Statistiques - Fluctuation - Estimation

Prise de décision

▷ **Exercice 1.** Le responsable de la maintenance des machines à sous d'un casino doit vérifier qu'un certain type de machine est bien réglé sur une fréquence de succès de 0,06. Il décide de régler chaque machine pour laquelle il aura observé, dans l'historique des jeux, une fréquence de succès se situant en dehors d'un intervalle de fluctuation au seuil de 95%. Lors du contrôle d'une machine, le technicien constate qu'elle a fourni 9 succès sur 85 jeux.

a) Déterminer la fréquence observée f de succès de cette machine. On a $f = \dfrac{9}{85}$

b) Déterminer un intervalle de fluctuation asymptotique au seuil de 95% de la fréquence du succès pour un échantillon de taille 85.

Les conditions $\begin{cases} n = 85 \geqslant 30 \\ p = 0,06 \\ n \times p = 5,1 \geqslant 5 \\ n \times (1-p) = 79,9 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{85} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,06 - 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{85}}\ ;\ 0,06 + 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{85}}\right] \simeq [0,0095\ ;\ 0,1105]$$

c) Le technicien va-t-il modifier le réglage de la machine ?

$f = \dfrac{9}{85} \simeq 0,1059 \in \mathrm{I}_{85}$ donc le technicien ne peut pas rejeter l'hypothèse selon laquelle la fréquence de succès est bien égale à $0,06$.

d) Quelle aurait été sa décision s'il y avait eu 21 succès sur 200 jeux ?

Les conditions $\begin{cases} n = 200 \geqslant 30 \\ p = 0,06 \\ n \times p = 12 \geqslant 5 \\ n \times (1-p) = 188 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{200} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,06 - 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{200}}\ ;\ 0,06 + 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{200}}\right] \simeq [0,0271\ ;\ 0,0930]$$

$f = \dfrac{21}{200} \simeq 0,105 \notin \mathrm{I}_{200}$ donc le technicien rejettera l'hypothèse selon laquelle la fréquence de succès est bien égale à $0,06$, avec une marge d'erreur de 5%.

▷ **Exercice 2.** Les enfants sont dits prématurés lorsque la durée gestationnelle est inférieure ou égale à 259 jours. La proportion de ces naissances est de 6%. Des chercheurs suggèrent que les femmes ayant eu un travail pénible pendant leur grossesse sont plus susceptibles d'avoir un enfant prématuré que les autres. Il est décidé de réaliser une enquête auprès d'un échantillon aléatoire de 400 naissances correspondant à des femmes ayant eu pendant leur grossesse un travail pénible. Le nombre d'enfants prématurés est de 50. Quelle est donc la conclusion ?

$f = \dfrac{50}{400} = 0,125$

Les conditions $\begin{cases} n = 400 \geqslant 30 \\ p = 0,06 \\ n \times p = 24 \geqslant 5 \\ n \times (1-p) = 376 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{400} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,06 - 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{400}}\ ;\ 0,06 + 1,96\frac{\sqrt{0,06 \times 0,94}}{\sqrt{400}}\right] \simeq [0,0367\ ;\ 0,0833]$$

$f = 0,125 \notin \mathrm{I}_{400}$ donc on peut rejeter l'hypothèse selon laquelle la proportion est de 6% dans l'ensemble des femmes qui ont eu un travail pénible, avec un risque d'erreur de 5%.

▷ **Exercice 3.** Dans le monde, la proportion de gauchers est 12%. Dans un club de tennis, il y a 21 gauchers parmi les 103 licenciés.

a) Déterminer la fréquence de gauchers dans ce club.

$f = \dfrac{21}{103} \simeq 0,2039$

b) Déterminer un intervalle de fluctuation asymptotique au seuil de 95%.

Les conditions $\begin{cases} n = 103 \geqslant 30 \\ p = 0,12 \\ n \times p = 12,36 \geqslant 5 \\ n \times (1-p) = 90,64 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{103} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,12 - 1,96\frac{\sqrt{0,12 \times 0,88}}{\sqrt{103}}\ ;\ 0,12 + 1,96\frac{\sqrt{0,12 \times 0,88}}{\sqrt{103}}\right] \simeq [0,0572\ ;\ 0,1828]$$

c) Peut-on dire que ce club est « représentatif » de la proportion de gauchers dans le monde ?

$f = 0,2039 \notin \mathrm{I}_{103}$ donc on peut rejeter l'hypothèse selon laquelle ce club est représentatif de la proportion de gauchers dans le monde, avec un risque d'erreur de 5%.

▷ **Exercice 4.** D'après les lois génétiques de Mendel, certains croisements de différentes variétés de pois devraient donner des pois jaunes et verts dans une proportion égale à 3 pour 1. Lors d'une expérience, on a obtenu un échantillon, que l'on peut considérer comme aléatoire, présentant 176 pois jaunes et 48 pois verts. Ces résultats sont-ils cohérents avec la théorie de Mendel ?

La proportion observée de pois jaunes dans cet échantillon est $f = \dfrac{176}{176+48} = \dfrac{176}{224} \simeq 0,7857$

D'après la théorie de Mendel, cette proportion est $p = \dfrac{3}{4} = 0,75$

Les conditions $\begin{cases} n = 224 \geqslant 30 \\ p = 0,75 \\ n \times p = 168 \geqslant 5 \\ n \times (1-p) = 56 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{224} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,75 - 1,96\frac{\sqrt{0,75 \times 0,25}}{\sqrt{224}}\ ;\ 0,75 + 1,96\frac{\sqrt{0,75 \times 0,25}}{\sqrt{224}}\right] \simeq [0,6933\ ;\ 0,8067]$$

$f = 0,7857 \in \mathrm{I}_{224}$ donc on ne peut pas rejeter l'hypothèse $p = 0,75$ avec un risque d'erreur de 5%. Rien d'incohérent avec la théorie de Mendel.

▷ **Exercice 5.** Une pièce équilibrée ?

On a lancé une pièce de monnaie 100 fois et obtenu 60 fois pile. On se demande si l'on peut considérer que la pièce est équilibrée.

a) Si la pièce était équilibrée, quelle serait la probabilité d'obtenir Pile ?

$p = \dfrac{1}{2} = 0,5$

b) Calculer l'intervalle de fluctuation asymptotique au seuil de 95% de la proportion de Pile obtenus pour un échantillon de taille 100 ?

Les conditions $\begin{cases} n = 100 \geqslant 30 \\ p = 0,5 \\ n \times p = 50 \geqslant 5 \\ n \times (1-p) = 50 \geqslant 5 \end{cases}$ sont vérifiées.

$$\mathrm{I}_{100} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0,5 - 1,96\frac{\sqrt{0,5 \times 0,5}}{\sqrt{100}}\ ;\ 0,5 + 1,96\frac{\sqrt{0,5 \times 0,5}}{\sqrt{100}}\right] = [0,402\ ;\ 0,598]$$

c) On a construit l'intervalle de fluctuation sous l'hypothèse selon laquelle la pièce est équilibrée. La règle de décision est la suivante :

- si la proportion de Pile dans l'échantillon n'appartient pas à l'intervalle de fluctuation, on rejette l'hypothèse ;
- si la proportion de Pile dans l'échantillon appartient à l'intervalle de fluctuation, on ne peut pas rejeter l'hypothèse.

Appliquer la règle de décision et conclure.

Dans notre échantillon, la fréquence observée est $f = \dfrac{60}{100} = 0,6 \notin \mathrm{I}_{100}$ donc on rejette l'hypothèse selon laquelle la pièce est équilibrée.

▷ **Exercice 6.** On souhaite utiliser l'intervalle de fluctuation asymptotique $\mathrm{J}_n = \left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

a) Pour $p = 0,02$, déterminer la plus petite valeur de n vérifiant les conditions d'utilisation : $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

$np \geq 5 \iff n \geq \dfrac{5}{p} = 250$ et $n(1-p) \geq 5 \iff n \geq \dfrac{5}{1-p} = 6,25$ donc la plus petite valeur de n qui vérifie les trois conditions est $n = 250$

b) Déterminer ensuite la plus petite valeur de n pour laquelle l'amplitude de l'intervalle de fluctuation est inférieure à 0,01.

On suppose que les trois conditions précédentes sont vérifiées.

L'amplitude de l'intervalle de fluctuation est $\dfrac{2}{\sqrt{n}}$ donc on résout l'inéquation $\dfrac{2}{\sqrt{n}} \leq 0,01$:

$\dfrac{2}{\sqrt{n}} \leq 0,01 \iff \sqrt{n} \geq \dfrac{2}{0,01} \iff \sqrt{n} \geq 200 \iff n \geq 200^2 \iff n \geq 40000$

Estimation

▷ **Exercice 7.** Une société veut estimer la taille du marché potentiel d'un nouveau produit auprès de femmes âgées de plus de 25 ans. Un premier sondage est effectué auprès de 40 femmes âgées de plus de 25 ans. 24 sont satisfaites de ce nouveau produit. Lors d'un deuxième sondage auprès de 225 femmes, 150 sont satisfaites du nouveau produit. Soit p la proportion de femmes de plus de 25 ans satisfaites de ce nouveau produit.

1. Étude du premier échantillon.

a) Calculer une estimation ponctuelle f de la proportion p, fournie par le premier échantillon.

$f_{obs} = \dfrac{24}{40} = 0,6$

b) A partir du 1er échantillon, déterminer une estimation de p par un intervalle de confiance au niveau de confiance 0,95.
Pour les bornes, on donnera les valeurs arrondies à 0,01 près.

Les conditions $\begin{cases} n = 40 \geqslant 30 \\ n \times f_{obs} = 24 \geqslant 5 \\ n \times (1 - f_{obs}) = 16 \geqslant 5 \end{cases}$ sont vérifiées.

L'intervalle de confiance au seuil de 95% de p est donc :

$\mathrm{J} = \left[f_{obs} - \dfrac{1}{\sqrt{n}} ; f_{obs} + \dfrac{1}{\sqrt{n}}\right] = \left[0,6 - \dfrac{1}{\sqrt{40}} ; 0,6 + \dfrac{1}{\sqrt{40}}\right] \approx [\, 0,44 ; 0,76 \,]$

c) A l'issue de ce sondage, la société peut-elle prétendre que plus de la moitié des femmes de plus de 25 ans sont satisfaites de son nouveau produit ?

D'après ce qui précède, on peut donc estimer que la proportion de clientes satisfaites se situe entre 44% et 76% environ. La société ne peut donc pas prétendre que plus de la moitié de ses clientes sont satisfaites de son nouveau produit

2. Répondre aux questions 1.a) et 1.b) pour le deuxième sondage.

a) Calculer une estimation ponctuelle f de la proportion p, fournie par le premier échantillon.

$f_{obs} = \dfrac{150}{225} = \dfrac{2}{30} \approx 0,67$

b) A partir du 1er échantillon, déterminer une estimation de p par un intervalle de confiance au niveau de confiance 0,95.
Pour les bornes, on donnera les valeurs arrondies à 0,01 près.

Les conditions $\begin{cases} n = 225 \geqslant 30 \\ n \times f_{obs} = 150 \geqslant 5 \\ n \times (1 - f_{obs}) = 75 geqslant 5 \end{cases}$ sont vérifiées.

L'intervalle de confiance au seuil de 95% de p est donc :

$\mathrm{J} = \left[f_{obs} - \dfrac{1}{\sqrt{n}} ; f_{obs} + \dfrac{1}{\sqrt{n}}\right] = \left[\dfrac{2}{3} - \dfrac{1}{\sqrt{225}} ; \dfrac{2}{3} + \dfrac{1}{\sqrt{225}}\right] \approx [\, 0,6 ; 0,73 \,]$

c) A l'issue de ce sondage, la société peut-elle prétendre que plus de la moitié des femmes de plus de 25 ans sont satisfaites de son nouveau produit?

D'après ce qui précède, on peut donc estimer que la proportion de clientes satisfaites se situe entre 60% et 73% environ. La société peut donc prétendre que plus de la moitié de ses clientes sont satisfaites de son nouveau produit, avec un risque d'erreurs de 5%.

3. Déterminer le nombre minimal de femmes de plus de 25 ans que l'on doit interroger pour obtenir une estimation de p à plus ou moins 1%.

La précision de l'intervalle de confiance donne une estimation de p avec une précision de $\pm\dfrac{1}{\sqrt{n}}$.

$\dfrac{1}{\sqrt{n}} = 0,1 \iff \sqrt{n} = \dfrac{1}{0,1} \iff \sqrt{n} = 10 \iff n = 10^2 = 100$. Il faut donc interroger au moins 100 personnes.

▷ **Exercice 8.** Lors d'essais cliniques avant la commercialisation d'un médicament, on a observé que sur un échantillon de 3000 personnes, 2941 n'ont présenté aucun effet secondaire. Le laboratoire est certain de ne pas demander l'agrément pour ce médicament si le pourcentage de personnes souffrant d'effets secondaires dépasse 3%.

Pensez-vous que ce médicament peut-être commercialisé?

Dans l'échantillon, la proportion observée de personnes souffrant d'effets secondaires est : $f_{obs} = \dfrac{3000-2941}{3000} = \dfrac{59}{3000} \simeq 0,0197$

Les conditions $\begin{cases} n = 3000 \geqslant 30 \\ n \times f_{obs} = 59 \geqslant 5 \\ n \times (1-f_{obs}) = 2941 \geqslant 5 \end{cases}$ sont vérifiées. L'intervalle de confiance au seuil de 95% de p est donc :

$$\mathrm{J} = \left[f_{obs} - \frac{1}{\sqrt{n}} \; ; \; f_{obs} + \frac{1}{\sqrt{n}} \right] = \left[\frac{59}{3000} - \frac{1}{\sqrt{3000}} \; ; \; \frac{59}{3000} + \frac{1}{\sqrt{3000}} \right] \approx [\, 0,0014 \; ; \; 0,0379 \,]$$

La proportion p se situe donc dans l'intervalle $[\, 0,0014 \; ; \; 0,0379 \,]$ et le laboratoire ne peut donc pas être certain que $p < 3\%$. Il ne peut donc pas commercialiser ce médicament.

Déterminer la taille d'un échantillon

▷ **Exercice 9.** Deux candidats se présentent à une élection. Un sondage portant sur n personnes ($n \geq 30$) a donné 51,5% des suffrages au candidat A et 48,5% des suffrages au candidat B.

a) Déterminer, au seuil de confiance 95%, un intervalle de confiance de la proportion p des votants pour la candidat A et de la proportion p' des votants pour le candidat B.

Candidat A : Dans l'échantillon, la proportion observée de personnes déclarant vouloir voter pour A est $f_{obs} = 0,515$

Les conditions $\begin{cases} n \geqslant 30 \\ 30 \times f_{obs} \geq 15,45 \geq 5 \\ n \times (1-f_{obs}) \geq 14,55 \geq 5 \end{cases}$ sont vérifiées. L'intervalle de confiance au seuil de 95% de p est donc :

$$\mathrm{J} = \left[f_{obs} - \frac{1}{\sqrt{n}} \; ; \; f_{obs} + \frac{1}{\sqrt{n}} \right] = \left[0,515 - \frac{1}{\sqrt{n}} \; ; \; 0,515 + \frac{1}{\sqrt{n}} \right]$$

Candidat B : Dans l'échantillon, la proportion observée de personnes déclarant vouloir voter pour B est $f'_{obs} = 0,485$

Les conditions $\begin{cases} n \geqslant 30 \\ 30 \times f'_{obs} \geq 14,55 \geq 5 \\ n \times (1-f'_{obs}) \geq 15,45 \geq 5 \end{cases}$ sont vérifiées. L'intervalle de confiance au seuil de 95% de p est donc :

$$\mathrm{J}' = \left[f'_{obs} - \frac{1}{\sqrt{n}} \; ; \; f'_{obs} + \frac{1}{\sqrt{n}} \right] = \left[0,485 - \frac{1}{\sqrt{n}} \; ; \; 0,485 + \frac{1}{\sqrt{n}} \right]$$

b) Combien de personnes ont-elles été interrogées pour que l'organisateur affirme, au niveau de confiance 0,95, que le candidat A va être élu?

Pour qu'il n'y ait pas d'incertitude sur le résultat à la lecture de ce sondage, les intervalles de confiance des deux candidats doivent avoir une intersection vide. Pour cela, il faut et il suffit que la condition $0,485 + \dfrac{1}{\sqrt{n}} < 0,515 - \dfrac{1}{\sqrt{n}}$ soit vérifiée :

$0,485 + \dfrac{1}{\sqrt{n}} < 0,515 - \dfrac{1}{\sqrt{n}} \iff \dfrac{2}{\sqrt{n}} < 0,03 \iff \sqrt{n} > \dfrac{2}{0,03} \iff n > \left(\dfrac{2}{0,03}\right)^2 \simeq 4444,4$ donc il faut avoir interrogé au moins 4445 personnes.

▷ **Exercice 10.** On réalise un sondage sur un échantillon de n personnes afin de connaître le pourcentage de personnes qui vont partir en vacances l'été suivant. Quelle doit être la taille minimale de l'échantillon afin que l'intervalle de confiance de cette proportion nous donne celle-ci à 2% près avec une probabilité au moins égale à 0,95 ?

L'intervalle de confiance nous donne une estimation de p avec une précision égale à $\pm\frac{1}{\sqrt{n}}$ donc on résout :
$\frac{1}{\sqrt{n}} \leq 0,02 \iff \sqrt{n} \geq \frac{1}{0,02} \iff \sqrt{n} \geq 50 \iff n \geq 50^2 \iff n > 2500$. Il faut donc interroger au moins 2500 personnes.