CORRIGE des exercices sur les intervalles de confiance

Exercice 1

Le parti d'un candidat commande un sondage réalisé à partir de 1 600 personnes à l'issue duquel il est donné gagnant avec 52% des voix. A-t-il des raisons d'être confiant ?

CORRIGE :

On a bien : $n \geq 25$

On peut répondre à la question si on montre p > 0,5 avec une grande probabilité. Le problème est que p est inconnu . . .

On remarque alors que : $p-\frac{1}{\sqrt{1600}} \leq f \leq p+\frac{1}{\sqrt{1600}} \quad \Leftrightarrow \quad f-\frac{1}{\sqrt{1600}} \leq p \leq f+\frac{1}{\sqrt{1600}}$

→ **c'est l'intervalle de confiance au risque de 5% :** $\left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}} \right]$

$\Leftrightarrow \quad 0,52-0,025 \leq p \leq 0,52+0,025$

$\Leftrightarrow \quad 0,495 \leq p \leq 0,535$

$I_C = [0,495 ; 0,535]$

$0,5 \in I_C$ donc on ne peut conclure. De plus, cette intervalle de confiance est précis à 95% (seulement).

Exercice 2

Lors d'un référendum, un sondage aléatoire simple pratiqué sur 1 000 personnes a donné 55% pour le "Oui" et 45% pour le "Non". Peut-on prévoir le résultat du référendum ?

CORRIGE :

On a bien : $n \geq 25$

L'**intervalle de confiance** est : $\left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}} \right] = \left[0,55-\frac{1}{\sqrt{1000}} ; 0,55+\frac{1}{\sqrt{1000}} \right] = [0,518 ; 0,582]$

Avec un risque d'erreur de 5%, on peut dire le "Oui" va l'emporter.

Exercice 3

Si, pour un référendum, on sait que "oui" se situe autour de 50%, combien de personnes faudrait-il interroger pour que la proportion de "Oui" soit connue à 1% près ? (en plus ou en moins)

CORRIGE :

L'**intervalle de confiance** est : $\left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}} \right]$

On veut : $\frac{1}{\sqrt{n}} \leq 0,01 \quad \Leftrightarrow \quad \sqrt{n} \geq 100 \quad \Leftrightarrow \quad n \geq 10000$

On peut aussi considérer que la proportion connue à 1% près traduit une amplitude de 0,02 :

$$\left(f+\frac{1}{\sqrt{n}} \right) - \left(f-\frac{1}{\sqrt{n}} \right) = 0,02 \quad \Leftrightarrow \quad f+\frac{1}{\sqrt{n}} - f + \frac{1}{\sqrt{n}} = 0,02 \quad \Leftrightarrow \quad \frac{2}{\sqrt{n}} = 0,02$$

$$\Leftrightarrow \quad \sqrt{n} \geq 100 \quad \Leftrightarrow \quad n \geq 10000$$

Exercice 4

Pourcentage de garçons à la naissance.

Dans un pays, sur 429 440 naissances, on a dénombré 221 023 garçons. Ce résultat est-il conforme à l'hypothèse selon laquelle il y a 50% de naissances masculines (et donc 50% de naissances féminines) ?

CORRIGE :

Intervalle de confiance de niveau 0,95 :

$$\left[\frac{221023}{429\,440}-\frac{1}{\sqrt{429\,440}};\frac{221023}{429\,440}+\frac{1}{\sqrt{429\,440}}\right]=[0,5132;0,5162]$$

soit entre 51.32% et 51.62% de naissances masculines donc non conformité avec l'hypothèse.

Exercice 5

Le dernier sondage de 2002 ne prévoyait pas la présence de Jean-Marie Le Pen au second tour.
Pouvait-on croire au sondage ?
21 Avril 2002 second tour de l'élection présidentielle en France.

Les sondages (1000 p) prévoient :	M Chirac : 19 %	M Jospin : 18 %	M Le Pen : 14 %
Les Résultats sont : Surprenant !	M Chirac : 19,88 %	M Jospin : 16,18 %	M Le Pen : 16,88 %

CORRIGE :

Les intervalles de confiance à 95 % sont :

M. Chirac : $\left[0,19-\frac{1}{\sqrt{1000}};0,19+\frac{1}{\sqrt{1000}}\right]=[0,158;0,222]$

M. Jospin : $\left[0,18-\frac{1}{\sqrt{1000}};0,18+\frac{1}{\sqrt{1000}}\right]=[0,148;0,212]$

M. Le Pen : $\left[0,14-\frac{1}{\sqrt{1000}};0,14+\frac{1}{\sqrt{1000}}\right]=[0,108;0,172]$

Il n'y a pas de surprise, seulement que M Jospin est dans la partie basse et M Le Pen dans la partie haute.

Exercice 6 **Qu'est-ce qu'un sondage ?**

Un maire voudrait connaître le pourcentage de personnes de sa commune favorables à un projet d'urbanisme, et ceci à partir d'une enquête portant sur un nombre restreint d'individus.
Il demande à quatre collaborateurs comment procéder.

- le 1er propose d'ouvrir à la mairie un registre pour recueillir l'avis des personnes désirant s'exprimer sur le sujet ;
- le 2e d'interroger les 1 350 habitants de son quartier ;
- le 3e d'interroger les 100 premières personnes rencontrées dans la rue, le mardi suivant à partir de 10 h.
- le 4e de sélectionner, de façon totalement aléatoire, 100 individus à interroger, à partir de la liste des habitants de la commune.

Tous les quatre pensent ensuite utiliser la formule bien connue : $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$

Ils affirment avoir une probabilité 0,05 de se tromper en disant que la proportion cherchée est dans cet intervalle.
À la place du maire, et indépendamment de toute considération de coût ou de difficulté de réalisation pratique (et en supposant que toutes les personnes interrogées répondent), quelle méthode choisiriez-vous ?

CORRIGE :

Celle du collaborateur :

1 : Non, car seules les personnes intéressées feront la démarche d'aller jusqu'à la mairie. Celles qui n'en ont pas le temps, ou qui ne se sentent pas assez concernées, ne seront pas consultées. L'échantillonnage ainsi réalisé ne sera pas représentatif de toute la population.

2 : Non, la taille de l'échantillon est grande, ce qui permettrait une bonne précision si on avait un échantillon vraiment aléatoire, mais il ne s'agit pas d'un tirage au hasard sur l'ensemble de la population, puisqu'on exclut du sondage tous les habitants des autres quartiers.

3 : Non, car on exclut du sondage toutes les personnes qui travaillent le mardi matin. On risque de n'interroger que des femmes au foyer, des retraités ou des chômeurs. L'échantillon ne serait pas représentatif de l'ensemble des habitants de la ville.

4 : **Oui**, c'est la seule démarche qui permette de justifier le recours à la formule donnant l'intervalle de confiance. Il est nécessaire d'avoir un échantillon aléatoire simple : tous les habitants ont la même chance d'être choisis, et de façon indépendante. Personne n'est exclu du sondage.

Le maire décide donc de choisir, à partir d'une liste de plusieurs milliers de noms, 100 personnes, "totalement au hasard". Mais comment faire pour être sûr d'agir "en toute objectivité" ?
Une solution est l'utilisation de tables de nombres au hasard, ou de procédés informatiques. À partir d'une liste numérotée de N noms, choisir les numéros de n personnes, de façon à ce que chacun ait la même probabilité d'être choisi, et de façon indépendante.
D'autre part, si le maire pense que son projet risque d'être ressenti différemment par les hommes et les femmes (implantation d'un stade de foot-ball par exemple), ou selon les tranches d'âge, et que la liste d'habitants dont il dispose mentionne le sexe et l'âge, que faire ?
Il peut améliorer la précision de son estimation en choisissant au hasard un certain nombre d'hommes, un certain nombre de femmes, un certain nombre d'individus par tranche d'âge. C'est ce que l'on appelle un sondage **stratifié**. De même, il peut être logique de procéder dans certains cas à des sondages à probabilités inégales : par exemple si les individus sont des entreprises, il peut être utile de les choisir avec des probabilités proportionnelles à leur chiffre d'affaire, ou au nombre de leurs salariés.

Remarque. Une méthode de sondage consiste à définir la façon dont on va prélever les individus dans la population afin de constituer un échantillon. Lorsque tous les individus ont la même probabilité d'appartenir à l'échantillon sélectionné, on parle de **sondage à probabilités égales**. Un sondage aléatoire est dit simple si tous les échantillons de taille n fixée sont réalisables avec la même probabilité. Il existe également des **sondages stratifiés** (s'appuyant sur des sous-populations appelées strates constituées à partir des données portant sur l'ensemble de la population), des **sondages par la méthode des quotas** (analogue aux sondages stratifiés mais avec probabilité inégales d'appartenir à l'échantillon sélectionné), . . .

Exercice 7

Sur un échantillon de 350 personnes, un candidat aux élections municipales a obtenu 54 % des intentions de vote.

a) Déterminer un intervalle de confiance.
b) Si les élections avaient eu lieu le jour de ce sondage et si les réponses au sondage étaient sincères, ce candidat aurait-il été élu au premier tour ?
c) Déterminer le nombre de personnes qu'il aurait fallu interroger afin d'être certain que ce candidat sera élu.

CORRIGE :

a) La proportion de visiteurs français est INCONNUE.
L'effectif de l'échantillon est $n = 350$. La fréquence de l'échantillon est 0,54.
On a bien : $0,2 \leq f \leq 0,8$ et $n \geq 25$
L'intervalle de confiance est donc :
$$I = \left[0,54 - \frac{1}{\sqrt{350}} ; 0,54 + \frac{1}{\sqrt{350}}\right] \text{ soit approximativement } I = [0,487 ; 0,593].$$
b) On ne peut pas affirmer que ce candidat a 95% de chances d'être élu.
c) Soit n le nombre cherché, et en supposant la même fréquence, l'intervalle de confiance est :
$$\left[0,54 - \frac{1}{\sqrt{n}} ; 0,54 + \frac{1}{\sqrt{n}}\right]$$
Il faut que $0,54 - \frac{1}{\sqrt{n}} > 0,5$
$$-\frac{1}{\sqrt{n}} > 0,5 - 0,54$$
$$-\frac{1}{\sqrt{n}} > -0,04$$
$$\frac{1}{\sqrt{n}} < 0,04$$

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

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

$$\sqrt{n} > 25$$

$$n > 625$$

Exercice 8

Un institut de sondage communique à un candidat aux élections régionales l'intervalle de confiance au niveau 0,95 de son futur score. Cet intervalle a une amplitude de 0,04.

Combien de personnes a interrogé l'institut de sondage ?

CORRIGE :

a) La proportion de visiteurs français est INCONNUE.
L'effectif de l'échantillon est n. La fréquence de l'échantillon n'est pas communiquée.
On suppose que : $0,2 \leq f \leq 0,8$ et $n \geq 25$
L'intervalle de confiance est donc :

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

Ainsi la taille de l'échantillon est :

$$\left(f + \frac{1}{\sqrt{n}} \right) - \left(f - \frac{1}{\sqrt{n}} \right) = f + \frac{1}{\sqrt{n}} - f + \frac{1}{\sqrt{n}} = \frac{2}{\sqrt{n}}$$

L'amplitude de l'intervalle est égale à 0,04 donc :

$$\frac{2}{\sqrt{n}} = 0,04$$

$$2 = 0,04 \times \sqrt{n}$$

$$\frac{2}{0,04} = \sqrt{n}$$

$$\frac{2}{0,04} = \sqrt{n}$$

$$\sqrt{n} = 50$$

$$n = 2500$$