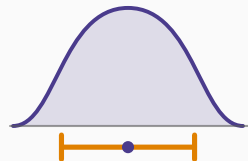


Statistique inférentielle

Chapitre 11 : Proportion et variance



Jaouad Madkour

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

L'intervalle pour une proportion

L'intervalle pour une variance

Ce qu'il faut retenir

L'intervalle pour une proportion

Intervalle de confiance d'une proportion

$$f \pm z_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}$$

valable si $nf \geq 5$ et $n(1-f) \geq 5$.

Une entorse assumée

L'écart-type théorique fait intervenir p , que l'on cherche justement à estimer. On le remplace donc par f .

C'est une approximation, non un théorème. Elle est acceptable dès que n dépasse quelques dizaines, et c'est la formule que tout le monde emploie.

La part de diplômés Bac+5

Sur les 50 salariés de l'échantillon, 13 sont titulaires d'un Bac+5.

$$f = \frac{13}{50} = 0,26 \quad \sqrt{\frac{0,26 \times 0,74}{50}} = 0,0620$$

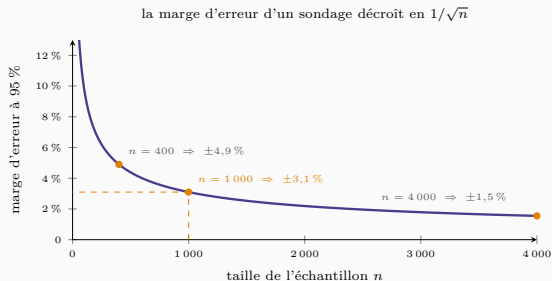
$$IC_{95\%} = 0,26 \pm 1,96 \times 0,0620 = [0,138 ; 0,382]$$

La vérification, et la leçon

La proportion réelle vaut 0,214 : elle est bien dans l'intervalle.

Mais celui-ci s'étend de 14 % à 38 % — presque du simple au triple. **Cinquante observations ne permettent pas d'estimer une proportion avec finesse**, et c'est un ordre de grandeur qu'il faut avoir en tête avant de commander une enquête.

La marge d'erreur, telle qu'on l'annonce



D'où vient le « plus ou moins trois points »

$$1,96 \sqrt{\frac{0,5 \times 0,5}{1000}} = 0,031$$

On prend $p = 0,5$ parce que c'est le **cas le plus défavorable** : $p(1-p)$ y est maximal. La marge annoncée est donc toujours une majoration.

La courbe s'aplatit. Passer de 1 000 à 4 000 personnes quadruple le coût et ne gagne que 1,5 point.

C'est pourquoi tous les instituts, dans tous les pays, s'arrêtent autour du millier — ce n'est pas une convention, c'est un calcul.

Pourquoi les proportions exigent de gros échantillons

Comparaison des deux paramètres

Paramètre	Estimation	Demi-largeur	En part de la valeur
Moyenne	7 005,80	± 640	9 %
Proportion	0,26	$\pm 0,122$	47 %

L'explication

Une variable indicatrice ne prend que deux valeurs : elle est extrêmement pauvre en information. Chaque observation n'apporte qu'un bit.

C'est pourquoi les sondages politiques interrogent mille personnes là où une enquête de salaires en interroge cent : **estimer une proportion coûte structurellement plus cher.**

L'intervalle pour une variance

La loi utilisée

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

L'intervalle s'obtient en encadrant cette quantité par les deux quantiles du khi-deux, puis en isolant σ^2 .

La formule

$$\left[\frac{(n-1)s^2}{\chi_{1-\alpha/2; n-1}^2} ; \frac{(n-1)s^2}{\chi_{\alpha/2; n-1}^2} \right]$$

Trois différences à repérer

1. L'intervalle **n'est pas symétrique** autour de s^2 : la loi du khi-deux ne l'est pas.
2. Il n'y a **pas de $\pm$** : les deux bornes se calculent séparément.
3. Les quantiles sont **inversés** : le grand quantile va au dénominateur de la borne inférieure.

Sur nos données

$n = 50$, $s^2 = 5\,066\,894$, $\nu = 49$.

$$\chi_{0,975;49}^2 = 70,222 \quad \chi_{0,025;49}^2 = 31,555$$

$$IC_{95\%}(\sigma^2) = \left[\frac{49 \times 5\,066\,894}{70,222} ; \frac{49 \times 5\,066\,894}{31,555} \right] = [3\,535\,592 ; 7\,868\,118]$$

On prend la racine des deux bornes

$$IC_{95\%}(\sigma) = [1\,880,3 ; 2\,805,0]$$

La valeur réelle, $\sigma = 2\,598,10$, est bien dans l'intervalle.

Prendre la racine est ici légitime

La racine est une fonction croissante : elle conserve l'ordre des bornes, donc le niveau de confiance.

Attention toutefois : cet intervalle est correct pour σ , mais son centre n'est pas s — la transformation déplace le milieu.

Une condition beaucoup plus stricte

La loi χ^2 pour s^2 exige que la **population soit normale**. Contrairement à l'intervalle de la moyenne, aucun théorème central limite ne vient sauver le cas général.

Sur une population à queues épaisses — rendements financiers, montants de sinistres — cet intervalle peut être gravement faux, même avec un grand n .

Le contraste à retenir

$\bar{X}$ est robuste à la non-normalité s^2 ne l'est pas

C'est un point que beaucoup d'utilisateurs ignorent, et c'est l'une des raisons pour lesquelles l'estimation de la volatilité est un sujet difficile en finance.

Ce qu'il faut retenir

Six points

1. **Proportion** : $f \pm z_{\alpha/2} \sqrt{f(1-f)/n}$, valable si $nf \geq 5$ et $n(1-f) \geq 5$.
2. Sur 50 observations : $[0,138 ; 0,382]$ — très large. Une proportion exige un gros échantillon.
3. **Variance** : intervalle **asymétrique**, construit sur le khi-deux, sans $\pm$.
4. Les quantiles y sont **croisés** : le grand au dénominateur de la borne basse.
5. Pour σ , on prend la racine des deux bornes : $[1\,880,3 ; 2\,805,0]$.
6. L'intervalle de la variance exige la **normalité de la population** ; celui de la moyenne, non.

La suite

Tous ces intervalles se sont révélés larges. La question qui vient naturellement est celle du gestionnaire : **combien d'observations faut-il pour atteindre la précision voulue ?**

Le chapitre 12 y répond par une formule, et par un calcul de budget.