

Oral presentation

Multivariate extensions of the coefficient of variation with applications

Lixin Zhang, MSc

PhD student

Medical Informatics and Biostatistics
Department of Public Health Sciences
University of Liège, Belgium

The coefficient of variation, $CV = \sigma/|\mu|$, is a widely used parameter to measure the relative variability of a random variable X to its mean. It also allows comparing the reproducibility of measurement techniques/equipments: the lower the CV , the better the analytical precision. Consider a p -dimensional Normal random vector $\mathbf{X}^T = (X_1, \dots, X_p)$ with mean $\boldsymbol{\mu} \neq \mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The problem of extending the univariate CV to the multivariate setting is not merely a theoretical one. In laboratory medicine, there are assay techniques that provide profiles of test measurements (e.g. serum protein electrophoresis) rather than single analytical results. When comparing two or more such techniques, it would seem natural to choose the one yielding the smallest multivariate CV . The talk starts with reviewing the existing methods with some amendments, including the CV derived by Reyment (1960) $CV_{RR} = (|\boldsymbol{\Sigma}|^{1/p}/\boldsymbol{\mu}^T\boldsymbol{\mu})^{1/2}$, by Van Valen (2005) $CV_{VV} = [\text{tr}(\boldsymbol{\Sigma})/\boldsymbol{\mu}^T\boldsymbol{\mu}]^{1/2}$ and by Voinov-Nikulín (1996) $CV_{VN} = (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1/2}$. Then, we focus on introducing a more general definition of the multivariate CV , which writes $CV_m = [\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}/(\boldsymbol{\mu}^T\boldsymbol{\mu})^2]^{1/2}$. The novel method features some interesting theoretical properties which will be discussed. In particular, it is simple and easy to calculate requiring no matrix inversion (unlike most other methods) but only simple quadratic forms. This property is particularly valuable in fields like genomics or proteomics where the matrix $\boldsymbol{\Sigma}$ is singular due to the overdimensionality of the dataset. The various multivariate CV s are illustrated on serum protein electrophoresis data from external quality assessment and on a published microarray dataset.

Key words: Analytical variability; multivariate statistics; dispersion measure; test profiles.

Statistical area: Quality Improvement.

References

- Reyment, R. A. (1960). Studies on Nigerian Upper Cretaceous and Lower Tertiary ostracoda: part 1. Senonian and Maastrichtian ostracoda. *Stockholm Contributions in Geology* **7**, 1-238.
- Van Valen, L. (2005). The statistics of variation. In *Variation: a central concept in biology* (eds B. Hallgrímsson & B. K. Hall), 29-47. Elsevier, Burlington.
- Voinov, V. G. and Nikulin, M. S. (1996). *Unbiased estimators and their applications. Vol. 2, multivariate case*. Kluwer, Dordrecht.