

Première partie

Étude de quelques définitions, pour comprendre le lexique de l'échantillonnage



Définitions. (Source : Wikipedia)

1. L'*échantillonnage* [...] désigne les méthodes de sélection d'un sous-ensemble d'individus (un échantillon) à l'intérieur d'une population pour estimer les caractéristiques de l'ensemble de la population.
2. Un *échantillon* est un ensemble d'individus représentatifs d'une population.



Question 1. Selon vous, quels pourraient être les avantages de l'échantillonnage lorsque l'on réalise un sondage ?



Question 2. Plaçons-nous dans le cadre d'un sondage qui aurait lieu à la veille d'une élection politique d'une grande importance. Quels critères proposeriez-vous pour choisir des individus représentatifs de la population française ?

Deuxième partie

Quelques mises en situation pratiques : sondages, résultats et interprétations







Situation 1. En dessous de cette publicité, il est indiqué *Scorage clinique, 53 femmes.*



Question 1. Que penser du choix de l'échantillon ?



Situation 2. (*Source : Wikipedia*) Emmanuel Macron ayant été élu pour un second mandat au second tour de l'élection présidentielle de 2022, la Constitution lui interdit de renouveler son mandat une deuxième fois consécutive. La question de la succession au président de la République se pose alors au sein de la majorité présidentielle, ce qui amène les sondeurs à conduire des enquêtes d'opinions sur le sujet. Le tableau ci-dessous ne comprend que les sondages les plus récents pour chaque institut.

Sondeur	Date	Échantillon						
				Attal (REN)	Bayrou (MoDem)	Darmanin (REN)	Le Maire (REN)	Philippe (HOR)
Ifop [archive]	3-4 avril 2024	1 028	Ensemble des Français	63 %	31 %	49 %	50 %	55 %
			Partisans de la majorité	92 %	54 %	68 %	75 %	87 %
Odoxa [archive]	14-15 février 2024	1 005	Ensemble des Français	44 %	19 %	23 %	30 %	49 %
			Partisans de la majorité	82 %	23 %	39 %	70 %	86 %



Question 2. D'un sondage à l'autre, les résultats sont tout de même assez différents. Comment expliquez-vous ces différences ?



Situation 3. Un site de commerce en ligne a lancé une enquête auprès de 300 clients, choisis au hasard parmi ceux ayant effectué des achats sur leur site. Le temps de livraison a été jugé « satisfaisant » par 160 des personnes interrogées.



Question 3. Calculer la fréquence f de clients ayant jugé « satisfaisant » le temps de livraison dans cet échantillon. Donner un intervalle de confiance au seuil de 95% de la proportion p de clients de ce site Internet satisfaits par le temps de livraison.



Situation 4. En 1976, dans un comté du Texas, Rodrigo Partida était condamné à huit ans de prison, après une suspicion de cambriolage dans une résidence privée du Texas. Il attaqua ce jugement, au motif que la désignation des jurés de ce comté était discriminante à l'égard des américains d'origine mexicaine. En effet, 79% de la population de ce comté est d'origine mexicaine, et sur les 870 personnes convoquées pour être jurés, seules 339 d'entre elles étaient d'origine mexicaine.



Question 4. Peut-on dire que la constitution des jurys est faite de façon aléatoire ?

Troisième partie

Retour aux mathématiques



Pour un intervalle I donné, on définit sa *longueur* $\ell(I)$ comme l'écart entre ses deux bornes, peu importe l'orientation des crochets. Si l'une des bornes est infinie, on conviendra que sa longueur est égale à $+\infty$. Par exemple, la longueur de l'intervalle $I =]2 ; 7]$ est égale à $\ell(I) = 7 - 2 = 5$.



Question 1. Déterminer la longueur des intervalles $I = [0 ; 1]$, $J =]-2 ; 1]$ et $K = [0 ; +\infty[$.



On rappelle que si un sondage produit une estimation f pour la présence d'un caractère dans un échantillon, alors il y a (au moins) 95% de chances que la proportion p de présence d'un caractère dans la population totale appartienne à l'intervalle de confiance :

$$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$$



Question 2. Déterminer la longueur de l'intervalle de confiance.



Question 3. Commenter la phrase suivante : « Quand le nombre n de personnes sondées augmente, alors l'intervalle de confiance se resserre autour de f ». Que peut-on en déduire lorsqu'un sondage est réalisé ?



On souhaiterait « contrôler l'erreur d'estimation », c'est-à-dire faire en sorte que l'intervalle de confiance soit le plus resserré possible. Imaginons par exemple que l'on souhaite avoir une erreur qui soit au maximum égale à 10%. Alors il faudrait que $\frac{2}{\sqrt{n}} \leq 0,1$, c'est-à-dire que $2 \leq 0,1\sqrt{n}$, autrement dit $\sqrt{n} \geq \frac{2}{0,1}$, ce qui donne $\sqrt{n} \geq 20$, ou encore $n \geq 400$. Il faudrait interroger 400 personnes pour que l'erreur soit au maximum égale à 10%.



Question 4. En vous inspirant de l'exemple ci-dessus, résoudre une résolution pour déterminer quel serait le nombre minimum de personnes à interroger pour avoir :

1. Une erreur maximale de 5% ?
2. Une erreur maximale de 2% ?
3. Une erreur maximale de 1% ?
4. Une erreur maximale de 0,5% ?