

Covariance - Formules

On se reportera, si nécessaire, au site <http://piaget.psych.univ-paris5.fr/LMD/> (rubrique Bases Math) pour quelques exercices préalables sur le signe Σ .

Soit deux variables numériques X et Y observées sur un échantillon de n individus

Démontrer l'équivalence des deux formules suivantes de la covariance :

$$\text{Cov}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

$$\text{Cov}(X, Y) = \frac{\sum (xy)}{n} - \bar{x} \bar{y}$$

On a :

$$\text{Cov}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

$$= \frac{\sum (xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})}{n}$$

$$= \frac{\sum xy - \sum x\bar{y} - \sum \bar{x}y + \sum \bar{x}\bar{y}}{n} \quad (\text{distributivité de la somme})$$

$$= \frac{\sum xy - \bar{y} \sum x - \bar{x} \sum y + \sum \bar{x}\bar{y}}{n} \quad \text{car } \bar{y} \text{ et } \bar{x} \text{ sont des constantes (sur le protocole)}$$

$$= \frac{\sum xy - \bar{y}n\bar{x} - \bar{x}n\bar{y} + \sum \bar{x}\bar{y}}{n} \quad \text{car } \sum x = n\bar{x} \text{ (idem pour } Y)$$

$$= \frac{\sum xy - \bar{y}n\bar{x} - \bar{x}n\bar{y} + n\bar{x}\bar{y}}{n} \quad \text{car } \sum cste = n \cdot cste$$

$$= \frac{\sum xy - n\bar{x}\bar{y}}{n}$$

$$= \frac{\sum xy}{n} - \bar{x}\bar{y}$$