

المستخلص

تناولت هذه الدراسة مقارنة بين طريقتين من طرق التحليل العنقودي الحاد ، الذي ينحدر من تحليل متعدد المتغيرات وهما طريقة المتوسطات وطريقة التحليل الهرمي حيث تم تناول منهجية عمل كل من الطريقتين والزمن المستغرق لكل طريقة للحصول علي النتائج النهائية للتحليل باعتبار زمن الحصول علي النتائج من العوامل المهمة عند اختيار طرق التحليل، وتتمثل اهمية هذه الدراسة في انها تهتم بقياس ومقارنة طريقة التحليل العنقودي الحاد مع مراعات الاستخدام الامثل للموارد المتاحة في الحواسيب للوصول الي النتائج وتكون هي الاسرع من حيث الزمن المستغرق في التحليل بعد التأكد من صحة كل طريقة في الحصول علي العناقيد، تمثلت فرضيات هذه الدراسة في انها لا تعتبر طريقة المتوسطات افضل في العينات الكبيرة من الطريقة الهرمية من حيث اعطاء النتائج كما لا يوجد فرق معنوي بين متوسطي الزمن المستغرق لتنفيذ كل من الطريقتين ، تم اتباع المنهج الاحصائي الوصفي لوصف متغيرات الدراسة، والمنهج الاحصائي التحليلي للوصول الي النتائج باستخدام حزمة البرمجيات المتوفرة في برنامج (R) و SPSS 20 و MS-Excel ، من أهم النتائج التي توصلت إليها هذه الدراسة أن الطريقة الهرمية لا تصلح لتحليل العينات الكبيرة نسبة لاستهلاكها مدة زمنية طويلة مقارنة بطريقة المتوسطات وكذلك زيادة حجم العينة لا يقلل من كفاءة طريقة المتوسطات ، وطريقة المتوسطات افضل في العينات الكبيرة من طريقة الهرمية من حيث السرعة في اعطاء النتائج والزمن المستغرق كما توصي الدراسة باستخدام طريقة المتوسطات في العينات كبيرة الحجم وعدم استخدام الطريقة الهرمية في حالة العينات كبيرة الحجم لعدم كفاءتها واستخدام طريقة الهرمية اذا اردنا تفصيل واضح لمراحل تكون العناقيد في حالة البيانات صغيرة الحجم وعمل مزيد من المقارنات بين طريقة المتوسطات وطرق التحليل العنقودي لمعرفة ماهي الطريقة الأكفاء.

ABSTRACT

This study compared two methods of hard cluster analysis, which is a part of multi-variables Analysis and the methods is K-means Clustering and Hierarchical clustering where the expense of the time it takes is calculated for each method to obtain the final results of the analysis through the software package, the importance of this study that it is the first study of its kind that of interest in comparing statistical Analysis, taking into consideration the optimal use of the available resources in the computers to reach results that will be the best in terms of time spent in the analysis after making sure of the validity of each method in obtaining clusters, the assumptions of this study are the following, K-means clustering is better than Hierarchical Clustering when the data is considered as big data, There is a significant difference between the means of time it takes for the implementation of both methods , the search followed of the statistical descriptive approach to describe the study variables, the statistical methodology of the analytical work and access to the results of the study using the SPSS 20, R Programming and MS-Excel, some important results of the findings of this study are hierarchical Clustering is not suitable to analyze big data because it take much time to obtain the final results, changing the size of the sample does not reduce the efficiency of k-means Clustering, K-means significantly work better than Hierarchical Clustering in big data, the study recommends that is Better to Use K-means in big data , Better not to use the Hierarchical Clustering in the case of big data , find comparisons between K-means analysis and other clustering algorithms to know what the best qualified Method in Large Sample size is.