EXERCICES
DES STATS INFÉRENTIELLES POUR VOUS PORTER BONHEUR !

Exercice 1 : On admet que la proportion de femmes dans la population française est de $51,4\%$.

1. Si l'on considère 577 personnes tirées aléatoirement avec remise, à quel intervalle appartiendra la proportion de femmes parmi ces 577 personnes avec une probabilité de 95% ?
2. A l'assemblée nationale, on dénombre 107 femmes sur 577 députés. Peut-on considérer qu'il y a une sous représentation des femmes à l'Assemblée Nationale ?

Exercice 2 : Dans un pays tropical, l'office national de tourisme déclare que seulement 8% des plages sont atteintes par des algues toxiques.
Un organisme indépendant souhaite vérifier la véracité de ceci et teste à son tour 70 plages (tirage considéré aléatoire et avec remise). 15 d'entre elles s'avèrent être atteintes par ces algues toxiques.

1. Donner l'intervalle de fluctuation asymptotique de la fréquence de plages atteintes par ces algues au seuil de 99% sur un échantillon aléatoire de taille 70 ?
2. Que va conclure l'organisme indépendant ?
3. Mêmes questions avec un intervalle de fluctuation asymptotique au seuil de 95%. Commenter.

Exercice 3 : Les enfants sont dits prématurés lorsque la durée gestationnelle est inférieure ou égale à 259 jours. La proportion de ces naissances est de 6%.
Des chercheurs suggèrent que les femmes ayant eu un travail pénible pendant leur grossesse sont plus susceptibles d'avoir un enfant prématuré que les autres. Il est décidé de réaliser une enquête auprès d'un échantillon aléatoire de 400 naissances correspondant à des femmes ayant eu pendant leur grossesse un travail pénible.
Les chercheurs décident a priori que si la proportion d'enfants nés prématurés dans cet échantillon est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 0,95 alors leur hypothèse sera acceptée. Finalement le nombre d'enfants prématurés est de 50. Quelle est donc la conclusion ?

Exercice 4 : Une compagnie aérienne possède des A340 (longs courriers) d'une capacité de 300 places.
Cette compagnie a vendu n billets pour le vol 2013.
La probabilité pour qu'un acheteur se présente à l'embarquement est p et les comportements des acheteurs sont indépendants les uns des autres.
On note X_n la variable aléatoire désignant le nombre d'acheteurs d'un billet se présentant à l'embarquement.
La compagnie cherche à optimiser le remplissage de l'avion en vendant éventuellement plus de places que la capacité totale de l'avion (surréservation ou surbooking) soit ici $n > 300$.
Comme il y a évidemment un risque que le nombre de passagers munis d'un billet se présentant à l'embarquement excède 300, la compagnie veut maîtriser ce risque.

1. Déterminer la loi de X_n.
2. On suppose que $0,5 \leq p \leq 0,95$. Écrire l'intervalle de fluctuation asymptotique I_n de $\dfrac{X_n}{n}$ au seuil de 0,95.
3. Montrer que si $I_n \subset \left[0, \dfrac{300}{n}\right]$ alors la probabilité que le nombre de passagers se présentant à l'embarquement excède 300 est proche de 0,05.
4. On cherche à déterminer la valeur de n maximale permettant de satisfaire la condition $I_n \subset \left[0, \dfrac{300}{n}\right]$.
   a. Montrer que $I_n \subset \left[0, \dfrac{300}{n}\right] \Longrightarrow pn + 1.96\sqrt{np(1-p)} - 300 \leq 0$
   b. On estime que $p = 0.9$ et on pose $f(x) = px + 1,96\sqrt{xp(1-p)} - 300$.
   Montrer qu'il existe un entier n_0 unique tel que si $n \leq n_0$ alors $f(n) \leq 0$ et si $n > n_0$ alors $f(n) > 0$ et le déterminer.
   c. Interpréter.

Exercice 5 : Vérifier que l'intervalle $[-2,576\ ;\ 1,696]$ peut être considéré comme un intervalle de fluctuation au seuil de 95% d'une variable X suivant une loi $N(0,1)$ (c'est-à-dire que $P(X \in I) \geq 0,95$).

Exercice 6 : Une société qui produit des jus de fruits propose au service commercial deux nouveaux mélanges : OPK (Orange Pamplemousse Kiwi) et OMA (Orange Mangue Ananas). Dans un sondage aléatoire réalisé sur 60 consommateurs, 56% préfère le mélange OPK. On fait l'hypothèse (H) que, dans l'ensemble des consommateurs, 50% préfère le mélange OPK.

1. Peut-on rejeter l'hypothèse (H) au risque de 5% ? *Bien détailler la démarche.*
2. Déterminer toutes les valeurs de p que l'on ne peut pas rejeter ?

Exercice 7 : Une ferme piscicole possède plusieurs bassins dans lesquels grandissent les alevins. Lorsque les poissons ont une masse supérieur ou égal à $650g$, ils sont prêts à être commercialisés.

1. On extrait au hasard un échantillon de 50 poissons. Le nombre de poissons est suffisamment grand pour considérer qu'il s'agit d'un tirage avec remise. On constate que 14 d'entre eux ont une masse inférieur à 650 g et 8 une masse supérieure ou égale à 1 kg. Donner un intervalle de confiance au niveau 95% de la proportion de poissons commercialisables dans ce bassin.
2. On suppose désormais que la variable aléatoire M qui à tout poisson du bassin associe sa masse en grammes suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Déterminer μ et σ en utilisant les données de la questions 1.
3. La ferme expédie les poissons par boîte de 20. On remplit les boîtes en prélevant au hasard 20 poissons dans le bassin et on appelle X la variable aléatoire qui à toute boîte associe le nombre de poissons dont la masse est inférieure à 650 g.
   a. Déterminer la loi de X.
   b. Quelle est la probabilité que la boîte contienne 20 poissons dont la masse est supérieure ou égale à 650 g.
   c. Quelle est en moyenne, par boîte, le nombre de poissons dont la masse est insuffisante pour être commercialisés ?

Exercice 8 : Dans une ville de 45 000 habitants, le maire souhaite restructurer le centre-ville. Pour convaincre l'ensemble du conseil municipal, il s'appuie sur une fourchette de sondage dans la commune qui estime, avec un intervalle de confiance à 95%, qu'entre 60 et 80% des habitants sont favorables à cette restructuration.

1. Combien de personnes sont interrogées dans ce sondage ?
2. Combien faudrait-il interroger de personnes au minimum pour que la fourchette de sondage soit d'une amplitude d'au plus 10% ?

Exercice 9 : Prenons un cas très classique : un sondage politique.
Le 18 avril 2002, l'institut IPSOS 13 effectue un sondage dans la population en âge de voter.
On constitue un échantillon de 1000 personnes (inscrites sur les listes électorales) que l'on suppose choisies ici de manière aléatoire. Ce n'est pas le cas en pratique [1], mais le principe reste le même que dans cet exemple.
Les résultats partiels en sont les suivants : sur les 1000 personnes

- 135 ont déclaré vouloir voter pour Jean-Marie Le Pen
- 195 ont déclaré vouloir voter pour Jacques Chirac
- 170 ont déclaré vouloir voter pour Lionel Jospin.

1. Pour chacun des trois candidats ci-dessus, déterminer l'intervalle de confiance au seuil de 95%.
2. L'institut ne donne qu'une valeur unique en pourcentage à la presse pour chaque candidat. Lesquelles ?
   Evaluer l'imprécision.
3. Les vrais résultats ont respectivement été : 16,9%, 19.9% et 16.2%. Commenter.
4. L'institut CSA lui donnait 14% pour Jean-Marie Le Pen (dans les mêmes conditions d'échantillonnage).
   Retrouver l'intervalle de confiance correspondant et commenter par rapport au score réel [2].

1. On peut consulter le site www.ipsos.fr/faq pour des détails sur les méthodes utilisées par cet institut.
2. En réalité, les instituts font des calculs bien plus complexes, notamment ils tentent de tenir au mieux compte du mensonge possible des gens (surtout pour les votes du FN, mais aussi des changements d'avis possibles jusqu'au jour de l'élection.

Solution :

1. On peut déterminer trois intervalles de confiance au niveau de confiance de 95% 14 :
   - Jean-Marie Le Pen $[0,135-0,032;0,135+0,032]=[0,103;0,167]$
   - Jacques Chirac $[0,195-0,032;0,195+0,032]=[0,163;0,227]$
   - Lionel Jospin $[0,170-0,032;0,170+0,032]=[0,138;0,202]$.
2. Les valeurs centrales des intervalles en pourcentage, et donc une imprécision de $\pm3\%$.
3. Ok pour Jacques et Lionel, pas pour Jean-Marie. Voir note de bas de page.
4. $[0.108;0.172]$ donc contient le score réel.