Plaquette 5 – Problèmes de synthèse : santé, médias et sport

Échantillonnage et fluctuation — Seconde

Corrections détaillées

Boîte à outils

Trois questions à se poser devant une fréquence.

1. **Sur quelle taille ?** Une fréquence calculée sur peu d'individus fluctue beaucoup ; sur un grand échantillon, elle est, sauf exception, proche de la proportion réelle (loi des grands nombres). 2. **Quel échantillon ?** Seul un échantillon tiré au hasard dans la population permet de généraliser ; un biais de sélection ne se corrige pas en interrogeant plus de monde. 3. **Quel écart ?** Comparer l'écart observé à la fluctuation habituelle, connue par simulation (étendue ou écart type des fréquences simulées).

Combiner des proportions. Si une population est formée de deux groupes de parts a et b (avec $a + b = 1$), où les proportions d'un caractère sont p_1 et p_2, la proportion globale vaut

$$p = a\,p_1 + b\,p_2$$

Rédiger. « Les données sont compatibles avec… », « l'écart peut s'expliquer par la fluctuation d'échantillonnage », « l'écart est bien plus grand que ce que produit le hasard dans les simulations » : jamais « c'est prouvé » à partir d'un seul échantillon.

Correction de l'exercice 1 – Couverture vaccinale de deux départements

a) Dans le département A, $f_A = \dfrac{312}{400} = 0{,}78$; dans le département B, $f_B = \dfrac{51}{60} = 0{,}85$.

b) L'écart est de 7 points, mais l'échantillon du département B ne compte que 60 personnes : une seule personne y pèse

$$\frac{1}{60} \approx 0{,}017$$

soit près de 2 points. Avec un si petit échantillon, la fréquence fluctue fortement d'une enquête à l'autre, et l'écart peut s'expliquer par le hasard. Le titre affirme plus que ce que les données permettent.

c) Il faudrait interroger dans le département B un échantillon aussi grand que dans le département A, tiré lui aussi au hasard. Si l'écart persiste avec deux échantillons de plusieurs centaines de personnes, la comparaison deviendra crédible.

Réponse. a) 0,78 et 0,85 ; b) non : l'échantillon de B (60 personnes) fluctue trop ; c) un échantillon plus grand en B.

Correction de l'exercice 2 – Le petit et le grand hôpital

a) Beaucoup de personnes répondent « autant dans les deux », car la probabilité d'un garçon est la même. C'est oublier que la taille de l'échantillon joue un rôle.

b) Petit hôpital : $\dfrac{70}{365} \approx 0{,}19$. Grand hôpital : $\dfrac{29}{365} \approx 0{,}08$. Le petit hôpital connaît plus de deux fois plus de jours « à plus de 60 % de garçons ».

c) Chaque jour fournit un échantillon de naissances : 15 au petit hôpital, 45 au grand. Plus l'échantillon est petit, plus sa fréquence de garçons fluctue autour de 0,512, et plus souvent elle s'en écarte au point de dépasser 0,6. Au petit hôpital, il suffit de 10 garçons

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).

sur 15 naissances, ce qui arrive assez souvent ; au grand, il en faut au moins 28 sur 45, bien plus rare. La loi des grands nombres explique que les grands échantillons « s'écartent » moins.

Réponse. b) $\approx 0{,}19$ et $\approx 0{,}08$; c) le petit hôpital, car ses échantillons quotidiens sont plus petits, donc fluctuent davantage.

Correction de l'exercice 3 – Mesurer l'audience d'une émission

a) La part d'audience dans le panel vaut $\dfrac{1\,150}{5\,000} = 0{,}23$, soit 23 %.

b) Le panel étant représentatif, on suppose que la même proportion de foyers regarde l'émission dans tout le pays :

$$30\,000\,000 \times 0{,}23 = 6\,900\,000$$

soit environ 6,9 millions de foyers. C'est une estimation : un autre panel de même taille donnerait une fréquence un peu différente.

c) L'écart est de 0,4 point, soit 0,004. Sur un panel de 5 000 foyers, la fluctuation d'échantillonnage est du même ordre, voire plus grande : un écart aussi faible ne permet pas d'affirmer que la seconde émission a été plus regardée. On peut seulement dire que les deux audiences sont très proches.

Réponse. a) 23 % ; b) environ 6,9 millions de foyers ; c) non, l'écart de 0,4 point est trop faible.

Correction de l'exercice 4 – Les tirs au but

a) Sur 500 tirs, la fréquence de réussite est $\dfrac{400}{500} = 0{,}8$: on estime la probabilité de réussite à 0,8.

b) En match, la fréquence vaut $\dfrac{3}{5} = 0{,}6$. D'après la simulation, rater au moins deux tirs sur cinq arrive à un tireur à 80 % dans environ une séance sur quatre : c'est un résultat banal. Avec 5 tirs seulement, la fréquence fluctue énormément : chaque tir pèse 20 points. Rien ne permet d'affirmer que le joueur a « perdu ses moyens ».

c) Il faudrait observer ses tirs en match sur un grand nombre de séances, plusieurs dizaines ou centaines de tirs, pour que la fréquence soit, d'après la loi des grands nombres, proche de sa vraie probabilité de réussite en match et puisse être comparée à

$$p \approx 0{,}8$$

Réponse. a) 0,8 ; b) affirmation non fondée : ce résultat arrive dans environ 26 % des séances ; c) sur un grand nombre de tirs.

Correction de l'exercice 5 – Un sondage dans deux régions

a) Les habitants favorables du Nord représentent 55 % de 60 % de la population, et ceux du Sud 35 % de 40 % :

$$0{,}6 \times 0{,}55 + 0{,}4 \times 0{,}35 = 0{,}33 + 0{,}14 = 0{,}47$$

On estime donc à 47 % la proportion d'habitants favorables dans le pays.

b) Le journaliste donne le même poids aux deux régions, alors que le Nord est plus peuplé. Il faut pondérer chaque proportion par la part de la population qu'elle concerne ; sa moyenne simple sous-estime l'avis du Nord, plus favorable.

c) Parmi les 47 % d'habitants favorables, 33 % habitent le Nord. La part cherchée est $\dfrac{0{,}33}{0{,}47} \approx 0{,}70$: environ 70 % des personnes favorables sont du Nord.

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) 47 % ; b) il ne pondère pas par la taille des régions ; c) environ 70 %.

Correction de l’exercice 6 – Le sondage qui s’est trompé

a) Dans l’Amérique de 1936, en pleine crise économique, posséder une voiture ou le téléphone, ou s’abonner à un magazine, était le signe d’une certaine aisance. L’échantillon du *Literary Digest* surreprésentait les électeurs aisés, plus favorables à Landon : c’est un **biais de sélection**. En outre, seules les personnes qui ont bien voulu répondre sont comptées. Interroger plus de monde dans une population mal choisie reproduit la même erreur à plus grande échelle.

b) La loi des grands nombres suppose que l’échantillon est tiré **au hasard** dans la population étudiée. Elle garantit alors que, pour un grand échantillon, la fréquence est, sauf exception, proche de la proportion réelle. Elle ne dit rien d’un échantillon biaisé : la qualité du tirage compte plus que la taille. L’institut, avec 50 fois moins de réponses mais un échantillon représentatif, a vu juste.

Réponse. a) Échantillon biaisé (abonnés, propriétaires de voitures et de téléphones, plus aisés) ; b) la loi des grands nombres ne vaut que pour un échantillon tiré au hasard.

Correction de l’exercice 7 – « Le chômage baisse de 0,1 point »

a) L’écart vaut $7{,}4 - 7{,}3 = 0{,}1$ point, soit 0,001 en fréquence.

b) La simulation montre que, même si le taux de chômage ne change pas, deux sondages de 1 000 personnes donnent couramment des résultats éloignés de près d’un point (l’écart type vaut 0,8 point), et parfois de plusieurs points. Un écart de 0,1 point est donc huit fois plus petit que la fluctuation habituelle :

$$0{,}1 \ll 0{,}8$$

Il ne permet pas d’affirmer que le chômage a reculé ; le titre honnête serait « Le chômage est stable ». Pour mesurer des variations aussi fines, il faut des échantillons beaucoup plus grands, comme ceux des enquêtes officielles.

Réponse. a) 0,1 point ; b) non, l’écart est bien plus petit que la fluctuation d’un sondage de 1 000 personnes.

Correction de l’exercice 8 – Choisir la taille d’une enquête

a) Avec 25 personnes, seuls 79 % des échantillons tombent à moins de 10 points de 0,5 : environ un sur cinq s’en écarte davantage, l’exigence de 95 % n’est pas satisfaite. Avec 100 personnes, la proportion atteint 95,2 % : la taille 100 convient, de justesse.

b) On a observé par simulation (plaquette 1) que la dispersion des fréquences est à peu près divisée par 2 quand la taille de l’échantillon est multipliée par 4. Pour diviser la marge par 2 (de 10 à 5 points) en gardant la même proportion de 95 % d’échantillons, il faut donc environ

$$4 \times 100 = 400$$

personnes. La réponse de l’institut est cohérente avec les simulations.

Réponse. a) 25 : non (79 %) ; 100 : oui (95,2 %) ; b) oui : diviser la marge par 2 demande environ 4 fois plus de personnes.

Correction de l’exercice 9 – Les naissances d’un village

a) Dans le village, $f = \dfrac{31}{52} \approx 0{,}596$, soit environ 60 % de garçons.

b) D’après la simulation, un village de 52 naissances obtient au moins 31 garçons environ une fois sur sept, sans aucune cause particulière : c’est la fluctuation d’échantillonnage

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).

sur un petit échantillon. Le résultat du village n'a rien d'extraordinaire, et il n'y a pas besoin de chercher une explication dans l'eau.

c) Si toutes les communes avaient environ 52 naissances par an, on pourrait s'attendre à environ

$$35\,000 \times 0{,}14 = 4\,900$$

communes avec au moins 60 % de garçons. Chercher « la » commune où il naît beaucoup de garçons, c'est être sûr d'en trouver, par le seul effet du hasard.

Réponse. a) ≈ 0,596 ; b) résultat banal (environ un village sur sept) ; c) environ 4 900 communes.

Correction de l'exercice 10 – Une machine qui se dérègle ?

a) Le lundi, $f = \dfrac{6}{200} = 0{,}03$; le dimanche, $f = \dfrac{19}{200} = 0{,}095$.

b) De lundi à jeudi, les nombres de pièces défectueuses (6, 5, 8, 7) sont entre 1 et 12, comme dans 99,4 % des contrôles simulés d'une machine bien réglée : ils sont compatibles avec un bon réglage. Vendredi, samedi et dimanche, on compte 13, 15 puis 19 pièces défectueuses. D'après la simulation, 13 défauts ou plus n'arrivent que dans

$$\frac{4}{1\,000} = 0{,}4\,\%$$

des contrôles d'une machine bien réglée ; l'observer trois jours de suite, avec des valeurs qui augmentent, serait extrêmement improbable par le seul hasard.

c) La machine s'est très probablement déréglée à partir de vendredi : il faut l'arrêter et la régler, et contrôler plus attentivement les pièces produites depuis vendredi.

Réponse. a) 0,03 et 0,095 ; b) lundi–jeudi : compatibles ; vendredi–dimanche : très improbables pour une machine bien réglée ; c) la régler.

Correction de l'exercice 11 – Comparer deux tireurs de lancers francs

a) Hugo : $\dfrac{9}{10} = 0{,}9$; Yanis : $\dfrac{160}{200} = 0{,}8$.

b) La fréquence d'Hugo est plus élevée, mais elle repose sur 10 lancers seulement : chaque lancer y pèse 10 points, et un tireur à 80 % peut très bien réussir 9 lancers sur 10 par chance. La fréquence de Yanis, calculée sur 200 lancers, est une estimation bien plus fiable de son adresse. Les données ne permettent pas de dire qu'Hugo est meilleur ; il faudrait l'observer sur beaucoup plus de lancers.

c) Sur 20 lancers, une fréquence de 0,8 correspond à $0{,}8 \times 20 = 16$ réussites. Hugo en a déjà 9 ; il lui en faudrait donc

$$16 - 9 = 7$$

sur les 10 suivants. Cela montre à quel point sa fréquence actuelle est fragile.

Réponse. a) 0,9 et 0,8 ; b) la fréquence d'Hugo, sur 10 lancers, n'est pas fiable ; c) 7 réussites.

Correction de l'exercice 12 – Trois années d'enquête

a) La plus petite fréquence simulée est 0,285 et la plus grande 0,331 :

$$0{,}331 - 0{,}285 = 0{,}046$$

b) Dans la simulation, la proportion réelle ne change jamais, et pourtant les fréquences varient de près de 5 points d'une enquête à l'autre. Les variations observées (+2 points puis −3 points) sont du même ordre : elles peuvent s'expliquer entièrement par la

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).

fluctuation d'échantillonnage. Le rapport surinterprète les chiffres ; on peut seulement dire que la pratique musicale concerne environ 30 % des lycéens et semble stable.

Réponse. a) 0,046 ; b) non : ces variations sont de l'ordre de la fluctuation d'échantillonnage.

Correction de l'exercice 13 – Rédiger une conclusion

La fréquence observée vaut

$$f = \frac{305}{500} = 0{,}61$$

Elle est en dehors de la plage $[0{,}65\,;\,0{,}75]$ où tombent presque toutes les simulations d'une population à 70 % de satisfaits.

— A : à éviter. Un résultat très improbable n'est pas impossible ; un sondage ne « prouve » rien avec certitude.
— B : correct. Elle compare l'observation à la fluctuation habituelle et conclut avec prudence.
— C : faux. 0,61 est la fréquence dans l'échantillon, pas la proportion exacte chez tous les clients, qui reste inconnue.
— D : correct. C'est un simple constat sur l'échantillon (et même exact : 61 %).

On retient B et D.

Réponse. B et D ($f = 0{,}61$, hors de $[0{,}65\,;\,0{,}75]$).

Correction de l'exercice 14 – Pas de chance au tirage au sort ?

a) La fréquence des tirages perdus vaut $\frac{8}{10} = 0{,}8$.

b) Chaque équipe a environ 5,5 % de chances d'être dans ce cas, d'où, sur 20 équipes :

$$20 \times 0{,}055 = 1{,}1$$

Il est donc normal qu'environ une équipe du championnat ait perdu au moins 8 tirages sur 10.

c) Non. Le résultat est rare pour une équipe donnée, mais quand on regarde tout un championnat, on s'attend à ce que cela arrive à une équipe : celle-ci a l'impression d'être visée alors qu'elle subit simplement la fluctuation sur 10 lancers. Pour soupçonner la pièce, il faudrait la tester sur un grand nombre de lancers.

Réponse. a) 0,8 ; b) environ 1 équipe ; c) non, c'est une fluctuation ordinaire à l'échelle du championnat.

Correction de l'exercice 15 – Synthèse : une étude sur le sommeil

a) Seconde : $\frac{75}{250} = 0{,}3$; première : $\frac{135}{350} \approx 0{,}386$; terminale : $\frac{94}{200} = 0{,}47$.

b) On additionne les effectifs :

$$f = \frac{75 + 135 + 94}{800} = \frac{304}{800} = 0{,}38$$

c) Non. Sur l'ensemble, 38 % des lycéens dorment moins de 7 heures, soit un peu plus d'un sur trois. « Près d'un sur deux » ne vaut que pour les terminales (47 %) : l'article généralise à tous les lycéens le chiffre du niveau le plus touché.

d) L'écart entre la seconde et la terminale est de 17 points, bien plus grand que les fluctuations de l'ordre de 3 points observées en simulation pour des échantillons de cette taille. Cet écart ne s'explique vraisemblablement pas par le hasard : la hausse du manque

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).

de sommeil avec le niveau est sérieuse. L’écart entre seconde et première (8,6 points) l’est aussi, mais moins nettement.

e) Si les répondants sont volontaires, l’échantillon n’est plus tiré au hasard : les élèves qui dorment mal sont peut-être plus enclins à répondre à une enquête sur le sommeil, ou au contraire moins disponibles. Ce biais de sélection peut fausser toutes les fréquences, quelle que soit la taille de l’échantillon. Les conclusions doivent être présentées avec cette réserve.

Réponse. a) 0,3 ; ≈ 0,386 ; 0,47 ; b) 0,38 ; c) non : 38 % au total ; d) oui, 17 points ≫ 3 points ; e) volontaires : risque de biais de sélection.

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).