

Master Santé -S1- Biostatistique

Travaux pratiques (TP2)

Martin LARSEN

Distribution de X et $\bar{X}$

X est une variable aléatoire continuée (VAC)

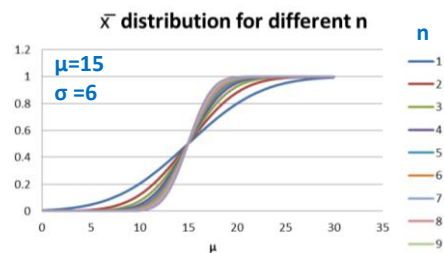
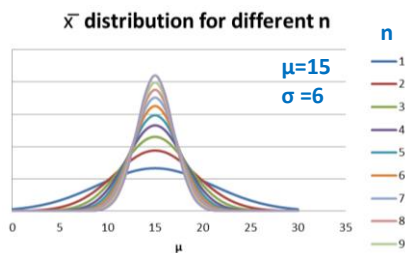
➤ Distribution de Echantonnage de Moyenne:

$$X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$$

Pour $n=1$, nous avons $X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$

Fonction de densité de probabilité

Fonction de répartition (Proba accumuler)



Distribution de X et $\bar{X}$

LOI.NORMALE.N(x;μ;σ/RACINE(N);cumulative)

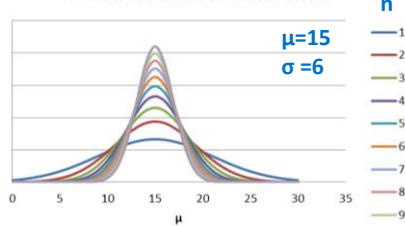
	A	B	C	D	E	F	G	H
mu		15						
sigma		6						
n		1	2	3	4	5	6	7
X	1	2	3	4	5	6	7	
	0	0.002921383	0.000182	9.77E-06	4.96E-07	2.43E-08	1.17E-09	5.56E-11
	1	0.004370315	0.000406	3.27E-05	2.48E-06	1.82E-07	1.31E-08	9.32E-10
	2	0.006358771	0.00086	0.000101	1.11E-05	1.19E-06	1.25E-07	1.29E-08
	3	0.008998494	0.001722	0.000285	4.46E-05	6.75E-06	1E-06	1.46E-07
	4	0.012385194	0.003263	0.000744	0.00016	3.33E-05	6.8E-06	1.37E-06
	5	0.016579523	0.005847	0.001785	0.000514	0.000143	3.91E-05	1.05E-05
	6	0.021586266	0.009911	0.003941	0.001477	0.000536	0.000191	6.69E-05
	7	0.027335012	0.015893	0.008002	0.003799	0.001746	0.000786	0.000349

Pour $n=1$, nous avons $X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$

Fonction de densité de probabilité

Cumulative = FAUX

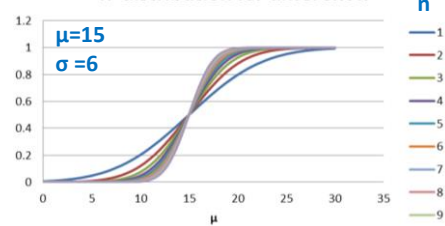
$\bar{x}$ distribution for different n



Fonction de répartition (Proba accumuler)

Cumulative = VRAI

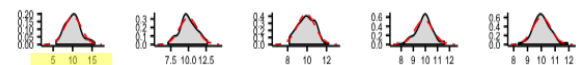
$\bar{x}$ distribution for different n



Représentation graphique - distribution normale theorique

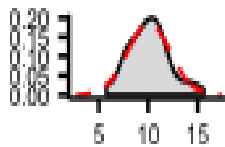
$\bar{X}$ distribution: 100 échantillons a taille n éléments chaque

N=1



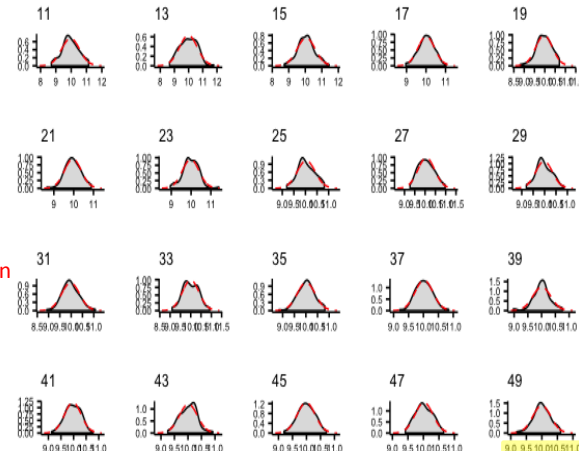
$X \sim \mathcal{N}(\mu=10; \sigma=2)$

N=100



Observed

Theoretical normal distribution



$X \sim \mathcal{N}(\mu=10; \sigma=2)$

Distribution d'échantillonnage
de la Moyenne

$\bar{X}_{bar} \sim \mathcal{N}(\mu; \sigma/\sqrt{n})$

Distribution de X et $\bar{X}$

X est une variable aléatoire continuée (VAC)

➤ Distribution de Echantonnage de Moyenne:

$$X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$$

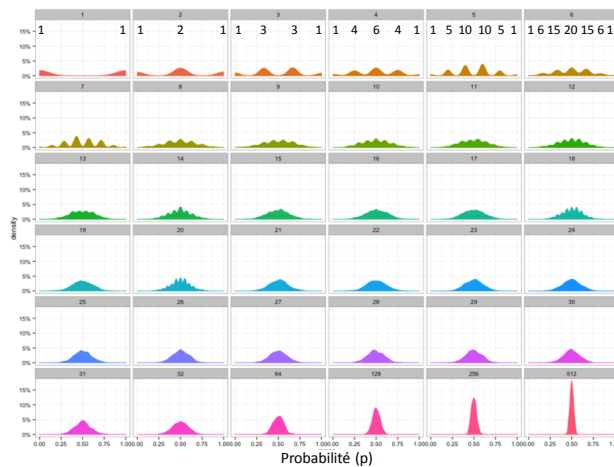
➤ Distribution binomiale – convergence distribution normale

Distribution binomiale -> convergence distribution normale

E.g. distribution binomial : $X \sim B(n, \pi) \Rightarrow \left[\begin{matrix} n\pi > 5 \\ n(1-\pi) > 5 \end{matrix} \right] \Rightarrow N(\mu = n\pi, \sigma^2 = n\pi(1-\pi))$
 $\pi = 0.5$ and $n \in [1-512]$ $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k}$

Fréquence de succès au cours de n épreuves de Bernoulli

Le moyenne de 1000 échantillons était analyse



k	$\frac{n!}{k! (n-k)!}$
3	1
2	3
1	3
0	1

k	$\frac{n!}{k! (n-k)!}$
4	1
3	4
2	6
1	4
0	1

Distribution de X et $\bar{X}$

X est une variable aléatoire continuée (VAC)

➤ Distribution de Echantonnage de Moyenne:

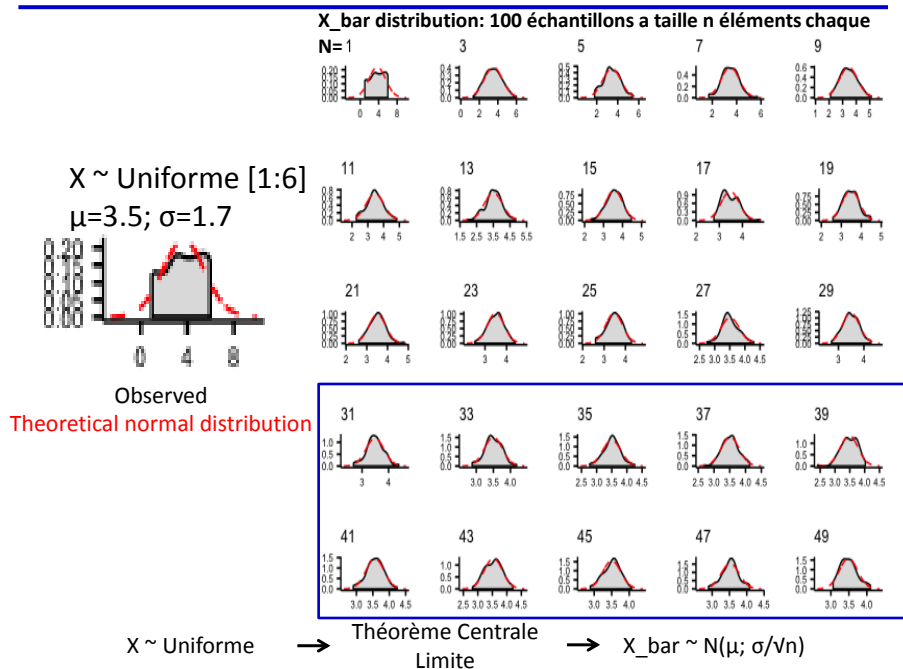
$$X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$$

➤ Distribution binomiale – convergence distribution normale

➤ Théorème centrale limite (TCL):

X suit n'importe quelle distribution. Pour $n > 30 \Rightarrow \text{TCL} \Rightarrow \bar{X} \sim \mathcal{N}(\mu_x, \frac{\sigma_x}{\sqrt{n}})$

Représentation graphique - distribution uniforme theorique



Distribution de X et $\bar{X}$

X est une variable aléatoire continuée (VAC)

➤ Distribution de Echantonnage de Moyenne:

$$X \sim \mathcal{N}(\mu_x, \sigma_x) \Rightarrow \bar{X} \sim \mathcal{N}\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$$

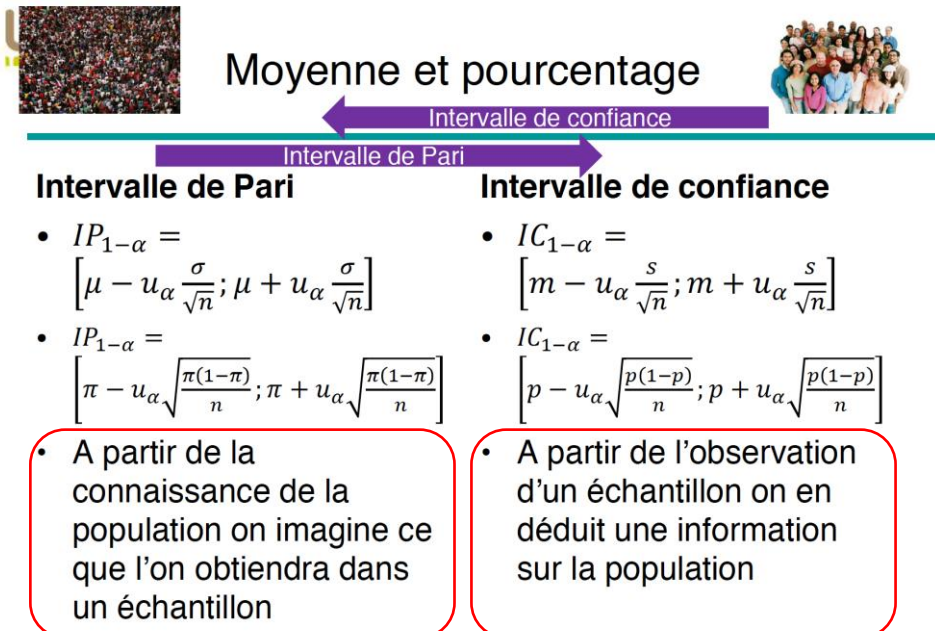
➤ Distribution binomiale – convergence distribution normale

➤ Théorème centrale limite (TCL):

X suit n'importe quelle distribution. Pour $n > 30 \Rightarrow \text{TCL} \Rightarrow \bar{X} \sim \mathcal{N}\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$

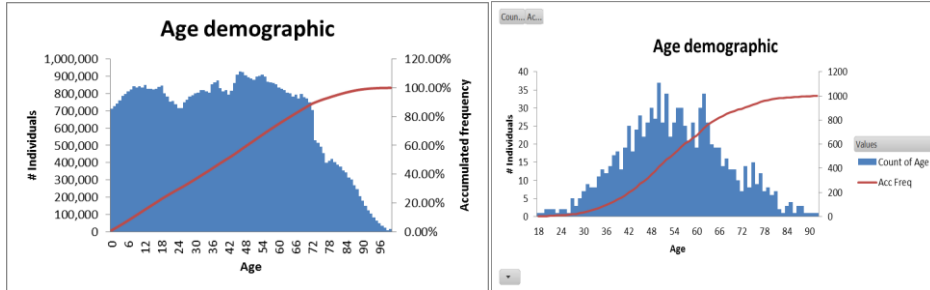
➤ Autre

En l'absence de distribution connue pour X et $n < 30$, nous ne pouvons pas identifier une distribution pour $\bar{X} \sim ??$



Distribution d'Age (statistique descriptive)

Démographie de la population française (Tableau 1) et une échantillon inconnu (Tableau 2):



Descriptive statistics of population:

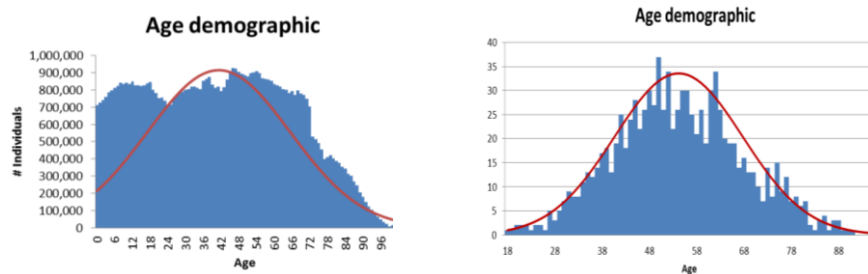
n 66,992,699
 μ 41.2 years
 σ^2 586.4 years²
 σ 24.2 years

Descriptive statistics of sample

n 1000 =NB(A2:A1001)
Mean 54.2 years =moyenne(A2:A1001)
 s^2 184.3 years² =Var.S(A2:A1001)
s 13.6 years =ECARTYPE(A2:A1001)

Distribution d'Age (statistique descriptive)

Démographie de la population française (Tableau 1) et une échantillon inconnu (Tableau 2):



Descriptive statistics of population:

n 66,992,699
 μ 41.2 years
 σ^2 586.4 years²
 σ 24.2 years

Descriptive statistics of sample

n 1000
Mean 54.2 years
 s^2 184.3 years²
s 13.6 years

=IFERROR(VLOOKUP(G2,\$M\$3:\$N\$77,2,FALSE),0)						
G	H	I	J	K	L	M
	norm	Count Of A	Rel Count of Age			N
18	0.000842	1	0.001		Row Labels	Count of Age
19	0.001022	1	0.001		18	1
20	0.001233	2	0.002		19	1
21	0.00148	2	0.002		20	2
22	0.001768	2	0.002		21	2
23	0.002099	1	0.001		22	2
24	0.00248	2	0.002		23 (Age)	1
25	0.002913	2	0.002		24 (Row: 22)	2



Moyenne et pourcentage



Intervalle de confiance

Intervalle de Pari

Intervalle de Pari

- $IP_{1-\alpha} = \left[\mu - u_{\alpha} \frac{\sigma}{\sqrt{n}}; \mu + u_{\alpha} \frac{\sigma}{\sqrt{n}} \right]$
- $IP_{1-\alpha} = \left[\pi - u_{\alpha} \sqrt{\frac{\pi(1-\pi)}{n}}; \pi + u_{\alpha} \sqrt{\frac{\pi(1-\pi)}{n}} \right]$
- A partir de la connaissance de la population on imagine ce que l'on obtiendra dans un échantillon

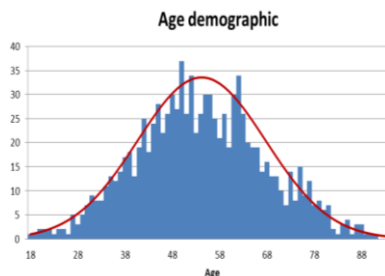
Intervalle de confiance

- $IC_{1-\alpha} = \left[m - u_{\alpha} \frac{s}{\sqrt{n}}; m + u_{\alpha} \frac{s}{\sqrt{n}} \right]$
- $IC_{1-\alpha} = \left[p - u_{\alpha} \sqrt{\frac{p(1-p)}{n}}; p + u_{\alpha} \sqrt{\frac{p(1-p)}{n}} \right]$
- A partir de l'observation d'un échantillon on en déduit une information sur la population

Intervalle de pari vs. Intervalle de confiance

Démographie de la population française (Tableau 1) et une échantillon inconnu (Tableau 2):

Statistique descriptive (échantillon)
 n 1000
 Mean 54.2 years
 s² 184.3 years²
 s 13.6 years
 $IC_{1-\alpha} : \mu \in [\bar{x} - u_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + u_{\alpha/2} \frac{s}{\sqrt{n}}]$
 = [53.3; 55.0] avec risque α
 Valider précision de s²:
 précise (n>30) (loi normale)
 Peu- précise (n<30) (loi de Student, programme)



Statistique descriptive (population):
 n 66,992,699
 μ 41.2 years
 σ^2 586.4 years²
 σ 24.2 years
 X ne suis pas $\mathcal{N}(\mu, \sigma)$ (Visuellement)
 Mais Théorème Central Limite (TCL, n>30):
 $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$
 $IP_{1-\alpha} : \bar{x} \in [\mu - u_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu + u_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$
 = [39.7; 42.7] avec risque $\alpha=5\%$ (n=1000)

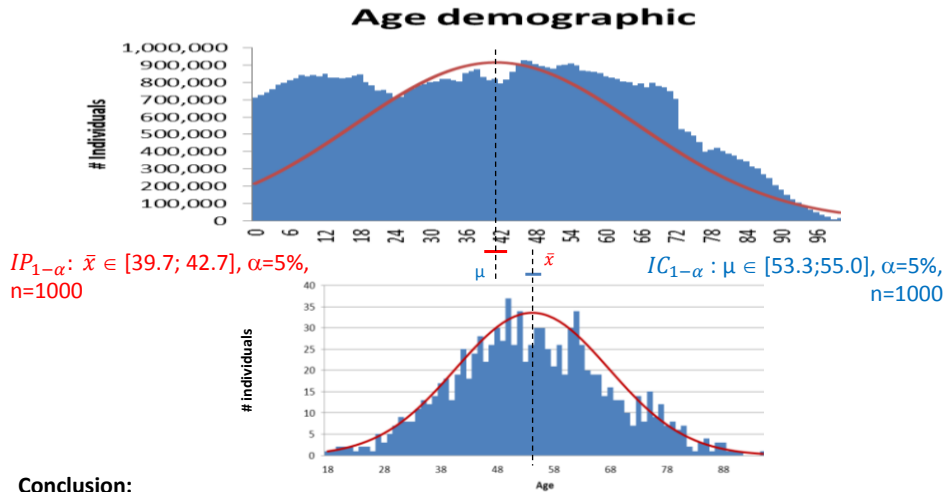
Statistique descriptive (échantillon)
 n 1000
 Mean 54.2 years
 s² 184.3 years²
 s 13.6 years
 $IC_{1-\alpha} : \mu \in [\bar{x} - u_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + u_{\alpha/2} \frac{s}{\sqrt{n}}]$
 = [53.3; 55.0] avec risque $\alpha=5\%$
 Valider précision de s²:
 précise (n>30) (loi normale)
 Peu- précise (n<30) (loi de Student, hors programme)

N → ∞

Intervalle rétréci

Intervalle de pari vs. Intervalle de confiance

Démographie de la population française (Tableau 1) et une échantillon inconnu (Tableau 2):

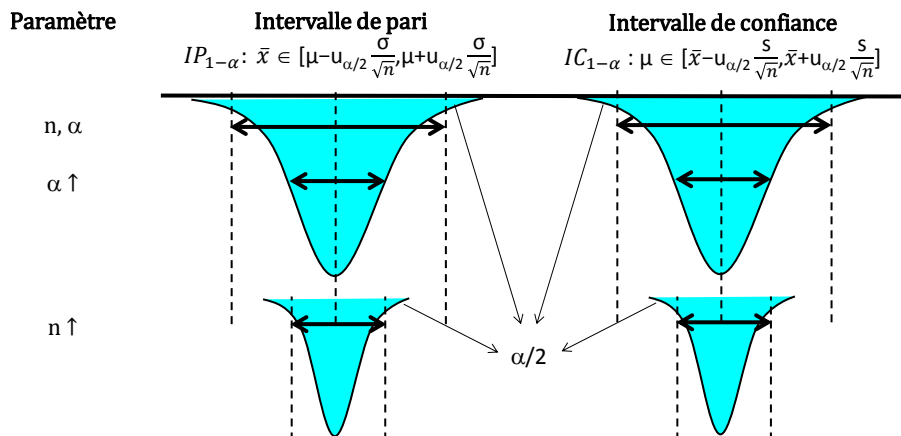


Conclusion:

- $\mu \notin IC_{1-\alpha}$: **Inférence:** l'échantillon a été sélectionné au hasard dans une autre population avec $\mu \in [53.3; 55.0]$ au risque 5%
- $\bar{x} \notin IP_{1-\alpha}$: **Prédiction:** l'échantillon n'a pas été sélectionné au hasard dans la population avec $\mu = 41.2$ au risque 5%

Intervalle de pari vs. Intervalle de confiance

Démographie de la population française (Tableau 1) et une échantillon inconnu (Tableau 2):



Conclusion

- Statistique descriptive
 - Fournissez les informations nécessaires pour calculer les intervalles statistiques pour la prédiction ou l'inférence.
- Prédiction:
 - Intervalle de Pari (IP)
 - A partir de la connaissance de la population on prédit ce que l'on obtiendra dans un échantillon.
- Inférence:
 - Intervalle de Confiance
 - A partir de l'observation d'un échantillon on en déduit une information sur la population.