Échantillonnage - Estimation

Échantillonnage - Estimation

Règle générale

Échantillonnage - Estimation

Règle générale

On utilise un **intervalle de fluctuation** lorsque la proportion p dans la population est connue ou si l'on fait une hypothèse sur sa valeur (prise de décision à partir d'un échantillon).

Échantillonnage - Estimation

Règle générale

On utilise un **intervalle de fluctuation** lorsque la proportion p dans la population est connue ou si l'on fait une hypothèse sur sa valeur (prise de décision à partir d'un échantillon).

On utilise un **intervalle de confiance** lorsque l'on veut estimer une proportion inconnue p dans une population à partir de la fréquence f observée dans un échantillon (estimation, par exemple dans le cadre d'un sondage).

Échantillonnage - Estimation

Règle générale

On utilise un **intervalle de fluctuation** lorsque la proportion p dans la population est connue ou si l'on fait une hypothèse sur sa valeur (prise de décision à partir d'un échantillon).

On utilise un **intervalle de confiance** lorsque l'on veut estimer une proportion inconnue p dans une population à partir de la fréquence f observée dans un échantillon (estimation, par exemple dans le cadre d'un sondage).

À retenir

Intervalle de fluctuation d'une **fréquence**
Intervalle de confiance d'une **proportion**

Échantillonnage - Estimation

La notion d'intervalle de fluctuation est un "fil rouge" des programmes de lycée.

Intervalle de fluctuation, intervalle de confiance dans les programmes (résumé) :

	Interv. de fluctuation	*Interv. de confiance*
Seconde	$\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$	Sensibilisation
Première	Avec la loi binomiale	xxx
Terminale	$\left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$	$\left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right]$

Échantillonnage

Dans le sens commun (sondages par exemple), un **échantillon** est un sous-ensemble obtenu par prélèvement aléatoire dans une population.
Dans le programme (seconde) :
un **échantillon de taille** n est constitué des résultats de n répétitions indépendantes de la même expérience.

Échantillonnage

Dans le sens commun (sondages par exemple), un **échantillon** est un sous-ensemble obtenu par prélèvement aléatoire dans une population.
Dans le programme (seconde) :
un **échantillon de taille** n est constitué des résultats de n répétitions indépendantes de la même expérience.

La fréquence f des individus possédant le caractère dans l'échantillon varie d'un échantillon à l'autre : c'est la **fluctuation d'échantillonnage**.
Les fluctuations diminuent lorsque la taille des échantillons augmente.

Échantillonnage

Définition (programme de seconde)
Intervalle de fluctuation au seuil de 95 % : centré autour de p (proportion du caractère dans la population), contient, avec une probabilité (au moins) égale à 0,95, la fréquence observée dans un échantillon de taille n.

Échantillonnage

Définition (programme de seconde)
Intervalle de fluctuation au seuil de 95 % : centré autour de p (proportion du caractère dans la population), contient, avec une probabilité (au moins) égale à 0,95, la fréquence observée dans un échantillon de taille n.

Il existe, pour un même seuil, **plusieurs intervalles de fluctuation**.
On peut vérifier que, pour une même valeur de p, ces intervalles sont de plus en plus proches lorsque n augmente.

Échantillonnage

En fonction de l'appartenance ou non de f à l'intervalle de fluctuation à 0,95 que l'on a déterminé, on prend **une décision** concernant la conformité de l'échantillon :

Échantillonnage

En fonction de l'appartenance ou non de f à l'intervalle de fluctuation à 0,95 que l'on a déterminé, on prend **une décision** concernant la conformité de l'échantillon :

- si f n'appartient pas à l'intervalle, **on rejette, au risque d'erreur de 5 %**, l'hypothèse que l'échantillon est compatible avec le modèle ;

Échantillonnage

En fonction de l'appartenance ou non de f à l'intervalle de fluctuation à 0,95 que l'on a déterminé, on prend **une décision** concernant la conformité de l'échantillon :

- si f n'appartient pas à l'intervalle, **on rejette, au risque d'erreur de 5 %**, l'hypothèse que l'échantillon est compatible avec le modèle ;
- dans le cas contraire, **on ne peut pas rejeter** l'hypothèse.

(*d.r. 2nde pages 14 à 18*)

Échantillonnage - Intervalle vu en seconde

D'après le théorème de Moivre-Laplace (approximation par la loi normale), environ 95 % des échantillons de taille n fournissent une fréquence f appartenant à l'intervalle

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

Échantillonnage - Intervalle vu en seconde

D'après le théorème de Moivre-Laplace (approximation par la loi normale), environ 95 % des échantillons de taille n fournissent une fréquence f appartenant à l'intervalle

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

En seconde, on majore $1,96\sqrt{p(1-p)}$ par 1 ; l'intervalle $\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$ contient l'intervalle précédent.

(*d.r. collège page 24*)

Échantillonnage - Loi binomiale (première)

Avec la notion de variable aléatoire et la loi binomiale, il n'est plus nécessaire d'approximer par la loi normale.

L'**intervalle de fluctuation à 95 % d'une fréquence** correspondant à la réalisation, sur un échantillon aléatoire, d'une variable aléatoire X de loi $\mathcal{B}(n\,;\,p)$ est l'intervalle

$$\left[\frac{a}{n}\,;\,\frac{b}{n}\right]$$

défini par :

Échantillonnage - Loi binomiale (première)

Avec la notion de variable aléatoire et la loi binomiale, il n'est plus nécessaire d'approximer par la loi normale.

L'**intervalle de fluctuation à 95 % d'une fréquence** correspondant à la réalisation, sur un échantillon aléatoire, d'une variable aléatoire X de loi $\mathcal{B}(n\,;\,p)$ 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 $P(X \leqslant b) \geqslant 0,975$.

Échantillonnage - Terminale

F_n = variable aléatoire qui, à tout échantillon de taille n, associe la fréquence d'apparition du caractère.

$\left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ = **intervalle de fluctuation asymptotique** au seuil $1-\alpha$ de F_n.

Il contient F_n avec une probabilité d'autant plus proche de $1-\alpha$ que n est grand.

En terminale ES/L, STI2D, STL, STMG :
$\alpha = 0,05$; $1-\alpha = 0,95$; $u_\alpha = 1,96$. (seuil 95 %).

Estimation

Estimation d'une proportion inconnue p grâce à un échantillon aléatoire

Estimation

Estimation d'une proportion inconnue p grâce à un échantillon aléatoire
On peut faire une **estimation ponctuelle** en posant $p = f$.
Cette estimation varie d'un échantillon à l'autre du fait de la fluctuation d'échantillonnage.

Estimation

Estimation d'une proportion inconnue p grâce à un échantillon aléatoire
On peut faire une **estimation ponctuelle** en posant $p = f$.
Cette estimation varie d'un échantillon à l'autre du fait de la fluctuation d'échantillonnage.

Mieux : On cherche un **intervalle de confiance de la proportion** p (c'est-à-dire un intervalle contenant « très vraisemblablement » p) à partir de la fréquence f mesurée dans un échantillon de taille n.

(*d.r. 3ème p.25, 2nde p.18, terminale p.30 ...*)

Estimation

Si $n \geqslant 30$ et si $nf \geqslant 5$ et $n(1-f) \geqslant 5$, un **intervalle de confiance** de p au niveau de confiance 0,95 est $\left[f - \frac{1}{\sqrt{n}}\,;\, f - \frac{1}{\sqrt{n}}\right]$.

Parmi tous les échantillons de taille n possibles, 95 % des intervalles associés $\left[f - \frac{1}{\sqrt{n}}\,;\, f - \frac{1}{\sqrt{n}}\right]$ contiennent p.

Une fois l'échantillon tiré, l'intervalle de confiance associé est entièrement fixé, il n'y a plus d'aléatoire à ce stade.
Il est donc incorrect de dire que p a une probabilité 0,95 d'appartenir à cet intervalle (p est inconnu mais pas aléatoire).

(*document ressource 2nde, pages 18 et 19*)