

Comment choisir N et n_1 ?

Quelques notations :

	Les billes dans le biberon	Echantillon déterminé par la classe
Effectif total	$N = 5$ (ou on peut prendre une plus grande valeur)	n = nombre de tirages attendus = 50, 100, 150, 200
Nombre de billes noires	n_1 : suivant la composition choisie	X = nombre de billes noires comptées
Fréquence des noires	$p = \frac{n_1}{N}$	$x = \frac{X}{n}$

On note α la probabilité correspondant au risque d'erreur choisi : $\alpha = 0,01$ pour un risque d'erreur de 1%, $\alpha = 0,05$ pour un risque d'erreur de 5%. Et oui, si les élèves n'ont pas de "chance", ils pourraient avoir un échantillon qui ferait partie des 1% (ou 5%) les plus "rares".

Les variables aléatoires :

Voir une bille parmi 5 et noter sa couleur est une expérience de Bernoulli. La couleur noire de la bille pouvant correspondre au succès et la couleur blanche à l'échec.

La probabilité p d'obtenir une bille noire est ce que les élèves cherchent et que le professeur connaît a priori.

Si l'on répète n fois de manière indépendante une expérience de Bernoulli, le nombre de succès au cours de ces n épreuves est une variable aléatoire X qui suit la loi binomiale (de paramètres n et p). Son espérance est np et son écart type est $\sqrt{np(1-p)}$.

La fréquence $F = \frac{X}{n}$ peut être approchée par la loi normale $N(p, \frac{\sigma}{\sqrt{n}})$ avec $\sigma = \sqrt{p(1-p)}$.

Donc :

x appartient à l'intervalle $[p - 1,96 \frac{\sigma}{\sqrt{n}} ; p + 1,96 \frac{\sigma}{\sqrt{n}}]$ avec un risque de 5%.

x appartient à l'intervalle $[p - 2,58 \frac{\sigma}{\sqrt{n}} ; p + 2,58 \frac{\sigma}{\sqrt{n}}]$ avec un risque de 1%.

Voir la partie "Un peu de théorie".

Les N et n_1 possibles :

Combien les élèves doivent-ils regarder de billes pour que l'intervalle de confiance ne contienne qu'une seule valeur de p possible?

Pour que les élèves aient la possibilité de trouver p (la valeur de x trouvée doit être proche de la valeur p), il faut que l'intervalle n'ait pas une amplitude trop grande :

Les valeurs p possibles sont $0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}, 1$. Pour que l'intervalle de confiance ne contienne qu'une valeur possible de p , il faut donc que l'amplitude de l'intervalle de confiance soit strictement inférieure à $\frac{1}{N}$.

On pose $t = 1,96$ si le risque choisi est $\alpha = 5\%$, $t = 2,58$ si le risque choisi est $\alpha = 1\%$.

Il faut donc que $2t \frac{\sigma}{\sqrt{n}}$ (l'amplitude de l'intervalle) soit strictement inférieure à $\frac{1}{N}$.

Donc il faut $N < \frac{\sqrt{n}}{2t\sigma}$. (avec n : le nombre d'expériences faites par les élèves et N : le nombre de billes)

Cela revient à exprimer une condition sur n : $n > 4t^2 N^2 p(1-p)$ car $\sigma = \sqrt{p(1-p)}$

Si on appelle k le nombre de billes noires, alors $p = \frac{k}{N}$ et la condition devient $n > 4t^2 k(N-k)$

Le maximum de $k(N-k)$ étant $\frac{N^2}{4}$, en prenant $n > t^2 N^2$, on est sûr d'avoir une seule valeur possible de p dans l'intervalle de confiance.

Si $N = 5$, il faut $n > 96$ (5% d'erreur) ou $n > 166$ (1% d'erreur)

Si $N = 6$, il faut $n > 138$ (5% d'erreur) ou $n > 239$ (1% d'erreur)

Si $N = 7$, il faut $n > 199$ (5% d'erreur) ou $n > 327$ (1% d'erreur)

Si $N = 8$, il faut $n > 246$ (5% d'erreur) ou $n > 427$ (1% d'erreur)