II- Estimation d'une proportion (sert à estimer une proportion inconnue)

1) Intervalle de confiance

Pour des raisons de coût et de faisabilité, on ne peut pas toujours étudier toute la population. (C'est par exemple le cas lorsque l'on réalise un sondage).

On chercher à « estimer » la proportion inconnue p d'individus ayant une certaine propriété. (Par exemple la proportion de client satisfait par un service).

Pour cela, on sélectionne aléatoirement un échantillon de taille n de cette population et on calcule la fréquence f des individus qui ont cette propriété.

Enfin, on pourra en déduire que la proportion inconnue p appartient à l'intervalle $\left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}}\right]$ avec un risque d'erreur de 5%.

Définition : L'intervalle $\left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}}\right]$ est appelé un « intervalle de confiance de la proportion inconnue p avec un niveau de confiance 0,95.

Conditions : $n \geq 30$, $n \times f \geq 5$ et $n \times (1-f) \geq 5$.

2) Taille minimale de l'échantillon pour avoir une précision donnée

Avec un niveau de confiance de 0,95 l'amplitude de l'intervalle de confiance est de : $\frac{2}{\sqrt{n}}$ (écart entre les 2 bornes).

Ainsi, plus n est grand et plus l'intervalle devient précis (les bornes se resserrent).

On peut vous demandez de déterminer la valeur de n à partir de laquelle la longueur de l'intervalle devient plus petit qu'une certaine valeur « a » .

Il faudra alors résoudre :

$$\frac{2}{\sqrt{n}} \leq a$$

$\div 2 > 0$

$$\frac{1}{\sqrt{n}} \leq \frac{a}{2}$$

on inverse
⚠ fonction ↘ sur $]0;+\infty[$

$$\sqrt{n} \geq \frac{2}{a}$$

on met au carré
fonction ↗ sur $[0;+\infty[$

$$n \geq \left(\frac{2}{a}\right)^2$$

$$n \geq \frac{4}{a^2}$$

Exemple : avec $a = 0,08$

$$\frac{2}{\sqrt{n}} \leq 0,08$$

$$\frac{1}{\sqrt{n}} \leq 0,04$$

$$\sqrt{n} \geq 25$$

$$n \geq 625$$

45 Une clinique a proposé une nouvelle opération chirurgicale et a connu 40 échecs sur 200 tentatives.
On note p le pourcentage de réussite de cette nouvelle opération.

1. Déterminer un intervalle de confiance pour p au seuil de confiance 0,95.

2. Combien d'opérations la clinique devrait-elle réaliser pour connaître le pourcentage de réussite avec une précision de plus ou moins 1 %, au seuil de confiance 0,95 ?

1) D'après l'échantillon, on a : $n = 200$ et $f = \frac{160}{200} = 0{,}8$ (Attention on cherche à estimer le pourcentage p de réussite)

Conditions : $n = 200 \geq 30$, $n \times f = 160 \geq 5$ et $n \times (1 - f) = 40 \geq 5$

L'intervalle de confiance pour p au seuil de confiance de 95% est :

$$\left[0{,}8 - \frac{1}{\sqrt{200}}; 0{,}8 + \frac{1}{\sqrt{200}}\right] \approx [0{,}729; 0{,}871]$$

Soit un pourcentage de réussite compris entre 72,9% et 87,1% au seuil de confiance 0,95 d'après cet échantillon.

2) Il faut résoudre l'inéquation :

$$\frac{2}{\sqrt{n}} \leq 0{,}01$$

$$\frac{1}{\sqrt{n}} \leq 0{,}005$$

$$\sqrt{n} \geq 200$$

$$n \geq 40\,000$$

C'est à partir de 40 000 opérations réalisées que l'on connaîtrait le pourcentage de réussite avec une précision de plus ou moins 1%.

49 L'équipe de France joue la finale de la Coupe du Monde.

1. On interroge un groupe d'amis sur un réseau social : au bout d'une journée, on obtient 250 réponses dont 30 ont confiance dans la victoire des Bleus. Donner un intervalle de confiance à 95 % de la proportion précédente.

2. On décide d'attendre quelques jours avant de parier afin d'avoir plus de réponses. Combien de réponses doit-on attendre afin d'obtenir un résultat à 3 % près, au seuil de 0,95 ?

1) D'après l'échantillon, on a : $n = 250$ et $f = \frac{30}{250} = 0{,}12$

Conditions : $n = 250 \geq 30$, $n \times f = 30 \geq 5$ et $n \times (1 - f) = 220 \geq 5$

L'intervalle de confiance ce cette proportion au seuil de confiance de 95% est :

$$\left[0{,}12 - \frac{1}{\sqrt{250}} ; 0{,}12 + \frac{1}{\sqrt{250}}\right] \approx [0{,}056; 0{,}184]$$

2) Il faut résoudre l'inéquation :

$$\frac{2}{\sqrt{n}} \leq 0{,}03$$

$$\frac{1}{\sqrt{n}} \leq 0{,}015$$

$$\sqrt{n} \geq \frac{200}{3}$$

$$n \geq \frac{40\,000}{9} \approx 4\,444{,}4$$

C'est à partir de $4\,445$ de réponses obtenues que l'on connaîtrait ce pourcentage avec une précision de plus ou moins 3%.