Chapitre 18 : Loi normale et estimation.

I-Rappels de seconde

1. Echantillons

Lorsqu'on travaille sur une population de grande taille, il est rarement possible d'avoir accès aux données relatives à l'ensemble de la population.

On utilise alors un échantillon de cette population.

Définition Un échantillon de taille n est une sélection de n individus choisis "au hasard" dans une population.

Exemple
On étudie la répartition mâle/femelle d'une population de truites peuplant une rivière.
Il est pratiquement impossible de recenser toutes les truites de la rivière. On décidera donc de travailler sur un échantillon en prélevant, par exemple, 100 truites.
La taille de l'échantillon doit être suffisamment élevée pour fournir des résultats fiables (mais pas trop pour ne pas entrainer un surcroit de travail important !)

2. Intervalle de fluctuation et intervalle de confiance

Si l'on effectue plusieurs échantillonnage de même taille sur une même population, on obtiendra en général des fréquences légèrement différentes pour un caractère donné.
Voici, par exemple, les résultats que l'on pourrait obtenir en prélevant 5 échantillons de 100 truites :

Echantillons	échantillon n°1	échantillon n°2	échantillon n°3	échantillon n°4	échantillon n°5
Pourcentage de truites femelles	52%	55%	42%	50%	48%

Ce phénomène s'appelle la fluctuation d'échantillonnage.

Le résultat suivant précise cette notion :

Théorème définition On note p la proportion d'un caractère dans une population donnée.
On prélève un échantillon de taille n de cette population et on note f la fréquence du caractère dans l'échantillon.
Si $0{,}2 \leq p \leq 0{,}8$ et si $n \geq 25$ alors, dans au moins 95% des cas, f appartient à l'intervalle : $\boxed{I = \left[p - \frac{1}{\sqrt{(n)}} ; p + \frac{1}{\sqrt{(n)}}\right]}$.
I est appelé l'intervalle de fluctuation au seuil 95%

Remarques

- On applique le théorème ci-dessus si on connaît la proportion p du caractère dans la population.
- Bien retenir la signification de chacune des variables :
  - p = proportion du caractère dans l'ensemble de la population
  - f = fréquence du caractère dans l'échantillon
  - n = taille de l'échantillon
- Au niveau Seconde, les intervalles de fluctuation seront toujours demandés au seuil de 95%.
  Ce seuil a été choisi car :
  - il conduit à une formule assez simple
  - on peut considérer comme "raisonnablement fiable" un résultat validé dans 95% des cas
- En terminale, on verra un intervalle de fluctuation plus précis !

Autrement dit :
Si l'on fait des simulations et que la probabilité d'un résultat est p, 95 % des résultats trouvés sont dans l'intervalle $\boxed{I = \left[p - \frac{1}{\sqrt{(n)}} ; p + \frac{1}{\sqrt{(n)}}\right]}$. Cet intervalle s'appelle l'intervalle de fluctuation au seuil 95%.

Réciproquement :
Si on fait une simulation pour estimer une probabilité p, on aura 95% de chances que p se trouve dans l'intervalle $\boxed{I = \left[f - \frac{1}{\sqrt{(n)}} ; f + \frac{1}{\sqrt{(n)}}\right]}$ où f est le résultat de la simulation sur un échantillon de dimension n. Cet intervalle s'appelle l'intervalle de confiance au seuil de 95%.

Cet intervalle de confiance sera réutilisé en terminale.

Exemple 1 : Intervalle de fluctuation

Supposons que notre rivière contienne 50% de truites femelles (et donc 50% de mâles...). Pour nos échantillons de taille 100, $n=100\geq25$; par ailleurs $p=0,5 \in [0,2 ; 0,8]$ Donc l'intervalle de fluctuation au seuil de 95% sera :

I=

Exemple 2 : Intervalle de _ _ _ _ _ _ _ _.

Supposons qu'on lance un dé bien équilibré, et qu'on compte le nombre de résultats supérieurs à 4. Cet événement aura 2 chances sur 6, donc 1 sur 3 de se produire, soit une probabilité $p = 1/3 \approx 33\%$. Si l'on fait un millier d'échantillons de 100 lancers aléatoires, l'intervalle de fluctuation sera I=_ _ _ _ _ _ _ _ _=_ _ _ _ _ _ _ _
donc 95% des résultats seront entre _ _ _% et _ _ _%.

Exemple 3 : Intervalle de _ _ _ _ _ _ _ _.

Supposons qu'un sondage sur 100 personnes donne 35 voix pour le candidat A, quel sera l'intervalle de _ _ _ _ _ _ _ _ _ _ correspondant ?

I=

Même question si on interroge 1 000 personnes.

Exercice 1

Deux entreprises recrutent leur personnel dans un bassin d'emploi où il y a autant d'hommes que de femmes. Dans l'entreprise A il y a 100 employés dont 43 femmes. Dans l'entreprise B, il y a 2500 employés dont 1150 femmes (soit 46%). Il semble que l'entreprise B respecte mieux la parité que l'entreprise A.

Il s'agit en réalité de savoir si les entreprises sont composées de personnes choisies au hasard dans la population sans favoriser un sexe au détriment de l'autre. Pour le savoir, on considère que chaque entreprise est un échantillon de la population.

1. Quelle est la proportion p de femmes dans la population ?
2. Quelles sont les tailles de ces échantillons ?
3. Quels sont leurs intervalles de fluctuation ?
4. Expliquer pourquoi ces intervalles permettent de dire que l'entreprise A respecte mieux la parité que l'entreprise B.

Exercice 2

Une société fabrique des écrans plasma. En moyenne, 21% des écrans sont défectueux. Lors d'un contrôle d'un lot de 40 écrans, 14 sont défectueux.

1. Calculer la proportion d'écrans défectueux dans ce lot.
2. Faut-il s'en inquiéter ?
3. Afin de tester une nouvelle machine, on contrôle 400 écrans au hasard. 28% d'entre eux sont défectueux.

   Ce résultat semble satisfaisant. Qu'en pensez-vous ?

Exercice 3

Une rhino-pharyngite guérit naturellement en moins de 5 jours dans 60% des cas. On veut tester un médicament censé abréger la durée de la maladie. Pour cela, on administre le médicament à 1000 personnes. Pour 63% d'entre elles, la guérison a eu lieu en moins de 5 jours. Que penser de l'efficacité de ce médicament ?

Exercice 4

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

Exercice 5

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.