

Statistique descriptive

Chapitre 12 : Distributions conjointe, marginales, conditionnelles



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Trois familles de distributions

Les distributions conditionnelles

Le théorème de la moyenne totale

La décomposition de la variance

Les moyennes et variances marginales

Le tableau de synthèse

Ce qu'il faut retenir

Trois familles de distributions

Les trois familles		
Distribution	Ce qu'elle décrit	Combien
conjointe	le couple $(x ; y)$	une seule
marginale	une variable, l'autre oubliée	deux
conditionnelle	une variable, l'autre fixée	autant que de modalités

Où elles se lisent dans le tableau

- La **conjointe** est l'intérieur du tableau : les n_{ij} .
- Les **marginales** sont la dernière colonne et la dernière ligne.
- Les **conditionnelles** sont les lignes ou les colonnes prises **une par une**, chacune ramenée à son propre total.

Trois questions distinctes

- Conjointe : « Quelle part des salariés est Bac+5 et gagne plus de 10 000 dirhams ? »
- Marginale : « Quel est le salaire moyen de l'entreprise ? »
- Conditionnelle : « Quel est le salaire moyen **des Bac+5** ? »

La troisième est celle qui intéresse presque toujours le décideur. C'est elle qui permet de comparer des groupes.

Les distributions conditionnelles

Distribution conditionnelle

La distribution de y sachant $x = x_i$ est la répartition de y restreinte aux individus pour lesquels x vaut x_i .

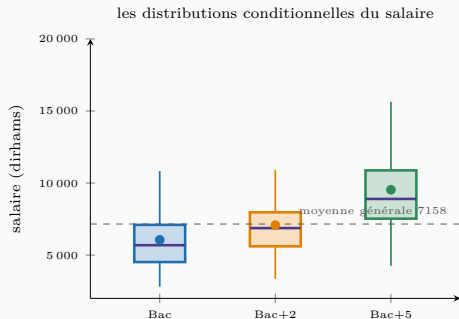
Ses fréquences sont les profils-lignes du chapitre 11 :

$$f_{j|i} = \frac{n_{ij}}{n_{i\bullet}} \quad \text{avec} \quad \sum_j f_{j|i} = 1$$

Le geste, en une phrase

Conditionner, c'est réduire la population, puis recalculer.

On ne travaille plus sur 500 salariés mais sur les 107 Bac+5. Tous les indicateurs des chapitres 6 à 10 s'appliquent alors à cette sous-population.



Les trois moyennes conditionnelles

$$\bar{y}_{\text{Bac}} = 6\,059 \quad \bar{y}_{\text{Bac}+2} = 7\,081 \quad \bar{y}_{\text{Bac}+5} = 9\,534$$

Leur moyenne, *pondérée par les effectifs*, redonne exactement la moyenne générale 7 158 : c'est le **théorème de la moyenne totale**.

Et la variance se décompose

$$\underbrace{6\,750\,122}_{\text{totale}} = \underbrace{5\,010\,426}_{\text{intra, résiduelle}} + \underbrace{1\,739\,696}_{\text{inter, expliquée}}$$

$\eta^2 = 1\,739\,696 / 6\,750\,122 = 0,258$: le diplôme explique 25,8 % de la variance des salaires.

Le salaire selon le diplôme				
Diplôme	n_j	Moyenne	Écart-type	Médiane
Bac	219	6 059	2 021	5 690
Bac+2	174	7 081	2 045	6 875
Bac+5	107	9 534	2 872	8 900
Ensemble	500	7 158	2 598	6 770

La lecture, en trois constats

1. **Le niveau monte avec le diplôme** : +17 % du Bac au Bac+2, +35 % du Bac+2 au Bac+5. La prime au diplôme n'est pas linéaire, elle s'accélère.
2. **La dispersion aussi** : l'écart-type des Bac+5 dépasse celui des deux autres groupes. Les carrières de cadres sont plus différenciées.
3. **La moyenne d'ensemble n'est celle d'aucun groupe** : 7 158 dirhams tombe entre le Bac+2 et le Bac+5, et ne décrit personne.

Ce troisième point justifie à lui seul le conditionnement.

Le théorème de la moyenne totale

Théorème de la moyenne totale

$$\bar{y} = \frac{1}{n} \sum_j n_j \bar{y}_j = \sum_j f_j \bar{y}_j$$

La moyenne générale est la moyenne des moyennes conditionnelles, pondérée par les effectifs.

Vérification

$$\frac{219 \times 6\,059 + 174 \times 7\,081 + 107 \times 9\,534}{500} = \frac{3\,579\,000}{500} \simeq 7\,158 \quad \checkmark$$

Moyenne **simple** des trois : $(6\,059 + 7\,081 + 9\,534)/3 = 7\,558$. Faux de 400 dirhams.

L'effet de structure

Ce théorème explique un phénomène qui déroute les non-statisticiens : **la moyenne générale peut évoluer alors qu'aucune moyenne conditionnelle ne bouge.**

Supposons que l'entreprise recrute 50 Bac+5 de plus. Aucun salaire individuel ne change, aucune moyenne par diplôme ne change — et pourtant la moyenne générale augmente, parce que les **poids** f_j ont changé.

C'est l'**effet de structure**, et il est au cœur de toute analyse de l'évolution des salaires, des prix ou de la productivité. Une hausse du salaire moyen ne signifie pas que quelqu'un a été augmenté.

La décomposition de la variance

Décomposition de la variance

$$\underbrace{V(y)}_{\text{totale}} = \underbrace{\frac{1}{n} \sum_j n_j V_j(y)}_{\text{variance intra-groupe}} + \underbrace{\frac{1}{n} \sum_j n_j (\bar{y}_j - \bar{y})^2}_{\text{variance inter-groupes}}$$

Ce que chaque terme représente

- La variance **intra** est la moyenne des variances internes aux groupes : c'est ce que le diplôme **n'explique pas**. On l'appelle aussi variance **résiduelle**.
- La variance **inter** mesure l'écartement des moyennes de groupe : c'est ce que le diplôme **explique**. On l'appelle variance **expliquée**.

Autrement dit : la dispersion totale se partage entre **ce qui sépare les groupes** et **ce qui subsiste à l'intérieur** de chacun.

Le calcul sur nos données

$$\underbrace{6\,750\,122}_{\text{totale}} = \underbrace{5\,010\,426}_{\text{intra}} + \underbrace{1\,739\,696}_{\text{inter}}$$

Rapport de corrélation

$$\eta^2 = \frac{\text{variance inter}}{\text{variance totale}} \quad 0 \leq \eta^2 \leq 1$$

C'est la **part de la variance expliquée** par la variable qualitative.

Sur nos données

$$\eta^2 = \frac{1\,739\,696}{6\,750\,122} = 0,258$$

Le diplôme explique 25,8 % de la variance des salaires.

Comment on commente ce nombre

Un quart de la dispersion salariale s'explique par le seul diplôme. **Les trois quarts restants tiennent à autre chose** : l'ancienneté, le poste, la région, la performance individuelle — et une part irréductible d'arbitraire.

C'est un résultat mesuré, et il vaut mieux qu'une opinion. Il permet de répondre à « le diplôme détermine-t-il le salaire ? » par « il pèse pour un quart », ce qui est autrement plus utile qu'un oui ou un non.

Les deux cas extrêmes

- $\eta^2 = 0$: toutes les moyennes conditionnelles sont égales. La variable qualitative **n'apporte rien**.
- $\eta^2 = 1$: aucune variance à l'intérieur des groupes. Connaître le groupe donne **exactement** la valeur.

(suite)

Une remarque importante

η^2 ne suppose **aucune forme** de relation : il fonctionne avec une variable explicative purement nominale, sans ordre.

C'est ce qui le distingue du coefficient de corrélation du chapitre 14, qui n'existe qu'entre deux variables quantitatives et ne capte que la liaison **linéaire**.

Les moyennes et variances marginales

Distribution marginale

Elle s'obtient en **oubliant** l'autre variable : ce sont les distributions étudiées dans toute la partie III.

$$f_{i\bullet} = \sum_j f_{ij} \qquad f_{\bullet j} = \sum_i f_{ij}$$

Le lien entre les trois familles

$$f_{ij} = f_{i\bullet} \times f_{j|i}$$

Conjointe = **marginale** $\times$ **conditionnelle**. Cette relation reviendra en probabilités sous la forme $P(A \cap B) = P(A) P(B|A)$: c'est le même énoncé, écrit dans un autre langage.

Le cas de l'indépendance

x et y sont indépendantes si et seulement si

$$f_{ij} = f_{i\bullet} \times f_{\bullet j} \quad \text{pour toutes les cases}$$

Toutes les conditionnelles sont alors égales à la marginale : conditionner n'apporte **rien**.

Le tableau de synthèse

Les trois familles, récapitulées

Ce qu'il faut savoir de chacune			
	Conjointe	Marginale	Conditionnelle
Où	intérieur du tableau	les marges	une ligne, ramenée à son total
Fréquences	f_{ij}	$f_{i\bullet}, f_{\bullet j}$	$f_{j i}$
Somme	1 sur tout le tableau	1	1 par ligne
Combien	une	deux	k
Ce qu'elle donne	la liaison	le profil d'ensemble	la comparaison de groupes

Les deux théorèmes

$$\bar{y} = \sum_j f_j \bar{y}_j \qquad V(y) = \underbrace{\overline{V_j(y)}}_{\text{intra}} + \underbrace{V(\bar{y}_j)}_{\text{inter}}$$

La moyenne des moyennes redonne la moyenne. La moyenne des variances ne redonne pas la variance — il manque la variance des moyennes.

La phrase à retenir

C'est la formulation la plus compacte du chapitre, et elle vaut d'être apprise telle quelle.

Ce qu'il faut retenir

Six points

1. Trois familles : **conjointe, marginales, conditionnelles**.
2. Conditionner, c'est **réduire la population puis recalculer**.
3. **Moyenne totale** : moyenne des moyennes conditionnelles, pondérée par les effectifs. D'où l'**effet de structure**.
4. **Décomposition de la variance** : totale = intra + inter.
5. η^2 : part de variance expliquée par la variable qualitative. Ici 25,8 % pour le diplôme.
6. $f_{ij} = f_{i\bullet} f_{\bullet j}$ caractérise l'**indépendance**.

La suite

Quand les **deux** variables sont quantitatives, on peut faire mieux qu'une comparaison de groupes : mesurer directement dans quel sens et avec quelle force elles varient ensemble.

Cet indicateur s'appelle la **covariance**, et sa formule se déduit directement de celle de la variance. C'est le chapitre 13.