

5th International Conference on

Biometrics & Biostatistics

October 20-21, 2016 Houston, USA

Model-based clustering with certainty estimation: Implication for clade assignment of influenza viruses

Shunpu Zhang

University of Central Florida, USA

Clustering is a common technique used by molecular biologists to group homologous sequences and identify co-expressed genes. There remain issues such as how to cluster molecular sequences accurately and in particular how to evaluate the certainty of a cluster. We presented a model-based clustering method to analyze molecular sequences, described a subset bootstrap scheme to evaluate a certainty of the clusters, and showed an effective way using 3D visualization to examine clusters. The above methods were applied for the clade assignment of influenza viral hemagglutinin (HA) sequences. For the high pathogenic avian influenza (HPAI) A (H5N1) HA sequences, nine clusters were obtained using the model-based method, which agrees with previous findings; the certainties for sequences assigned to a cluster were all 1.0, the certainties for clusters were also very high (0.92-1.0), with an overall clustering certainty of 0.95. For influenza A (H7) HA sequences, 10 HA clusters were assigned and the vast majority of sequences could be assigned to a cluster with a certainty of more than 0.99; the certainties for clusters, however, varied from 0.40 to 0.98. We suspect such certainty variation is attributed to the dissimilar homogeneity of sequence data within cluster. In both cases, the certainty values estimated using the subset bootstrap method are all higher than those calculated based upon the standard bootstrap method, suggesting our bootstrap scheme is more robust for the estimation of clustering certainty.

shunpu.zhang@ucf.edu

A statistical outlook on factors affecting the infant mortality in South Asia

Amal Krishna Dey

Senior Freelance Journalist, India

The success of the lowering infant mortality rate has been a matter of controversy. A common view is that the lowering infant mortality has been a success in some comparatively economically sound areas, but not in the economically weaker areas. The question however needs to be examined statistically with a trend analysis of different regions of South Asia, namely Bangladesh, India, Nepal, Pakistan and Sri Lanka. This paper takes a fresh look at the discussion on the rates of changes of infant mortality rates in the light of the data which has become available up to the end of the year 2015. The trend analysis is carried out mainly with the two types regression based curves ---viz. the exponential and Gompertz and goodness fit of these two trend fitting exercise has been statistically established by the DW (Durbin-Watson) statistic. The exponential, which may be written as $Y_t^* = ab^t$ and the equation for the Gompertz curve as $Y_t^* = Ka^{bt}$ which may be put in logarithmic form $\log Y_t^* = \log K + (\log a)^{bt}$. A correlation matrix also constructed to study the effects of female literacy, GDP rates, states' expenditure on health and family planning etc. The findings of this study are that, the so-called country with better economy notwithstanding, there has been a tendency towards a slowing down of the rate of percentage changes in the infant mortality in the regions with the better economy, whereas the regions with comparative lower economy has shown better results than the regions with better economy in controlling the growth of infant mortality rate and of course, the female literacy rate has a greater say in lowering the infant mortality rate.

amalkdey@gmail.com