TESTING MEANS OF SEVERAL PARETO DISTRIBUTIONS

Lakhana Watthanacheewakul

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Statistics)
School of Applied Statistics
National Institute of Development Administration
2007

TESTING MEANS OF SEVERAL PARETO DISTRIBUTIONS

Lakhana Watthanacheewakul

School of Applied Statistics

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Statistics).

Professor Prachoom Suwattee .....Committee Chairperson

(Prachoom Suwattee, Ph.D.)

Associate Professor Jirawan Jitthavech .....Committee

(Jirawan Jitthavech, Ph.D.)

Associate Professor Pachitjanut Siripanich .....Committee

(Pachitjanut Siripanich, Ph.D.)

Assistant Professor W. Bodhisuwan .....Committee

(Winai Bodhisuwan, Ph.D.)

Associate Professor S. Auwatanamongkol .....Dean

(Surapong Auwatanamongkol, Ph.D.)

Date Aug 28, 2007

ABSTRACT

Title of Dissertation	Testing Means of Several Pareto Distributions
Author	Miss Lakhana Watthanacheewakul
Degree	Doctor of Philosophy (Statistics)
Year	2007

To test the equality of the means of non-normal data, both parametric and non-parametric tests may be applied. Choosing an appropriate test usually depends upon the sample size and the power of the test. In this dissertation, analysis of variance (ANOVA), the likelihood ratio test and the Kruskal-Wallis test for Pareto populations are investigated. Since Pareto data sets are non-normal distributions, they must be transformed to normal distributions with a constant variance. A proposed transformation to achieve this is in the form

$$Y_{ij} = \begin{cases} \dfrac{X_{ij}^{\lambda} - (X_{ij} - 0.99\theta_i)^{\lambda}}{\lambda}, & \lambda \neq 0 \\ \ln\left(\dfrac{X_{ij}}{X_{ij} - 0.99\theta_i}\right), & \lambda = 0, \end{cases}$$

where X_{ij} is a random variable in the *jth* trial from the *ith* Pareto distribution, Y_{ij} the transformed variable of X_{ij}, θ_i the minimum value of X_{ij} from the *ith* Pareto distribution and λ a transformation parameter, which is developed from the folded-power transformation. Where values of X_{ij} are equal to θ_i, the logarithm of $(X_{ij} - \theta_i)$ does not exist, so $0.99\theta_i$ can be considered as an adjustment. The proposed transformation performs better than the Box-Cox transformation in cases where there is both normality and homogeneity of variances. Moreover, an alternative numerical method was applied to estimate the value of the transformation parameter for a particular set of observed data.

The powers of the ANOVA, likelihood ratio and Kruskal-Wallis tests were compared in a number of different situations and differing sample sizes. It was found that the results depended on the location and shape parameters. The powers of the tests are almost the same when the coefficients of variation of the population means are large. In order to choose the appropriate test, the location parameter and the shape parameter for each group should be checked. In cases of the same location parameter but a different shape parameter, sample size did not affect the choice of test. If the differences among the population means are small, the likelihood ratio test should be selected but where they are large, all three tests can be used. When the location parameter and shape parameter are different, sample sizes are small and the numbers unequal, and the differences among the population means are small, the likelihood ratio test should be selected. In the same scenario but where the differences among the population means are large, any of the tests can be selected. When the sample size is increased and the differences among the population means are small, the ANOVA test was appropriate, but again if the differences among the population means are large, all three tests performed almost identically. For a different location parameter but using the same shape parameter, sample size did not affect the choice of test. If the differences among the population means are small, the Kruskal-Wallis test should be selected whereas large differences among the population means show that both the Kruskal-Wallis test or the likelihood ratio test are suitable. It seems that the likelihood ratio test was a good choice in almost every case, but it is difficult to find a test based upon the likelihood ratio.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this dissertation. I am greatly indebted to Professor Dr. Prachoom Suwattee, my advisor, for his valuable suggestions, patience and encouragement.

I am very grateful to the other committee members, Associate Professor Dr. Jirawan Jitthavech, Associate Professor Dr. Pachitjanut Siripanich and Assistant Professor Dr. Winai Bodhisuwan, who have made many helpful suggestions.

Furthermore, I would like to thank the Agriculture office, Amphoe Sansai, Chiang Mai province for supplying me with the data that helped in the completion of this dissertation.

Finally, I particularly wish to give my special thanks to my parents for their help and support.

Lakhana Watthanacheewakul

June 2007