ROBUST RATIO TYPE ESTIMATORS IN SIMPLE RANDOM SAMPLING USING HUBER M ESTIMATION

M. Subzar*[1] C., N. Bouza**, S. Maqbool*, T A Raja** and B. A. Para***
*Division of Agricultural Statistics, SKUAST-Kashmir (190025), India.
**Facultad de Matemática y Computación, Universidad de La Habana, Cuba
***Post Graduate Department of Statistics, University of Kashmir, Srinagar.

ABSTRACT
In survey sampling a very dominating problem is to obtain the better ratio estimators of the population mean and population variance. A very striking question, in every researcher's mind, is that if there are outliers in the data how to estimate the population mean and variance, as outliers mislead the results. So keeping the above problem under consideration, we propose some new modified ratio estimators using robust regression, which are robust against the outliers and give accurate results even in the presence of the outliers; also the properties are studied. It has been shown that the proposed class of estimators is more efficient than the existing classes of estimators. An empirical study has been carried out to examine the merits of the proposed class of estimators over others.

KEYWORDS: Quartiles, Non-Conventional location parameters, Median, M-Estimation, Bias, Mean Square Error, Efficiency.

MSC: 62D05, 62G35

RESUMEN
En encuestas por muestreo un problema muy dominante es el de obtener el mejor estimador de razón de la media y la varianza. Una pregunta muy inquietante, en la mente de todo investigador, es cuando hay observaciones aberrantes en la data como estimar los parámetros sin conllevar una mala interpretación de los resultados. Con esto en mente, proponemos una nueva clase de estimadores de razón modificados, usando una regresión robusta, cuyos resultados sean adecuados incluso ante la existencia de outliers; también se estudian sus propiedades. Se ha probado que la clase de estimadores propuesta es más eficiente que la de otros estimadores existentes. Un estudio empírico ha sido llevado a cabo para examinar los méritos de esta clase respecto a otras.

PALABRAS CLAVE: Cuartiles, Parámetros Non-Convencionales de Posición, Mediana, M-Estimación, Sesgo, Error Cuadrático Medio, Eficiencia

1. INTRODUCTION

Since from 1960, various theoretical efforts have been devoted to develop statistical procedures that are resistant to small deviations from the assumptions, i.e. robust with respect to outliers and stable with respect to small deviations from the assumed parametric model. In fact, it is well-known that classical optimum procedures behave quite poorly under slight violations of the strict model assumptions.
It is also well-known that to screen the data, to remove outliers and then to apply classical inferential procedures is not a simple and good way to proceed. First of all, in multivariate or highly structured data, it can be difficult to single out outliers or it can be even impossible to identify influential observations. Second, in place of rejecting an observation, it could be better to down-weight uncertain observation, although we may wish to reject completely wrong observations. Moreover, rejecting outliers reduces the sample size, could affect the distribution theory, and variance could be underestimated from the cleaned data. So even one outlying observation can destroy least squares estimation, resulting in parameter estimates that do not provide useful information for the majority of the data. So Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown out. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. The properties of efficiency, breakdown point, and bounded influence are used to define the measure of robust technique performance in a theoretical sense. Efficiency can tell us how well a robust technique performs relative to least squares on clean data (without outliers). High efficiency is mostly desired on estimation. The breakdown point is a

[1] subzarstat@gmail.com