I FLUCTUATION D'ÉCHANTILLONNAGE (VIDÉO 1)

1 DÉCISION À PARTIR DE LA FRÉQUENCE D'UN ÉCHANTILLON (VIDÉO 2)

Quand les critères d'approximation sont vérifiés, l'intervalle de fluctuation asymptotique I_n permet de déterminer des seuils de décision :

— pour accepter ou rejeter l'hypothèse selon laquelle p est la proportion d'un caractère dans la population ;

— pour déterminer si un échantillon issu de la population est représentatif.

On formule l'hypothèse que la proportion d'un caractère dans la population est p.
On prélève dans la population un échantillon de taille n et on note f la fréquence observée du caractère étudié.
Lorsque $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ on pose :

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

— Si la fréquence observée f n'appartient pas à l'intervalle I_n, alors on rejette l'hypothèse selon laquelle p est la proportion du caractère étudié dans la population avec un risque d'erreur de 5 %.

— Si la fréquence observée f appartient à l'intervalle I_n, alors l'hypothèse selon laquelle p est la proportion du caractère étudié dans la population est acceptée.

EXEMPLE : EXTRAIT SUJET BAC

Enoncé :
Le site internet « ledislight.com » spécialisé dans la vente de matériel lumineux vend deux sortes de rubans LED flexibles : un premier modèle dit d'« intérieur » et un deuxième modèle dit d'« extérieur ». Le site internet dispose d'un grand stock de ces rubans LED.
Le fournisseur affirme que, parmi les rubans LED d'extérieur expédiés au site internet, 5 % sont défectueux. Le responsable internet désire vérifier la validité de cette affirmation. Dans son stock, il prélève au hasard 400 rubans LED d'extérieur parmi lesquels 25 sont défectueux.
Ce contrôle remet-il en cause l'affirmation du fournisseur ?

Correction :
On prend un échantillon de taille 400 donc $n = 400$. Le fournisseur affirme que 5 % des rubans sont défectueux donc la probabilité qu'un ruban soit défectueux est $p = 0{,}05$.
$n = 400 \geqslant 30$, $np = 20 \geqslant 5$ et $n(1-p) = 380 \geqslant 5$ donc on peut déterminer l'intervalle de fluctuation asymptotique de la proportion de rubans défectueux au seuil de 95 % :

$$\begin{aligned} I &= \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}} ; p + 1{,}96\sqrt{\frac{p(1-p)}{n}} \right] \\ &= \left[0{,}05 - 1{,}96\sqrt{\frac{0{,}05 \times 0{,}95}{400}} ; 0{,}05 + 1{,}96\sqrt{\frac{0{,}05 \times 0{,}95}{400}} \right] \approx [0{,}029 ; 0{,}071] \end{aligned}$$

La fréquence de rubans défectueux dans l'échantillon considéré est $f = \dfrac{25}{400} = 0{,}0625$.
Or $f \in I$ donc il n'y a pas de raison de remettre en cause l'affirmation du fournisseur.

EXERCICE CORRIGÉ : VIDÉO 3

Enoncé :
Dans un forum on a constaté que 28 personnes sur 100 ont regardé la série dont la part d'audience a été estimée à 34 %. Ce résultat remet-il en question l'estimation de la part d'audience de la série ?

Correction :
La fréquence observée de la part d'audience dans l'échantillon de taille $n = 100$ est : $f = \dfrac{28}{100} = 0,28$.
On connaît la proportion dans la population totale : $p = 0,34$

$n = 100 \geqslant 30$, $np = 34 \geqslant 5$ et $n(1-p) = 66 \geqslant 5$ donc on peut déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la part d'audience de la série dans les échantillons de taille 100 :

$$\begin{aligned} I_{100} &= \left[p - 1,96\sqrt{\frac{p(1-p)}{n}}\,;\, p + 1,96\sqrt{\frac{p(1-p)}{n}}\right] \\ &= \left[0,34 - 1,96\sqrt{\frac{0,34 \times 0,66}{100}}\,;\, 0,34 + 1,96\sqrt{\frac{0,34 \times 0,66}{100}}\right] \\ &\approx [0,247\,;\,0,433] \end{aligned}$$

d'où : $I_{100} = [0,247;0,433]$.
Comme $0,28 \in [0,247;0,433]$, l'estimation d'une part d'audience de 34 % pour la série n'est pas remise en cause.

II INTERVALLE DE CONFIANCE (VIDÉO 4)

On cherche à estimer avec un certain niveau de confiance, la proportion p **inconnue** d'un caractère au sein d'une population à partir d'un échantillon de taille n.

1 DÉFINITION

Soit f la fréquence observée d'un caractère dans un échantillon de taille n.
Sous les conditions usuelles d'approximation $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}; f + \dfrac{1}{\sqrt{n}}\right]$ est un intervalle de confiance au niveau de confiance 0,95 de la proportion inconnue p dans la population.

REMARQUES

— En pratique, les conditions de validité de la formule peuvent être vérifiées à posteriori.

— La précision de l'intervalle de confiance est donnée par son amplitude $\dfrac{2}{\sqrt{n}}$.
Plus la taille de l'échantillon est grande, plus les intervalles de confiance obtenus sont précis.

EXEMPLE : SUITE DU SUJET BAC PRÉCÉDENT

Enoncé :
Le fournisseur n'a donné aucune information concernant la fiabilité des rubans LED d'intérieur. Le directeur du site souhaite estimer la proportion de rubans LED d'intérieur défectueux. Pour cela, il prélève un échantillon aléatoire de 400 rubans d'intérieur, parmi lesquels 38 sont défectueux. Donner un intervalle de confiance de cette proportion au seuil de confiance de 95 %.

Correction :
Le fournisseur n'a donné aucune information concernant la fiabilité des rubans LED d'intérieur. Le directeur du site souhaite estimer la proportion de rubans LED d'intérieur défectueux. Pour cela, il prélève un échantillon aléatoire de 400 rubans d'intérieur, parmi lesquels 38 sont défectueux ce qui fait une fréquence $f = \dfrac{38}{400} = 0{,}095$.
$n = 400 \geqslant 30$, $nf = 38 \geqslant 5$ et $n(1-f) = 362 \geqslant 5$ donc on peut établir un intervalle de confiance de cette proportion au seuil de confiance de 95 % :

$$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right] = \left[0{,}095 - \frac{1}{\sqrt{400}} ; 0{,}095 + \frac{1}{\sqrt{400}}\right] = [0{,}045 ; 0{,}145].$$

EXERCICE CORRIGÉ (Vidéos 5-6-7)

Enoncé :
On interroge au hasard 100 clients ayant effectué des achats à la sortie d'une grande surface. Le temps d'attente aux caisses a été jugé raisonnable par 52 personnes interrogées.

1. Peut-on considérer que plus de 50% des clients de cette grande surface estiment que le temps d'attente aux caisses est raisonnable ? **(Vidéo 5)**
2. Déterminer le nombre minimal de clients qu'il faut interroger pour estimer la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable avec une précision inférieure à 0,1.**(Vidéo 6)**
3. À fréquence observée égale à 0,52, quel nombre de clients aurait-il fallu interroger pour estimer que plus de 50% des clients trouvent que le temps d'attente aux caisses est raisonnable ? **(Vidéo 7)**

Correction question 1 :(Vidéo 5) ***Détermination d'un intervalle de confiance. Validation d'une hypothèse.***
Soit $f = \dfrac{52}{100} = 0{,}52$ la fréquence des clients qui estiment que le temps d'attente aux caisses est raisonnable.
Les bornes de l'intervalle de confiance au niveau de confiance 95% de la proportion des clients qui estiment que le temps d'attente aux caisses est raisonnable sont :

$$f - \frac{1}{\sqrt{n}} = 0{,}52 - \frac{1}{\sqrt{100}} = 0{,}42 \text{ et } f + \frac{1}{\sqrt{n}} = 0{,}52 + \frac{1}{\sqrt{100}} = 0{,}62$$

On a : $n = 100$, $0{,}42 \leqslant p \leqslant 0{,}62$, $100 \times 0{,}42 \leqslant np \leqslant 100 \times 0{,}62$ et $100 \times (1 - 0{,}62) \leqslant n\left(1-p\right) \leqslant 100 \times (1 - 0{,}42)$.
Soit $n \geqslant 30$, $42 \leqslant np \leqslant 62$ et $38 \leqslant n\left(1-p\right) \leqslant 58$. Les conditions d'approximation d'un intervalle de confiance au niveau de confiance 95 % sont vérifiées.
Un intervalle de confiance au niveau de confiance 95 % de la proportion de clients qui estiment que le temps d'attente aux caisses est raisonnable est $[0{,}42 ; 0{,}62]$.
La borne inférieure de l'intervalle de confiance est 0,42, il est donc possible que moins de 50% des clients trouvent que le temps d'attente aux caisses est raisonnable.

Correction question 2 :(Vidéo 6)

Détermination de la taille de l'échantillon à partir de l'amplitude d'un intervalle de confiance.

La précision de l'estimation de p est $\dfrac{2}{\sqrt{n}}$. Pour tout entier naturel n :

$$\begin{aligned}\frac{2}{\sqrt{n}} < 0{,}1 &\iff \frac{1}{\sqrt{n}} < 0{,}05\\ &\iff \sqrt{n} > \frac{1}{0{,}05}\\ &\iff \sqrt{n} > 20\\ &\iff n > 400\end{aligned}$$

Il faut interroger plus de 400 clients pour obtenir une estimation de la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable avec une précision inférieure à 0,1.

Correction question 3 :(Vidéo 7)

Détermination de la taille de l'échantillon à partir d'une fréquence et de la proportion.

La borne inférieure de l'intervalle de confiance au niveau de confiance 0,95 sur un échantillon de taille n est $0{,}52 - \dfrac{1}{\sqrt{n}}$ d'où n est solution de l'inéquation :

$$\begin{aligned}0{,}52 - \frac{1}{\sqrt{n}} \geqslant 0{,}5 &\iff -\frac{1}{\sqrt{n}} \geqslant -0{,}02\\ &\iff \frac{1}{\sqrt{n}} \leqslant 0{,}02\\ &\iff \sqrt{n} \geqslant \frac{1}{0{,}02}\\ &\iff \sqrt{n} \geqslant 50\\ &\iff n \geqslant 2500\end{aligned}$$

Avec une fréquence observée égale à 0,52, il faudrait un échantillon de taille supérieure à 2500 pour que la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable appartienne à un intervalle de confiance dont la borne inférieure est supérieure à 0,5.