

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

Échantillonnage — Première spé maths

Corrections détaillées

Boîte à outils

Les trois questions d'un problème d'échantillonnage.

1. Connaît-on p ? → **fluctuation** (on teste une hypothèse). 2. Cherche-t-on p ? → **confiance** (on estime). 3. La taille est-elle suffisante ? → dimensionnement en $\frac{1}{\delta^2}$.

Formules à mobiliser.

$$\sigma(X) = \sqrt{np(1-p)} \quad \sigma(F) = \sqrt{\frac{p(1-p)}{n}}$$

$$I_f = \left[p \pm \frac{2\sqrt{p(1-p)}}{\sqrt{n}} \right] \quad I_c = \left[f \pm \frac{1}{\sqrt{n}} \right]$$

Grille de critique d'une étude statistique.

À vérifier	Question à se poser
taille n	la marge $\frac{1}{\sqrt{n}}$ est-elle annoncée ?
représentativité	l'échantillon est-il vraiment aléatoire ?
formulation	« ne peut pas rejeter » ou « est vrai » ?
comparaisons	deux sondages : a-t-on tenu compte des deux fluctuations ?
sens du test	unilatéral ou bilatéral, décidé avant les données ?

Écart type d'une différence de deux fréquences (échantillons indépendants de même taille) :

$$\sigma(F_1 - F_2) = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{2} \sigma \quad \text{si } n_1 = n_2$$

Ce qu'une statistique ne dira jamais. Elle ne prouve pas une hypothèse, elle ne corrige pas un biais de sélection, et elle n'établit pas une **causalité** — seulement une corrélation ou un écart improbable.

Correction de l'exercice 1 - Sondage et marge d'erreur

- **Relation** : la marge vaut $\frac{1}{\sqrt{n}} = 0,03$.
- **Résolution** : $\sqrt{n} = \frac{1}{0,03} \approx 33,333$, donc

$$n \approx 1111$$
- **Réponse** : environ 1111 personnes — l'institut a très probablement interrogé 1000 ou 1100 personnes ✓
- **Contrôle pour $n = 1000$** : $\frac{1}{\sqrt{1000}} \approx 0,031623$, soit $\pm 3,16$ points, arrondi à ± 3 ✓
- **Intervalle de confiance** : $[0,35; 0,41]$.

- **Version affinée** pour $f = 0,38$ et $n = 1000$: $2\sqrt{\frac{0,38 \times 0,62}{1000}} \approx 2 \times 0,015347 \approx 0,030694$
— pratiquement la même marge ✓
- **Pourquoi les instituts utilisent 1000** : c'est le meilleur compromis entre coût et précision, la marge de ± 3 points étant jugée acceptable ✓
- **Marge pour** $n = 500$: $\pm 4,5$ points ; pour $n = 2000$: $\pm 2,2$ points ✓
- **Réflexe de lecture** : une marge de ± 3 points signifie que tout écart inférieur à 3 points entre deux sondages est **sans signification**.

Réponse. Environ 1111 personnes (soit un échantillon de 1000 à 1100).

Correction de l'exercice 2 - Contrôle de conformité

- **Hypothèse** : $p = 0,95$.
- **Paramètres** : $E(X) = 380$ et

$$\sigma(X) = \sqrt{400 \times 0,95 \times 0,05} = \sqrt{19} \approx 4,3589$$
- **Intervalle de fluctuation** ($\pm 2\sigma$) : $[380 - 8,72 ; 380 + 8,72] \approx [371,3 ; 388,7]$.
- **Observation** : $370 \notin [371,3 ; 388,7]$.
- **Conclusion** : on **rejette** l'hypothèse — la norme n'est probablement **pas** respectée ✓
- **En écarts types** : $\frac{370 - 380}{4,3589} \approx -2,294 \sigma$ ✓
- **Probabilité exacte** : $P(X \leq 370) \approx 0,019034$, soit 1,9 % ✓
- **Fréquence observée** : $\frac{370}{400} = 0,925$, contre un intervalle $[0,9282 ; 0,9718]$ ✓
- **Intervalle de confiance affiné** : $2\sqrt{\frac{0,925 \times 0,075}{400}} \approx 2 \times 0,013161 \approx 0,026322$, soit $[0,8987 ; 0,9513]$.
- **Nuance** : cet intervalle contient 0,95 **de très peu** — la conclusion est nette mais pas écrasante ✓
- **Remarque sur le sens du test** : ici seul le **défaut** de conformité importe, donc un test **unilatéral** serait plus approprié. Le seuil unilatéral à 5 % serait $380 - 1,645 \times 4,3589 \approx 372,8$, et $370 < 372,8$ conduirait au même rejet, avec plus de puissance ✓
- **Recommandation** : le taux réel est plutôt autour de 92,5 % ; il manque 10 pièces conformes sur 400 pour atteindre la norme.

Réponse. Non : $370 \notin [371,3 ; 388,7]$ ($-2,29 \sigma$, probabilité exacte 1,9 %).

Correction de l'exercice 3 - Essai clinique

- **Hypothèse à réfuter** : $p = 0,4$ (le nouveau traitement ne fait pas mieux).
- **Paramètres** : $E(X) = 48$ et

$$\sigma(X) = \sqrt{120 \times 0,4 \times 0,6} = \sqrt{28,8} \approx 5,3666$$
- **Test unilatéral** (on cherche si le nouveau est **meilleur**) : seuil à $E + 1,645\sigma \approx 48 + 8,83 = 56,83$.
- **Observation** : $58 > 56,83$.
- **Conclusion** : on **rejette** $p = 0,4$ en faveur de $p > 0,4$ — le nouveau traitement semble meilleur ✓
- **En écarts types** : $\frac{58 - 48}{5,3666} \approx 1,864 \sigma$ ✓
- **Probabilité exacte** : $P(X \geq 58) \approx 0,039215$, soit moins de 5 % ✓

- **Fréquence observée** : $\frac{58}{120} \approx 0,48333$.
- **Intervalle de confiance affiné** : $2\sqrt{\frac{0,48333 \times 0,51667}{120}} \approx 2 \times 0,045615 \approx 0,091230$, soit $[0,3921 ; 0,5745]$ — il contient 0,4, ce qui invite à la **prudence** ✓
- **Test bilatéral** : l'intervalle serait $[37,3 ; 58,7]$, qui contient 58 — pas de rejet ✓
- **Verdict honnête** : le résultat est **encourageant mais non concluant**. Avec $1,86\sigma$, il y a encore 3,6 % de chances d'observer cela par hasard.
- **Taille nécessaire** : pour détecter un passage de 0,40 à 0,48 avec une puissance de 80 %, il faudrait environ 600 patients ✓
- **Précaution méthodologique** : un essai clinique sérieux exige un **groupe témoin** simultané, pas une comparaison à un taux historique.

Réponse. Encourageant ($P(X \geq 58) \approx 0,039$ en unilatéral) mais non concluant — un test bilatéral ne rejeterait pas.

Correction de l'exercice 4 - Comparaison de deux groupes

- **Fréquences** : $f_A = \frac{130}{200} = 0,65$ et $f_B = \frac{110}{200} = 0,55$.
- **Écart observé** : 0,10.
- **Fréquence globale** (pour estimer σ) : $\frac{240}{400} = 0,6$.
- **Écart type d'une fréquence** : $\sqrt{\frac{0,6 \times 0,4}{200}} = \sqrt{0,0012} \approx 0,034641$.
- **Écart type de la différence** :

$$\sqrt{2} \times 0,034641 \approx 0,048990$$
- **En écarts types** : $\frac{0,10}{0,048990} \approx 2,041\sigma$ ✓
- **Conclusion** : l'écart est **tout juste significatif** au seuil de 5 % ✓
- **Erreur à éviter** : comparer 0,65 à l'intervalle de fluctuation de 0,55 seule — cela négligerait la fluctuation du groupe A ✓
- **Seuil de significativité** : il faut un écart supérieur à $2 \times 0,048990 \approx 0,098$, soit environ 10 points — on y est tout juste ✓
- **Intervalles de confiance** : $[0,5793 ; 0,7207]$ et $[0,4793 ; 0,6207]$ — ils se **chevauchent** entre 0,579 et 0,621 ✓
- **Nuance importante** : des intervalles de confiance qui se chevauchent n'impliquent **pas** une absence de différence significative — c'est un test plus fin qui tranche ✓
- **Avec des groupes de 500** : $\sigma(F_1 - F_2) \approx 0,030984$ et l'écart vaudrait $3,23\sigma$ — largement significatif ✓
- **Verdict** : différence à la limite ; il faudrait doubler les effectifs pour trancher sans hésiter.

Réponse. Tout juste significative : $2,04\sigma$ (seuil à environ 10 points d'écart).

Correction de l'exercice 5 - Dimensionner un contrôle industriel

- **Condition approximative** : les deux espérances doivent être séparées d'au moins 4σ (soit 2σ de chaque côté).
- **Espérances** : $0,02n$ et $0,04n$, d'écart $0,02n$.
- **Écart type** (sous $p = 0,02$) : $\sqrt{0,02 \times 0,98n} \approx 0,14\sqrt{n}$.
- **Condition** : $0,02n \geq 4 \times 0,14\sqrt{n}$, soit $0,02\sqrt{n} \geq 0,56$ et $\sqrt{n} \geq 28$.

$$n \geq 784$$

- **Réponse** : environ 800 pièces ✓
- **Vérification pour $n = 800$** : sous $p = 0,02$, $E(X) = 16$ et $\sigma \approx 3,9598$, donc l'intervalle est $[8,1 ; 23,9]$ — rejet si $X \geq 24$.
- **Sous $p = 0,04$** : $E(X) = 32$ et $\sigma \approx 5,5426$, donc 24 est à $\frac{24 - 32}{5,5426} \approx -1,443 \sigma$.
- **Puissance** : $P(X \geq 24) \approx 0,94270$, soit environ 94 % de détection ✓
- **Excellent compromis** : on détecte le doublement du taux dans 94 % des cas, avec seulement 5 % de fausses alertes ✓
- **Avec $n = 400$** : le seuil serait $E + 2\sigma = 8 + 2 \times 2,7999 \approx 13,6$, donc rejet si $X \geq 14$; sous $p = 0,04$, $E = 16$ et $\sigma \approx 3,9192$, donc $P(X \geq 14) \approx 0,73$ — puissance nettement moindre ✓
- **Coût du contrôle** : 800 pièces prélevées, éventuellement détruites — c'est le prix de la fiabilité, à mettre en regard du coût d'un lot défectueux livré.

Réponse. Environ 800 pièces (puissance de détection d'environ 94 %).

Correction de l'exercice 6 - Critique d'une conclusion

- **Hypothèse à tester** : $p = 0,4$ sur un échantillon de 50.
- **Paramètres** : $E(X) = 20$ et $\sigma(X) = \sqrt{50 \times 0,4 \times 0,6} = \sqrt{12} \approx 3,4641$.
- **Intervalle de fluctuation ($\pm 2\sigma$)** : $[20 - 6,93 ; 20 + 6,93] \approx [13,1 ; 26,9]$.
- **Observation** : $0,6 \times 50 = 30$ sportifs, et $30 \notin [13,1 ; 26,9]$.
- **Premier constat** : l'écart **est** statistiquement significatif ($\frac{30 - 20}{3,4641} \approx 2,887 \sigma$) ✓
- **Mais les critiques restent nombreuses** :
- **Taille** : 50 personnes seulement, d'où une marge de $\frac{1}{\sqrt{50}} \approx 0,141$, soit ± 14 points !
L'intervalle de confiance est $[0,459 ; 0,741]$ — très large ✓
- **Représentativité** : comment ces 50 sportifs ont-ils été choisis ? Un recrutement en salle de musculation biaiserait tout ✓
- **Définition** : que compte-t-on comme « complément » ? Une vitamine occasionnelle ?
- **Causalité** : même si l'écart est réel, rien ne dit que le sport **cause** la consommation ✓
- **Déclaratif** : les réponses sont auto-déclarées, donc sujettes à minimisation.
- **Conclusion mesurée** : l'écart est probablement réel, mais le mot « nettement » est excessif au vu de l'intervalle $[45,9 \% ; 74,1 \%]$ ✓
- **Reformulation correcte** : « la proportion est estimée à 60 %, avec une marge de ± 14 points ; elle est significativement supérieure aux 40 % de la population générale, sur cet échantillon dont la représentativité reste à établir ».
- **Amélioration du protocole** : interroger 500 sportifs tirés au sort sur les licences fédérales ramènerait la marge à $\pm 4,5$ points ✓

Réponse. L'écart est significatif ($2,89 \sigma$) mais l'intervalle $[0,459 ; 0,741]$ est très large, et la représentativité n'est pas établie.

Correction de l'exercice 7 - Estimation d'un effectif total

- **Principe (capture-recapture)** : la proportion de marqués dans l'échantillon estime celle de la population.
- **Fréquence observée** : $f = \frac{12}{300} = 0,04$.
- **Proportion théorique** : $\frac{200}{N}$ où N est la population.
- **Équation** : $\frac{200}{N} = 0,04$, donc

$$N = \frac{200}{0,04} = 5000 \text{ poissons}$$

- **Intervalle de confiance pour f** : $\left[0,04 - \frac{1}{\sqrt{300}}; 0,04 + \frac{1}{\sqrt{300}}\right] \approx [-0,0177; 0,0977]$ — la borne négative montre que cette version est **inutilisable** ici ✓
- **Version affinée** : $2\sqrt{\frac{0,04 \times 0,96}{300}} \approx 2 \times 0,011314 \approx 0,022627$, soit $[0,0174; 0,0626]$ ✓
- **Intervalle pour N** : $\left[\frac{200}{0,0626}; \frac{200}{0,0174}\right] \approx [3195; 11\,494]$ ✓
- **Conclusion honnête** : la population est estimée à 5000 poissons, mais très imprécisément — entre 3200 et 11 500 ✓
- **Cause de l'imprécision** : seulement 12 poissons marqués retrouvés. La précision relative dépend du **nombre de succès**, pas de la taille de l'échantillon ✓
- **Amélioration** : pour une précision de $\pm 20\%$, il faudrait environ 100 recaptures, donc pêcher 2500 poissons ✓
- **Contrôle** : $2\sqrt{\frac{0,04 \times 0,96}{2500}} \approx 0,00784$, soit $[0,0322; 0,0478]$ et $N \in [4184; 6211]$ ✓
- **Hypothèses du modèle** : mélange complet des marqués, pas de mortalité ni de migration entre les deux pêches, égale capturabilité — autant de conditions rarement parfaitement remplies ✓

Réponse. $N \approx 5000$ poissons, mais avec un intervalle très large $[3200; 11\,500]$.

Correction de l'exercice 8 - Test sur un intervalle asymétrique

- **Paramètres** : $E(X) = 15$ et $\sigma(X) = \sqrt{300 \times 0,05 \times 0,95} = \sqrt{14,25} \approx 3,7749$.
- **Approximation** : $[15 - 7,55; 15 + 7,55] \approx [7,45; 22,55]$, soit $[8; 22]$ en entiers.
- **Contrôle exact** : $P(X \leq 7) \approx 0,015968$ et $P(X \geq 23) \approx 0,029183$.
- **Somme** : $\approx 0,045151$, soit un niveau réel de $95,5\%$ ✓
- **Asymétrie** : la probabilité de la queue **haute** ($2,9\%$) dépasse celle de la queue **basse** ($1,6\%$) ✓
- **Explication** : pour p petit, la loi binomiale est **asymétrique** — elle est bornée à gauche par 0 mais s'étale à droite ✓
- **Intervalle exact optimal** : $[8; 22]$ reste le meilleur choix en entiers, mais l'asymétrie signale que l'approximation normale est imparfaite.
- **Contrôle de validité** : $np = 15 \geq 5$ et $n(1 - p) = 285 \geq 5$ ✓ — l'approximation est acceptable, sans être excellente.
- **Comparaison avec $p = 0,5$** : pour $n = 300$, l'intervalle serait $[133,7; 166,3]$ et les deux queues seraient parfaitement **symétriques** ✓
- **Conséquence pratique** : pour de petites proportions, un test bilatéral par $\pm 2\sigma$ est légèrement trop **indulgent** du côté haut ; mieux vaut calculer les queues exactes.
- **Ordre de grandeur de l'erreur** : le vrai niveau est $95,5\%$ au lieu de 95% — négligeable pour une conclusion tranchée, gênant pour un cas limite ✓

Réponse. $[8; 22]$ dans les deux cas, mais le niveau réel est $95,5\%$ et les queues sont asymétriques ($1,6\% / 2,9\%$).

Correction de l'exercice 9 - Sondage à deux tours

- **Attention** : la question porte sur p connaissant f — c'est une question d'**estimation**, non de fluctuation ✓
- **Écart type** : $\sigma \approx \sqrt{\frac{0,51 \times 0,49}{1000}} \approx 0,015808$.

- **Écart entre f et 0,5** : 0,01, soit $\frac{0,01}{0,015808} \approx 0,633\sigma$.
- **Ordre de grandeur de la réponse** : la probabilité que $p < 0,5$ est d'environ 26,3 % — soit plus d'une chance sur quatre ✓
- **Interprétation** : le candidat a environ **trois chances sur quatre** d'être majoritaire, ce qui est loin d'une certitude ✓
- **Intervalle de confiance** : $\left[0,51 \pm \frac{1}{\sqrt{1000}}\right] \approx [0,4784; 0,5416]$ — il contient largement 0,5 ✓
- **Conclusion** : le sondage **ne permet pas** de départager ✓
- **Écart nécessaire pour une quasi-certitude** : il faudrait 2σ , soit environ 3,2 points au-dessus de 50 %, donc $f \geq 53,2\%$ ✓
- **Effet de la taille** : avec $n = 10\,000$ et la même fréquence 0,51, on aurait $\sigma \approx 0,005$ et un écart de 2σ — la conclusion « majoritaire » deviendrait solide ✓
- **Précision sur la rigueur** : parler de « probabilité que $p < 0,5$ » relève de l'approche bayésienne et dépasse le programme. Au lycée, on se contente de constater que 0,5 est dans l'intervalle de confiance.
- **Leçon électorale** : 51 % contre 49 % dans un sondage de 1000 personnes est un **match nul** statistique.

Réponse. Environ une chance sur quatre d'être minoritaire : le sondage ne départage pas ($0,5 \in [0,478; 0,542]$).

Correction de l'exercice 10 - Échantillon stratifié

- **Problème** : l'échantillon comporte 80 % d'urbains au lieu de 60 % — il n'est **pas représentatif** ✓
- **Conséquence** : si les opinions diffèrent entre urbains et ruraux, la fréquence globale sera **biaisée** vers l'opinion urbaine ✓
- **Correction par redressement** : on pondère chaque sous-groupe selon son poids réel.

$$f_{\text{corrigée}} = 0,6f_{\text{urbains}} + 0,4f_{\text{ruraux}}$$
- **Exemple numérique** : si $f_u = 0,55$ et $f_r = 0,35$, la fréquence brute vaut $0,8 \times 0,55 + 0,2 \times 0,35 = 0,44 + 0,07 = 0,51$.
- **Fréquence corrigée** : $0,6 \times 0,55 + 0,4 \times 0,35 = 0,33 + 0,14 = 0,47$ ✓
- **Biais évité** : 4 points — considérable pour un sondage électoral ✓
- **Coût du redressement** : les 200 ruraux ont un poids doublé, donc leur fluctuation compte davantage.
- **Écart type après redressement** :

$$\sqrt{0,6^2 \times \frac{0,55 \times 0,45}{800} + 0,4^2 \times \frac{0,35 \times 0,65}{200}} \approx \sqrt{1,1138 \times 10^{-4} + 1,8200 \times 10^{-4}} \approx 0,017128$$
 ✓
- **Comparaison** : un échantillon **bien stratifié** (600 urbains, 400 ruraux) donnerait

$$\sqrt{0,36 \times \frac{0,2475}{600} + 0,16 \times \frac{0,2275}{400}} \approx \sqrt{1,4850 \times 10^{-4} + 9,1000 \times 10^{-5}} \approx 0,015363$$
 — un peu mieux ✓
- **Leçon** : le redressement **corrige** le biais mais **dégrade** la précision. Mieux vaut construire l'échantillon correctement dès le départ ✓
- **Pratique réelle** : les instituts utilisent la **méthode des quotas** sur plusieurs critères croisés (âge, sexe, région, profession).

Réponse. Échantillon non représentatif ; on redresse par $0,6f_u + 0,4f_r$, au prix d'une légère perte de précision.

Correction de l'exercice 11 - Contrôle séquentiel

- **Premier prélèvement** : sous $p = 0,03$, $E(X) = 3$ et $\sigma(X) = \sqrt{100 \times 0,03 \times 0,97} = \sqrt{2,91} \approx 1,7059$.
- **Intervalle de fluctuation** : $[3 - 3,41; 3 + 3,41] \approx [0; 6,4]$, soit acceptation si $X \leq 6$.
- **Zone ambiguë proposée** : $[4; 8]$ — elle chevauche la zone d'acceptation (≤ 6) et la zone de rejet (≥ 7) ✓
- **Probabilités** : $P(4 \leq X \leq 8) \approx 0,34953$ — plus d'un tiers des lots conformes seraient renvoyés au second prélèvement ✓
- **Coût** : 34 % de contrôles doublés, ce qui est lourd.
- **Second prélèvement** : sur 200 pièces au total, $E(X) = 6$ et $\sigma \approx 2,4125$, donc l'intervalle devient $[1,2; 10,8]$ ✓
- **Avantage du protocole séquentiel** : il tranche les cas nets rapidement (économie de contrôles) et ne double l'effort que pour les cas douteux ✓
- **Inconvénient statistique majeur** : les **deux** tests successifs cumulent chacun leur risque de 5 % — le risque global de rejet à tort dépasse 5 % ✓
- **Ordre de grandeur du risque cumulé** : environ $1 - 0,95^2 = 0,0975$, soit près de 10 % si l'on applique naïvement le même seuil deux fois.
- **Correction nécessaire** : il faut **durcir** le seuil de chaque étape (par exemple à 2,5 % chacune) pour que le risque global reste à 5 % ✓
- **Nom** : c'est le problème des **tests multiples**, et la correction porte le nom de **correction de Bonferroni**.
- **Verdict sur le protocole** : l'idée est bonne, la zone $[4; 8]$ est trop large, et les seuils doivent être recalculés ✓

Réponse. Zone ambiguë trop large (34 % des lots) et risque cumulé proche de 10 % : il faut durcir les seuils.

Correction de l'exercice 12 - Précision et budget

- **Tailles nécessaires** : $n = \frac{1}{\delta^2}$.
- **Pour ± 5 points** : $n = 400$, coût $2000 + 1200 = 3200$ €.
- **Pour ± 3 points** : $n = 1112$, coût $2000 + 3336 = 5336$ €.
- **Pour ± 2 points** : $n = 2500$, coût $2000 + 7500 = 9500$ €.
- **Comparaison** : passer de ± 5 à ± 2 points **triple** le budget ✓
- **Coût marginal de la précision** : gagner les 2 derniers points (de ± 3 à ± 2) coûte 4164 €, alors que gagner les 2 premiers (de ± 5 à ± 3) n'en coûtait que 2136 € ✓
- **Loi du rendement décroissant** : le coût croît comme $\frac{1}{\delta^2}$, donc chaque point de précision supplémentaire coûte de plus en plus cher ✓
- **Pour ± 1 point** : $n = 10\,000$, coût 32 000 € — soit dix fois le premier scénario ✓
- **Marge pour un budget de 5000 €** : $n = \frac{3000}{3} = 1000$, donc $\pm 3,2$ points ✓
- **Optimum pratique** : le standard de 1000 personnes (≈ 5000 €) correspond au point où la précision devient « suffisante » pour la plupart des usages ✓
- **Contrôle** : $\frac{1}{\sqrt{1000}} \approx 0,0316$ ✓
- **Arbitrage réel d'un commanditaire** : mieux vaut souvent **deux** sondages de 1000 personnes à deux dates (10 000 €) qu'un seul de 2500 — on gagne l'information sur l'**évolution**, plus précieuse qu'un point de précision.

Réponse. 3200 €, 5336 € et 9500 € — le coût croît comme $\frac{1}{\delta^2}$.

Correction de l'exercice 13 - Décision avec deux hypothèses

- **Fréquence observée** : $f = \frac{76}{200} = 0,38$.
- **Sous** $p = 0,30$: $E(X) = 60$ et $\sigma(X) = \sqrt{200 \times 0,3 \times 0,7} = \sqrt{42} \approx 6,4807$.
- **Écart** : $\frac{76 - 60}{6,4807} \approx 2,469 \sigma$ — **hors** de l'intervalle $[47,0 ; 73,0]$, donc on **rejette** ✓
- **Sous** $p = 0,45$: $E(X) = 90$ et $\sigma(X) = \sqrt{200 \times 0,45 \times 0,55} = \sqrt{49,5} \approx 7,0356$.
- **Écart** : $\frac{76 - 90}{7,0356} \approx -1,990 \sigma$ — **tout juste dans** l'intervalle $[75,9 ; 104,1]$, donc on ne rejette pas ✓
- **Conclusion** : on retient $p = 0,45$, mais **de très peu** ✓
- **Rapport de vraisemblance** : $P(X = 76)$ vaut environ 0,00314 sous $p = 0,3$ et 0,00781 sous $p = 0,45$, soit un rapport de 2,49 ✓
- **Interprétation** : $p = 0,45$ est environ **deux fois et demie** plus vraisemblable — un avantage réel mais modeste.
- **Position de l'observation** : 0,38 est presque **à mi-chemin** entre 0,30 et 0,45, d'où l'ambiguïté ✓
- **Contrôle** : le milieu est 0,375, et l'on observe 0,38 ✓
- **Taille nécessaire pour trancher** : pour séparer 0,30 et 0,45 à 4σ , il faut $0,15 \geq \frac{4 \times 0,5}{\sqrt{n}}$, soit $n \geq 178$ — la taille est théoriquement suffisante, mais l'observation tombe pile au milieu ✓
- **Recommandation** : refaire 200 tirages. Une fréquence à nouveau proche de 0,38 suggérerait qu'**aucune** des deux hypothèses n'est correcte.

Réponse. $p = 0,45$, mais de justesse ($-1,99 \sigma$ contre $+2,47 \sigma$) ; l'observation est à mi-chemin.

Correction de l'exercice 14 - Interprétation d'un taux de réponse

- **Fréquence parmi les répondants** : $\frac{210}{300} = 0,7$.
- **Taux de réponse** : $\frac{300}{2000} = 15\%$ — très faible ✓
- **Problème majeur** : les répondants ne sont **pas** un échantillon aléatoire de la population, mais un échantillon **auto-sélectionné** ✓
- **Biais de non-réponse** : les personnes concernées ou motivées répondent davantage. Si les favorables sont plus motivés, 0,7 **surestime** la vraie proportion ✓
- **Ce que l'on peut dire** : la proportion de favorables **parmi les répondants** est estimée à 70 %, avec une marge de $\frac{1}{\sqrt{300}} \approx 0,0577$, soit $[0,642 ; 0,758]$ ✓
- **Ce que l'on ne peut PAS dire** : que 70 % de la population est favorable ✓
- **Encadrement extrême** : si les 1700 non-répondants étaient tous défavorables, la proportion réelle serait $\frac{210}{2000} = 0,105$; s'ils étaient tous favorables, $\frac{1910}{2000} = 0,955$ ✓
- **Amplitude de l'incertitude réelle** : $[0,105 ; 0,955]$ — l'information est donc très faible ✓
- **Comparaison éclairante** : la marge statistique ($\pm 5,8$ points) est **ridicule** devant l'incertitude due à la non-réponse (± 42 points) ✓
- **Correction possible** : relancer les non-répondants, ou effectuer un sondage **téléphonique** sur un sous-échantillon aléatoire des non-répondants pour estimer leur

opinion ✓

- **Leçon** : un taux de réponse faible ruine toute inférence, quelle que soit la taille de l'envoi.

Réponse. Non : auto-sélection des répondants. On ne peut estimer que la proportion **parmi les répondants**, soit $[0,642 ; 0,758]$.

Correction de l'exercice 15 - Test répété et faux positifs

- **Risque par test** : 5 % de rejeter à tort un dé équilibré.
- **Nombre attendu de faux positifs** :

$$20 \times 0,05 = 1$$
- **Réponse** : un dé en moyenne sera déclaré truqué à tort ✓
- **Loi du nombre de faux positifs** : $B(20 ; 0,05)$, avec $E = 1$ et
 $\sigma = \sqrt{20 \times 0,05 \times 0,95} \approx 0,97468$.
- **Probabilité d'au moins un faux positif** : $1 - 0,95^{20} \approx 1 - 0,35849 = 0,64151$ ✓
- **Interprétation** : dans près de **deux cas sur trois**, au moins un dé sera injustement accusé ✓
- **Probabilité de deux faux positifs ou plus** :
 $1 - 0,35849 - 20 \times 0,05 \times 0,95^{19} \approx 1 - 0,35849 - 0,37735 = 0,26416$ ✓
- **Phénomène** : c'est le problème des **tests multiples** — multiplier les tests multiplie les fausses alertes ✓
- **Correction de Bonferroni** : pour garder un risque global de 5 %, il faut tester chaque dé au seuil de $\frac{0,05}{20} = 0,0025$, soit environ $\pm 3\sigma$ ✓
- **Contrôle** : avec 20 tests à 0,25 %, le risque global vaut $1 - 0,9975^{20} \approx 0,0489$, bien proche de 5 % ✓
- **Coût de la correction** : le seuil passe de 2σ à 3σ , ce qui **réduit la puissance** — on détectera moins bien les dés réellement truqués ✓
- **Application réelle** : ce phénomène explique de nombreux résultats scientifiques non reproductibles, quand des dizaines d'hypothèses sont testées sur un même jeu de données.

Réponse. Un dé en moyenne, et au moins un dans 64 % des cas — d'où la nécessité de corriger le seuil (Bonferroni).

Correction de l'exercice 16 - Synthèse : audit complet

a) **Test.** Sous $p = 0,98$:

$$E(X) = 588 \quad \sigma(X) = \sqrt{600 \times 0,98 \times 0,02} = \sqrt{11,76} \approx 3,4293$$

- **Intervalle de fluctuation** ($\pm 2\sigma$) : $[588 - 6,86 ; 588 + 6,86] \approx [581,1 ; 594,9]$.
- **Observation** : $576 \notin [581,1 ; 594,9] \rightarrow$ **rejet** ✓
- **En écarts types** : $\frac{576 - 588}{3,4293} \approx -3,499 \sigma$ — un écart **très significatif** ✓
- **Probabilité exacte** : $P(X \leq 576) \approx 0,00131$, soit environ une chance sur 770 ✓

b) **Estimation.** $f = \frac{576}{600} = 0,96$.

- **Version affinée** : $2\sqrt{\frac{0,96 \times 0,04}{600}} \approx 2 \times 0,008000 \approx 0,016000$, soit
 $[0,944 ; 0,976]$

- **Conclusion** : le vrai taux est très probablement entre 94,4 % et 97,6 % — l'affirmation de 98 % est **exclue** ✓
- **Version simple** : $\left[0,96 \pm \frac{1}{\sqrt{600}} \right] \approx [0,9192; 1,0008]$ — inutilisable (elle dépasse 1) ✓
- c) **Trancher entre 0,96 et 0,98** demande une demi-amplitude inférieure à 0,01.
- **Avec** $f \approx 0,97$: $2\sqrt{\frac{0,97 \times 0,03}{n}} \leq 0,01$ donne $\frac{0,34118}{\sqrt{n}} \leq 0,01$, soit $\sqrt{n} \geq 34,118$ et

$$n \geq 1164$$
- **Réponse** : environ 1200 commandes ✓
- **Contrôle** : pour $n = 1200$, la demi-amplitude vaut $\frac{0,34118}{34,641} \approx 0,00985$ ✓

d) **Conclusion rigoureuse.**

- « Sur un audit de 600 commandes, le taux de livraison à l'heure est estimé à 96,0 %, avec un intervalle de confiance à 95 % de [94,4 % ; 97,6 %]. L'affirmation de 98 % est rejetée au seuil de 95 % (écart de 3,5 écarts types, probabilité de 0,13 %). »
- **Réserves à mentionner** : les 600 commandes ont-elles été tirées au hasard ? Couvrent-elles toutes les périodes (les fêtes de fin d'année dégradent les délais) et toutes les zones géographiques ? ✓
- **Ce que l'audit ne dit pas** : la **cause** du déficit de 2 points, ni s'il est stable dans le temps ✓
- **Impact chiffré** : 2 points sur, disons, 500 000 commandes annuelles, cela représente 10 000 livraisons tardives de plus qu'annoncé — l'écart statistiquement « petit » est commercialement **majeur** ✓

Réponse. a) rejet ($-3,50\sigma$) · b) $[0,944; 0,976]$ · c) environ 1200 commandes · d) 98 % réfuté, sous réserve de représentativité.