popKorn: An R package for Inference on Selected Populations

Vik Gopal^{1,*}, Claudio Fuentes², Andrew Womack³

National University of Singapore
Oregon State University
University of Florida
*Contact author: stavg@nus.edu.sg

Keywords: confidence intervals, selected means, selected populations, asymmetric intervals

Consider an experiment in which p independent populations π_i , with corresponding unknown means θ_i are available and suppose that for every $1 \leq i \leq p$, we can obtain a sample $X_{i1}, X_{i2}, \ldots, X_{in}$ from π_i . In this context, researchers are sometimes interested in selecting the populations that give the largest sample means as a result of the experiment, and to estimate the corresponding population means θ_i 's. In [1], the authors present an approach to the problem and discuss how to construct confidence intervals for the mean of $k \geq 1$ selected populations, assuming the π_i are independent and normally distributed with a common variance σ^2 . The R package **popKorn** implements this approach, which is based on minimisation of the coverage probability.

The **popKorn** package is the next generation of **kPop**, which was presented at useR! 2013 [2]. The new version contains a better implementation of functions for estimating the optimal asymmetric intervals to be used. This provides a practical yet formal tool for estimating (simultaneously) the mean of several selected populations.

In this talk, we shall motivate the problem, introduce this package, and demonstrate how its main functions can be used. We shall compare it with traditional methods, some of which do not account for the selection phase.

References

- [1] Fuentes, C., G. Casella, and M. Wells (2013). Interval estimation for the mean of the selected populations. *Submitted*.
- [2] Gopal, V. and C. Fuentes (2013). **kPop**: An *R* package for the interval estimation of the mean of selected populations. In *useR!* 2013, *The R User Conference* (*Albacete*, *Spain*).