وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم الإحصاء والمعلوماتية

التقـديــر الحصــين لمتــوسط المـجتمع
باستخدام المعاينة الطبقية المرتبة
بـوجود القيــم الشـاذة

رسالة مقدمة

الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل كجزء من
متطلبات نيل شهادة ماجستير علوم في الإحصاء

من قبل

أيمن محمود خـلف احمد

بإشراف

د. ريكان عبد العزيز أحمد

1446هـ 2025م

المستخلص

تضمنت هذه الرسالة تقديم ودراسة مقدرات مقترحة قائمة على الانحدار (المنفصل والمشترك) (Separate and Combined Regression) لتقدير متوسط المجتمع المحدود في اطار المعاينة المرتبة الوسيطية الطبقية (Stratified Median Ranked Set Sampling S_tMRSS) تُقارن هذه المقدرات بمقدرات الانحدار (المنفصل والمشترك) لتقدير متوسط المجتمع في اطار المعاينة العشوائية الطبقية (Stratified Random Sampling S_tRS) باستخدام الأساليب الحصينة لتقدير معلمات الانحدار اضافةً الى المقدرات التقليدية، حيث تمت المقارنة من خلال معيار متوسط مربعات الخطأ (Mean Square Error MSE) الذي يتم حسابه من خلال اجراء المحاكاة للنموذج، بالإضافة الى معيار الكفاءة النسبية (RE) Relative Efficiency . علاوةً على ذلك، تُدرس خصائص وفعالية المقدر المقترح مقارنةً بنظيراتها لإثبات دقتها في تمثيل متوسط المجتمع. حيث تضمن مقدر الانحدار الخطي المنفصل تقدير متوسط المجتمع لكل طبقة بشكل مستقل، ثم حساب المتوسط المرجح لهذه الطبقات، وتكون الأوزان متناسبة مع أحجام الطبقات. في المقابل، يتضمن مقدر الانحدار الخطي المشترك نمذجة العلاقة بين متغير الاستجابة والمتغيرات التفسيرية، مع الأخذ في الاعتبار مؤشر الطبقة، باستخدام نموذج الانحدار الخطي، ثم تقدير متوسط المجتمع باستخدام النموذج المناسب.

أظهرت دراستنا أن استخدام أسلوب المعاينة المرتبة المتوسطة الطبقية (S_tMRSS) في التقدير ساهم بشكل كبير في تحسين جودة تقديرات الانحدار (المنفصل والمشترك) للوسط الحسابي للمجتمع من خلال تقليل تأثير القيم الشاذة. تم ذلك من خلال مقارنة تقديرات هذا الأسلوب من المعاينة مع عدة طرائق للانحدار الحصين، مثل (Huber M, Huber MM, Least Median of Squares Regression LMS, Least Trimmed Square LTS) ومصفوفات التباين والتباين المشترك الحصينة (Minimum Minimum Covariance Determinant MCD Volume Ellipsoid MVE,) في إطار المعاينة العشوائية الطبقية، بهدف الوصول إلى أفضل مقدر عبر إجراء دراسة محاكاة. وقد أظهرت النتائج أن المقدر الناتج من خلال استخدام المعاينة المرتبة المتوسطة الطبقية (S_tMRSS) كان الأفضل في الكثير من الحالات المفترضة وعلى وجه الخصوص عند استخدام أسلوب الانحدار المشترك حيث كان الاكفأ من بين جميع الطرائق

المستخدمة في هذه الرسالة ولجميع أحجام العينات المفترضة (n = 100, 250, 400, 600) ومع النسب المختلفة للقيم الشاذة (10%, 15%, 20%, 25%)، وذلك باستخدام معيار متوسط مربعات الخطأ (MSE) والكفاءة النسبية (RE). وتم تطبيق النتائج النظرية على بيانات حقيقية من مديرية زراعة نينوى لتقدير متوسط غلة القمح المتوقعة لمحافظة نينوى.

Ministry of Higher Education
and Scientific Research

University of Mosul

College of Computer Science and
Mathematics

Department of Statistics and
Informatics

Robust estimation of the population mean using stratified ranked set sampling in the presence of outliers

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master of Science
in
Statistics**

By
Ayman Mahmoud Khalaf Ahmed

Supervised by
Dr. Rikan Abd Al-Aziz Ahmed

2025 A.D. **1446 A.H**

Abstract

This thesis included the presentation and study of proposed estimators based on regression (separate and combined) to estimate the mean of a finite population within the framework of Stratified Median Ranked Set Sampling (S_tMRSS).These estimators are compared with regression estimators (separate and combined) for estimating the population mean within the framework of Stratified Random Sampling (S_tRS) using robust methods for estimating regression parameters in addition to traditional estimators. The comparison is made through the Mean Square Error (MSE) criterion, which is calculated by simulating the model, in addition to the Relative Efficiency (RE) criterion. Moreover, the properties and effectiveness of the proposed estimator are studied in comparison with its counterparts to demonstrate its accuracy in representing the population mean. The separate linear regression estimator ensures the estimation of the community mean for each stratum independently, then calculates the weighted average of these strata, with weights proportional to the stratum sizes. In contrast, the common linear regression estimator involves modeling the relationship between the response variable and the explanatory variables, taking into account the class indicator, using the linear regression model, and then estimating the population mean using the appropriate model.

Our study showed that using the stratified median rank sampling (S_tMRSS) method in estimation significantly contributed to improving the quality of regression estimates (both separate and joint) for the population mean by reducing the impact of outliers. This was achieved by comparing the estimates of this sampling method with several robust regression methods, such as Huber M, Huber MM, Least Median of Squares Regression (LMS), Least Trimmed Square (LTS), and robust covariance and minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) matrices within the framework of stratified random sampling, with the aim of reaching the best estimator through a simulation study. The results showed that the estimator obtained through the use of stratified median rank sampling (S_tMRSS) was the best in many assumed cases, particularly when using the joint regression method, as it was the most efficient among all the methods used in this thesis for all assumed sample sizes (n = 100, 250, 400, 600) and with different proportions of outliers (10%, 15%, 20%, 25%), using the Mean Squared Error (MSE) and Relative Efficiency (RE) criteria. The theoretical results were applied to real data from the Nineveh Agriculture Directorate to estimate the expected average wheat yield for Nineveh Governorate.