemple

Population ($\mathcal{P}$): **moyenne μ, variance σ^2 de la variable X.**



Soit un échantillon de 10 patients, les résultats ont été obtenus:.

28,40; 26,44; 29,49; 31,55; 31,40; 32,47; 24,63; 28,53; 31,19; 26,83

Quelle est l'estimation de l'IMC moyen de la population des diabétiques en utilisant l'intervalle de confiance de 95%?

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE PETITS ÉCHANTILLONS (n<30)

- **Quelle est l'estimation de l'IMC moyen de la population des diabetiques en utilisant un intervalle de confiance de 95%?**
- Réponse: **Estimateur ponctuel sera m=**(28,4+26,44+29,49+31,55+31,4+32,47+24,63+28,53+31,19+26,83)/10=29,1
- Variance de l'IMC sur l' échantillon: s^2 =[(28,4-29,1)^2 +....+(26,83-29,1)^2]/10=6,02
- L' écart type de l' échantillon: $s = \sqrt{s^2}$= 2,45
- **95% IC:**

$$\left[m - t_{n-1,\alpha/2} \frac{s}{\sqrt{n-1}} ; m + t_{n-1,\alpha/2} \frac{s}{\sqrt{n-1}} \right]$$

- **95% IC:** $[29{,}1 - 2{,}26 \cdot 0{,}82; 29{,}1 + 2{,}26 \cdot 0{,}82]$
- **95% IC: [27,25; 30,94]**

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE PETITS ÉCHANTILLONS (n<30)

- On considéré un échantillon aléatoire de nouveau-nés choisi d'une maternité à Cluj pour lesquels nous avons mesuré le poids (grammes):
- 2990, 2500, 3100, 2000, 2100, 3200, 3350, 3400, 3250, 2900.
- **Quelle est l'estimation du poids moyen de la population de nouveau-nés en utilisant un intervalle de confiance de 95%?**
- Réponse: **Estimateur ponctuel m**=(2990+2500+…+2900)/10=2879 g
- Variance de poids sur l'echantillon:
- $s^2 = [(2990-2879)^2 + \dots + (2900-2879)^2]/10 = 257632$
- L' écart type de l' échantillon: $\mathbf{s = \sqrt{s^2} = \sqrt{257632} = 507.57}$
- **95% IC:** $$\left[m - t_{n-1,\alpha/2} \frac{s}{\sqrt{n-1}} ; m + t_{n-1,\alpha/2} \frac{s}{\sqrt{n-1}} \right]$$

$$\left[2879 - t_{9,0.975} \frac{507.57}{\sqrt{9}} ; 2879 + t_{9,0.975} \frac{507.57}{\sqrt{9}} \right] = \left[2879 - 2.262 \frac{507.57}{\sqrt{9}} ; 2879 + 2.262 \frac{507.57}{\sqrt{9}} \right]$$

95% IC : [2516 ;3242]

ESTIMATION DE LA FREQUENCE DE POPULATION CAS DE GRANDS ÉCHANTILLONS (np≥10 OU n(1-p)≥10)

100 (1-α)% IC:

$$[f - Z_{\alpha/2}\sqrt{\frac{f\cdot(1-f)}{n}}\,;\, f + Z_{\alpha/2}\sqrt{\frac{f\cdot(1-f)}{n}}]$$

95% IC:

$$[f - 1{,}96\sqrt{\frac{f\cdot(1-f)}{n}}\,;\, f + 1{,}96\sqrt{\frac{f\cdot(1-f)}{n}}]$$

Ou f = la fréquence relative calculée sur l'échantillon.

ESTIMATION DE LA FREQUENCE DE POPULATION
CAS DE GRANDS ÉCHANTILLONS (np≥10 OU n(1-p)≥10)

EXEMPLE

Sur un échantillon aléatoire de 100 diabétiques, il a été observé que 35% étaient en surpoids.

Quelle est l'estimation de la prévalence du surpoids dans la population en utilisant l'intervalle de confiance de 95%?

Réponse: on connait que f=0,35

Condition d' applicabilité: np=100*0,35=35≥10

95% IC: $$[f - 1{,}96\sqrt{\frac{f \times (1-f)}{n}}\,;\, f + 1{,}96\sqrt{\frac{f \times (1-f)}{n}}]$$

95% IC: [0,26;0,44]

Estimation d'un paramètre (moyenne, fréquence) de la population

Intervalle de confiance pour

- Moyenne $\boldsymbol{\mu}$
  - σ connue ou n≥30
    - **100 (1−α)% IC:** $[m - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, m - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
    - **100 (1−α)% IC:**$[m - Z_{\alpha/2}\frac{S}{\sqrt{n}}, m - Z_{\alpha/2}\frac{S}{\sqrt{n}}$
    - **100 (1−α)% IC:**$[m - Z_{\alpha/2}\frac{s}{\sqrt{n-1}}, m - Z_{\alpha/2}\frac{s}{\sqrt{n-1}}$
  - σ inconnue sau n<30
    - **100 (1−α)% IC:** $[m - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}, m + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}]$
    - **100 (1−α)% IC:** $[m - t_{n-1;\alpha/2}\frac{s}{\sqrt{n-1}}, m + t_{n-1;\alpha/2}\frac{s}{\sqrt{n-1}}]$
- Fréquence **p**
  - SI n·f ≥10 et n·(1-f) ≥10
  - **100 (1−α)% IC:** $[f - Z_{\alpha/2}\sqrt{\frac{f\cdot(1-f)}{n}}, f - Z_{\alpha/2}\sqrt{\frac{f\cdot(1-f)}{n}}]$

DÉTERMINER LA TAILLE À L'AIDE D'INTERVALLE DE CONFIANCE (IC)

- Si, pour un risque (niveau du signification) donné α, on souhaite obtenir une précision donnée (c'est-à-dire une certaine marge d'erreur) de l'intervalle de confiance, c'est-à-dire de la forme:[f - ME , f + ME]
- en tenant compte de la formule de l'intervalle de confiance:

$$\left[f - Z_{\alpha/2} \cdot \sqrt{\frac{f(1-f)}{n}},\ f + Z_{\alpha/2} \cdot \sqrt{\frac{f(1-f)}{n}} \right]$$

$$\text{ME} = Z_{\alpha} \cdot \sqrt{\frac{f(1-f)}{n}} \Rightarrow n = \frac{Z_{\alpha}^{2} \cdot f(1-f)}{ME^{2}}$$

EXEMPLE

- On veut d'évaluer la fréquence de l'échec à l'examen de Biostat (prévalence) dans une certaine population d' étudiants.
- Dans les années précédentes, on sait qu'il est d'environ 30%. Si une précision de l'estimation de ME = 4% est souhaitée pour un risque α = 0,05 (auquel Z = 1,96 correspond), cette estimation doit être effectuée à l'aide d'un échantillon dont la taille doit être supérieure ou égale à

$$n=\frac{(1.96)^2 \times 0.3 \times 0.7}{(0.04)^2}$$

- On considerer que 1.96 $\cong$ 2 donc n = 525.

DIFFERENCE de moyennes
DIFFERENCE de fréquences

Intervalle de confiance pour:

Difference de moyennes
$d = \mu_1 - \mu_2$

échantillons indépendants

100 (1-α)% IC: $[d - t_{n-1,\frac{\alpha}{2}} S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}; d + t_{n-1,\frac{\alpha}{2}} S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$

Ou $d = m_1 - m_2$; $S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

échantillons dépendants

100 (1−α)% IC: $[m_d - Z_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}; m_d + Z_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}]$

Ou m_d = moyenne des différences
S_d = l' ecart type des différences

Difference de frequences
$d = p_1 - p_2$

100 (1-α)% IC: $[d - Z_{\alpha/2}\, ES; d + Z_{\alpha/2}\, ES]$

$$ou\ ES = \sqrt{\frac{f_1\times(1-f_1)}{n_1} + \frac{f_2\times(1-f_2)}{n_1}}$$

L’intervalle de confiance pour la Différence

- **Interpretation**:

❑ Si la valeur zéro est contenue dans l'intervalle de confiance, la différence entre les moyennes n'est pas significativement différente de zéro.

❑ Si la valeur zéro n’est pas dans l'intervalle de confiance la différence entre les moyennes est significativement différente de zéro.

Difference de moyennes
Echantillons indépendants

Genre=M	7	7	8	8	8	6	9	6	5
Genre=F	8	10	9	6	10	8	9	7	8

	Genre (M)	Genre(F)
Moyenne	7,11	8,33
Variance	1,27	1,32
Deviation standard (ecart type)	1,61	1,75

df=16

$\alpha = 0{,}05 \longrightarrow t_{18,0,0025} = 2{,}10$

95% IC: [(7,11-8,33)−2,10*1,14; (7,11-8,33)+ 2,10*1,14]=[-3,41; 1,37]

On estime une différence entre les moyennes des notes au Biostat chez les étudiants de genre masculin et féminin égale à -1,22 et nous sommes sûrs à 95% que l'intervalle [-3,41; 1.37] contiendra la différence entre les moyennes des notes. Comme IC contient 0-> il n'y a pas de différence significative entre les moyens des notes chez les étudiants de genre M et F

Difference: frequences

- Fréquence des maladies cardiovasculaires chez les fumeurs et les non-fumeurs

	MCV-	MCV+	Total	Proportions (freq.relatives)
Fumeurs	60	90	150	$f_1 = 90/150 = 0{,}6$
Nonfumeurs	170	30	200	$f_2 = 30/200 = 0{,}15$
Total	230	120	350	

- ES=$\sqrt{0{,}6*(1-0{,}6))/150+(0{,}15*(1-0{,}15))/200}$= 0,05
- Intervalle de confiance de 95% pour la difference p_1-p_2:
- 95% IC: [(0,6-0,15)-1,96*0,05; (0,6-0,15)+1,96*0,05] = [0,45-1,96*0,05; 0,45+1,96*0,05] = [0,35 ;0,55]

On estime une différence de fréquence de MCV égale à 0,45 et nous sommes sûrs à 95% que [0,35; 0,55] contient la différence de fréquences pour les 2 populations. De plus, l'IC à 95% ne contient pas 0 →la différence significative de fréquence de MCV entre les deux populations (fumeurs et non-fumeurs).

Synthèse de cours

Échantillon

proportion <u>observée</u> – f ou
moyenne <u>observée</u> – m
ou
différence des 2
proportions
ou
différence des 2 moyennes

Sont des **estimateurs ponctuels** du paramètre dans la population

Population

proportion <u>réelle</u> p
ou
moyenne <u>réelle</u> μ
ou
différence des 2 proportions
ou
différence des 2 moyennes
Et **intervalle de confiance 95%**
(limite inferieure – limite supérieure)

Limite inferieure = **m** ou **f** – **ES** * **Z** ou **t** ...
Limite supérieure = **m** ou **f** + **ES** * **Z** ou **t** ...

(Z ou t sont pris depuis des tableaux des lois des distributions
ES = erreur standard a différentes formules en fonction de la situation)

Synthèse de cours

L' interprétation d'un estimateur ponctuel

d'un proportion, ou une moyenne, ou une différence des 2 proportions, ou une différence des 2 moyennes

Notre meilleure estimation de la → proportion ou moyenne ou différence des 2 proportions ou différence des 2 moyennes → dans la population cible est …

Exemple:

Le pourcentage des fumeurs dans un échantillon est 25%

=> Notre meilleure estimation de la vrai valeur du pourcentage des fumeurs dans la population cible est 25%

La moyenne de la glycémie dans un échantillon est 87,5 mg/dL

⇒ Notre meilleure estimation de la vrai valeur de la moyenne de la glycémie dans la population cible est 87,5 mg/dL

Mais toujours est important de montrer aussi l'intervalle de confiance pas seulement l'estimateur ponctuel pour les résultats importants de l' étude

Synthèse

L' interprétation d'un intervalle de confiance: d'un proportion, ou une moyenne, ou une différence des 2 proportions, ou une différence des 2 moyennes (95% IC: [a, b])

La vraie valeur du

proportion
ou
moyenne
ou
différence des 2 proportions
ou
différence des 2 moyennes

dans la population cible est entre a et b (les limites de l'intervalle de confiance) avec une probabilité de 95%

Exemple:
Intervalle de confiance du pourcentage des fumeurs: 95% IC: [20%; 30%]
=> Donc, l'intervalle compris entre 20% et 30% contiendra à 95% du temps (avec une probabilité égale a 0.95) la vrai fréquence relative des fumeurs dans la population cible.
Intervalle de confiance de la moyenne de la glycémie (95% IC: 82 mg/dL– 97 mg/dL)
=> l'intervalle compris entre 82 et 97 mg/dL contiendra à 95% du temps (avec une probabilité égale a 0.95) la vrai valeur de la moyenne de la glycémie dans la population cible.

Synthèse de cours

Type variable	Nombre echantillons	Estimateur ponctuel	Conditions application	Formule
qualitative	une	fréquence: f	grands échantillons: nf ≥ 10 et nq ≥ 10	$[f - Z_{\alpha/2}\,ES; f + Z_{\alpha/2}\,ES]$ $ES = \sqrt{\frac{f(1-f)}{n}}$
	une	fréquence: f	petits échantillons: nf < 10 ou nq <10	on ne va pas calculée
	deux	différence entre les fréquences: $f_1 - f_2$	grands échantillons: $f_1n_1 \geq 10$, $(1-f_1)n_1 \geq 10$, $f_2n_2 \geq 10$, $(1-f_2)n_2 \geq 10$	$ES = \sqrt{\frac{f_1 \times (1-f_1)}{n_1} + \frac{f_2 \times (1-f_2)}{n_2}}$ $[(f_1 - f_2) - Z_{\alpha/2}\,ES; (f_1 - f_2) + Z_{\alpha/2}\,ES]$
	deux	différence entre les fréquences:	petits échantillons: np < 10 ou nq <10	on va pas calculée
quantitative	un	moyenne: m	grands échantillons: n ≥ 30, σ - connue	$[m - Z_{\alpha/2}\,ES; m + Z_{\alpha/2}\,ES]$ $ES = \frac{\sigma}{\sqrt{n}}$
	un	moyenne: m	grands échantillons: n ≥ 30, σ – non connue	$[m - t_{n-1,\frac{\alpha}{2}}ES; m + t_{n-1,\frac{\alpha}{2}}ES]$ $ES = \frac{S}{\sqrt{n}}$
	un	moyenne: m	petits échantillons: n < 30, σ – non connue	$[m - t_{n-1,\frac{\alpha}{2}}ES; m + t_{n-1,\frac{\alpha}{2}}ES]$ $ES = \frac{S}{\sqrt{n}}$
	deux	différence entre les moyennes: $m_1 - m_2$	variances égales	$ES = \sqrt{\frac{(n_1-1) \times s_1^2 + (n_2-1) \times s_2^2}{n_1 + n_2 - 2}}$ $[(m_1 - m_2) - t_{n-1,\frac{\alpha}{2}}ES; (m_1 - m_2) + t_{n-1,\frac{\alpha}{2}}ES]$

(ou n, n_1, n_2 - nombre des sujets; f, f_1, f_2 - fréquence observée; q = 1 – f; m, m_1, m_2 – moyennes; s, s_1, s_2 – déviations standard descriptive de l'échantillon; S – déviation standard d' échantillonage; $S = \sqrt{\frac{n}{n-1}}s$; $s = \sqrt{\frac{n-1}{n}}S$; σ - déviation standard dans la population; ES=erreur standard; pour α=0,05, $Z_{\alpha/2}$=1,96)

Exemples d’articles scientifiques avec intervalles de confiance (IC)

Clin Orthop Relat Res. 2014 Sep 9. [Epub ahead of print]

Can Therapy Dogs Improve Pain and Satisfaction After Total Joint Arthroplasty? A Randomized Controlled Trial.

Harper CM[1], Dong Y, Thornhill TS, Wright J, Ready J, Brick GW, Dyer G.

IC pour variables **quantitatives**:
Score du douleur après la prothèse totale d’articulation avec <u>traitement avec chiens</u> 1,7 [95% IC: 1.4-2.0] , <u>et sans</u> traitement 4,1 [95% IC: 3.8-4.4]
IC pour variables **qualitatives**:
Satisfaction des patients avec <u>traitement avec chiens</u> 92% [95% IC: 78%-98%] , <u>et sans</u> traitement 69% [95% IC: 52%-84%]

Abstract

BACKGROUND:

The use of animals to augment traditional medical therapies was reported as early as the 9
this study was to evaluate the role of animal-assisted therapy using therapy dogs in the po

QUESTIONS/PURPOSES:

We asked: (1) Do therapy dogs have an effect on patients' perception of pain after total jo
their hospital stay after total joint arthroplasty as measured by the Hospital Consumer Ass

METHODS:

A randomized controlled trial of 72 patients undergoing primary unilateral THA or TKA was conducted. Patients were randomized to a 15-minute visitation with a therapy dog before physical therapy or standard postoperative physical therapy regimens. Both groups had similar demographic characteristics. Reduction in pain was assessed using the VAS after each physical therapy session, beginning on postoperative Day 1 and continuing for three consecutive sessions. To ascertain patient satisfaction, the proportion of patients selecting top-category ratings in each subsection of the HCAHPS was compared.

RESULTS:

Patients in the treatment group had lower VAS scores after each physical therapy session with a final VAS score difference of 2.4 units (animal-assisted therapy VAS, 1.7; SD, 0.97 **[95% CI, 1.4-2.0]** versus control VAS, **4.1**; SD, 0.97 **[95% CI, 3.8-4.4**], $p < 0.001$) after the third physical therapy session. Patients in the treatment group had a higher proportion of top-box HCAHPS scores in the following fields: nursing communication (33 of 36**, 92% [95% CI, 78%-98%]** versus 69%, 25 of **36 [95% CI, 52%-84%],** p = 0.035; risk ratio, 1.3 [95% CI of risk ratio, 1.0-1.7]; risk difference, 23% [95% CI of risk difference, 5%-40%]), pain management (34 of 36, 94% [95% CI, 81%-99%], versus 26 of 36, 72% [95% CI, 55%-86%], p = 0.024; risk ratio, 1.3 [95% CI of risk ratio, 1.1-1.6]; risk difference, 18% [95% CI of risk difference, 5%-39%]). The overall hospital rating also was greater in the treatment group (0-10 scale) (9.6; SD, 0.7 [95% CI, 9.3-9.8] versus 8.6, SD, 0.9 [95% CI, 8.3-8.9], $p < 0.001$).

CONCLUSIONS:

The use of therapy dogs has a positive effect on patients' pain level and satisfaction with hospital stay after total joint replacement. Surgeons are encouraged to inquire about the status of volunteer-based animal-assisted therapy programs in their hospital as this may provide a means to improve the immediate postoperative recovery for a select group of patients having total joint arthroplasty.

Exemples d'articles scientifiques avec intervalles de confiance (IC)

Clin Orthop Relat Res. 2014 Sep 9. [Epub ahead of print]

Can Therapy Dogs Improve Pain and Satisfaction After Total Joint Arthroplasty? A Randomized Controlled Trial.

Harper CM[1], Dong Y, Thornhill TS, Wright J, Ready J, Brick GW, Dyer G.

- **Exemple:** IC pour variables **quantitatifs**:
- **Extrait d'article**
- Score du douleur après la prothèse totale d'articulation avec traitement avec chiens 1,7 [95% CI, 1.4-2.0] , et sans traitement 4,1 [95% CI, 3.8-4.4]
- **Exemple:** IC pour variables **qualitatifs**:
- **Extrait d'article**
- Satisfaction des patients avec traitement avec chiens 92% [95% CI, 78%-98%] , et sans traitement 69% [95% CI, 52%-84%]
- **Interprétation:**
- L'estimateur ponctuel est 1,7, l'intervalle de confiance est de 1,4 a 2,0.
- Dans l' étude ils ont trouve une moyenne du score du douleur après la prothèse totale d'articulation avec traitement avec chiens de 1,7. Leurs meilleure estimation de la vrai valeur de la moyenne du score du douleur après la prothèse totale d'articulation avec traitement avec chiens dans la population cible est 1,7. La vrais valeur du score du douleur après la prothèse totale d'articulation avec traitement avec chiens dans la population cible est entre 1,4 et 2,0 avec une probabilité de 95%
- **Interprétation:**
- L'estimateur ponctuel est 92%, l'intervalle de confiance est de 78% a 98%.
- Dans l' étude ils ont trouve une pourcentage de satisfaction avec traitement avec chiens de 92%. Leurs meilleure estimation de la vrai valeur du pourcentage de satisfaction avec traitement avec chiens dans la population est 92%. La vrais valeur du pourcentage de satisfaction avec traitement avec chiens dans la population est entre 78% et 98% avec une confiance de 95%.

Exemples d'articles scientifiques avec intervalles de confiance (IC)

J Antimicrob Chemother. 2014 Sep 10. pii: dku352. [Epub ahead of print]

Randomized non-inferiority trial to compare trimethoprim/sulfamethoxazole plus r

Harbarth S[1], von Dach E[2], Pagani L[3], Macedo-Vinas M[4], Huttner B[5], Olearo F[6], Emonet

IC pour **différence entre deux échantillons** - variables **qualitatives**:
Différence entre guérison de l'infection avec traitement avec linezolid et traitement avec trimethoprim/sulfamethoxazole et rifampicin 6.3%, 95% IC: [-6.8%, 19.2%].

Abstract

OBJECTIVES:

The therapeutic arsenal for MRSA infections is limited. The aim of this study was to as ... olid alone for the treatment of MRSA infection.

METHODS:

We conducted a randomized, open-label, single-centre, non-inferiority trial comparing trimethoprim/sulfamethoxazole (160 mg/800 mg three times daily) plus rifampicin (600 mg once a day) versus linezolid (600 mg twice a day) alone in adult patients with various types of MRSA infection. Patients were allocated 1: 1 to either regimen. The primary outcome was clinical cure at 6 weeks after the end of treatment (non-inferiority margin 20%) assessed by both ITT and PP analyses. Secondary outcomes included the microbiologically documented persistence of MRSA in clinical cultures, mortality and adverse events. The study protocol has been registered with ClinicalTrials.gov (NCT00711854).

RESULTS:

Overall, 150 patients were randomized to one of the two treatment arms between January 2009 and December 2013 and were included in the ITT analysis. Of these 56/75 (74.7%) in the linezolid group and 59/75 (78.7%) in the trimethoprim/sulfamethoxazole and rifampicin group experienced clinical success (risk difference 4%, 95% CI -9.7% to 17.6%). The results were confirmed by the PP analysis, with 54/66 (81.8%) cured patients in the linezolid group versus 52/59 (88.1%) in the trimethoprim/sulfamethoxazole and rifampicin group (**risk difference 6.3%, 95% CI -6.8% to 19.2%).** There were no statistically significant differences between the two groups in any of the secondary outcomes, including microbiologically documented failure. Four adverse drug reactions attributed to the study medication occurred in the linezolid group versus nine in the trimethoprim/sulfamethoxazole and rifampicin group.

CONCLUSIONS:

Compared with linezolid, trimethoprim/sulfamethoxazole and rifampicin seems to be non-inferior in the treatment of MRSA infection.

KEYWORDS: Switzerland; adults; drug therapy; humans; multidrug-resistant organisms; prospective clinical studies; staphylococcal infections

Exemples d'articles scientifiques avec intervalles de confiance (IC)

J Antimicrob Chemother. 2014 Sep 10. pii: dku352. [Epub ahead of print]

Randomized non-inferiority trial to compare trimethoprim/sulfamethoxazole plus rifampicin versus linezolid for the treatment of MRSA infection.

Harbarth S[1], von Dach E[2], Pagani L[3], Macedo-Vinas M[4], Huttner B[5], Olearo F[6], Emonet S[7], Uçkay I[6].

- **Exemple:** IC pour **différence entre deux échantillons** - variables **qualitatives**:
- Extrait d'article
- Différence entre guérison de l'infection avec traitement avec linezolid et traitement avec trimethoprim/sulfamethoxazole et rifampicin 6.3%, 95% CI : [6.8%, 19.2%].
- Interprétation:
- L'estimateur ponctuel est 6,3%, l'intervalle de confiance est de -6,8% a 19,2%.

Dans l' étude ils ont trouve une différence des pourcentages de guérison de l'infection entre ceux qui ont été recu le traitement avec linezolid et ceux qui ont recu le traitement avec trimethoprim/sulfamethoxazole et rifampicin de 6,3%. Leurs meilleure estimation de la vrai différence des pourcentages de guérison de l'infection entre ceux qui ont été recu le traitement avec linezolid et ceux qui ont recu le traitement avec trimethoprim/sulfamethoxazole et rifampicin dans la population cible est de 6,3%. La vraie valeur de la différence des pourcentages de guérison de l'infection entre ceux qui ont été recu le traitement avec linezolid et ceux qui ont recu le traitement avec trimethoprim/sulfamethoxazole et rifampicin dans la population cible est entre -6,8% et 19,2% avec une probabilité de 0.95.

Exemples d’articles scientifiques avec intervalles de confiance (IC)

- **Periodontal health of palatally displaced canines treated with open or closed surgical technique: A multicenter, randomized controlled trial**
- Nicola A. Parkin, Richard S. Milner, Chris Deery, David Tinsley, Anne-Marie Smith, Peter Germain, Jennifer V. Freeman, Sarah J. Bell, Philip E. Benson
- Received: December 2012; Received in revised form: March 2013;
- DOI: http://dx.doi.org/10.1016/j.ajodo.2013.03.016
- **Introduction**
- The aim of this study was to investigate differences in the periodo open or a closed surgical technique.

IC pour la différence des moyennes entre deux échantillons – pour variables **quantitatives**:
La différence des moyennes entre ceux qui ont été soumis a l’intervention chirurgicale des canins et ceux sans intervention a été 0.5 mm; (96% CI, 0.4-0.7)

- **Methods:** A multicenter, randomized controlled trial was undertaken in 3 hospitals in the United Kingdom, involving 2 parallel groups. Patients with unilateral PDC were randomly allocated to receive either an open or a closed surgical exposure. Periodontal health was assessed 3 months after removal of fixed appliances. Parameters measured included clinical attachment levels, recession, alveolar bone levels, and clinical crown height.
- **Results:** Data from 62 participants (closed, 29; open, 33) were analyzed. There was no difference between PDC exposed with an open vs a closed surgical technique (**mean difference, 0.1 mm; 95% confidence interval [CI], −0.2-0.5**). There was, however, a statistical difference in mean attachment loss between the operated and unoperated (contralateral) canines (**mean difference, 0.5 mm; 96% CI, 0.4-0.7;** *P* <0.001). Twenty of the 62 subjects had some recession on the palatal aspect of the operated canine, whereas only 4 subjects had some visible root surface on the palatal aspect on the unoperated side (*P* = 0.001).
- **Conclusions:** There is a periodontal impact when a unilateral PDC is exposed and aligned. This impact is small and unlikely to have clinical relevance in the short term; however, the long-term significance is unknown. When the open and closed techniques were compared, no difference in periodontal health was found.

Exemples d’articles scientifiques avec intervalles de confiance (IC)

- **Periodontal health of palatally displaced canines treated with open or closed surgical technique: A multicenter, randomized controlled trial**
- Nicola A. Parkin, Richard S. Milner, Chris Deery, David Tinsley, Anne-Marie Smith, Peter Germain, Jennifer V. Freeman, Sarah J. Bell, Philip E. Benson
- Received: December 2012; Received in revised form: March 2013; Accepted: March 2013;
- DOI: http://dx.doi.org/10.1016/j.ajodo.2013.03.016

Exemple: IC pour la **différence des moyennes entre deux échantillons** – pour variables **quantitatifs**:

Extrait d’article

- La différence des moyennes entre ceux qui ont été soumis a l’intervention chirurgicale des canins et ceux sans intervention a été 0.5 mm; (96% CI, 0.4-0.7)

Interprétation:

L’estimateur ponctuel est 0,5 mm, l’intervalle de confiance est de 0,4 a 0,7.

Dans l’ étude ils ont trouve une différence des moyennes entre ceux qui ont été soumis a l’intervention chirurgicale des canins et ceux sans intervention de 0,5 mm. Leurs meilleure estimation de la vrai différence entre ceux qui ont été soumis a l’intervention chirurgicale des canins et ceux sans intervention dans la population cible est 0,5 mm. La vrais valeur de la différence entre ceux qui ont été soumis a l’intervention chirurgicale des canins et ceux sans intervention dans la population cible est entre 0,4 et 0,7 avec une confiance de 95% (ou un risque de 5%).

Exemples des questions

Nombre de réponses exactes par question: 1-4.

E1. Lesquelles des réponses suivantes sont correctes :

A. La moyenne empirique (moyenne de l' échantillon) est un estimateur sans biais de la moyenne d'une loi.

B. La variance empirique (variance descriptive de l' échantillon) est un estimateur sans biais de la variance d'une loi.

C. On peut toujours trouver un intervalle de confiance qui contient le paramètre avec une probabilité égale à 1.

D. Un intervalle de confiance pour une moyenne d'une loi normale avec l'écart type est connu, a une longueur fixe pour un niveau donné et une taille d'échantillon n donnée

E. La variance empirique (variance descriptive de l' échantillon) est un estimateur biaisé de la variance d'une loi.

R1. A, D, E

Exemples des questions pour l’examen

E2. * On veut calculer l'intervalle de confiance pour la moyenne de la circonférence abdominale dans une population, si on a les données sur un échantillon de 26 sujets. La valeur de la quantile d’ intérêt est 2,06. La moyenne de la circonférence abdominale est égale à 63 cm et l'écart type descriptif est égal à 7 cm.

Quelle est l'intervalle de confiance au niveau 95 % pour la circonférence abdominale moyenne de la population?

A. [60.12; 65.88]

B. [60.26; 65.74]

C. [42.41; 83.59]

D. [60.06; 65.94]

E. [60.17; 65.83]

R1.A

Solution de l’exercice -> voir la diapo suivante

ESTIMATION DE LA MOYENNE DE POPULATION
CAS DE PETITS ÉCHANTILLONS (n<30)

95% IC:

$$\left[m - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} ; m + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right]$$

s (minuscule) =déviation standard de l' échantillon.

95% IC:

$$\left[63 - t_{25,0975} \frac{7}{\sqrt{26-1}} ; 63 + t_{25,0975} \frac{7}{\sqrt{26-1}}\right]$$

95% IC:

$$\left[63 - 2{,}06 \frac{7}{\sqrt{26-1}} ; 63 + 2{,}06 \frac{7}{\sqrt{26-1}}\right]$$

95% IC:

$$[63{,}12; 65{,}88]$$

Exemples des questions pour l’examen

E3. * On veut calculer l'intervalle de confiance pour la moyenne de la température maximale, des enfants avec grippe, si on a les données sur un échantillon de 11 sujets. La valeur de la quantile d’ intérêt est 2,23. La moyenne de la température est égale à 39 dégrées et l'écart type d’ échantillonnage est égal à 0.8.

Quelle est l'intervalle de confiance à 95 % pour la température moyenne de la population ?

A. [38.41; 39.59]

B. [38.46; 39.54]

C. [38.5; 39.5]

D. [38.53; 39.47]

E. [38.44; 39.56]

R3. B; *Solution de l’exercice* -> voir la diapo suivante

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE PETITS ÉCHANTILLONS (n<30)

95% IC: $\left[m - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} ; m + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$

s (minuscule) =déviation standard de l' échantillon.

95% IC: $\left[39 - t_{25,0975} \frac{0,8}{\sqrt{11}} ; 39 + t_{25,0975} \frac{0,8}{\sqrt{11}}\right]$

95% IC: $\left[39 - 2,23 \frac{0,8}{\sqrt{11}} ; 39 + 2,23 \frac{0,8}{\sqrt{11}}\right]$

95% IC: $[38,46; 39,54]$

Exemples des questions pour l'examen

E4. * On veut calculer l'intervalle de confiance pour la perte de poids moyenne de la population en connaissant les données mesurées sur un échantillon de 176 étudiantes de l'UMF Cluj qui ont suit un régime alimentaire au cours d'une année.

On a trouvé que la perte de poids moyenne était égale a 6.2 kg et la deviation standard pour la population est de 3 kg.

Quelle est l'intervalle de confiance au risque 5 % pour la perte de poids moyenne de la population?

A. [5.752; 6.644]

B. [5.752; 6.648]

C. [5.753; 6.646]

D. [5.757; 6.643]

E. [5.756; 6.644]

R4.D

Exemples des questions pour l’examen

E5. * On veut calculer l'intervalle de confiance pour l’indice de mass corporelle (IMC) moyen de la population d’ étudiants de l’UMF Cluj en connaissant les données mesurées sur un échantillon de 250 étudiants .

On a trouvé que la moyenne de l’IMC était égale a 26 kg/m^2 et la deviation standard de la population est de 3.5 kg/m^2 . La valeur de la quantile d’ intérêt est 1,96.

Quelle est l'intervalle de confiance au risque 5 % pour l’ IMC moyen pour la population d’ intérêt?

A. [25.565; 26.435]

B. [25.579; 26.434]

C. [25.566; 26.434]

D. [25.707; 26.293]

E. [25.564; 26.434]

R5.C; *Solution de l’exercice* -> voir la diapo suivante

ESTIMATION DE LA MOYENNE DE POPULATION CAS DE GRANDS ÉCHANTILLONS (n≥30)

- Le plus souvent utilise est le niveau de signification $\alpha = 0.05$.
- Alors $Z_{\alpha/2}$=1.96 et l'intervalle de confiance:

L'ecart type (σ) sur la population est connu?

OUI

95% IC: $\left[m - 1{,}96\frac{\sigma}{\sqrt{n}}, m + 1{,}96\frac{\sigma}{\sqrt{n}}\right]$

95% IC: $\left[26 - 1{,}96\frac{3{,}5}{\sqrt{250}}, 26 + 1{,}96\frac{3{,}5}{\sqrt{250}}\right]$

$[25{,}57; 26{,}43]$

Exemples des questions pour l’examen

E6. On veut comparer la moyenne de la pression artérielle systolique (PAS mmHg) chez les sujets hypertendus qui ont utilise un médicament A et la moyenne de PAS chez les sujets hypertendus qui ont utilise un médicament B. 50 sujets qui ont reçu le médicament A ont eu en moyenne la PAS égale a 156 mmHg et déviation standard descriptive de l'échantillon de 50 mmHg tandis que les 14 sujets qui ont reçu le médicament B ont eu en moyenne la PAS de 140 mmHg, déviation standard de l’ échantillon de 14 mmHg. La valeur de la quantile est 1,999.

Quelle est l'intervalle de confiance à 95 % pour la différence de moyenne pour toute la population des hypertendus?

A. [-73.78; 105.78]

B. [-72.18; 107.88]

C. [73.78; 105.78]

D. [23.98; 150.78]

E. [74.78; 205.78]

R6. A

Exemples des questions pour l'examen

E7. * On veut calculer l'intervalle de confiance de la fréquence d'avoir grippe dans une population, si on a les données sur un échantillon de 100 sujets.

Dans l'échantillon il y a 10 sujets qui ont eu la grippe. La valeur de la quantile est 1,96.

Quelle est l'intervalle de confiance à 95 % pour la prévalence de la grippe ?

A. [0.04; 0.16]

B. [-0.01; 0.21]

C. [-0.01; 0.21]

D. [0; 0.2]

E. [0.07; 0.13]

R7. A

Solution de l'exercice -> voir la diapo suivante

ESTIMATION DE LA FREQUENCE DE POPULATION CAS DE GRANDS ÉCHANTILLONS (np≥10 OU n(1-p)≥10)

95% IC: $$[f - Z_{\alpha/2}\sqrt{\frac{f \times (1-f)}{n}}; f + Z_{\alpha/2}\sqrt{\frac{f \times (1-f)}{n}}]$$

95% IC: $$[f - 1{,}96\sqrt{\frac{f \times (1-f)}{n}}; f + 1{,}96\sqrt{\frac{f \times (1-f)}{n}}]$$

- f est la fréquence relative observée sur l'échantillon=10/100=0,10.

95% IC: $$[0.10 - 1.96\sqrt{\frac{0.10 \times 0.90}{100}}; 0.10 + 1.96\sqrt{\frac{0.10 \times 0.90}{100}}]$$

95% IC: $[0.04; 0.16]$

Exemples des questions pour l'examen

E8. * On veut comparer la fréquence d'avoir grippe chez les garçons et la fréquence de la grippe chez les filles.

On a observé 41 malades parmi 320 garçons et 25 malades parmi 300 fille. La valeur de la quantile est 1,96.

Quelle est l'intervalle de confiance à 95 % pour la différence de fréquence de la grippe ?

A. [-0.003; 0.093]

B. [-0.10; 0.211]

C. [-0.001; 0.021]

D. [0; 0.093]

E. [0.007; 0.13]

R8. A

Exemples des questions pour l'examen

E9. Le nombre de consultations en urgence au cabinet dentaire dans un intervalle de temps donné X={0,1,2,...,n,...} est une:

A. variable aléatoire discrete

B. variable aléatoire continue

R9: A

> The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.
>
> — *John Tukey, Exploratory Data Analysis*

Les Graphiques (avec des simulations numériques) du cours ont été réalisés par Mihaela Iancu avec:
R software envinronment for statistical computing and graphics