Chapitre 7

Relations entre deux variables quantitatives

Pour les relations entre deux variables quantitatives, la notion de probabilités conditionnelles intervient. L'objectif de ce chapitre est d'expliquer la manière de les obtenir et de les étudier. Soient X et Y les deux caractères quantitatifs étudiés, p le nombre de modalités prises par X, q le nombre de modalités prises par Y, et n le nombre total d'observations.

7.1 Couple de variables aléatoires discrètes

Soient X et Y des variables aléatoires définies sur un même univers S, et ayant pour espaces images respectifs :

$$X(S) = \{x_1, x_2, \ldots, x_n\} \tag{7.1}$$

et

$$Y(S) = \{y_1, y_2, \ldots, y_n\} \tag{7.2}$$

On transforme l'ensemble produit :

$$X(S) \times Y(S) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{7.3}$$

en un espace probabilisé, en définissant la probabilité du couple ordonné (x_i, y_i) par $\Pr(X = x_i, Y = y_i)$, que l'on écrit $k(x_i, y_i)$. Cette fonction h sur $X(S) \times Y(S)$, définie par $h(x_i, y_i) = \Pr(X = x_i, Y = y_i)$ est appelée la **distribution jointe**, où la loi de probabilité produit de X et Y, et est habituellement représentée sous la forme d'un tableau (Tab. 7.1).

	y_1	y_2	...	y_n	**SOMME**
x_1	$h(x_1, y_1)$	$h(x_1, y_2)$	...	$h(x_1, y_n)$	$f(x_1)$
x_2	$h(x_2, y_1)$	$h(x_2, y_2)$	...	$h(x_2, y_n)$	$f(x_2)$
...	...	...	...	...	•••
x_n	$h(x_n, y_1)$	$h(x_n, y_2)$	...	$h(x_n, y_n)$	$f(x_n)$
SOMME	$g(y_1)$	$g(y_2)$	•••	$g(y_n)$	

TABLE 7.1 – Loi de probabilité produit entre deux variables aléatoires X et Y

Les fonctions f et g sont définies par :

$$f(x_i) = \sum_{j=1}^{q} = h(x_i, y_j) \tag{7.4}$$

et

$$g(y_j) = \sum_{i=1}^{p} = h(x_i, y_j) \tag{7.5}$$

La loi du couple (X, Y) est également appelée **loi de probabilité simultanée** (ou loi conjointe). Elle est définie par l'ensemble des nombres $n_{ij} = h(x_i, y_j)$ avec $(0 < n_{ij} < 1)$ tels que :

$$n_{ij} = \Pr(X = x_j \text{ et } Y = y_j) \tag{7.6}$$

Les lois de probabilités $n_{ij} = h(x_i, y_i)$ vérifient les relations :

1. $h(x_i, y_i) \geq 0$;
2. $\sum_{i=1}^{p} \sum_{j=1}^{q} h(x_i, y_i) = 1$

Dans le cas fini, la loi conjointe est affichable sous la forme d'un **tableau de contingence**. Les probabilités n_{ij} y figurant définissent la loi du couple et toutes les lois associées.

Dans le tableau 7.1, $f(x_i)$ est la somme des éléments de la i-ième ligne, tandis que $g(y_j)$ est la somme des éléments de la j-ième colonne. Ces fonctions portent le nom de **lois de probabilités marginales**. Ce sont les distribution individuelles respectives de X et Y.

Les **lois de probabilité marginales** sont les lois de probabilité des variables X et Y prises séparément. Par définition, il en existe deux :

1. la loi de probabilité marginale de la variable X : $\Pr(X = x_i) = \sum_{j=1}^{q} n_{ij} = p_{i.}$;
2. la loi de probabilité marginale de la variable Y : $\Pr(Y = y_j) = \sum_{i=1}^{p} n_{ij} = p_{.j}$.

Les quantités $p_{i.}$ et $p_{.j}$ constituent les marges du tableau de contingence et vérifient les relations :

$$\sum_{i=1}^{p} n_{i.} = \sum_{j=1}^{q} n_{.j} = 1 \tag{7.7}$$

Les **lois conditionnelles** sont les deux familles de lois suivantes :

1. la loi conditionnelle de X sachant $Y = y_j$ c'est-à-dire que la valeur de la variable Y est connue :

$$\Pr\left(X = x_i \setminus Y = y_j\right) = \frac{n_{ij}}{n_{.j}} = \frac{\Pr\left(X = x_i \text{ et } Y = y_j\right)}{\Pr\left(Y = y_j\right)} \tag{7.8}$$

2. la loi conditionnelle de Y sachant $X = x_i$ c'est-à-dire que la valeur de la variable X est connue :

$$\Pr\left(Y = y_j \setminus X = x_i\right) = \frac{n_{ij}}{n_{i.}} = \frac{\Pr\left(X = x_i \text{ et } Y = y_j\right)}{\Pr\left(X = x_i\right)} \tag{7.9}$$

Remarque 1. Ces lois sont parfaitement définies si les quantités $\Pr\left(Y = y_j\right)$ ou $\Pr\left(X = x_i\right)$ sont différentes de 0.

Remarque 2. Si on connaît les lois conditionnelles, on peut inversement en déduire la loi du couple.

Remarque 3. Grâce à la formule de Bayes, on peut exprimer une loi conditionnelle en fonction de l'autre. Par exemple :

$$\Pr\left(X = x_i \setminus Y = y_j\right) = \frac{\Pr\left(X = x_i \text{ et } Y = y_j\right)\Pr\left(X = x_i\right)}{\sum_{i=1}^{p}\Pr\left(Y = y_j \setminus X = x_i\right)\Pr\left(X = x_i\right)} \tag{7.10}$$

7.1.1 Propriétés des couples de variables aléatoires discrètes

Si X et Y sont des variables aléatoires définies sur le même univers S, $X+Y$, $X+k$, kX et XY avec $k \in \mathbb{R}$, sont des fonctions sur S définies par :

— $(X+Y)(s) = X(s) + Y(s)$;

— $(kX)(s) = kX(s)$;

— $(X+k)(s) = X(s) + k$;

— $(XY)(s) = X(s)\,Y(s)$;

avec $\forall s \in S$.

N.B. Toutes ces fonctions sont également des variables aléatoires.

7.1.2 Tableaux statistiques

Dans le cas de deux variables quantitatives discrètes, les tableaux statistiques portent le nom de **tableaux croisés** ou **tableaux de contingence** (Tab. 7.2). Dans chaque case du tableau, on inscrit l'effectif n_{ij} de l'échantillon, c'est-à-dire le nombre de données tel que $X = x_i$ et $Y = y_i$.

On définit les fréquences absolues suivantes.

1. L'**effectif** et la **fréquence marginale** n_i est le nombre d'individus possédant la modalité i du caractère X quelle que soit la distribution du caractère Y.

$$n_{i.} = \sum_{j=1}^{q} n_{ij} \tag{7.11}$$

et

$$f_{i.} = \frac{n_{i.}}{n} \tag{7.12}$$

De même, n_j est le nombre d'individus possédant la modalité j du caractère Y quelle que soit la distribution du caractère X.

$$n_{.j} = \sum_{i=1}^{p} n_{ij} \tag{7.13}$$

et

$$f_{.j} = \frac{n_{.j}}{n} \tag{7.14}$$

2. L'**effectif** et la **fréquence conditionnelle** $n_{j\backslash i}$ est la distribution de la variable Y lorsque l'on a fixé la modalité i pour la variable X. Elle est définit par :

$$n_{j\backslash i} = \frac{n_{ij}}{n_{i.}} \tag{7.15}$$

On définit de la même façon la fréquence conditionnelle $n_{i\backslash j}$ par :

$$n_{i\backslash j} = \frac{n_{ij}}{n_{.j}} \tag{7.16}$$

On définit les **fréquences relatives** f_{ij}, $f_{i.}$ et $f_{.j}$ par la division des effectifs n_{ij} et les **fréquences marginales** $n_{i.}$ et $n_{.j}$ par l'effectif total n.

Le tableau de contingence permet de vérifier si les deux variables sont bien dépendantes l'une de l'autre. Toutefois, pour représenter graphiquement deux variables quantitatives, on utilise un nuage de points dans $\mathbb{R}^2$.

XY	x_1	...	x_i	...	x_p	**Effectif marginal**
y_1	n_{11}	...	n_{j1}	...	n_{p1}	$n_{.1}$
...	...	...	...	...	...	...
y_j	n_{1j}	...	n_{ij}	...	n_{pj}	$n_{.j}$
...	...	...	...	...	...	...
y_q	n_{1q}	...	n_{iq}	...	n_{pq}	$n_{.q}$
Effectif marginal	$n_{1.}$	...	$n_{i.}$	...	$n_{p.}$	n

TABLE 7.2 – Tableau de contingence entre la variable aléatoire X et la variable aléatoire Y

7.1.3 Les caractéristiques marginales

Moyennes marginales

La moyenne arithmétique, c''est-à-dire l'espérance mathématique, de X vaut :

$$\mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{p} n_{i.}x_i \tag{7.17}$$

La moyenne arithmétique, c''est-à-dire l'espérance mathématique, de Y vaut :

$$\mathbb{E}(Y) = \frac{1}{n}\sum_{j=1}^{q} n_{.j}y_j \tag{7.18}$$

Le point G de coordonnées $(\mathbb{E}(X), \mathbb{E}(Y))$ est appelé **point moyen**.

Variances marginales

La variance de X vaut :

$$\mathbb{V}(X) = \frac{1}{n}\sum_{i=1}^{p} n_{i.}\left(x_i - \mathbb{E}(X)\right)^2 = \left[\frac{1}{n}\sum_{i=1}^{p} n_{i.}{x_i}^2\right] - \mathbb{E}(X) \tag{7.19}$$

La variance de Y vaut :

$$\mathbb{V}(Y) = \frac{1}{n}\sum_{j=1}^{q} n_{.j}\left(y_j - \mathbb{E}(Y)\right)^2 = \left[\frac{1}{n}\sum_{j=1}^{q} n_{.j}{y_j}^2\right] - \mathbb{E}(Y) \tag{7.20}$$

7.1.4 Les caractéristiques conditionnelles

Moyennes conditionnelles

$$\mathbb{E}(X_j) = \frac{1}{n_{.j}} \sum_{i=1}^{p} n_{ij} x_i \tag{7.21}$$

et

$$\mathbb{E}(Y_i) = \frac{1}{n_{i.}} \sum_{j=1}^{q} n_{ij} y_j \tag{7.22}$$

Variances conditionnelles

$$\mathbb{V}(X_j) = \frac{1}{n_{.j}} \sum_{i=1}^{p} n_{ij} (x_i - \mathbb{E}(X_j))^2 \tag{7.23}$$

et

$$\mathbb{V}(Y_i) = \frac{1}{n_{i.}} \sum_{j=1}^{q} n_{ij} (y_j - \mathbb{E}(Y_i))^2 \tag{7.24}$$

7.1.5 Les relations entre les caractéristiques marginales et les caractéristiques conditionnelles

Moyenne

La moyenne marginale est la moyenne pondérée des moyennes conditionnelles :

$$\mathbb{E}(X) = \frac{1}{n} \sum j = 1 q n_{.j} \mathbb{E}(X_j) \tag{7.25}$$

et

$$\mathbb{E}(Y) = \frac{1}{n} \sum i = 1 p n_{i.} \mathbb{E}(Y_i) \tag{7.26}$$

C'est le **théorème de la moyenne conditionnée**.

Variance

La variance marginale est la somme de la moyenne pondérée des variances conditionnelles et de la variance pondérée des moyennes conditionnelles.

$$\mathbb{V}(X) = \mathbb{V}(\bar{X}) + \mathbb{V}\left(\bar{X}_j\right) \tag{7.27}$$

$$\mathbb{V}(X) = \frac{1}{n}\sum_{j=1}^{q} n_{.j}\mathbb{V}(X_j) + \frac{1}{n}\sum_{j=1}^{q} n_{.j}\left(\bar{X}_j - \bar{X}\right)^2 \tag{7.28}$$

et

$$\mathbb{V}(Y) = \mathbb{V}(\bar{Y}) + \mathbb{V}\left(\bar{Y}_i\right) \tag{7.29}$$

C'est le **théorème de la variance conditionnée**.

La variance traduit la dispersion de la distribution. Dans le cas de la distribution marginale de X ou de Y, la dispersion résulte de deux facteurs :

1. la dispersion des distribution conditionnées autour de leurs moyennes $\mathbb{V}(\bar{X})$ ou $\mathbb{V}(\bar{Y})$ que l'on appelle **variance intra-population** (ou variance résiduelle) que l'on note souvent $s_w^2(X)$ ou $s_w^2(Y)$, w signifiant *within*.
2. la dispersion des moyennes conditionnelles autour de la moyenne $\mathbb{V}\left(\bar{Y}_i\right)$ ou $\mathbb{V}\left(\bar{X}_j\right)$ que l'on appelle **variance inter-population** (ou variance expliquée) que l'on note $s_b^2(X)$ ou $s_b^2(Y)$, b signifiant *between*.

7.1.6 Covariance

Lors du chapitre exposant les paramètres statistiques, la notion de **covariance** avait été posée sans être définie clairement. Cette partie propose de préciser la signification de cette notion importante.

La covariance entre deux variables aléatoires mesure la façon dont deux variables aléatoires X et Y varient **simultanément**. Si elle est nulle, cela signifie que les deux variables ne sont pas corrélées, c'est-à-dire qu'elles sont indépendantes.

La covariance correspond au paramètre suivant :

$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum_{\substack{1\leq i\leq p\\1\leq j\leq q}}^{p\ q} n_{ij}\left(x_i - \mathbb{E}(X)\right)\left(y_j - \mathbb{E}(Y)\right) = \left[\frac{1}{n}\sum_{\substack{1\leq i\leq p\\1\leq j\leq q}}^{p\ q} n_{ij}x_iy_j\right] - \mathbb{E}(X)\,\mathbb{E}(Y) \tag{7.30}$$

ou

$$\operatorname{cov}(X,Y) = \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)\left(Y - \mathbb{E}(Y)\right)\right] = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) \tag{7.31}$$

Propriété 1. $\operatorname{cov}(aX+b, cY+d) = ac\operatorname{cov}(X,Y)$

Propriété 2. $\operatorname{cov}(X,X) = \mathbb{V}(X)$

Propriété 3. $|\operatorname{cov}(X,Y)| = \sigma(X)\,\sigma(Y)$

Propriété 4. Si les variables X et Y sont indépendantes, alors $\operatorname{cov}(X,Y) = 0$. Par contre, la réciproque est fausse.

La covariance est une notion fondamentale qui permet l'étude d'une corrélation à n variables. C'est grâce à elle que l'on sait immédiatement si les variables étudiées sont corrélées, ou pas. Cet aspect fera l'objet d'un chapitre spécifique.

7.1.7 Corrélation

On définit la **corrélation** de X et Y, que l'on écrit $\rho(X, Y)$, par :

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{7.32}$$

ρ est une quantité sans dimension. ρ a les propriété suivantes :

1. $\rho(X, Y) = \rho(Y, X)$;
2. $\rho \in [-1, +1]$;
3. $\rho(X, X) = 1$ et $\rho(X, -X) = -1$;
4. $\rho(aX + b, cY + d) = \rho(X, Y)$ si $a \neq 0$ et $c \neq 0$.

Covariance et corrélation sont des mesures de la relation qui existe entre X et Y.

N.B. La notion de loi de probabilité produit se généralise à un nombre fini quelconque de variables aléatoires X, Y, ..., Z. h est alors une fonction sur l'ensemble produit $X(S) \times Y(S) \times \ldots \times Z(S)$, définie par :

$$h(x_i, y_j, \ldots, z_k) = \Pr(X = x_i, Y = y_j, \ldots, Z = z_k) \tag{7.33}$$

Cela sera développé dans le cadre des analyses multivariées.

7.1.8 Mesure de la dépendance

À l'aide de différents coefficients, l'étude de la distribution simultanée de deux variables permet de préciser le **type de liaison** pouvant exister entre ces deux variables, la nature et l'intensité de cette liaison.

On dit qu'un nombre fini des variables aléatoires X, Y, ..., Z sur un univers S sont **indépendantes** si :

$$\Pr(X = x_i, Y = y_j, \ldots, Z = z_k) = \Pr(X = x_i) \Pr(Y = y_j) \ldots \Pr(Z = z_k) \tag{7.34}$$

pour toutes les valeurs x_i, y_j, ..., z_k. En particulier, X et Y sont indépendants si :

$$\Pr(X = x_i, Y = y_j) = \Pr(X = x_i) \Pr(Y = y_j) \tag{7.35}$$

Le cadre multivarié sera utilisé plus tard. En attendant, voici quelques particuliers du cadre bivarié.

Si X et Y ont des distributions respectives f et g, et une loi de probabilité produit h, l'équation peut s'écrire :

$$h(x_i, y_j) = f(x_i)\, g(y_j) \tag{7.36}$$

Dit autrement, X et Y sont indépendants si chaque élément $h(x_i, y_j)$ est le produit de ses éléments marginaux.

Théorème. Soient X et Y des variables aléatoires indépendantes, on a alors :

1. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$;
2. $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$;

   N.B. Cela se généralise dans le cas multivarié. Soient X_1, X_2, ..., X_n des variables aléatoires, alors :

$$\mathbb{V}(X_1 + X_2 + \ldots + X_n) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + \ldots + \mathbb{V}(X_n) \quad (7.37)$$

3. $\operatorname{cov}(X, Y) = 0$.

Corollairement, les distributions X et Y sont **statistiquement indépendantes** si et seulement si :

$$f_{ij} = f_{i\cdot} f_{\cdot j} \quad (7.38)$$

pour toutes les valeurs des indices i et j. Dit autrement,

$$n_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n} \quad (7.39)$$

Pour qu'il y ait indépendance, il faut que l'égalité ait toujours lieu. Pour démontrer qu'il a **dépendance**, il suffit de fournir un seul cas pour lequel l'égalité n'a pas lieu. Mathématiquement, l'indépendance statistique correspond au fait que les lignes sont proportionnelles, ainsi que les colonnes. L'indépendance statistique de X et de Y correspond à la fois : 1. à l'indépendance de Y par rapport à X c'est-à-dire que les fréquences conditionnelles de Y pour $X = x_i$ ne dépendent pas de i ; 2. à l'indépendance de X par rapport à Y c'est-à-dire que les fréquences conditionnelles de X pour $Y = y_j$ ne dépendent pas de j. Différents **tests statistiques** peuvent être mis en œuvre pour vérifier l'indépendance de deux variables statistiques. Le plus utilisé est le test du χ^2.

7.1.9 Fonctions d'une variable aléatoire

Soient X et Y des variables aléatoires sur le même univers S, on dit alors que Y est une fonction de X si Y peut se mettre sous la forme $Y = \Phi(X)$ pour une certaine fonction Φ d'une variable réelle à valeurs réelles, c'est-à-dire si $Y(s) = \Phi[X(X)]$ pour tout $s \in S$. Par exemple, kX, X^2, $X+k$ et $(X+k)^2$ sont toutes des fonctions de X avec respectivement $\Phi(x) = kx$, x^2, $x+k$ et $(x+k)^2$.

Théorème. Soient X et Y des variables aléatoires sur un même univers S avec $Y = \Phi(x)$, on a alors :

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \Phi(x) f(x_i) \quad (7.40)$$

avec f la fonction de distribution de X.

De la même manière, on dit qu'une variable aléatoire Z est une fonction de X et Y si l'on peut représenter Z sous la forme $Z = \Phi(X, Y)$ avec Φ une fonction à valeurs réelles de deux variables réelles, c'est-à-dire si :

$$Z(s) = \Phi[X(s), Y(s)] \tag{7.41}$$

pour tout $s \in S$.

Théorème. Soient X, Y et Z des variables aléatoires sur un même univers S avec $Z = \Phi(X, Y)$, on a alors :

$$\mathbb{E}(Z) = \sum_{i,j}^{p,q} \Phi(x_i, y_j)\, h(x_i, y_j) \tag{7.42}$$

avec h la loi de probabilité produit de X et Y.

7.1.10 Combinaison linéaire de deux variables aléatoires

La combinaison linéaire de deux variables aléatoires a des conséquences sur la moyenne, la variance et l'écart type. Les propriétés ont été vues lors du chapitre exposant les paramètres statistiques.

7.2 Couple de variables aléatoires continues

7.2.1 Lois associées

La fonction de répartition conjointe F du couple (X, Y) est une application de $\mathbb{R}^2$ dans $[0, 1]$ définie par :

$$\forall (a, b) \in \mathbb{R}^2, F(a, b) = \Pr(X < a \text{ et } Y < b) \tag{7.43}$$

Le couple (X, Y) est absolument continu, s'il existe une fonction f continue des deux variables X et Y, appelée **densité de probabilité conjointe du couple** (X, Y) telle que, pour tout domaine D du plan, on ait :

$$\Pr[(X, Y) \in D] = \int_D \int f(x, y)\, \mathrm{d}x\mathrm{d}y \tag{7.44}$$

Si le domaine D est l'ensemble des couples (x, y) tels que $x \leq a$ et $y \leq b$, on obtient :

$$F(a, b) = \int_{-\infty}^{b} \mathrm{d}y \int_{a}^{-\infty} f(x, y)\, x \tag{7.45}$$

Entre la densité f et la fonction de répartition F, il existe la relation suivante :

$$f(x,y) = \frac{\mathrm{d}^2 F(x,y)}{\mathrm{d}x\mathrm{d}y} \tag{7.46}$$

Les **lois marginales** sont les lois des variables aléatoires X et Y prises séparément.

1. La fonction de répartition marginale de la variable X vaut :

$$\Pr(-\infty < X < a \text{ et } -\infty < Y < +\infty) = H(a) = \int_{-\infty}^{a} \mathrm{d}x \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y \tag{7.47}$$

2. La fonction de répartition marginale de la variable Y vaut :

$$\Pr(-\infty < Y < b \text{ et } -\infty < Y < +\infty) = G(b) = \int_{-\infty}^{b} \mathrm{d}y \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x \tag{7.48}$$

Remarque. On peut aussi écrire les lois marginales tels que :

$$H(a) = F(a, +\infty) \tag{7.49}$$

et

$$G(b) = F(-\infty, b) \tag{7.50}$$

Il est alors possible de calculer les **densités marginales** :

$$h(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y \tag{7.51}$$

et

$$g(y) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x \tag{7.52}$$

Il est alors possible de calculer les **densités conditionnelles** :

$$h(x \setminus y) = \frac{f(x,y)}{g(y)} \tag{7.53}$$

et

$$g(y \setminus x) = \frac{f(x,y)}{h(x)} \tag{7.54}$$

7.2.2 Indépendance

Soient f, h et g les densités de probabilités du couple (X, Y) et des variables X et Y prises séparément, alors les variables aléatoires continues X et Y sont indépendantes si et seulement si :

$$f(x, y) = h(x)\, g(y) \tag{7.55}$$

En analyse statistique, la question qui se pose souvent est de savoir, à la vue d'un tableau répartissant une population de taille n selon des modalités définies par deux variables aléatoires X et Y, si ces variables sont indépendantes ou non. Le test le plus utilisé est le test du χ^2 qui consiste à comparer le tableau observé et le tableau théorique que l'on obtiendrait si les variables X et Y étaient indépendantes.

7.2.3 Moments d'une distribution d'un couple de deux variables aléatoires

Le moment d'ordre p par rapport à la variable X et d'ordre q par rapport à la variable Y est l'espérance mathématique m_{pq} de la variable aléatoire $X^p Y^q$. Il est défini, sous réserve de l'existence de l'intégrale et en désignant par D le domaine de définition des variables aléatoires X et Y, par :

$$m_{pq} = \int_D \int x^p y^q f(x, y)\, \mathrm{d}x\mathrm{d}y \tag{7.56}$$

Dans les relations de dépendance, il apparaît souvent clairement que l'une des variables permet d'estimer l'autre. On établit de fait une relation fonctionnelle entre les deux dont il convient d'estimer et de tester la nature (droite, parabole, *etc*.). C'est là qu'intervient la notion de régression, dont la principale méthode est celle des moindres carrés.

Bibliographie