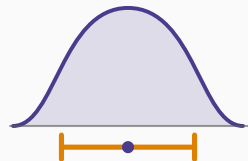


Statistique inférentielle

Chapitre 3 : La distribution d'échantillonnage de la moyenne



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

L'objet central du cours

Les deux premiers moments

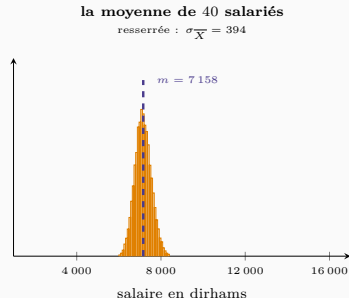
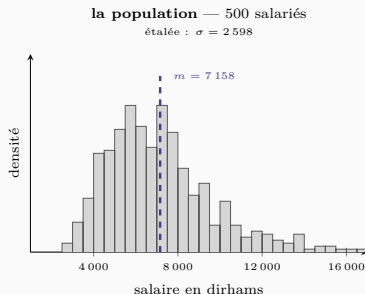
La forme de la distribution

Un premier usage

Ce qu'il faut retenir

L'objet central du cours

Deux graphiques qui n'ont pas le même objet



Les deux graphiques n'ont pas le même objet. À gauche, des *salariés*. À droite, des *moyennes d'échantillons*.

Trois tirages de 40 salariés ont donné 6863, 7001 et 7895 dirhams : jamais la même valeur, toujours autour de 7158. En répétant vingt mille fois, on obtient la cloche de droite — c'est la **distribution d'échantillonnage** de $\bar{X}$, et c'est l'objet central de tout le cours.

Elle est 6.6 fois moins étalée que la population, exactement $\sqrt{40}$ à la correction de population finie près.

Distribution d'échantillonnage

La **distribution d'échantillonnage** d'une statistique est la loi de probabilité des valeurs qu'elle prend lorsqu'on répète le tirage d'échantillons de même taille dans la même population.

Le changement d'objet, à saisir une fois pour toutes

La population des salaires décrit des **salariés**. La distribution d'échantillonnage décrit des **moyennes d'échantillons**.

Ce ne sont pas les mêmes individus, pas les mêmes unités de raisonnement, et surtout pas le même étalement. Confondre les deux est l'erreur qui rend le reste du cours incompréhensible.

Les deux premiers moments

Le résultat, démontré au chapitre 20 des probabilités

Pour un échantillon aléatoire de taille n tiré dans une population d'espérance m et d'écart-type σ :

$$E(\bar{X}) = m \qquad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Erreur-type

La quantité $\sigma/\sqrt{n}$ porte un nom propre : l'**erreur-type** de la moyenne.

Ce n'est pas l'écart-type des données. C'est l'écart-type des **moyennes** — combien la moyenne d'échantillon se trompe, typiquement.

Vingt mille échantillons de taille 40

Grandeur	Théorie	Simulation
Espérance de $\bar{X}$	7 158,26	7 159,81
Erreur-type $\sigma/\sqrt{n}$	410,80	392,61

L'écart sur l'erreur-type n'est pas une erreur

392,61 au lieu de 410,80 : la simulation est trop précise, et systématiquement.

La raison est simple : nous tirons **sans remise** dans une population de seulement 500 individus. Un échantillon de 40 en représente déjà 8 % : il ne peut pas se tromper autant qu'un tirage dans une population infinie.

Le facteur d'exhaustivité

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

Ici : $410,80 \times \sqrt{460/499} = 394,42$ — la valeur simulée, à un dixième près.

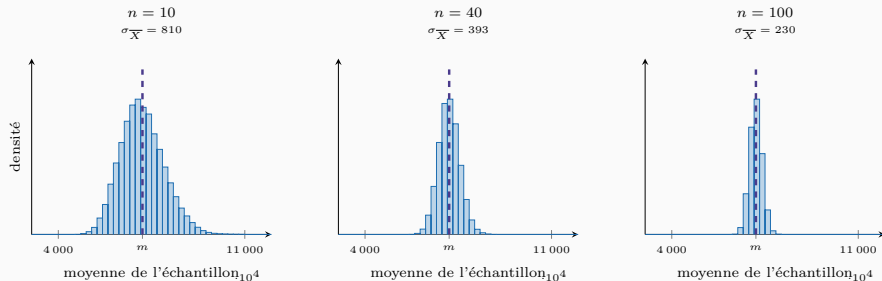
Quand faut-il l'appliquer

La règle usuelle : dès que n dépasse 5 % de N .

Elle joue toujours dans le bon sens — elle **réduit** l'erreur-type — et elle s'annule quand $n = N$: sur un recensement complet, la moyenne d'échantillon est la moyenne, et son erreur est nulle. La formule le dit toute seule.

La forme de la distribution

L'effet de la taille de l'échantillon



Deux choses se produisent en même temps quand n grandit, et il faut les distinguer.

1. **La forme devient normale** — c'est le théorème central limite. La population des salaires est nettement étirée à droite ; dès $n = 40$, la cloche est là.
2. **L'étalement se resserre en $\sigma/\sqrt{n}$** — c'est la loi des grands nombres. De 10 à 100, l'écart-type est divisé par $\sqrt{10} \approx 3,2$.

Le premier point autorise à utiliser la table normale ; le second dit combien on gagne à interroger plus de monde.

Ce qui se produit quand n grandit

1. **La forme devient normale** — c'est le théorème central limite.
2. **L'étalement se resserre en $1/\sqrt{n}$** — c'est la loi des grands nombres.

Ils ne disent pas la même chose

Le premier autorise à **lire une table normale**, donc à calculer des probabilités.

Le second dit **de combien** on gagne à interroger davantage de monde. Le premier rend l'inférence possible; le second en fixe le prix.

Le théorème, sous sa forme utile

Les deux cas

Population normale : $\bar{X}$ suit exactement une loi normale, quel que soit n .

Population quelconque, n grand : $\bar{X}$ suit approximativement $\mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)$.

$$Z = \frac{\bar{X} - m}{\sigma/\sqrt{n}} \approx \mathcal{N}(0; 1)$$

La règle des trente, et sa vraie portée

$n \geq 30$ est une convention commode, pas un théorème. Elle suffit largement pour une population modérément dissymétrique — comme nos salaires.

Pour une distribution très étirée — patrimoines, sinistres, durées de chômage — il faut bien davantage.
Le bon réflexe est de regarder l'histogramme avant d'invoquer la règle.

Un premier usage

Que peut donner un échantillon de 50?

Avec $m = 7\,158,26$ et $\sigma = 2\,598,10$, l'erreur-type vaut $2\,598,10/\sqrt{50} = 367,44$.

$$P(7\,158,26 - 1,96 \times 367,44 \leq \bar{X} \leq 7\,158,26 + 1,96 \times 367,44) = 0,95$$

$$\bar{X} \in [6\,438,08; 7\,878,44] \text{ dans } 95\% \text{ des cas}$$

Attention au sens de lecture

Cet encadrement porte sur $\bar{X}$, **connaissant** m . C'est encore du calcul des probabilités : on descend de la population vers l'échantillon.

Notre moyenne observée, $7\,005,80$, tombe bien dans cet intervalle. Le chapitre 10 retournera l'inégalité pour encadrer m à partir de $\bar{x}$ — et ce sera enfin de l'inférence.

L'ordre de grandeur qu'il faut avoir en tête

Quelques erreurs-types, sur nos salaires

n	$\sigma/\sqrt{n}$	Écart typique
10	821,6	énorme
50	367,4	acceptable
200	183,7	confortable
500	116,2	tout le monde est interrogé

Ce que le gestionnaire doit en conclure

Sur dix bulletins de paie, la moyenne peut se tromper de huit cents dirhams : elle ne permet aucune décision.

Sur deux cents, elle se trompe d'environ deux cents dirhams : on peut négocier une grille salariale. **La taille d'échantillon n'est pas une contrainte technique, c'est une variable de décision** — et le chapitre 12 la calculera.

Ce qu'il faut retenir

Six points

1. La **distribution d'échantillonnage** est la loi des valeurs prises par une statistique d'un échantillon à l'autre.
2. Elle décrit des **moyennes**, non des individus : ce ne sont pas les mêmes objets.
3. $E(\bar{X}) = m$ et $\sigma(\bar{X}) = \sigma/\sqrt{n}$: cette dernière s'appelle l'**erreur-type**.
4. Au-delà de 5 % de N , la **correction de population finie** réduit l'erreur-type.
5. Le **théorème central limite** rend $\bar{X}$ approximativement normale, quelle que soit la population de départ.
6. La règle $n \geq 30$ est une convention, non un théorème.

La suite

Une moyenne n'est pas la seule statistique utile. Un taux de rebut, un taux d'équipement, une dispersion de qualité : ce sont des **proportions** et des **variances**.

Le chapitre 4 établit leurs distributions d'échantillonnage — et rencontre au passage l'énigme du $n - 1$.