

## CORRECTION DE GROUPEMENT (C2, F)

(25 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **caractéristique** empirique (eg **moment**) est généralement calculée à partir d'un certain nombre de « **valeurs individuelles** », ie à partir des **observations** de diverses **variables** effectuées sur des **unités statistiques** (cf aussi **théorique, empirique, statistique naturelle**).

On appelle **correction de groupement** la **correction** (ou **modification**) de cette caractéristique empirique consistant à la calculer sur un **groupement de classes** des observations précédentes.

Le cas typique est celui d'une caractéristique calculée à partir d'un **histogramme** : la valeur retenue pour chaque classe de l'histogramme est souvent eg le centre de la classe, au lieu d'être une caractéristique moyenne de chaque classe (ie calculée sur l'ensemble des valeurs figurant dans cette classe).

Autrement dit, dans le cas d'un histogramme à une dimension dans **R**, si  $p$  désigne le nombre de classes tq  $[x_{j-1}, x_j[$ , où les  $x_j$  ( $j = 2, \dots, p$ ) désignent les limites (ou extrémités) de classes, le « centre »  $x_{c(j)}$  d'une classe  $[x_{j-1}, x_j[$  est calculé selon :

$$(1) \quad x_{c(j)} = (x_{j-1} + x_j) / 2$$

et non pas selon :

$$(2) \quad x_{c(j)} = \bar{X}_j = (\sum_{n=1}^N \mathbf{1}(X_n \in [x_{j-1}, x_j[))^{-1} \cdot \sum_{n=1}^N \mathbf{1}(X_n \in [x_{j-1}, x_j[) \cdot X_n,$$

ie la moyenne des observations (valeurs de l'échantillon)  $X_n$  ( $n = 1, \dots, N$ ) appartenant à cette classe, où  $\mathbf{1}(R)$  désigne l'**indicatrice** d'une partie  $R \subset \mathbf{R}$ .

Ce type d'**agrégation** entraîne, en effet, une perte de « **précision** » ou d'« **information** ».

Selon les hypothèses retenues, le type de correction peut varier.

Un exemple « historique » est la **correction de SHEPPARD**.