

Statistique descriptive

Chapitre 6 : Les quantiles

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Généraliser la médiane	2
L'idée	2
Le vocabulaire	2
Le calcul	2
Sur données brutes : la méthode	2
Un exemple à la main	3
Les quantiles de nos 500 salaires	3
Sur données groupées	3
La formule, dessinée	4
Groupé contre exact	4
La boîte à moustaches	5
La construction	5
Les éléments	5
Les valeurs atypiques des salaires	5
Ce qu'une boîte montre d'un coup d'œil	6
Deux boîtes côte à côte	6
Les mesures dérivées	6
Les indicateurs fondés sur les quantiles	6
Le rapport interdécile	7
Des mesures robustes	7
Le partage des rôles	8
S'entraîner	8
Exercice 1 : quantiles à la main	8
Exercice 1 : solution	8
Exercice 2 : des seuils pour une politique salariale	8
Exercice 2 : solution	9
Ce qu'il faut retenir	9
Le résumé du chapitre	9

Généraliser la médiane

L'idée

Pourquoi s'arrêter à deux parts

La médiane coupe la population en deux moitiés. Rien n'oblige à s'arrêter là. Couper en quatre donne les **quartiles**, en dix les **déciles**, en cent les **centiles**. On obtient une description bien plus fine que trois indicateurs de tendance centrale.

Quantile d'ordre p

Le **quantile d'ordre** p , noté q_p , est la valeur telle qu'une proportion p des individus lui est inférieure :

$$F(q_p) = p$$

C'est la lecture de la courbe cumulative **à l'envers** : on entre par les fréquences, on sort par la variable.

Le vocabulaire

Les trois familles

Nom	Nombre de parts	Notation
Quartiles	4	$Q_1, Q_2 = Me, Q_3$
Déciles	10	$D_1, \dots, D_9$
Centiles	100	$C_1, \dots, C_{99}$

Un piège de vocabulaire

Pour couper en quatre, il faut **trois** traits : il y a trois quartiles, neuf déciles, quatre-vingt-dix-neuf centiles. Et $Q_2 = D_5 = C_{50} = Me$: quatre noms de la même valeur.

Le calcul

Sur données brutes : la méthode

Le rang du quantile

On range les n valeurs par ordre croissant, puis on cherche celle de **rang** $r = 1 + p(n - 1)$. Si r n'est pas entier, on interpole entre les deux valeurs qui l'encadrent : avec $r = k + g$, où k est entier et $0 < g < 1$,

$$q_p = x_{(k)} + g \times (x_{(k+1)} - x_{(k)})$$

Une convention parmi d'autres

Pour $p = 0,5$, on retrouve le rang $(n + 1)/2$ de la médiane du chapitre 5. C'est la convention de R et de Python, ce qui permettra de retrouver tous les nombres aux chapitres 14 et 15. D'autres conventions existent : sur 500 valeurs, elles diffèrent de quelques dirhams.

Un exemple à la main

Les douze salaires de la PME de l'introduction

3 200 ; 3 500 ; 3 800 ; 4 100 ; 4 300 ; 4 600 ; 5 000 ; 5 400 ; 6 200 ; 7 500 ; 9 800 ; 15 600.

Quantile	Rang $1 + p \times 11$	Calcul	Valeur
Q_1	3,75	$3\,800 + 0,75 \times 300$	4 025
Me	6,5	$4\,600 + 0,5 \times 400$	4 800
Q_3	9,25	$6\,200 + 0,25 \times 1\,300$	6 525

La lecture

La moitié centrale des salariés gagne entre 4 025 et 6 525 dirhams : un intervalle de 2 500 dirhams, alors que l'étendue totale en couvre 12 400. Le salaire de 15 600 n'y est pour rien.

Les quantiles de nos 500 salaires

Cinq nombres pour toute la distribution

Quantile	Valeur (dirhams)	Lecture
D_1	4 348	10 % gagnent moins
Q_1	5 318	25 % gagnent moins
Me	6 770	la moitié gagne moins
Q_3	8 330	75 % gagnent moins
D_9	10 473	90 % gagnent moins

Une première lecture de la forme

De Q_1 à la médiane : 1 452 dirhams. De la médiane à Q_3 : 1 560. De Q_3 à D_9 : 2 143. Les écarts **s'allongent vers le haut** — c'est encore l'étalement à droite.

Sur données groupées

La formule générale

$$q_p = e_{i-1} + a_i \times \frac{p - F_{i-1}}{F_i - F_{i-1}}$$

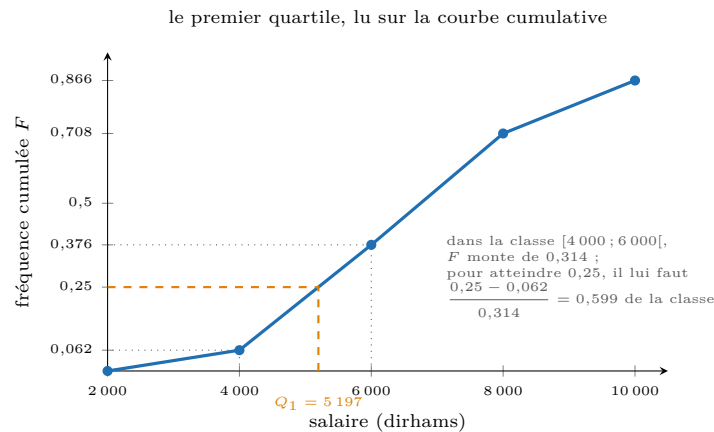
où $[e_{i-1}; e_i[$ est la classe dans laquelle F franchit p . C'est la formule de la médiane du chapitre 5, avec p à la place de 0,5.

Le premier quartile, en détail

$p = 0,25$ est franchi dans $[4\,000; 6\,000[$, où F passe de 0,062 à 0,376 :

$$Q_1 = 4\,000 + 2\,000 \times \frac{0,250 - 0,062}{0,376 - 0,062} = 4\,000 + 2\,000 \times 0,599 = 5\,197$$

La formule, dessinée



Une règle de trois sur un segment

Entre deux bornes, la courbe cumulative est un segment : le quantile se trouve à la même **fraction** de la classe que p de la montée de F .

Grouper contre exact

Comparer les deux calculs

	Groupé	Exact	Écart
D_1	4 242	4 348	-2,4 %
Q_1	5 197	5 318	-2,3 %
Me	6 747	6 770	-0,3 %
Q_3	8 532	8 330	+2,4 %
D_9	11 193	10 473	+6,9 %

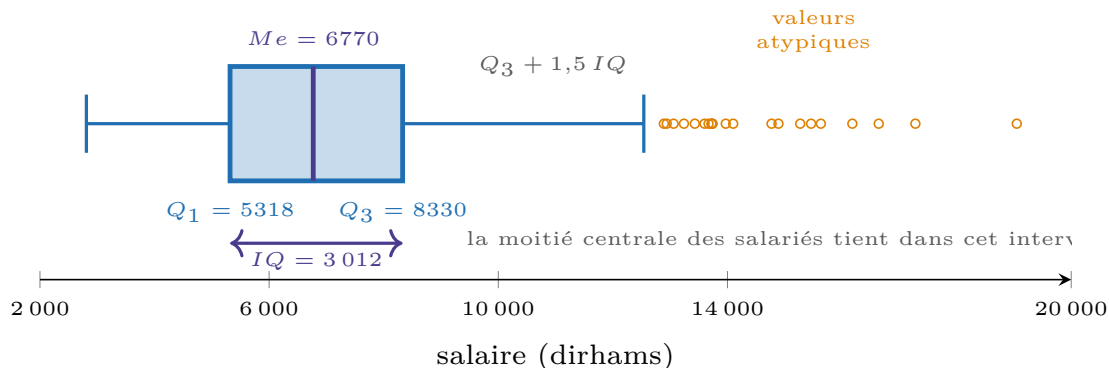
La limite du travail sur données groupées

L'approximation est excellente au centre et se dégrade vers les extrémités — là où les classes sont larges et le remplissage inégal. D_9 tombe dans $[10\,000 ; 14\,000[$, dont les salariés se pressent près de la borne basse.

La boîte à moustaches

La construction

la boîte à moustaches : cinq nombres résument la distribution



Les éléments

Comment on la lit

- La **boîte** va de Q_1 à Q_3 : elle contient la moitié centrale des individus.
- Le trait intérieur est la **médiane**.
- Les **moustaches** s'étendent jusqu'aux valeurs les plus extrêmes situées à moins de $1,5 IQ$ des quartiles.
- Les points au-delà sont marqués individuellement : ce sont les **valeurs atypiques**.

La règle des 1,5 IQ

$$\text{seuil bas} = Q_1 - 1,5 IQ$$

$$\text{seuil haut} = Q_3 + 1,5 IQ$$

Les valeurs atypiques des salaires

Le calcul des seuils

$IQ = 8330 - 5318 = 3012$, donc $1,5 IQ = 4518$.

Seuil haut : $8330 + 4518 = 12848$. Seuil bas : $5318 - 4518 = 800$, en dessous du minimum observé.

21 salariés dépassent le seuil haut, soit 4,2% de l'effectif. Aucun en dessous.

Atypique ne veut pas dire faux

Ces 21 salariés ne sont pas des erreurs de saisie : ce sont des cadres expérimentés, parfaitement réels. La règle des $1,5 IQ$ est une **convention de repérage**, pas un test.

Que les valeurs atypiques soient **toutes du même côté** est en soi une information : c'est la signature de l'asymétrie à droite.

Ce qu'une boîte montre d'un coup d'œil

Quatre informations

1. Le **niveau** — la position de la médiane ;
2. la **dispersion** — la largeur de la boîte ;
3. l'**asymétrie** — la médiane décentrée dans la boîte, une moustache plus longue ;
4. les **valeurs atypiques**, individuellement.

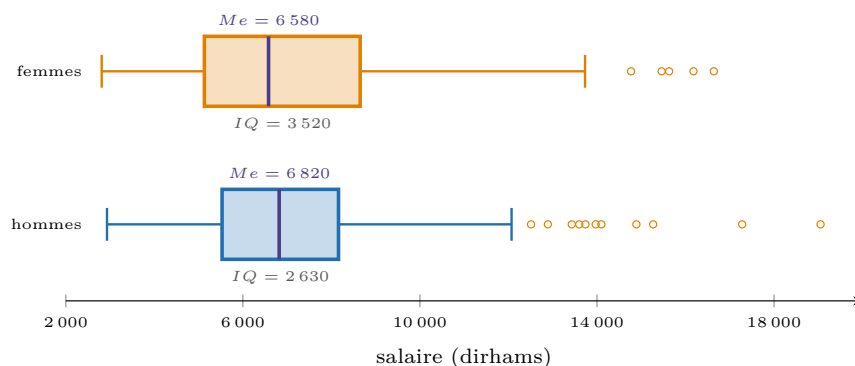
Quatre informations dans un dessin de deux centimètres de haut.

Son usage professionnel : comparer

Sa véritable force est la **comparaison**. Placer côte à côte les boîtes de plusieurs groupes montre instantanément lesquels diffèrent en niveau, en dispersion, en asymétrie. Le chapitre 11 le fera pour le diplôme.

Deux boîtes côte à côte

deux boîtes côte à côte : le salaire des femmes et des hommes



Même moyenne, deux distributions

Moyennes égales (7 158 dirhams), mais la boîte des femmes est **plus large** — IQ de 3 520 contre 2 630 — et sa médiane plus basse. La boîte étroite des hommes fait apparaître plus de valeurs atypiques : **atypique se juge par rapport à sa propre boîte**.

Les mesures dérivées

Les indicateurs fondés sur les quantiles

Trois mesures

Mesure	Formule	Ce qu'elle dit
Intervalle interquartile	$IQ = Q_3 - Q_1$	dispersion de la moitié centrale
Écart interquartile relatif	IQ/Me	dispersion relative , sans unité

Rapport interdécile	D_9/D_1	inégalité entre les extrêmes
---------------------	-----------	------------------------------

Sur nos données

$$IQ = 3\,012 \quad \frac{IQ}{Me} = \frac{3\,012}{6\,770} = 0,445 \quad \frac{D_9}{D_1} = \frac{10\,473}{4\,348} = 2,41$$

Le rapport interdécile

L'indicateur d'inégalité le plus lu

« Les 10 % les mieux payés gagnent au moins 2,41 fois plus que les 10 % les moins payés. » Cette phrase est immédiatement compréhensible par un non-spécialiste : c'est pour cela que ce rapport figure dans les bilans sociaux d'entreprise et dans les publications sur les inégalités. À titre de repère, il est de l'ordre de 3 pour les salaires en France et dépasse 5 dans les pays très inégalitaires : notre entreprise est plutôt homogène.

Par sexe

	D_1	D_9	D_9/D_1
Femmes	4 074	10 790	2,65
Hommes	4 542	10 276	2,26

Les salaires des femmes sont **plus inégaux** entre eux que ceux des hommes.

Des mesures robustes

Leur avantage décisif

Toutes ces mesures ne dépendent que des **rangs**.
Ajouter un dirigeant à 200 000 dirhams ne déplace Q_1 , Q_3 et D_9 que de quelques dirhams. L'écart-type, lui, serait plus que triplé : de 2 598 à 8 990 dirhams.
C'est pourquoi les organismes statistiques privilégient les quantiles pour décrire les revenus.

Leurs limites

1. Elles **ignorent** ce qui se passe à l'intérieur de chaque part : deux distributions très différentes peuvent avoir les mêmes quantiles.
2. Elles ne se prêtent pas au **calcul algébrique** : l'écart-type intervient dans des formules — covariance, corrélation, intervalles de confiance —, pas l'intervalle interquartile.

Le partage des rôles

Deux familles complémentaires

Quantiles pour décrire et publier ; moyenne et écart-type pour calculer et modéliser.

Un bon tableau de résultats donne les deux : niveau et dispersion par la médiane et l'intervalle interquartile, puis par la moyenne et l'écart-type. Leur désaccord est lui-même une information sur la forme.

S'entraîner

Exercice 1 : quantiles à la main

Les notes de onze étudiants

Notes sur 20, rangées : 4 ; 6 ; 7 ; 8 ; 9 ; 10 ; 11 ; 12 ; 12 ; 14 ; 17.

1. Calculer D_1 , Q_1 , Me , Q_3 et D_9 avec la règle du rang $1 + p(n - 1)$.
2. Calculer l'intervalle interquartile et les seuils de la règle des $1,5 IQ$. Y a-t-il des notes atypiques ?
3. Un douzième étudiant obtient 20. Recalculer la médiane et la moyenne. Laquelle a le plus bougé ?

Exercice 1 : solution

Les rangs et les valeurs

1. Avec $n = 11$, le rang vaut $1 + 10p$:

	D_1	Q_1	Me	Q_3	D_9
Rang	2	3,5	6	8,5	10
Valeur	6	$(7 + 8)/2 = 7,5$	10	$(12 + 12)/2 = 12$	14

Atypiques et robustesse

2. $IQ = 12 - 7,5 = 4,5$; seuils $7,5 - 6,75 = 0,75$ et $12 + 6,75 = 18,75$: **aucune** note atypique.
3. Avec $n = 12$: la médiane devient $(10 + 11)/2 = 10,5$; la moyenne passe de $110/11 = 10$ à $130/12 = 10,83$. La moyenne a bougé presque deux fois plus que la médiane.

Exercice 2 : des seuils pour une politique salariale

Avec le tableau en six classes du fil rouge

Classe	$[2 ; 4[$	$[4 ; 6[$	$[6 ; 8[$	$[8 ; 10[$	$[10 ; 14[$	$[14 ; 20[$
F_i	0,062	0,376	0,708	0,866	0,980	1

Salaires en milliers de dirhams. La direction veut :

1. une prime pour les 20 % les moins payés : sous quel salaire la verser ?
2. une enquête auprès des 5 % les mieux payés : au-dessus de quel salaire ?
3. Combien de salariés se trouvent entre ces deux seuils ?

Exercice 2 : solution

Deux quantiles par interpolation

1. D_2 : 0,20 est franchi dans $[4\,000; 6\,000[$:

$$D_2 = 4\,000 + 2\,000 \times \frac{0,20 - 0,062}{0,314} = 4\,000 + 879 = 4\,879 \text{ dirhams}$$

2. C_{95} : 0,95 est franchi dans $[10\,000; 14\,000[$:

$$C_{95} = 10\,000 + 4\,000 \times \frac{0,95 - 0,866}{0,114} = 10\,000 + 2\,947 = 12\,947 \text{ dirhams}$$

3. Par construction, $95\% - 20\% = 75\%$ des salariés, soit 375.

Et sur les données exactes ?

$D_2 = 5\,010$ et $C_{95} = 12\,295$: comme au tableau « groupé contre exact », l'écart grandit dans la classe large du haut.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. Un **quantile** q_p est la valeur telle qu'une proportion p lui est inférieure : $F(q_p) = p$.
2. Trois quartiles, neuf déciles, quatre-vingt-dix-neuf centiles — et $Q_2 = D_5 = C_{50} = Me$.
3. Données brutes : valeur de rang $1 + p(n - 1)$, en interpolant.
4. Données groupées : une seule formule d'**interpolation linéaire**, celle de la médiane avec p .
5. La **boîte à moustaches** montre niveau, dispersion, asymétrie et valeurs atypiques — et sert à **comparer**.
6. Règle des $1,5\,IQ$: une **convention de repérage**, pas un test.
7. IQ , IQ/Me et D_9/D_1 sont **robustes** ; ils décrivent bien mais ne se calculent pas.

La suite

IQ mesure la dispersion à partir des rangs. Une autre famille repose sur les **écarts à la moyenne** — et elle donnera l'écart-type, l'indicateur le plus utilisé de toute la statistique. C'est le chapitre 7.