Problèmes statistiques

Exercice 1

Dans une usine automobile, on contrôle les défauts de peinture de type « grains ponctuels sur le capot ».

Lorsque le processus est sous contrôle, on a 20 % de ce type de défauts.
Lors du contrôle aléatoire de 50 véhicules, on observe 13 véhicules ayant des défauts.
Faut-il s'inquiéter ?

Ce type de contrôle de qualité a effectivement été pratiqué par un constructeur d'automobiles français. Il s'agissait de détecter une amélioration significative du procédé de peinture grâce à cet indicateur de défaut, quasiment invisible pour le client.

$\frac{13}{50} \times 100 = 26$. Non ~~Oui~~ il ne faut pas qu'elle s'inquiète car il y a une différence de 5 % entre la réalité et leur approximation
Non il ne faut pas s'inquiéter car ce n'est qu'un échantillon

exercice 1 :

$\frac{13}{50} = 0,26$

il y a donc sur l'échantillon 26% de ce type de défaut. Mais il ne faut pas qu'elle s'inquiète car l'échantillon de 50 véhicules n'est pas assez large et lorsqu'on dit qu'on a 20 % de ce type de défaut, c'est qu'on a environ 20 % avec une certaine marge d'erreur.
Cette échantillon est donc à peu près valable.

Comme $p = 0,2$ et que $n = 50$, les hypothèses du théorème sont vérifiées

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

c'est à dire $\left[0,2 - \frac{1}{\sqrt{50}} ; 0,2 + \frac{1}{\sqrt{50}} \right]$

soit $[0,06 ; 0,34]$

0,26 étant ~~entre~~ dans l'intervalle $[0,06 ; 0,34]$ alors, il ne faut pas s'inquiéter car la marge de confiance est $[0,06 ; 0,34]$.

Exercice 1 :

Dans un echantillon de taille $n = 50$, la fréquence f pour le nomb de voiture ayant des défaut est $\frac{13}{50} = 0,26$.

La proportion dans toute la proportion de voiture se trouve dans l'intervalle $\left[0,26 - \frac{1}{\sqrt{50}} ; 0,26 + \frac{1}{\sqrt{50}}\right]$

$\cancel{\simeq [0,12 ; 0,40]}$

donc $\left[0,2 - \frac{1}{\sqrt{50}} ; 0,2 + \frac{1}{\sqrt{50}}\right]$

$[0,06 ; 0,34]$

$0,06 < 0,26 < 0,34$

Cette donnée n'est pas inquiétante car la fréquence de voitur ayant des défauts est comprise dans l'intervalle $[0,06 ; 0,34]$

Exercice 1

$f = \frac{13}{50} = 0,26$

$p = 0,2 \qquad n = 50$

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

$\left[0,2 - \frac{1}{\sqrt{50}} ; 0,2 + \frac{1}{\sqrt{50}}\right]$

$[0,06 ; 0,34]$

Réponse : Il ne faut pas s'inquiéter parce que 0,26 se trouve dans l'intervalle $[0,06 ; 0,34]$. $0,06 < 0,26 < 0,34$

Exercice 1 : $20\% = \frac{1}{5}$

$50 \div 5 = 10$

Donc oui il faut s'inquiéter car quand le processus est sous contrôle le nombre des voitures avec un défaut est de 10 voitures, alors qu'ici il est de 13.

$P = 0{,}20 + \frac{1}{\sqrt{50}} \simeq 0{,}39$

$P = 0{,}20 - \frac{1}{\sqrt{50}} \simeq 0{,}058$

$f = 0{,}26$

$P \leq F \leq P$

$0{,}058 \leq 0{,}26 \leq 0{,}39$

Donc non il ne faut pas s'inquiéter car la frequence appartient à l'intervalle de confiance.

Exercice 2

Pour connaître la proportion *p* de fumeurs dans une entreprise, on effectue un sondage portant sur 160 employés pris au hasard. On constate que 40 employés sont fumeurs.

1) Donner une estimation *f* de cette proportion *p*.

2) Donner l'intervalle de confiance de *p* pour un risque a de 5% (donner les bornes de cet intervalle avec 2 chiffres significatifs).

3) **Quelle doit être la taille minimale de l'échantillon à prélever pour connaître *p* au risque 5% avec une erreur absolue (ou une précision) inférieure à 2% ?**

Quel problème statistique est alors mis en évidence ?

c) $0{,}25 - \frac{1}{\sqrt{n}} = 0{,}23 \quad | -0{,}25$

$-\frac{1}{\sqrt{n}} = -0{,}02$

$\sqrt{n} = \frac{1}{0{,}02} \quad | ()^2$

$n = 2500$

L'échantillon de 2500 est très grand. Il faut les moyens de réaliser la statistique avec un échantillon de 2500.

Exercice 3

On veut évaluer le nombre de Français porteurs d'un caractère génétique donné. On effectue une analyse sur 1 000 personnes choisies au hasard, et on observe que 384 personnes sont porteuses de ce caractère génétique. Évaluez par un intervalle le nombre de Français porteurs du caractère génétique sur une population de 64 millions de Français.

exercice 2

$\frac{384}{1000} \times 100 = 38{,}4$

64 millions x 0.384 ≈ 24,5 millions

Exercice 2.

6 / 1000 , 384/1000

64 000 × 384 = 24 576 000

24 576 000 / 64 000 000
de personnes sont porteurs
de ce caractère.

$64\,000\,000 \in \left[0,4 + \frac{1}{\sqrt{1000}} = 0,4 \; ; \; 0,4 - \frac{1}{\sqrt{1000}} = 0,35\right]$

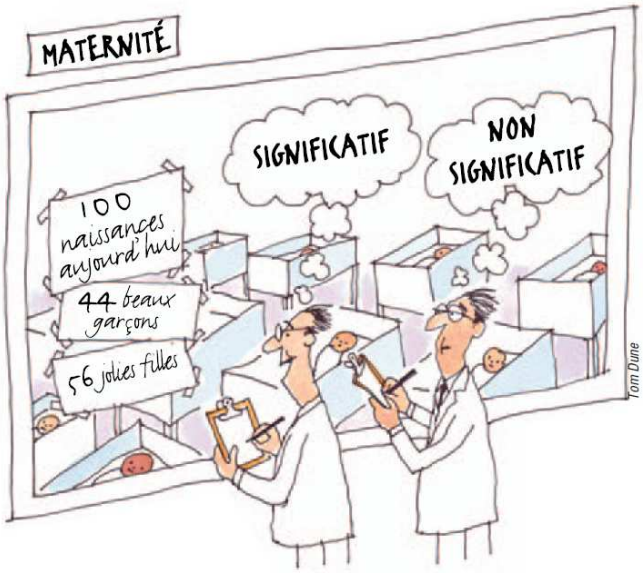
$0,35 < p < 0,4$

$22\,400\,000 < \cancel{64\,000\,000}\; 24\,576\,000 < 26\,880\,000$

Exercice 4

Commentez, en argumentant avec tous les moyens dont vous disposez, les textes de l'image ci-contre.

Dressez la liste de ces moyens dont vous disposez sur l'analyse de données dans le but d'évaluer une situation dans une population.



exercice 3 :

chaque jour, il y a 100 naissance.
aujourd'hui, nous avons un certain échantillon.
pour fille :

$p = \frac{56}{100} = 0,56$

$\left[0,56 - \frac{1}{\sqrt{100}} ; 0,56 + \frac{1}{\sqrt{100}}\right]$

soit

$p \in \left[\frac{44}{100} - \frac{1}{\sqrt{100}} ; \frac{44}{100} + \frac{1}{\sqrt{100}}\right]$

$[0,44 - 0,1 ; 0,44 + 0,1]$

$[0,34 ; 0,54]$

L'intervalle de confiance comprend p entre 0,34 et 0,54 pour la naissance des garçons sur 100 nouveaux-nés.

Dans une population très grand dans le 95% des cases; le frequence d'apparition des garçons se trovera entre $0,44 - \frac{1}{\sqrt{100}}$ et $0,44 + \frac{1}{\sqrt{100}}$, c'est a dire entre 0,34 et 0,54, et alors le frequence d'apparition des filles se trovera entre 0,66 et 0,46.

La frequence theorique (probabilité) d'avoir un fille ou un garçon c'est 0,5. $0,5 \in [0,34 ; 0,54]$ et $0,5 \in [0,46, 0,66]$, alors le event et normal et c'est pas significatif.

Il y a deux personnes qui pensent à ce qu'il y a écrit devant la maternité : 44 beaux garçons et 56 jolies filles pour 100 naissances.

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

$44 \in \left[44 + \frac{1}{\sqrt{100}} = 44,1 ; 44 - \frac{1}{\sqrt{100}} = 43,9 \right]$

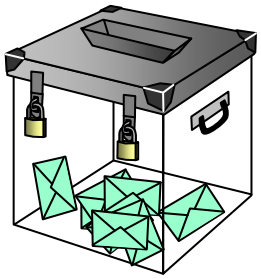
$44 \in [44,1 ; 43,9]$

Le nombre de beaux garçons appartient bien à l'intervalle.

Exercice 5

Voici un extrait d'article, publié dans le journal « Le Monde » par le statisticien Michel Lejeune, après le premier tour de l'élection présidentielle de 2002.

« Pour les rares scientifiques qui savent comment sont produites les estimations, il était clair que l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé. En effet, certains des derniers sondages indiquaient 18 % pour Jospin et 14 % pour Le Pen. Si l'on se réfère à un sondage qui serait effectué dans des conditions idéales [...], on obtient sur de tels pourcentages une incertitude de plus ou moins 3 % étant donné la taille de l'échantillon [...]. »



1) Si l'on tient compte de l'incertitude liée au sondage, entre quels pourcentages pourraient se situer réellement (à 95 % de confiance) les deux candidats lorsque le sondage donne 18 % pour l'un et 14 % pour l'autre ?

2) Représenter sur un même graphique les deux « fourchettes » calculées à la question précédente. Peut-on prévoir l'ordre des candidats ?

3) Au premier tour de l'élection présidentielle de 2002, L. Jospin a obtenu 16,18 % des voix et J.-M. Le Pen 16,86 %.

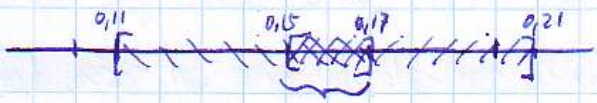
Expliquer la phrase « l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé ».

1)

Pour Jospin ; $p \in [0,18 - 3 ; 0,18 + 3] \Rightarrow p \in [0,15 ; 0,21]$

Pour Le Pen ; $p \in [0,14 - 3 ; 0,14 + 3] \Rightarrow p \in [0,11 ; 0,17]$

2)



Exercice 6 (source : Groupe « statistique et citoyenneté », IREM de Paris Nord.)
En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison pour cambriolage d'une résidence et tentative de viol.
Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population du comté était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des 11 années précédentes, seules 339 d'entre elles étaient d'origine mexicaine.
Lors du procès, un statisticien produisit des arguments pour convaincre la Cour Suprême du bien fondé de la requête de l'accusé (dont les juges votèrent à 5 contre 4 en faveur de la requête).
En vous situant dans le rôle de ce statisticien, produisez à votre tour des calculs, des raisonnements, des graphiques... pour montrer que le hasard ne peut pas « raisonnablement » expliquer à lui seul la sous-représentation des américains d'origine mexicaine dans les jurys de ce comté.

Sur 870 personnes seules 339 étaient d'origine mexicaine ce qui représente 39% des jurés alors que 79,1% de la population total du Texas est d'origine mexicaine. Il y a donc une certaine forme d'injustice car d'après le condamné il faudrait qu'il y est la même part de la population mexicaine dans le jury.