

Rappels de statistiques ; nuage de points et corrélation

EXERCICE 1: [Indications] - Cours - Calculer -

L'espérance de vie à la naissance des français (hors Mayotte) est résumée approximativement ici de 1950 à 2015 (données suivant INSEE). (Les années, en ordre croissant, sont résumées par les deux derniers chiffres.)

Année	50	55	60	65	70	75	80	85	90	95	00	05	10	15
Hommes	63	65.5	67	67.5	68.5	69	70	71.5	73	74	75.5	77.5	78	80
Femmes	69	71.5	74	75	76	77	78.5	79.5	81	82	82.5	84	84.5	85.5

On note A , H et F les différentes valeurs respectives correspondant aux années, à l'espérance de vie des hommes et à celle des femmes.

1. a) Déterminer les moyennes et les écarts-type de F et H .
b) Déterminer $\sigma_{A,H}$ la covariance de A , H ainsi que $\sigma_{A,F}$ et expliquer son signe.
c) En déduire les coefficients de corrélation $r_{A,H}$ et $r_{A,F}$.
2. a) Déterminer les droites d'ajustement linéaire $y = ax + b$ et $y = a'x + b'$ exprimant respectivement H en fonction de A et F en fonction de A .
b) Ces droites sont-elles pertinentes ?
c) Peut-on en déduire une espérance de vie éventuelle pour 2030 ?
3. a) Y-a-t-il un rapport entre l'évolution des hommes et des femmes ?
b) Déterminer le coefficient directeur d'une des deux droites d'ajustement. Qu'en déduire ?

EXERCICE 2: [Indications] - Calculer - Chercher -

D'après le rapport de concours de 2021, (période où l'oral de maths au concours Agro était réparti en 2 matières maths/info distinctes) la répartition des notes (en pourcentage de candidats admissibles) était la suivante : (les notes sont sur 10 pour chaque partie Mathématiques et Informatique, et sur 20 au total.)

classe	[0; 2[	[2; 4[	[4; 6[	[6; 8[	[8; 10]
Maths	10	26	30	23	11
Info	3	21	37	31	9

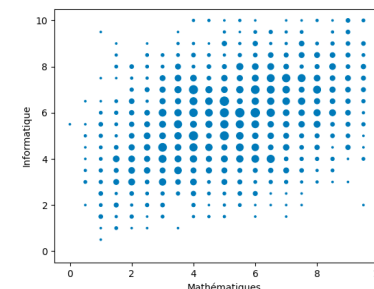
et

classe	[0; 4[	[4; 8[	[8; 12[	[12; 16[	[16; 20]
Note Finale	2	20	42	29	7

1. a) Calculer la moyenne, l'écart-type, la classe médiane et une médiane (on supposera que les effectifs sont répartis uniformément à l'intérieur de chaque classe).
Les résultats annoncés par le concours sont les suivants :

	Moyenne	Écart-type	médiane
Maths	5,28	2,24	5
Info	5,72	1,85	6
Final	11	3,44	11

- b) Expliquez la différence avec les moyennes et écart-types obtenus dans le calcul précédent.
- c) Commentez les résultats sur la médiane par-rapport à ceux obtenus.
2. Le nuage suivant donne la corrélation entre les notes de mathématiques et d'informatique. La taille du cercle de centre (x, y) est proportionnelle au nombre de candidats ayant eu la note x en mathématiques et y en Informatique.



- a) Le coefficient de corrélation est-il, à votre avis, très proche de ± 1 ?
- b) Le coefficient de corrélation est-il positif ou négatif ?
- c) Les candidats ayant obtenu une bonne note en Maths ont-ils généralement obtenus une bonne note en Informatique ?
- d) Les candidats ayant obtenu une bonne note en Informatique ont-ils généralement obtenus une bonne note en maths ?

EXERCICE 3: [Indications] – Mobiliser – Modéliser –

☆

Montrer en utilisant l'inégalité de Bienaymé-Tchebychev que, pour tout $x > 0$,

$$\frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du \geq \frac{1}{2} \left(1 - \frac{1}{x^2}\right)$$

Pour quels x cette inégalité a-t-elle un sens ?

EXERCICE 4: [Indications] – Cours – Mobiliser – Calculer –

Une urne contient une proportion inconnue de boules blanches. On y effectue une série de n tirages d'une boule avec remise et on approche p par la proportion Y_n de boules blanches obtenue dans les n tirages.

- Montrer que, pour tout $\epsilon > 0$, on a $P(|Y_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$.
- En déduire une condition suffisante sur n pour que l'approximation de p par Y_n donne une valeur approchée de p à 0,01 près avec une probabilité supérieure ou égale à 95%.
- a) Montrer qu'en utilisant le théorème central limite, on trouve comme condition suffisante

$$n \geq 196^2 p(1-p)$$

- En prenant n_0 minimal vérifiant la condition, montrer que $n_0 \leq 9604$. Commentez.

Approximations

EXERCICE 5: [Indications] – Cours – Modéliser – Calculer – Dans les ruches de Robert, la longévité d'une abeille ouvrière née au printemps suit une variable aléatoire T définie par la densité de probabilités suivante :

$$\forall t \geq 0, \quad f(t) = \lambda t^2 e^{-\alpha t}$$

- Sachant que la longévité moyenne est de 50 jours, calculer λ et α .
- Soit Z la longévité moyenne des 2000 ouvrières nées le premier mai. Déterminer la valeur de t pour laquelle :

$$P(|Z - 50| \geq t) \leq 0,01$$

- en utilisant l'inégalité de Bienaymé-Tchebychev.
- en effectuant une approximation grâce au TCL.

EXERCICE 6: [Indications] – Modéliser – Cours –

Les fourmis rouges du haut plateau du Katanga ont les yeux bleus, excepté les albinos qui ont les yeux jaunes. Chaque fourmi a une chance sur dix d'être albinos. Un éléphant vient s'abreuver au bord du lac et écrase malencontreusement 453 fourmis rouges.

Soit X le nombre de fourmis aux yeux jaunes victimes du pachyderme.

- Quelle est la loi de probabilité de X ? Déterminer $\mathbb{E}[X]$ et $\sigma(X)$.
- Par quelle loi de probabilité peut-on approcher la loi de X ?
- a) Calculer la probabilité $P(20 < X < 50)$
b) Déterminer un intervalle I fermé et centré au tour de $E[X]$ tel que $P(X \in I) \simeq 0,95$.

EXERCICE 7: [Indications] – Modéliser – Cours –

Dans l'eau rejetée par une centrale nucléaire, un pisciculteur daltonien croit élever des poissons rouges. En réalité, 2% de cette espèce de poissons sont verts. Un incident de fonctionnement du réacteur provoque une brusque élévation de la température qui est fatale à certains poissons.

Le pisciculteur ramasse 150 poissons morts au hasard pour en mesurer la radioactivité. Soit X la variable aléatoire égale au nombre de poissons verts de l'échantillon.

- Quelle est la loi de probabilité de X ? Déterminer son espérance.
- Par quelle(s) probabilité(s) peut-on approcher la loi de X ?
- Calculer $P(X > 3)$ en utilisant l'approximation précédente. (Vérifiez à la calculatrice ou avec Python si ce résultat est proche du résultat sans approximation.)

EXERCICE 8: [Indications] – Modéliser – Calculer –

Lors d'une soirée d'Halloween, un couple a décidé d'offrir au choix des bonbons au chocolat et au citron. On admet qu'un enfant a une probabilité de 0,6 d'avoir envie d'un bonbon au chocolat (et qu'il ne peut choisir qu'un seul bonbon). Sachant qu'il y a 200 enfants dans le quartier, en utilisant une approximation de loi bien choisie, combien faut-il acheter au minimum de bonbons au chocolat pour avoir

- 90% de chances de contenter tous les enfants qui aiment le chocolat ?
- 99% de chances de contenter tous les enfants qui aiment le chocolat ?
- Et en ce qui concerne les bonbons au citrons ?
- Commenter.

EXERCICE 9: [Indications] – Modéliser – Calculer –

Les "fortune Cookies" contiennent des messages qui annoncent la richesse une fois sur 10. Combien doit-on acheter de ces petits gateaux pour avoir 90% de chance de tomber au moins 5 fois sur l'annonce de la fortune ?

EXERCICE 10: [Indications] – Chercher – Modéliser –

☆☆

Soit $0 < p < 1$, $p + q = 1$.

- Montrer que si $p > q$, alors, pour n assez grand,

$$\sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} p^k (1-p)^{n-k} \simeq 1$$

- Supposons qu'une élection oppose deux candidats A et B . Chaque électeur n'ayant pas voté blanc a une probabilité p de voter pour A et une probabilité q de voter pour B . Interpréter le résultat de la question précédente par-rapport à cette situation.

EXERCICE 11: [Indications] – Chercher – Mobiliser –

Soit n un entier naturel non nul et λ un nombre réel fixé. On pose

$$P_n(\lambda) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}$$

Soient $X_1, \dots, X_n$ n variables aléatoires définies sur le même espace probabilisé, indépendantes et suivant toutes la loi de Poisson de paramètre 1.

1. Exprimer $P_n(n)$ en fonction de la fonction de répartition de $S_n = X_1 + \dots + X_n$.
2. A l'aide du théorème de la limite centrée, démontrer que $\lim_{n \rightarrow \infty} P_n(n) = \frac{1}{2}$.

Complément : convergence en loi

EXERCICE 12: [Indications] – Cours – Raisonner – Calculer –

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de lois uniformes sur $[0; 1]$. Soit N_n une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ et indépendante des X_i . Soit $T_n = n \min_{1 \leq i \leq N_n} X_i$ et $X \hookrightarrow \mathcal{E}(p)$. On note F_n la fonction de répartition de T_n et F la fonction de répartition de X .

1. Calculer $F_n(x)$ pour tout $x \in \mathbb{R}$.
2. Montrer que pour tout $x \in \mathbb{R}$, $F_n(x) \xrightarrow{n \rightarrow +\infty} F(x)$. (Ceci s'appelle la convergence en loi.)
3. En déduire que pour tout $a, b \in \mathbb{R}$, $P(a < T_n \leq b) \xrightarrow{n \rightarrow +\infty} P(a < X \leq b)$

EXERCICE 13: [Indications] – Cours – Raisonner – Calculer –

Chaque côté de la galère royale est occupé par n rameurs. Ces galériens rament sans arrêt, jusqu'à épuisement de toute l'équipe, pendant X_n jours.

La variable aléatoire X_n suit une loi exponentielle de paramètre $\frac{1}{n}$. (Attention, X_n n'est pas forcément un nombre entier !)

La galère quitte Toulon pour un long voyage le 31 du mois d'août à minuit. On note Y_n l'heure de la journée à laquelle les galériens cessent de ramer, traduite en jour décimal. (par exemple, midi = 0.5 jour ...)

1. Justifier que

$$Y_n \leq t \iff (\text{il existe } k \in \mathbb{N} \text{ tel que } k \leq X_n \leq k + t)$$

2. Montrer que la fonction de répartition F_n de Y_n est définie par :

$$F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{1 - e^{-\frac{t}{n}}}{1 - e^{-\frac{1}{n}}} & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases}$$

3. En déduire que Y_n converge en loi vers une variable aléatoire de loi uniforme sur $[0; 1]$.

test de conformité

EXERCICE 14: [Indications] – Cours – Calculer – Raisonner –

Le propriétaire d'un bar soupçonne que sa machine à café est déréglée. Sa machine ne délivre jamais exactement la même quantité de café, mais ses expressos devraient faire en moyenne 40ml. Au service du midi, il mesure chacun des $n = 50$ cafés apportés au client. On note v le volume moyen en ml que délivre en réalité la machine. On pose les hypothèses suivantes :

$$\mathcal{H}_0 : "v = 40" \quad ; \quad \mathcal{H}_1 : "v \neq 40"$$

On note V_n le volume empirique moyen observé sur les n cafés et S_n la variance observée.

1. Sous l'hypothèse $\mathcal{H}_0$, déterminer x tel que

$$P\left(\left|\frac{V_n - 40}{\sqrt{S_n}/\sqrt{n}}\right| \leq x\right) \simeq 0.95$$

2. Les mesures prises expérimentalement donnent $S_n = 2$, En déduire un intervalle I sous la forme $I = [40 - a, 40 + a]$ tel que

$$P(V_n \in I) \simeq 0.95$$

3. Les données expérimentales donnent également $V_n = 39$. Peut-on conclure sur le fait que la machine est déréglée ?

EXERCICE 15: [Indications] – Cours – Calculer – Raisonner –

On souhaite tester si une pièce est truquée. On note p_0 la probabilité pour cette pièce de tomber sur "Pile". On pose les hypothèses suivantes :

$$\mathcal{H}_0 : "p_0 = \frac{1}{2}" \quad ; \quad \mathcal{H}_1 : "p_0 \neq \frac{1}{2}"$$

On note F_n la proportion empirique de "Pile" observée dans une série de n lancers.

1. Sous l'hypothèse $\mathcal{H}_0$, déterminer x tel que

$$P\left(\left|\frac{F_n - \frac{1}{2}}{\sqrt{\frac{1}{4}}/\sqrt{n}}\right| \leq x\right) \simeq 0.95$$

2. En déduire a tel que

$$P\left(F_n \in \left[\frac{1}{2} - a, \frac{1}{2} + a\right]\right) \simeq 0.95$$

3. a) Les données expérimentales sont 29 piles. Peut-on conclure sur le fait que la pièce est truquée ?
b) Un autre dé donne 32 Piles. Peut-on conclure ?

EXERCICE 16: [Indications] – Cours – Modéliser –

Dans un bar, le manager aimerait connaître la probabilité p pour un client de laisser un pourboire. Il fait une étude sur une journée de 100 consommateurs. On note X_i la variable de Bernoulli dont le succès est "le $i^{\text{ème}}$ client laisse un pourboire." et X le nombre total de client qui laissent un pourboire.

Le serveur constate que sur le 100-échantillon de clients, 38 ont laissé un pourboire.

1. Donner une variable aléatoire qui approche p (justifier vos dires) ?
2. Déterminer la moyenne et l'écart-type empiriques des X_i .
3. a) Déterminer un nombre u tel que $P\left(\overline{X}_n - u \frac{S_n}{\sqrt{n}} \leq p \leq \overline{X}_n + u \frac{S_n}{\sqrt{n}}\right) = 0,99$
 b) En déduire un intervalle I tel que $P(p \in I) = 0,99$.
 c) Chercher de même un intervalle I tel que $P(p \in I) = 0,95$.
4. Peut-on conclure que la proportion de clients laissant un pourboire est de 50% ?
5. On raisonne maintenant avec une confiance de 95% et On se place le lendemain de la journée test. On note X le nombre de clients laissant un pourboire sur les $n = 120$ clients. ☆
 a) Encadrez la probabilité pour que moins de 100 clients laissent un pourboire (et en déduire une valeur approchée de cette probabilité)
 b) Encadrez de même s'il s'agit de moins de la moitié des clients.

EXERCICE 17: [Indications] – Cours – Calculer –

On souhaite comparer les tailles d'un même type de plants de tomates effectués sur des sols de différentes natures.

On désigne par A, B, C, D les populations constituées par les plants sur le sol de type respectifs 1,2,3,4. Constitués chacun de plus de 10 000 plants. On note X_i la variable statistique : *taille de l'individu (en mm) i à 5 mois après plantation.*

L'ensemble des questions suivantes sera résolu en utilisant une approximation bien choisie et en justifiant les théorèmes utilisés.

1. Calculs sur la population A :
 Les données sur la population A sont déjà connues. On sait que la moyenne et l'écart type (théoriques) de la taille d'un plant de A sont : $\mu_A = 570$ et $\sigma_A = 152$.
 On mesure au hasard dans A un échantillon de $n = 250$ individus. Soit M_n la moyenne aléatoire associée à X .
 a) Calculer la probabilité d'avoir une moyenne M_n supérieure ou égale à 550.
 b) Déterminer le nombre h tel que $P[M_n > h] = 0,95$.
 c) Déterminer le nombre a tel que $P(|M_n - \mu_A| < a) = 0,95$ et en déduire l'intervalle I_A centré autour de μ_A pour lequel $P(M_n \in I_A) = 0,95$.
 (Intervalle de confiance de M_n de niveau de confiance 0,95.)

- d) Déterminer de même l'intervalle de confiance I'_A pour la probabilité $p = 0,90$.
- e) Quelle devrait être la taille minimale de l'échantillon à considérer pour que M_n soit une estimation de μ_A à 10mm près avec une confiance $\alpha = 0,99$?

2. La population B :

Dans la population B , la taille moyenne empirique d'un échantillon $\mathcal{E}_B$ de 250 individus est $m_B = 555$.

Peut-on dire a priori qu'il y a une différence significative entre les tailles moyennes des deux catégories A et B ? (Utiliser c)

3. La population C :

La taille moyenne μ_C dans la population C n'est pas connue. Pour l'estimer, on a choisi au hasard un échantillon $\mathcal{E}_C$. Le tableau suivant représente la distribution statistique sur $\mathcal{E}_C$ de la taille en mm.

x	100 – 300	300 – 400	400 – 500	500 – 600	600 – 700	700 – 900
n_i	10	19	30	38	31	22

- a) Quel est le nombre total d'individus étudiés ?
- b) Calculer m_C et s_C les moyennes et écarts-type de C .
- c) Donner une estimation de la taille moyenne μ_C de la population C avec la confiance $\alpha = 0,95$.
- d) Peut-on dire, avec une confiance de 95% qu'il y a-t-il une différence significative de taille entre les populations A et C ?
4. la population D :
 Concernant la population D , quelqu'un de bien intentionné avait déjà fait les calculs avant nous. La personne en question a obtenu une moyenne $m_D = 507$, un écart-type $s_D = 170$ et un intervalle de confiance à 95% de la moyenne théorique $\mu_D : I_D = [480, 534]$.
 Peut-on affirmer qu'il y a une différence de taille moyenne entre les populations C et D ?

Indications

Exercice 3

Le membre de gauche est positif...

Exercice 4

1. On montrera que $p(1-p) \leq \frac{1}{4}$ pour tout $p \in [0, 1]$.

Exercice 5

1. On peut reconnaître le moment d'ordre 2 d'une variable suivant une loi exponentielle.

Exercice 11

2. Interpréter à l'aide d'une moyenne empirique.