ECHANTILLONNAGE EN PREMIÈRE ES ET L

DÉFINITION

Pour réaliser un **échantillon** de taille n dans une population, on sélectionne "au hasard" n individus avec ou sans remise.

REMARQUE

"Avec remise" signifie qu'un même individu peut être sélectionné plusieurs fois

PROPRIÉTÉ

Dans un échantillon avec remise, la variable aléatoire X qui compte le nombre d'individus présentant un certain caractère suit une loi binomiale $\mathscr{B}(n; p)$ où n est la taille de l'échantillon et p la proportion du caractère étudié dans la population.

REMARQUE

Si l'échantillon est réalisé *sans remise* on peut encore considérer que la propriété précédente est vraie si la taille de l'échantillon est nettement inférieure à l'effectif total.

EXEMPLE

La proportion des Français ayant des yeux bleus est estimée à 25%. On choisit 100 personnes au hasard dans la population française. La variable aléatoire comptant le nombre de personnes aux yeux bleus dans un tel échantillon suit une loi binomiale $\mathscr{B}(100; 0,25)$

DÉFINITION

L'intervalle de fluctuation à 95 % d'une fréquence correspondant à la réalisation, sur un échantillon de taille n, d'une variable aléatoire X de loi binomiale $\mathscr{B}(n; p)$, est l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$, défini par :

- a est le plus petit entier tel que $p(X \leqslant a) > 0,025$;
- b est le plus petit entier tel que $p(X \leqslant b) \geqslant 0,975$.

REMARQUE

Pour $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, on observe que l'intervalle trouvé est sensiblement le même que l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ vu en classe de Seconde.

EXEMPLE

On lance 100 fois une pièce équilibrée. La variable aléatoire correspondant au nombre de *Pile* obtenus suit la loi binomiale $\mathscr{B}(100;0.5)$. En consultant une table de loi binomiale ou à l'aide d'un tableur ou d'une calculatrice on trouve $a=40$ et $b=60$ donc l'intervalle

$I=\left[\frac{40}{100};\frac{60}{100}\right]=[0,4;0,6]$.

La formule $I=\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]=\left[0,5-\frac{1}{\sqrt{100}};0,5+\frac{1}{\sqrt{100}}\right]$ donne le même résultat.

VALIDATION D'UNE HYPOTHÈSE

On fait l'hypothèse qu'un caractère est présent dans une proportion p de la population.

Pour valider cette hypothèse, on prélève un échantillon de taille n et on note f la fréquence du caractère étudié dans cet échantillon. **La règle de décision** est alors la suivante : si la fréquence observée f appartient à l'intervalle de fluctuation à 95 %, on **accepte** l'hypothèse selon laquelle la proportion est p dans la population avec un risque d'erreur inférieur à 5 % ; sinon, on **rejette** cette hypothèse (avec un risque d'erreur inférieur à 5 %).

INTERVALLE DE FLUCTUATION POUR UN SEUIL DIFFÉRENT DE 95 %

On peut généraliser le résultat précédent pour un seuil différent de 95% : **L'intervalle de fluctuation à t %** est alors l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$, défini par :

- a est le plus petit entier tel que $p(X\leqslant a)>\frac{1}{2}\left(1-\frac{t}{100}\right)$;
- b est le plus petit entier tel que $p(X\leqslant b)\geqslant\frac{1}{2}\left(1+\frac{t}{100}\right)$.

EXEMPLE

Une variable aléatoire X suit une loi binomiale de paramètres $n=500$ et $p=0,30$.

Pour trouver l'intervalle de fluctuation au seuil de 99% on recherche les valeurs de a et b telles que :

- a est le plus petit entier vérifiant $p(X\leqslant a)>0,005$;
- b est le plus petit entier vérifiant $p(X\leqslant b)\geqslant 0,995$.

La table de loi binomiale $\mathscr{B}(500;0,30)$ (ou une calculatrice) donne au seuil de 99% : $a=124$ et $b=177$ donc l'intervalle de fluctuation à 99 % est $I=\left[\frac{124}{500};\frac{177}{500}\right]$