ANNEXE À LA SECONDE PARTIE :
RAPPEL DE QUELQUES FORMULES COURANTES EN STATISTIQUE

Pour une présentation plus complète, le lecteur est référé à Gilles (1994), chapitres 3 à 5, ainsi que sections 6.1, 6.2 et 8.1. On peut aussi consulter le chapitre 2 de Wonnacott et Wonnacott (1992).

Nous nous contentons ici de reproduire quelques formules de mesures parmi les plus couramment utilisées en statistique. Pour certaines d'entre elles, nous donnons deux formules : celle qui s'applique à la population et celle qui s'applique à un échantillon. La raison d'être de cette distinction est exposée en rapport avec l'induction statistique (on utilise pour l'échantillon une formule qui produit un estimateur *non biaisé* du paramètre correspondant de la population). Puisque la statistique descriptive ne distingue pas entre la population et l'échantillon, c'est généralement la formule de la population qui est employée dans ce cas.

<u>Notation</u>

n = nombre d'observations

x_i = valeur de la variable X à la $i^{\text{ème}}$ observation

y_i = valeur de la variable Y à la $i^{\text{ème}}$ observation

3.1 Mesures de tendance centrale

- Moyenne

  (1) population [1] : $\mu_X = \left(\frac{1}{n}\right)\sum_i x_i$

  (2) échantillon : $m_X = \left(\frac{1}{n}\right)\sum_i x_i$

- Médiane : c'est la valeur $\breve{x}$ de la variable X telle que 50 % de la population ou de l'échantillon ont des valeurs inférieures à $\breve{x}$, alors que 50 % ont des valeurs qui lui sont supérieures.

[1] Les formules données ici s'appliquent à des populations finies. Ces formules peuvent se généraliser à des populations infinies au moyen du concept d'espérance mathématique.

- Mode : dans un échantillon ou une population finie, c'est la valeur la plus fréquente de la variable X ; lorsque les observations sont regroupées par classes, c'est la classe qui a la fréquence la plus élevée ; dans une population infinie, c'est la valeur à laquelle correspond la plus grande densité de probabilité.

3.2 Mesures de dispersion

- Variance

  (1) population : $\sigma_x^2 = \frac{1}{n}\sum_i (x_i - \mu_x)^2$

  (2) échantillon : $s_x^2 = \frac{1}{n-1}\sum_i (x_i - m_x)^2$

- Écart type

  (1) population : $\sigma_x = \sqrt{\sigma_x^2}$

  (2) échantillon : $s_x = \sqrt{s_x^2}$

- Coefficient de variation

  (1) population : $C_x = \frac{\sigma_x}{\mu_x}$

  (2) échantillon : $C_x = \frac{s_x}{m_x}$

3.3 Mesures d'association

- Covariance

  (1) population : $\sigma_{xy} = \frac{1}{n}\sum_i (x_i - \mu_x)(y_i - \mu_y)$

  (2) échantillon : $s_{xy} = \frac{1}{n-1}\sum_i (x_i - m_x)(y_i - m_y)$

- Coefficient de corrélation simple

(1) population : $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \, \sigma_y}$, avec $-1 < \rho_{xy} < +1$

(2) échantillon : $r_{xy} = \dfrac{s_{xy}}{s_x \, s_y}$, avec $-1 < r_{xy} < +1$

où $r_{xy} = \rho_{xy}$, puisque $s_{xy} = \sigma_{xy} \dfrac{n}{n-1}$ et $s_x^2 = \sigma_x^2 \dfrac{n}{n-1}$