SciencesPo

STATISTIQUES APPLIQUÉES AUX SCIENCES HUMAINES & SOCIALES

Niveau introductif

EXERCICES SUR LES INTERVALLES DE FLUCTUATION

Exercice 1

Un candidat lors une élection souhaite savoir s'il pourra être élu dès le premier tour (c'est à dire récolter plus de 50% des voix). Il organise un sondage portant sur un échantillon représentatif comportant 500 votants.

1. En supposant que 50% de la population souhaite voter pour ce candidat, donner l'intervalle de fluctuation au seuil de 95% pour un échantillon de 500 personnes.

2. Sur les 500 personnes interrogées, 223 disent qu'elles voteront pour ce candidat. Peut-il espérer être élu dès le premier tour ?

Solution

1. On suppose que la proportion de la population qui votera pour ce candidat est connue et vaut $p = 0,5$

L'effectif de l'échantillon est $n = 500$.

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 500 \geq 30 \quad np = 250 \geq 5 \quad n(1-p) = 250 \geq 5$$

L'intervalle de fluctuation au seuil de 95% demandé est donc

$$I = \left[0,5 - 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{500}} ; 0,5 + 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{500}}\right] \approx [0,456; 0,544]$$

2. 223 personnes par rapport aux 500 interrogées représentent une fréquence de

$$f = \frac{223}{500} = 0,446$$

Cette fréquence n'appartient pas à l'intervalle de fluctuation : Il est donc très peu vraisemblable que ce candidat soit élu dès le premier tour.

Exercice 2

On cherche à savoir si une pièce est truquée à partir d'un échantillon de lancers de pièces : on obtient 2 050 fois piles en lançant 4 000 fois cette pièce et on veut tester l'hypothèse $p = 0,5$.

Solution

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 4\,000 \geq 30 \quad np = 2\,000 \geq 5 \quad n(1-p) = 2\,000 \geq 5$$

On sait que si $n = 4\,000$ alors pour au moins 95% des expériences (qui consistent à lancer 4 000 pièces), les fréquences appartiendront à l'intervalle de fluctuation

$$I = \left[0,5 - 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{4\,000}} ; 0,5 + 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{4\,000}}\right] \approx [0,485; 0,515]$$

Ici, on sait que

$$f = \frac{2\,050}{4\,000} = 0,5125$$

donc $f \in I$. La pièce n'est certainement pas truquée.

Exercice 3

Dans la réserve indienne d'Aamjiwnaag, située au Canada il est né entre 1999 et 2003, 132 enfants dont 46 garçons. Est-ce le fruit du hasard ? On considèrera que la proportion théorique est p = 0,5.

Solution

Il y a 46 garçons sur 132 naissances alors que la proportion théorique de garçons est de $p = 0,5$.

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 132 \geq 30 \quad np = 66 \geq 5 \quad n(1-p) = 66 \geq 5$$

donc l'intervalle de fluctuation au seuil de 95% est

$$I = \left[0,5 - 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{132}} ; 0,5 + 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{132}}\right] \approx [0,415; 0,585]$$

Il n'y a donc que 5% de chances d'obtenir une valeur en dehors de cet intervalle. La fréquence de l'échantillon est égale à

$$f = \frac{46}{132} = 0,348 \notin I$$

Elle n'appartient pas à l'intervalle de fluctuation. Les naissances ne sont donc pas le fruit du hasard. On peut suspecter une influence génétique ou environnemental dans le sexe des bébés, en majorité des filles.

Exercice 4

Les entreprises sont censées ne pas faire de discrimination quant au sexe des personnes employées. Deux entreprises A et B ont respectivement 41 femmes pour 100 employés et 4 850 femmes sur 10000 employés. Pour chacune des entreprises, la sélection est-elle équitable ?

Solution

- La proportion théorique $p = 0,5$ est connue. On a bien $\{n = 100 \geq 30 \; np = 50 \geq 5 \; n(1-p) = 50 \geq 5$

Pour l'entreprise A l'intervalle de fluctuation de la fréquence au seuil de 95% est :

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

or $f = \frac{41}{100} = 0,41$

$f \in [0,402; 0,598]$ donc l'échantillon peut représenter une **situation de parité**.

- Pour l'entreprise B, l'intervalle de fluctuation est

$$I = \left[0,5 - 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{10\,000}} ; 0,5 + 1,96 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{10\,000}}\right] = [0,4902; 0,5098]$$

or $f = \frac{4\,850}{10\,000} = 0,485$

$f \notin [0,4902; 0,5098]$, donc l'échantillon n'est **pas représentatif d'une situation de parité**.

Exercice 5

Dans la ville F, on considère qu'il y a 265 jours de soleil par an. Déterminer l'intervalle de fluctuation de la fréquence de jours de soleil sur une période de 6 mois. On considérera qu'une année comporte 365 jours et 6 mois comportent 183 jours.

Solution

La proportion moyenne de jours de soleil est connue

$$p = \frac{265}{365} \approx 0,726$$

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 183 \geq 30 \; np = 183 \times \frac{265}{365} \geq 5 \; n(1-p) = 183 \times \frac{100}{365} \geq 5$$

L'intervalle de fluctuation à 95% pour 6 mois est donc

$$I = \left[\frac{265}{365} - 1,96 \times \frac{\sqrt{\frac{265}{365} \times \frac{100}{365}}}{\sqrt{183}} ; \frac{265}{365} + 1,96 \times \frac{\sqrt{\frac{265}{365} \times \frac{100}{365}}}{\sqrt{183}}\right] \approx [0,661; 0,791]$$

Pour 6 mois, soit 183 jours, l'intervalle à 95 % devient

$$[0,661 \times 183 \,; 0,791 \times 183] \approx [121 \,; \; 145]$$

On peut espérer **entre 121 et 145 jours de soleil** sur 6 mois.

Exercice 6

Si on lance un dé, la proportion d'avoir une valeur supérieure ou égale à 5 est de 1/3.

1. Déterminer les intervalles de fluctuation à 95% si on lance le dé 50 fois, 250 fois, 1000 fois.

2. Combien de fois faudrait-il lancer le dé pour que l'intervalle de fluctuation correspondant à la sortie d'un nombre supérieur ou égal à 5 ait une amplitude inférieure à 0,01 ?

Solution

1. $p = 1/3$. Dans tous les cas, les conditions sont respectées

$$\{n \geq 30 \; np \geq 5 \; n(1-p) \geq 5$$

- Si on lance 50 fois le dé, on obtient l'intervalle de fluctuation au seuil de 95% suivant

$$I = \left[\frac{1}{3} - 1,96 \times \frac{\sqrt{\frac{1}{3} \times \frac{2}{3}}}{\sqrt{50}} ; \frac{1}{3} + 1,96 \times \frac{\sqrt{\frac{1}{3} \times \frac{2}{3}}}{\sqrt{50}}\right] \approx [0,203; 0,464]$$

- Si on lance 250 fois le dé, on obtient l'intervalle de fluctuation au seuil de 95% suivant

$$I = \left[\frac{1}{3} - 1,96 \times \frac{\sqrt{\frac{1}{3} \times \frac{2}{3}}}{\sqrt{250}} ; \frac{1}{3} + 1,96 \times \frac{\sqrt{\frac{1}{3} \times \frac{2}{3}}}{\sqrt{250}}\right] \approx [0,275; 0,392]$$

- Si on lance 1000 fois le dé, on obtient l'intervalle de fluctuation au seuil de 95% suivant

$$I = \left[\frac{1}{3} - 1,96 \times \frac{\sqrt{\frac{1}{3}\times\frac{2}{3}}}{\sqrt{1\,000}}; \frac{1}{3} + 1,96 \times \frac{\sqrt{\frac{1}{3}\times\frac{2}{3}}}{\sqrt{1\,000}}\right] \approx [0,304; 0,363]$$

2. L'amplitude, différence entre la borne inférieure et la borne supérieure, doit être inférieure à 0,01.
Soit $n \in N$, l'amplitude de l'intervalle

$$\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

est de

$$2 \times 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} = 3,92\frac{\sqrt{\frac{1}{3}\times\frac{2}{3}}}{\sqrt{n}} = \frac{3,92\sqrt{2}}{3\sqrt{n}}$$

$$\frac{3,92\sqrt{2}}{3\sqrt{n}} \leq 0,01 \Leftrightarrow \frac{3,92\sqrt{2}}{3} \leq 0,01 \times \sqrt{n} \Leftrightarrow \frac{3,92\sqrt{2}}{3\times 0,01} \leq \sqrt{n} \Leftrightarrow \left(\frac{3,92\sqrt{2}}{3\times 0,01}\right)^2 \leq n$$

Or,

$$\left(\frac{3,92\sqrt{2}}{3\times 0,01}\right)^2 \approx 34\,147,6$$

On devra donc lancer le dé au moins **34 148 fois**.

Exercice 7

Au Royaume-Uni, 31% de collégiens souffrent d'asthme. Dans un collège de 284 élèves, 81 ont mentionné souffrir d'asthme. Ce collège présente-t-il des statistiques inquiétantes par rapport à l'ensemble de la population ?

Solution

Pour $n = 284$ élèves et $p = 0,31$, les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 284 \geq 30 \quad np = 88,04 \geq 5 \quad n(1-p) = 195,96 \geq 5$$

L'intervalle de fluctuation 95% est donc

$$I = \left[0,31 - 1,96 \times \frac{\sqrt{0,31\times 0,69}}{\sqrt{284}}; 0,31 + 1,96 \times \frac{\sqrt{0,31\times 0,69}}{\sqrt{284}}\right] \approx [0,256; 0,364]$$

La fréquence des élèves du collège souffrant d'asthme

$$f = \frac{81}{284} \approx 0,285$$

est dans cet intervalle donc la population du collège n'est pas significativement différente de la population du Royaume-Uni.

Exercice 8

La fréquence des yeux bleus en France est d'environ 0,31.
On a prélevé un échantillon de 50 individus dont 15 ont les yeux bleus.
1. Quel est l'intervalle de fluctuation au seuil de 95 % ?
2. Cet échantillon est-il représentatif de la population pour ce caractère ?
3. Votre classe est-elle un échantillon représentatif de la population ?

Solution

1. La proportion des yeux bleus est connue $p = 0,31$
L'effectif de l'échantillon est $n = 50$
Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 50 \geq 30 \quad np = 15,5 \geq 5 \quad n(1-p) = 34,5 \geq 5$$

L'intervalle de fluctuation demandé est donc

$$I = \left[0,31 - 1,96 \times \frac{\sqrt{0,31\times 0,69}}{\sqrt{50}}; 0,31 + 1,96 \times \frac{\sqrt{0,31\times 0,69}}{\sqrt{50}}\right] \approx [0,182; 0,438]$$

2. La fréquence des yeux bleus dans cet échantillon est

$$f = \frac{15}{50} = 0,3 \in I$$

Cet échantillon est **représentatif de la population**.
3. !!!

Exercice 9

Depuis 1996, l'accès au Musée du Louvre est gratuit le premier dimanche de chaque mois. La proportion des visiteurs français ces jours-là est de 0,59. Une toute nouvelle exposition est proposée. On souhaite connaître son impact sur la fréquentation des visiteurs les jours de gratuité. Sur un échantillon de visiteurs un jour de gratuité, 67 % sont français. Décider si la nouvelle exposition a eu un impact si l'échantillon comporte 50 visiteurs, 500 visiteurs.

Solution

- La proportion de visiteurs français est connue $p = 0,59$

L'effectif de l'échantillon est $n = 50$

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 50 \geq 30 \; np = 29,5 \geq 5 \; n(1-p) = 20,5 \geq 5$$

L'intervalle de fluctuation demandé est donc

$$I = \left[0,59 - 1,96 \times \frac{\sqrt{0,59 \times 0,41}}{\sqrt{50}}; 0,59 + 1,96 \times \frac{\sqrt{0,59 \times 0,41}}{\sqrt{50}}\right] \approx [0,454; 0,726]$$

$0,67 \in I$ donc cette expo n'a pas eu d'influence déterminante.

- L'effectif de l'échantillon est $n = 500$

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 500 \geq 30 \; np = 295 \geq 5 \; n(1-p) = 205 \geq 5$$

L'intervalle de fluctuation demandé est donc

$$I = [0,59 - 1,96 \times \frac{\sqrt{0,59 \times 0,41}}{\sqrt{500}}; 0,59 + 1,96 \times \frac{\sqrt{0,59 \times 0,41}}{\sqrt{500}}] \approx [0,547; 0,633]$$

$0,67 \notin [0,545; 0,635]$ donc cette exposition semble **déterminante** au regard de cet effectif (et semble inciter les français à venir).

Exercice 10

La fréquence d'utilisation du diesel dans les moteurs de voiture est aujourd'hui de 0,52 en France. On a étudié des échantillons dans différents départements.

- **Échantillon 1** Taille : 30 Fréquence : 0,6.
- **Échantillon 2** Taille : 35 Fréquence : 0,3.

Pour chaque échantillon, indiquer si la fréquence observée appartient ou non à l'intervalle de fluctuation au seuil de 95 %.

Solution

- La proportion de diesel est connue $p = 0,52$

L'effectif de l'échantillon 1 est $n = 30$

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 30 \geq 30 \; np = 15,6 \geq 5 \; n(1-p) = 14,4 \geq 5$$

L'intervalle de fluctuation demandé est donc

$$I = \left[0,52 - 1,96 \times \frac{\sqrt{0,52 \times 0,48}}{\sqrt{30}}; 0,52 + 1,96 \times \frac{\sqrt{0,52 \times 0,48}}{\sqrt{30}}\right] \approx [0,341; 0,699]$$

$0,6 \in I$ donc l'échantillon 1 est **représentatif** de la population au seuil de 95%.

- La proportion de diesel est connue $p = 0,52$

L'effectif de l'échantillon 2 est $n = 35$

Les conditions pour utiliser un intervalle de fluctuation sont bien réalisées.

$$\{n = 35 \geq 30 \; np = 18,2 \geq 5 \; n(1-p) = 16,8 \geq 5$$

L'intervalle de fluctuation demandé est donc

$$I = \left[0,52 - 1,96 \times \frac{\sqrt{0,52 \times 0,48}}{\sqrt{35}}; 0,52 + 1,96 \times \frac{\sqrt{0,52 \times 0,48}}{\sqrt{35}}\right] \approx [0,354; 0,686]$$

$0,3 \notin I$ donc l'échantillon 2 n'est **pas représentatif** de la population au seuil de 95%.