Plaquette 5 – Problèmes de synthèse : sondages, contrôles et esprit critique

Échantillonnage — Première spé maths

Corrections détaillées

Boîte à outils

Modéliser. Variable aléatoire X, loi, espérance $\mu = E(X)$, variance $V(X) = E(X^2) - \mu^2$, écart type $\sigma = \sqrt{V(X)}$; linéarité $E(aX + b) = aE(X) + b$; pour une variable de Bernoulli, $\mu = p$ et $\sigma = \sqrt{p(1-p)}$.

Estimer. La moyenne m d'un échantillon de taille n est une estimation de μ ; la fréquence observée est une estimation de p. L'estimation est d'autant meilleure que n est grand et que l'échantillon est tiré au hasard.

Juger un écart. D'après les simulations, dans environ 95 % des échantillons,

$$|m - \mu| \leq \frac{2\sigma}{\sqrt{n}}$$

Un écart plus grand est **inhabituel** (environ 5 % des cas) : il invite à douter du modèle, sans le réfuter avec certitude.

Dimensionner. $\frac{2\sigma}{\sqrt{n}} \leq e \Leftrightarrow n \geq \left(\frac{2\sigma}{e}\right)^2$; pour une proportion, $n \geq \frac{1}{e^2}$ suffit toujours.

Biais. Aucune formule ne corrige un échantillon mal choisi (volontaires, mauvais lieu, mauvais moment).

Correction de l'exercice 1 – Un sondage et sa précision

a) Comme $\sigma = \sqrt{p(1-p)} \leq \frac{1}{2}$, on a

$$\frac{2\sigma}{\sqrt{1\,200}} \leq \frac{1}{\sqrt{1\,200}} \approx \frac{1}{34{,}64} \approx 0{,}029$$

b) Si $p = 0{,}5$, l'écart observé vaut $0{,}54 - 0{,}5 = 0{,}04$, supérieur à l'écart de référence ($0{,}029$ au plus). D'après les simulations, un tel écart n'apparaît que dans environ 5 % des sondages : il est inhabituel. Le titre est donc plausible, mais le sondage ne le prouve pas avec certitude, et il suppose l'échantillon correctement tiré.

c) Il suffit que $\frac{1}{\sqrt{n}} \leq 0{,}01$, c'est-à-dire $\sqrt{n} \geq 100$, soit $n \geq 10\,000$ personnes.

Réponse. a) $\leq 0{,}029$; b) non, écart de $0{,}04$ inhabituel si $p = 0{,}5$; c) $n \geq 10\,000$.

Correction de l'exercice 2 – Le poids des paquets de pâtes

a) $\frac{2 \times 8}{\sqrt{64}} = \frac{16}{8} = 2$ g. Pour une machine bien réglée, le poids moyen de 64 paquets est, dans environ 95 % des cas, entre 498 et 502 g.

b) L'écart observé vaut $|497{,}2 - 500| = 2{,}8$ g, supérieur à 2 g. Ce résultat est inhabituel pour une machine bien réglée (environ 5 % des contrôles au plus) : il est raisonnable de soupçonner un déréglage et de recommencer le contrôle.

c) Le poids d'un paquet isolé varie beaucoup (écart type 8 g) : un paquet à 510 g est banal, même si la machine verse en moyenne trop peu. C'est la **moyenne** sur 64 paquets, bien moins dispersée (écart type d'environ $\frac{8}{8} = 1$ g), qui renseigne sur le réglage.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Réponse. a) 2 g ; b) non : écart de 2,8 g, inhabituel ; c) un paquet isolé varie beaucoup ; c'est la moyenne qui compte.

Correction de l'exercice 3 – Estimer un temps de trajet

a) $\dfrac{2 \times 11}{\sqrt{100}} = \dfrac{22}{10} = 2{,}2$ minutes.

b) Si $\mu = 20$, l'écart avec la moyenne observée vaut $24{,}5 - 20 = 4{,}5$ minutes, plus de deux fois l'écart de référence. Un tel écart serait très inhabituel : l'affirmation du proviseur paraît peu crédible, le temps moyen est plutôt de l'ordre de 24 à 25 minutes.

c) On veut $\dfrac{22}{\sqrt{n}} \leq 1$, soit $\sqrt{n} \geq 22$, donc

$$n \geq 484$$

Il aurait fallu interroger environ 484 élèves, soit près d'un élève sur trois.

Réponse. a) 2,2 min ; b) non, écart de 4,5 min très inhabituel ; c) 484 élèves.

Correction de l'exercice 4 – Comparer deux groupes

a) Groupe 1 : $\dfrac{2 \times 1{,}2}{\sqrt{200}} \approx \dfrac{2{,}4}{14{,}14} \approx 0{,}17$ h. Groupe 2 : $\dfrac{2{,}4}{\sqrt{50}} \approx \dfrac{2{,}4}{7{,}07} \approx 0{,}34$ h.

b) L'écart entre les deux moyennes vaut 0,4 h. Mais chaque moyenne peut s'écarter de l'espérance de son groupe d'un écart de l'ordre de son écart de référence, soit jusqu'à $0{,}17 + 0{,}34 = 0{,}51$ h à eux deux. L'écart observé est du même ordre : il pourrait s'expliquer par la fluctuation, surtout du petit groupe. Les données suggèrent une différence, sans permettre de l'affirmer.

c) Même avec un écart net, l'étude ne montrerait pas que le téléphone **cause** le manque de sommeil : les élèves qui gardent leur téléphone peuvent différer des autres par l'âge, les horaires, les activités. Une liaison observée n'est pas une relation de cause à effet.

Réponse. a) $\approx 0{,}17$ h et $\approx 0{,}34$ h ; b) non, l'écart de 0,4 h est de l'ordre de la fluctuation ; c) corrélation n'est pas causalité.

Correction de l'exercice 5 – Dimensionner un contrôle

a) On veut $\dfrac{2 \times 0{,}9}{\sqrt{n}} \leq 0{,}1$, soit $\sqrt{n} \geq \dfrac{1{,}8}{0{,}1} = 18$, donc

$$n \geq 18^2 = 324$$

Il faut tester au moins 324 batteries.

b) Le contrôle coûte $324 \times 15 = 4\,860$ €.

c) Avec 81 batteries, $\dfrac{1{,}8}{\sqrt{81}} = \dfrac{1{,}8}{9} = 0{,}2$ h. En divisant le nombre de tests par 4, l'écart de référence est multiplié par 2 : le contrôle coûte quatre fois moins cher mais il est deux fois moins précis.

Réponse. a) 324 batteries ; b) 4 860 € ; c) 0,2 h.

Correction de l'exercice 6 – Critiquer une conclusion

a) $\dfrac{2 \times 1{,}5}{\sqrt{4}} = \dfrac{3}{2} = 1{,}5$ g.

b) L'écart observé vaut $13{,}1 - 12 = 1{,}1$ g, inférieur à l'écart de référence : même si l'annonce est exacte, un tel écart est tout à fait habituel sur 4 paquets. La conclusion n'est pas justifiée ; avec si peu de paquets, il faudrait un écart de plus de 1,5 g pour qu'il

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

soit inhabituel.

c) Les 4 paquets ont sans doute été achetés au même endroit, peut-être dans le même lot de fabrication : ils ne forment pas un échantillon tiré au hasard dans toute la production. Leurs teneurs peuvent se ressembler pour une raison commune, ce qui fausse le raisonnement.

Réponse. a) 1,5 g ; b) non : 1,1 < 1,5, écart habituel ; c) échantillon non tiré au hasard (même magasin, même lot).

Correction de l'exercice 7 – Fixer la mise d'un jeu

a) Le joueur reçoit 20 €, 5 € ou 0 € avec les probabilités $\frac{1}{20}$, $\frac{4}{20}$ et $\frac{15}{20}$, et paie x dans tous les cas. Par linéarité,

$$E(G) = 20 \times \frac{1}{20} + 5 \times \frac{4}{20} - x = 1 + 1 - x = 2 - x$$

b) Le jeu est équitable si $E(G) = 0$, c'est-à-dire pour $x = 2$ €.

c) Pour $x = 2{,}5$, $E(G) = -0{,}5$ €. L'écart type ne dépend pas de la mise : en notant R la somme reçue, $E(R) = 2$, $E(R^2) = 400 \times 0{,}05 + 25 \times 0{,}2 = 25$, donc $V(R) = 25 - 4 = 21$ et $\sigma(G) = \sigma(R) = \sqrt{21} \approx 4{,}58$ €.

d) Sur 400 parties, le gain moyen des joueurs est, dans environ 95 % des cas, à moins de $\frac{2 \times 4{,}58}{20} \approx 0{,}46$ € de $-0{,}5$ €, donc entre $-0{,}96$ € et $-0{,}04$ €. Le bénéfice du forain, opposé du gain total des joueurs, est donc entre $400 \times 0{,}04 = 16$ € et $400 \times 0{,}96 = 384$ € : environ 200 € en moyenne, avec une forte incertitude.

Réponse. a) $E(G) = 2 - x$; b) $x = 2$ € ; c) $E(G) = -0{,}5$, $\sigma \approx 4{,}58$; d) entre environ 16 € et 384 €.

Correction de l'exercice 8 – Estimer un effectif

a) La fréquence observée est $\frac{24}{300} = 0{,}08$. En l'égalant à la proportion $\frac{120}{N}$, on obtient $N \approx \frac{120}{0{,}08} = 1\,500$ papillons.

b) Avec $N = 2\,400$, la proportion de papillons marqués serait $p = \frac{120}{2\,400} = 0{,}05$. Alors $\sigma = \sqrt{0{,}05 \times 0{,}95} \approx 0{,}218$ et

$$\frac{2 \times 0{,}218}{\sqrt{300}} \approx \frac{0{,}436}{17{,}32} \approx 0{,}025$$

L'écart entre la fréquence observée et 0,05 vaut 0,03, supérieur à 0,025 : la fréquence observée serait inhabituelle. L'hypothèse de 2 400 papillons est donc peu crédible, sans être exclue.

c) La proportion de papillons marqués est petite, et N s'en déduit par une division : une petite erreur sur p se traduit par une grande erreur relative sur N. Marquer davantage de papillons au départ augmenterait p et rendrait l'estimation plus précise.

Réponse. a) $p \approx 0{,}08$, $N \approx 1\,500$; b) $p = 0{,}05$, écart de référence $\approx 0{,}025$: 0,08 serait inhabituel ; c) p petit : erreur relative forte sur N.

Correction de l'exercice 9 – Le coût des sinistres

a) $E(C) = 500 \times 0{,}08 + 5\,000 \times 0{,}02 = 40 + 100 = 140$ €. Puis $E(C^2) = 250\,000 \times 0{,}08 + 25\,000\,000 \times 0{,}02 = 20\,000 + 500\,000 = 520\,000$, d'où

$$V(C) = 520\,000 - 140^2 = 500\,400 \qquad \sigma(C) \approx 707 \text{ €}$$

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

b) $\dfrac{2 \times 707}{\sqrt{2\,500}} = \dfrac{1\,414}{50} \approx 28$ €.

c) Dans environ 95 % des cas, le coût moyen par contrat est entre 112 € et 168 €. La prime de 160 € couvre presque tout cet intervalle, mais pas sa partie haute : l'assureur sera bénéficiaire la plupart du temps, sans que ce soit garanti dans 95 % des cas.

d) Avec 100 contrats, l'écart de référence devient $\dfrac{1\,414}{10} \approx 141$ € : le coût moyen par contrat peut aller, dans 95 % des cas, jusqu'à environ 281 €. Deux ou trois gros sinistres suffisent à mettre la mutuelle en déficit : les petits portefeuilles sont bien plus risqués, ce qui explique le regroupement des assureurs.

Réponse. a) $E(C) = 140$ €, $\sigma \approx 707$ € ; b) ≈ 28 € ; c) le plus souvent, mais pas garanti ($[112\,;168]$) ; d) non : écart de référence ≈ 141 €.

Correction de l'exercice 10 – Lancers francs et confiance

a) $\dfrac{340}{400} = 0{,}85$: on estime $p \approx 0{,}85$.

b) $\sigma = \sqrt{0{,}85 \times 0{,}15} = \sqrt{0{,}1275} \approx 0{,}357$, d'où

$$\frac{2 \times 0{,}357}{\sqrt{25}} \approx 0{,}143$$

La fréquence de cette saison vaut $\dfrac{18}{25} = 0{,}72$, soit un écart de 0,13 avec 0,85, inférieur à l'écart de référence : ce résultat reste habituel pour une joueuse à 85 %.

c) Il est trop tôt pour conclure : 25 tentatives forment un petit échantillon. Il faut attendre davantage de lancers ; avec 100 tentatives, l'écart de référence tomberait à environ 0,07, et une fréquence de 0,72 deviendrait alors inhabituelle.

Réponse. a) 0,85 ; b) $\approx 0{,}143$; 0,72 reste habituel ; c) attendre davantage de tentatives.

Correction de l'exercice 11 – Les retards de livraison

a) $E(R) = 0{,}2 + 0{,}3 = 0{,}5$ jour. $E(R^2) = 0{,}2 + 9 \times 0{,}1 = 1{,}1$, d'où $V(R) = 1{,}1 - 0{,}25 = 0{,}85$ et $\sigma(R) \approx 0{,}92$ jour.

b) L'écart de référence vaut $\dfrac{2 \times 0{,}92}{\sqrt{50}} \approx 0{,}26$ jour. L'écart observé, $0{,}82 - 0{,}5 = 0{,}32$ jour, le dépasse : ce client a eu une année inhabituellement mauvaise, ce qui peut justifier une vérification de son circuit de livraison.

c) Non. Même si le modèle est exact, le retard moyen d'un client de 50 livraisons dépasse 0,6 jour dès qu'il s'écarte de plus de 0,1 de l'espérance, soit moins de la moitié de l'écart de référence. Cela arrivera à une part importante des clients, bien plus que 5 %. Un engagement prudent serait de l'ordre de

$$0{,}5 + 0{,}26 \approx 0{,}8 \text{ jour}$$

Réponse. a) $E(R) = 0{,}5$, $\sigma \approx 0{,}92$; b) non : écart $0{,}32 > 0{,}26$; c) non, il faudrait viser environ 0,8 jour.

Correction de l'exercice 12 – Un taux de réponse trompeur

a) Le taux de réponse vaut $\dfrac{640}{8\,000} = 0{,}08$, soit 8 %. La fréquence de satisfaits parmi les répondants est $\dfrac{576}{640} = 0{,}9$.

b) $\sigma = \sqrt{0{,}09} = 0{,}3$, d'où

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

$$\frac{2\times 0{,}3}{\sqrt{640}} \approx \frac{0{,}6}{25{,}3} \approx 0{,}024$$

c) Le calcul de l'écart de référence suppose que les 640 répondants forment un échantillon tiré **au hasard** parmi les usagers. Ce n'est pas le cas : ce sont des volontaires, qui ne représentent que 8 % des usagers, et rien ne dit que les satisfaits et les mécontents répondent dans les mêmes proportions. La précision « à 2 points près » n'a donc pas de sens ici : le biais de non-réponse peut être bien plus grand que la fluctuation d'échantillonnage.

Réponse. a) 8 % de réponses, 0,9 de satisfaits ; b) $\approx 0{,}024$; c) l'échantillon n'est pas tiré au hasard : précision illusoire.

Correction de l'exercice 13 – Un dé de casino

a) $\dfrac{2\times 1{,}708}{\sqrt{3\,000}} \approx \dfrac{3{,}416}{54{,}77} \approx 0{,}062$.

b) L'écart observé vaut $3{,}58 - 3{,}5 = 0{,}08$, supérieur à $0{,}062$: c'est un résultat inhabituel pour un dé équilibré, qui justifie un examen plus poussé.

c) La fréquence des « 6 » vaut $\dfrac{560}{3\,000} \approx 0{,}187$. Pour un dé équilibré, la variable « obtenir 6 » est de Bernoulli de paramètre $\dfrac{1}{6}$, d'écart type $\sqrt{\dfrac{1}{6}\times\dfrac{5}{6}} = \dfrac{\sqrt{5}}{6} \approx 0{,}373$, d'où

$$\frac{2\times 0{,}373}{\sqrt{3\,000}} \approx 0{,}014$$

L'écart $0{,}187 - 0{,}167 = 0{,}02$ dépasse aussi l'écart de référence. Les deux observations vont dans le même sens : le dé semble favoriser le 6, et le casino ferait bien de le retirer.

Réponse. a) $\approx 0{,}062$; b) non ($0{,}08 > 0{,}062$) ; c) $\approx 0{,}187$, écart $0{,}02 > 0{,}014$: le dé semble favoriser le 6.

Correction de l'exercice 14 – Multiplier les contrôles

a) Pour chaque commune conforme, un écart supérieur à $\dfrac{2\sigma}{\sqrt{n}}$ se produit dans environ 5 % des cas. Sur 60 communes, on peut donc s'attendre à environ

$$60 \times 0{,}05 = 3$$

communes déclarées « anormales » chaque mois, par le seul effet du hasard.

b) Une alerte isolée ne prouve rien : elle fait peut-être partie des trois fausses alertes attendues chaque mois. Le titre est alarmiste ; il faudrait savoir si la commune a été de nouveau signalée lors des contrôles suivants.

c) Par exemple, ne déclencher une alerte que si une commune dépasse l'écart de référence deux mois de suite, ou refaire immédiatement un contrôle avec un échantillon plus grand avant toute annonce.

Réponse. a) Environ 3 communes ; b) une alerte isolée peut être due au hasard ; c) exiger deux dépassements successifs ou un second contrôle.

Correction de l'exercice 15 – Synthèse : l'audit d'un centre d'appels

a) $E(A) = 0{,}5 + 0{,}6 + 0{,}9 = 2$ minutes, et les attentes de moins de 3 minutes ($A = 1$ ou $A = 2$) ont la probabilité $0{,}5 + 0{,}3 = 0{,}8$: le modèle respecte l'engagement. Puis $E(A^2) = 0{,}5 + 1{,}2 + 4{,}05 = 5{,}75$, d'où $V(A) = 5{,}75 - 2^2 = 1{,}75$ et $\sigma(A) \approx 1{,}32$.

b) $\dfrac{2\times 1{,}32}{\sqrt{144}} = \dfrac{2{,}64}{12} = 0{,}22$ minute.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

c) L'écart avec l'espérance vaut $2{,}45 - 2 = 0{,}45$ minute, environ deux fois l'écart de référence : la durée moyenne observée est très inhabituelle si l'engagement est tenu.

d) La fréquence vaut $\dfrac{106}{144} \approx 0{,}736$. Pour une proportion de 0,8, $\sigma = \sqrt{0{,}8 \times 0{,}2} = 0{,}4$ et l'écart de référence vaut

$$\frac{2 \times 0{,}4}{12} \approx 0{,}067$$

L'écart $0{,}8 - 0{,}736 = 0{,}064$ est juste en dessous : ce résultat reste dans les cas habituels, mais à la limite.

e) On simule chaque attente selon le modèle, puis on calcule les deux indicateurs :

```
from random import random

def attente():
    u = random()
    if u < 0.5:
        return 1
    elif u < 0.8:
        return 2
    return 4.5

def audit():
    L = [attente() for i in range(144)]
    rapides = [a for a in L if a < 3]
    return sum(L) / 144, len(rapides) / 144
```

En appelant *audit()* un grand nombre de fois, on pourrait vérifier qu'environ 95 % des audits donnent une moyenne entre 1,78 et 2,22 minutes.

f) « Sur 144 appels tirés au hasard, l'attente moyenne (2,45 minutes) est nettement supérieure à ce que donnerait le prestataire s'il tenait son engagement : un tel écart n'apparaît que dans environ 5 % des audits. La proportion d'appels décrochés en moins de 3 minutes (73,6 %) est inférieure aux 80 % promis, sans que l'écart soit, à lui seul, inhabituel. Ces résultats invitent à demander des explications au prestataire et à renouveler l'audit le mois prochain. »

Réponse. a) $E(A) = 2$, $P(A < 3) = 0{,}8$, $\sigma \approx 1{,}32$; b) 0,22 min ; c) non ($0{,}45 > 0{,}22$) ; d) $\approx 0{,}736$, écart $0{,}064 < 0{,}067$: habituel, à la limite ; e) liste de 144 attentes, moyenne et fréquence ; f) durée inhabituelle, à surveiller.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).