Statistiques Descriptifs

Cours n°1

19/01/2012

**Deuxième partie :**

*Les distributions à deux caractères*

*Ajustement, régression et corrélation*

La 1ère partie traitait une seule dimension (colonne xi et ni). Dans cette partie, on va doubler l’information (xi, yj et ni). Ce tableau porte le nom de tableau de contingence.

On va chercher à calculer des moyennes (pour x et y).

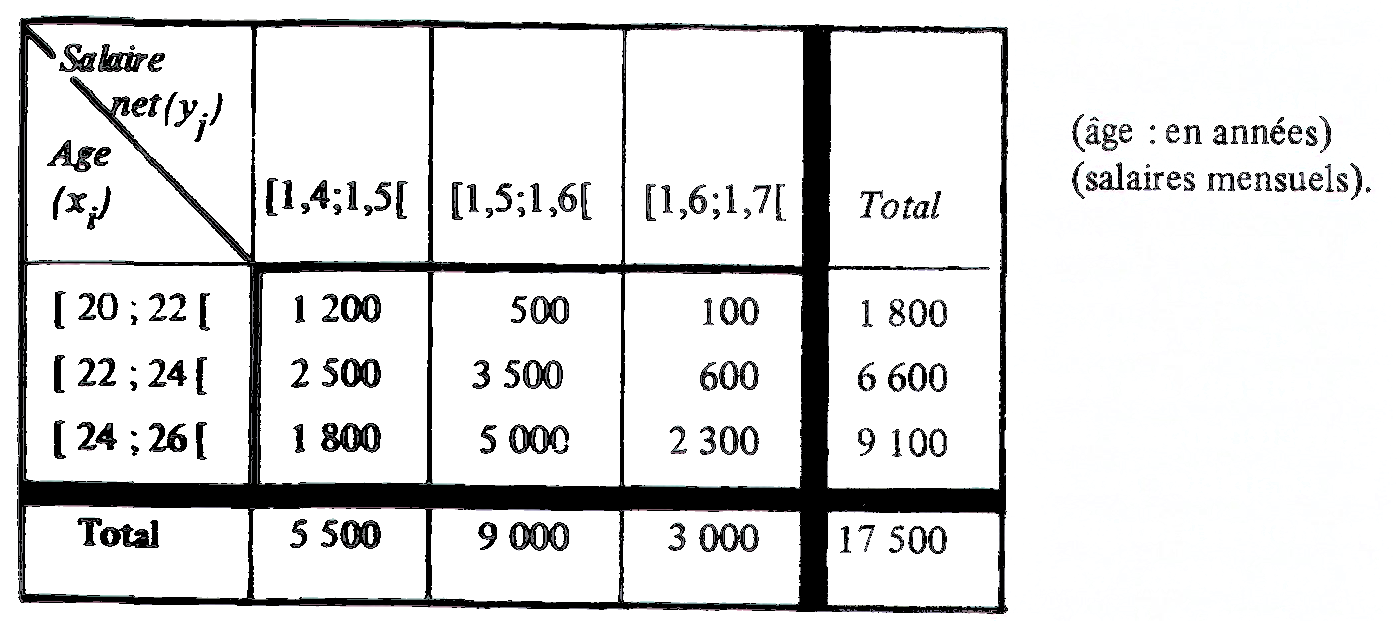
Chapitre I : Présentation des données

Section 1 : Les tableaux de contingence

**Paragraphe 1 : Construction des tableaux de contingence**

1. Les différents types de distribution à deux caractères

On a plusieurs types car les variables présentes plusieurs combinaisons de caractères (quantitatifs discrets, continues ou qualitatifs).



1. Exemples de répartition

On a la répartition de 17500 diplômés. On remarque que deux tableaux à une dimension n’équivalent pas un tableau à deux dimensions car on ne connait pas le nombre de diplômés dans chaque classe.

La ligne et la colonne appelées ici «  Total », prennent le nom de « marges ». Si l’on associe la marge du bas et la ligne du haut (yj), on obtient la distribution des 17500 jeunes salariés selon leur salaire net (en milliers d’euros). Cette distribution est à une dimension.

On l’appelle distribution marginale du caractère y. De même ; la dernière colonne (Total) associée à la première, n’est autre que la distribution marginale des individus selon leur âge.

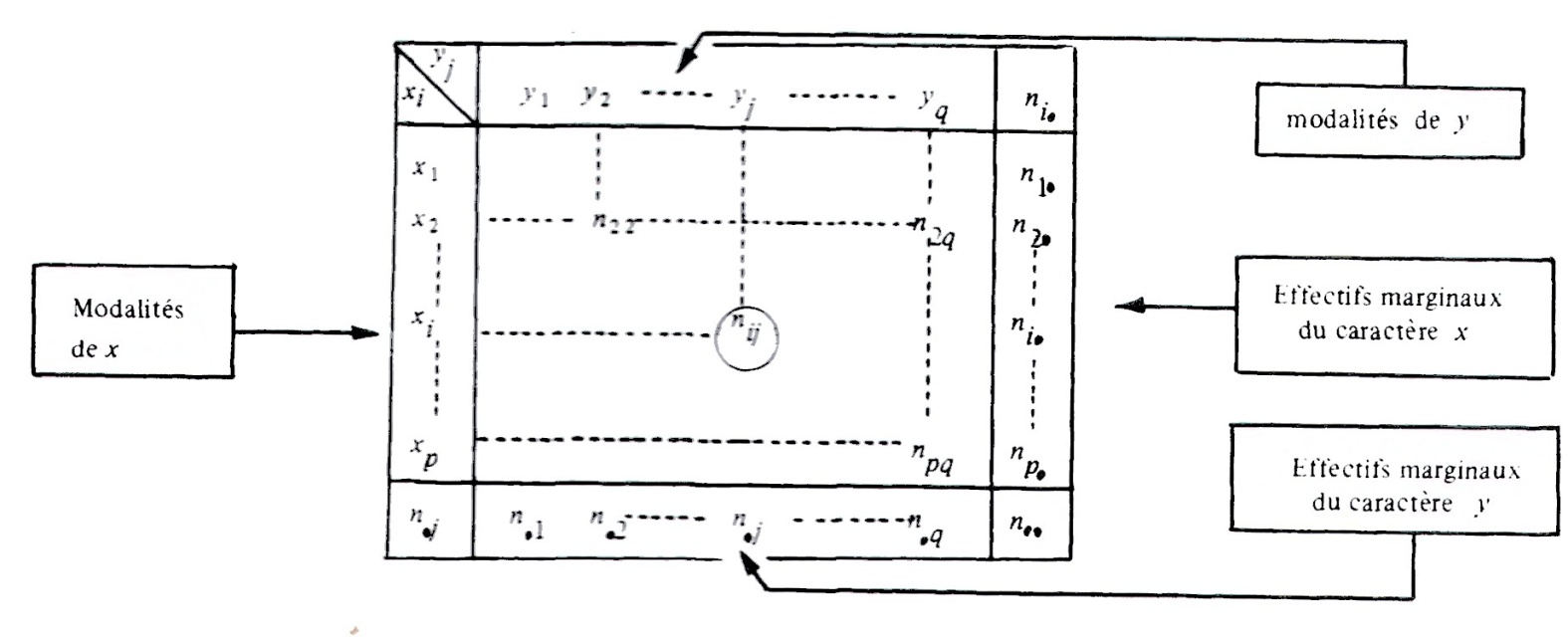
Si l’on s’en tenait uniquement à cette double représentation à une dimension, c’est-à-dire, en utilisant seulement les distributions marginales, rien ne nous permettrait de déterminer si l’âge a une influence sur le niveau de salaire, ou l’inverse.

Le tableau de contingence, par contre, nous permet de voir comment se distribuent les effectifs de chaque modalité d’un caractère, suivant les modalités de l’autre. On a, en quelque sorte, « croisé » l’information.

Dans le totale des colonnes, on remplace par un point, l’indice qui varie.

Les marges correspondent aux distributions marginales.

1. Contenus et Lectures des tableaux de contingences

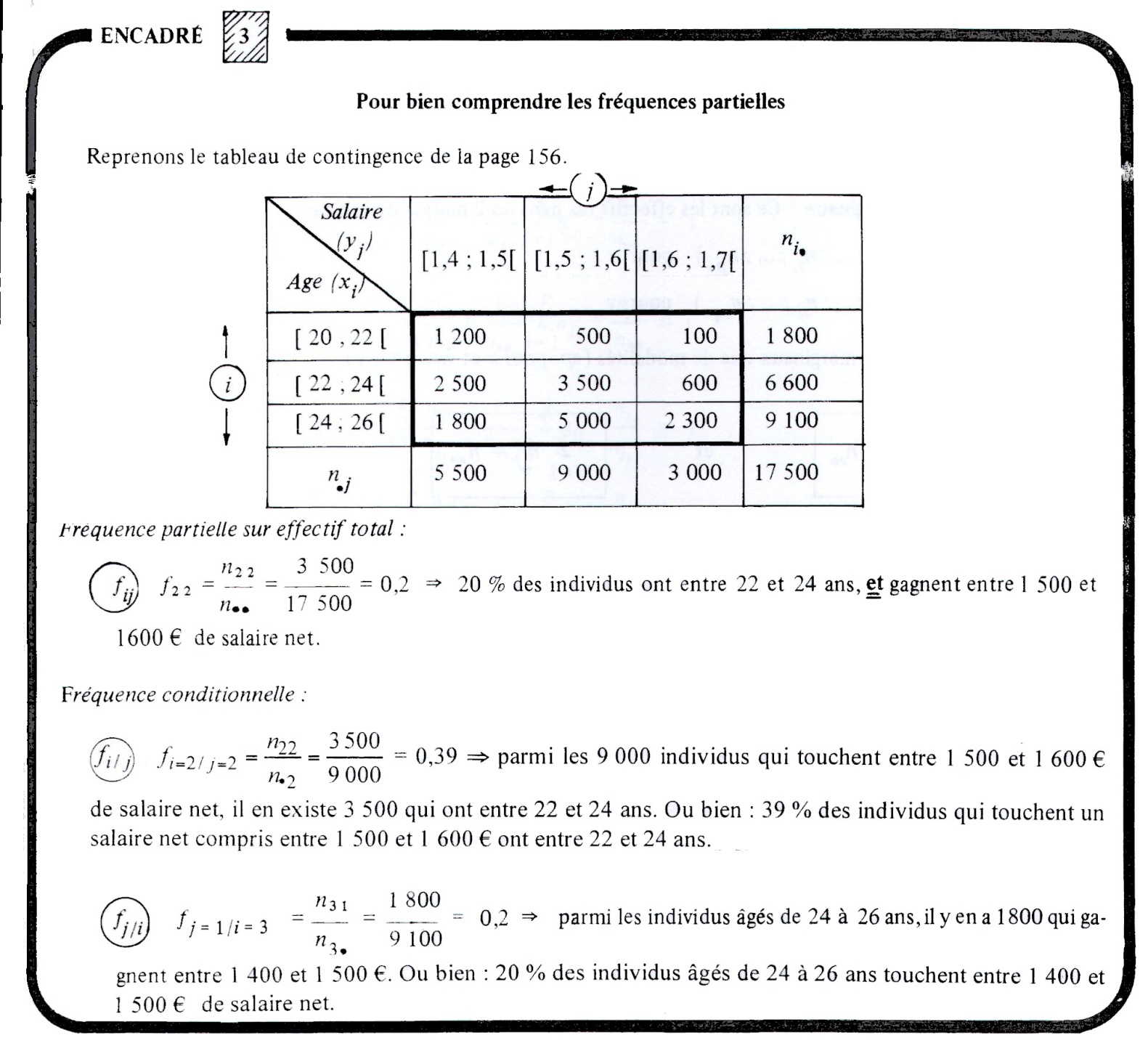


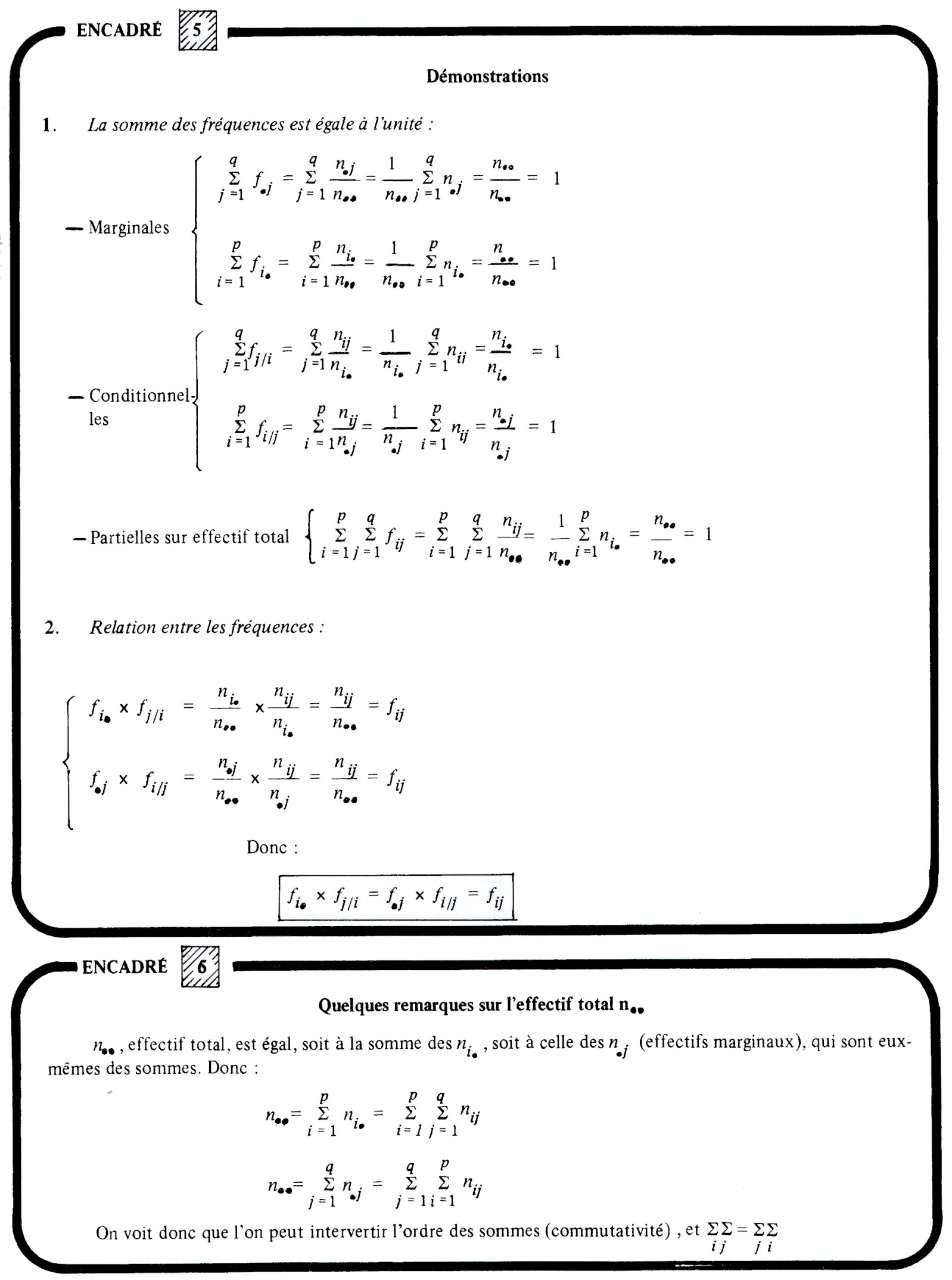
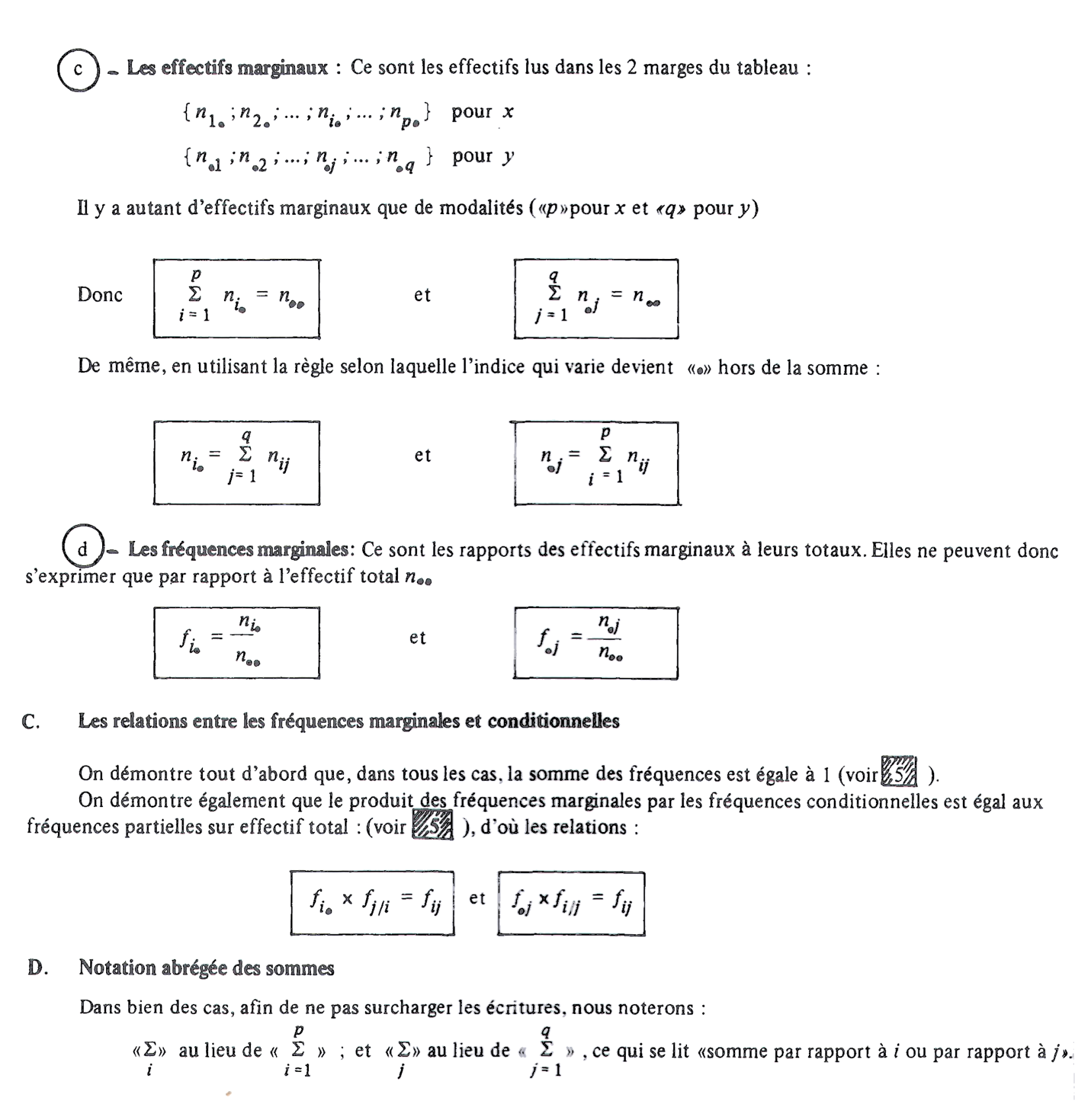
**Les éléments du tableau :**

*Les effectifs partiels :* Ce sont les effectifs lus à l’intérieur du tableau de contingence (repérés par nij). Ils correspondent au nombre d’individus présentant à la fois la modalité xi et la modalité yj.

*Les fréquences partielles :* Il existe deux types de fréquences relatives partielles

* Celles qui correspondent au rapport de l’effectif partiel sur l’effectif total : On les appelle fréquences partielles sur effectifs total
* Celles qui correspondent au rapport de l’effectif partiel sur l’effectif marginal : on les appelle fréquences conditionnelles.







Section 2 : Les caractéristiques des séries à deux caractères

On se limitera aux moyennes arithmétiques et aux variances dans un premier temps. Il faut bien comprendre que, puisqu’il existe des fréquences marginales et conditionnelles, on déterminera des moyennes et des variances marginales et conditionnelles.

Ainsi, si nous reprenons l’exemple de la section 1, l’âge moyen et la variance des âges seront calculés à partir de la distribution marginale de x. De la même manière, le salaire moyen et la variance des salaires seront tirés de la distribution marginale de y.

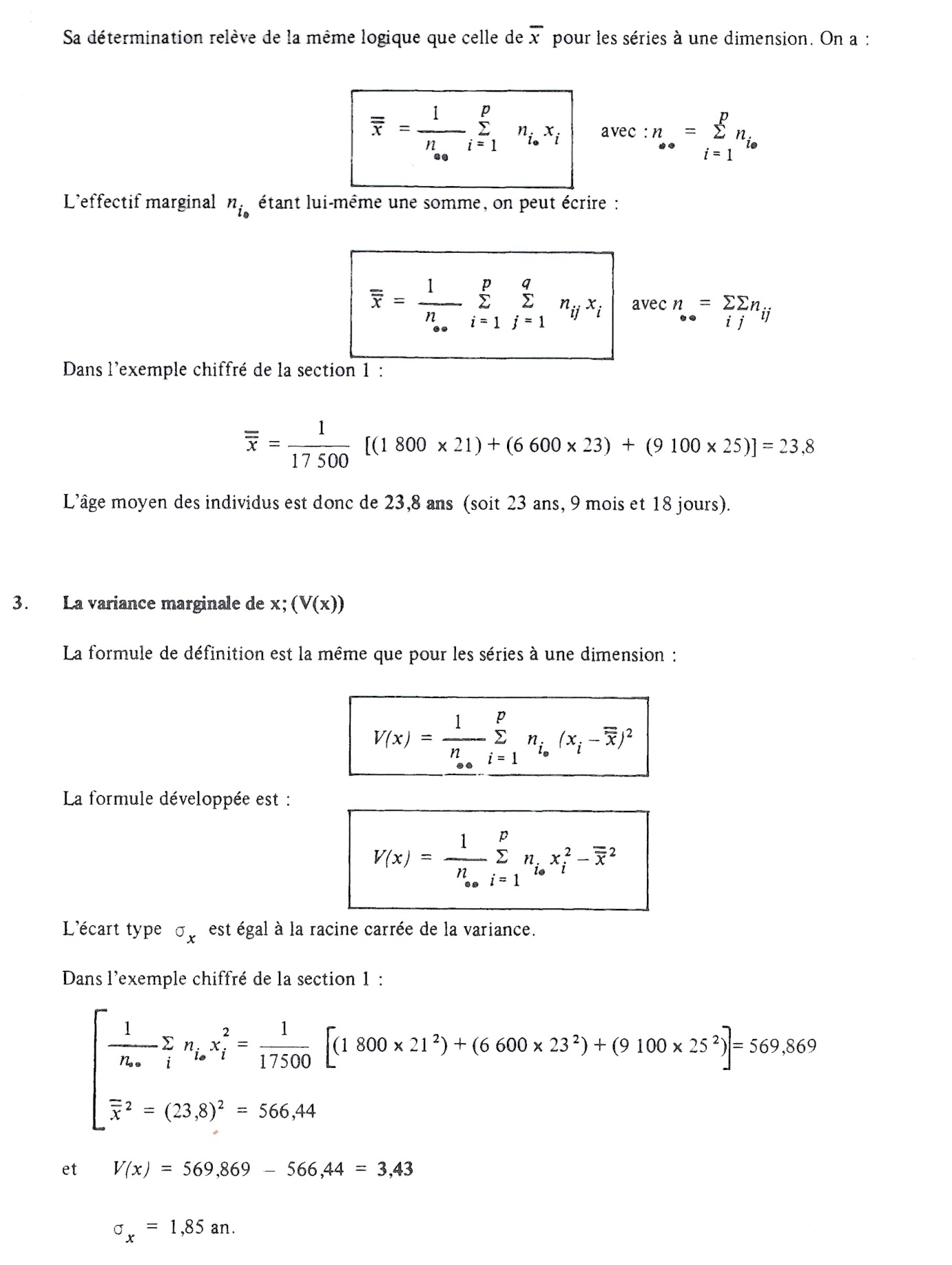
L’âge moyen et la variance des âges pour chaque classe de salaire, seront tirés des distributions conditionnelles de x en y (« i » varie, « j » est fixé), et inversement. On emploi indifféremment le terme de « distribution » ou de « loi ».

**Paragraphe 1 : Caractéristiques des lois marginales**

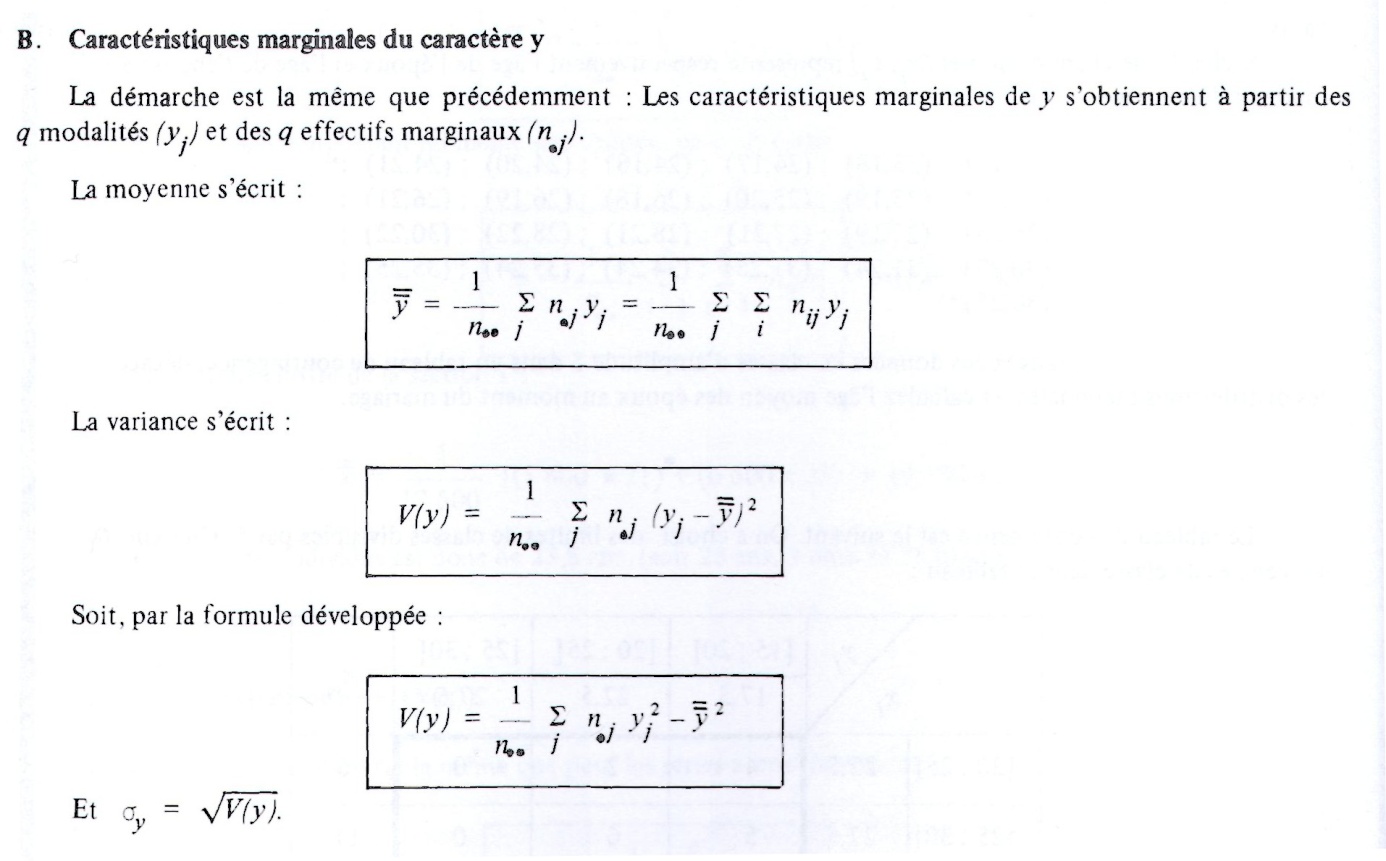
1. Caractéristiques des lois marginales

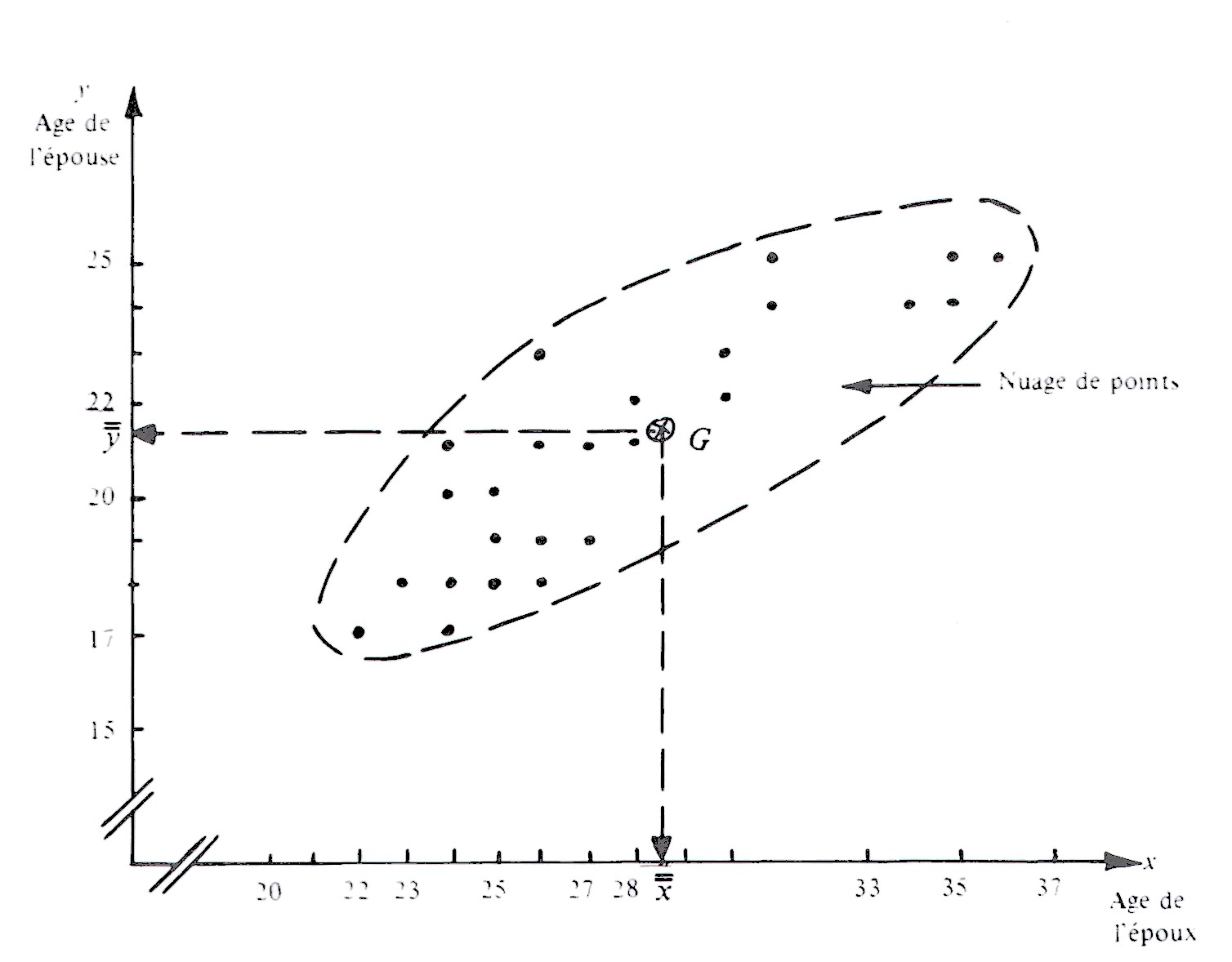
Les caractéristiques marginales de x (moyenne et variance) s’obtiennent à partir de la distribution marginale de x, c’est-à-dire à partir des « p » modalités (xi) et des « p » effectifs marginaux (ni.).

La moyenne marginale est une moyenne arithmétique.



1. Caractéristiques marginales du caractère y

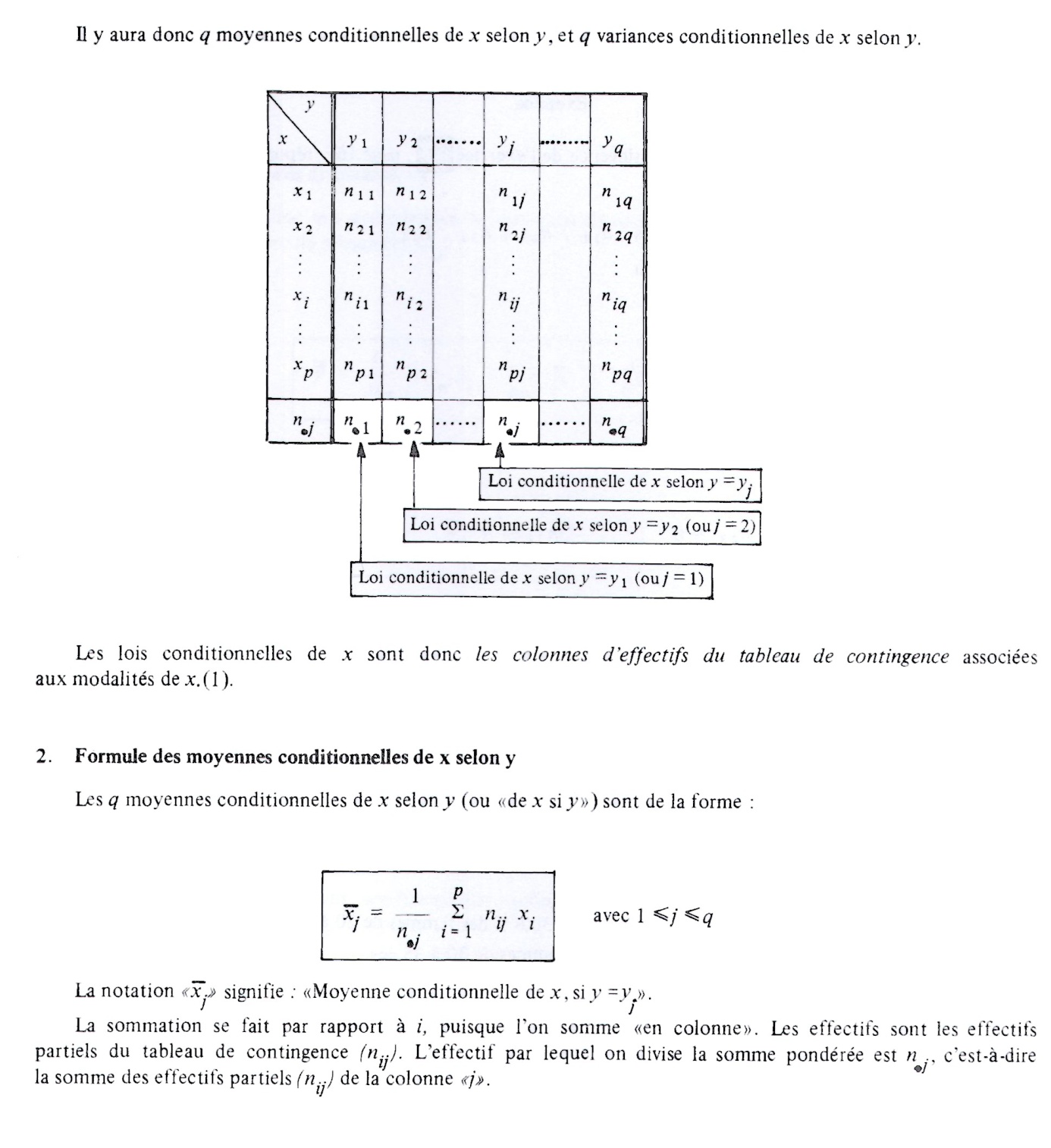




**Paragraphe 2 : Caractéristiques des lois conditionnelles**

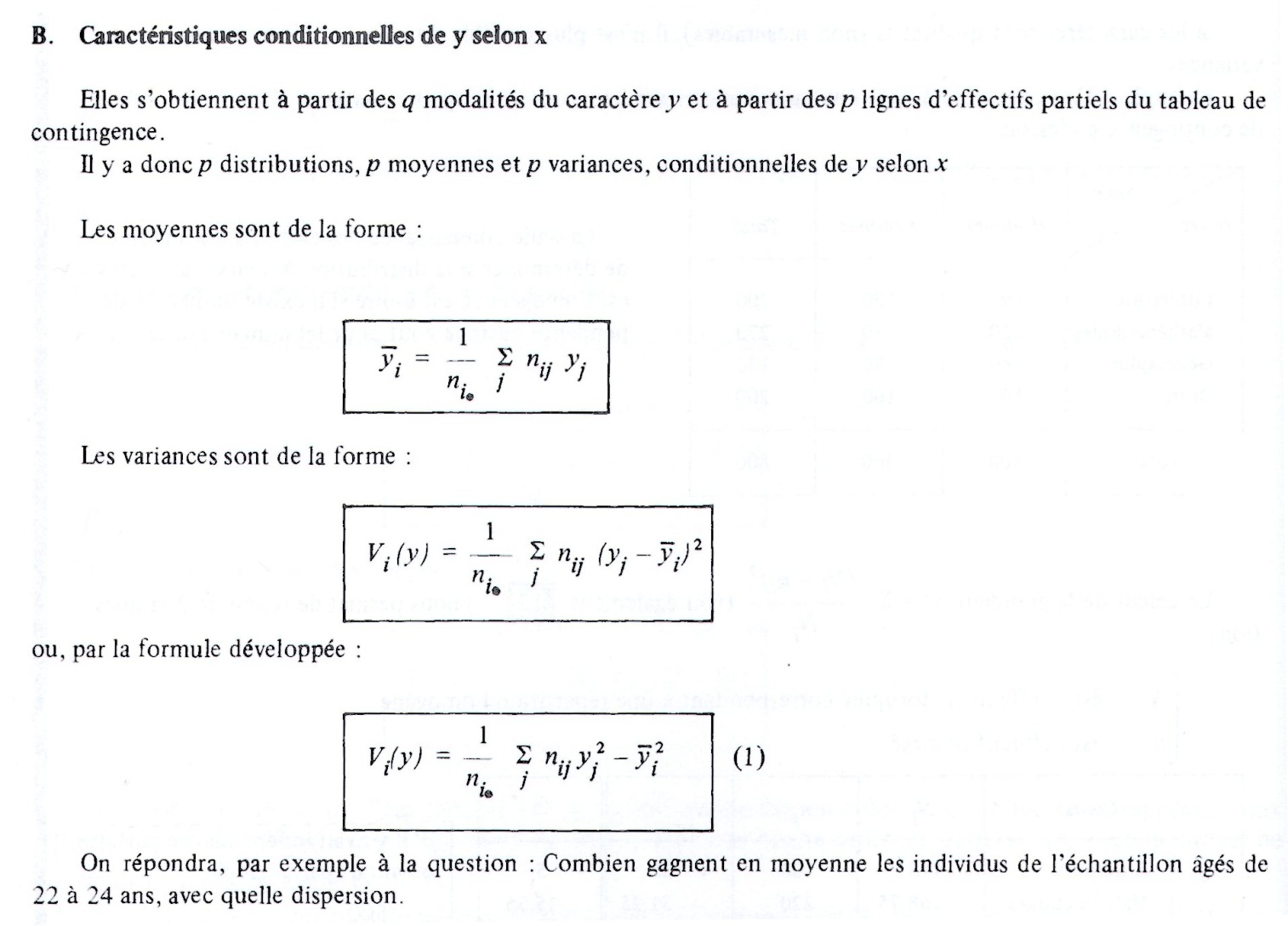
1. Caractéristiques conditionnelles de x selon y
2. *Définitions*

Elles s’obtiennent à partir des p modalités du caractère x et à partir des q colonnes d’effectifs qui correspondent à y. On a q moyennes conditionnelles de x selon y, et q variances conditionnelles de x selon y.



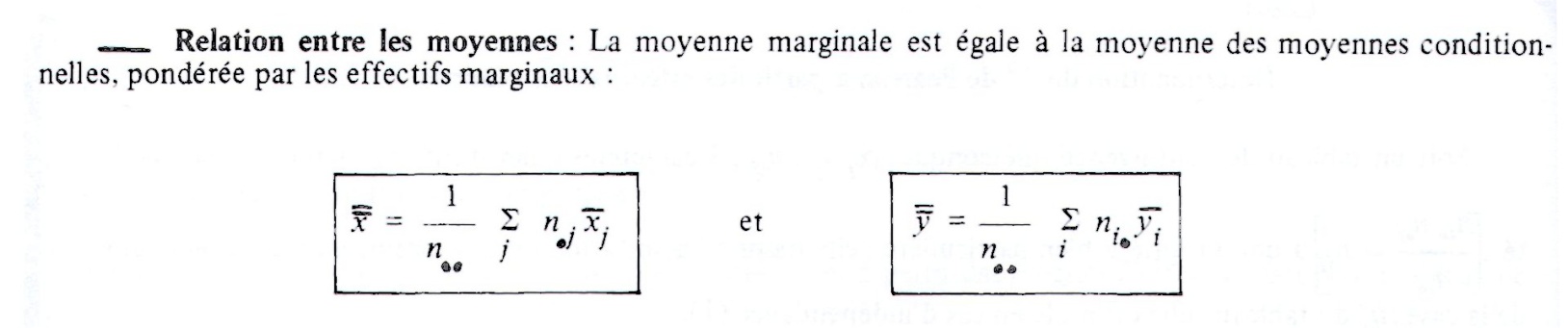
1. Caractéristiques de y selon x
2. *Définitions*

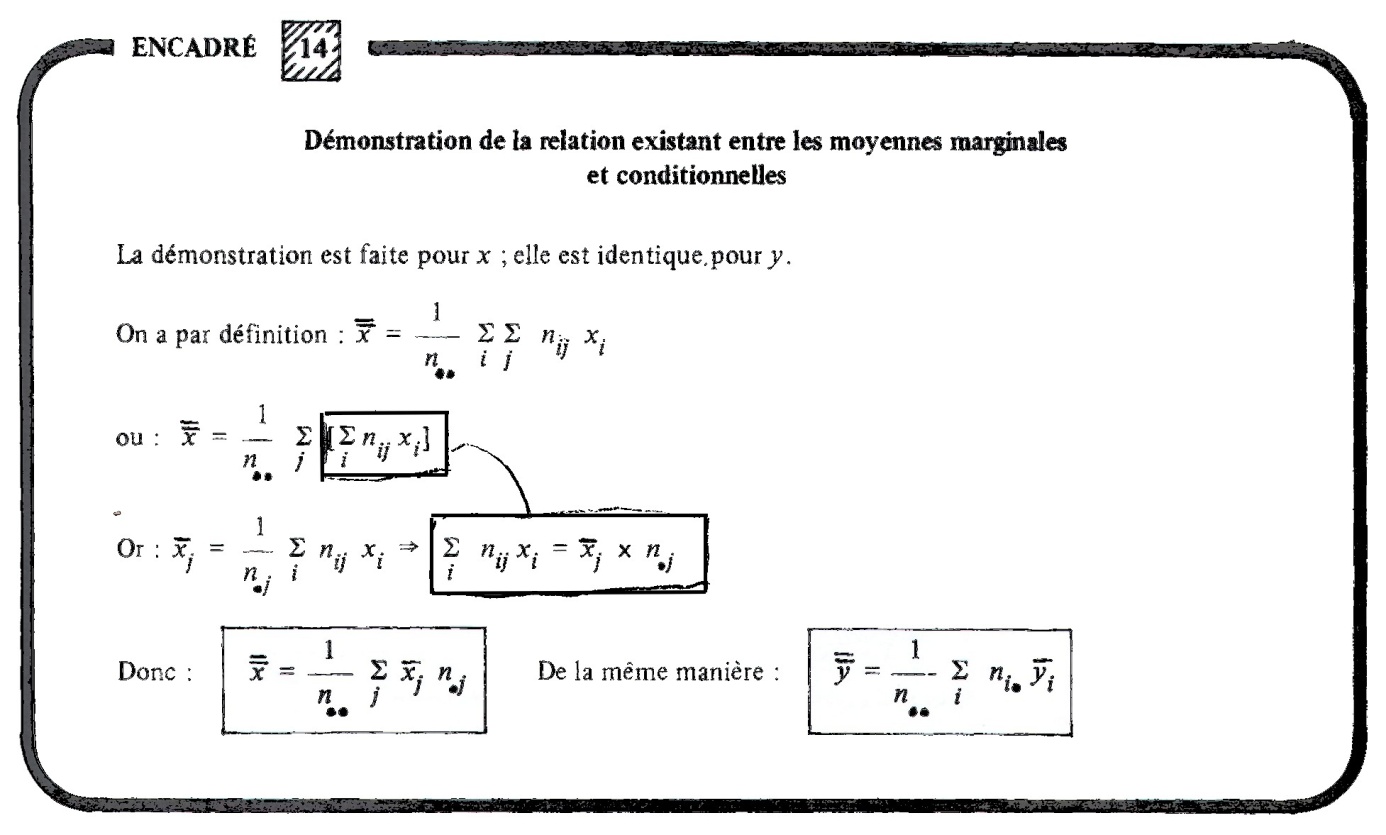
Elles s’obtiennent à partir des q modalités du caractère y et à partir des p lignes d’effectifs partiels du tableau de contingence. On a p moyennes conditionnelles.



**Paragraphe 3 : Relations entre les caractéristiques marginales et conditionnelles**

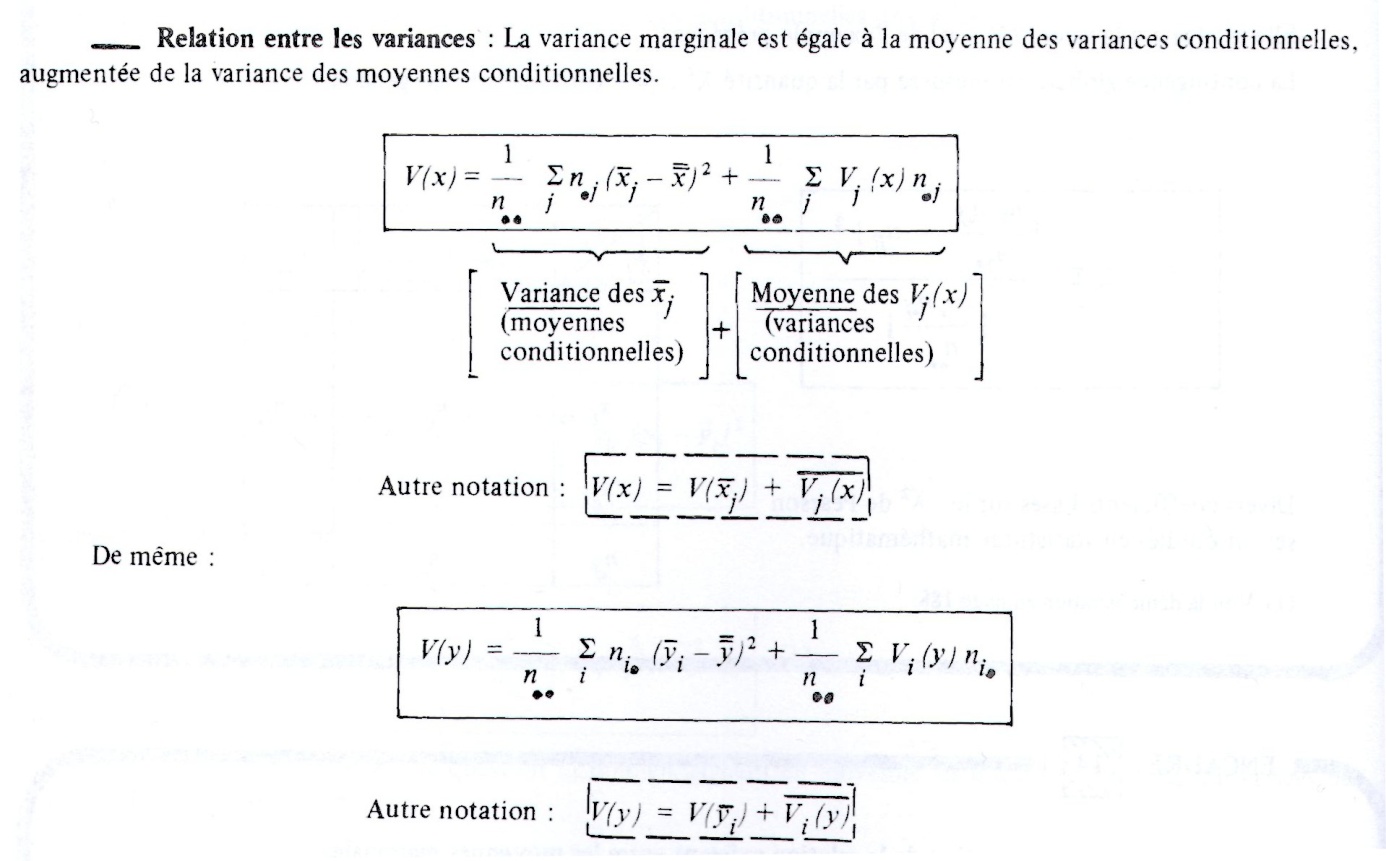
1. Relation entre les moyennes

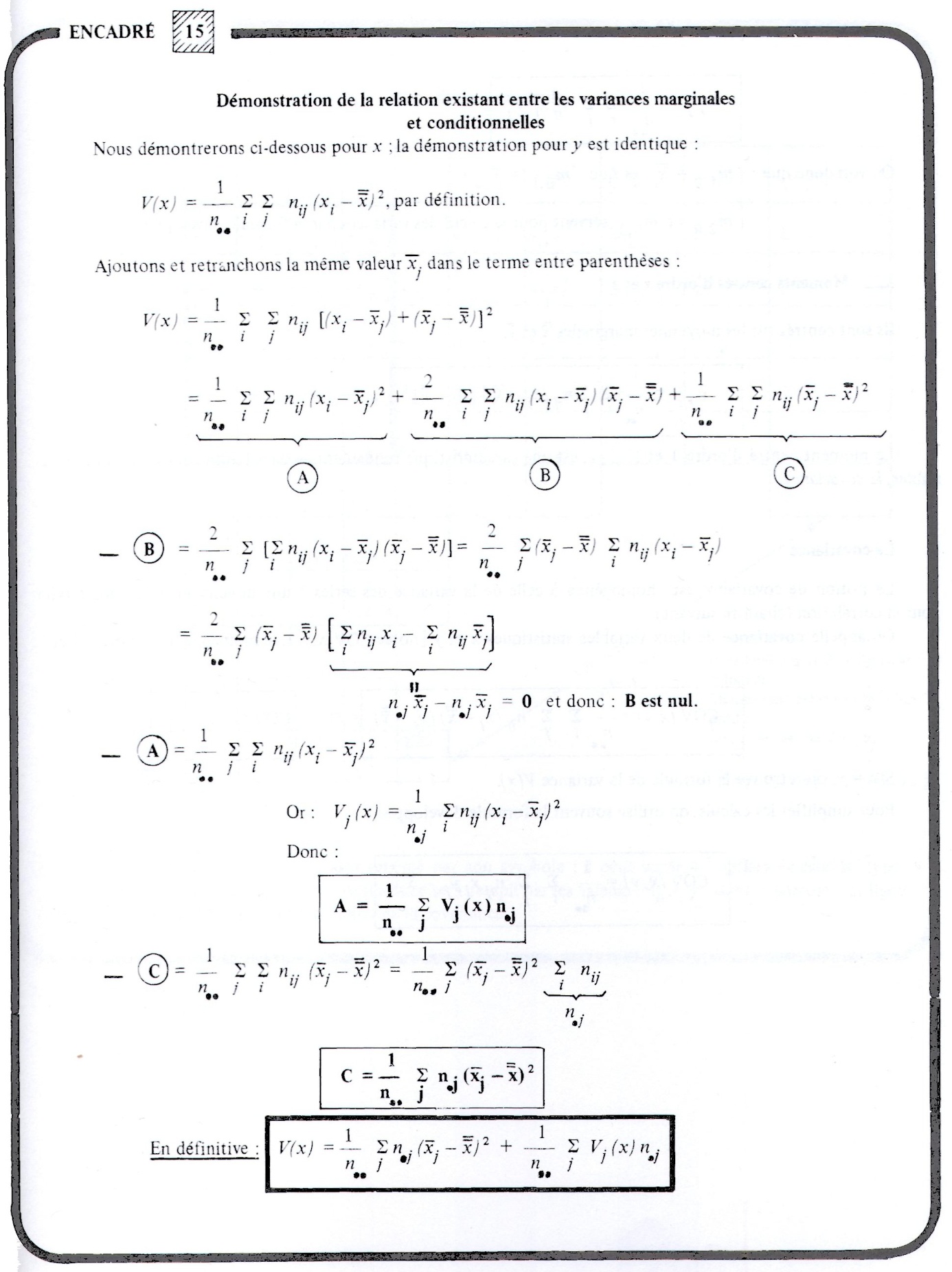




La moyenne marginale est égale à la moyenne de la moyenne conditionnelle pondérée par les effectifs marginaux.

1. Relation entre les variances

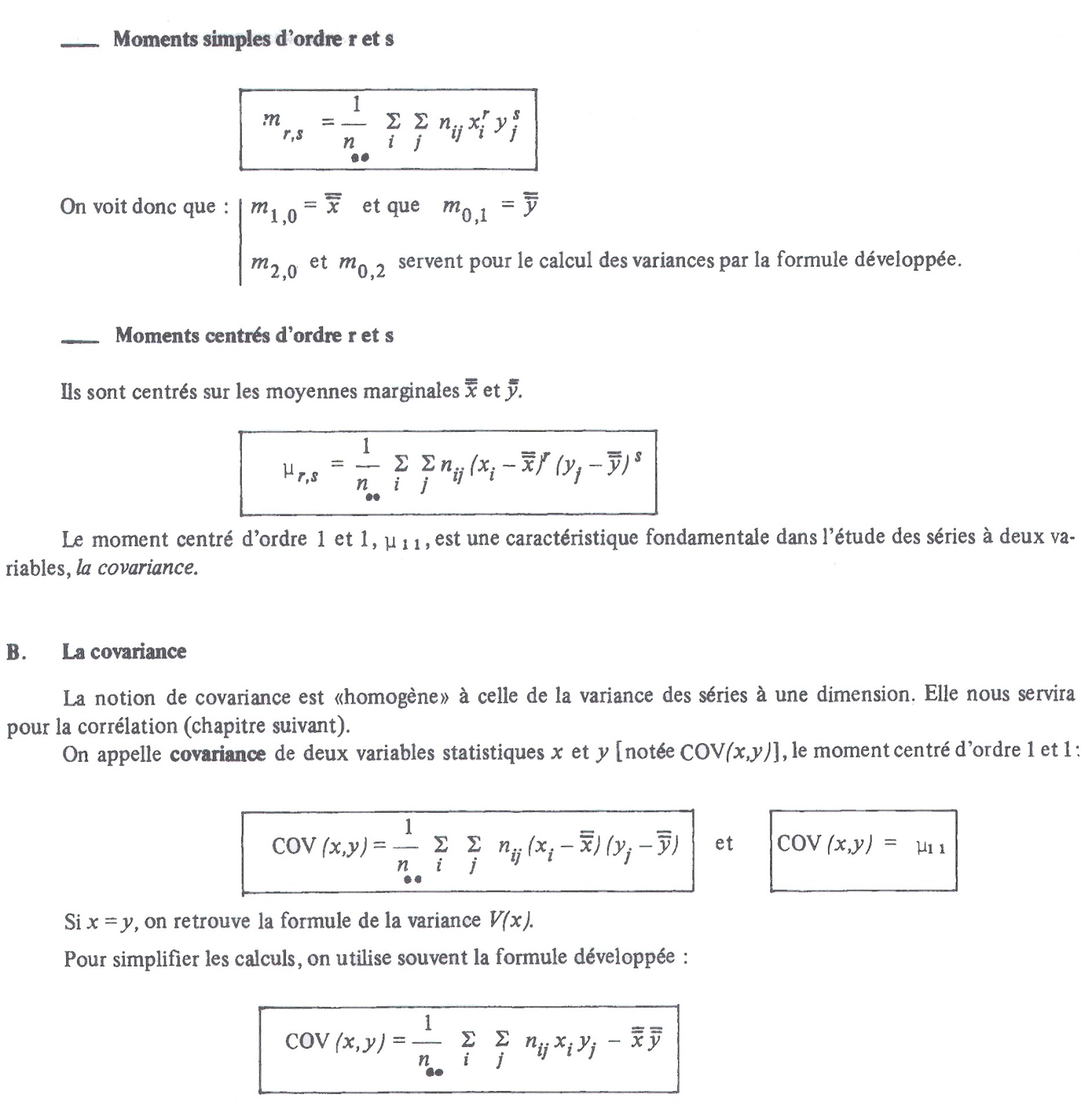




Cours n°2

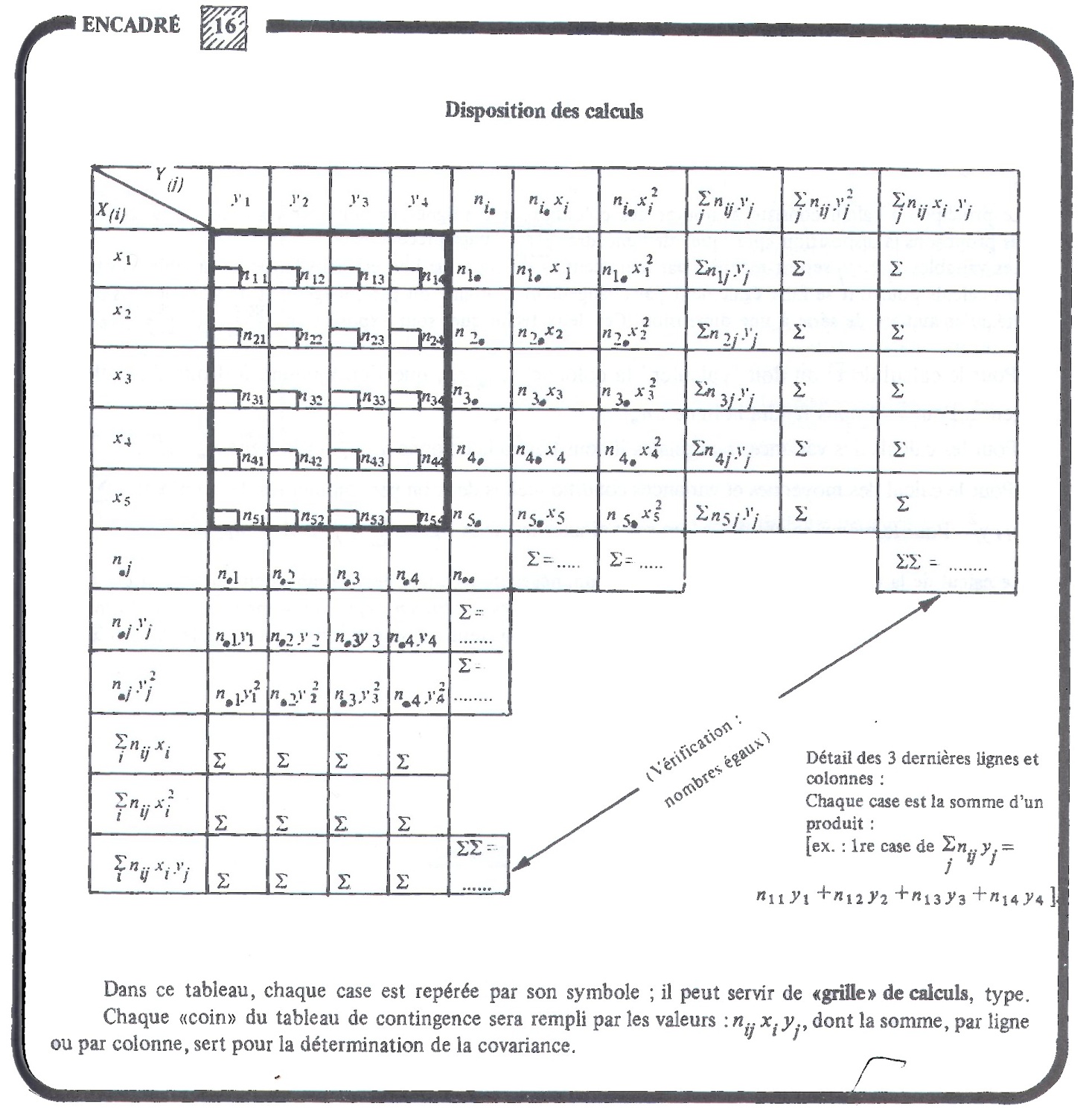
26/01/2012

**Paragraphe 4 : Les moments et la covariance**



Lorsque x = y, la covariance égale à la variance de x.

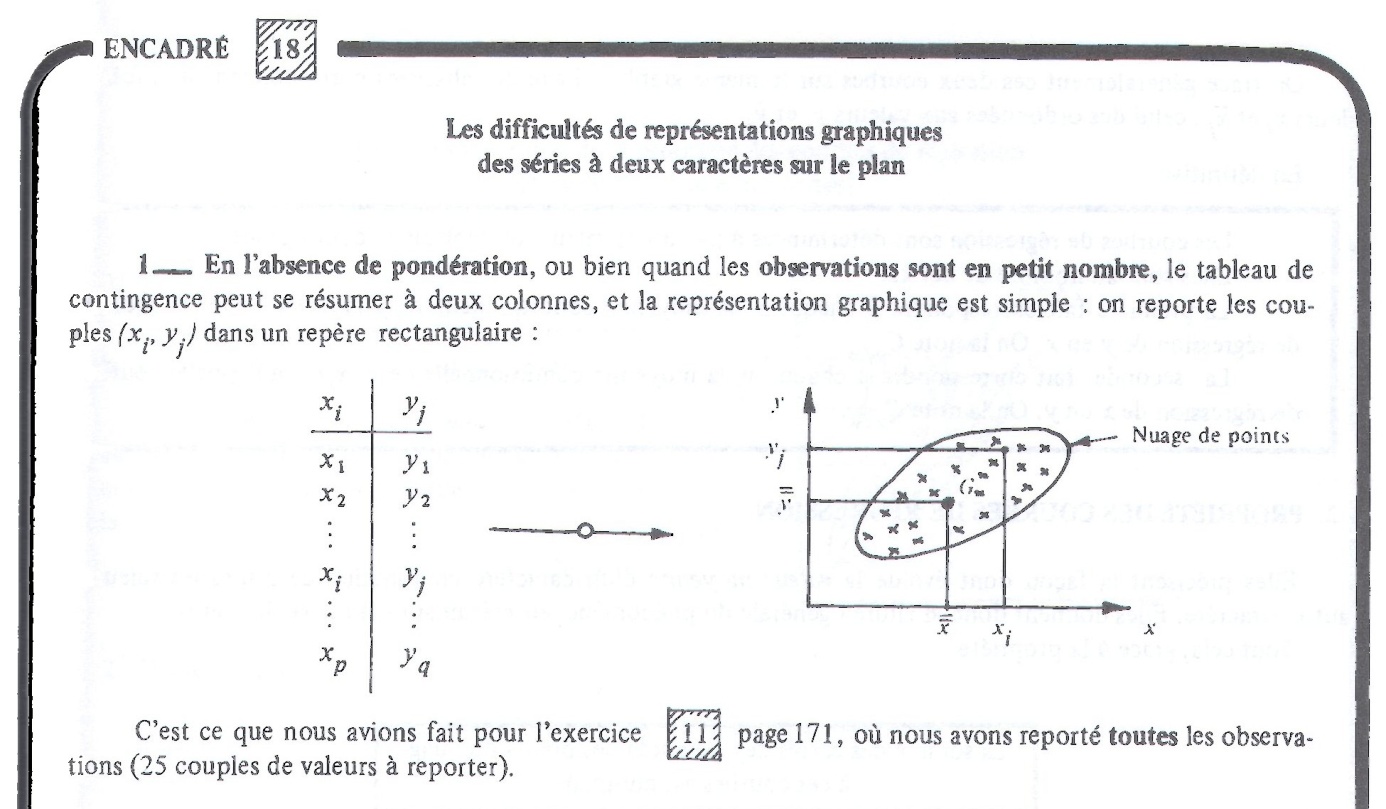
**Paragraphe 5 : Calcul pratique des caractéristiques des séries à deux caractères**



Section 3 : Représentations graphiques et notions de courbes de régressions

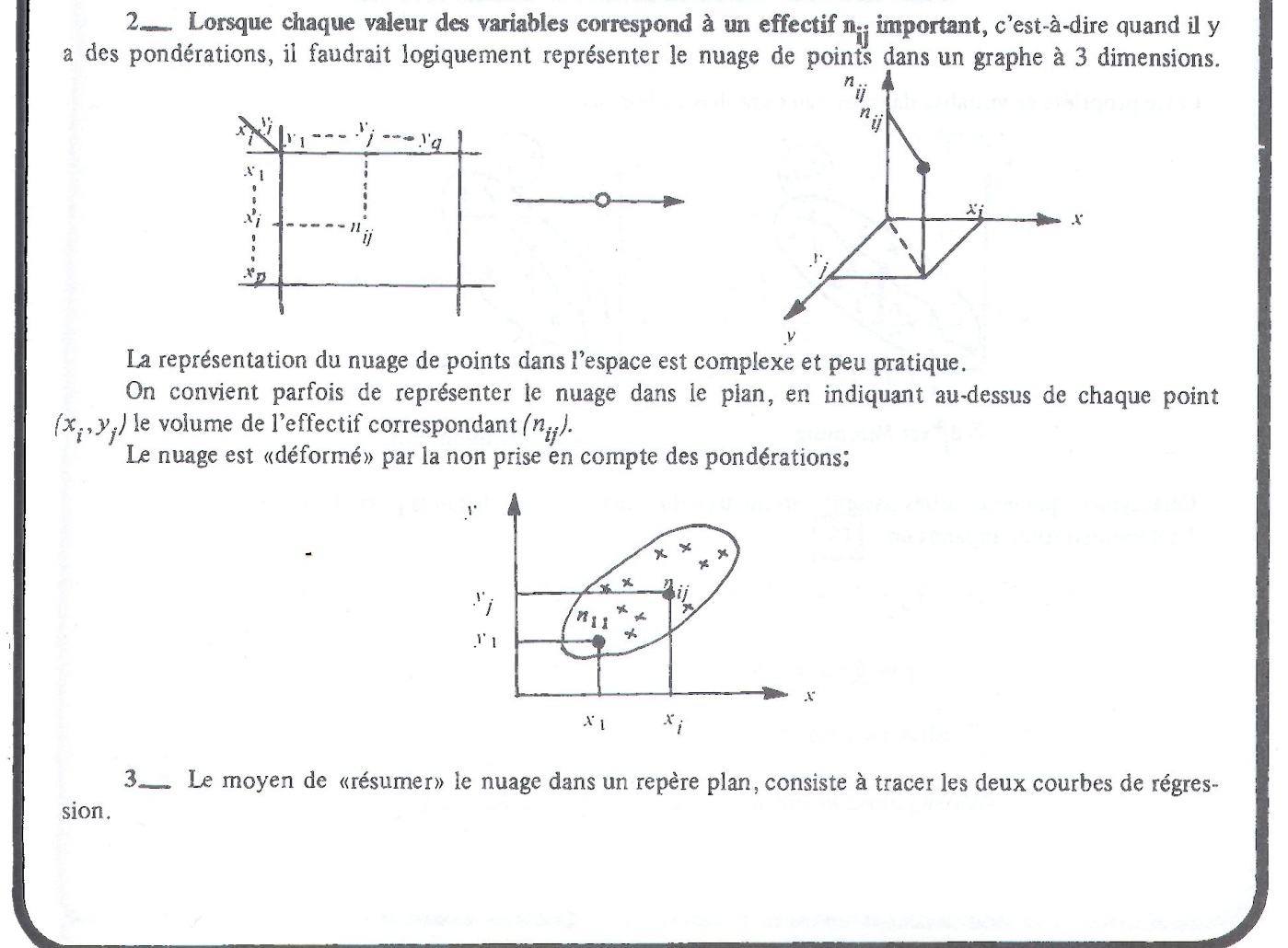
**Paragraphe 1 : Représentation à partir des courbes xi et yj**

1. Absence de pondération :



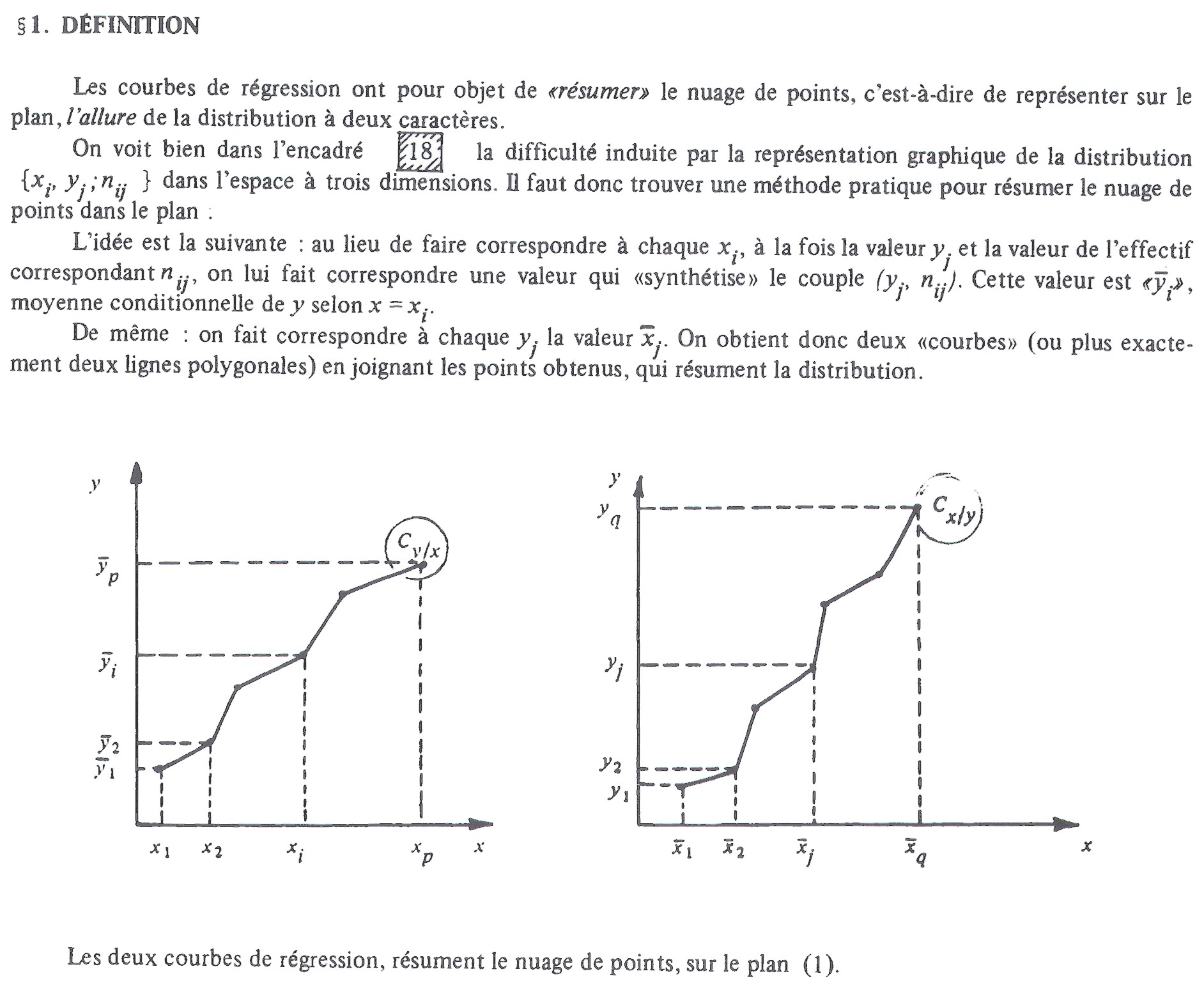
La plupart du temps, il y a des pondérations.

1. Présence de pondération :



**Paragraphe 2 : Représentation par les courbes de régression**

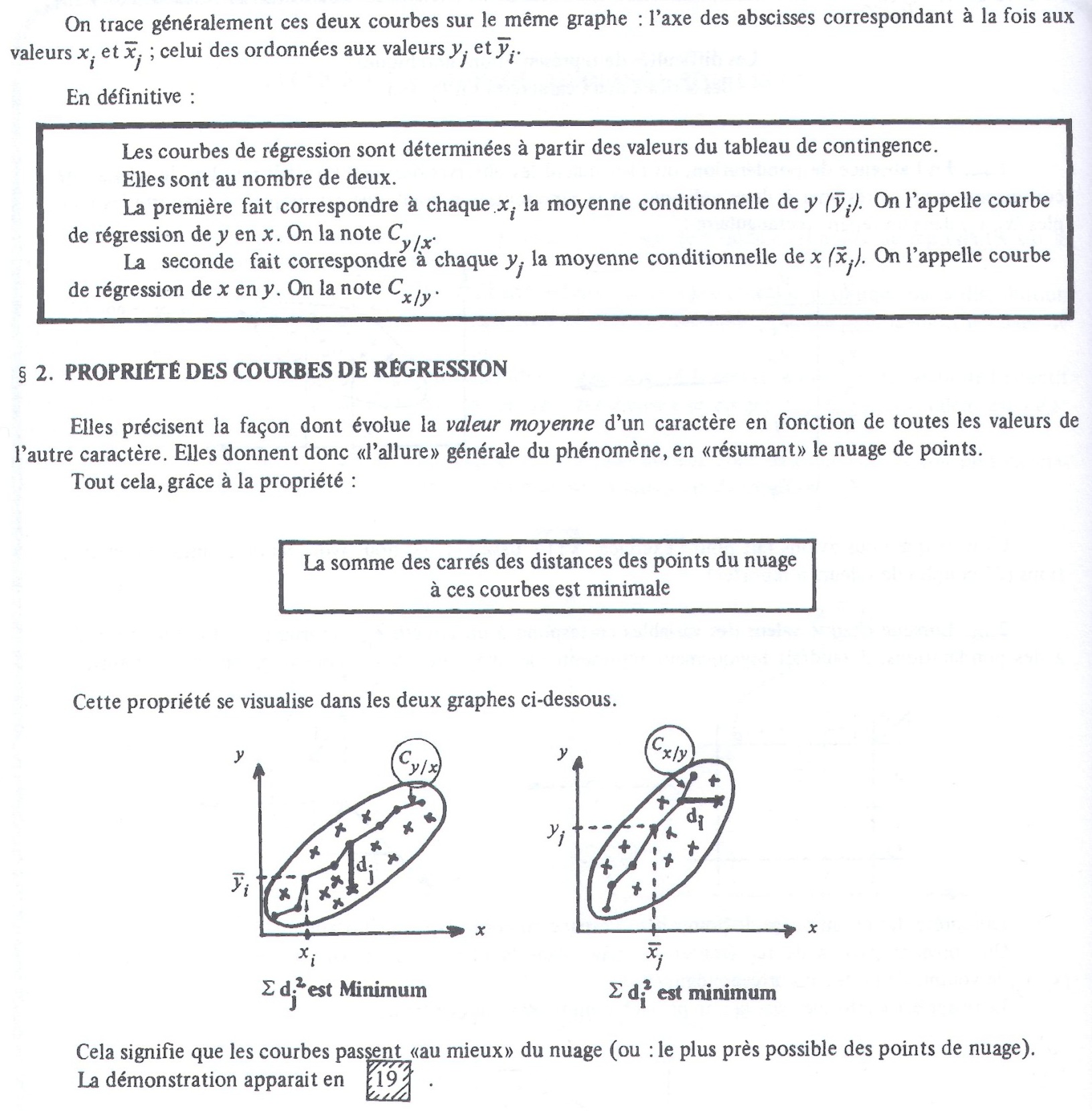
1. Définition



Au lieu de faire correspondre à chaque xi, un yj et un nij, on va faire correspondre une valeur qui résume nij et yi. 🡺 Y barre i (moyenne condition de y selon x).

On obtiendra 2 lignes brisées.

1. Propriété des courbes de régression



Méthodes des moindres carrés