Math93.com

TD n°1 - Terminale ES/L

Échantillonnage et estimation

Les exercices suivants dont l'intitulé est suivi du symbole (c) sont corrigés intégralement en fin du présent TD. Les autres présentent des éléments de réponses et un lien vers une correction détaillée sur www.math93.com

Intervalle de fluctuation asymptotique : Prise de décision

Exercice 1. D'après Nouvelle calédonie, mars 2017

À l'occasion de la fête des Mères, un fleuriste décide de proposer à ses clients plusieurs types de bouquets spéciaux.

En se basant sur les ventes réalisées l'année précédente, ce fleuriste suppose que 85 % de ses clients viendront ce jour-là acheter un des bouquets pour la fête des Mères.
Quelques semaines avant de préparer ses commandes, il décide de vérifier son hypothèse en envoyant un questionnaire à 75 de ses clients, ces derniers étant supposés représentatifs de l'ensemble de sa clientèle.
Les réponses reçues montrent que, parmi les 75 clients interrogés, 16 déclarent qu'ils ne lui achèteront pas de bouquet pour la fête des Mères. Le fleuriste doit-il rejeter son hypothèse ?

Réponses
$I_{75} \approx [0,76\ ;\ 0,94]$ *et* $f \approx 0,79 \in I_{75}$.
Le corrigé complet sur www.math93.com.

Exercice 2. (c) Dans un cabinet d'assurance

Dans un cabinet d'assurance, une étude est réalisée sur la fréquence des sinistres déclarés par les clients ainsi que leur coût. Une enquête affirme que 20 % des clients ont déclaré un sinistre au cours de l'année. Un expert indépendant interroge un échantillon de 200 clients choisis au hasard dans l'ensemble des clients du cabinet d'assurance.
L'expert constate que 28 clients ont déclaré un sinistre au cours de l'année. Déterminer, en justifiant, si l'affirmation du cabinet d'assurance : « 20 % des clients ont déclaré un sinistre au cours de l'année » peut être validée par l'expert.

Intervalle de confiance : Estimation d'une proportion

Exercice 3. (c) Temps de livraison

On interroge au hasard 1 500 clients ayant effectué des achats sur un site internet et s'étant fait livrer le produit à domicile. Le temps de livraison a été jugé excessif par 795 personnes interrogées.
Peut-on considérer que plus de 50% des clients de ce site estiment que le temps de livraison est excessif ?

Exercice 4. (c) Temps d'attente

On interroge au hasard 100 clients ayant effectué des achats à la sortie d'une grande surface. Le temps d'attente aux caisses a été jugé raisonnable par 52 personnes interrogées.

1. Peut-on considérer que plus de 50% des clients de cette grande surface estiment que le temps d'attente aux caisses est raisonnable ?
2. Déterminer le nombre minimal de clients qu'il faut interroger pour estimer la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable avec une précision inférieure à à 0,04 soit 4% (en plus ou en moins).
3. À fréquence observée égale à $0,52$, quel nombre de clients aurait-il fallu interroger pour estimer que plus de 50% des clients trouvent que le temps d'attente aux caisses est raisonnable ?

Corrections

Correction de l'exercice 2 : Dans un cabinet d'assurance

Dans un cabinet d'assurance, une étude est réalisée sur la fréquence des sinistres déclarés par les clients ainsi que leur coût. Une enquête affirme que 20 % des clients ont déclaré un sinistre au cours de l'année. Un expert indépendant interroge un échantillon de 200 *clients choisis au hasard dans l'ensemble des clients du cabinet d'assurance. L'expert constate que* 28 *clients ont déclaré un sinistre au cours de l'année. Déterminer, en justifiant, si l'affirmation du cabinet d'assurance : « 20 % des clients ont déclaré un sinistre au cours de l'année » peut être validée par l'expert.*

- **Analyse des données** :
  - « Sur un échantillon de $n = 200$ clients. Il est constaté que 28 d'entre eux ont déclaré un sinistre. ». Donc la fréquence observée clients qui ont déclaré un sinistre est
  $$f = 28 \div 200 = 0{,}14 \text{ soit } \underline{f = 0{,}14}$$
  - On veut tester l'hypothèse : « la proportion de clients qui ont déclaré un sinistre est $\underline{p = 20\%}$ ».
- **Intervalle de fluctuation** :
  On a pour le cas étudié, $n = 200$, $p = 20$ %. Vérifions les conditions d'application du théorème :
  $$\begin{cases} \checkmark & n = 200 \geq 30 \\ \checkmark & np = 200 \times 0{,}2 = 40 \geq 5 \\ \checkmark & n(1-p) = 200 \times 0{,}8 = 160 \geq 5 \end{cases}$$
  Un intervalle fluctuation asymptotique au seuil de confiance de 95% de la fréquence f dans un échantillon de taille $n = 200$: est alors :
  $$I_n = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = \left[0{,}2 - 1{,}96\ \frac{\sqrt{0{,}2 \times 0{,}8}}{\sqrt{200}}\ ;\ 0{,}2 + 1{,}96\ \frac{\sqrt{0{,}2 \times 0{,}8}}{\sqrt{200}}\right]$$
  Soit puisque les borne sont :
  - $p - 1{,}96\ \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}14456$. On arrondit la borne inférieure par défaut à 10^{-3} près soit $\underline{0{,}144}$.
  - $p + 1{,}96\ \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}25544$. On arrondit la borne supérieure par excès à 10^{-3} près soit $\underline{0{,}256}$.
  $$\boxed{I_{200} \approx [0{,}144\ ;\ 0{,}256]}$$
- **Conclusion**
  La fréquence observée $f = 0{,}14$ n'appartient pas à l'intervalle de fluctuation asymptotique I_{200}, donc le résultat du contrôle remet en question l'hypothèse, avec un risque d'erreur de 5%.

Correction de l'exercice 3 : Intervalle de confiance et temps de livraison

On interroge au hasard 1 500 clients ayant effectué des achats sur un site internet et s'étant fait livrer le produit à domicile. Le temps de livraison a été jugé excessif par 795 personnes interrogées.
Peut-on considérer que plus de 50% des clients de ce site estiment que le temps de livraison est excessif?

- **Analyse des données** :
  - « *Sur un échantillon de* $n = 1500$ *clients. Il est constaté que* 795 *sont insatisfaits du temps de livraison.* ». Donc la fréquence observée de clients insatisfaits du temps de livraison est

    $$f = 795 \div 1500 = 0,53 \text{ soit } f = \underline{0,53}$$

  - On a veut savoir si plus de 50% des clients sont insatisfaits du temps de livraison donc si la proportion p de clients insatisfaits du temps de livraison est supérieure à 50% .

- **Intervalle de confiance** :
  On a pour le cas étudié, $n = 1500$, $f = 0,53$. Vérifions les conditions d'application du théorème :

  $$\begin{cases} \checkmark & n = 1500 \geq 30 \\ \checkmark & nf = 1500 \times \dfrac{795}{1500} = 795 \geq 5 \\ \checkmark & n(1-f) = 1500 \times \dfrac{705}{1500} = 705 \geq 5 \end{cases}$$

  Un intervalle de confiance au seuil de confiance de 95% de la proportion p est alors :

  $$I_n = \left[f - \frac{1}{\sqrt{n}} \ ; \ f + \frac{1}{\sqrt{n}}\right] = \left[\frac{795}{1500} - \frac{1}{\sqrt{1500}} \ ; \ \frac{795}{1500} + \frac{1}{\sqrt{1500}}\right]$$

  Soit puisque les borne sont :

  - $f - \dfrac{1}{\sqrt{n}} = \dfrac{795}{1500} - \dfrac{1}{\sqrt{1500}} \approx 0,50418$. On arrondit la borne inférieure par défaut à 10^{-3} près soit $\underline{0,504}$.
  - $f + \dfrac{1}{\sqrt{n}} = \dfrac{795}{1500} + \dfrac{1}{\sqrt{1500}} \approx 0,55582$. On arrondit la borne supérieure par excès à 10^{-3} près soit $\underline{0,556}$.

  $$\boxed{I_{1500} \approx [0,504 \ ; \ 0,556]}$$

- **Conclusion**
  Cet intervalle contient la proportion p des clients insatisfaits du temps de livraison , au niveau de confiance 95% (ou au risque d'erreur de 5%). La proportion des clients insatisfaits du temps de livraison se situe donc, au niveau de confiance 95%, entre 50.4% et 55.6%.
  On ne peut donc pas invalider l'affirmation.

Correction de l'exercice 4 : Intervalle de confiance et temps d'attente

On interroge au hasard 100 clients ayant effectué des achats à la sortie d'une grande surface. Le temps d'attente aux caisses a été jugé raisonnable par 52 personnes interrogées.

1. **Peut-on considérer que plus de 50% des clients de cette grande surface estiment que le temps d'attente aux caisses est raisonnable ?**

   - **Analyse des données** :
     - « *Sur un échantillon de* $n = 100$ *clients. Il est constaté que* 52 *sont satisfaits du temps d'attente.* ».
       Donc la fréquence observée de clients satisfaits du temps d'attente est
       $$f = 52 \div 100 = 0{,}52 \text{ soit } f = \underline{0{,}52}$$
     - On a veut savoir si plus de 50% des clients sont satisfaits du temps d'attente donc si la proportion p de clients satisfaits du temps d'attente est supérieure à 50% .
   - **Intervalle de confiance** :
     On a pour le cas étudié, $n = 100$, $f = 0{,}52$. Vérifions les conditions d'application du théorème :
     $$\begin{cases} \checkmark \quad n = 100 \geq 30 \\ \checkmark \quad nf = 100 \times \dfrac{52}{100} = 52 \geq 5 \\ \checkmark \quad n(1-f) = 100 \times \dfrac{48}{100} = 48 \geq 5 \end{cases}$$
     Un intervalle de confiance au seuil de confiance de 95% de la proportion p est alors :
     $$I_n = \left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right] = \left[\frac{52}{100} - \frac{1}{\sqrt{100}}\ ;\ \frac{52}{100} + \frac{1}{\sqrt{100}}\right]$$
     Soit puisque les borne sont :
     $$\left| \begin{array}{l} \blacksquare\ f - \dfrac{1}{\sqrt{n}} = \dfrac{52}{100} - \dfrac{1}{\sqrt{100}} = 0{,}42\ . \\ \blacksquare\ f + \dfrac{1}{\sqrt{n}} = \dfrac{52}{100} + \dfrac{1}{\sqrt{100}} = 0{,}62\ . \end{array} \right.$$
     $$\boxed{I_{100} \approx [0{,}42\ ;\ 0{,}62]}$$
   - **Conclusion**
     Cet intervalle contient la proportion p des clients satisfaits du temps d'attente , au niveau de confiance 95% (ou au risque d'erreur de 5%). La proportion des clients satisfaits du temps d'attente se situe donc, au niveau de confiance 95%, entre 42% et 62%.

2. **Déterminer le nombre minimal de clients qu'il faut interroger pour estimer la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable avec une précision inférieure à 0,04 soit 4% (en plus ou en moins).**

   La précision de l'estimation de p est $\dfrac{1}{\sqrt{n}}$. Pour tout entier naturel n :
   $$\frac{1}{\sqrt{n}} < 0{,}04 \iff \frac{1}{\sqrt{n}} < \frac{4}{100}$$
   On compose par la fonction inverse, strictement décroissante sur $]0\,;+\infty[$, l'ordre change
   $$\iff \sqrt{n} > \frac{100}{4} = 25$$
   On compose par la fonction carrée, strictement croissante sur $[0\,;+\infty[$, l'ordre est inchangé
   $$\iff n > 25^2 = 625$$
   Il faut interroger plus de 625 clients pour obtenir une estimation de la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable avec une précision inférieure à 0,04.

3. À fréquence observée égale à $0,52$, quel nombre de clients aurait-il fallu interroger pour estimer que plus de 50% des clients trouvent que le temps d'attente aux caisses est raisonnable ?

La borne inférieure de l'intervalle de confiance au niveau de confiance 0,95 sur un échantillon de taille n est $0,52-\frac{1}{\sqrt{n}}$ d'où n est solution de l'inéquation :

$$\begin{aligned} 0,52-\frac{1}{\sqrt{n}} \geqslant 0,5 &\iff -\frac{1}{\sqrt{n}} \geqslant -0,02 \\ &\iff \frac{1}{\sqrt{n}} \leqslant 0,02 \end{aligned}$$

On compose par la fonction inverse, strictement décroissante sur $]0\,;\,+\infty[$, l'ordre change

$$\begin{aligned} &\iff \sqrt{n} \geqslant \frac{1}{0,02} \\ &\iff \sqrt{n} \geqslant 50 \end{aligned}$$

On compose par la fonction carrée, strictement croissante sur $[0\,;\,+\infty[$, l'ordre est inchangé

$$\iff n \geqslant 50^2 = 2500$$

Avec une fréquence observée égale à 0,52, il faudrait un échantillon de taille supérieure à 2500 pour que la proportion p de clients qui trouvent le temps d'attente aux caisses raisonnable appartienne à un intervalle de confiance dont la borne inférieure est supérieure à 0,5.