Démonstration de la formule de travail de la variance

Jean-Etienne Poirrier*

14 juillet 2005

Soit l'échantillon $\{x_1, x_2, x_3, ..., x_n\}$. L'effectif est donc n (variables quantitatives) et la moyenne est $\overline{x}$.

La théorie dit que la formule de la variance est $s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$. Essayons de trouver quelque chose de plus fonctionnel ...

$$\begin{aligned}
s^2 &= \frac{1}{n-1} \cdot \left[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2 \right] && (1) \\
&= \frac{1}{n-1} \cdot \left[\left(x_1^2 + \overline{x}^2 - 2x_1\overline{x}\right) + \left(x_2^2 + \overline{x}^2 - 2x_2\overline{x}\right) + ... + \left(x_n^2 + \overline{x}^2 - 2x_n\overline{x}\right) \right] && (2) \\
&= \frac{1}{n-1} \cdot \left[\left(x_1^2 + x_2^2 + ... + x_n^2\right) + n.\overline{x}^2 - 2\overline{x}\left(x_1 + x_2 + ... + x_n\right) \right] && (3) \\
&= \frac{1}{n-1} \cdot \left[\sum \left(x_i^2\right) + n \left(\frac{\sum x_i}{n} \right)^2 - 2 \left(\frac{\sum x_i}{n} \right) \left(\sum x_i \right) \right] && (4) \\
&= \frac{1}{n-1} \cdot \left[\sum \left(x_i^2\right) + \frac{\left(\sum x_i\right)^2}{n} - \frac{2\left(\sum x_i\right)^2}{n} \right] && (5) \\
&= \frac{1}{n-1} \cdot \left[\sum \left(x_i^2\right) - \frac{\left(\sum x_i\right)^2}{n} \right] && (6) \\
&= \frac{\sum \left(x_i^2\right) - \frac{\left(\sum x_i\right)^2}{n}}{n-1} && (7)
\end{aligned}$$

Cqfd ! Il suffit donc de connaître n et de calculer $\sum x_i$ et $\sum x_i^2$ pour avoir la variance.

*Ces notes de cours de statistiques sont largement inspirées de mes notes prises au cours du professeur A. Albert pour les deuxièmes candidatures en biologie à l'Université de Liège (Belgique). Je ne suis donc pas statisticien et, si vous aviez des remarques, des suggestions ou si vous trouviez des erreurs, vous pouvez m'en faire part à l'adresse suivante : jepoirrier@gmail.com. La dernière version de ces notes se trouvent ici : http ://www.epot.org/notes/.