1
Statistiques descriptives élémentaires

Dans tous les rapports récents, les jurys insistent sur deux choses : 1) les paramètres de description élémentaires des données statistiques doivent être parfaitement maîtrisés, et 2) **il faut savoir les calculer à l'aide d'une calculatrice**. Il est hors de question de vous amuser à calculer la moyenne empirique d'un échantillon de 15 données à la main, et encore moins l'écart-type.

Dans cette section, il n'y a pas de probabilités. Il n'y a que des outils mathématiques qui permettent de *décrire* des données connues et accessibles : on ne dit rien sur le modèle statistique qui aurait pu générer ces données. C'est pour cela qu'on parle de *statistiques descriptives*, par opposition aux *statistiques inférentielles* qui ont pour objectif de rechercher, à partir des données, les processus qui auraient pu les générer.

1.1 — Moyennes, écarts-types et coefficient de variation

Si $x_1, \dots, x_n$ sont des données statistiques numériques, leur moyenne empirique est

$$\mu = \frac{x_1 + \cdots + x_n}{n}$$

et leur variance empirique (naïve [1]) est

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Exercice 1.1. Vérifier que la variance empirique peut aussi se calculer via la formule

$$\sigma^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \mu^2.$$

L'écart-type empirique est σ, la racine carrée de la variance empirique. Elle est plus utile que la variance, car elle a la même dimension que les données : si les données sont en mètres, la moyenne est en mètres mais la variance est en « mètres au carré ».

DÉFINITION 1.2. — *Le **coefficient de variation** d'un échantillon statistique est $\rho = \sigma/\mu$.*

Ce coefficient s'interprète comme un écart-type relatif et en soi, c'est une excellente mesure de dispersion des données. L'intérêt du coefficient de variation est qu'il permet de mieux jauger la dispersion de certaines données. Par exemple, supposons qu'une série statistique ait un écart-type de 10. En soi, cela ne dit absolument rien sur la dispersion de ces données. Si les données ont tendance à être très grandes (disons que leur moyenne est 10000) alors une variation de ± 10 représente un écart de seulement $\rho = 0,1\%$ de la valeur moyenne. En revanche, si les données sont d'amplitude modérée (disons que la moyenne est 100), alors un écart type de 10 représente une variation de $\rho = 10\%$ par rapport à la moyenne.

Attention, le principal inconvénient du coefficient de variation est qu'il n'a plus d'intérêt lorsque la moyenne est proche de zéro ; il n'est même pas défini pour des variables centrées. Ce qu'il faut retenir, c'est qu'un coefficient de variation très petit signifie que l'écart-type est beaucoup plus petit que l'écart-type, donc les données sont très concentrées. Lorsque le coefficient de variation est grand, cela peut avoir deux interprétations : soit les données sont en moyenne très proches de zéro, et dans ce cas ρ ne dit rien, soit ce n'est pas le cas et dans ce cas les données ne sont pas très concentrées.

Exercice 1.3. Les notes de Jean-Patrick en latin sont les suivantes : 11, 11, 11, 11. Celles de Marie-Ursuline sont : 10, 10, 12, 12. Celles de Charles-Jason sont 15, 15, 1, 15. Quant à celles de Timéo-Kylian, elles sont de 17, 17, 19, 19. Calculer les coefficients de variation de chacun.

1. On reviendra plus tard sur un meilleur estimateur de la variance des données lorsqu'elles proviennent de réalisations de variables aléatoires.

1.2 — Histogrammes

Un histogramme permet de représenter et de visualiser d'un coup d'oeil la répartition de données numériques. Pour construire un histogramme, on identifie d'abord des classes (les « boîtes ») ; on compte le nombre d'observations qui appartiennent à chaque classe (l'*effectif*), puis on trace un rectangle dont la hauteur [2] est proportionnelle à l'effectif.

Mathématiquement, on choisit les bords des intervalles qui vont représenter nos classes, disons

$$b_1 < b_2 < \cdots < b_h$$

où h est le nombre de classes. La classe i sera composée de toutes les observations entre b_i et b_{i+1}. S'il y a n_i observations dans cette classe, on dessine le rectangle de base $[b_i, b_{i+1}]$ et de hauteur n_i. L'*amplitude* d'une classe est la longueur de sa base, c'est-à-dire $b_{i+1} - b_i$.

REMARQUE 1.4. La question du choix des classes est cruciale. En particulier, insistons sur un point : il est possible de choisir des classes qui n'ont pas la même amplitude. Ce sera le cas par exemple lorsque les données comportent des valeurs aberrantes : supposons que dans un jeu de données, 1000 valeurs sont entre -1 et 1, mais quelques-unes (disons, 3 ou 4) sont très éloignées de cet intervalle. Il peut être intéressant de diviser l'intervalle $[-1,1]$ en classes de même amplitude pour bien visualiser la répartition des données dans cet intervalle, et d'ajouter deux classes $]-\infty,-1[$ et $[1,+\infty[$ qui regrouperont les données aberrantes.

Un même jeu de données peut avoir une allure bien différente selon le choix des classes, comme en témoigne la figure 1.

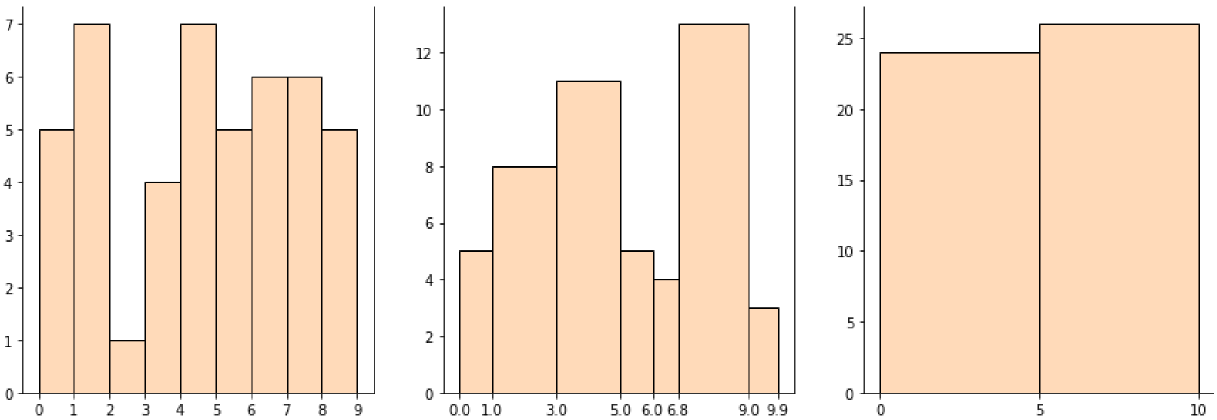

FIGURE 1 – Ces trois histogrammes représentent les mêmes 50 données visibles ci-dessous... mais évidemment, les boîtes ne sont pas les mêmes. es coordonnées des boîtes sont sur le graphique. Pour le premier histogramme on a utilisé des boîtes de même amplitude, c'est-à-dire qu'on a regroupé les données qui étaient entre 0 et 1, entre 1 et 2, etc. Il est par exemple possible de voir qu'il y a 7 observations dans l'intervalle [4,5] alors qu'il n'y en a qu'une dans l'intervalle [2,3], et vous pouvez vérifier cela dans le tableau ci-dessous. Attention, ici les histogrammes ne sont pas normalisés, c'est-à-dire que leur hauteur est égale à leur effectif.

```
7.73758879 8.55486039 5.57317639 7.40163887 8.90398262 0.49244671
1.3229121  1.88289874 5.53974594 5.50729498 1.28480919 7.52289461
6.89911707 7.53623245 3.59944876 0.04543076 8.86375818 5.67736335
4.76600856 2.53338482 4.52923367 4.48551095 3.89973424 1.45357723
1.52385813 9.82116848 4.79383897 4.56044908 9.53507463 6.45979969
3.06989745 6.22678533 4.44466447 7.88809468 0.34304044 6.92446781
8.44946871 1.42276336 3.96894957 6.03034629 9.98164926 1.81290755
5.1177333  6.13538887 7.60008069 0.61181801 4.85440412 0.16406851
8.12691921 9.60018777
```

On dit qu'un histogramme d'un jeu de n données est *normalisé* lorsque les hauteurs de tous les rectangles s'additionnent à 1. Autrement dit, la hauteur du rectangle qui représente la classe i n'est pas n_i, mais n_i/n. Lorsqu'un histogramme est normalisé, il donne une bonne approximation de la densité empirique des données. Si les

2. Ce n'est pas la seule façon de faire. On aurait pu demander à ce que l'aire soit proportionnelle à l'effectif.

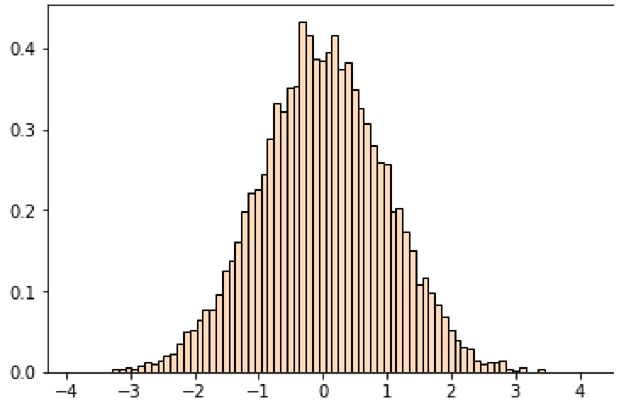

FIGURE 2 – Histogramme de 10000 réalisations d'une loi normale centrée réduite. Cette fois, l'histogramme est normalisé (cf l'axe des ordonnées).

données sont les réalisations de variables iid, cela se voit très bien. Par exemple, dans l'histogramme suivant, on a représenté un histogramme de 10000 réalisations d'une variable gaussienne centrée réduite.

REMARQUE 1.5 (histogramme et convergence)**.** Supposons que les données x_i sont des réalisations de variables aléatoires indépendantes de même loi X_i. La loi des grands nombres, lorsqu'elle s'applique, dit que la proportion d'observations entre les valeurs a et b converge vers la probabilité que la variable aléatoire sous-jacente soit entre a et b. Autrement dit, si l'histogramme est normalisé, la hauteur de la classe correspondant à l'intervalle $[a,b[$ est asymptotiquement très proche de $\mathbf{P}(X \in [a,b[)$. Ceci est valable pour une classe fixée, et donc aussi pour un nombre fini de classes, mais pas nécessairement pour un quantité « infinie » de classes. En réalité, c'est la convergence en loi qui correspond à ce type de convergence. Il est même possible de voir que la convergence en loi d'une suite de variables aléatoires est en fait équivalence à la « convergence de ses histogrammes ».

1.3 — Quantiles, médiane, médiale

Les quantiles d'un échantillon de données numériques x_i sont des nombres qui divisent l'échantillon en parties de même taille. Par exemple, la médiane est un nombre q tel qu'il y a autant d'observations inférieures à q que d'observations supérieures à q. Mathématiquement, la définition est la suivante.

DÉFINITION 1.6. — *Soient $x_1, \dots, x_n$ des nombres réels. Des quantiles d'ordre m de ces observations sont des nombres $q_{m,1}, \dots, q_{m,m}$, avec $q_{0,m}$ éventuellement égal à $-\infty$ et $q_{m,m}$ éventuellement égal à $+\infty$, tels que le nombre d'observations x_i dans chaque intervalle $I_k = [q_{k,m}, q_{k+1,m}[$ est toujours le même, c'est-à-dire n/m.*

Les quantiles d'ordre 3 portent le nom de terciles, ceux d'ordre 4 quartiles, ceux d'ordre 5 quintiles, ceux d'ordre 6 sexiles, ceux d'ordre 7 septiles, ceux d'ordre 8 octiles, deux d'ordre 9 noniles, ceux d'ordre 10 déciles, ceux d'ordre 11 ondéciles, ceux d'ordre douze duodéciles, ceux d'ordre 13 tridéciles, deux d'ordre 100 centiles, ceux d'ordre 1000 millésimiles, mais ce sont surtout les mots *quartiles*, *quintiles*, *déciles* et *centiles* qui sont utilisés — et évidemment, les quantiles d'ordre 2 sont très utilisés et s'appellent médiane.

Les quantiles définis ainsi ne sont pas forcément uniques. Par exemple, dans l'échantillon donné par $0,0,0,10,10,10$, le nombre 5 est une médiane, mais le nombre 6 également, ainsi que n'importe quel nombre entre 0 et 10 (strictement). Avec cette définition, les quantiles ne sont même pas toujours définis : par exemple, si le nombre d'observations est impair, il n'y a pas de médiane au sens de la définition ci-dessous. Il faut donc un peu adapter la définition. Dans le cas de la médiane, c'est facile : il suffit de dire que le nombre « du milieu ». En d'autres termes, si l'on ordonne nos observations de façon croissante

$$x_{i_1} \leqslant \cdots \leqslant x_{i_n}$$

alors la médiane est n'importe quel nombre entre $x_{i_{n/2}}$ et $x_{i_{1+n/2}}$ si n est pair, et sinon on convient de dire que c'est le nombre $i_{[n/2]}$.

Par ailleurs, les quantiles les plus utilisés sont ceux d'ordre 4. On note souvent Q_1 le premier et Q_3 le dernier. Celui du milieu est une médiane (pourquoi ?).

Exercice 1.7. Trouver une médiane pour les données suivantes :

$$84, 58, 7, 12, 63, 99, 588, 6, 65, 23, 74, 25, 5, 56545, 6, \pi, 55.$$

Calculer la moyenne.

Exercice 1.8. La moyenne et la médiane ne sont pas forcément égales. Comment interpréter le fait qu'une moyenne soit, par exemple, inférieure à la médiane ?

Exercice 1.9. Dans la bibliothèque de l'agrégation, on a compté les livres en fonction de leur nombre de pages :

pages	[0, 100[	[100, 250[	[250, 400]	[400, ∞[
nombre de livres	19	29	48	12

Quelles sont les médianes possibles ? Peut-on réellement identifier une médiane ?

Il existe aussi des façons rigoureuses de définir les quantiles d'ordre m même si n n'est pas divisible par m. Nous n'en parlerons pas [3]. En revanche, il peut être utile d'avoir en tête la définition des quantiles pour variables aléatoires. Soit donc Z une variable aléatoire.

DÉFINITION 1.10. — *Si m est un entier naturel, les quantiles d'ordre m pour la loi de Z sont $m+1$ nombres $q_{m,1}, \dots, q_{m,m}$, avec $q_{0,m}$ éventuellement égal à $-\infty$ et $q_{m,m}$ éventuellement égal à $+\infty$, tels que pour chaque $i \in \{1, \dots, m\}$,*

$$\P(q_{i-1,m} < Z \leqslant q_{i,m}) = \frac{1}{m}. \tag{1}$$

La médiane est n'importe quel nombre q tel que $\mathbf{P}(Z > q) = \mathbf{P}(Z < q) = 1/2$.

Exercice 1.11. Les quantiles ne sont pas uniques en général. Montrer toutefois qu'ils le sont lorsque Z possède une densité continue sur $\mathbb{R}$ et strictement positive.

Exercice 1.12. Montrer que $\P(Z \leqslant q_{k,m}) = k/m$.

Exercice 1.13. Quelle est la médiane d'une variable aléatoire centrée ?

Dans le cas d'un jeu de données numériques $(x_1, \dots, x_n)$, les quantiles et la médiane se définissent de façon analogue : des nombres $q_{i,m}$ sont des quantiles d'ordre m pour ce jeu de données si le nombre des observations entre $q_{m,i-1}$ et $q_{m,i}$ est égal à n/m, autrement dit si la fréquence des observations entre ces deux nombres est $1/m$.

Écart interquartile. Il s'agit de l'écart entre le premier et le troisième quartile d'une distribution statistique, c'est-à-dire de la quantité $Q_3 - Q_1$. Il s'agit d'un indicateur de dispersion : un grand écart interquartile indique (dans une certaine mesure) une distribution dispersée. Cet indicateur est souvent présent dans les sorties des logiciels statistiques (par exemple, pour analyser les résidus dans une régression linéaire). Si l'écart interquantile est très proche de zéro, cela signifie que les valeurs sont très concentrées, puisque la moitié des données sont comprises entre les deux valeurs très proches que sont Q_1 et Q_3.

Médiale. Beaucoup moins utilisée que la médiane [4], la médiale d'une distribution est la valeur de la variable qui sépare la masse totale de cette variable en deux parties de même poids [5].

La médiale est donc la valeur qui sépare la *masse totale* de la variable en deux parties, tandis que la médiane est la valeur qui sépare les *effectifs* en deux. Cette distinction est plus digeste avec un exemple :

3. Go wikipédia
4. En fait, la médiale est strictement et rigoureusement inutilisée en dehors du champ restreint des membres du jury de l'oral de mathématiques de l'agrégation de sciences sociales (à confirmer).
5. Ce n'est pas tout à fait exact : il est possible qu'une telle valeur n'existe pas ou ne soit pas unique. Comme dans le cas de la médiane, on définit rigoureusement la médiale m' comme la *plus petite valeur* de x telle que la somme des observations x_i strictement inférieures ou égales à x soit supérieure ou égale à la somme des obervations strictement supérieures à x, c'est-à-dire que

$$m' = \inf \left\{ x \in : \sum_{x_i \leqslant x} x_i \geqslant \sum_{x_i > x} x_i \right\}.$$

	Salaire	Cumul
M. Acarien	1 200	1 200
M. Moucheron	1 220	2 420
Mlle Araignée	1 250	3 670
Mme Musaraigne	1 300	4 970
Mme Souris	1 350	6 320
M. Rat	1 450	7 770
M. Cobaye	1 450	9 220
M. Lapin	1 560	10 780
M. Chat	1 600	12 380
Mme Gazelle	1 800	14 180
M. Daim	1 900	16 080
M. Cheval	2 150	18 230
M. Élan	2 310	20 540
M. Bison	2 600	23 140
M. Rhinocéros	3 000	26 140
M. Éléphant	3 400	29 540
Mme Baleine	4 800	34 340

Sur le tableau ci-dessus, la médiane de la distribution des salaires est le salaire de M. Chat (1600 euros). La médiale, quant à elle, est le plus petit salaire correspondant à un cumul supérieur à la moitié de la masse salariale totale, c'est-à-dire, puisque cette masse totale est de 34340 euros, le premier salaire correspondant à un cumul supérieur à 17170 euros, donc celui de M. Cheval (18230 euros).

Exercice 1.14. Montrer que la médiale d'une variable positive est toujours supérieure à sa médiane.

Exercice 1.15. Montrer que la médiale est l'abscisse du premier point d'ordonnée supérieure ou égale à $0,5$ sur la courbe de Lorenz de la distribution.

1.4 — Polygone des effectifs, ou fréquences cumulées

Supposons que l'on dispose de n observations $(x_1,\dots,x_n)$ d'une variable quantitative : salaire, taille, nombre total de cheveux, longueur de l'intestin grêle.

Ce que le programme nomme pompeusement « polygone des effectifs ou fréquences cumulées » n'est rien d'autre que la fonction de répartition des x_i, ou sa courbe représentative. Autrement dit, c'est la fonction f continue par morceaux de la façon suivante :

$$f(t) = \text{ nombre d'observations } x_i \text{ telles que } x_i < t. \tag{2}$$

La courbe représentative d'une telle fonction est une fonction « en escalier », et il y a des gens qui n'aiment pas les fonctions en escalier parce qu'elles ont des trous. En effet, supposons que les x_i sont tous distincts et notons $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ le réordonnement croissant des x_i. À chaque $x_{(i)}$, la fonction f saute de la valeur $i-1$ à la valeur i.

Pour circonvenir ce problème, on introduit le polygone des effectifs, qui est tout simplement le graphe que l'on obtient en reliant les points $(x_{(i)},i)$. Lorsque certains des $x_{(i)}$ sont égaux, c'est la même chose, sauf que les « trous » dans le graphe de f peuvent prendre les valeurs $1,2,3...$ si par exemple plusieurs x_i ont la même valeur.

Qu'il s'agisse du polygone ou du graphe de f, ce genre de graphiques permet de saisir d'un coup d'oeil la répartition d'une variable quantitative. Comme la fonction f prend des valeurs allant de 0 à n, on la divise parfois par n afin d'obtenir de vraies fréquences. Le maximum des x_i est le plus petit x tel que $f(x) = 1$; le minimum est le plus grand x tel que $f(x) = 0$.

Exercice 1.16. Comment calculer les quantiles d'un jeu de données en regardant ses fréquences cumulées ?

1.5 — Courbe de Lorenz

Soient $x_1,\dots,x_n$ des observations numériques, par exemple des revenus pécuniaires[6]. Si les revenus étaient équitablement distribués dans la population, la moitié de la population devrait recevoir à peu près la moitié de

6. C'est l'exemple originellement utilisé par Max Lorenz pour étudier les inégalités de revenus aux États-Unis.

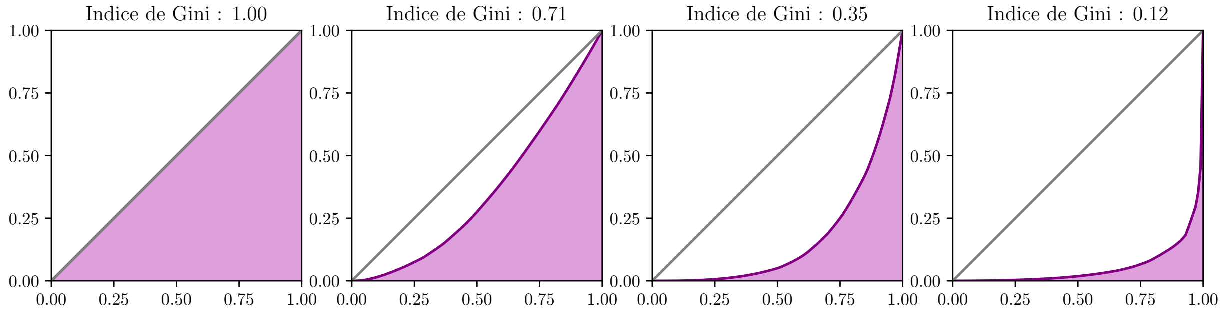

FIGURE 3 – Quelques courbes de Lorenz avec leurs indices correspondants.

la masse totale des revenus. Ce n'est évidemment pas le cas : en règle générale, la moitié des revenus totaux est allouée à une part de la population beaucoup plus faible. La courbe de Lorenz permet de visualiser cela. Cette courbe représente la part cumulée des revenus en fonction de la part cumulée de la population.

Mathématiquement, on note $S = x_1 + \cdots + x_n$ la quantité totale de la variable mesurée (dans le cas des revenus, c'est la richesse totale de la population). On ordonne par ordre croissant toutes les observations, disons

$$x_{i_1} \leqslant \cdots \leqslant x_{i_n}.$$

Les k premières observations sont $x_{i_1}, \cdots, x_{i_k}$. Elles représentent donc $k/n\%$ de la population, mais la part de variable qui lui revient est $x_{i_1} + \cdots + x_{i_k}$, ce qui représente $(x_{i_1} + \cdots + x_{i_k})/S\%$ de la masse totale. La courbe de Lorenz relie les points d'abscisse k/n et d'ordonnée $(x_{i_1} + \cdots + x_{i_k})/S$. Elle passe donc nécessairement par les points $(0,0)$ et $(1,1)$, et elle est toujours inférieure à la droite qui relie ces deux points (pourquoi ?). On notera souvent $\ell(t)$ la fonction de Lorenz, celle qui est représentée par la courbe de Lorenz.

Exercice 1.17. Comment trouver la courbe de Lorenz à partir du polygone des fréquences cumulées ?

Si les variables étaient équitablement réparties dans la population, la courbe de Lorenz serait précisément cette droite reliant $(0,0)$ et $(1,1)$. L'aire entre ces deux droites est donc d'autant plus grande que les données sont inégalement réparties dans la population. Il existe plusieurs façons de mesurer la différence entre ces deux aires, ce qui donne naissance aux divers indices de Gini. La définition exigée pour l'agrégation est celle du *rapport* entre ces deux aires.

DÉFINITION 1.18. — *Le coefficient ou indice de Gini est égal défini par*

$$\text{Gini} = \frac{\textit{Aire sous la courbe de Lorenz}}{\textit{Aire si les données étaient équitablement réparties}} = \frac{\int_0^1 \ell(t)\mathrm{d}t}{1/2}. \tag{3}$$

Avec cette définition, **plus l'indice est faible, plus la répartition est inégalitaire, et plus l'indice est proche de 1, plus la répartition est égalitaire**.

1.6 — Boîtes à moustache

Ce qu'on appelle coquettement « boîte à moustache » est une représentation de données numériques qui fait clairement apparaître tous les indicateurs dont nous avons parlé plus tôt : moyenne, médiane, quantiles, écart-type. Pour des données $x_1, \ldots, x_n$, les données sont représentées sous la forme d'une boîte, dont la largeur n'a aucune importance, mais dont le segment supérieur est Q_3, le segment inférieur Q_1, et dans laquelle on a représenté la médiane et/ou la moyenne. Enfin, deux traits (les moustaches) indiquent les valeurs maximales et minimales prises par les données. Parfois, certaines valeurs extrêmes sont représentées.

L'intérêt des box-plots (leur petit nom étranger) est principalement de comparer des données entre différentes populations.

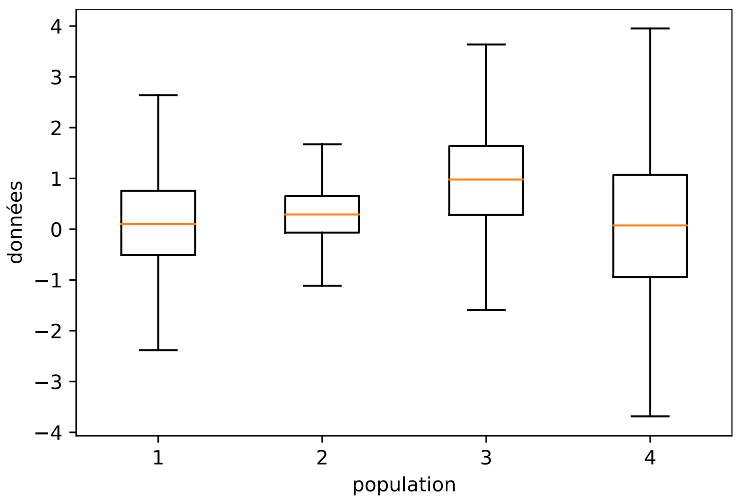

FIGURE 4 – Un exemple de boîtes à moustache. On voit d'un coup d'oeil la différence de dispersion des données entre les quatre populations.

1.7 — Indices simples et synthétiques

On note $p_{t,i}$ le prix du bien i au temps t et $q_{t,i}$ la quantité de bien i faisant partie du panier de biens au temps t. Les indices de références du programme sont les suivants.

1.7.1 Indices des prix

Dans une économie, les prix des produits suivent des évolutions diverses mais il peut être intéressant, notamment pour les autorités monétaires, de calculer un indice synthétique de prix pour suivre l'évolution du niveau général des prix. Pour ce faire, les indices des prix de Laspeyres-Paasche fixent un panier de biens de référence et comparent leur prix (valeur monetaire) entre deux périodes.

1. **Indice de Paasche des prix**.
$$\Delta P_P = 100 \times \frac{\sum_i p_{t,i} q_{t,i}}{\sum_i p_{0,i} q_{t,i}}. \tag{4}$$
C'est la variation de prix (en pourcentage) d'un panier de biens typique de la période courante depuis la période de référence. Cet indice est insensible au changement des habitudes de consommation, c'est-à-dire aux changements de composition du panier.
2. **Indice de Laspeyres des prix**.
$$\Delta P_L = 100 \times \frac{\sum_i p_{t,i} q_{0,i}}{\sum_i p_{0,i} q_{0,i}}$$
C'est la variation de prix (en pourcentage) d'un panier de biens typique de la période de référence depuis cette période. Cet indice est celui utilisé dans le calcul de l'inflation en France ; il est insensible au changement des habitudes de consommation.
3. **Indice de Fisher des prix**.
$$\Delta P_F = \sqrt{\Delta P_P \Delta P_L}$$
Cet indice incorpore les indices de Paasche et de Laspeyres dont il est la moyenne géométrique[7]. Il est notamment utilisé par l'institut Statistique Canada.

1.7.2 Indices des quantités

Les indices des quantités mesurent de la croissance de l'activité économique. S'il n'y avait qu'un produit, par exemple le blé, la situation serait relativement simple : pour mesurer la croissance de la production, il suffirait de suivre l'évolution des quantités produites. Mais l'économie comporte plus d'un produit ! Supposons donc que l'économie produise non seulement du blé mais aussi du poisson. Si entre deux périodes, les quantités de blé

7. On appelle *moyenne géométrique* de deux nombres positifs a et b la quantité $\sqrt{ab}$, tandis que la quantité $\frac{a+b}{2}$, habituellement appelée tout simplement « moyenne » de ces deux nombres, en est la *moyenne arithmétique*.

et de poisson évoluent toutes deux de 20%, il semble donc raisonnable de dire que la croissance de l'économie s'établit, elle aussi, à 20%. Mais, si entre deux périodes, les quantités de blé et de poisson évoluent différemment, par exemple si les quantités de blé baissent de 17% alors que celles de poisson augmentent de 25%. Que dire de la croissance globale ? Il pourrait être tentant de considérer les quantités globales produites mais additionner des quantités de produits de nature différente n'a pas de sens.

Pour résoudre cette difficulté on utilise la valeur monetaire : $v_i = p_i \times q_i$.

Si, dans les comparaisons de valeurs entre deux periodes, on veut éliminer l'effet de l'inflation, un moyen simple consiste à valoriser les quantités des deux périodes aux mêmes prix.

1. **Indice de Paasche des quantités**.
$$\Delta Q_P = 100 \times \frac{\sum_i p_{t,i} q_{t,i}}{\sum_i p_{t,i} q_{0,i}}$$
C'est la variation (en pourcentage) de valeur monetaire entre un panier de biens de la période de référence et un panier de biens de la période courante, calculé par rapport aux prix actuels.

2. **Indice de Laspeyres des quantités**.
$$\Delta Q_L = 100 \times \frac{\sum_i p_{0,i} q_{t,i}}{\sum_i p_{0,i} q_{0,i}}$$
C'est la variation (en pourcentage) de valeur monetaire entre un panier de biens de la période de référence et un panier de biens de la période courante, calculé par rapport aux prix de la période de réference.

3. **Indice de Fisher des quantités**.
$$\Delta Q_F = \sqrt{\Delta Q_P \Delta Q_L}$$

Tous ces indices sont *composites*, c'est-à-dire qu'ils intègrent des informations sur plusieurs grandeurs (les quantités et les prix aux deux périodes considérées). À l'inverse, un indice *simple* est un indice qui ne prend en compte qu'une seule grandeur, par exemple l'évolution en pourcentage du prix d'un bien unitaire donné entre deux périodes.