Analyse des données Licence Pro 2024-2025

Cours n°2- l'analyse univariée

Florian Bayer

Objectifs du cours

L'analyse d'une série de données est une étape obligatoire de **toutes** productions scientifiques.

Si on s'intéresse aux caractères pris un à un, on parle d'analyse **univariée**.

Elle permet :

- De mieux **comprendre** les données les unes par rapport aux autres (série homogène ? Hétérogène ?).
- De mettre en évidence les **valeurs remarquables**.
- Elle permet **résumer** l'information.
- Elle donne les éléments scientifiques pour **justifier et reproduire** ses choix.

Ce qui se traduit aussi en cartographie par :

- Trouver le meilleur **compromis** entre information statistique et information géographique.
- **Résumer** l'information tout en conservant la **forme de la distribution**.
- Si besoin de mettre en évidence les **valeurs remarquables** et de les faire apparaître sur la carte.
- Choisir la méthode de **discrétisation** la plus adaptée à vos données.

Objectifs du cours

L'analyse univariée se fait à l'aide de la combinaison **graphiques** et **calculs** :

- Les graphiques permettent de visualiser instantanément les caractéristiques d'une ou de plusieurs séries.
- Les calculs statistiques donnent des éléments factuels et reproductibles.

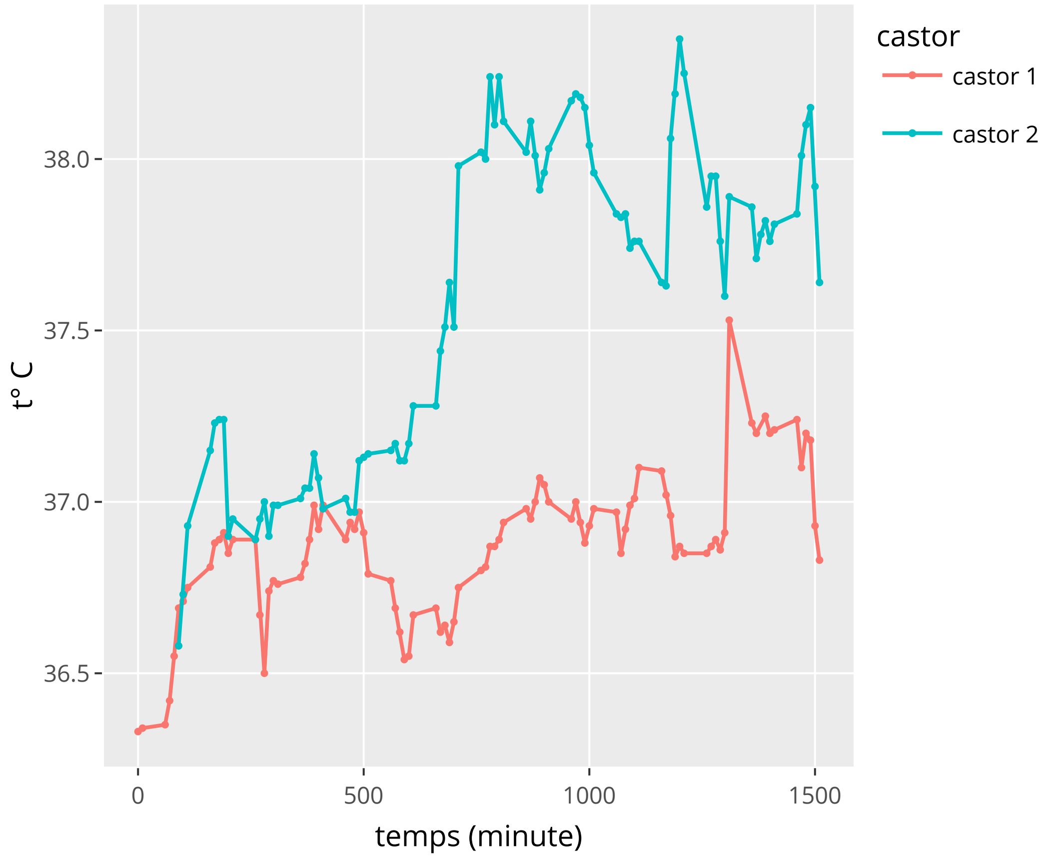
1- Analyse univariée : les graphiques

Intérêt

« La représentation des données sous forme numérique n'a souvent qu'une assez faible lisibilité immédiate. Il est alors important de pouvoir utiliser des moyens qui améliorent l'appréhension, l'interprétation et la communication des données statistiques » [Dumolard et al.,2003]

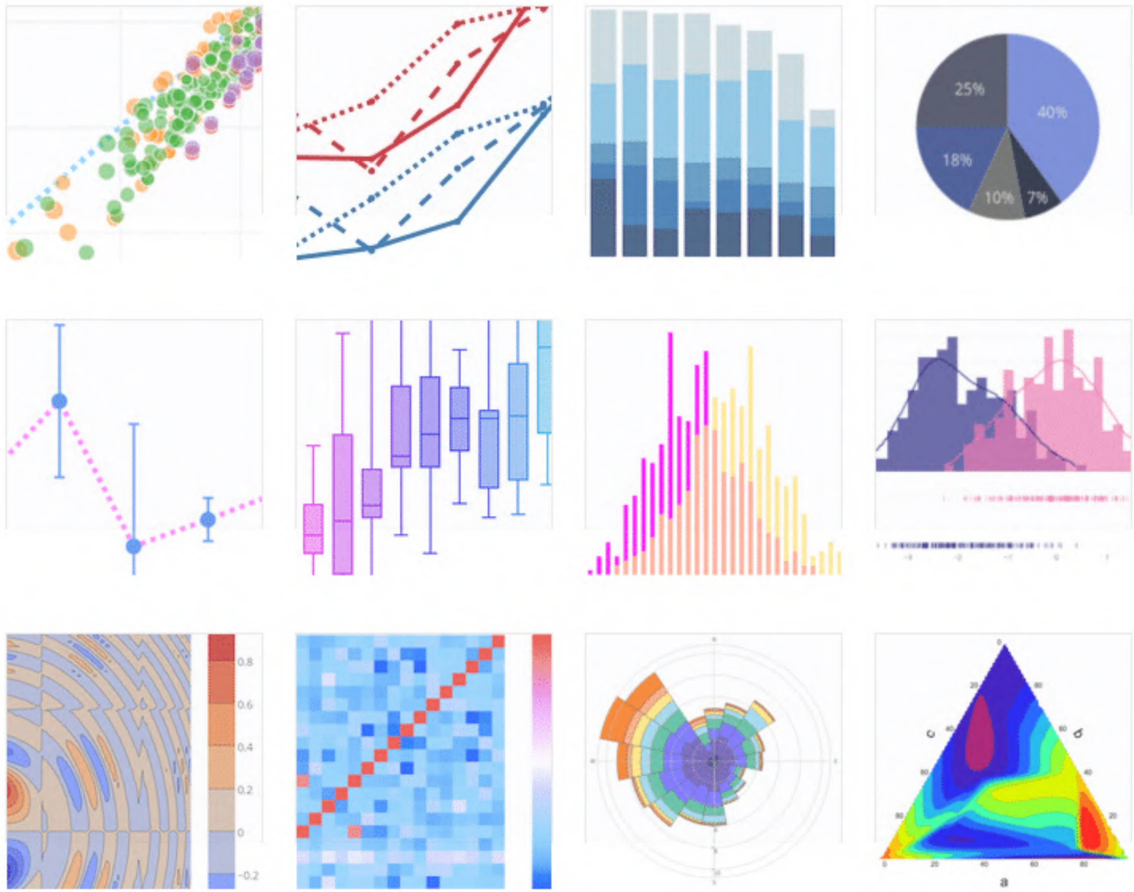
Mesure de la t° de 5 castors

time	temp	castor
0	36.33000	castor 1
2	37.09129	castor 3
7	36.80903	castor 4
10	36.34000	castor 1
12	36.54754	castor 4
13	37.04813	castor 3
13	36.65537	castor 4
16	37.88445	castor 5
35	37.99484	castor 5
39	36.80177	castor 3
40	37.91439	castor 5



Des graphiques en fonction des données

Selon le type de données (qualitatives / quantitatives, discrètes ou continues), les outils permettant de décrire l'information statistique ne sont pas les mêmes.



Les diagrammes en bâton

On représente généralement les **données qualitatives discrètes / exhaustives** à l'aide d'un diagramme en bâton. Il faut réaliser un **dénombrage** à l'aide un tableau de contingence.

Le diagramme en bâton :

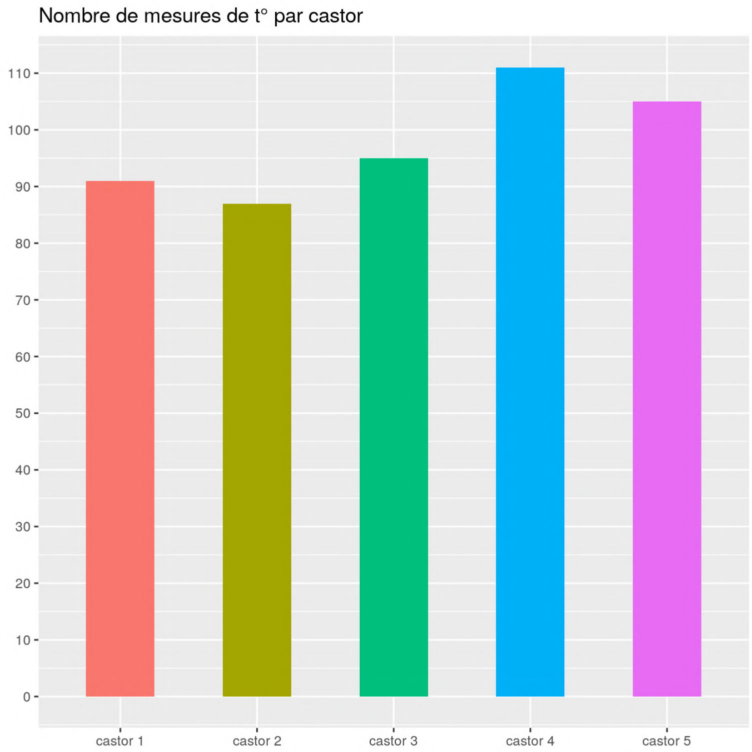
- La hauteur de chaque bâtons (y) est égale au nombre d'individus dans la modalité correspondante.
- Les variables discrètes sont en abscisse (x).

C'est donc la représentation d'une variable qualitative pondérée par le nombre d'occurence de la valeur dans le tableau.

Les diagrammes en bâton : exemple 1

On peut compter puis représenter le nombre de mesures de température pour les castors à l'aide d'un dénombrement

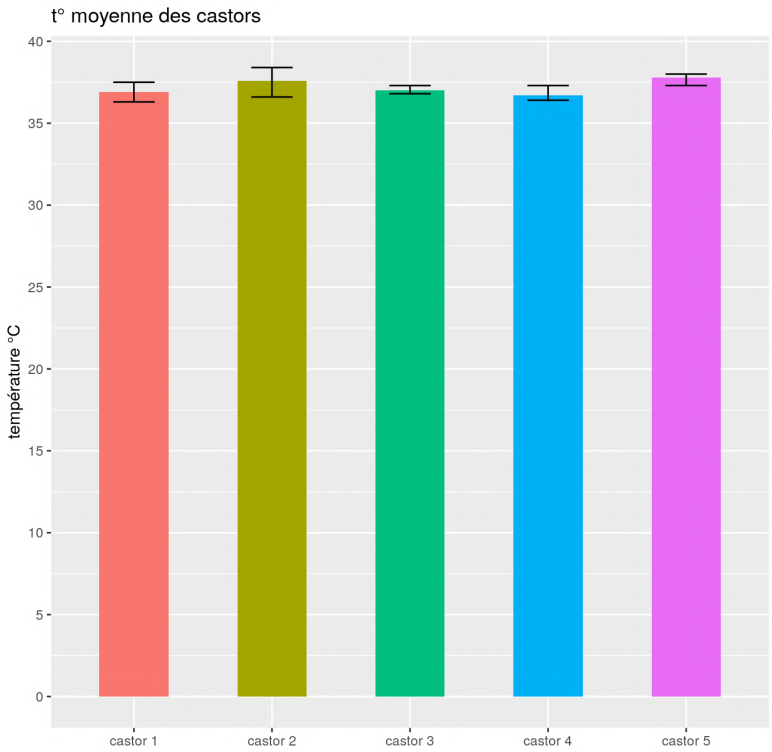
castor	Nombre de mesures
castor 4	111
castor 5	105
castor 3	95
castor 1	91
castor 2	87



Les diagrammes en bâton : exemple 2

On peut compter également calculer la moyenne, le minimum, le maximum etc. des températures

castor	Moyenne	Min	Max
castor 1	36.9	36.3	37.5
castor 2	37.6	36.6	38.4
castor 3	37.0	36.8	37.3
castor 4	36.7	36.4	37.3
castor 5	37.8	37.3	38.0

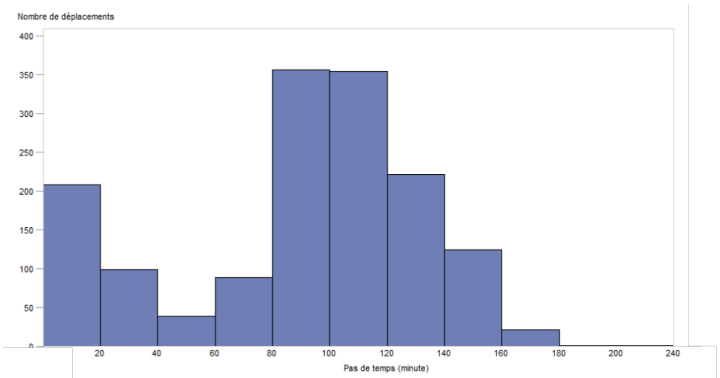


Les histogrammes

Les données de stock se représentent par un **histogramme** et non un diagramme en bâton. La surface à représenter doit être proportionnelle à la quantité contenue dans la classe.

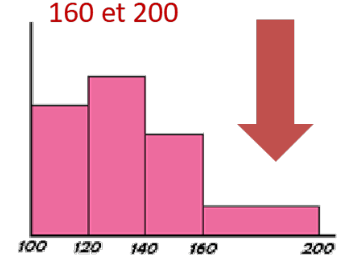
L'idéal est de créer des **classes d'amplitudes égales** (les classes vont de 10 en 10 ou de 20 en 20). Le graphique ressemblera alors à un diagramme en bâton, sans discontinuité.

Si les classes ne sont pas de même amplitudes (de 10 à 20, puis de 20 à 120), la surface de la colonne (et non sa hauteur) doit être proportionnelle au nombres d'individus représentés. Dans la cas contraire, l'impression visuelle ne sera pas conforme à la réalité. Peu de logiciels proposent cette foncitonnalité, il est donc préférable de faire des classes d'amplitudes égales.



Amplitudes égales

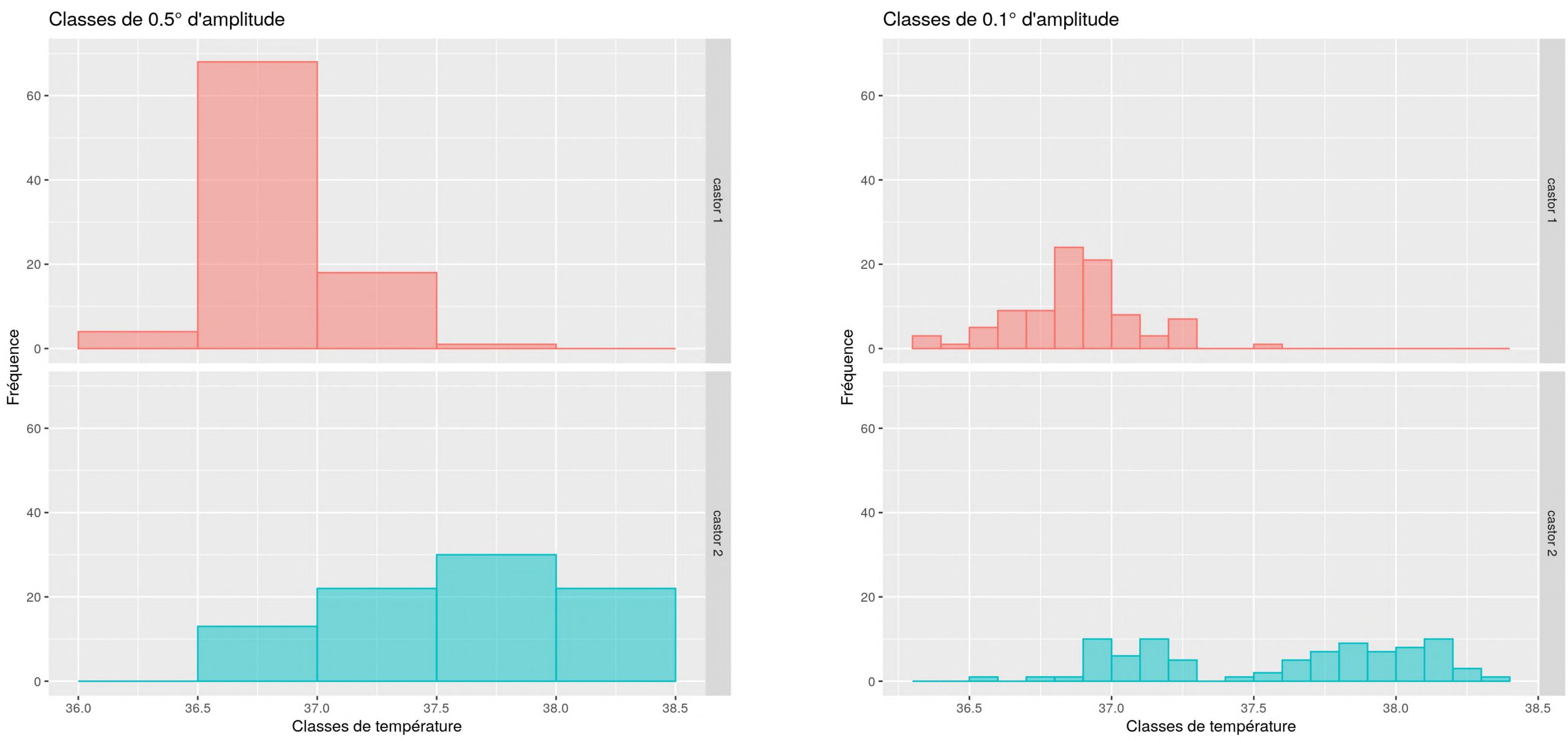
Les classes vont de 20 en 20, sauf la dernière entre 160 et 200



Amplitudes non égales, mais le graphique est juste

Les histogrammes : exemples

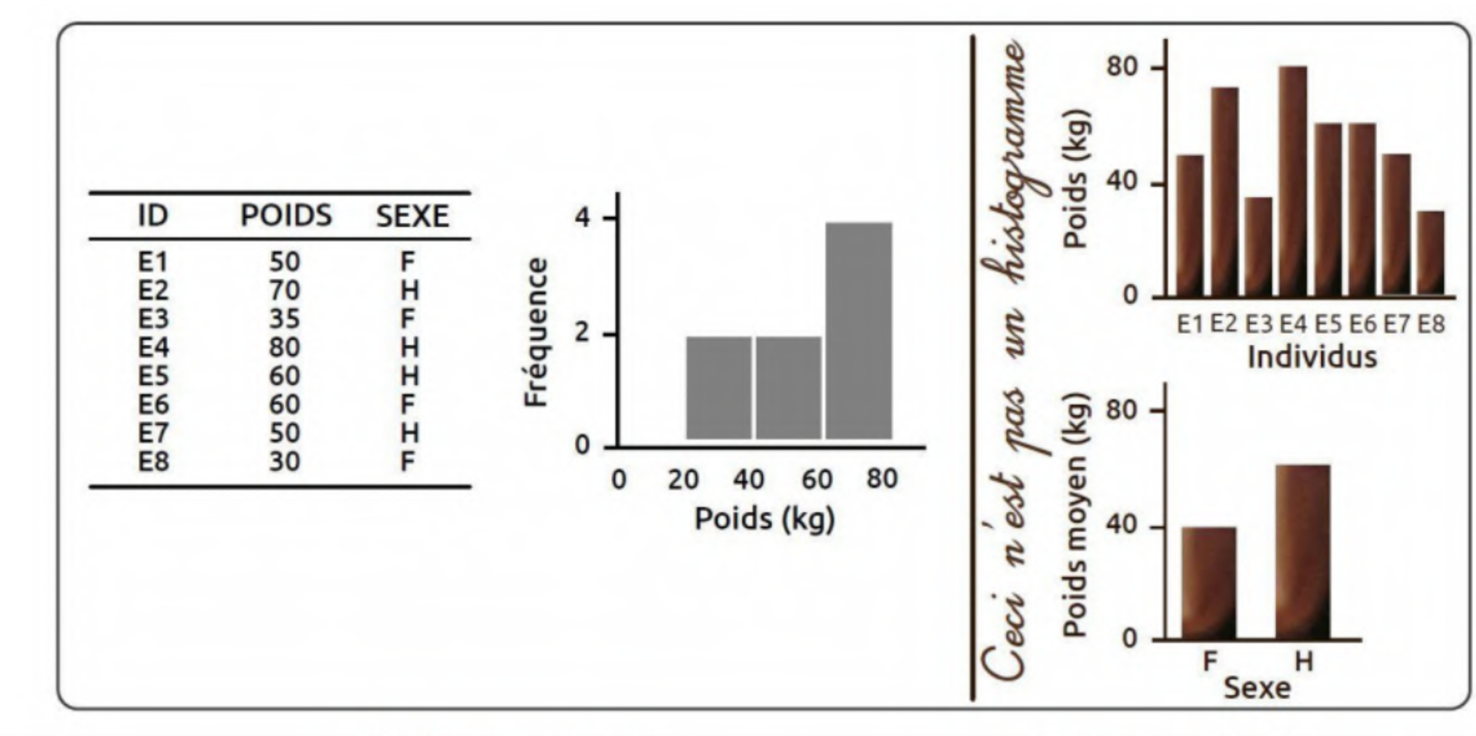
Effet du nombre de classes sur la représentation.



Les données sont les mêmes pour les deux séries, mais les **amplitudes** des classes des histogrammes sont différentes : de 0.5° en 0.5° à gauche, de 0.05° en 0.1° à droite.

Diagramme en bâton ? Histogramme ?

Attention à ne pas confondre histogramme et diagramme en bâton. Sous Excel, l'histogramme correspond en réalité au diagramme en bâton…



ID	POIDS	SEXE
E1	50	F
E2	70	H
E3	35	F
E4	80	H
E5	60	H
E6	60	F
E7	50	H
E8	30	F

Histogramme :
quantitatif x quantitatif

Diag. en bâton :
qualitatif x quantitatif

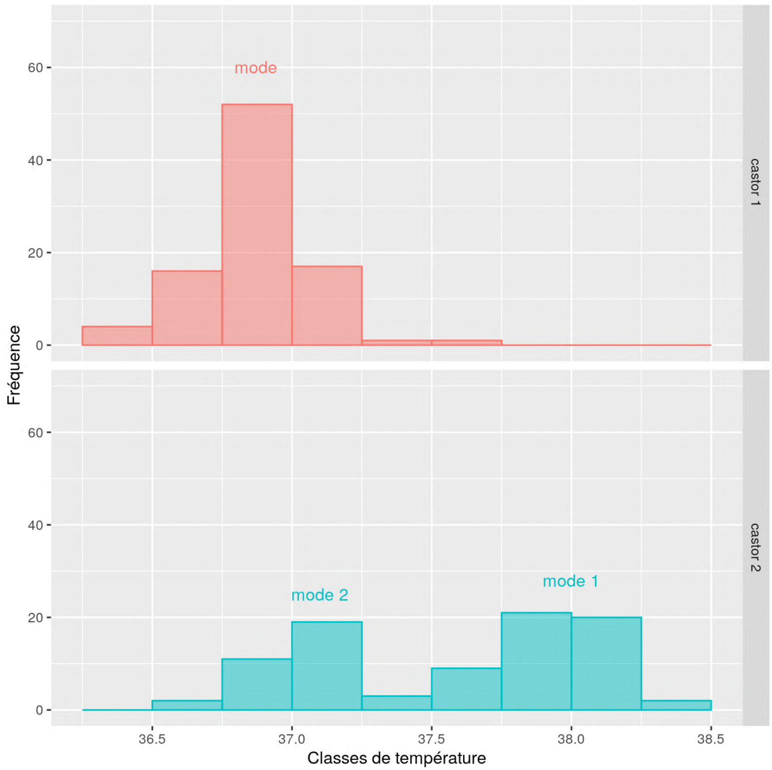
La forme de la distribution statistique

Les histogrammes permettent de décrire la **forme** d'une série statistique. On peut la caractériser par

- le mode
- la symétrie

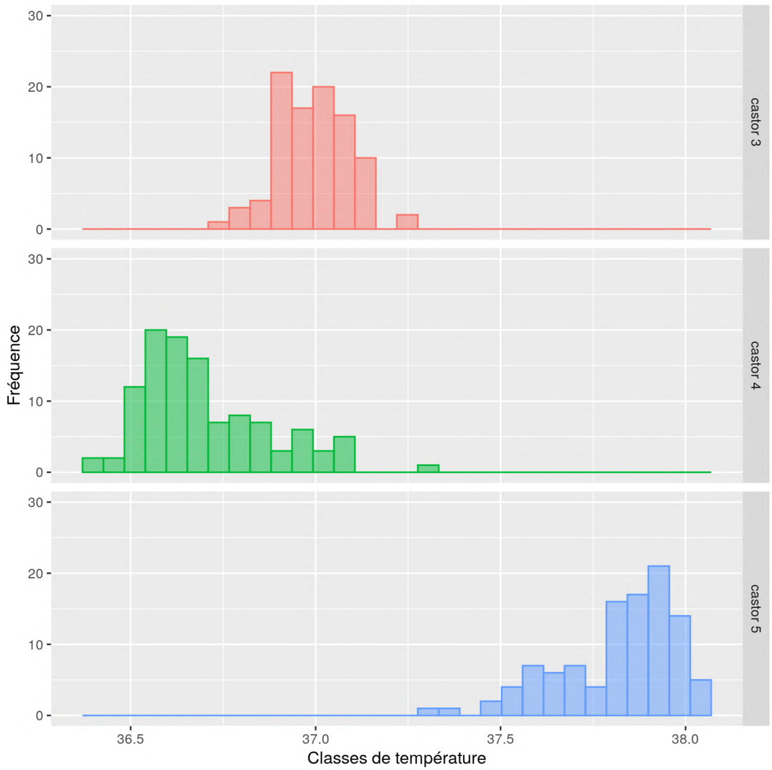
Le mode de la série

- La série est **unimodale** : les données sont concentrées autours d'un seul point.
- La série est **multimodale** : les données sont concentrées autours d'un mode principale et un ou plusieurs modes secondaires.



La symétrie de la série

- La série est symétrique : toutes les valeurs sont proches, il y a peu ou pas de valeurs extrêmes.
- Dissymétrique à gauche : beaucoup de valeurs faibles, peu de valeurs élevées.
- Dissymétrique à droite : beaucoup de valeurs élevées, peu de valeurs faibles.



Cette approche graphique permet déjà de poser des hypothèses sur nos données. Le castor 3 a une température normale (autours de 37°C) alors que le castor 4 semble en hypothermie, à l'inverse du castor 5 qui pourrait faire une réaction inflammatoire.

Les graphiques : premières notions clés

Pour caractériser une série, il est possible dans un premier temps de la représenter graphiquement :

- par un diagramme en bâton (qualitatif x quantitatif).
- Ou un histogramme (quantitatif x quantitatif).
- D'autres méthodes sont possibles (diagramme de distribution, courbe de fréquence cumulée), l'une d'elle sera abordée dans la partie suivante (boxplot).

Cette approche graphique permet de déterminer si la série possède des valeurs plutôt proches les unes des autres ou au contraire très hétérogènes.

Un graphique est simple à comprendre, mais son interprétation peut-être **subjective**. Il faut donc compléter cette approche par des données statistiques.

2- Analyse univariée : les valeurs centrales

Intérêt

Une série de données **doit être caractérisée** par des calculs **scientifiques**, c'est-à-dire une méthode :

- objective
- factuelle
- et reproductible.

Les principaux outils de l'analyse univariée peuvent être classer en 2 familles :

- Les valeurs centrales.
- Les paramètres de dispersions.

2.1 - Les valeurs centrales

Les valeurs centrales

Elles permettent de **résumer** une série en une seule valeur. Il existe 3 types de valeurs centrales :

- Moyenne
- Médiane
- Mode (classe modale)

Elles n'ont pas la même signification et doivent donc être toutes calculées lorsque l'on souhaite résumer une série statistique (quitte à conserver la plus adaptée à la fin).

A noter que la classe modale est la seule valeur centrale utilisable pour les données qualitatives, après dénombrement (cf. diagramme en baton)

Moyenne

La somme de l’ensemble des valeurs de la série statistique, divisée par le nombre d’éléments de la série (moyenne arithmétique)

- Fréquemment utilisée, la moyenne se base sur **l'ensemble des données** de la série.
- Formule Excel : =moyenne().
- Elle présente l'inconvénient d'être **influencée par les valeurs extrêmes**. Elle n'est donc pas toujours représentative de la série.

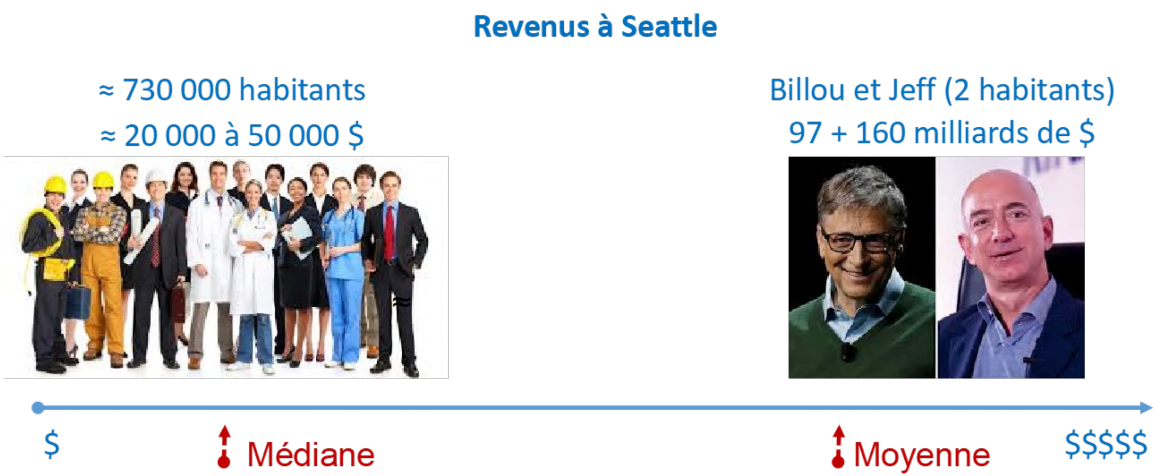
Il existe beaucoup de variations de la moyenne. Par exemple :

- la moyenne tronquée qui consiste à supprimer du calcul le minimum et le maximum afin d'éliminer l'influence des valeurs extrêmes. On peut aussi supprimer les 10% les plus faibles et fortes.
- La moyenne pondérée, en multipliant chaque valeur par un poids, puis diviser la somme par la somme des poids. Très utile si les individus statistiques sont très hétérogènes (population des départements français).

Médiane

La valeur qui partage une série ordonnée en **deux sous ensembles d'effectifs égaux** : 50 % des éléments ont des valeurs supérieures à la médiane et 50% ont des valeurs inférieures. Pour la série [1;2;3;4;5] médiane = 3).

- Si il y a un nombre paire de valeurs dans la série, la médiane est la moyenne des deux valeurs centrales (pour la série [1;2;3;4] médiane = 2,5).
- Contrairement à la moyenne, elle **n'est pas influencée par les valeurs extrêmes** puisqu'elle n'est dépendante que des valeurs centrales de la série ordonnée.
- On dit que la médiane est **robuste** aux valeurs extrêmes (outliers).
- Formule Excel : =mediane()

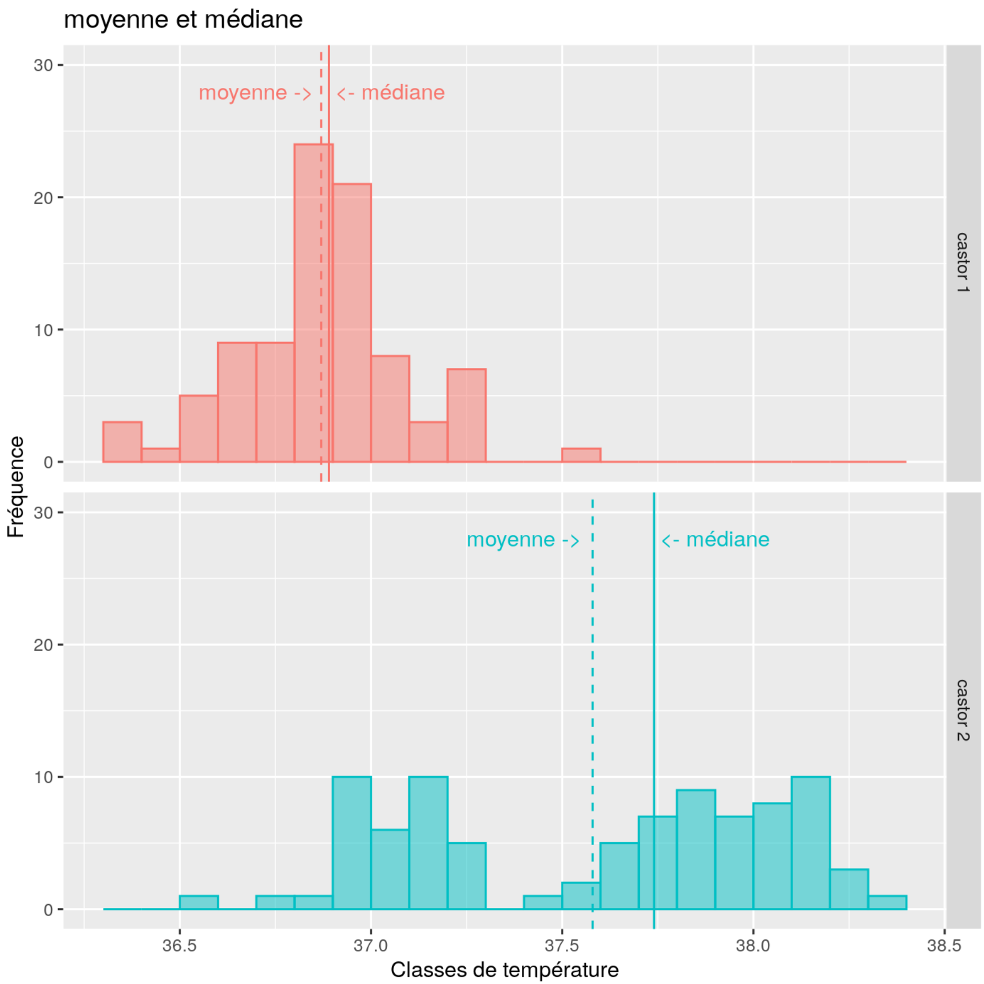


Mode

Mode ou classe modale / *Formule Excel : =mode()*

- Pour les variables qualitatives ou quantitatives discrètes.
- Valeur ou modalité la plus fréquente d'une série. Elle est issue d'un dénombrement.
- Pour les données continues : hasard si plusieurs valeurs sont identiques. Il faut donc discrétiser la série continue pour trouver le mode, qui sera influencé par la méthode utilisée.

Moyenne, mode et médiane : exemple



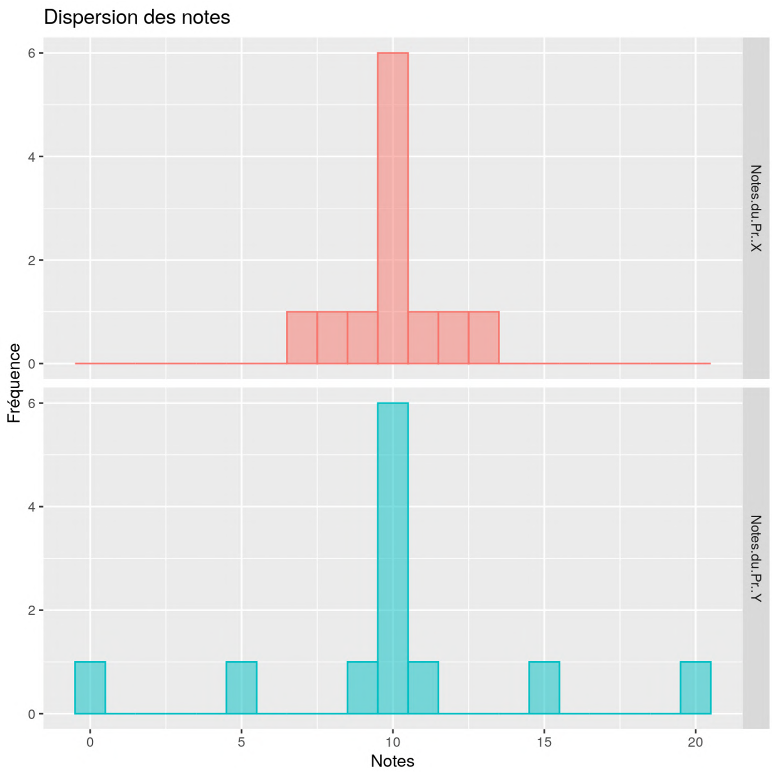
Moyenne, mode et médiane : limites

Le résumé d'une distribution que donne une valeur centrale ne nous renseigne pas sur la ***dispersion*** *des valeurs autour de cette valeur centrale, c'est-à-dire sur la tendance qu'elles-ont à se concentrer ou se disperser autour de celle-ci. (Claude Grasland)*

Etudiants	Notes.du.Pr..X	Notes.du.Pr..Y
A	7	0
B	8	5
C	9	9
D	10	10
E	10	10
F	10	10
G	11	11
H	12	15
I	13	20
Mode	**10**	**10**
Moyenne	**10**	**10**
Médiane	**10**	**10**

Moyenne, mode et médiane : limites 2

Les valeurs centrales sont identiques pour les notes des deux professeurs, mais leur **dispersion** est différente



2.2 - Les paramètres de dispersion

Les paramètres de dispersion

Contrairement aux valeurs centrales, les paramètres de dispersion permettent de mesurer la **la tendance qu'ont les valeurs de la distribution à se répartir autours d'une valeur ou bien les unes par rapport aux autres.**

La mesure de cette **variabilité** est au coeur de nombreuses analyses statistique :

- mesure de la variabilité
- distinguer un effet aléatoire d'une *vrai* variabilité

Il existe deux type de paramètres de dispersion :

- dispersion absolue : même unité que le caractère étudié.
- dispersion relative : sans unité mais facilement comparable avec d'autres séries d'unités différentes.

L'étendue

- La différence entre la plus grande valeur de la série et la plus petite.
- Elle dépend uniquement de deux valeurs.
- Elle est seulement intéressante lorsque les extrêmes ont une signification (exemple : amplitude thermique annuelle sur des moyennes mensuelles).

Ecart absolu moyen

La distance moyenne à la moyenne (moyenne arithmétique de la valeur absolue des écarts à la moyenne).

Etudiants i	**Xi**	**|Xi-Xmoy|**
A	7	3
B	8	2
C	9	1
D	10	0
E	10	0
F	10	0
G	11	1
H	12	2
I	13	3
Total	**90**	**12**
Moyenne	**10**	**12/9=1,33**

Les notes s’écartent de 1,33 par rapport à la moyenne : **relativement homogène**.
Si on fait le calcul pour les notes du professeur *y*, on trouvera 3.6 : **notes hétérogènes**.
Cet indicateur est facile à comprendre mais n’a pas les propriétés statistiques intéressante. On préfère donc utiliser l'écart-type.

Variance et écart-type

La variance n'est pas un paramètre de dispersion en elle-même, mais une mesure **GLOBALE** de la variation autour de la moyenne (quantité d'information).

- Comme l'écart absolu moyen c'est une moyenne des écarts à la moyenne
- mais calculé avec le théorême de Pythagore et non plus avec la valeur absolue

Pour obtenir un paramètre de dispersion absolue, on effectue la racine carrée de la variance, appelée **écart-type**.

Contrairement à l'écart absolu moyen, l'écart type a un signification **probabiliste** :

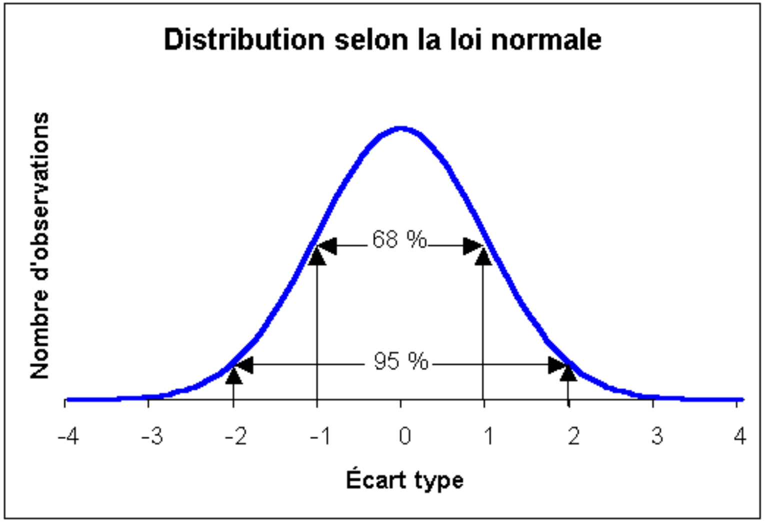
- La théorie des probabilités permet d'estimer la chance qu'a une valeur d'être éloignée de la moyenne de plus d'un certain nombre d'écart-types, si la distribution est gaussienne (normale).
- Formule Excel : =ecartype()
- N.B. Sur des échantillons de population, on utilisera l'estimation de l'écart-type : =ecartypep()

La loi normale

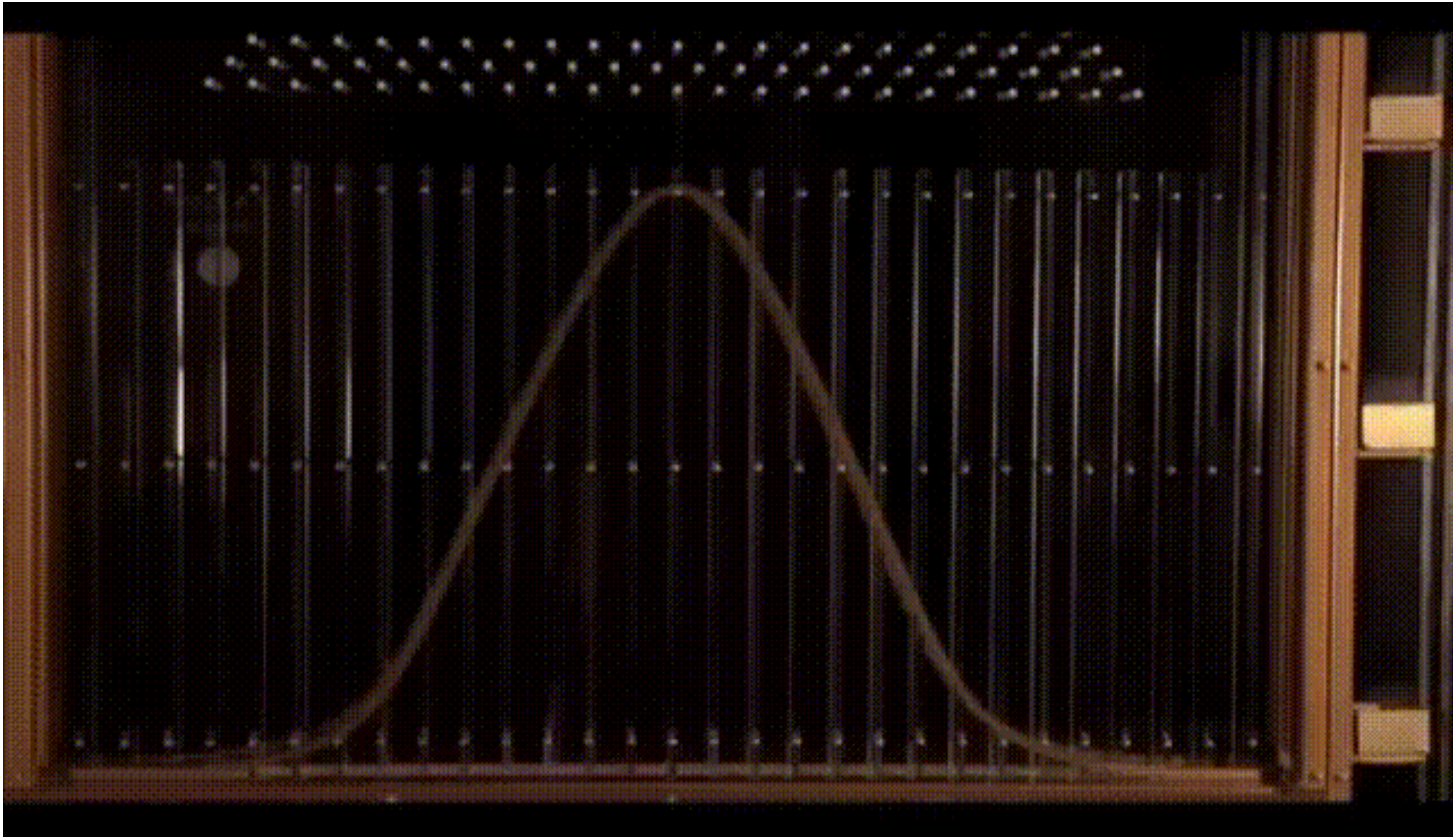
La **loi normale** est une loi de probabilité, très courante en statistique. Elle modélise très bien de nombreux phénomènes naturels aléatoires dont la probabilité d'apparition est proche d'une valeur centrale. Elle permet de vérifier si un phénomène est en dehors de la normale - anormale, un des **tests statistique** de base.

Lorsqu'une distribution est normale (on dit aussi gaussienne), les probabilités de trouver les valeurs a une distance donnée de la moyenne sont les suivantes :

- 68 % des valeurs sont comprises entre -1 écart type et +1 écart type
- 95 % des valeurs sont comprises entre -2 écart type et +2 écart type
- 99 % des valeurs sont comprises entre -3 écart type et +3 écart type



La loi normale : exemple



Ecart-type

Etudiants i	Xi	(Xi-Xmoy)²
A	7	9
B	8	4
C	9	1
D	10	0
E	10	0
F	10	0
G	11	1
H	12	4
I	13	9
Total	**90**	**28**
Moyenne	**10**	**28/9=3,11**

La variance est de 3.11, on en déduit la valeur de l'écart-type ($\sqrt{3.11} = 1.8$). Elle est considérée comme une mesure de l'ordre de grandeur de la dispersion des notes autour de la moyenne.

Si la distribution des notes du Pr X était gaussienne (ce qui est difficile à affirmer au vu de la faiblesse de l'échantillon) on devrait trouver :

- environ deux tiers des notes dans l'intervalle [8.2 ; 11.8] qui correspond à la moyenne plus ou moins un écart-type
- 95% des notes dans l'intervalle [6.4 ; 13.6] qui correspond à la moyenne plus ou moins deux écarts-type.

Le même calcul effectué sur les notes du Pr Y aboutit à un écart-type de **5.3** qui est beaucoup plus importante que celui des notes du professeur X. Les notes du Pr Y sont **beaucoup plus dispersées** que celles du Pr X. (Claude Grasland)

Ecart-type : illustration graphique

3 valeurs : 4 ; 5,5 ; 10

Individu	Valeur	Ecart à la moyenne	Ecart à la moyenne au carré
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25

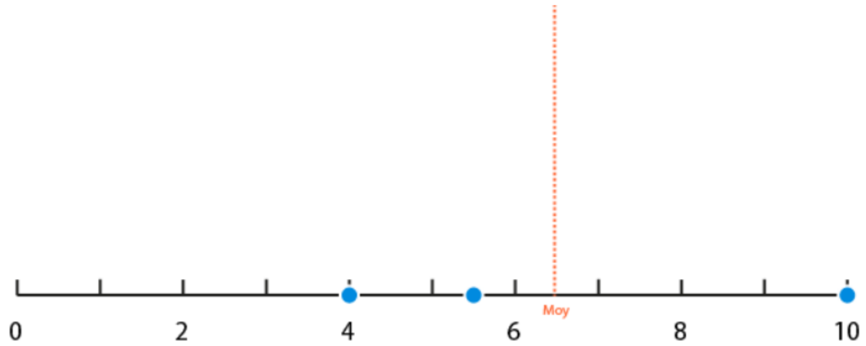
Somme	19,5	0	19,5
Moyenne	6,5	0	6,5

Ecart-type : illustration graphique

Moyenne

Individu	Valeur	Ecart à la moyenne	Ecart à la moyenne au carré
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25

Somme	19,5	0	19,5
Moyenne	6,5	0	6,5

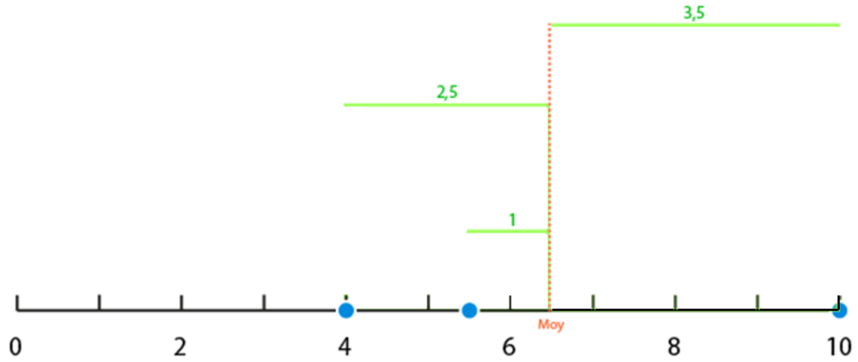


Ecart-type : illustration graphique

Ecarts à la moyenne

Individu	Valeur	Ecart à la moyenne	Ecart à la moyenne au carré
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25

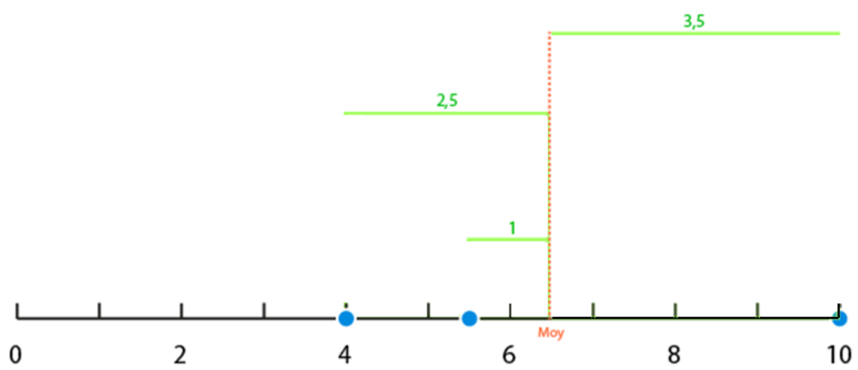
Somme	19,5	0	19,5
Moyenne	6,5	0	6,5



Ecart-type : illustration graphique

Ecart absolu moyen

Individu	Valeur	Ecart à la moyenne	Ecart à la moyenne au carré
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25
Somme	19,5	0	19,5
Moyenne	6,5	0	6,5



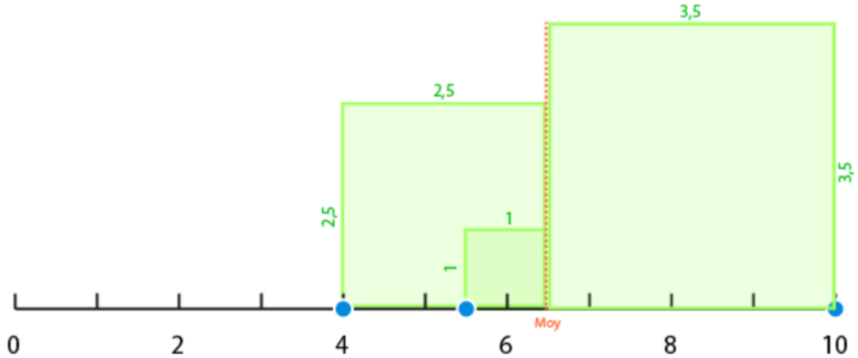
L'écart absolu moyen correspond à la moyenne des distances à la moyenne : (|2,5| + |1| + |-3,5|) / 3 = 2,33

Ecart-type : illustration graphique

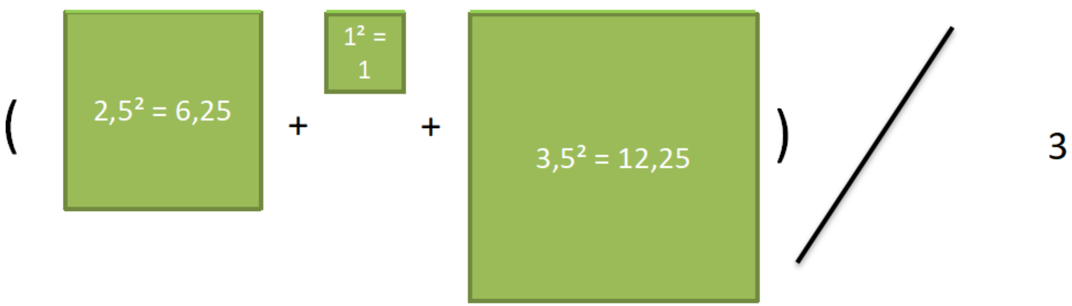
Variance

Individu	Valeur	Ecart à la moyenne (coté des carrés)	Ecart à la moyenne au carré (aire des carrés)
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25

Somme	19,5	0	19,5
Moyenne	6,5	0	6,5



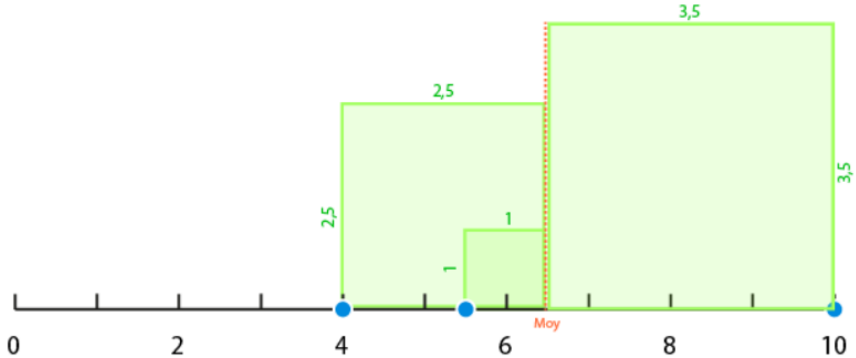
Pour la variance, on va élever au carré chaque distance à la moyenne, puis en faire la moyenne.
Le résultat va donner une aire moyenne



Ecart-type : illustration graphique

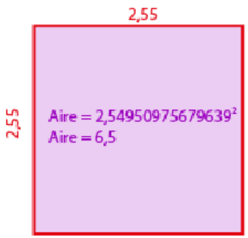
Variance

Individu	Valeur	Ecart à la moyenne (coté des carrés)	Ecart à la moyenne au carré (aire des carrés)
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25
Somme	19,5	0	19,5
Moyenne	6,5	0	6,5



Pour la variance, on va élever au carré chaque distance à la moyenne, puis en faire la moyenne. Le résultat va donner une aire moyenne

La variance est la moyenne des distances à la moyenne au carré , soit le carré dont l'aire est :
(6,25 + 1 +12,25)/3 = **6,5**

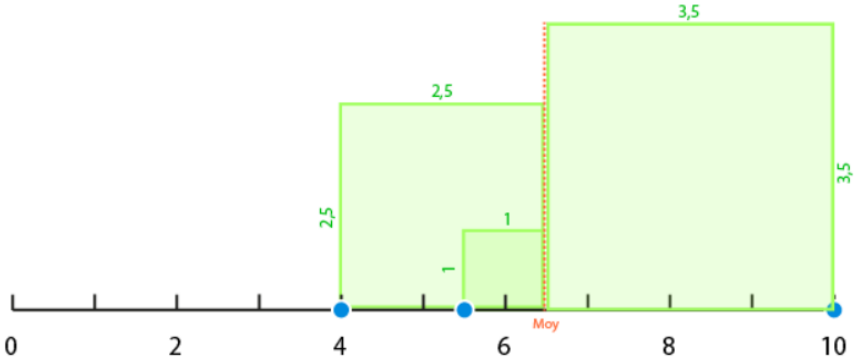


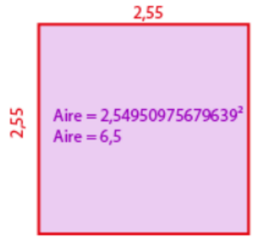
Ecart-type : illustration graphique

Ecart-type

Individu	Valeur	Ecart à la moyenne (coté des carrés)	Ecart à la moyenne au carré (aire des carrés)
X1	4	2,5	6,25
X2	5,5	1	1
X3	10	-3,5	12,25

Somme	19,5	0	19,5
Moyenne	6,5	0	6,5





La variance est la moyenne des distances à la moyenne au carré , soit le carré dont l'aire est :
(6,25 + 1 +12,25)/3 = **6,5**

L'écart-type est la racine carrée de la variance, soit le côté du carré précédent : $\sqrt{6,5} = 2,55$
A l'aide de la racine carrée, on repasse donc d'une aire à une distance de 2,55 :

Ecart-type : interprétation

La variance s'interprète donc comme une surface :

- plus elle est grande, plus il y a de dispersion loin de la moyenne.

Puisque l'écart-type est la racine carrée de la variance, on passe d'une interprétation de **surface** pour la variance (les carrés de l'exemple précédent) à une interprétation de **distance** pour l'écart-type (on passe de deux à une dimension).

Il s'interprète alors comme une moyenne des distances à la moyenne. **Ces « distances » sont mesurées à la « Pythagore »** : une racine carrée des différences au carré .

L'écart-type a l'avantage d'être exprimé dans la même unité de mesure que la variable de départ.

Les paramètres de dispersion relatifs

Les paramètres de dispersion absolus ne permettent pas de comparer des séries statistiques de natures différentes. Dans ce cas il faut utiliser des paramètres de dispersion relatifs :

- Pas d'effet de taille.
- Indépendamment de l'unité de mesure.

Le plus utilisé est le coefficient de variation (CV). Il peut être exprimé en %. Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande.

- CV = Ecart-type / moyenne

2.3 - Autres outils de l'analyse univariée

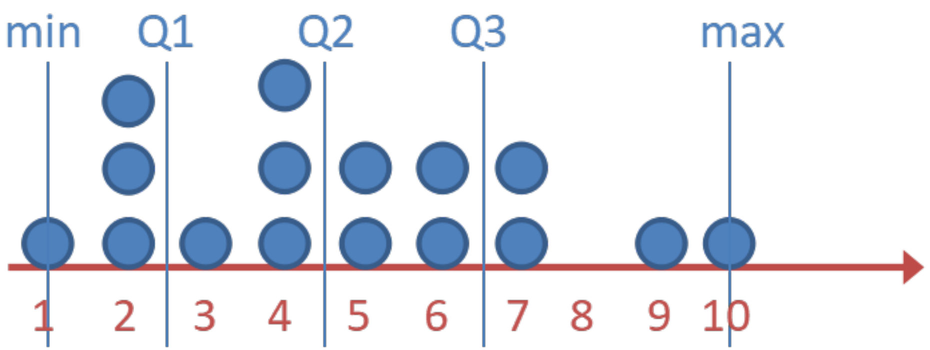
Médiane et quantile

la médiane est souvent associée dans les analyses univariées aux **quantiles** :

- Les valeurs qui définissent les bornes d'une partition en **classes d'effectifs égaux**.
- Il y aura le **même nombre d'éléments** dans chaque classe.
- Mais cette approche ne permet pas de voir les outliers si le nombre de classes est faible.

Pour des quartiles (4 classes), on note Min, Q1, Q2, Q3, Max.

Entre Min et Q1, on a 25% de la fréquence des effectifs de la série, les 25% suivants entre Q1 et Q2, 25% suivants entre Q2 et Q3, même chose entre Q3 et Max. A noter que Q2 est ici la médiane.



Min	1
Q1	2,5
Q2	4,5
Q3	6,5
Max	10

L’intervalle interquartile

Comme pour la moyenne avec l'écart-type, il existe un paramètre de dispersion associé à la médiane : l'intervalle interquartile.

- Basé sur les quantiles
- Pour avoir un paramètre de dispersion, on calcule l'intervalle interquartile : on conserve 50 % des valeurs les plus proches de la médiane, c’est-à-dire la différence entre Q3 et Q1.
- L’intervalle interdéciles est également utilisé (on conserve 80 % des valeurs).
- Fréquemment utilisé pour les distributions non symétrique (et non normale)

Pour obtenir un paramètre de dispersion relatif, on peut calculer le coefficient interquartile (CI) :

- CI = (Q3-Q1)/Q2

Les boxplots

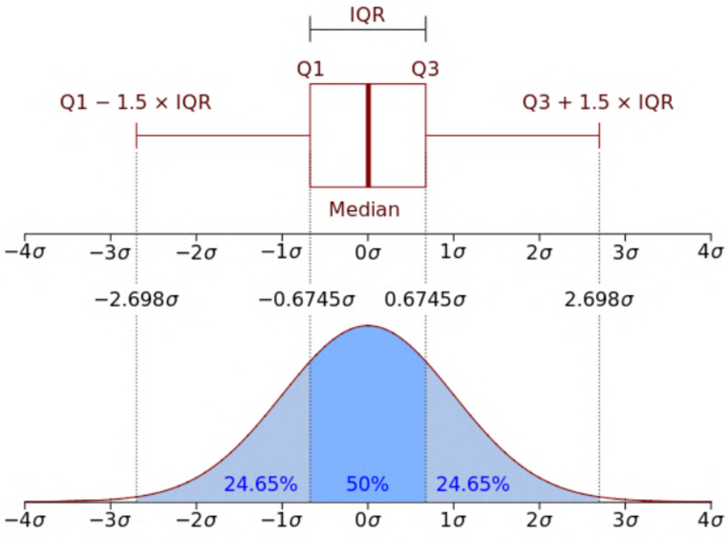
L'utilisation des quantiles permet une autre représentation graphique d'une série quantitative : **la boite à moustache (boxplot)**.

Elle consiste à mettre sur un graphique les valeurs des quantiles, le plus souvent des quartiles.

Par défaut, de nombreux boxplots n'affichent pas le min et le max mais 1,5x l'intervalle interquartile (IQR,soit les valeurs comprises entre Q1 et Q3 = 50% des valeurs autours de la médiane). Cette méthode permet de couvrir 99,3% de la série.

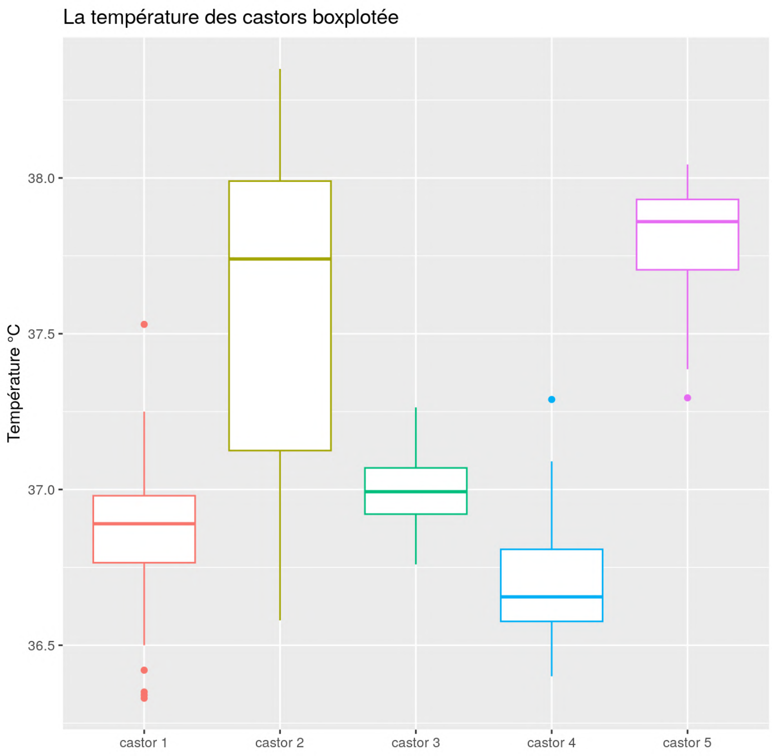
Les valeurs restantes sont appelées **outliers** et correspondent aux valeurs extrêmes min et max.

Box plot basé sur l'intervalle interquartile (IQR).
On voit bien que 99,3% de la série est couverte.



24,65% + 50% + 24,65% = 99,3%

Les boxplots : exemple

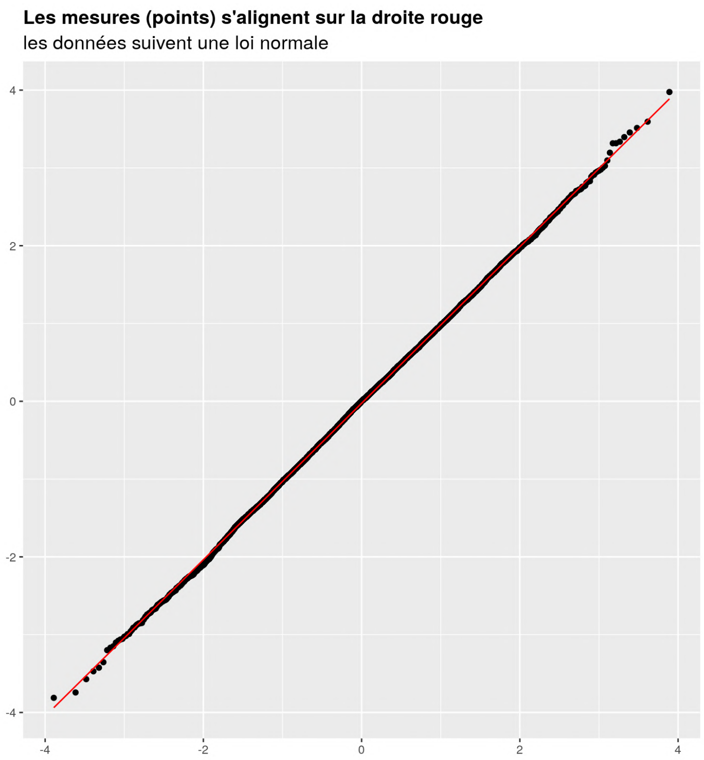


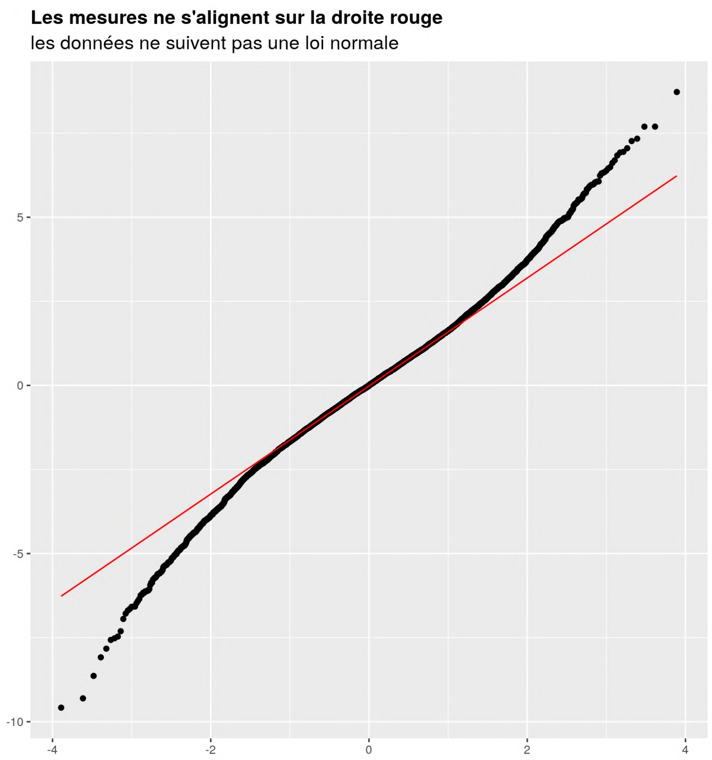
La comparaison des distributions et de leur forme est très facile avec le box plot (ici en 1,5x l'intervalle interquartile, on le reconnait aux outliers -valeurs hors norme- en forme de points).

L'avantage des box plots est qu'ils sont construits de manière **objective** (quartile). C'est-à-dire que la seule intervention du concepteur sur le graphique est le choix de cette méthode de représentation...

Le Quantile-Quantile Plot : qq-plot

Certains logiciels de statistiques vous permettent de créer un QQPLOT, idéal pour montrer la normalité d'une série :



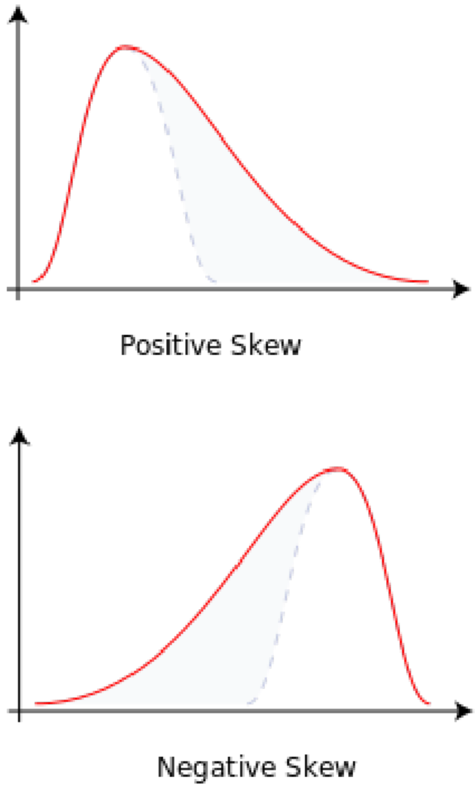


Mesure de la symétrie

La plupart des outils statistiques, dont Excel, donnent un indicateur appelé coefficient d'asymétrie (Skew). Il s'agit d'une mesure standardisée de l'étalement de la queue de distribution.

- S'il est égal ou très proche de 0 : symétrie
- Plus il s'éloigne de 0 positivement, plus la série est étirée à droite
- Plus il s'éloigne de 0 négativement plus la série est étirée à gauche

Il est de nouveau préférable de confirmer cet indicateur graphiquement.



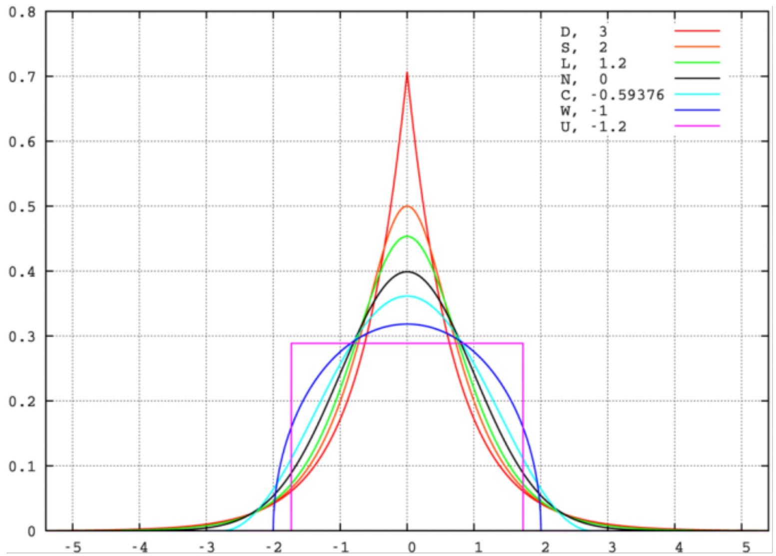
Source : Wikipédia

Mesure de l'aplatissement

On associe souvent au coefficient d'asymétrie un second indicateur : **le kurtosis** qui permet de mesurer l'aplatissement d'une distribution.

L'interprétation combinée des deux indicateurs permet de justifier factuellement la loi de probabilité d'une série. Par exemple pour justifier qu'une série s'approche d'une loi normale.

Attention, le kurtosis d'une loi normale est normalement égale à 3, mais pour des raisons pratiques on soustrait 3 au résultats du calcul (kurtosis normalisé). Excel et la plupart des outils de statistique fournissent un kurtosis normalisé.



Loi de probabilité	Kurtosis normalisé	Symbole
Loi de Laplace	3	D
Loi sécante hyperbolique	2	S
Loi logistique	1,2	L
Loi normale	0	N
Loi du cosinus surélevé	-0,593762...	C
Loi du demi-cercle	-1	W
Loi uniforme continue	-1,2	U

Source : Wikipédia

Tests

Il existe également des tests statistiques pour déterminer si une série suit une loi normale.

Pour la loi normale :

- le test de Shapiro-Wilk. Si la p-value n'est pas significative (rejet de H0, généralement à 5%), vous données suivent une loi normale
- le test de Jarque-Bera (utilisé dans Arcgis)

Conclusions

Concepts clés

L'analyse univariée d'une série statistique est essentielle pour comprendre le comportement des données.

Cette analyse se fait à l'aide de deux approches complémentaires :

- la représentation graphique, subjective mais qui donne des informations rapidement.
- Le calcul mathématique, objectif, qui fourni un résumé de la distribution (valeur centrale), des informations sur la dispersion des données (paramètres de dispersion) et sur la forme de la série (asymétrie, kurtosis).

Pour les graphiques :

- La représentation des données qualitatives se fait avec un diagramme en baton. Il faut souvent *dénombrer* les données à l'aide d'une tableau de contingence.
- La représentation des données quantitatives se fait avec un histogramme ou un boxplot.

Concepts clés

Pour résumer une série, on utilise les valeurs centrales:

- La moyenne est souvent rencontrée, mais est sensible aux outliers contrairement à la médiane.
- La médiane peut se combiner avec les quantiles, qui décrivent de manière objective la série. Ils peuvent néanmoins *cacher* les outliers.

Résumer une série avec une seule valeur centrale n'est pas suffisant. Il faut également mesurer sa variabilité à l'aide d'un paramètre de dispersion :

- à l'aide de l'écart-type, qui avec la moyenne possède une signification probabiliste pour les distributions qui suivent une loi normale.
- A l'aide de l'interval interquartile.
- Ces paramètres de dispersion sont absolus : ils s'expriment dans la même unité que la série. Des paramètres de dispersion relatifs permettent de supprimer l'unité et donc de comparer des variabilités de différentes séries (coefficient de variation, coefficient interquartile)

Concepts clés

Il faut toujours avoir conscience que vos données reflètent une réalité :

- directement observable (un nombre d'habitants)
- ou calculée (une densité de population, soit le nombre d'habitants en supprimant l'effet de taille du maillage observé)

Attention, faire un graphique et donner le résultat d’un calcul n’est pas une plus-value : un logiciel de statistique peu le faire automatiquement.

Ce sont vos capacités à les interpréter dans le contexte de votre étude qui priment !