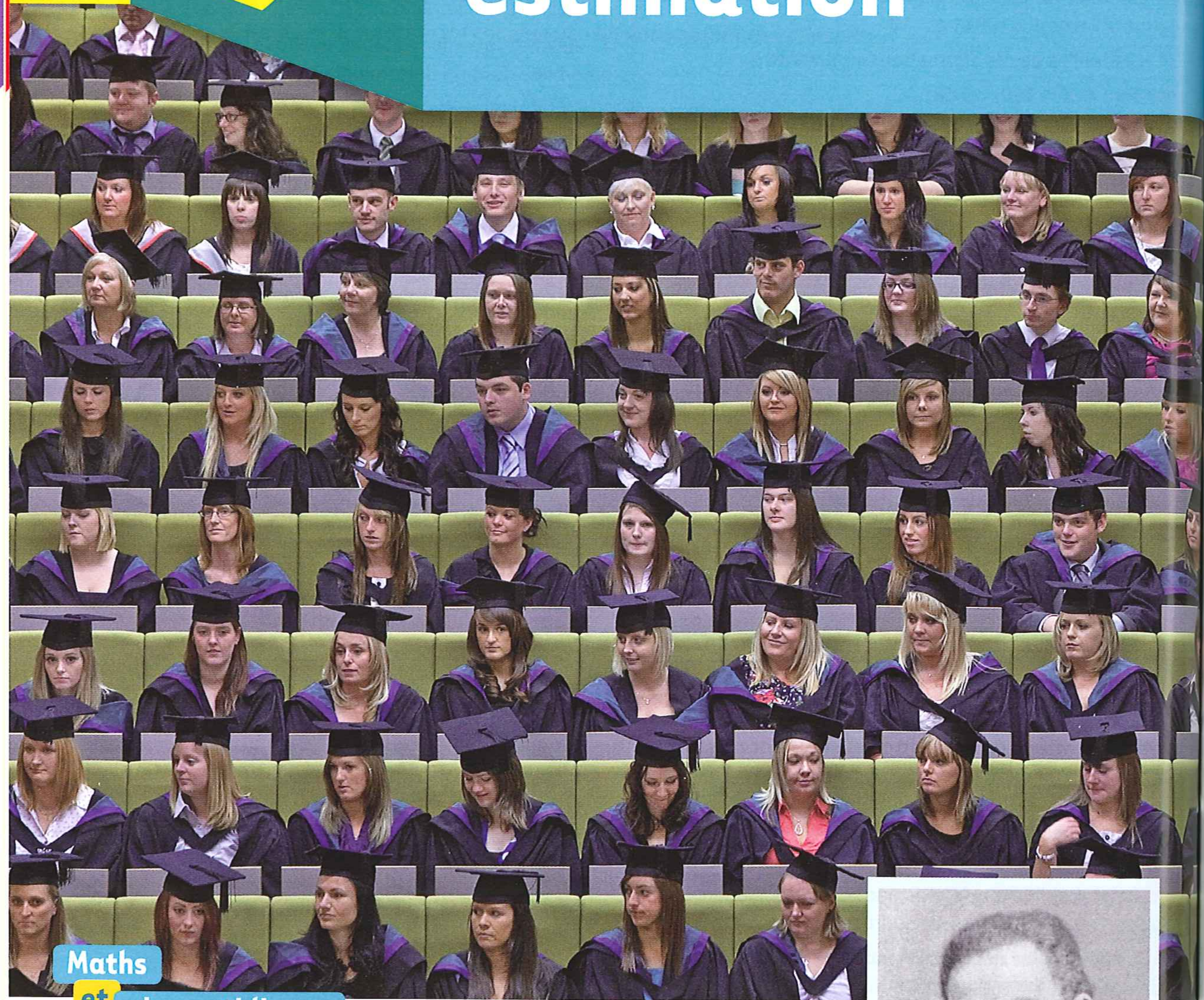
CHAPITRE 9

Fluctuation, estimation



Maths et vie quotidienne

Ces étudiants constituent un échantillon de la population dans lequel la proportion de femmes est supérieure à 80 %. Peut-on en déduire pour autant que la proportion de femmes est de 80 % dans la population de ce pays ? Dans cette université ?

Lorsqu'on utilise un échantillon pour estimer une proportion, il faut se rappeler que, même quand l'échantillon est très grand, l'estimation n'est valable qu'avec une certaine probabilité. Dans le cas d'un échantillon non représentatif, il est toujours possible d'obtenir une fréquence très éloignée de la proportion réelle.

Karl Pearson
1857-1936

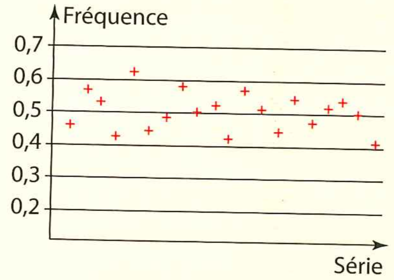
→ Chercheurs d'hier p. 258

Rappels & Questions-tests

Compléments numériques

Fluctuation

Une urne contient 50 % de boules blanches. On effectue, par simulation, 20 séries de 100 tirages avec remise, et on détermine pour chacune d'elles la fréquence de sortie d'une boule blanche.
Le graphique ci-dessous donne les 20 fréquences ainsi obtenues.



Ici la proportion p de boules blanches est connue, $p = 0{,}5$ et le nombre n de boules tirées pour chaque série est égal à 100.
Les fréquences obtenues sont différentes, elles fluctuent.
On sait que, si $n \geqslant 25$ et $0{,}2 \leqslant p \leqslant 0{,}8$, il y a 95 % de chances pour que la fréquence dans l'échantillon apparaisse dans l'intervalle :

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

Cet intervalle est appelé intervalle de fluctuation au seuil de 95 %.

1 Dans la situation ci-contre déterminez l'intervalle de fluctuation au seuil de 95 %.

2 D'après le schéma ci-contre, les résultats obtenus sont-ils en accord avec les résultats théoriques ?

3 Une urne contient 30 % de boules rouges. On effectue 200 séries de 100 tirages avec remise et on calcule, pour chaque série, la fréquence des boules rouges obtenues.
Combien de fréquences devraient être situées dans l'intervalle [0,2 ; 0,4] pour que le résultat de cette expérience soit en accord avec les résultats théoriques ?

Estimation – Sondages

• Dans une population contenant un « grand nombre » N d'individus, un pourcentage p de ces individus possède une certaine propriété $\mathcal{P}$, par exemple, « ont l'intention de voter pour tel candidat ».
Dans la pratique, pour avoir une idée de p, on ne considère qu'une partie de la population, contenant un nombre n d'individus, n étant nettement inférieur à N.
On peut alors calculer, dans cette partie de la population, le pourcentage f d'individus ayant la propriété $\mathcal{P}$. Si cette partie n'est pas « trop petite » on conçoit intuitivement que le pourcentage réel p, relatif à toute la population, devrait être « voisin » de f.

4 On considère une urne contenant des boules blanches et des boules noires. Proposez une expérience, particulièrement « malchanceuse », où la fréquence f de boules blanches observée sur n tirages serait égale à 100 %, alors qu'en réalité, le pourcentage p réel de boules blanches dans l'urne est égal à 1 %.

5 Proposez d'autres situations où on peut être amené à estimer un pourcentage p inconnu.

→ Voir les corrigés p. 361

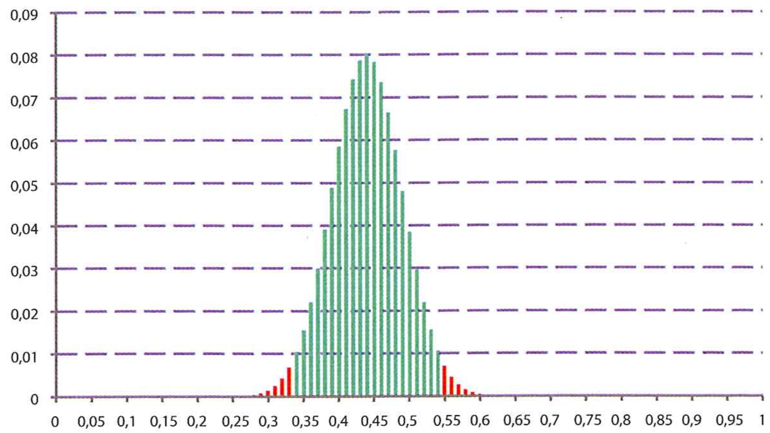
Activité LA LOI BINOMIALE POUR JUSTIFIER

1 Une situation

Une urne contient 44 % de boules blanches. On effectue 100 tirages avec remise. On note X la variable aléatoire comptant le nombre de boules blanches tirées.

1. Expliquez pourquoi X suit une loi binomiale $\mathcal{B}(n\,;p)$.
Précisez la valeur de n et celle de p.

2. On considère la variable aléatoire $F = \dfrac{X}{100}$, c'est-à-dire la variable aléatoire des fréquences observées.
On peut obtenir le diagramme en bâtons de la variable aléatoire F à l'aide d'un logiciel par exemple.



Notez que, sur le diagramme, certains bâtons ne sont pas visibles car leur hauteur est trop petite.

À l'aide de ce graphique, répondez aux questions suivantes :

a) Quelle est la fréquence qui a la plus forte probabilité ? Expliquez pourquoi ce résultat était prévisible.

b) Quelle est approximativement la probabilité de la fréquence 0,5 ?

c) Expliquez pourquoi la probabilité de l'événement $F \notin [0{,}34\,;0{,}54]$ peut être obtenue en faisant la somme des hauteurs des bâtons colorés en rouge. Vérifiez que cette somme est approximativement égale à 0,05.

2 Interpréter un résultat

Reprenons l'exemple ci-dessus. Supposons que, lors des 100 tirages, on obtienne 55 boules blanches.

1. Quelle est alors la fréquence F des boules blanches obtenues ?

2. D'après la question **1.2.c)** ci-dessus, expliquez pourquoi cette fréquence F était « très peu probable ».

Commentaire

Dans cette situation il est légitime de penser (sans en être certain) que le tirage n'a pas été effectué vraiment au hasard, ou qu'une erreur a été commise lors du calcul de F…

1 Notion d'échantillon

1.1 Définition

Définition 1 Un **échantillon** de taille n est constitué des résultats de n répétitions indépendantes de la même expérience.

1.2 Exemples

- **Cas d'une urne**

Un échantillon de taille n est constitué des résultats de n tirages avec remise.

Notons que si les tirages s'effectuent sans remise, les tirages successifs ne sont pas indépendants. En effet, dans ce cas, la composition de l'urne varie après chaque tirage.

- **Cas d'un sondage**

Supposons que l'on interroge n personnes dans une population de N personnes.

A priori, il s'agit d'une situation analogue à un tirage sans remise dans une urne ; en effet, les personnes interrogées ne sont interrogées qu'une fois.

En assimilant les personnes interrogées aux boules d'une urne, ceci signifie que l'on ne remet pas la personne interrogée « dans l'urne » avant d'interroger la suivante.

1.3 Une convention

On convient que si le nombre n de personnes interrogées est nettement inférieur au nombre total N de personnes, un sondage peut être considéré comme un tirage avec remise dans une urne.

En effet, dans ce cas, les pourcentages des diverses catégories de personnes dans la population sont très peu modifiés par la « suppression » de certaines d'entre elles, en nombre négligeable par rapport au nombre total de personnes.

2 Fluctuation d'échantillonnage

2.1 Un exemple

Reprenons l'exemple de l'urne dans laquelle la proportion de boules blanches est égale à p. Supposons que p est connu, par exemple $p = 0{,}6$.

Voici les fréquences de boules blanches obtenues, par simulation, à partir de 20 échantillons, chacun de taille 100.

0,51 – 0,62 – 0,68 – 0,55 – $\boxed{0{,}47}$ – 0,6 – 0,69 – 0,58 – 0,61 – 0,67
0,55 – 0,63 – 0,53 – 0,54 – 0,52 – 0,68 – 0,69 – 0,54 – 0,55 – 0,59

On constate sur cet exemple que les fréquences observées fluctuent. Ce phénomène est appelé **fluctuation d'échantillonnage**.

Plus précisément, on peut constater que, pour la plupart des échantillons, la fréquence de sortie d'une boule blanche se trouve dans l'intervalle [0,5 ; 0,7]. On dispose ainsi d'un ordre de grandeur du nombre d'échantillons dont la fréquence appartient à l'intervalle [0,5 ; 0,7].

Dans l'exemple, on peut vérifier qu'il y en a 19 sur 20, c'est-à-dire 95 %.

2.2 Intervalle de fluctuation asymptotique

Les résultats observés dans l'exemple précédent sont en accord avec la propriété générale suivante, démontrée en théorie des probabilités et des statistiques.

Théorème 1

Notons F_n la variable aléatoire qui à tout échantillon de taille n associe la fréquence d'un caractère.

Posons $I_n = \left[p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où p désigne la proportion de ce caractère dans la population.

Alors F_n prend ses valeurs dans I_n avec une probabilité qui s'approche de 0,95 quand n devient grand.

On pratique cette approximation dès que **$n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.**

Définition 2

L'intervalle $I_n = \left[p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé **intervalle de fluctuation asymptotique au seuil de 95 %.**

Exemple

Dans l'exemple du paragraphe **2.1** ci-dessus, $n = 100$; $p = 0{,}6$. Donc ici la variable aléatoire est égale à F_{100} et l'intervalle de fluctuation asymptotique au seuil de 95 % noté I_{100} est l'intervalle [0,503 9 ; 0,696 1]. On peut constater que pour 95 % des échantillons, la fréquence observée appartient à l'intervalle I_{100}. Ce résultat est en accord avec le théorème 1.

2.3 Retour sur la classe de seconde

On a vu en classe de seconde que pour un échantillon de taille n, l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$ est un intervalle de fluctuation de la fréquence au seuil de 95 %.

Le théorème 1 ci-dessus donne en général un résultat un peu plus précis que le théorème analogue vu en seconde où l'intervalle utilisé est $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$ car l'intervalle I_n est inclus dans l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$.

Voir exercice 54 p. 259.

3 Estimation d'une proportion inconnue à partir d'un échantillon

On suppose maintenant que la proportion p n'est pas connue.

On va voir comment il est possible d'obtenir une estimation de la valeur de p à partir de la fréquence observée dans l'échantillon.

3.1 Une propriété importante

Théorème 2

Notons f la fréquence observée dans un échantillon de taille n et p le pourcentage que l'on veut estimer.

Alors **l'intervalle** $\left[f - \dfrac{1}{\sqrt{n}} ; f + \dfrac{1}{\sqrt{n}}\right]$ **contient p avec une probabilité d'au moins 0,95.**

- **Conditions d'application.** Ce sont les mêmes que celles du théorème 1 :

$$n \geqslant 30,\ np \geqslant 5 \text{ et } n(1-p) \geqslant 5.$$

Définition 3

On dit que l'intervalle $\left[f - \dfrac{1}{\sqrt{n}} ; f + \dfrac{1}{\sqrt{n}}\right]$ est **l'intervalle de confiance de p au niveau de confiance 0,95.**

(On dit aussi avec un risque de 5 %.)

3.2 Exemple d'application

À l'aide d'un échantillon de taille 100, on souhaite estimer un pourcentage p. On suppose que $p \in [0{,}2 ; 0{,}8]$.

La fréquence observée dans cet échantillon est 0,45.

1. Vérifions que les conditions d'application du théorème 2 sont réalisées. On a :

- $n = 100$, donc $n \geqslant 30$;
- $0{,}2 \leqslant p \leqslant 0{,}8$ d'où $20 \leqslant np \leqslant 80$, donc $np \geqslant 5$;
- $0{,}2 \leqslant p \leqslant 0{,}8$ d'où $0{,}2 \leqslant 1 - p \leqslant 0{,}8$ et $20 \leqslant n(1-p) \leqslant 80$ donc $n(1-p) \geqslant 5$.

2. Les conditions étant remplies, on peut appliquer le théorème 2. On peut donc dire que l'intervalle de confiance de p au niveau de confiance 0,95 est l'intervalle $\left[0{,}45 - \dfrac{1}{\sqrt{100}} ; 0{,}45 + \dfrac{1}{\sqrt{100}}\right]$, c'est-à-dire [0,35 ; 0,55]. Cet intervalle contient p avec une probabilité d'au moins 0,95.

Application

Compléments numériques

OBJECTIF 1 Connaître l'intervalle de fluctuation asymptotique au seuil de 95 %

Dans une population, la proportion p d'un caractère est connue. La variable aléatoire F_n qui, à tout échantillon de taille n, associe la fréquence de ce caractère, prend ses valeurs dans l'intervalle de fluctuation asymptotique au seuil de 95 %, $I_n = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$, avec une probabilité qui s'approche de 0,95 quand n devient grand. Dans la pratique, on utilise cet intervalle dès que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

EXERCICE RÉSOLU A

La proportion de naissances d'enfants prématurés 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. On réalise une enquête auprès d'un échantillon aléatoire de 400 naissances correspondant à des femmes ayant eu pendant leur grossesse un travail pénible. Les chercheurs décident *a priori* que si la proportion d'enfants nés prématurés dans cet échantillon est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 0,95 alors leur hypothèse sera acceptée. On sait que dans l'échantillon, il y a 50 enfants prématurés.

1. Déterminez l'intervalle de fluctuation asymptotique I associé à cette situation au seuil de 0,95.

2. Quelle est donc la conclusion ?

Méthode

1. Nous utilisons la formule donnant l'intervalle de fluctuation asymptotique au seuil de 0,95. Il convient de bien analyser l'énoncé pour savoir quelle est la valeur de n et celle de p.
Ne pas oublier de vérifier les conditions de validité : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

2. On calcule la fréquence de prématurés dans l'échantillon et on compare f à la borne supérieure de I.

On prend en compte la règle de décision choisie.

Solution

1. $I = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
Ici $p = 0{,}06$ et $n = 400$.
Les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont satisfaites.
En effet $n = 400$, $np = 24$ et $n(1-p) = 376$.
Donc $I = [0{,}037\,;\,0{,}083]$.

2. Notons f la proportion de prématurés dans l'échantillon : $f = \frac{50}{400} = 0{,}125$.
La borne supérieure de I est égale à 0,083.
f est supérieure à la borne supérieure de I.
Les chercheurs concluent que la proportion d'enfants prématurés est plus élevée chez les femmes ayant eu un travail pénible pendant la grossesse.

Mise en pratique

1 Une rhino-pharyngite guérit naturellement en moins de cinq jours dans 60 % des cas.
On veut tester un médicament censé abréger la durée de la maladie. Pour cela, on administre le médicament à 1 000 personnes.
Pour 63 % d'entre elles, la guérison a eu lieu en moins de cinq jours.

1. Déterminez l'intervalle de fluctuation asymptotique au seuil de 0,95 pour un échantillon de taille 1 000 associé à cette situation, après avoir justifié les conditions de validité.

2. Que peut-on penser de l'efficacité de ce médicament ?

OBJECTIF 2 Estimer une proportion inconnue à partir d'un échantillon

Notons f la fréquence observée dans un échantillon de taille n et p le pourcentage que l'on veut estimer. Alors l'intervalle $\left[f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}\right]$ contient p avec une probabilité d'au moins 0,95.
Cet intervalle est appelé intervalle de confiance de p au niveau de confiance 0,95.
On utilise cet intervalle dès que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

EXERCICE RÉSOLU B

Un candidat à une élection fait effectuer un sondage. Sur 100 personnes interrogées, 63 déclarent vouloir voter pour lui.
On suppose que les électeurs ne changent pas d'avis le jour du vote.
On notera p le pourcentage de voix obtenues par le candidat. On suppose que $0{,}5 \leqslant p \leqslant 0{,}7$.

1. Déterminez l'intervalle de confiance de p au niveau de confiance 0,95.

2. Énoncez le résultat ci-dessus en langage courant.

Méthode

1. On utilise la notion d'intervalle de confiance au niveau de confiance 0,95.
On vérifie que les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont bien satisfaites.

2. La borne inférieure de l'intervalle $[0{,}53\,;\,0{,}73]$ joue un rôle essentiel. Si cette borne inférieure est supérieure à 0,5, alors le candidat a de fortes chances d'être élu.

Solution

1. La fréquence obtenue à l'aide de l'échantillon est égale à 0,63. On sait que $0{,}5 \leqslant p \leqslant 0{,}7$ donc $-0{,}7 \leqslant -p \leqslant -0{,}5$ et $0{,}3 \leqslant 1 - p \leqslant 0{,}5$.
Ici $n = 100$ donc $50 \leqslant np \leqslant 70$ et $30 \leqslant n(1-p) \leqslant 50$.
Les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont satisfaites. Donc l'intervalle de confiance au niveau 0,95 est l'intervalle $\left[0{,}63 - \frac{1}{\sqrt{100}}\,;\, 0{,}63 + \frac{1}{\sqrt{100}}\right]$, c'est-à-dire $[0{,}53\,;\,0{,}73]$.

2. Il y a donc 95 chances sur 100 pour que cet intervalle contienne p. Or $0{,}53 > 0{,}5$.
Le candidat a donc au moins 95 chances sur 100 de gagner l'élection.

Mise en pratique

2 Lors d'une épidémie de grippe, 13 des élèves d'une classe de terminale ES qui compte 34 élèves ont contracté la maladie. On suppose que cette classe constitue un échantillon représentatif de l'ensemble des 850 élèves du lycée.

1. Donnez la fréquence d'élèves malades dans la classe.

2. a) Justifiez que les conditions pour utiliser un intervalle de confiance au niveau de confiance 0,95 sont satisfaites.

b) Donnez cet intervalle de confiance.

3. Emma, élève de la classe de terminale ES, estime qu'il est possible que plus de la moitié des élèves du lycée aient eu la grippe. Qu'en pensez-vous ?

3 Une urne contient des boules blanches et des boules noires. On effectue 100 tirages avec remise dans cette une et on obtient 39 boules noires.
Peut-on affirmer, avec une probabilité d'au moins 0,95, qu'il y a plus de boules blanches que de boules noires dans l'urne ?

EXERCICES

OBJECTIF 3 Déterminer une taille d'échantillon suffisante pour obtenir une estimation avec une précision donnée

Notons f la fréquence observée dans un échantillon de taille n, et p le pourcentage que l'on veut estimer.
Dès que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, l'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ contient p avec une probabilité d'au moins 0,95.

EXERCICE RÉSOLU C

On veut estimer la proportion p de foyers disposant en France d'un abonnement internet. On sait que p est compris entre 50 % et 70 %.

1. Quelle doit être la taille minimale de l'échantillon pour obtenir un résultat avec une précision de 3 % au seuil de 0,95 ?

2. Quelle doit être la taille minimale de l'échantillon pour obtenir un résultat avec une précision de 1 % au seuil de 0,95 ?

Méthode

1. On est dans le cas où p n'est pas connu.
On veut en trouver une estimation à partir d'un échantillon.
En écrivant la définition de l'intervalle de confiance, on est ramené à la résolution d'une inéquation.
La fonction $x \mapsto \frac{1}{x}$ est décroissante sur $]0 ; +\infty[$ donc $0 < a \leqslant b \Leftrightarrow \frac{1}{a} \geqslant \frac{1}{b} > 0$.

Solution

1. L'intervalle de confiance au seuil de 0,95 est l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ si les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont réalisées.
Ici $0{,}5 \leqslant p \leqslant 0{,}7$ donc $0{,}3 \leqslant 1 - p \leqslant 0{,}5$.
Donc si $n \geqslant 30$ alors $np \geqslant 15$ et $n(1-p) \geqslant 9$ et les trois conditions seront réalisées.
On veut trouver n tel que $\frac{1}{\sqrt{n}} \leqslant 0{,}03$.
On obtient $\sqrt{n} \geqslant \frac{1}{0{,}03}$, c'est-à-dire
$n \geqslant \frac{1}{(0{,}03)^2}$. Or $\frac{1}{(0{,}03)^2} \approx 1\,111{,}11$
donc $n \geqslant 1\,112$.
Les trois conditions sont donc réalisées.

Méthode

2. Dans la question 2, on obtient une valeur de n supérieure à celle de la question 1. Ce résultat était prévisible car on impose une plus grande précision.

Solution

2. On procède de façon analogue. Cette fois n doit vérifier l'inégalité $\sqrt{n} \geqslant \frac{1}{0{,}01}$
c'est-à-dire $n \geqslant \frac{1}{(0{,}01)^2}$. Donc $n \geqslant 10\,000$.
Les trois conditions sont bien réalisées.

Mise en pratique

4 La semaine précédant une élection qui oppose deux candidats A et B, on interroge 100 personnes pour connaître leur intention de vote. 45 personnes indiquent qu'elles vont voter pour le candidat A.

1. Justifiez que les conditions d'utilisation d'un intervalle de confiance au niveau de confiance 0,95 sont satisfaites.

2. Peut-on affirmer avec une probabilité de 0,95 que le candidat A ne sera pas élu ?

3. En supposant que le pourcentage d'intention de vote pour le candidat A reste égal à 45 %, quel nombre minimum de personnes faut-il interroger pour pouvoir affirmer avec une probabilité supérieure à 0,95 que le candidat A ne sera pas élu ?

Pour se tester

Exercices interactifs

EXERCICES

5 Questions sur le cours

Complétez comme il convient.
Dans les questions qui suivent, on suppose que $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$.

1. Pour n assez grand, si p désigne la proportion dans la population, l'intervalle de fluctuation asymptotique au seuil de 95 % est [...... , ...].

2. Si f désigne la fréquence observée sur un échantillon de taille n, la proportion p dans la population totale est élément de l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ avec un niveau de confiance de ....

6 Vrai ou faux

Les affirmations suivantes sont-elles vraies ou fausses ? Justifiez votre réponse.

1. Pour un échantillon de taille $n = 100$ extrait d'une population dans laquelle la proportion d'un caractère est égale à 0,5, l'intervalle de fluctuation asymptotique de la fréquence au niveau 0,95 est égal à [0,402 ; 0,598].

2. Si dans un échantillon de taille 100 la fréquence observée est égale à 0,4, alors la probabilité que la proportion p dans la population totale soit comprise entre 0,3 et 0,5 est au moins égale à 0,95.

3. Si l'on veut obtenir, au niveau de confiance 0,95, un intervalle de confiance d'étendue 0,1 alors il suffit d'étudier des échantillons de taille supérieure ou égale à 400.

7 QCM Une seule réponse exacte

Pour chaque affirmation, une seule réponse est exacte. Identifiez-la en justifiant votre réponse.

1. Dans une population, la proportion d'un caractère est égale à 0,4. Dans un échantillon aléatoire de taille 100 extrait de cette population, la fréquence de ce caractère appartient à l'intervalle [0,3 ; 0,5].
a) Toujours.
b) Jamais.
c) Avec une probabilité de 95 %.

2. Lorsque la taille de l'échantillon augmente, la longueur de l'intervalle de confiance au seuil de 95 % :
a) Diminue. **b)** Reste constante.
c) Augmente.

3. Dans un échantillon de taille $n \geqslant 30$, et pour une proportion p telle que $np \geqslant 5$ et $n(1-p) \geqslant 5$, la fréquence appartient à l'intervalle de fluctuation au seuil de 95 % :
a) Avec une probabilité inférieure à 0,05.
b) Avec une probabilité au moins égale à 0,95.
c) Toujours.

8 QCM Au moins une réponse exacte

Pour chaque affirmation, plusieurs réponses peuvent être exactes. Identifiez-les en justifiant votre réponse.

1. La longueur de l'intervalle de fluctuation asymptotique au seuil de 95 % :
a) dépend de la taille de l'échantillon ;
b) augmente quand la taille de l'échantillon augmente ;
c) diminue quand la taille de l'échantillon augmente.

2. Si la fréquence d'un caractère d'un échantillon de taille 100 est égale à 0,3, alors :
a) l'intervalle [0,2 ; 0,4] est un intervalle de confiance au niveau de confiance de 95 %.
b) Dans un échantillon de taille 200 la fréquence serait égale à 0,6.
c) Dans la population totale, la probabilité que la fréquence soit inférieure à 0,2 ou supérieure à 0,4 est inférieure à 0,05.

3. [0,55 ; 0,65] est l'intervalle de confiance au niveau de confiance 0,95 obtenu à partir d'un échantillon de taille n. On suppose que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$. Alors :
a) La taille de l'échantillon est égale à 400.
b) Si on multiplie la taille de l'échantillon par 4, alors la longueur de l'intervalle de confiance sera égale à 0,05.
c) Il y a moins de 5 % de chances que la fréquence dans la population totale soit inférieure à 0,55.

Voir les corrigés p. 361

Utiliser un algorithme

→ Pour étudier la fluctuation d'échantillonnage et les intervalles de confiance

9 Simulation d'échantillons

L'algorithme ci-dessous simule *k* échantillons de taille N d'un lancer d'une pièce bien équilibrée. Pour chaque échantillon, on calcule la fréquence d'apparition de « pile » et on dessine un nuage de points dont les ordonnées sont les différentes fréquences obtenues.

Algorithme	Commentaires
Entrées	
Saisir K	• Lecture du nombre K d'échantillons
Saisir N	• Lecture du nombre N de lancers par échantillon.
Traitement	
Pour J variant de 1 à K	• Début de la boucle extérieure, où J est le numéro de l'échantillon.
S prend la valeur 0	
Pour I variant de 1 à N	Traitement de l'échantillon n° J avec :
S prend la valeur : S + entier aléatoire 0 ou 1	– réalisation des N lancers, où S contient, à la sortie de la boucle, le nombre total de « pile » ;
FinPour	– fin de la boucle intérieure ;
F prend la valeur S/N	– calcul de la fréquence de l'échantillon N° J ;
Tracer point (J, F)	– affichage du point d'abscisse J et d'ordonnée la fréquence de l'échantillon n° J.
FinPour	• Fin de la boucle extérieure.

A Étude d'un cas particulier

Pour bien comprendre le fonctionnement de l'algorithme, prenons par exemple K = 3 et N = 2.

1. Combien y aura-t-il d'échantillons ?
Combien y aura-t-il de lancers dans un échantillon ?

2. Ci-dessous, on a écrit le contenu des mémoires I et J au fur et à mesure du déroulement de l'algorithme. Recopiez et complétez le tableau en faisant tourner l'algorithme à la main.

J	1	1	2	2	…
I	1	2	1	…	…

3. À la fin du traitement correspondant à J = 1, c'est-à-dire juste avant le deuxième « FinPour » :

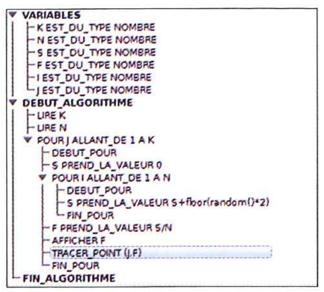
a) que va contenir F ?

b) quelles seront les coordonnées du point tracé ?

4. Répondez aux mêmes questions pour J = 2.

5. Lorsque l'algorithme se sera entièrement déroulé, combien y aura-t-il de points tracés ?

Algobox



TI

Casio

B Intervalle de fluctuation

1. Expliquez pourquoi l'intervalle de fluctuation asymptotique au seuil de 95 % est ici égal à :

$$\left[0{,}5 - \frac{1{,}96}{2\sqrt{N}} ; 0{,}5 + \frac{1{,}96}{2\sqrt{N}}\right].$$

2. Tapez sur votre ordinateur ou votre calculatrice le programme adapté.

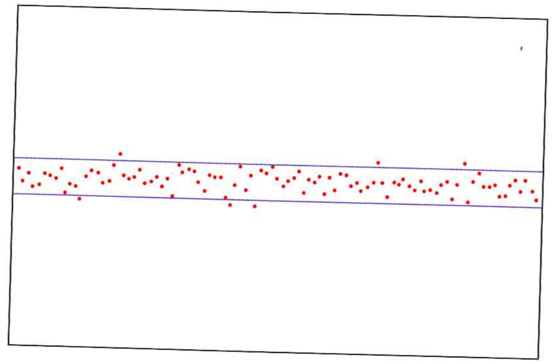
a) Faites tourner ce programme pour K = 100 et N = 10.

b) Constatez qu'avec des échantillons de taille aussi petite, les points sont dispersés.

Note N'oubliez pas de gérer les dimensions de la fenêtre d'affichage.

c) Pour mieux vous retrouver sur l'écran graphique, dessinez la droite d'équation $y = 0{,}5$.

3. a) Faites à présent tourner le programme pour K = 100 et N = 1 000.
Voici un exemple de ce que l'on peut obtenir :



b) Dessinez les droites d'équation $y = 0{,}5 - \frac{1{,}96}{2\sqrt{N}}$ et $y = 0{,}5 + \frac{1{,}96}{2\sqrt{N}}$ (en écrivant dans le programme, ou bien directement).

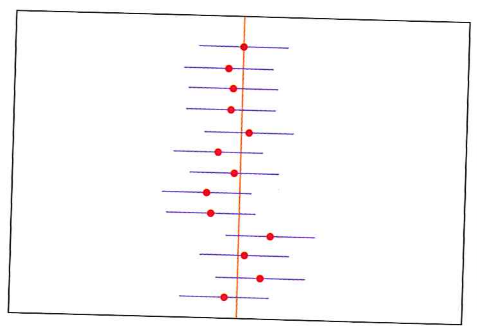
c) Comptez le nombre d'échantillons situés à l'extérieur de la bande horizontale limitée par les deux droites précédentes.

d) Recommencez une autre simulation, toujours avec K = 100 et N = 1 000.

e) D'après le résultat du cours sur l'intervalle de fluctuation au niveau 0,95 combien, environ, devrait-on observer de points extérieurs à la bande horizontale ?
Combien en comptez-vous sur l'écran ?

C Intervalle de confiance

L'écran ci-dessus est le résultat de la simulation de 13 échantillons de taille 100 pour une pièce équilibrée où l'on compte le nombre de « pile ».
Pour chaque échantillon, on a marqué la fréquence F obtenue avec un point, ainsi que l'intervalle de confiance au niveau 0,95 : $\left[F - \frac{1}{\sqrt{100}} ; F + \frac{1}{\sqrt{100}}\right]$, c'est-à-dire [F – 0,1 ; F + 0,1].



1. a) Modifiez l'algorithme étudié précédemment pour obtenir un tel résultat.

b) Traduisez cette modification sur votre ordinateur ou votre calculatrice.

2. a) Faites tourner le programme plusieurs fois et comptez à chaque fois le nombre d'intervalles de confiance au niveau 0,95 qui contiennent *p*, c'est-à-dire ici 0,5.

b) Quel est le résultat du cours illustré ici ?

Entraînement

DE TÊTE

10 Ordonnez les nombres suivants.

1. $a = \dfrac{1}{15}$; $b = \dfrac{1}{\sqrt{81}}$; $c = \dfrac{1}{\sqrt{400}}$.

2. $a = \dfrac{1}{\sqrt{900}}$; $b = \dfrac{1}{15}$; $c = \dfrac{\sqrt{100}}{100}$.

11 Écrivez sous la forme $[a\,;b]$ les intervalles suivants.

1. I intervalle fermé de centre 0,5 et de rayon $\dfrac{1}{\sqrt{100}}$.

2. K intervalle fermé de centre 0,4 et de rayon $\dfrac{1}{\sqrt{10\,000}}$.

12 On note $I = \left[0{,}15 - \dfrac{1}{\sqrt{400}}\,;0{,}15 + \dfrac{1}{\sqrt{400}}\right]$

et $J = \left[0{,}15 - \dfrac{1}{\sqrt{100}}\,;0{,}15 + \dfrac{1}{\sqrt{100}}\right]$.

A-t-on $I \subset J$ ou $J \subset I$?

FLUCTUATION D'ÉCHANTILLONNAGE

13 Donnez un exemple d'échantillon de taille 20 que l'on peut obtenir, ainsi qu'un exemple du caractère que l'on peut observer avec :

a) une pièce de monnaie ;

b) un dé en forme de tétraèdre, marqué 1, 2, 3, 4 ;

c) une urne contenant des boules bleues, blanches ou rouges ;

d) un jeu de 52 cartes.

14 Comment simuler un échantillon statistique de taille 50 dans un établissement scolaire de 1 252 élèves ?

15 La proportion d'un caractère dans une population est égale à p.
Peut-on utiliser le théorème 1 pour déterminer un intervalle de fluctuation asymptotique de taille n au seuil de 95 % dans les cas suivants :

a) $n = 36$; $p = 0{,}5$;

b) $n = 100$; $p = 0{,}04$;

c) $n = 50$; $p = 0{,}95$;

d) $n = 400$; $p = 0{,}05$.

16 La proportion d'un caractère dans une population est égale à 0,1.
Quelle taille minimale devra avoir l'échantillon utilisé pour calculer la fréquence, afin de répondre aux conditions d'application du théorème sur l'intervalle de fluctuation asymptotique au seuil de 95 % ?

17 La proportion d'un caractère dans une population est égale à 0,92.
Quelle taille minimale devra avoir l'échantillon utilisé pour calculer la fréquence, afin de répondre aux conditions d'application du théorème sur l'intervalle de fluctuation asymptotique au seuil de 95 % ?

18 La proportion p d'un caractère dans une population est égale à 0,65.
Déterminez l'intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence d'un échantillon :

a) de taille 100 ;

b) de taille 1 000.

19 Une urne opaque contient 60 % de boules rouges. On effectue 100 tirages avec remise.
On note f la fréquence des boules rouges tirées.
Au seuil de 95 %, à quel intervalle devrait appartenir f ?

20 La proportion p d'un caractère dans une population est égale à 0,42.
Dans un échantillon issu de la population, on a trouvé 0,49 comme fréquence de ce caractère.
Cet échantillon est-il « représentatif » de la population pour ce caractère, sachant que l'échantillon est :

a) de taille 100 ?

b) de taille 1 000 ?

Indication
On pourra utiliser la notion d'intervalle de fluctuation asymptotique au seuil de 95 %.

21 Un candidat à un poste de député a été élu avec 51 % des voix. Dans un village de sa circonscription, après dépouillement, sur 1 253 votes, 49 % étaient en sa faveur.
Ce village est-il « représentatif » de la couleur politique de la circonscription ?

Indication
On pourra utiliser la notion d'intervalle de fluctuation asymptotique au seuil de 95 %.

22 Beaucoup de garçons
Au cours de l'année 1961, dans un village d'Asie, il est né 52 enfants, parmi lesquels 37 garçons.

1. Quel est le pourcentage de garçons nés cette année-là dans ce village ?

2. Intrigués par ce résultat, on se propose d'effectuer une simulation où l'on prendra pour modèle l'équiprobabilité de naissance des garçons et des filles, et où l'on pourra observer les fluctuations dues au hasard.

a) Simulez, dans la ligne 1 du tableur, 52 naissances. Le 0 représentera une fille et le 1 représentera un garçon.

b) Dans la cellule BA1, tapez =SOMME(A1:AZ1)/52. Expliquez ce que calcule cette formule.

c) Tirez sur la ligne A pour obtenir 100 échantillons de 52 naissances et les fréquences correspondantes.

3. a) En utilisant le résultat du cours, dites quel est l'intervalle de fluctuation asymptotique au seuil de 95 % pour la fréquence d'un échantillon de taille 52.

BB1 =SI(ET(BA1>=0,362;BA1<=0,638);0;1)

	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB
1	0	1	0	1	1	0	0	1	0	0	1	1	0	0,462	0
2	0	0	0	1	0	0	1	0	0	1	0	0	1	0,442	0
3	0	0	1	1	0	1	1	0	1	0	0	0	1	0,385	0
4	0	1	0	0	0	0	1	0	0	1	1	1	1	0,442	0
5	0	1	0	0	0	1	0	1	1	1	0	1	1	0,519	0
6	0	1	0	1	1	0	0	1	1	0	0	0	0	0,596	0
7	1	1	0	0	0	0	0	0	1	0	0	1	1	0,423	0
8	1	1	1	1	0	1	0	1	0	0	1	0	1	0,462	0
9	0	0	0	0	1	0	1	0	1	0	0	0	1	0,481	0
10	0	1	0	1	0	1	0	0	0	1	1	1	1	0,385	0

b) En utilisant la formule écrite dans la capture d'écran ci-dessus, faites apparaître un « 0 » lorsque la fréquence appartient à l'intervalle de fluctuation asymptotique et un « 1 » sinon. Affichez le total de la colonne correspondante.
Que représente le nombre obtenu ?
Utilisez la touche F9 pour obtenir d'autres résultats.

4. Quelle conclusion pouvez-vous émettre sur les naissances dans ce village ?

23 Pratique du sport
Le maire d'une grande ville affirme que, d'après ses informations, 55 % des habitants majeurs pratiquent un sport.
On interroge 100 habitants majeurs au hasard et 43 % disent pratiquer un sport. On suppose que l'affirmation du maire est vraie et que la population de la ville est suffisamment grande pour considérer que les 100 « tirages » sont effectués avec remise.

1. Précisez l'intervalle de fluctuation asymptotique de la fréquence au seuil 0,95 adapté à cette situation.

2. Comment peut-on interpréter l'affirmation du maire ?

24 Accès à internet
Un fournisseur d'accès à Internet (FAI) affirme que, sur sa hotline, seuls 20 % des clients attendent plus de 5 minutes pour obtenir un interlocuteur.
Une association de consommateurs, ayant reçu de nombreuses doléances de la part de ses adhérents, décide de faire une enquête et interroge au hasard 200 personnes ayant eu à s'adresser à la hotline de ce FAI. 53 d'entre elles ont dû attendre plus de 5 minutes.
On se demande si ces résultats permettent de mettre en doute l'affirmation de ce fournisseur d'accès à Internet.

1. Quel est le pourcentage de personnes qui ont attendu plus de 5 minutes dans cet échantillon ?

2. Précisez l'intervalle de fluctuation asymptotique de la fréquence au seuil de 95 % adapté à cette situation.

3. Que concluez-vous ?

25 Au ski
Une station de ski familiale n'attire que 25 % de skieurs habitant hors du département. Souhaitant élargir sa clientèle, la station fait réaliser des travaux au cours de l'été suivant : nouveau télésiège débrayable à six places, canons à neige.
L'hiver suivant, 500 skieurs sont interrogés : 172 d'entre eux habitent hors du département.

1. Donnez l'intervalle de fluctuation asymptotique au seuil de 95 % adapté à cette situation.

2. Peut-on affirmer que les travaux de l'été ont eu un impact sur la fréquentation des skieurs habitant hors du département ?

26 Trafic routier
Les habitants d'un village traversé par un axe routier à fort trafic réclament au Conseil général de leur département la construction d'une déviation. Leur argument : « Les camions sont plus bruyants et plus polluants que les autres véhicules. Or, chez nous, 80 % des véhicules traversant notre village sont des camions ».
Le Conseil général, étonné par un tel pourcentage, fait procéder à un comptage : sur 1 000 véhicules circulant sur cet axe et pris au hasard, 70 % sont des camions.
On se demande si, au niveau 0,95, le « chiffre » avancé par les habitants du village est exact.
On suppose vrai le pourcentage avancé par les habitants du village.
On note F la variable aléatoire donnant la fréquence des camions sur un échantillon de 1 000 véhicules traversant le village.

1. Quel est alors l'intervalle de fluctuation asymptotique au seuil de 95 % associé à F ?

2. La fréquence obtenue lors du comptage appartient-elle à cet intervalle ? Concluez.

27 Prise de décision

Cet exercice est extrait du document d'accompagnement du programme.

On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13 %.

Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux-ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.

Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.

La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 95 %, alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.

1. Déterminez l'intervalle de fluctuation asymptotique au seuil de 95 % de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100.

2. L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?

3. Le médecin n'est pas convaincu par cette conclusion et déclare que le nombre de personnes interrogées était insuffisant pour mettre en évidence qu'il y avait plus de jeunes ayant eu des crises d'asthme que dans le reste du département.

Combien faudrait-il prendre de sujets pour qu'une proportion observée de 19 % soit en dehors de l'intervalle de fluctuation asymptotique ?

INTERVALLES DE CONFIANCE

Pour les exercices de cette rubrique : conformément au document d'accompagnement du programme, lorsque l'énoncé ne donne **pas d'encadrement de p**, pour un échantillon de taille $n \geqslant 30$, on peut **remplacer les conditions $np \geqslant 5$ et $n(1 - p) \geqslant 5$ par les conditions $nf \geqslant 5$ et $n(1 - f) \geqslant 5$.**

Pour les exercices 28 à 32

Dans un échantillon de taille n, on a trouvé une fréquence égale à f. Indiquez si les conditions usuelles d'utilisation de l'intervalle de confiance au niveau de confiance 0,95 sont remplies et donnez cet intervalle.

28 $f = 0{,}71$; $n = 25$.

29 $f = 0{,}57$; $n = 100$.

30 $f = 0{,}15$; $n = 36$.

31 $f = 0{,}88$; $n = 40$.

32 $f = 0{,}51$; $n = 200$.

33 Pour évaluer une proportion p, on s'appuie sur les fréquences f_1 et f_2 observées sur deux échantillons de taille 100. On trouve $f_1 = 0{,}71$ et $f_2 = 0{,}63$.
Les intervalles de confiance associés sont-ils disjoints ?

34 Pour évaluer une proportion p, on s'appuie sur les fréquences f_1 et f_2 observées sur deux échantillons de taille 900. On trouve $f_1 = 0{,}32$ et $f_2 = 0{,}39$.
Les intervalles de confiance associés sont-ils disjoints ?

35 Lors d'un sondage, l'intervalle de confiance au niveau 0,95 a pour longueur 0,08.
Quelle était la taille de l'échantillon ?

36 **1.** Dans un échantillon de taille 400, la fréquence d'un caractère est de 0,62.

a) Déterminez l'intervalle de confiance I_1 au seuil de 95 % de la proportion de ce caractère dans la population tout entière.

b) Dessinez sur une droite graduée cet intervalle.

2. a) Reprenez les questions précédentes avec un échantillon de taille 1 000, qui donnerait la même fréquence, à savoir 0,62.
On notera I_2 l'intervalle de confiance obtenu.

b) Quelle relation peut-on écrire entre I_1 et I_2 ?

37 Une urne contient des boules blanches et des boules noires. On aimerait connaître la proportion p des boules blanches. Pour cela, on effectue 100 tirages avec remise dans cette urne. On obtient 32 boules blanches. Estimez p à l'aide de l'intervalle de confiance au niveau 0,95.

38 Dans une population, on note p la proportion d'individus aux yeux bleus. Dans un échantillon de 400 personnes choisies au hasard dans cette population on a trouvé une fréquence égale à 0,3 pour les yeux bleus.

Donnez un intervalle de confiance au niveau de confiance 0,95 pour la proportion p.

39 Dans un échantillon de 100 personnes, 67 % possèdent un ordinateur. Donnez une estimation de la proportion p de personnes de la population entière possédant un ordinateur, à l'aide de l'intervalle de confiance au niveau 0,95.

40 Devenir Président

Lors du second tour des élections présidentielles, un institut de sondage a interrogé 1 000 personnes pour connaître leurs intentions de vote.

La candidat A a été crédité de 49 % des voix, et la candidat B de 51 % des voix.

On suppose que les personnes interrogées ne vont pas changer d'avis le jour du vote.

1. Déterminez l'intervalle de confiance au niveau 0,95 :

a) du pourcentage de voix du candidat A ;

b) du pourcentage de voix du candidat B.

2. Sur un axe gradué, dessinez ces deux intervalles avec des couleurs différentes.

3. Après avoir pris connaissance des résultats de ce sondage, le candidat B peut-il être sûr d'être élu président de la République ?

41 Audimat

L'audience d'une chaîne de télévision pour les informations de 20 heures de lundi dernier a été mesurée sur 1 500 appareils : 28 % de l'ensemble des téléspectateurs ont regardé le journal télévisé de cette chaîne.

Déterminez l'intervalle de confiance au seuil de 95 % de l'audience de la chaîne lors de ce journal de 20 heures.

42 Assurances

Dans un groupe d'assurances, on s'intéresse aux sinistres susceptibles de survenir, une année donnée, aux véhicules de la flotte d'une importante entreprise de maintenance de chauffage collectif.

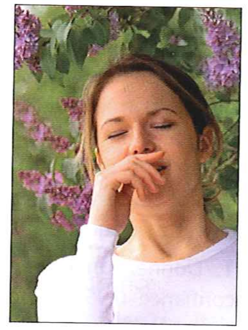
On considère un échantillon de 100 véhicules prélevés au hasard dans le parc de véhicules mis en service depuis 6 mois. On constate que 91 véhicules n'ont pas eu de sinistre.

1. Donnez le pourcentage f de véhicules de cet échantillon qui n'ont pas eu de sinistre 6 mois après leur mise en service.

2. p est la proportion inconnue de véhicules du parc qui n'ont pas eu de sinistre 6 mois après leur mise en service. Déterminez un intervalle de confiance de p avec le niveau de confiance 95 %.

3. On considère l'affirmation suivante :
« la proportion p est obligatoirement dans l'intervalle de confiance obtenu à la question **b.** »
Est-elle vraie ? Justifiez votre réponse.

43 Allergies printanières



On se propose de comparer la sensibilité aux allergies printanières entre la population urbaine et la population rurale d'une région.

Dans un échantillon de 320 personnes de la population urbaine, on a trouvé 121 personnes allergiques. Dans un échantillon de 296 personnes de la population rurale, on a trouvé 126 personnes allergiques.

1. Donnez, pour chacune des deux populations, un intervalle de confiance au niveau de confiance 0,95 pour le pourcentage de personnes allergiques.

2. Peut-on, au niveau de confiance 0,95, déduire des résultats des deux échantillons que la population rurale est plus sensible que la population urbaine ?

3. En supposant que dans les deux échantillons la fréquence observée de personnes allergiques reste la même, quelle même taille minimum d'échantillon aurait-il fallu étudier, dans chacune des deux populations, pour pouvoir conclure à une différence de sensibilité entre la population urbaine et la population rurale, avec un niveau de confiance de 0,95 ?

44 Dans deux régions

Dans un échantillon E_1 de 225 personnes de la population d'une région R_1, on a trouvé 30 personnes atteintes d'une maladie M. On note p_1 la proportion des personnes de cette région atteintes de cette maladie.

Dans un échantillon E_2 de 100 personnes de la population d'une région R_2, on a trouvé 25 personnes atteintes de la maladie M. On note p_2 la proportion de personnes de cette région atteintes de cette maladie.

1. a) Donnez le pourcentage de personnes atteintes de la maladie M dans l'échantillon E_1.

b) Donnez un intervalle de confiance au niveau 0,95 de la proportion p_1.

2. a) Donnez le pourcentage de personnes atteintes de la maladie M dans l'échantillon E_2.

b) Donnez un intervalle de confiance au niveau 0,95 de la proportion p_2.

3. a) Peut-on conclure de cette étude qu'il y a plus de personnes atteintes par la maladie M dans la région R_2 que dans la région R_1 ?

b) En supposant que le pourcentage de personnes atteintes reste inchangé dans l'échantillon E_2, quelle aurait dû être la taille minimum de l'échantillon E_2 pour pouvoir conclure à une différence entre les deux régions avec un niveau de confiance de 95 % ?

45 Les gauchers

L'étude de 122 squelettes d'adultes de la population de Wharram Percy, un village isolé d'Angleterre au Moyen Âge, a permis de constater que 16 % des personnes étaient gauchères. On considère que cette population permet d'obtenir une image représentative du pourcentage réel de vrais gauchers dans une population de cette époque.

1. Donnez un intervalle de confiance au niveau de confiance 0,95 du pourcentage de gauchers dans la population de Wharram Percy.

2. En Angleterre, de nos jours, la proportion de gauchers se situe entre 10 et 13 %.
Peut-on, au niveau de confiance 0,95, dire que de nos jours la population de gauchers est moins élevée qu'au Moyen Âge ?

3. En supposant que les pourcentages soient restés inchangés, quelle aurait dû être la taille minimum de squelettes étudiés pour pouvoir conclure, au niveau de confiance 0,95, que le nombre de gauchers est moins élevé de nos jours qu'au Moyen Âge ?

Les exercices 46 à 48 sont extraits du document d'accompagnement du programme.

46 Dépistage de la bronchiolite

Dans le but d'évaluer la prise en charge de la bronchiolite du nourrisson dans un hôpital de la région Aquitaine, une étude rétrospective a été mise en place.

1. Il est recommandé de coucher l'enfant de manière très inclinée (couchage en proclive) dans le cadre de la prise en charge de la bronchiolite. On évalue cette pratique à partir d'un échantillon de 134 dossiers. 106 des enfants ont été couchés en proclive.
Déterminez un intervalle de confiance au niveau de confiance de 95 % de la proportion d'enfants dont le couchage respecte la recommandation.

2. Une étude plus fine permet de comparer les pratiques entre les différents services ayant admis des enfants atteints de bronchiolite :

Couchage proclive	En service des urgences	En service hospitalier	Total
Oui	45	52	97
Non	29	8	37
Total	74	60	134

a) Déterminez un intervalle de confiance au seuil de 95 % de la proportion de couchage en proclive pour chaque type de service.

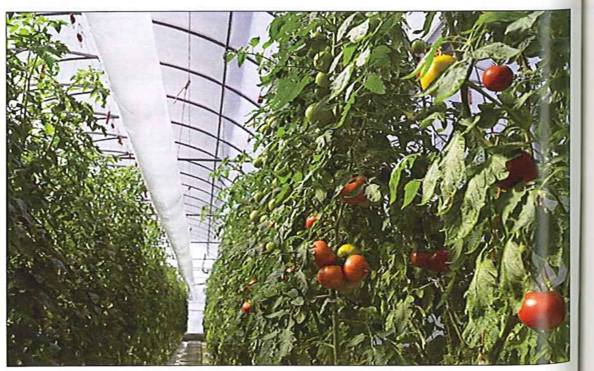
b) Peut-on conclure selon vous au seuil de 95 % que la pratique de couchage n'est pas identique selon le service ?

47 Taux de germination

Un maraîcher achète un lot de semences de tomates pour produire ses plants de tomate. Il lui reste des semences de l'année passée, dont il doit contrôler le taux de germination pour pouvoir les utiliser avec les autres. En effet, des taux de germination trop différents provoquent des trous dans les plates bandes de production, ce qui génère un coup de manutention plus élevé. Il faut donc comparer les taux de germination des semences des deux années.
Une stratégie consiste à calculer et à comparer les intervalles de confiance des taux de germination des plants de l'année et de l'année précédente. Si les deux intervalles sont disjoints, on peut conclure à une différence de taux de germination entre les semences des deux origines. Il faudra alors les semer séparément.
Pour faire cette comparaison, le maraîcher prélève, aléatoirement dans les semences de l'année, un échantillon de 200 graines qu'il met à germer. Il constate que 185 graines germent.
Il prélève ensuite, aléatoirement dans les semences de l'année précédente, un échantillon de 200 graines qu'il met à germer. Il constate que 150 graines germent.

1. Déterminez un intervalle de confiance, au niveau de confiance de 95 %, du taux de germination p_a du lot de semences de l'année.

2. Déterminez (par la même méthode qu'à la question 1.) un intervalle de confiance de niveau 95 %, du taux de germination p_b du lot de semences de l'année précédente.



3. Concluez.

48 Diagnostic de la jaunisse

Un test de diagnostic rapide effectué sur des sujets ictériques (coloration jaune de la peau et des muqueuses) doit permettre d'estimer si l'ictère est d'origine virale ou non, sans avoir besoin de faire des analyses longues et compliquées. Cependant il est important de pouvoir s'assurer que ce test est de bonne qualité, c'est-à-dire qu'il doit pouvoir indiquer correctement si l'ictère est viral ou non. Il doit être capable d'identifier correctement le type d'ictère : il doit être positif chez les sujets dont l'ictère est viral et négatif sinon.
Une étude est effectuée sur 100 personnes ayant un ictère viral et 100 personnes ayant un ictère d'origine non virale.
Les résultats obtenus sont présentés dans le tableau ci-dessous :

	Hépatite virale	Ictère d'origine non virale
Test positif	85	20
Test négatif	15	80

1. a) Déterminez le pourcentage de sujets ayant un test positif parmi ceux ayant un ictère viral.

b) Déterminez un intervalle de confiance à 95 % de la proportion de tests positifs lorsque l'ictère est viral. Cette proportion est appelée *sensibilité du test diagnostic*, c'est-à-dire la probabilité qu'une personne ayant un ictère viral réagisse au test. Un test diagnostic sera d'autant meilleur que la sensibilité est importante.

2. a) Déterminez le pourcentage de sujets ayant un test négatif parmi celles ayant un ictère non viral.

b) Déterminez un intervalle de confiance à 95 % de la proportion de tests négatifs lorsque l'ictère est non viral. Cette proportion est appelée *spécificité du test diagnostic*, c'est-à-dire la probabilité qu'une personne ayant un ictère non viral ne réagisse pas au test. Un test diagnostic sera d'autant meilleur que la spécificité est importante.

49 ALGORITHMIQUE Longueur au hasard

On choisit au hasard deux points A et B sur un segment de longueur 1.
On veut déterminer la probabilité p que la longueur AB soit supérieure ou égale à 0,5.
Faute de savoir calculer directement p, nous allons simuler cette expérience un grand nombre de fois.

1. Si A a pour abscisse a dans le repère $(O ; \overrightarrow{OI})$ et si B a pour abscisse b, quelle est la distance AB ?

2. Voici un algorithme pour simuler n expériences :

```
Entrée
Saisir n
Traitement
S prend la valeur 0
Pour i variant de 1 à n
  a prend la valeur nombre aléatoire de
  [0:1[
  b prend la valeur nombre aléatoire de
  [0;1[
  ℓ prend la valeur √(a−b)²
  Si ℓ… alors S prend la valeur…
  FinSi
FinPour
f prend la valeur S/n
Sortie
Afficher f
```

a) Dans la boucle « pour », que contient la variable ℓ ?

b) Complétez les parties manquantes de l'algorithme.

c) Que contiennent les variables S et f à la sortie de la boucle « pour » ?

3. a) Traduisez l'algorithme précédent en un programme adapté à votre ordinateur ou à votre calculatrice.

b) Faites tourner ce programme et notez les résultats obtenus pour :
- $n = 100$;
- $n = 1000$;
- $n = 10\,000$ (si votre équipement le permet !).

4. a) À l'aide des résultats précédents, donnez une estimation de p en utilisant l'intervalle de confiance au niveau 0,95.

b) Si l'on veut une estimation de p au niveau 0,95 avec un intervalle d'amplitude 2×10^{-2}, quelle valeur de n faut-il prendre ? **Note** On peut montrer que p est égal à 0,25.

POUR LA LOGIQUE

Pour les exercices 50 à 52
On veut estimer une proportion p d'un caractère en calculant une fréquence f de ce caractère dans un échantillon de taille n. On suppose que $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$. Dites si la propriété indiquée est vraie ou fausse et justifiez votre réponse.

50 $\exists x \in \mathbb{N}$ tel que $\left[f-\frac{1}{\sqrt{x}} ; f+\frac{1}{\sqrt{x}}\right]$ contient p avec une probabilité au moins égale à 95 %.

51 $\exists x \in \mathbb{N}$ tel que $\left[f-\frac{1}{\sqrt{x}} ; f+\frac{2}{\sqrt{x}}\right]$ contient p avec une probabilité au moins égale à 95 %.

52 $x \in \,]0 ; n[\, \Rightarrow \left[f-\frac{1}{\sqrt{x}} ; f+\frac{1}{\sqrt{x}}\right]$ contient p avec une probabilité au moins égale à 95 %.

Un peu d'histoire

Chercheurs d'hier

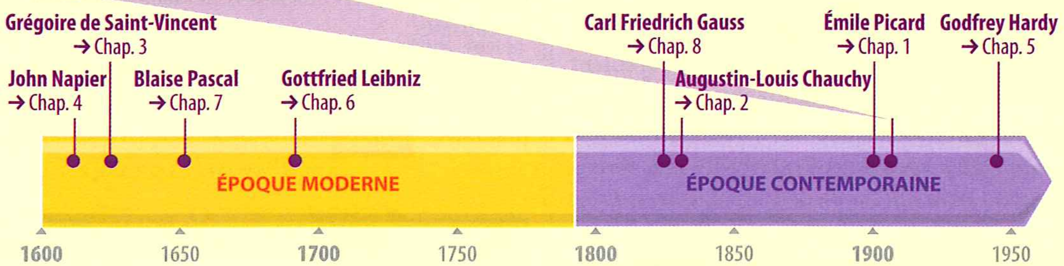
Pearson Karl et l'étude des échantillons

Karl Pearson est un mathématicien anglais, célèbre pour ses travaux en statistique. Il fait ses études à Oxford. A 27 ans, il obtient la chaire de mathématiques appliquées et de mécanique au University College de Londres. En 1901, il fonde le laboratoire de biométrie et la revue *Biometrica* dont il est le directeur jusqu'à sa mort.


Pearson Karl
1857 - 1936

Poursuivant l'œuvre de Galton, Pearson est l'un des fondateurs de la statistique moderne. Il consacre une partie de ses recherches à l'étude des échantillons ; c'est lui qui introduit les coefficients qui portent désormais son nom. Il propose des applications de ses travaux à la théorie de l'évolution.

Karl Pearson est également connu pour ses activités littéraires et politiques. Parmi les publications de Pearson citons *The Grammar of Science* (1892), *Mathematical Contribution to the Theory of Evolution* (recueil de dix-huit articles publiés de 1893 à 1912) et, dans le domaine politique et littéraire, *The new Werther* (1880) et *The Trinity: A nineteenth century passion-play* (1882).



À la même époque

Au Royaume-Uni

En 1887, la reine Victoria fête son jubilé d'or, célébrant cinquante années d'un règne sous lequel l'empire britannique est devenu la première puissance mondiale.

La reine Victoria
(1819 - 1901)


À la même époque

En sciences

Les travaux d'Einstein dans le domaine de la mécanique quantique et de la relativité restreinte (1905) offrent de nouveaux fondements à la physique moderne.

Albert Einstein
(1879 - 1955)


Sur le Web
http://serge.mehl.free.fr/chrono/Pearson.html

Accompagnement personnalisé

Soutien

53 À l'aide d'un échantillon de taille 100, on souhaite estimer un pourcentage p. On suppose que $p \in [0{,}2\,;\,0{,}8]$. On suppose que f, la fréquence observée, est égale à 0,45.
On se propose d'analyser cette situation.

➡ *Il convient avant tout de bien comprendre en quoi consiste le problème. Supposons par exemple qu'une urne contienne un certain pourcentage de boules blanches et que l'on ne connaisse pas ce pourcentage que l'on note usuellement p. Il s'agit alors de « localiser » p. Pour cela, on effectue cent tirages avec remise et on constate que la fréquence d'apparition d'une boule blanche est égale à 0,45.*
On se doute qu'il est probable que p soit « voisin » de 0,45. On va voir qu'il est possible de préciser cette notion de voisinage.

1. Dans l'expérience de l'urne ci-dessus combien de boules blanches sont apparues dans l'échantillon considéré ?

2. À votre avis, intuitivement aurions-nous eu plus d'informations si on avait obtenu la même fréquence, 0,45 à partir d'un échantillon de 1 000 tirages ?

➡ *Dans cette question, on demande une réponse intuitive. On verra ci-dessous que l'on dispose de résultats précis à ce sujet.*

3. Déterminez l'intervalle de confiance I de p au niveau 0,95.

➡ ***Indication.*** $I = \left[f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}\right]$, *où n est le nombre de tirages et f la fréquence observée. Ici n = 100 et f = 0,45.*

4. Un théorème fondamental

➡ *Le théorème suivant montre l'intérêt de la notion d'intervalle de confiance au niveau 0,95 : si $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, alors cet intervalle de confiance contient p avec une probabilité d'au moins 0,95, c'est-à-dire qu'il y a au moins 95 chances sur 100 que cet intervalle contienne p.*

Vérifiez que les conditions d'application de ce théorème sont satisfaites. Concluez.

5. a) Supposons à présent que l'on obtienne $f = 0{,}45$ à l'aide d'un échantillon de taille 10 000.
Dans ce cas, effectuez une étude analogue à celle de la question **3.**

b) Expliquez pourquoi le résultat obtenu est en accord avec la question **2.**

Approfondissement

54 APPRENDRE À CHERCHER

On se propose de démontrer que le théorème concernant la fluctuation d'échantillonnage énoncé cette année (théorème 1 p. 244) est plus précis que celui vu en seconde.

➡ *En classe de seconde, l'intervalle de fluctuation au seuil de 95 % est l'intervalle* $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$.

Cette année l'intervalle de fluctuation asymptotique au seuil de 95 % est l'intervalle :

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

1. Expliquez pourquoi le problème revient à comparer les deux intervalles précédents, tous deux centrés en p. Plus précisément, dire que le théorème vu cette année est plus précis que celui de seconde revient à dire que :

$I_n \subset \left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$. Pourquoi ?

2. a) Étudiez les variations de la fonction f définie par $f(x) = x \mapsto x(1-x)$.

b) Déduisez-en que :

$$I_n \subset \left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right].$$

55 Une urne contient des boules rouges et des boules noires. On voudrait estimer le pourcentage de boules noires.
Pour cela on effectue n tirages consécutifs avec remise.

1. a) Quelle doit être la valeur minimale de n pour obtenir un résultat avec une précision de 10 % au seuil de 0,95 ?

b) Quelle doit être la valeur minimale de n pour obtenir un résultat avec une précision de 5 % au seuil de 0,95 ?

2. On a effectué 100 tirages et on a obtenu 55 boules noires.

a) Justifiez que les conditions d'utilisation d'un intervalle de confiance au niveau 0,95 sont satisfaites.

b) Peut-on affirmer avec une probabilité supérieure à 95 % que l'urne contient plus de boules noires que de boules rouges ?

Le jour du BAC

56 Exercice commenté

→ **Rédigez la solution de cet exercice en vous aidant des conseils.**

On effectue un sondage auprès de 1 000 habitants d'une ville de 1 000 000 d'habitants, pour savoir s'ils voteront oui ou non à un référendum.

1. Expliquez pourquoi un tel sous-ensemble de 1 000 habitants peut être considéré comme un échantillon de taille 1 000, c'est-à-dire comme le résultat de 1 000 répétitions indépendantes de la même expérience.

2. On suppose que les personnes interrogées ne changeront pas d'avis le jour du vote. 54 % des personnes interrogées répondent « oui » lors du sondage.

Comment peut-on interpréter ce résultat, en utilisant l'intervalle de confiance au seuil de 0,95 associé à ce sondage ?

Analyser l'énoncé

Il s'agit d'un exercice classique sur les sondages.

Analyser l'énoncé

En termes de probabilité, on cherche quelle conclusion on peut tirer du résultat du sondage.

Conseils

1. *A priori* les expériences ne sont pas indépendantes puisqu'une même personne ne peut pas être interrogée deux fois. Il convient de préciser que la taille de la population, de 1 000 000, est très grande par rapport à la taille du sous-ensemble interrogé (1 000) ; on peut donc considérer que les expériences sont indépendantes.

2. • L'intervalle de confiance au seuil de 0,95 associé à ce sondage est l'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$; ici $f = 0{,}54$ et $n = 1000$.

• L'intervalle obtenu est [0,508 ; 0,572]. On peut donc dire qu'il y a au moins 95 % de chances pour que le « oui » l'emporte le jour du vote.

Remarque

Des tirages sans remise dans une urne peuvent permettre de modéliser ce type de situation : dans une urne de N boules, quand on en tire *n* sans remise, la composition de l'urne, après chaque tirage, n'est pratiquement pas modifiée si *n* est très petit devant N.
On peut donc considérer que les tirages s'effectuent avec remise.

Remarque

Ce qui importe ici, c'est la position de la borne inférieure de l'intervalle de confiance, ici 0,508, par rapport à 0,5.
Si la borne inférieure de l'intervalle de confiance au seuil de 0,95 avait été inférieure à 0,5, on n'aurait pas pu conclure de la même manière.

Voir les corrigés p. 361

57 A. Loi normale

Une entreprise fabrique des chaudières.

Soit X la variable aléatoire qui, à chaque chaudière prélevée au hasard dans la production, associe sa durée de fonctionnement en années.

On admet que X suit la loi normale de moyenne 15 et d'écart-type 3.

Une chaudière est dite « amortie » si sa durée de fonctionnement est supérieure ou égale à 10 ans.

Calculez la probabilité qu'une chaudière prélevée au hasard dans la production soit « amortie » ; arrondir à 10^{-3}.

B. Intervalle de confiance

On considère un échantillon de 100 chaudières prélevées au hasard dans un stock important.

Ce stock est assez important pour qu'on puisse assimiler ce tirage à un tirage avec remise.

On constate que 94 chaudières sont sans aucun défaut.

1. Donnez la fréquence f des chaudières de cet échantillon qui sont sans aucun défaut.

2. p est la proportion inconnue des chaudières du stock qui sont sans aucun défaut.

Déterminez un intervalle de confiance de la proportion p avec le niveau de confiance 95 %. Arrondir les bornes à 10^{-2}.

3. On considère l'affirmation suivante : « la proportion p est obligatoirement dans l'intervalle de confiance obtenu à la question **2** ». Est-elle vraie ?

58

Un maraîcher a constaté, sur plusieurs années, que 75 % des pieds de tomates qu'il n'a pas traités sont plus ou moins gravement atteints par le mildiou. Il décide de tester deux traitements biologiques. À la fin de la saison, il a relevé sur trois échantillons de 100 pieds les résultats suivants :

Échantillon	① Non traités	② Traitement A	③ Traitement B
Nombre de pieds atteints	81	66	54

1. Donnez, pour chacune des trois catégories de pieds de tomate, un intervalle de confiance au niveau 0,95 pour la proportion de pieds atteints par le mildiou.

2. Donnez un intervalle de fluctuation asymptotique au seuil de 95 % pour la fréquence des pieds malades dans un échantillon de pieds non traités de taille 100.

Les résultats sur l'échantillon ① sont-ils compatibles avec les observations des années précédentes ?

3. Quelles remarques sur l'efficacité des traitements peut-on faire à partir de l'observation de ces échantillons ?

59

Le pourcentage de personnes de groupe sanguin O et A en France et en Autriche est donné dans le tableau ci-dessous :

	O	A
France	43 %	45 %
Autriche	38 %	41 %

1. On choisit au hasard un échantillon de 100 personnes en France. Donnez un intervalle de fluctuation asymptotique au niveau 0,95 pour la fréquence des catégories « groupe O » et « groupe A ».

2. On choisit au hasard un échantillon de 100 personnes en Autriche. Donnez un intervalle de fluctuation asymptotique au niveau 0,95 pour la fréquence des catégories « groupe O » et « groupe A ».

3. Un laboratoire a reçu, en provenance de France ou d'Autriche, un échantillon de 100 doses de sang pour tester la présence d'un virus. Pour en déterminer la provenance, qui a été perdue (France ou Autriche), on étudie le groupe sanguin et on trouve 45 % de groupe O.

a) Peut-on conclure au niveau de confiance 0,95 que cet échantillon a été prélevé en France ?

b) Quelle aurait dû être la taille minimum de l'échantillon étudié, en supposant que la fréquence de groupe O reste à 45 %, pour pouvoir conclure au niveau de confiance 0,95 que le prélèvement venait de France ?

4. Dans l'échantillon reçu il y a 21 doses qui contiennent le virus.

a) Donnez au niveau de confiance 0,95 un intervalle de confiance du taux de personnes dont le sang contient le virus dans le pays concerné.

b) Les responsables sanitaires pensaient avant cette étude qu'un quart de la population était porteuse de ce virus. Leur prévision était-elle en contradiction avec l'étude de l'échantillon ?

60

Cet exercice est un questionnaire à choix multiple. Pour chaque question, une seule réponse est exacte.

1. On lance 400 fois une pièce bien équilibrée. La fréquence d'apparition de FACE :

a) est égale à 0,5 ;

b) est comprise entre 0,45 et 0,55 avec une probabilité supérieure ou égale à 95 % ;

c) est comprise entre 0,45 et 0,55.

2. Dans un échantillon de taille $n = 200$ extrait d'une population dans laquelle la proportion d'un caractère est égale à 0,4, l'intervalle de fluctuation asymptotique de la fréquence au niveau 0,95 est égal à :

a) [0,329 ; 0,471] ; **b)** [0,332 ; 0,468] ; **c)** [0,3 ; 0,5].