Chapitre 13 : Echantillonnage

EXERCICE 1

Dans une entreprise, il y a 28 cadres et 92 ouvriers. Le salaire moyen des cadres est de 3 450 € et celui des ouvriers est de 1 320 €.

1. Calculer le salaire moyen de l'ensemble des salariés de cette entreprise.
2. a) Quel est le pourcentage d'augmentation du salaire moyen si on verse une prime de 35 € à chaque salarié ?

   b) On augmente le salaire de chaque cadre de 2 % et celui de chaque ouvrier de 4 %.
   Le salaire moyen dans l'entreprise a-t-il augmenté de 3% ?

1. $\bar{x} = \dfrac{28 \times 3450 + 92 \times 1320}{28 + 92} = \dfrac{218040}{120} = 1817$

Le salaire moyen est de 1817 €

2. a) Si on augmente tous les salaires de 35 €, le salaire moyen augmente de 35 €.

Le taux d'évolution du salaire moyen est donc $\dfrac{35}{1817} = 1{,}93\%$

b) Les pourcentages d'évolution ne s'ajoutent pas ! De surcroît, il n'y a pas le même nombre d'ouvriers que de cadres.

Augmenter de 2% revient à multiplier par 1,02. La masse salariale des cadres est donc $1{,}02 \times 28 \times 3450 = 98532$

Augmenter de 4% revient à multiplier par 1,04. La masse salariale des ouvriers est donc $1{,}04 \times 92 \times 1320 = 126297{,}6$

La masse salariale de l'entreprise est donc : $98532 + 126297{,}6 = 224829{,}6$ soit un salaire moyen :

$$\bar{x}_1 = \frac{224829{,}6}{120} = 1873{,}58.$$

Calculons l'augmentation du salaire moyen en pourcentage : $t = \dfrac{1873{,}58 - 1817}{1817} = 3{,}11\%$

EXERCICE 2

Un concours est organisé dans deux centres d'examens. Dans le premier centre, les garçons ont obtenu 13 de moyenne et les filles 12 de moyenne. Dans le second centre, les garçons ont obtenu 9 de moyenne et les filles 8 de moyenne.
Le président du jury en déduit que les garçons ont eu de meilleurs résultats que les filles. Est-ce si sûr?
Sachant qu'il y avait 58 garçons et 104 filles dans le premier centre, et 87 garçons et 32 filles dans le second centre, calculer la moyenne générale des garçons puis celle des filles. Conclure.

58 garçons ont 13 de moyenne dans le 1er centre soit $58 \times 13 = 754$ points pour les garçons du centre 1
87 garçons ont 9 de moyenne dans le second centre soit $87 \times 9 = 783$ points pour les garçons du centre 2
soit un total de 1537 points pour ces 145 garçons.
La moyenne des garçons est donc $\frac{1537}{145} = 10{,}6$.
104 filles ont 12 de moyenne dans le 1er centre soit $104 \times 12 = 1248$ points pour les filles du centre 1
32 filles ont 8 de moyenne dans le second centre soit $32 \times 8 = 256$ points pour les filles du centre 2
soit un total de 1504 points pour ces 136 filles.
La moyenne des filles est donc $\frac{1504}{136} = 11{,}06$: sur le regroupement des 2 échantillons, les filles obtiennent une moyenne plus élevée que les garçons.

Attention au regroupement d'échantillons, et aux populations de références différentes.

EXERCICE 3

Chez un fabricant de lames de parquet en chêne rustique on indique :

— longueur moyenne des lames : 45 cm ;

— lames de longueur inférieure à 35 cm : 23 %.

On prélève un échantillon de 60 lames dans le stock, pour vérification. On constate que 18 lames ont une longueur inférieure à 35 cm.

1. Déterminer un intervalle de fluctuation au seuil de 95 % de la fréquence des lames dont la longueur est inférieure à 35 cm dans les échantillons de taille 60.
2. L'affirmation « 23% des lames ont une longueur inférieure à 35 cm » est-elle remise en cause ?

$$I_{f_{95}} = \left[p_{théo} - \frac{1}{\sqrt{n}} \, ; \, p_{théo} + \frac{1}{\sqrt{n}} \right]$$

$$I_{f_{95}} = [\, 0{,}100 \, ; \, 0{,}360 \,]$$

valeur arrondie par défaut

valeur arrondie par excès.

1. On a un échantillon de taille $N = 60 > 30$.

$p_{théo} = 0{,}23$

$n \times p_{théo} = 60 \times 0{,}23 = 13{,}8 > 5$.

Les conditions d'application de l'intervalle de fluctuation d'échantillonnage sont réunies.

2. 95% des échantillons de taille 60 ont une fréquence observée de lames de longueur inférieure à 35 cm dans l'intervalle $I_{f_{95}}$. Pour cet échantillon on a :

$f_{obs} = \frac{18}{60} = 0{,}3$ $\quad f_{obs} \in I_{f_{95}}$.

L'échantillon est donc conforme à l'affirmation

EXERCICE 4

Le directeur commercial affirme que 80 % des consommateurs sont satisfaits de la qualité des produits commercialisés par son entreprise.
On réalise une étude de satisfaction sur un échantillon de 120 personnes. Parmi les personnes interrogées, 90 déclarent être satisfaites des produits.
Déterminer, en justifiant, si l'on doit remettre en question l'affirmation du directeur commercial.

Dans cet échantillon de taille $N = 120$ on relève une fréquence observée de consommateurs satisfaits $f_{obs} = \frac{90}{120} = 0{,}75 = 75\%$

Il s'agit d'estimer si cette mesure est significativement différente de la proportion théorique $p_{théo} = 0{,}8$.

$n > 30$

$n \times p_{théo} = 120 \times 0{,}8 = 96 > 5$

Les conditions d'application de l'intervalle de fluctuation au seuil de 95% sont réunies.

$$I_{f_{95}} = \left[p_{théo} - \frac{1}{\sqrt{n}} ; p_{théo} + \frac{1}{\sqrt{n}} \right]$$

$$I_{f_{95}} = [0{,}708 ; 0{,}892] \qquad f_{obs} \in I_{f_{95}}$$

donc cet échantillon ne permet pas de mettre en défaut l'affirmation du directeur commercial.

EXERCICE 5

Selon une publication de l'INSEE, 28 % des ménages comprennent une famille avec au moins un enfant mineur.

1. On interroge un échantillon de 100 ménages choisis au hasard, et on constate que dans cet échantillon 35 ménages comprennent une famille avec au moins un enfant mineur.
   Cet échantillon est-il représentatif de la population ?
2. On interroge au hasard 300 ménages qui résident dans le même arrondissement d'une grande agglomération, et on constate également que 105 ménages comprennent une famille avec au moins un enfant mineur.
   Cet échantillon est-il représentatif de la population ?

1. La taille de l'échantillon est $n = 100$

$n > 30$.

$p_{théo} = 0{,}28$

$f_{obs} = \frac{35}{100} = 0{,}35$

$n \times p_{théo} = 28 \quad > 5$

Les conditions d'aplication de l'intervalle de fluctuation au seuil de 95% sont réunies.

$I_{f_{95}} = \left[p_{théo} - \frac{1}{\sqrt{n}} ; p_{théo} + \frac{1}{\sqrt{n}} \right]$

$I_{f_{95}} = [0{,}18 ; 0{,}38]$

$f_{obs} \in I_{f_{95}}$ donc l'échantillon est représentatif de la population.

2. $n = 300 \quad f_{obs} = \frac{105}{300} = 0{,}35 \; ; \; p_{théo} = 0{,}28$

$I_{f_{95}} = \left[p_{théo} - \frac{1}{\sqrt{n}} ; p_{théo} + \frac{1}{\sqrt{n}} \right]$

$= [0{,}222 ; 0{,}338]$

$f_{obs} \notin I_{f_{95}}$ donc cet échantillon n'est pas représentatif de la population.

EXERCICE 6

Une urne contient des boules de différentes couleurs dont 75% de boules rouges.
Cyril tire une boule au hasard, note la couleur et la remet dans l'urne. Il prétend avoir effectué cette expérience 60 fois et avoir obtenu 35 boules rouges. Son frère Paulo affirme qu'il n'a pas fait l'expérience sérieusement.
On se propose de vérifier s'il a de bonnes raisons de l'affirmer.
1) Déterminer la proportion théorique p et la taille n de l'échantillon.
2) Calculer la fréquence observée f.
3) Calculer l'intervalle de fluctuation au seuil de 95% I_f.
4) Vérifier si la fréquence observée f appartient à l'intervalle de fluctuation I_f et conclure.

Organisons l'information :

En prélevant 60 fois une boule dans des conditions identiques (remise de la boule dans l'urne), on constitue ainsi un échantillon de taille $N = 60$.

La proportion théorique de boules rouges dans l'urne est $p_{théo} = 0,75$.

La fréquence observée de boules rouges dans l'échantillon est $f_{obs} = \frac{35}{60} = 0,5833$

1.) $p_{théo} = 0,75 \quad n = 60$

2) $f = \frac{35}{60} = 0,5833$

3) Vérifions les conditions d'application de l'intervalle de fluctuation au seuil de 95%.

$n > 30 \quad n \times p_{théo} > 5$

$I_f = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$

$I_f = [0,6209 ; 0,8791]$

4) $f_{obs} \notin I_f$ donc, avec un risque de se tromper inférieur à 5% on peut estimer que l'échantillon constitué n'est pas représentatif de la population (Les boules n'ont peut être pas été bien mélangées entre les tirages).

EXERCICE 7

La proportion de personnes aux cheveux châtains en France est d'environ 50%.
On a observé un échantillon de 150 personnes dont 89 ont les cheveux châtains.
Cet échantillon est-il représentatif de la population ?

Organisons l'information.

$p_{théo} = 0,5$ $\quad n = 150 > 30$

$f_{obs} = \dfrac{89}{150}$ $\quad n \times p_{théo} = 75 > 5$

$f_{obs} = 0,5933$

$$I_{f_{95}} = \left[p_{théo} - \frac{1}{\sqrt{n}} ; p_{théo} + \frac{1}{\sqrt{n}} \right]$$

$$I_{f_{95}} = [0,418 ; 0,582]$$

$f_{obs} \notin I_{f_{95}}$ donc avec un risque de se tromper inférieur à 5% on peut estimer que cet échantillon n'est pas représentatif. La fréquence observée peut être considérée comme significativement trop élevée.

EXERCICE 8

Dans une classe de 37 élèves, un délégué de classe a été élu avec 60% des voix.
Parmi les 17 filles, 11 d'entre elles ont voté pour ce délégué.
Les filles sont-elles représentatives des résultats des élections de délégués ?

L'échantillon constitué par les 17 filles est-il représentatif de la classe ?

Organisons l'information

11 filles sur 17 ont voté pour le délégué élu.

$\dfrac{11}{17} = 0,6471 = 64,71\%$

On cherche à établir si cette fréquence observée est significativement différente de la proportion de 60%.

Tout d'abord $n < 30$ donc les conditions d'application de l'intervalle de fluctuation ne sont pas satisfaites. $I_f = \left[0,6 - \frac{1}{\sqrt{17}} ; 0,6 + \frac{1}{\sqrt{17}}\right] = [0,357 ; 0,843]$

Plus la taille de l'échantillon est faible, plus large est l'intervalle de fluctuation.

EXERCICE 9

En 1976, dans un comté du Texas, Rodrigo Partida était condamné à huit ans de prison.
Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Hispano-américains.
En effet, 79,1% de la population de ce comté était d'origine hispanique.
Alors que, sur les 870 personnes convoquées pour être juré, il n'y eut que 339 personnes d'origine hispanique. Qu'en pensez-vous ? (D'après IREM de Paris Nord)

Organisons l'information :

il s'agit de savoir si la population hispano-américaine est significativement sous représentée dans ce jury.

$p_{\text{théo}} = 0{,}791$

taille de l'échantillon, $n = 870$

$f_{\text{obs}} = \dfrac{339}{870} = 0{,}3897$

On utilise l'intervalle de fluctuation pour établir si l'échantillon de taille 870 est représentatif de la population.

$$I_{f_{95}} = \left[p - \frac{1}{\sqrt{n}} \,;\, p + \frac{1}{\sqrt{n}}\right] = [0{,}757 \,;\, 0{,}825]$$

$0{,}3897 \ll 0{,}757$: la communauté hispano américaine est clairement sous représentée dans l'échantillon convoqué. Cet échantillon n'est pas représentatif de la population et on comprend que la désignation des jurés ait pu être qualifiée de discriminante à l'égard des Hispano-américains.

EXERCICE 10

Une maladie guérit naturellement dans 70% des cas. Un laboratoire souhaite tester l'efficacité d'un nouveau médicament.
Pour cela, on administre ce médicament à 500 personnes. Pour 77% d'entre elles, la guérison a eu lieu.
Que penser de l'efficacité de ce médicament ?

Organisons l'information :

$p_{\text{théo}} = 0,7$

on constitue un échantillon de taille $n = 500$

$f_{obs} = 0,77$

$n > 30$ $n \times p_{\text{théo}} > 5$.

On peut donc vérifier la conformité d'un échantillon à l'aide de l'intervalle de fluctuation d'échantillonnage

Il s'agit de savoir si la fréquence observée 0,77 est significativement plus élevée que 0,7.

$$I_f = \left[p_{\text{théo}} - \frac{1}{\sqrt{n}} ; p_{\text{théo}} + \frac{1}{\sqrt{n}} \right] = [0,655 ; 0,745]$$

$f_{obs} \gg 0,745$ donc on peut estimer que le médicament a une efficacité permettant d'augmenter le taux de guérison.

EXERCICE 11

Un centre commercial n'attire que 22% de clients hors de la communauté urbaine.
Souhaitant élargir sa clientèle, le centre commercial s'agrandit.
Après les travaux, voulant connaître l'impact de ses investissements, 300 clients sont interrogés :
72 d'entre eux habitent hors de la communauté urbaine.
Peut-on affirmer que l'agrandissement a eu un impact sur la fréquentation des clients hors communauté urbaine ?

<u>Organisons l'information :</u>

$n = 300$ $\qquad p_{théo} = 0,22$

$f_{obs} = \dfrac{72}{300} = 0,24$

$n > 30$ et $n \times p_{théo} = 300 \times 0,22 = 66 > 5$.

Les conditions d'application de l'intervalle de fluctuation sont réunies.

$$I_f = \left[p_{théo} - \frac{1}{\sqrt{n}} \,;\, p_{théo} + \frac{1}{\sqrt{n}}\right] = [0,162 \,;\, 0,278]$$

$f_{obs} \in I_f$ donc cette fréquence est conforme à la proportion théorique de fréquentation du magasin avant l'agrandissement. Les travaux n'ont pas eu d'impact sur la fréquentation du magasin.

EXERCICE 12

Un fournisseur d'accès à Internet disposait de 25% de part de marché avant l'arrivée d'un nouveau concurrent.
Après l'arrivée de ce concurrent, il effectue une enquête sur un échantillon de 200 foyers et obtient 19% de part de marché.
Peut-il considérer que l'arrivée de ce nouveau concurrent lui a fait perdre des parts de marché ?

Organisons l'information :

$p_{théo} = 0{,}25$

$n = 200 \quad n \times p_{théo} = 200 \times 0{,}25 = 50$

$n \times p_{théo} > 5$

$n > 30$

Les conditions d'application de l'intervalle de fluctuation sont réunies.

$f_{obs} = 0{,}19$

Il s'agit de savoir si les 19% observés sont significativement inférieurs aux 25% théoriques.

$$I_{f_{95}} = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] = [0{,}179 ; 0{,}321]$$

$0{,}19 \in I_{f_{95}}$ donc on peut estimer que l'arrivée de ce concurrent ne lui a pas fait significativement perdre de parts de marché.

EXERCICE 13

Un musée national connaît une proportion de visiteurs français égale à 71%.
Durant l'été, le musée propose une exposition temporaire sur le thème de l'Égypte ancienne.
On souhaite connaître son impact sur la fréquentation des visiteurs français.
Décider si la nouvelle exposition a eu un impact dans les cas suivants :
1) Sur un échantillon de 50 visiteurs, 82% sont français.
2) Sur un échantillon de 500 visiteurs, 77% sont français.

Organisons l'information :

$p_{\text{théo}} = 0,71$

1) $n = 50 > 30$

$n \times p_{\text{théo}} = 50 \times 0,71 = 35,5 > 5$

$f_{obs} = 0,82$

$$I_{f_{95}} = \left[0,71 - \frac{1}{\sqrt{50}} ; 0,71 + \frac{1}{\sqrt{50}}\right] = [0,568 ; 0,852]$$

$f_{obs} \in I_{f_{95}}$: sur cet échantillon, on ne peut pas estimer que la nouvelle exposition a eu un impact sur les visiteurs français.

2) $n = 500 > 30$

$n \times p_{\text{théo}} = 500 \times 0,71 = 355 > 5$

$f_{obs} = 0,77$

$$I_f = \left[0,71 - \frac{1}{\sqrt{500}} ; 0,71 + \frac{1}{\sqrt{500}}\right] = [0,665 ; 0,755]$$

$0,77 \gg 0,755$ donc sur cet échantillon, on peut estimer que la nouvelle exposition a provoquer une augmentation significative de la fréquentation des visiteurs français.

EXERCICE 14

Léanne a lancé une pièce de monnaie 50 fois. Elle a obtenu 32 fois « pile ».
A-t-elle de bonnes raisons de s'étonner du résultat ?

En théorie, la proportion d'apparition de "pile" est $p_{théo} = 0,50$

Sur un échantillon de taille $n = 50$ on observe une fréquence $f_{obs} = \frac{32}{50} = 0,64$

$n > 30$
$n \times p_{théo} > 5$
donc les conditions d'application de l'intervalle de fluctuation sont réunies.

$$I_{f_{95}} = \left[0,50 - \frac{1}{\sqrt{50}} ; 0,50 + \frac{1}{\sqrt{50}}\right] = [0,358 ; 0,642]$$

$0,64 \in I_{f_{95}}$ donc l'échantillon constitué par Léanne est considéré conforme !