

## CHAPITRE 3 AUTOUR DE LA MOYENNE

### TABLE DES MATIÈRES

1. [La moyenne : un équilibre des forces](#)
  2. [L'écart type : mesurer la dispersion des données](#)
  3. [La courbe normale et l'écart type](#)
  4. [D'autres indicateurs de dispersion](#)
- [Exercices supplémentaires](#)

On dit parfois qu'une image vaut mille mots. De la même façon, un ou deux chiffres peuvent remplacer avantageusement plusieurs pages de tableaux. Le tout est de bien choisir ces chiffres. Dans ce chapitre, nous apprendrons justement à bien choisir et à bien interpréter ces quelques chiffres, qui résument d'un coup d'œil, l'ensemble d'une situation.

Avant de commencer, prenons un exemple tiré de l'hebdomadaire économique polonais *Gazeta Bankowa*, peu après l'effondrement du mur de Berlin. Les Polonais se trouvent à l'étroit, et à juste titre : ils ne disposent, en moyenne, que de 18 m<sup>2</sup> habitables par personne. C'est encore moins que les Japonais (28 m<sup>2</sup> en moyenne) qui n'ont pourtant pas la réputation d'occuper des logements spacieux. On voit ici qu'un simple chiffre, même s'il ne dit pas tout (il doit y avoir des Polonais plus à l'étroit que d'autres), nous dispense de consulter la liste des 40 millions d'habitants de cette république d'Europe de l'Est, si tant est qu'une telle liste existe. À titre de comparaison, l'Américain moyen dispose alors de 62 m<sup>2</sup> de surface habitable et l'Européen de l'Ouest de 37 m<sup>2</sup> en moyenne.

Comme on le dit vulgairement, les Américains se « pilent » moins sur les pieds que les Polonais. Il faut cependant faire quelques nuances. Même si l'Amérique est grande, tous les Américains ne logent pas à Beverly Hills. Il doit y avoir des écarts en fonction notamment du milieu social. Les disparités entre les gens sont encore plus élevées en Europe de l'Ouest, du moins en termes relatifs. Même si la moyenne d'espace de logement occupé par ces derniers est de 37 m<sup>2</sup> par habitant, la plupart des Européens de l'Ouest ont soit beaucoup plus, soit beaucoup moins. Les plus favorisés sont les gens de la campagne, ceux qui vivent dans des maisons de banlieue, ceux qui possèdent une résidence secondaire et plus généralement les personnes fortunées. Dans ce chapitre, nous verrons aussi comment évaluer rapidement l'importance de ces écarts autour de la moyenne.

Nous en profiterons également pour établir la position d'un individu à l'intérieur d'un groupe. Où se situe par exemple l'Américain qui occupe 100 m<sup>2</sup> de surface habitable? Fait-il partie de l'élite? Est-il dans les 10 % meilleurs? Dans les 40 %? Fait-il partie de la moitié la moins bien lotie de la population? Toutes ces situations étant possibles (même si la moyenne générale du pays est de 62 m<sup>2</sup>), il est parfois bon de connaître non seulement la moyenne, mais de posséder également des informations sur les écarts entre les éléments du groupe, ainsi que sur la position d'un élément particulier.

Au terme de ce chapitre, vous devriez être en mesure de répondre aux questions suivantes :

- L'individu « moyen » est-il celui qu'on risque le plus de rencontrer dans le groupe qui est le sien?
- Quel est le rapport entre la moyenne et les données qu'elle veut résumer?
- Comment peut-on situer une valeur (et le sujet qu'elle désigne) par rapport à la moyenne? Par rapport aux autres valeurs du même groupe?
- Comment les autres mesures comme la médiane, le mode ou les quintiles peuvent-elles nous être utiles à rendre la moyenne plus significative?

## 1. LA MOYENNE : UN ÉQUILIBRE DES FORCES

---

Tout le monde a entendu parler des moyennes. Tout le monde s'en sert. Comme on pouvait s'y attendre, notre but principal consiste davantage à bien interpréter les moyennes qui nous sont servies régulièrement dans les livres et les journaux que de nous livrer à de savants calculs. Cependant, nous ne sommes pas toujours de simples récepteurs d'information : nous avons parfois nous aussi quelque chose à dire. C'est pourquoi nous verrons comment bien choisir et bien calculer les moyennes. Comme toujours, nous prendrons nos exemples dans la réalité, car l'humanité a ceci de particulier : mieux on la connaît et plus on la trouve intéressante.

### 1.1. Le jour J : des péniches et des hommes

Jour J. Le 6 juin 1944. 745 navires, groupés en 38 convois, approchent des côtes normandes. En tout, 4006 péniches de débarquement sont mises à l'eau, au large des plages, et 185 000 hommes débarquent sur le territoire français. Regardons les choses à une échelle un peu plus humaine : combien d'hommes y avait-il en moyenne sur chaque péniche de débarquement? Essayez d'imaginer la réponse sans faire de calcul, mais plutôt en visualisant ces péniches. Et rappelez-vous qu'il ne s'agit pas d'un film hollywoodien, mais bien d'une histoire vécue par des jeunes gens tels que vous.

La moyenne est une mesure qui correspond à la somme des valeurs que prend une variable divisée par le nombre de ces valeurs.

Notons que la variable que l'on cherche à évaluer ici est la capacité, en nombre d'hommes, des péniches de débarquement. Nous ne disposons pas de la liste des péniches avec leur nombre respectif de passagers, mais nous connaissons le nombre de péniches et le total des passagers. Chaque péniche peut être associée à une *valeur* (le nombre d'hommes qu'elle contient). Le *nombre des valeurs* (le nombre de péniches) et la *somme des valeurs* (le nombre total de passagers) nous suffisent pour obtenir la moyenne des individus présents dans les péniches.

185 000 hommes/4006 péniches = 46 hommes en moyenne par péniche

Moyenne = Somme des valeurs/Nombre de valeurs
---

Les péniches de débarquement contenaient donc en moyenne 46 hommes chacun. Nous vous demandons plus haut d'essayer de visualiser ce résultat avant de le calculer. Si votre intuition correspondait au résultat, félicitations. Si vous aviez imaginé des péniches avec beaucoup moins d'hommes, ou beaucoup plus (nous sommes dans ce cas), cela prouve qu'il n'est pas toujours inutile de faire un petit calcul lorsque l'on veut avoir une idée *objective* de la réalité.

C'est donc dire que la plupart des péniches contenaient 46 hommes?

La moyenne ne prétend pas que tous soient égaux (il y aurait alors eu 46 hommes dans chaque péniche). Elle ne fait que nous dire quelle valeur aurait eu chacune des données si celles-ci avaient été toutes pareilles. Cela dit, rien ne prouve que la plupart des péniches contenaient 46 hommes. Certaines étaient peut-être très petites, d'autres très grandes. L'exemple qui suit nous démontrera que la moyenne est une valeur *autour* de laquelle les autres se situent et qu'elle n'est pas nécessairement celle qu'on retrouve le plus souvent.

## 1.2. César et ses Romains

Jules César, malgré ses nombreuses conquêtes, son immense talent et tous ses autres mérites, n'a jamais réussi à se faire couronner empereur. Sa carrière tardive fut brusquement interrompue lors d'une visite au Sénat, le 15 mars de l'année 44 avant notre ère. Le Sénat de l'époque était plus fringant que le nôtre! Auguste, fils adoptif de César (et son *vrai* fils selon les mauvaises langues) devait cependant fonder une dynastie qui dura près d'un siècle. Pour consoler les étudiants qui traînent encore dans les classes malgré leur âge avancé, signalons que Jules César n'a véritablement amorcé sa carrière qu'à 40 ans, après avoir réalisé de façon dramatique « qu'il n'avait encore rien fait de mémorable à un âge où Alexandre (le Grand) avait déjà soumis toute la terre ».

Si César et Auguste furent de grands hommes d'État (dont l'héritage est toujours vivant), on ne peut pas en dire autant de ceux qui leur ont succédé. Tous furent célèbres pour leurs débauches et leurs excès. Selon Suétone (l'auteur de la *Vie des douze Césars*), Tibère s'était fait installer un « [jardin des plaisirs](#)\* » dans sa retraite de Caprée. Caligula, son successeur — et sans doute son meurtrier —, entretenait des relations incestueuses avec ses sœurs. Caligula nomma même son cheval sénateur (notez l'évolution du Sénat avec le temps). Il fut assassiné, tout comme ses successeurs Claude (homme plutôt lâche et influençable) et Néron (qui se prenait pour une grande vedette).

Avis aux étudiants qui n'aiment pas l'histoire : c'est le moment ou jamais de vous y intéresser!

Les présentations étant faites, revenons à nos chiffres. À quel âge, en moyenne, devenait-on empereur sous les Césars? Combien de temps restait-on sur le trône? À quel âge mourait-on?

Pour calculer ces moyennes, nous utiliserons la même formule que dans le [cas précédent](#). La seule différence ici est que nous devons calculer nous-mêmes la somme (à l'aide des données du tableau 3.1 figurant ci-après).

**TABEAU 3.1 - La dynastie des Césars**

	Naissance	Début du règne	Fin du règne et décès	Âge au début du règne	Durée du règne	Âge au décès
Auguste	-63	-24	14	39	37	76
Tibère	-42	14	37	55	23	78
Caligula*	12	37	41	25	4	29
Claude*	-10	41	54	50	13	63
Néron*	37	54	68	17	14	31
	Échelle d'intervalle			Échelle de rapport		

Source des trois premières colonnes de chiffres : Petit Robert 2.

\* Mort assassiné.

$$\text{Âge moyen au début du règne : } (39 + 55 + 25 + 50 + 17)/5 = 37,2 \text{ ans}$$

$$\text{Durée moyenne du règne : } (37 + 23 + 4 + 13 + 14)/5 = 18,2 \text{ ans}$$

$$\text{Âge moyen du décès : } (76 + 78 + 29 + 63 + 31) = 55,4 \text{ ans}$$

Il faut reconnaître qu'une bonne compréhension des [échelles de mesure](#) s'avère ici fort utile. Puisque les données figurant dans les trois premières colonnes du tableau appartiennent à une échelle d'intervalle, il était évidemment impossible d'en tirer des moyennes. Par contre, les trois dernières colonnes du tableau — que nous avons déduites des trois premières — appartiennent à une échelle de rapport. C'est pourquoi nous avons été en mesure de calculer les trois moyennes correspondantes.

Voici maintenant la preuve que les méthodes quantitatives requièrent avant tout un esprit *methodique* plutôt que des connaissances poussées en mathématiques. Pour calculer les valeurs des trois dernières colonnes du tableau, nous avons dû tenir compte du fait que la vie et le règne de certains empereurs chevauchaient le premier millénaire avant Jésus-Christ et le premier millénaire après Jésus-Christ (il va de soi que les Romains utilisaient une échelle d'intervalle dont le point de départ était différent). Or, le premier millénaire après Jésus-Christ a de toute évidence [commencé en l'an 1\\*](#) et non en l'an 0, qui n'a d'ailleurs jamais existé. Ainsi, pour calculer la durée du règne d'Auguste, nous ne pouvons pas simplement additionner 24 (années avant J.-C.) et 14 (années après J.-C.). En réalité, il s'est écoulé 37 ans — et non 38 — entre l'an -24 et l'an 14.

Notons que le problème aurait été évité si nous avions utilisé le calendrier romain, qui part de la fondation de Rome (l'an 753 av. J.-C. correspondant à l'an 1 de l'ère romaine). Selon le calendrier romain, Auguste devint empereur en 730 et mourut en 767, ce qui donne bien 37 ans de règne : ouf, le compte est bon!

Si le premier millénaire a commencé en l'an 1, il est clair que le troisième millénaire a commencé en l'an 2001, et non en l'an 2000, n'en déplaise aux « millénaristes », qui considèrent plus les chiffres comme des objets de culte que comme une source de connaissance.

Cela dit, une telle erreur de méthode ne porterait pas tellement à conséquence ici, puisque nous utilisons des chiffres arrondis à l'année près (les dates considérées peuvent se situer entre le 1<sup>er</sup> janvier et le 31 décembre d'une année), et que le phénomène étudié n'exige pas un [degré de précision](#) extrême. Par contre, dans d'autres circonstances (dossier criminel, médical ou bancaire), une erreur de ce genre serait inacceptable. Il est donc important de bien porter attention au point de départ de l'échelle lorsque celle-ci est une échelle d'intervalle.

### 1.3. La moyenne et les fréquences

Combien d'individus compte, en moyenne, un ménage montréalais? Il n'y eut que 5 empereurs dans la dynastie des Césars, mais on a dénombré 756 000 ménages dans la communauté urbaine de Montréal en 1993. Le principe de la moyenne demeure toujours le même (la somme des valeurs divisée par le nombre de valeurs), mais la masse imposante des données nous oblige ici à trouver une méthode plus efficace.

On ne va quand même pas additionner la valeur (en nombre de personnes) des 756 000 ménages et diviser par 756 000 pour connaître la moyenne!

Dans le tableau 3.2 ci-après (tiré d'un exemple du chapitre précédent), la variable est la *taille du ménage*. Cette variable peut prendre diverses valeurs déterminées par le nombre d'individus par ménage (colonne 1). Étant donné que les ménages de plus de 4 personnes sont rares, notre tableau met dans le même paquet tous les ménages comptant 4 personnes et plus. Pour simplifier, nous estimerons que les ménages de 4 personnes et plus comptent en moyenne 4,5 personnes. La colonne 2 contient le nombre de ménages dénombrés (la fréquence, en milliers) pour chaque valeur possible. La colonne 3 reprend la colonne 2 sous forme de proportions (les fréquences relatives).

**TABLEAU 3.2 - Répartition des ménages familiaux à Montréal selon le nombre de membres**

	Fréquence (nombre de ménages en milliers)	Fréquence relative	Fréquence relative cumulée
[1]	[2]	[3]	[4]
1 personne	255	33,7	<b>33,7</b>
2 personnes	243	32,1	<b>65,9</b>
3 personnes	118	15,6	<b>81,5</b>
4 personnes ou plus	140	18,5	<b>100,0</b>
<b>Total</b>	<b>756</b>	<b>100,0</b>	

Source : Statistique Canada, Cansim. Données de 1993.

Note : Il s'agit de la Communauté urbaine de Montréal.

Pour obtenir la moyenne, nous pourrions aligner 756 chiffres (255 chiffres 1, suivis de 243 chiffres 2, etc.), les additionner et diviser le tout par 756. Nous pourrions encore, ce qui reviendrait au même, effectuer l'opération suivante :

$$[(255 \times 1) + (243 \times 2) + (118 \times 3) + (140 \times 4,5)]/756 = 1725/756 = 2,28$$

On notera que nous avons attribué, comme convenu, 4,5 personnes aux 140 ménages de 4 personnes et plus. On constatera également que la somme des fréquences (255 + 243 + 118 + 140) est égale au dénominateur (756). En fin de compte, le ménage montréalais moyen compte 2,28 individus.

$$\text{Moyenne} = [(Fréquence 1 \times Valeur 1) + (Fréquence 2 \times Valeur 2) + \dots]/Fréquence totale$$

Cette manière de calculer la moyenne n'est qu'une généralisation de la formule précédente (dans laquelle les fréquences de chaque élément sont toutes égales à 1). Avez-vous remarqué, par ailleurs, que les chiffres du tableau 3.2 représentent en réalité des *milliers* de ménages? Il y a en effet 756 000 ménages à Montréal et non 756. Cette particularité ne change rien au résultat, puisque les proportions restent les mêmes. C'est un peu comme si on mesurait le poids moyen des poulets en kilogrammes plutôt qu'en grammes : ça ne rendrait pas les poulets plus légers ni plus lourds.

## 1.4. La moyenne pondérée : des poids relatifs

Dans l'exemple qui précède (la taille moyenne des ménages), nous aurions pu obtenir le même résultat en utilisant les fréquences relatives. Ce qui compte, après tout, c'est de respecter les proportions de chaque catégorie.

Si on utilise ces proportions sous leur forme décimale, le calcul de la moyenne se fait de la façon suivante :

$$(0,337 \times 1) + (0,321 \times 2) + (0,156 \times 3) + (0,186 \times 4,5) = 2,28 \text{ personnes par ménage}$$

Moyenne = (Pondération 1 × Valeur 1) + (Pondération 2 × Valeur 2) + ...
---

Si on préfère utiliser les proportions sous forme de pourcentage, il faudra diviser le résultat final par 100 pour obtenir la moyenne : n'oublions pas que 100 % = 100/100 = 1.

$$[(33,7 \times 1) + (32,1 \times 2) + (15,6 \times 3) + (18,5 \times 4,5)]/100 = 2,28 \text{ personnes par ménage}$$

Les fréquences relatives représentent le poids relatif (ou *pondération*) de chaque composante. La somme de ces pondérations est bien sûr égale à 100 % ou 1.

$$33,7 \% + 32,1 \% + 15,6 \% + 18,5 \% = 99,9 \% = 100 \% = 1$$
$$0,337 + 0,321 + 0,156 + 0,185 = 0,999 = 1$$

Lorsque les pondérations sont connues, il n'est même plus nécessaire de connaître les diverses fréquences (ou effectifs) pour calculer la moyenne.

Parfois, les pondérations sont déterminées à partir d'autres critères que les fréquences. La moyenne des notes accumulées par un étudiant dans un cours est souvent une moyenne pondérée : tel exercice peut compter pour 10 % de la note finale, tel examen pour 30 %, etc. L'enseignant peut fixer ces pondérations en fonction de l'importance qu'il accorde personnellement à chaque évaluation, pourvu que la somme des pondérations soit égale à 100 %.

## 1.5. Les limites de la moyenne

Peut-être avez-vous déjà entendu l'expression « la loi de la moyenne » dans la bouche d'un mordu du casino : « Le numéro 13 n'est quasiment pas sorti hier. Aujourd'hui, il devrait se rattraper. J'en suis sûr, mon vieux, c'est la loi de la moyenne. Mise tout sur le 13! ».

Devant la majesté de la loi, l'homme n'a plus qu'à s'incliner, n'est-ce pas? C'est oublier que la moyenne est une mesure qui *traduit* la réalité, elle ne la *commande* pas. Le hasard n'a pas de mémoire, et la soi-disant « loi de la moyenne » n'est qu'une superstition héritée de nos croyances primitives.

Même si la moyenne est un concept relativement simple, il arrive parfois qu'on l'utilise de façon abusive, et pas seulement pendant les parties de roulette ou les performances sportives. C'est pourquoi nous allons mentionner quelques erreurs typiques liées au calcul ou à l'utilisation des moyennes.

### Erreur de calcul : la moyenne des moyennes

Alors que chaque Américain consomme *en moyenne* l'équivalent de 6794 kg de pétrole sous forme d'énergie diverse par an, le Mexicain n'en consomme en moyenne que 1588 kg, et le Canadien, 7270 kg\*. Quelle est la moyenne d'énergie consommée en Amérique du Nord?

On ne peut additionner ici les 3 chiffres et diviser le total par 3 (cela donnerait  $[6784 + 1588 + 7270]/3 = 5217$  kg). Il faut tenir compte de la population de chaque pays (314 millions pour les États-Unis, 121 millions pour le Mexique et 35 millions pour le Canada, soit une population totale de 470 millions).

Source : Banque mondiale, *Indicateurs de développement dans le monde*. Données de 2012. À titre de comparaison, la consommation par habitant était, en 1992, de 7662 kg aux États-Unis, 1525 au Mexique, et 7912 au Canada. Les populations respectives de ces trois pays étaient de 255, 85 et 27 millions. La consommation moyenne d'énergie pour l'ensemble de l'Amérique du Nord était donc d'environ 6260 kg d'équivalent pétrole.

Dans ce cas, la véritable moyenne serait de :

$$\begin{aligned} & [(6794 \times 314) + (1588 \times 121) + (7270 \times 35)]/470 \\ & = [2\,133\,316 + 191\,148 + 254\,450]/470 \\ & = 2\,579\,914/470 = 5489 \text{ kg d'équivalent pétrole consommé en moyenne par habitant} \end{aligned}$$

$$\text{Moyenne} = [\text{Consommation totale des États-Unis} + \text{Consommation totale du Mexique} + \text{Consommation totale du Canada}]/\text{Population totale de l'Amérique du Nord}$$

### Erreur d'utilisation

Il ne suffit pas de faire de bons calculs. Il faut encore que la moyenne veuille dire quelque chose.

Saviez-vous que l'on parle 7 langues locales différentes en Italie : l'italien, bien sûr, mais aussi l'occitan (dans le Piémont), le français (dans le Val d'Aoste), l'allemand (dans le Tyrol), le frioulan (en Vénétie), le slovène (en Vénétie aussi) et le sarde (en Sardaigne). Tout cela sans compter les multiples dialectes de l'italien aux différences beaucoup plus grandes qu'entre, par exemple, le français parlé au Québec ou à Paris. Étant donné que l'Italie comptait 57,8 millions d'habitants en 1992, on pourrait en déduire que chacune des 7 langues parlées en Italie comptait en moyenne 8 257 000 locuteurs et des poussières (soit  $57,8/7$ ).

Le résultat précédent ne présente non seulement aucun intérêt, mais il est même trompeur. En réalité, 56 millions d'Italiens parlent l'italien standard ou un de ses dialectes. Le reste (1,8 million) se partage les 6 autres langues, soit en moyenne 0,3 million ou 300 000 individus par langue (car  $1,8/6 = 0,3$ ). On peut donc dire que les langues secondaires comptent en moyenne 300 000 locuteurs : ça au moins, c'est une moyenne qui veut dire quelque chose.

Avant de faire une moyenne, il faut d'abord déterminer les éléments qui devront être considérés. Des valeurs extrêmes viennent parfois « fausser » des moyennes. Par contre, plus le nombre d'éléments est grand et moins les éléments extrêmes risquent de brouiller les cartes. De plus, en interprétant une moyenne, il faut savoir avec précision qui y est inclus. Ainsi, la moyenne des notes d'un groupe d'étudiants contient-elle uniquement les notes de ceux qui ont suivi le cours jusqu'au bout? Les notes de ceux qui ont abandonné et de ceux qui ne se sont jamais présentés au cours sont-elles incluses? Si oui, comment sont calculées ces notes? Voilà bien des points à éclaircir avant de se lancer dans de savantes comparaisons.



## En bref : Comment calculer une moyenne

### 1) Quand on connaît toutes les valeurs

$$\text{Moyenne} = \frac{\text{Somme des valeurs}}{\text{Nombre de valeurs}}$$

Exemple : Calcul du nombre moyen d'élèves par groupe dans une école primaire.

$$\frac{36 + 36 + 32 + 32 + 32 + 24 + 22 + 20}{8} = \frac{234}{8} = 29,25$$

Nombre d'élèves dans chaque groupe
36
36
32
32
32
24
22
20

### 2) Quand on connaît déjà la somme des valeurs et qu'on a le nombre de valeurs

On applique la même formule que dans le cas précédent.

Ici, un total de 234 élèves est distribué en 8 groupes :  $\frac{234}{8} = 29,25$

### 3) Quand on connaît la fréquence de chacune des valeurs

$$\text{Moyenne} = \frac{(\text{Fréquence 1} \times \text{Valeur 1}) + (\text{Fréquence 2} \times \text{Valeur 2}) + \dots}{\text{Somme des fréquences}}$$

$$\frac{(2 \times 36) + (3 \times 32) + (1 \times 24) + (1 \times 22) + (1 \times 20)}{8} = \frac{72 + 96 + 24 + 22 + 20}{8}$$

$$= \frac{234}{8} = 29,25$$

Valeur (nombre d'élèves)	Fréquence (nombre de groupes)
36	2
32	3
24	1
22	1
20	1
<b>8</b>	

### 4) Quand on connaît la pondération associée à chaque valeur

$$\text{Moyenne} = (\text{Pondération 1} \times \text{Valeur 1}) + (\text{Pondération 2} \times \text{Valeur 2}) + \dots$$

$$(0,25 \times 36) + (0,375 \times 32) + (0,125 \times 24) + (0,125 \times 22) + (0,125 \times 20)$$

$$= 9 + 12 + 3 + 2,75 + 2,5 = 29,25$$

Valeur	Pondération
36	0,250
32	0,375
24	0,125
22	0,125
20	0,125
<b>1,000</b>	

*Notes :* Les pondérations peuvent aussi être exprimées en pourcentages (0,25 n'est autre que 25 % ou 25/100). On note que la somme des pondérations est toujours égale à 1 (ou 100 %). Chaque pondération représente la fréquence relative associée à chaque valeur (exemple : 2 classes sur un total de 8 contiennent 36 élèves, soit 2/8, ou 0,25 ou 25 %).

## EXERCICES 1

---

### 1. 1. Le jour J

Le 6 juin 1944. 745 navires, groupés en 38 convois, approchent des côtes normandes. En tout, 4006 péniches de débarquement sont mises à l'eau, au large des plages, et 185 000 hommes débarquent sur le territoire français. En même temps, 1087 avions de largage, protégés par 13 175 avions de combat, lancent 18 000 parachutistes sur le territoire normand. (Source : Livre Guinness des records, 1995.)

- Combien y avait-il d'hommes, en moyenne, sur chaque navire?
- Combien de navires comptaient les convois, en moyenne?
- Combien y avait-il de parachutistes, en moyenne, dans chaque avion de largage?
- Chaque avion de largage était protégé par combien d'avions de combat, en moyenne?

### 2. Si j'avais un char... ou un ordi

a) Dans le cahier automobile de *la Presse de Montréal* du 17 juillet 1995, six concessionnaires accompagnaient leur publicité du prix total (au comptant) des véhicules proposés (tableau 3.3). Quel est le prix moyen des véhicules proposés aux lecteurs de *la Presse*?

**TABLEAU 3.3 - Le prix des automobiles annoncées par six concessionnaires**

Modèle	Marque	Prix (en \$)
Civic	Honda	14 995
Integra RS	Acura	17 995
Protégé S	Mazda	11 995
Precidia	Mazda	14 495
Civic Special Edition	Honda	14 495
940	Volvo	26 900
Precidia RS	Mazda	14 995
Protégé S	Mazda	11 995
626 Cronos	Mazda	16 995

Source : Cahier automobile de La Presse, 17 juillet 1995.

b) En 2011, année qui marque un premier apogée dans le marché des micro-ordinateurs, le nombre d'unités vendues dans le monde s'élevait à 355,2 millions, pour un chiffre d'affaires total de 329 milliards de \$US. Aux États-Unis seulement, les données étaient respectivement de 95,4 millions d'unités et de 85,5 milliards de \$US (source : [Gartner](#)).

Montrez que le prix moyen d'un micro-ordinateur était alors moins élevé aux États-Unis que dans le reste du monde.

### 3. La moyenne et les fréquences

a) À partir des chiffres du tableau 3.4, calculez le taux d'intérêt moyen des banques canadiennes.

**TABLEAU 3.4 - Les taux d'intérêt sur les dépôts à court terme proposés par les 12 principales banques canadiennes**

<b>Taux proposé (en %)</b>	<b>Fréquence (nombre de banques)</b>	<b>Fréquence relative</b>
5,750	2	16,67
6,000	4	33,33
6,125	1	8,33
6,250	3	25,00
6,500	1	8,33
6,600	1	8,33
<b>Total</b>	<b>12</b>	<b>100,0</b>

Source : Les Affaires, 15 avril 1995.

Note : Dépôts de 30 à 59 jours.

b) Calculez la moyenne des notes de chacun des étudiants du tableau 3.5 (chiffres fictifs).

**TABLEAU 3.5 - Les résultats des étudiants**

<b>Épreuve</b>	<b>Pondération (en %)</b>	<b>Notes de Julie</b>	<b>Notes de Julot</b>
Exercice 1	10	40	80
Examen 1	20	50	80
Exercice 2	10	60	40
Examen 2	20	70	40
Examen final	40	80	60

## 2. L'ÉCART TYPE : MESURER LA DISPERSION DES DONNÉES

---

La moyenne est un outil pratique. En un seul chiffre, elle peut résumer plusieurs dizaines, centaines, voire plusieurs millions de données. Cependant, comme tout résumé, la moyenne néglige les détails. Souvent, ces détails présentent peu d'intérêt, mais parfois ils sont essentiels. On constate par exemple que la température moyenne estivale est de 26 degrés à Montréal (moyenne des maximums de juillet) contre 31 degrés à San Juan de Porto Rico. Et pourtant, il arrive souvent que les Montréalais aient à subir des températures plus élevées que les Portoricains. C'est que le thermomètre, très stable dans les Antilles, n'hésite pas à se balader d'un extrême à l'autre au Québec. Certains jours d'été, Montréal bat même le record de chaleur de toute l'Amérique du Nord.

La même moyenne peut recouvrir des données très proches ou encore des données très éparpillées. On peut imaginer, ainsi, deux groupes d'étudiants dont les moyennes sont égales à 67 %. Dans un cas, la plupart des étudiants ont des notes qui varient de 60 à 75 alors que dans l'autre, les résultats se répartissent à peu près uniformément entre 35 et 95. Ici, nos deux moyennes résument des réalités complètement différentes. Dans cette section, nous verrons comment mesurer ce degré de *dispersion* des données.

### 2.1. Les Beatles : plus que des musiciens

Qui a eu la plus grande influence dans le domaine de la musique populaire depuis la fin de la Deuxième Guerre mondiale? La question étant plutôt subjective, elle pourrait être un bon sujet d'enquête. En attendant, nous avons personnellement couronné les Beatles pour plusieurs raisons. Nés juste un peu avant les baby-boomers et arrivés à maturité dans les années 1960, ils font non seulement des ravages dans les palmarès, mais ils deviennent les idoles de toute une génération, influencent (ou reflètent) les mœurs des jeunes, inspirent même un grand nombre des musiciens populaires qui vont leur succéder. Leur carrière effective ne dure pourtant que 5 ans (de 1964, année de leurs premiers triomphes internationaux à 1969, année de la dissolution officielle du groupe). Toutefois, pendant ces 5 ans leur musique, et le monde, semblent se transformer à toute vitesse.

Mais revenons à nos moutons. Nous avons fait l'inventaire des 12 albums originaux des Beatles (en excluant les compilations) et du nombre de titres qu'ils contiennent (tableau 3.6). Notre but ici est de mesurer à quel point le nombre de titres varie d'un album à l'autre. Une fois que nous aurons construit notre indicateur de dispersion (ici : *l'écart type*), nous pourrons l'appliquer à d'autres situations.

TABLEAU 3.6 - Le nombre de titres sur les albums originaux des Beatles

Album	Année	Nombre de titres	Écart	Écart au carré	Écart absolu
[1]	[2]	[3]	[4]	[5]	[6]
1 <i>Please Please Me</i>	1963	14	0	0	0
2 <i>With the Beatles</i>	1963	14	0	0	0
3 <i>A Hard Day's Night</i>	1964	13	-1	1	1
4 <i>Beatles for Sale</i>	1964	14	0	0	0
5 <i>Help</i>	1965	14	0	0	0
6 <i>Rubber Soul</i>	1965	14	0	0	0
7 <i>Revolver</i>	1966	14	0	0	0
8 <i>Sgt. Pepper</i>	1967	13	-1	1	1
9 Album blanc 1	1968	17	3	9	3
10 Album blanc 2	1968	13	-1	1	1
11 <i>Abbey Road</i>	1969	16	2	4	2
12 <i>Let it Be</i>	1970	12	-2	4	2
<b>Total</b>		<b>168</b>	<b>0</b>	<b>20</b>	<b>10</b>
Nombre	12 (nombre total d'albums)				
Moyenne	14 (nombre moyen de titres par album)				
Variance	1,67				
Écart-type	1,29				

Source : collection personnelle.

Les 12 albums contiennent en tout 168 titres. Il y a donc 14 titres en moyenne par album (car  $168/12 = 14$ ). Cette moyenne va nous servir de point de repère pour évaluer la dispersion. Y a-t-il beaucoup d'albums dont le nombre de titres s'écarte de cette moyenne et, si oui, s'en écartent-ils beaucoup? À première vue, l'écart est plutôt faible (colonne 4 du tableau 3.6). Certains écarts sont positifs, d'autres sont négatifs. La somme des écarts est toujours nulle, puisque la moyenne représente en quelque sorte le *centre de gravité* de tous les éléments.

L'écart absolu mesure l'écart entre une valeur et la moyenne, sans considérer le fait que cet écart soit positif ou négatif.

Pour additionner les écarts sans qu'ils s'annulent, il faut donc se débarrasser de leur signe. C'est ce que nous avons fait dans la colonne 6. En termes absolus, la somme des écarts est de 10 (en bas de la colonne 6). Comme le calcul porte sur 12 éléments (les 12 albums), on peut dire qu'en moyenne, l'écart absolu est de 0,83 (soit  $10/12$ ). Ce simple chiffre suffirait pour nous indiquer que le nombre de titres varie peu d'un album à l'autre.

Pour calculer l'écart absolu moyen

- 1. Calculez la moyenne :  
 $168/12 = 14$
- 2. Mesurez les écarts entre chaque valeur et la moyenne :  
pour les deux premières valeurs :  $14 - 14 = 0$ ; pour la troisième valeur :  $13 - 14 = -1$ ; etc.

- 3. Mettez ces écarts en valeur absolue et additionnez-les :  
 $0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 + 3 + 1 + 2 + 2 = 10$
- 4. Divisez la somme des écarts absolus par le nombre de valeurs :  
 $10/12 = 0,83$

L'écart type est un indice de dispersion des valeurs autour d'une moyenne. 7

En pratique, l'écart absolu est peu utilisé. On lui préfère l'écart type qui possède des propriétés plus intéressantes et qui n'est pas tellement plus compliqué à calculer. Dans la colonne 5 du [tableau 3.6](#), on s'est débarrassé des signes négatifs en mettant au carré tous les écarts à la moyenne. La somme des écarts mis au carré est égale à 20 (au bas de la colonne 5). Pour terminer, et puisque nous avons mis les écarts au carré au point de départ, nous allons faire la racine carrée du résultat obtenu : l'écart type est égal à la racine carrée de 1,66, soit 1,29. Ouf! Cela semble compliqué? Pas du tout : regardez ce qui suit.

Pour calculer l'écart type

- 1. Calculez la moyenne :  
 $168/12 = 14$
- 2. Mesurez l'écart entre chaque valeur et la moyenne
- 3. Mettez chacun de ces écarts au carré
- 4. Additionnez ces écarts au carré :  
 $0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 + 9 + 1 + 4 + 4 = 20$
- 5. Divisez le résultat par le nombre de données :  
 $20/12 = 1,66$
- 6. Faites la racine carrée du résultat :  
 $\sqrt{1,66} = 1,29$

Les données sur lesquelles nous venons de travailler avaient le mérite d'être simples. Elles nous permettaient d'illustrer facilement la méthode de calcul de l'écart type. Elles n'étaient cependant pas d'une grande utilité au niveau de l'interprétation. Il suffisait en effet de jeter un coup d'œil rapide sur la colonne 3 du [tableau 3.6](#) pour constater que le nombre de titres variait peu d'un album à l'autre. Nous allons donc, sans quitter les Beatles, faire appel à un exemple plus significatif.

La question que nous posons est la suivante : la durée des chansons des Beatles a-t-elle évolué avec le temps? Nous avons choisi, pour y répondre, deux albums typiques des Beatles ([tableau 3.7](#)) et nous nous pencherons davantage sur l'analyse des résultats que sur la méthode de calcul. L'album *Help* se situe plus vers le début de la carrière des Beatles (la chanson *Yesterday* se trouve dessus et elle dure 2 minutes et 4 secondes) et *Abbey Road* est le dernier album fait en commun par le groupe (la chanson vedette, *Here Comes the Sun* dure 3 minutes et 4 secondes).

Cette fois, nous avons utilisé notre chiffrier électronique pour calculer les moyennes et les écarts types. Observons les résultats du [tableau 3.7](#), puis essayons de les expliquer. En moyenne, les chansons plus récentes sont plus longues. Mais ce qui est encore plus frappant, c'est que la durée des chansons plus récentes devient de plus en plus variable (l'écart type a beaucoup augmenté). En

1965, les chansons étaient toutes sur le même modèle et duraient entre 2 et 3 minutes. En 1969, c'est beaucoup plus disparate : il y en a de très longues (7:49) et de très courtes (1:06).

Voici notre interprétation de cette évolution, avec tout ce qu'elle peut avoir de subjectif. En 1965, les chansons étaient faites avant tout pour *danser*, les paroles étaient simples, la musique était efficace, énergique et construite selon quelques modèles standards, le public était jeune et insouciant. Il fallait que ça « roule ». En 1969, l'auditoire des Beatles a changé : il est un peu plus vieux, il est plus sérieux, plus intellectuel, plus politisé; la musique est plus sophistiquée, la construction est plus élaborée et plus variée, les paroles sont plus recherchées; les chansons sont plutôt faites pour être *écoutées*.

**TABLEAU 3.7 - Durée des chansons des Beatles selon l'album**

	<b>Help</b>	<b>Abbey Road</b>
(en minutes:secondes)	1965	1969
	2:18	4:16
	1:42	2:59
	2:57	3:24
	2:27	3:28
	2:05	2:49
	2:16	7:49
	3:03	3:04
	2:28	2:45
	1:55	3:57
	2:33	2:31
	2:36	1:06
	2:03	1:13
	2:04	1:58
	2:53	1:31
		1:37
		2:04
<b>Moyenne</b>	<b>2:22</b>	<b>2:54</b>
<b>Écart-type</b>	<b>0:23</b>	<b>1:33</b>

Source : collection personnelle.

Proposons, pour conclure, une autre hypothèse plus terre à terre. Jusqu'au début des années 1950, les disques (78 tours) ne pouvaient contenir plus de 3 minutes de musique. Cela explique pourquoi les compositeurs écrivaient toujours des chansons de 2 ou 3 minutes. Et puis l'habitude est restée pendant près de quinze ans après la mort du 78 tours. Comme quoi, l'homme est parfois esclave de lui-même...

## EXERCICES 2

### 1. Les empereurs romains : l'essayer c'est l'adopter

Quelques dizaines d'années après les [Césars](#), Rome a eu la chance d'être gouvernée par des empereurs « sérieux » : les Antonins. Ces derniers avaient pris l'habitude de se choisir un successeur qualifié en l'adoptant. Un seul fit exception à la règle : Marc-Aurèle, surnommé le *meilleur des empereurs*, qui commit l'erreur de désigner son véritable fils comme héritier. Ce sinistre jeune homme, Commode de son prénom, un mélange de Néron et Caligula, sonna le glas de la dynastie des Antonins et périt assassiné.

TABLEAU 3.8 - Les empereurs romains

(Années)	Naissance	Début du règne	Fin du règne et décès
<i>Les Césars</i>			
Auguste	-63	-24	14
Tibère	-42	14	37
Caligula*	12	37	41
Claude*	-10	41	54
Néron*	37	54	68
<i>Les Antonins</i>			
Nerva	26	96	98
Trajan	53	98	117
Hadrien	76	117	138
Antonin	86	138	161
Marc-Aurèle	121	161	180
Commode*	161	180	192

Source : Petit Robert 2.

\* Mort assassiné.

- À partir des chiffres du tableau 3.8 ci-joint, calculez l'âge moyen des empereurs de la dynastie des Césars au moment de leur accession au trône et de leur mort. Faites les mêmes calculs pour la dynastie des Antonins.
- Calculez la moyenne et l'écart type de la durée du règne pour les trois cas suivants : (1) la dynastie des Césars, (2) la dynastie des Antonins, (3) la dynastie des Antonins à l'exception de son fondateur Nerva.
- Comparez les deux dynasties.

### 2. Chansons pour danser, chansons pour écouter

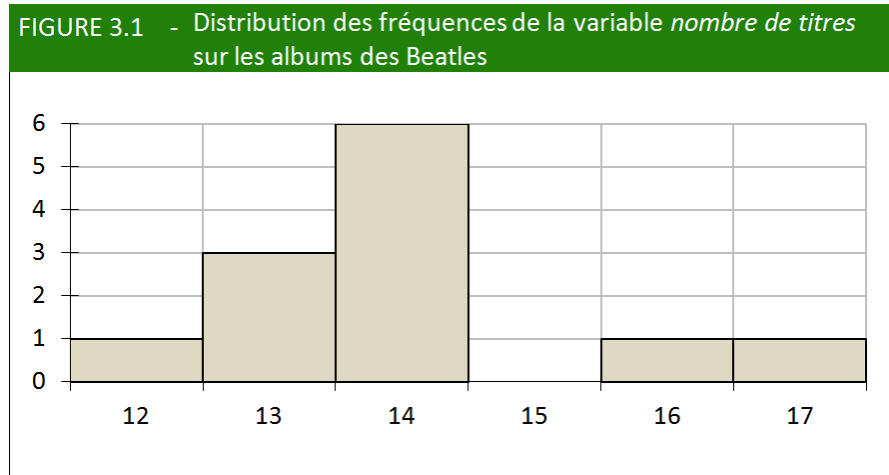
- Vérifiez les moyennes et les écarts types du tableau 3.7.
- Recherche. Choisissez deux disques (avec leur durée) et créez un tableau semblable au [tableau 3.7](#).

Note : vous pouvez remplacer les disques par des livres (avec la longueur de leurs chapitres) des listes de films proposés par des cinémas (avec la durée des films) ou par toute autre situation similaire.

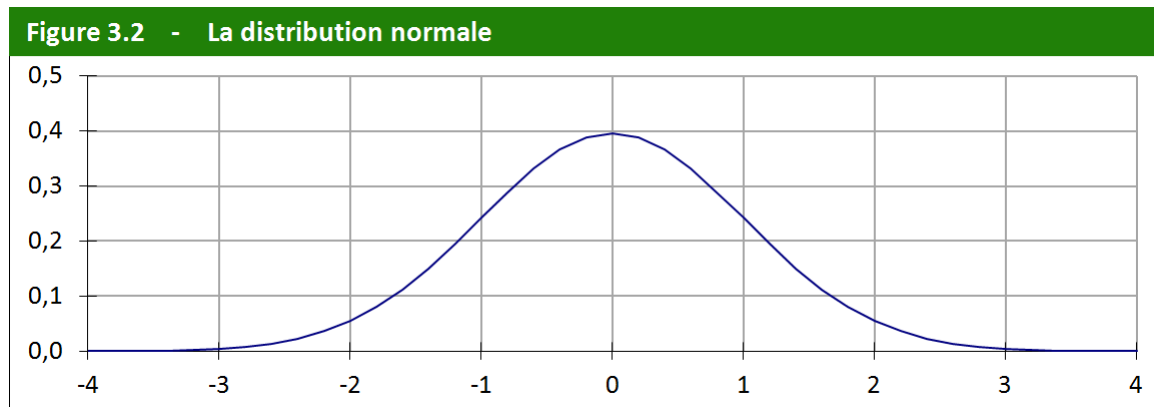


### 3. LA COURBE NORMALE ET L'ÉCART TYPE

Reprenons les données du tableau 3.6 et représentons-les graphiquement sous forme d'un histogramme de fréquences (figure 3.1). Sur l'axe horizontal, nous indiquons les valeurs que peut prendre la variable *nombre de titres* et sur l'axe vertical nous indiquons le nombre de fois (la fréquence) que la variable prend telle ou telle valeur.



Dans de nombreux cas, la distribution des valeurs que prend une variable suit une courbe en forme de cloche (ou de chapeau de Napoléon) : la courbe dite *normale* (figure 3.2). Lorsque la distribution est normale, le fait de connaître la moyenne et l'écart type permettent de trouver une foule d'informations.



#### 3.1. Quand la distribution est-elle normale?

Jusqu'ici, nos exemples étaient relativement simples et les données (les valeurs prises par les variables) étaient peu nombreuses. Notre but était surtout d'illustrer les nouveaux concepts présentés : la *moyenne* et l'*écart type*. Dans la pratique, lorsque les données sont trop peu nombreuses, certaines d'entre elles (celles qui s'éloignent trop de la tendance générale) peuvent jouer les trouble-fête et « fausser » la moyenne et l'écart type. Lorsque les éléments sont très nombreux, les trouble-fête ont par contre moins de chance de se faire remarquer : ils sont noyés dans la masse.

### La loi normale apparaît dans de très nombreux contextes.

Lorsque les données sont suffisamment nombreuses et qu'elles dépendent d'une multitude de facteurs dont aucun ne prédomine sur les autres, la distribution de ces données est souvent normale. Le nombre de clients à la banque le jeudi, le temps que nous passons sous la douche, la grosseur des œufs pondus par une poule X, l'épaisseur des anneaux d'un tronc d'arbre ont des bonnes chances de suivre la *distribution normale*. Par contre, la répartition de la population selon l'âge ne suit pas une courbe normale pas plus que la distribution des salaires dans une entreprise.

Même si le nombre de valeurs est peu élevé (il n'y a que 12 albums considérés), on sent déjà poindre la courbe normale dans la distribution du nombre de titres par album 33 tours chez les Beatles. Le nombre de titres est en effet lié à de multiples facteurs d'ordre commercial (l'attente des clients), artistique (l'inspiration des musiciens), matériel (la capacité d'un disque), économique (le coût de production), etc. À plus long terme cependant, il se peut qu'une variable (comme l'apparition du disque compact) vienne changer les règles du jeu de façon radicale et modifier complètement l'allure de la distribution.

La variable *taille des individus* possède les caractéristiques de la distribution normale. Plus on s'éloigne de la taille moyenne, que ce soit en montant ou en descendant, et moins on il y a d'individus. Nous allons utiliser cet exemple pour montrer comment, grâce à la courbe normale et deux simples chiffres (la moyenne et l'écart type), on peut répondre à un tas de questions intéressantes.

## 3.2. L'écart type et la courbe normale

Au Québec, la taille moyenne des hommes est de 176 cm. Évidemment, il y en a de plus grands et de plus petits. L'écart type, qui mesure cette dispersion, est de 7 cm. Pour les femmes, la moyenne est de 162 cm et l'écart type de 6 cm. Dans un cas comme celui-ci, les hommes diffèrent systématiquement des femmes et il serait maladroit de les mettre dans le même paquet. Notons par ailleurs que la taille moyenne a tendance à augmenter depuis un siècle.

Pour rendre les explications un peu plus simples, nous avons un peu arrondi les chiffres. Nous nous sommes basés sur les résultats (non publiés) de l'Enquête nationale sur la santé de la population menée en 1994-95 par Statistique Canada, selon laquelle la taille moyenne des Canadiens adultes (18 ans et plus) est estimée à 176,15 cm pour les hommes et à 161,93 cm pour les femmes. En passant, et selon la même enquête, le poids moyen des Canadiens adultes est de 80,3 kg pour les hommes et de 64,6 kg pour les femmes. À titre de comparaison, on pourra consulter le [tableau 3.11](#) qui contient la taille moyenne et le poids moyen des jeunes Australiennes. En ce qui concerne l'écart type de la taille des Canadiens, nous n'avons malheureusement pas été en mesure d'obtenir de données précises (et gratuites), malgré d'ardentes négociations avec les autorités. Nous avons donc dû procéder à nos propres estimations pour l'écart type, en nous basant sur des données américaines et des microdonnées ultérieures.

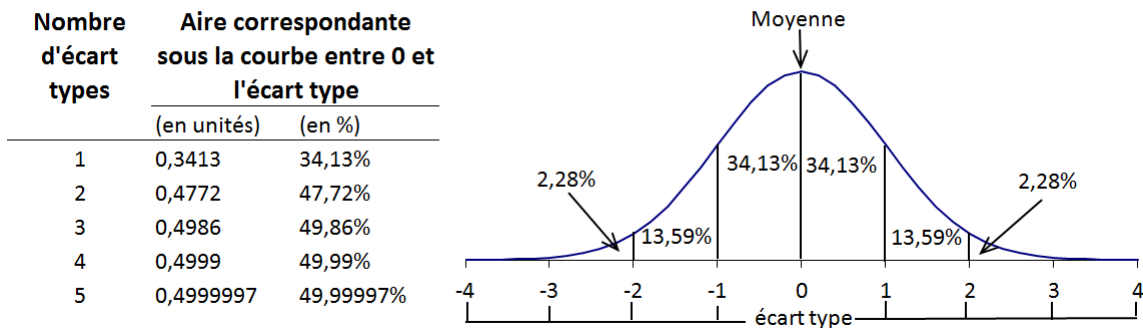
Les simples chiffres que nous venons de mentionner vont nous permettre de répondre aux questions suivantes :

- Julie mesure 168 cm. Est-elle une grande fille?
- Julot mesure 169 cm. Est-il un petit gars?

- Le châssis d'une porte a 2,11 m de haut. Sur les 1000 personnes différentes qui franchiront cette porte au cours de l'année, y en a-t-il beaucoup qui risquent de se cogner la tête?

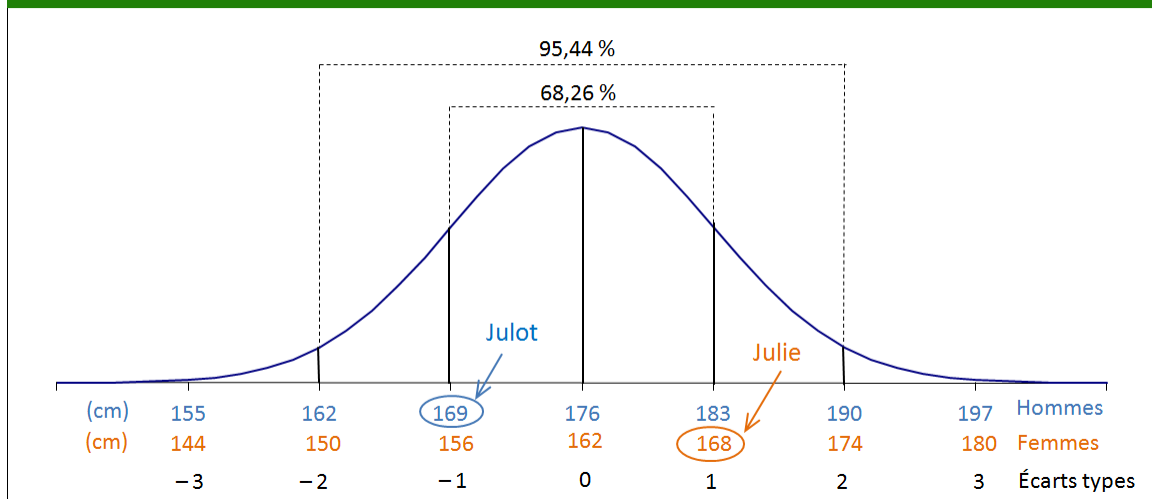
Il faut d'abord savoir que, dans une distribution normale, la proportion de sujets qui se trouve dans un intervalle donné est toujours la même. Pour être plus concret, disons que 34,13 % des sujets sont situés entre la moyenne (point central ou point 0 sur l'axe horizontal de la figure 3.3) et 1 écart type; 13,59 % des éléments sont situés entre 1 et 2 écarts types de la moyenne; le restant (2,28 %) est situé à plus de 2 écarts types de la moyenne. Il en va de même pour les sujets dont les valeurs sont inférieures à la moyenne (partie gauche de la courbe de la figure 3.3). La surface totale sous la courbe représente 100 % : 50 % à gauche et 50 % à droite de la moyenne.

**Figure 3.3 - La distribution normale et l'écart type**



Pour revenir à nos Québécoises, on peut affirmer que Julie est parmi les plus grandes (voir la figure 3.4) puisque 84,13 % des Québécoises sont plus petites qu'elle : les 50 % plus petites que la moyenne et les 34,13 % comprises entre la moyenne et 168 cm. Il reste quand même encore 14,87 % de femmes qui sont plus grandes que Julie. Quant à Julot, il est dans la situation inverse : 84,13 % des Québécois sont plus grands que lui.

**Figure 3.4 - La distribution normale et la taille des Québécois**



Réglons enfin le cas de la porte de 2,11 m. Utilisons l'écart type des hommes, qui sont le plus susceptibles de se cogner la tête. Cela nous met à 35 cm de la moyenne (211 cm – 176 cm = 35 cm), soit 5 écarts types (car 5 × 7 cm = 35 cm). Le tableau qui accompagne la [figure 3.3](#) montre que seulement 0,00003 % des sujets (soit 50 % – 49,99997 %) s'éloignent à plus de 5 écarts types de la moyenne. Cette proportion de 0,00003 % équivaut à 3 personnes sur 10 millions. Il y a donc fort peu

de chances pour qu'un tel individu, si tant est qu'il se présente, se cogne un jour au cadre de porte. À moins d'utiliser des talons particulièrement épais.

Le nombre d'écart types qu'il y a entre une valeur particulière et la moyenne de toutes les valeurs est souvent appelé la *cote z*. La cote z de Julie est égale à +1 et celle de Julot est égale à -1, celle du monsieur qui mesure 2,11 mètres et qui se cogne dans les châssis de porte est égale à +5.

$$\text{Cote } z = (\text{Valeur particulière} - \text{Moyenne}) / \text{Écart type}$$

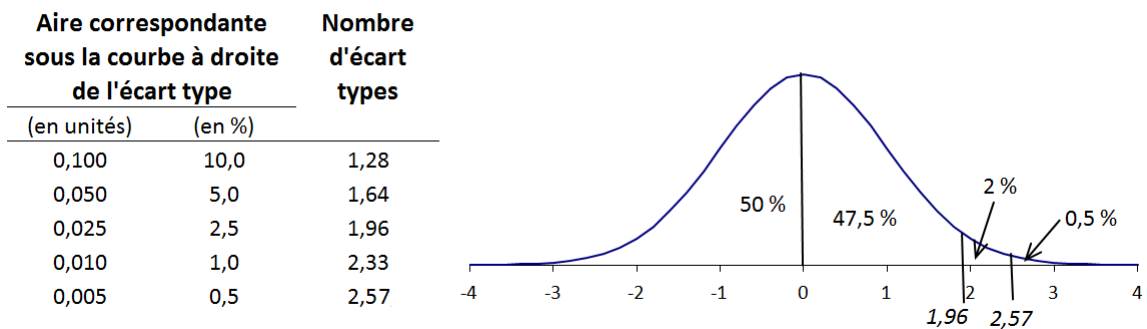
$$\begin{aligned} \text{Cote } z \text{ de Julie} &= (168 - 162) / 6 = 6 / 6 = 1 \\ \text{Cote } z \text{ de Julot} &= (169 - 176) / 7 = -7 / 7 = -1 \\ \text{Cote } z \text{ du grand monsieur} &= (211 - 176) / 7 = 35 / 7 = 5 \end{aligned}$$

Posons maintenant les questions à l'envers :

Si nous excluons les 0,5 % de Québécois les plus petits et les 0,5 % de Québécois les plus grands, quelle est la taille maximum et quelle est la taille minimum des 99 % de Québécois restants? (La question gagne à être posée séparément pour les hommes et pour les femmes.)

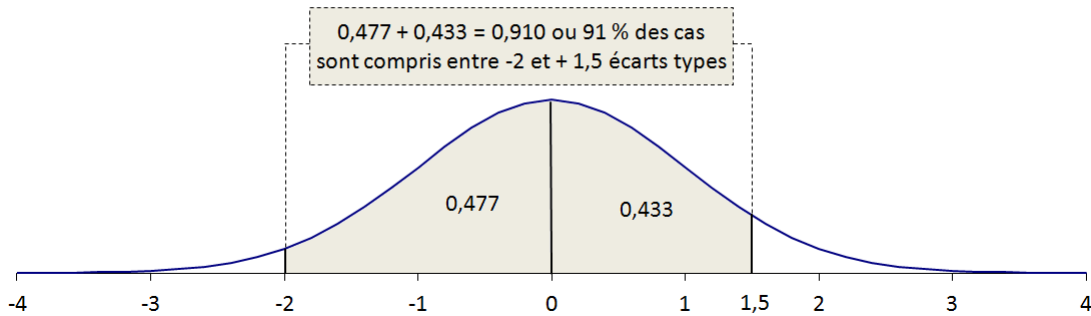
La figure 3.5 reprend la distribution normale sous l'angle que nous venons de lui donner. La proportion de 0,5 % correspond à 2,57 écart types de part et d'autre de la moyenne. Rappelons que la taille *moyenne* des hommes est de 176 cm et l'*écart type* correspondant, de 7 cm. Pour les hommes, 2,57 écart types correspondent à 18 cm (2,57 × 7 cm). Il y a donc 99 % des hommes qui mesurent entre 158 cm (soit 176 - 18) et 194 cm (soit 176 + 18).

**Figure 3.5 - Découpage de la distribution normale en chiffres ronds**



De la même manière, 99 % des femmes, dont la taille moyenne est de 162 cm avec un écart type de 6 cm, mesurent entre 147 cm (162 - [2,57 × 6]) et 177 cm (162 + [2,57 × 6]).

**En bref : Comment déterminer la position sur une courbe normale**



A) La valeur du sujet est connue : on cherche sa position dans la distribution

Exemple : Antonine mesure 171 cm. Comment se situe-t-elle par rapport aux autres adultes québécoises, dont la taille moyenne est de 162 cm avec un écart type de 6 cm?

1. Convertissez la distance entre la valeur du sujet et la moyenne en nombre d'écarts types. (Le résultat peut être positif ou négatif.)

$$\text{Nombre d'écarts types (cote } z) = (\text{Sujet} - \text{Moyenne}) / \text{Écart type}$$

$$(171 - 162)/6 = 9/6 = 1,5$$

2. Obtenez la valeur de l'aire correspondante en consultant la [table de distribution normale](#) ci-après (sans tenir compte du signe).

1,5 écart type correspond à une aire de 0,433 ou 43 %.

3. Ajoutez 0,5 (ou 50 %) au résultat.

Le chiffre obtenu représente la proportion de sujets inférieurs à notre sujet (si le résultat calculé en 1 était positif) ou supérieurs à notre sujet (dans le cas contraire).

$$0,433 + 0,5 = 0,933 \text{ ou } 93,3 \%$$

93,3 % des Québécoises sont plus petites qu'Antonine, et les autres (6,7 %) sont plus grandes qu'elle.

B) La position est connue : on cherche la valeur du sujet correspondant

Exemple : Césarine a une taille inférieure de deux écarts types à la moyenne québécoise.

1. Convertissez les écarts types en unités de mesure de la variable en question, et ajoutez ce résultat à la moyenne.

$$\text{Valeur du sujet} = \text{Moyenne} + (\text{Nombre d'écarts types} \times \text{Valeur de l'écart type})$$

$$162 \text{ cm} + (-2 \times 6 \text{ cm}) = 150 \text{ cm}$$

C) On cherche la proportion du nombre de sujets compris entre 2 sujets donnés

Exemple : Quelle est la proportion de Québécoises plus petites qu'Antonine et plus grandes que Césarine?

1. Convertissez la valeur des 2 sujets en nombre d'écarts types.
2. Obtenez la valeur de l'aire correspondante en consultant la table de distribution normale.
3. Si les deux écarts types sont de signe opposé : additionnez les aires obtenues à l'étape précédente; s'ils sont de même signe, soustrayez la plus petite aire de la plus grosse.

Antonine : +1,5 écart type; Aire = 0,433

Césarine : -2 écarts types; Aire = 0,477

Résultat :  $0,433 + 0,477 = 0,91$  (ou 91 %)

Conseil : Faites un dessin de la courbe normale et découpez cette dernière en tranches. Vous pourrez mieux délimiter la section qui vous intéresse.

#### Table de distribution normale

Les valeurs en bleu représentent l'aire comprise entre la moyenne et l'écart-type correspondant.

Écart type	Aire	Écart type	Aire	Écart type	Aire	Écart type	Aire
0,00	0,000	1,00	0,341	2,00	0,477	3,00	0,499
0,05	0,020	1,05	0,353	2,05	0,480	3,05	0,499
0,10	0,040	1,10	0,364	2,10	0,482	3,10	0,499
0,15	0,060	1,15	0,375	2,15	0,484	3,15	0,499
0,20	0,079	1,20	0,385	2,20	0,486	3,20	0,499
0,25	0,099	1,25	0,394	2,25	0,488	3,25	0,499
0,30	0,118	1,30	0,403	2,30	0,489	3,30	0,500
0,35	0,137	1,35	0,411	2,35	0,491	3,35	0,500
0,40	0,155	1,40	0,419	2,40	0,492	3,40	0,500
0,45	0,174	1,45	0,426	2,45	0,493	3,45	0,500
0,50	0,191	1,50	0,433	2,50	0,494	3,50	0,500
0,55	0,209	1,55	0,439	2,55	0,495	3,55	0,500
0,60	0,226	1,60	0,445	2,60	0,495	3,60	0,500
0,65	0,242	1,65	0,451	2,65	0,496	3,65	0,500
0,70	0,258	1,70	0,455	2,70	0,497	3,70	0,500
0,75	0,273	1,75	0,460	2,75	0,497	3,75	0,500
0,80	0,288	1,80	0,464	2,80	0,497	3,80	0,500
0,85	0,302	1,85	0,468	2,85	0,498	3,85	0,500
0,90	0,316	1,90	0,471	2,90	0,498	3,90	0,500
0,95	0,329	1,95	0,474	2,95	0,498	3,95	0,500

Vous trouverez une [table normale détaillée](#) en annexe.

## EXERCICES 3

---

### 1. La taille des Québécois

La taille des Québécois de sexe masculin est distribuée de façon normale avec une moyenne de 176 cm et un écart type de 7 cm.

- Quelle est la proportion de Québécois mesurant plus de 183 cm?
- Quelle est la proportion de Québécois mesurant plus de 185 cm?
- Quelle est la proportion de Québécois mesurant moins de 162 cm?
- Quelle est la proportion de Québécois mesurant entre 183 cm et 185 cm?
- César-Auguste Tremblay mesure 2 écarts types de moins que la moyenne de ses concitoyens. Quelle est sa taille?

### 2. Cote z

On suppose, dans l'exemple fictif suivant, que les notes sont distribuées de façon normale.

- Les 150 élèves du professeur X ont obtenu une note moyenne de 64 points à l'examen final (avec un écart type de 12 points). Julie, qui est également une élève du professeur X a obtenu pour sa part la note de 88 points. Calculez la cote z de Julie.
- Pour les élèves du professeur Y, la moyenne est de 60 points et l'écart type de 10. Julot, élève du professeur Y, a obtenu une note de 85 points. Calculez la cote z de Julot.
- Julie et Julot ont tous deux posé leur candidature à l'université. Le directeur de l'établissement où ces examens se sont déroulés recommande en premier lieu la candidature de Julot, estimant que ce dernier a obtenu de meilleurs résultats que Julie. Que pensez-vous de l'attitude de ce directeur?

### 3. Courbe normale et proportions

Les questions suivantes portent sur une variable distribuée de façon normale.

- Quelle est la proportion de valeurs situées entre 0 et 1 écart type?
- Quelle est la proportion de valeurs situées entre 0 et 2 écarts types?
- Quelle est la proportion de valeurs situées au-delà de 3 écarts types?
- Quelle est la proportion de valeurs situées entre -1 et 1 écart type?
- Quelle est la proportion de valeurs situées entre -1 et 2 écarts types?

## 4. D'AUTRES INDICATEURS DE MOYENNE ET DE DISPERSION

En dehors de la moyenne et de l'écart type, il existe quelques mesures pratiques et simples pour évaluer le niveau de dispersion des éléments d'une distribution.

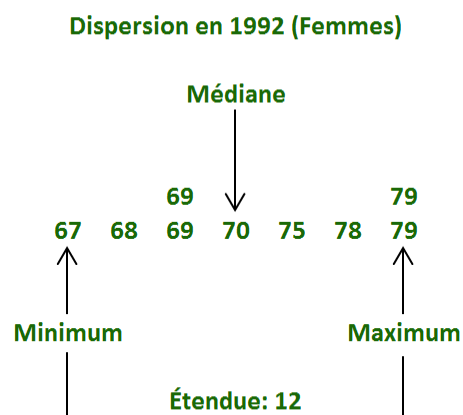
### 4.1. Le milieu et les bords : autres mesures

La figure 3.6 nous servira à illustrer ces mesures. Nous avons retenu dans ce tableau les pays de langue espagnole de l'Amérique centrale et des Antilles : ce sont des pays d'importance similaire (petits pays en superficie et en population) et homogènes (même région, même culture). Tout au long du commentaire qui va suivre, nous utiliserons les chiffres de la première colonne (les femmes en 1992).

**Figure 3.6 - Espérance de vie à la naissance (en années)**

#### Pays hispanophones d'Amérique centrale

	Femmes Hommes		Femmes Hommes	
	1992		2012	
Guatemala	67	62	75,3	68,2
Honduras	68	64	76,0	71,2
El Salvador	69	64	76,9	67,5
Nicaragua	69	65	77,6	71,5
R. Dominicaine	70	65	76,5	70,2
Panama	75	71	80,3	74,6
Porto Rico	78	71	82,4	74,9
Cuba	79	73	81,1	77,1
Costa Rica	79	74	82,0	77,5
<b>Moyenne</b>	<b>72,7</b>	<b>67,7</b>	<b>78,7</b>	<b>72,5</b>
<b>Écart-type</b>	<b>4,7</b>	<b>4,3</b>	<b>2,6</b>	<b>3,5</b>
<b>Coefficient de variation (en %)</b>	<b>6,5</b>	<b>6,3</b>	<b>3,3</b>	<b>4,8</b>



Source : Banque mondiale, IDM.

L'étendue mesure l'écart entre les deux valeurs extrêmes d'une distribution.

Les *valeurs extrêmes* sont le *minimum* (ici 67) et le *maximum* (ici 79). L'intervalle dans lequel la variable prend ses valeurs, ou *étendue*, est la différence entre les deux extrêmes (ici  $79 - 67 = 12$ ).

La *médiane* est la valeur autour de laquelle on retrouve deux groupes représentant chacun la moitié des données d'une distribution.

La médiane représente la valeur de l'élément du milieu de la distribution. Il faut évidemment que les valeurs soient, au préalable, classées par ordre de grandeur. Il y a ici 9 valeurs. La valeur de l'élément du milieu (le cinquième élément) est de 70 : c'est la médiane de cette distribution. Il existe autant de valeurs inférieures à 70 que de valeurs supérieures à 70 dans notre distribution. La médiane permet donc de couper la distribution en 2 parties égales.

Dans une certaine mesure, la médiane nous indique le milieu, mais elle n'est pas nécessairement égale à la moyenne. Dans notre exemple, la médiane est plutôt inférieure à la moyenne (qui est de 72,7 : vérifiez-le si le cœur vous en dit). Dans une distribution normale, par contre, la médiane



coïncide avec la moyenne. Enfin, lorsque le nombre de valeurs est pair (ici, il était impair), on coupe la poire en deux : la médiane est alors égale à la moyenne des 2 valeurs centrales.

Au Québec, l'âge médian de la population était de 25,6 ans en 1971. Cela signifie que la moitié des Québécois avaient moins de 25,6 ans. Vingt ans plus tard, cette médiane se situait à 34 ans. En 2012, au moment du « printemps érable », la médiane avait atteint 41,5 ans, et une bonne partie de la population se montrait peu sympathique aux revendications de la jeunesse. Par la suite, la médiane commença à plafonner : par définition, le « vieillissement de la population » ne peut pas être éternel!

Le mode est la (ou les) valeur(s) qui revient (qui reviennent) le plus souvent dans une distribution.

Le mode est la valeur qui revient le plus souvent dans la distribution. Ici, il y a deux modes : la valeur 69 et la valeur 79 reviennent 2 fois chacune. On dira que la distribution est bimodale (ou à deux bosses, comme un chameau). Avouons, pour nuancer, que notre distribution comporte peu d'éléments et qu'aucune valeur ne se détache vraiment nettement. Si on regroupait les éléments, on pourrait cependant observer nettement 2 catégories de pays : ceux qui tournent autour de 68 ou 69 et ceux qui tournent autour de 78 ou 79.

Le coefficient de variation mesure le rapport entre l'écart type et la moyenne.

Le coefficient de variation est une variante un peu plus raffinée de l'écart type. Dans le tableau 3.9, l'écart type est plus grand pour les femmes que pour les hommes (4,7 pour les femmes et 4,3 pour les hommes). Le coefficient de variation représente l'écart type relatif. Ici, le coefficient de variation est à peu près identique pour les femmes ( $4,7/72,7 = 0,065 = 6,5\%$ ) que pour les hommes ( $4,3 / 67,7 = 0,063 = 6,3\%$ ). À vous déterminer si le coefficient de variation convient mieux à la situation étudiée que l'écart type. Car les méthodes quantitatives sont plus une question de jugement que de calcul.

$$\text{Coefficient de variation} = \text{Écart type} / \text{Moyenne.}$$

$$\text{Coefficient de variation de l'espérance de vie des femmes} = 4,7/72,7 = 0,065 = 6,5\%$$

Au fait, pourquoi les femmes vivent-elles généralement plus longtemps que les hommes? Posez la question à votre entourage et vous récolterez sans doute de beaux spécimens de préjugés. La cause des écarts serait tout banalement de nature chimique : le métabolisme des hommes est plus élevé que celui des femmes (d'où, entre autres, des besoins supérieurs en calories ingérées) et la « machine » s'userait plus vite.

## 4.2. Les quintiles : cinq parts égales?

Par quintile on entend chacune des quatre valeurs qui partagent une distribution (rangée en ordre) en cinq groupes de même effectif et, par extension, ces cinq groupes eux-mêmes.

En plus d'identifier le milieu, la médiane permettait de couper la distribution de valeurs en 2 paquets, ou *tranches*. Eh bien, il arrive aussi que l'on coupe la distribution en 5 tranches (voir le tableau 3.10 sur lequel figurent les revenus d'une population divisée en 5 tranches : les plus pauvres, les assez pauvres, ceux du milieu, les assez riches et les plus riches). On parle alors de quintiles (qui n'est autre que le mot « cinquième » en latin).

Dans le tableau 3.9, les quintiles sont utilisés de façon particulière pour observer comment les revenus se distribuent dans la population. Aux Pays-Bas, le premier quintile le plus pauvre (c'est à dire les 20 % des Néerlandais les plus pauvres) se partage 8,2 % du revenu national (voir le tableau

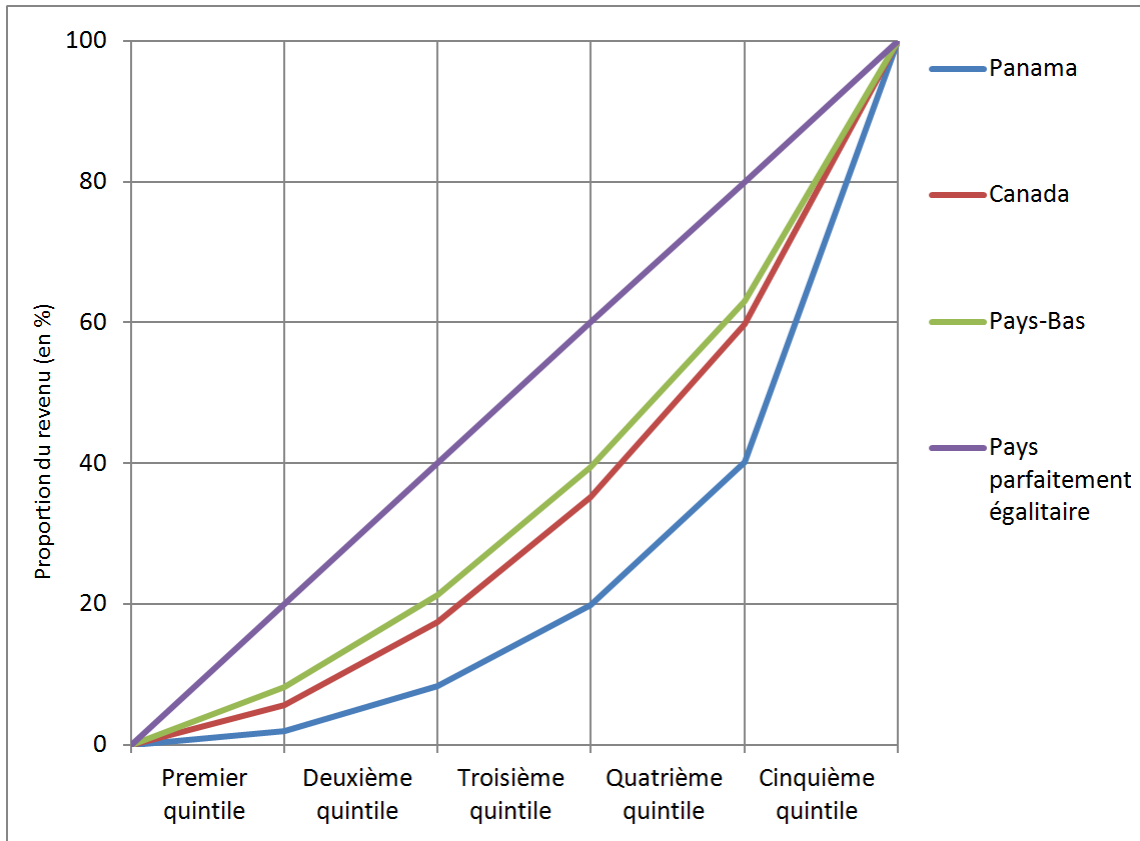
3.9). De l'autre côté, la tranche la plus riche de la population (le cinquième quintile) se partage 36,9 % du revenu. La figure qui accompagne le tableau permet d'évaluer d'un coup d'œil le degré d'inégalité de répartition du revenu dans un pays. Les courbes représentent les proportions *cumulées* du revenu au fur et à mesure qu'on ajoute des tranches. Le Panama est le pays qui s'éloigne le plus d'une répartition égalitaire. Cette représentation graphique s'appelle la courbe de Lorenz.

**TABLEAU 3.9 - Répartition du revenu dans divers pays**

(en %)		Quintile le				
		plus pauvre	Deuxième quintile	Troisième quintile	Quatrième quintile	Quintile le plus riche
Panama 1989	Proportion du revenu	2,0	6,3	11,6	20,3	59,8
	<i>Proportion cumulée</i>	2,0	8,3	19,9	40,2	100,0
Canada 1987	Proportion du revenu	5,7	11,8	17,7	24,6	40,2
	<i>Proportion cumulée</i>	5,7	17,5	35,2	59,8	100,0
Pays-Bas 1988	Proportion du revenu	8,2	13,1	18,1	23,7	36,9
	<i>Proportion cumulée</i>	8,2	21,3	39,4	63,1	100,0
Pays parfaitement égalitaire	Proportion du revenu	20,0	20,0	20,0	20,0	20,0
	<i>Proportion cumulée</i>	20,0	40,0	60,0	80,0	100,0

Source : Banque mondiale, Rapport sur le développement dans le monde 1994.

Note : On a ajusté le revenu afin de tenir compte du pouvoir d'achat dans chaque pays.

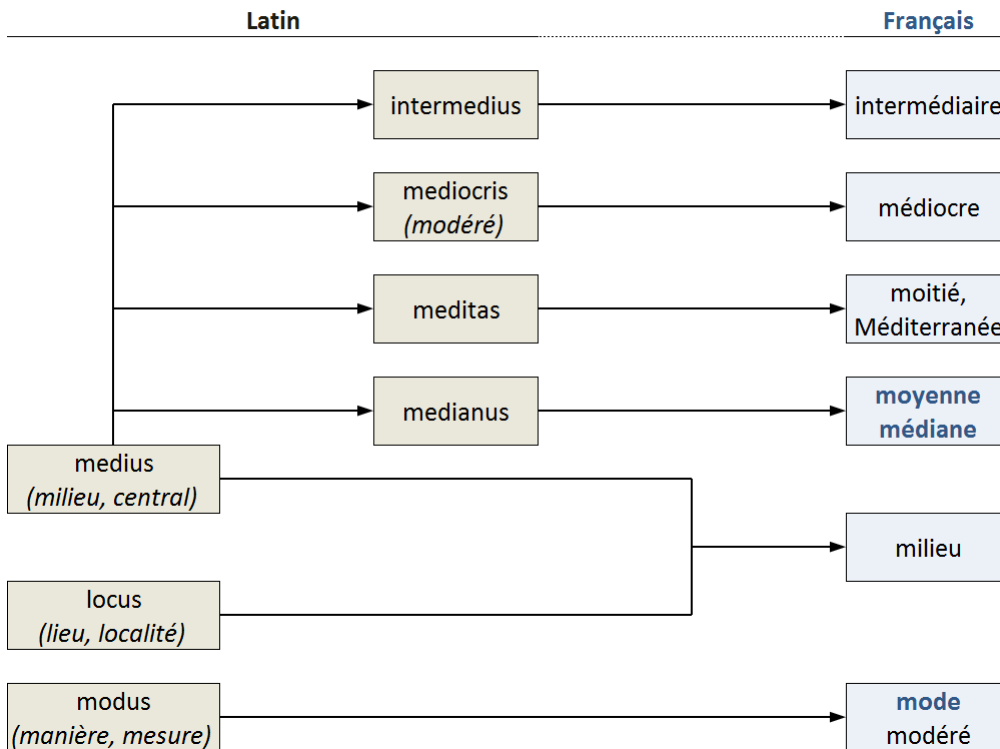


Parfois, on découpe une distribution en 4 tranches (on a alors des quartiles), en 10 tranches (les déciles), voire en 100 tranches (les centiles). Le principe reste le même.

### 4.3. Pour ne pas y perdre son latin

Nous venons de présenter beaucoup de mots : moyenne, médiane, milieu, moitié. Ce qui compte, c'est de bien comprendre chaque concept et non d'impressionner l'auditoire avec du jargon. D'ailleurs, les quatre mots que nous venons de citer ont tous la même origine : ils proviennent du mot latin *medius* qui veut dire *au milieu*. Observez votre main : votre médus (appelé aussi le majeur) n'est-il pas la médiane de votre collection de doigts? La figure 3.6 montre l'origine de tous ces mots.

Figure 3.7 - L'origine des mots



Parfois, le mot *moyenne* est utilisé à toutes les sauces. Le Québécois « moyen » n'est pas la moyenne des Québécois. Il s'agit plutôt du Québécois *ordinaire* (est-ce Elvis Gratton?) Le mot *moyenne* est utilisé dans le sens d'*intermédiaire* dans les expressions : classe moyenne, Moyen âge, Julot est de taille moyenne, placement à moyen terme. Ces exemples ne font que témoigner de la richesse du langage, reflet de la richesse humaine.

#### 4.4. Pour les curieux : la moyenne harmonique

La moyenne dont nous avons parlé jusqu'ici est la *moyenne arithmétique* (ou *moyenne* tout court). Dans la presque totalité des cas, cette moyenne est amplement suffisante (pour le vérifier, demandez à un de vos professeurs de sciences humaines s'il connaît la moyenne harmonique ou la moyenne géométrique).

La « colle » suivante nous permettra d'expliquer simplement la *moyenne harmonique* (nous verrons la moyenne géométrique dans le prochain chapitre). Notre but est uniquement de montrer, encore une fois, que l'utilisation des chiffres en sciences humaines est plus une question de jugement qu'une question de calcul.

Un automobiliste fait l'aller-retour entre Montréal et Gatineau (400 km) en parcourant les 100 premiers kilomètres à 50 km/h, les 100 kilomètres suivants à 100 km/h, les 100 suivants à 150 km/h et les 100 derniers à 200 km/h. Voilà un ex-chauffeur (son permis lui a été retiré suite à cette expérience) qui, une fois bien réchauffé, n'a pas froid aux yeux. Notre question est la suivante : quelle est la vitesse moyenne de l'automobile sur l'ensemble du parcours?

Étant donné que le parcours est divisé en 4 tronçons égaux, on serait tenté d'additionner les 4 vitesses et de les diviser par 4.

$$(50 \text{ km/h} + 100 \text{ km/h} + 150 \text{ km/h} + 200 \text{ km/h})/4 = 500/4 = 125 \text{ km/h.}$$

Mais si les 4 sections du parcours sont de *longueur égale*, le *temps* qu'il faut pour parcourir chacun d'entre eux est différent. Or, la vitesse dépend aussi du temps. Nous connaissons déjà la distance totale, il nous suffit de calculer le temps total du trajet pour obtenir la vitesse moyenne.

Il faut 2 heures pour parcourir les 100 premiers km (à 50 km/h), 1 heure pour le tronçon suivant, 40 minutes pour le troisième tronçon et 30 minutes pour le dernier, soit en tout 4 h 10 min pour parcourir 400 km. La vitesse moyenne est donc de 400 km/4 h 10 min. Notons qu'avant d'effectuer cette division, il faut convertir les minutes en fractions d'heures : 10 min = 10/60 d'heure = 0,166 heure. Exprimé sous forme décimale, le temps du parcours est de 4,166 heures. La vitesse moyenne est de 400 km/4,166 heures = 96 km/h.

Ce dernier résultat (le bon : 96 km/h) est bien inférieur à celui de notre premier calcul (le faux : 125 km/h). N'est-ce pas la preuve que « rien ne sert de courir, il faut partir à point »?

Pour calculer la moyenne harmonique :

- 1. Additionner l'inverse de chaque valeur  
 $1/50 + 1/100 + 1/150 + 1/200 = 12/600 + 6/600 + 4/600 + 3/600 = 25/600 = 1/24$
- 2. Diviser le nombre de valeurs par le résultat précédent.  
 $4/(1/24) = 4 \times (24/1) = 96.$

Moyenne harmonique = Nombre de valeurs / [(1/Valeur 1) + (1/Valeur 2) + ...]
--

Il y a une leçon à tirer de cet exemple. La formule ou même le nom de la moyenne harmonique ne parviendront peut-être pas à s'implanter dans votre mémoire, mais vous vous souviendrez sans doute de la chose suivante : il est plus important de réfléchir aux données du problème que de se lancer aveuglément dans des calculs.

La moyenne harmonique est employée lorsque les variables considérées sont en réalité des rapports (ici : la distance divisée par le temps) et lorsque c'est le dénominateur qui varie (ici : le temps).

## EXERCICES 4

---

### 1. Les hommes

À l'aide des données du [tableau 3.6](#), calculez l'étendue, la médiane, le mode et le coefficient de variation de l'espérance de vie des hommes des pays hispanophones d'Amérique centrale en 1992. Justifiez vos réponses avec quelques calculs si nécessaire.

## EXERCICES SUPPLÉMENTAIRES

### 1. Le train? Pas dangereux!

Le tableau suivant indique le nombre de personnes tuées dans des accidents de train dans la patrie du TGV (train à grande vitesse).

TABLEAU 3.10 - Nombre de personnes tuées dans des accidents de train en France entre 1972 et 1992

1972	1973	1974	1975	1978	1982	1983	1984	1985	1987	1988	1991	1992
107	3	10	9	1	3	7	1	78	2	52	14	0

Source : Quid 1995.

Note : Aucune victime en 1976, en 1977, entre 1979 et 1981, en 1986, en 1989, en 1990 et en 1992.

- Calculez la moyenne du nombre annuel de victimes.
- Calculez l'écart type du nombre annuel de victimes.

### 2. Une moyenne délicate

Il y a en Inde 880 millions d'habitants dont 340 millions parlent l'hindi. On dénombre en tout 845 langues différentes dans ce pays.

La Papouasie-Nouvelle-Guinée ne compte que 4,1 millions d'habitants. Le pays est très montagneux et presque entièrement recouvert de forêts. On y parle pas moins de [700 langues\\*](#).

Source : *L'état du monde 1994*. Données de 1992.

- Quel est le nombre moyen de locuteurs par langue en Inde?
- Quel est le nombre moyen de locuteurs par langue en Nouvelle-Guinée?

Note : un locuteur est une personne qui parle une langue.

### 3. Seul dans le noir

Selon *La Culture en perspective* (Statistique Canada 87-004-XPB, hiver 1995), il y avait en 1993-94 au Canada 664 cinémas disposant en tout de 1727 écrans. Les 78 812 milliers d'entrées payantes ont alors rapporté 400,5 millions de dollars de recettes.

- Quel était alors le nombre moyen d'écrans par cinéma?
- Quel était le nombre moyen de spectateurs (payants) par écran?
- Quelle était la recette moyenne par entrée?

### 4. Simon, José, Miguel et les autres

Un texte célèbre de José Martí, poète et patriote cubain du XIX<sup>e</sup> siècle, raconte comment un voyageur fourbu arrivant à Caracas s'enquiert, avant toute chose, de l'emplacement de la statue du libérateur Simon Bolivar. Puis, l'auteur fait le portrait et l'éloge de Bolivar, de José de San Martín et de Miguel Hidalgo. De ce dernier, Martí dit qu'il était de la race des hommes bons, c'est à dire des

hommes qui *veulent savoir*. N'est-ce pas là l'idéal de l'étudiant en sciences humaines? (Référence : José Martí, *Tres héroes*.)

- a) Cherchez la date de l'indépendance des pays d'Amérique hispanophones.
- b) Calculez la moyenne et l'écart type de ces dates.
- c) Commentez.

### 5. Une moyenne piégée

Il y a 1000 familles dans la ville. La moitié des familles de la ville compte trois enfants. Le reste des familles n'a pas d'enfant (chiffres fictifs).

- a) Quel est le nombre moyen d'enfants par famille?
- b) Quel est le nombre moyen de frères et sœurs que peut avoir un enfant?

### 6. Un comportement normal?

En 1993, une [enquête](#)\* menée auprès de 304 étudiants (118 garçons et 186 filles) dans deux universités de l'est des États-Unis a donné, entre autres, les résultats suivants : l'âge moyen de la première relation sexuelle était de 16,47 pour les garçons (avec un écart type de 2,4) et de 17,4 pour les filles (avec un écart type de 2,3). Nous supposons, pour les besoins de l'exercice, que la variable *âge de la première relation* était distribuée de façon normale, aussi bien pour les garçons que pour les filles.

Source : *Journal of Health Education*, mai-juin 1996, p. 144 à 152.

Pour faire cet exercice, vous pouvez utiliser la table de distribution normale présentée [dans ce chapitre](#), ou la table plus détaillée fournie [en annexe](#).

- a) Quelle est la proportion de filles ayant eu une première relation sexuelle avant l'âge de 17,4 ans? avant l'âge de 19,7 ans? après l'âge de 19,7 ans? avant l'âge de 21 ans? entre l'âge de 17,4 ans et de 19,7 ans? entre l'âge de 16 ans et 20 ans?
- b) Quel est l'âge à partir duquel on peut affirmer que 1 % des garçons ont eu une première relation sexuelle? 10 % des garçons? 10 % des filles? 20 % des garçons? 25 % des garçons? 50 % des garçons? 50 % des filles?

### 7. Un pays sans classes

En 1992, le centile le plus fortuné de la population américaine détenait 42 % du patrimoine (contre 21 % en 1979) et 46 % des actions cotées en Bourse. Le décile le plus fortuné détenait, quant à lui, 89 % des actions cotées en Bourse. Dans le même intervalle de temps (1979-1992), la famille médiane n'avait augmenté son patrimoine que de 10 %. (Sources : *Courrier international*, 8 février 1996.)

- a) Quelle était la proportion des actions détenues par les 1 % des Américains les plus fortunés en 1992? par les 9 % suivants? par les 90 % les moins fortunés?
- b) Quelle était la proportion d'Américains qui détenaient un patrimoine plus élevé que la famille médiane?
- c) Montrez que l'écart entre riches et pauvres s'est accru aux États-Unis entre 1979 et 1992.

d) Entre 1992 et 2012, le PIB américain par habitant en dollars constants a augmenté de 37 % (source : Banque mondiale, IDM). En combinant ce fait aux informations contenues dans la [figure 2.3](#), pouvez-vous déterminer si les inégalités se sont accrues ou non au cours de ces vingt ans aux États-Unis?

## 8. Changement de médiane

[Au Canada\\*](#), l'âge médian des personnes infectées par le VIH est passé de 32 ans en 1982 à 27 ans en 1983-84 et à 23 ans en 1994. Compte tenu de cela, que pensez-vous des affirmations suivantes?

Source de ces données : *Tendances sociales*, été 1996, SC. 11-008-XPF.

- La moitié des personnes infectées en 1994 avaient 23 ans ou plus.
- Les personnes infectées en 1994 avaient en moyenne 23 ans.
- La moyenne d'âge des personnes infectées a sensiblement diminué en l'espace de 12 ans.

## 9. À l'autre bout du monde

Répondez aux questions suivantes à l'aide du tableau 3.11.

**TABLEAU 3.11 - Les femmes australiennes**

		1964	1974	1984	1994
1 Poids moyen (16 à 20 ans)	kg	55,5	56,2	56,9	57,6
2 Taille moyenne (16-20 ans)	cm	162,2	162,9	163,6	164,3
3 Espérance de vie	années	74,2	74,8	78,3	81,0
4 Pourcentage de fumeuses adultes	%	28,0	30,0	31,0	23,3
5 Âge médian du mariage	années	24,3	21,4	24,2	26,6
6 Taux de mariage	‰	7,7	8,1	7,4	6,2
7 Durée médiane du mariage	années	14,3	10,7	7,7	7,7
8 Taux de divorce	%	0,7	1,3	2,8	2,7
9 Âge médian des maternités	années	26,4	25,7	27,1	29,0
10 Utilisation de la pilule contraceptive (20-24 ans)	%	25,0	46,0	50,0	52,2
11 Pourcentage des filles ayant réussi leur 12 <sup>e</sup> année	%	18,7	31,6	48,0	79,9
12 Pourcentage des femmes à l'université	%	25,9	38,9	46,6	53,5
13 Proportion des femmes dans la population active	%	25,0	34,0	40,0	43,0
14 Proportion des femmes mariées	%	18,0	39,8	43,5	52,0
15 Ratio des gains femmes/hommes	%	59,0	74,6	78,0	81,6

Source : The Bulletin, 16 avril 1996, Sydney, d'après une compilation de plusieurs études.

- Le tableau contient 3 moyennes, 3 médianes, 6 proportions et 3 autres types de rapports. Identifiez-les.
- Quelle est la proportion des femmes dont le mariage dure plus de 7,7 ans en 1994?
- Expliquez pourquoi les phrases suivantes ont des bonnes chances de « s'avérer » fausses! (Note : tous les chiffres concernent l'année 1994.)



- i) La moitié des femmes mesure 164,3 cm.
  - ii) Les femmes se marient en moyenne autour de 26,6 ans.
  - iii) 6,2 % des femmes sont mariées.
  - iv) La majorité des femmes utilise la pilule contraceptive.
  - v) Il y a environ deux fois plus d'hommes que de femmes à l'université.
  - vi) Les hommes gagnent 18,4 % de plus que les femmes.
- d) Tracez des courbes illustrant l'évolution du poids moyen et de la taille moyenne des Australiennes au cours des décennies.
- e) Quelle est l'évolution du ratio taille/poids au cours des décennies?

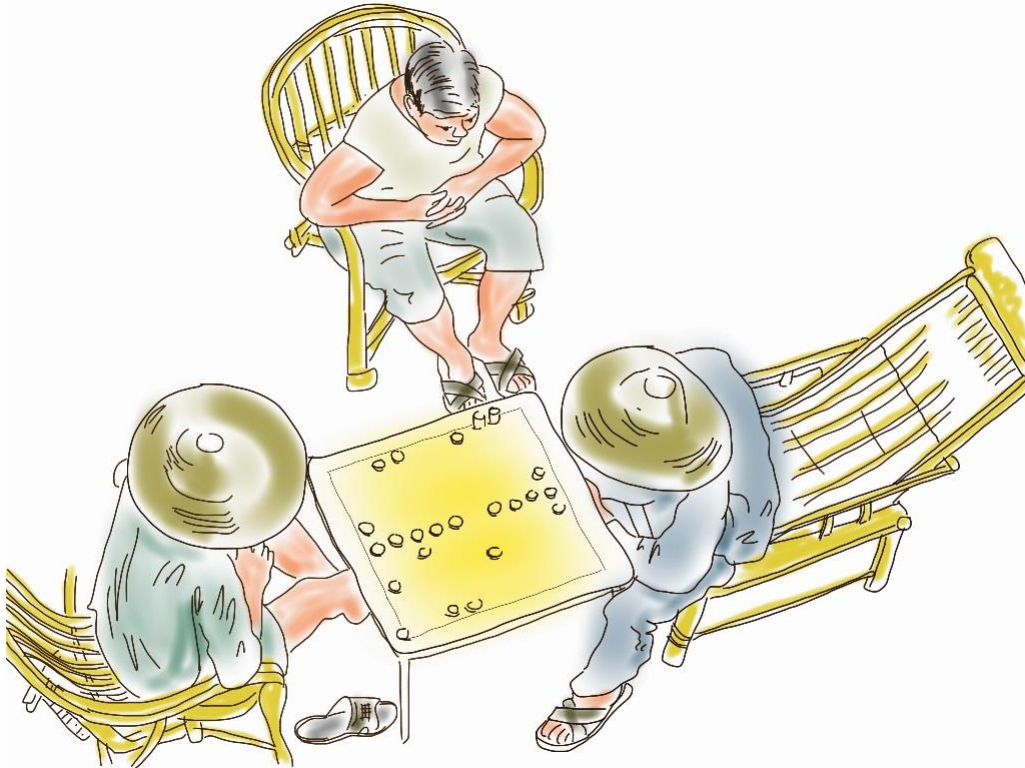
## 10. Recherche

Essayez d'obtenir l'âge des députés, des conseillers municipaux, des professeurs d'une école ou d'un département ou de tout autre groupe intéressant. Calculez la moyenne d'âge et l'écart type. Comparez les résultats entre eux : différences entre hommes ou femmes, entre périodes et entre lieux différents.

## DOSSIER 3 LA RÉFORME DE L'ÉCRITURE CHINOISE

Apprendre à lire et à écrire le chinois n'est pas chose facile, même en Chine.

À cause de sa complexité, l'écriture chinoise est longtemps demeurée l'apanage de la classe dominante. Depuis la proclamation de la république (en 1911) et plus encore depuis la révolution de 1949, la réforme de l'écriture chinoise est à l'ordre du jour. On a songé à transcrire le chinois en alphabet latin (le système *pinyin*), comme cela a déjà été fait, par exemple, pour le vietnamien. Un élève vietnamien peut en effet apprendre à écrire en 100 heures alors qu'il en faut 500 pour un élève chinois.



Toutefois, l'utilisation de l'alphabet latin pose de gros problèmes. Il y a en chinois beaucoup de mots différents qui ont la même prononciation et donc la même transcription en alphabet phonétique pinyin. Écrits en caractères chinois, ces mots perdent leur ambiguïté, un peu comme les mots français *ô, oh, eau, au, haut, os* qui ne se distinguent que par leur orthographe différente. C'est pourquoi l'alphabet phonétique pinyin n'a eu qu'un succès limité. Il sert surtout à orthographier les noms propres de façon standardisée (Mao Zedong, Sichuan et Beijing) et à enseigner le chinois aux étudiants étrangers débutants.

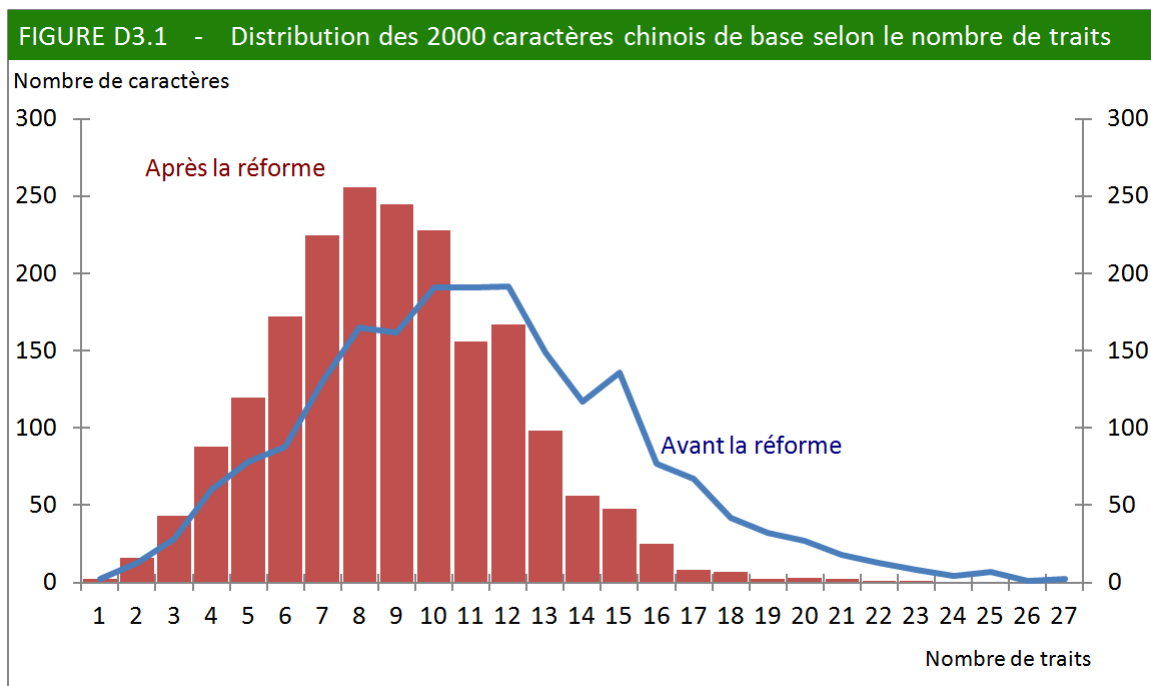
D'ailleurs, les caractères chinois possèdent un avantage évident. Ils permettent à un Pékinois, un Cantonais et un Japonais de communiquer — de façon élémentaire — par écrit, bien qu'ils parlent des langues différentes. Toutes proportions gardées, le Français, l'Anglais et l'Arabe en font autant

lorsqu'ils s'échangent des numéros de téléphone, qui s'écrivent de la même façon, mais qui se prononcent différemment.

### Au lieu d'une révolution, pourquoi pas une réforme?

Plutôt que d'abandonner les caractères chinois, on a alors cherché à les simplifier, en réduisant le nombre de traits nécessaires pour les tracer. Le comité mis en place dans les années 1950 a établi la liste des 2000 caractères de base que toute personne devrait connaître pour fonctionner convenablement dans la vie courante. Puis il a décidé d'en simplifier 795 et d'en supprimer 31, qui faisaient théoriquement double emploi.

La figure D3.1 indique la distribution des 2000 caractères de base en fonction du nombre de traits, avant et après la réforme. Le nombre moyen de traits par caractère a ainsi été réduit à 8,9, avec un écart type de 3,2. De plus, 71 % des caractères s'écrivent désormais avec 10 traits ou moins alors qu'avant la réforme, cette proportion n'était que de 46 %.



Calculer la moyenne et l'écart type de 2000 éléments n'est pas une mince affaire avec une calculatrice, même si ces éléments sont groupés en une vingtaine de catégories. Par contre, la chose est relativement facile avec un chiffrier électronique. C'est ce que nous avons fait pour les 1969 caractères d'après la réforme (voir tableau D3.1). À vous de trouver la moyenne et l'écart type pour les anciens caractères à l'aide d'un chiffrier électronique.

### Un petit casse-tête... chinois

Quel rapport existe-t-il entre feu le président Mao Zedong, mort à Beijing et l'illustre metteur en scène Kurosawa Akira, né à Tokyo? Aucun apparemment, du moins si on se fie uniquement à la transcription en alphabet latin. Toutefois, si on utilise les caractères chinois, on s'aperçoit que *ze* et *sawa* s'écrivent de la même façon et veulent donc dire la même chose : le *marais* ou l'*abondance*. Tokyo est simplement la capitale de l'est et Beijing la capitale du nord : *jing* et *kyo* s'écrivent de la même manière et signifient *capitale*. C'est pourquoi l'annonce dans le journal d'une visite de Kurosawa à Mao aurait pu être lue ainsi indifféremment par un Chinois ou un Japonais : M. Marais-

noir rencontre M. Poil Marais-de-l'est à sa descente de l'avion qui dessert la ligne capitale-de-l'est/capitale-du-nord (voir figure D3.2).

FIGURE D3.2 - Un petit casse-tête... chinois

Français	Chinois	Caractère simplifié	Caractère traditionnel	Japonais	Nombre de traits
noir	hei	黑	黑	kuro	12
marais	ze	泽	澤	sawa	8
nord	bei	北	北	kita	5
est	dong	东	東	to	5
capitale	jing	京	京	kyo	8
poil	mao	毛	毛	ke	4
quatre	si	四	四	yon	5
rivière	chuan	川	川	kawa	3
cap	qi	崎	崎	saki	11

Exemple : 8 traits

丶	一	一	一	一	一	一	一
---	---	---	---	---	---	---	---

À quand le jumelage entre Trois-Rivières et le Sichuan, et entre la rivière du Cap-rouge (en banlieue de Québec) et Kawasaki (en banlieue de Tokyo)?

## QUESTIONS

- Comparez les deux distributions de la figure D3.1 à la courbe normale.
- À première vue, quelle est la distribution qui possède la plus grande moyenne? Quelle est la distribution qui a le plus grand écart type?
- Calculez le nombre moyen de traits par caractère avant la réforme.
- Calculez l'écart type du nombre de caractères avant la réforme.
- Calculez la proportion de caractères de plus de 12 traits avant et après la réforme.