Dedication

I dedicate my thesis work to my family and my students.

Acknowledgement

I would take this opportunity to thank my research supervisor Dr. Amin Ibrahim, family and friends (Dr. Mohammed Suleman Gibreel , Dr. Zakria Mohamed Salih) for their support and guidance without them this research would not have been possible.

Abstract

The missing data in household health survey was a problem for the researchers because it leads to incomplete analysis. The statistical tool of cluster analysis methodology was implemented in the collected data of Sudan's household health survey in 2006.

This research focuses specifically on the analysis of the collected data and the objective is to deal with the missing values in cluster analysis. Two-Step Cluster Analysis is applied in which each participant is classified into one of the identified pattern and the optimal number of classes is determined using SPSS Statistics/IBM. Any observation with missing data is excluded in the Cluster Analysis as in the multi-variable statistical techniques. Therefore, before performing the cluster analysis, missing values is imputed using multiple imputations (SPSS Statistics/IBM). The clustering result is displayed in tables. The descriptive statistics and cluster frequencies are produced for the final cluster model, while the information criterion table displayed results for a range of cluster solutions.

Furthermore, the objective is extended to include the reduction of biases arising from the fact that non-respondents may be different from those who participate and to bring sample data up to the dimensions of the target population totals.

المستخلص

لقد شهدت مشكلة البيانات المفقودة اهتماما في السنوات الأخيرة ، ومع التطور السريع لأجهزة الحاسوب والبرمجيات في معالجة العمليات أصبح تطوير طرقتحليل البيانات المفقودة ممكن نظرياً وعلى الرغم من ذلك ما زال العديد منهـــا بحاجة للتطوير ويعاني من مشاكل عديدة.

من هنا تأتي أهمية تسليط الضوء على طرق معالجة القيم المفقودة لإيجاد أفضل الطرق التي تلائم البيانات من ناحية تقدير القيمة المفقودة، أخذين بعين الاعتبار نسب الفقدان و آلية الفقدان ونمط الفقدان.

ان البيانات المفقودة في مسح صحة الأسرة تمثل مشكلة للباحثين لأنه يؤدي إلى تحليل غير كامل. تم تنفيذ أداة إحصائية منهجية التحليل العنقودي في البيانات التي تم جمعها من مسح صحة الأسرة في السودان في عام 2006.

ويركز هذا البحث على وجه التحديد على تحليل البيانات التي تم جمعها والهدف من ذلك هو التعامل مع القيم المفقودة في التحليل العنقودي. تم تطبيق التحليل العنقودي ذو البعدين في كل صنف من النمازج المعرفه وحدد العدد الأمثل من الطبقات باستخدام حزم البرامج الاحصائية spss.

تم استبعاد أي ملاحظات مع البيانات المفقودة في التحليل العنقودي كما في التقنيات الإحصائية متعدده المتغيرات. لذلك، قبل تنفيذ التحليل العنقودي، نعوض القيم المفقودة باستخدام التعويض المتعدد باستخدام (IBM / SPSS).

تم عرض نتائج المجموعات في جداول ، ونتائج إحصاءات وصفية عنقوديه، كما أظهرت النتائج وجود عناقيد ذات جوده ما بين قوي وقوي جدا ولا يوجد عنقود ضعيف.

Table of contents

LIST OF TABLES

LIST OF FIGURES