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

توظيف التحليل العنقودي والانحدار اللوجستي لدراسة الإصابة بمرض الفشل الكلوي

رسالة مقدمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
وهي جزء من متطلبات نيل شهادة دبلوم عالي في الإحصاء

للطالب

صهيب بشار حميد احمد

بإشراف
م.د. محمود محمد طاهر جادر

١٤٤٦ هـ ٢٠٢٥ م

الملخص

تناولت الدراسة استراتيجية متكاملة تجمع بين التحليل العنقودي والانحدار اللوجستي لتحديد العوامل الرئيسة المؤثرة على مرض الفشل الكلوي. اعتمدت الدراسة على بيانات 142 مريضاً تم جمعها من مستشفى ابن سينا.

أظهرت الدراسة أهمية استخدام التحليل العنقودي لتجميع المتغيرات وتحسين دقة النماذج الإحصائية. وشملت المتغيرات المستقلة التي خضعت للتحليل (العمر ومستويات اليوريا والكرياتين في الدم والبوتاسيوم والالبومين والكالسيوم)، واختبار تأثيرها باستخدام نماذج انحدار لوجستي. وخلصت الدراسة إلى تحديد المتغيرات ذات التأثير الأكبر على احتمالية الإصابة بالفشل الكلوي.

تم تجميع المتغيرات باستخدام التحليل العنقودي إلى أربعة عناقيد كل عنقود يمثل مجموعة متغيرات يتم تحليلها باستخدام الانحدار اللوجستي، أظهرت النتائج بعض المتغيرات تأثيرًا واضحًا على احتمالية الإصابة، وتم استخدام نماذج الانحدار اللوجستي لتفسير وتقدير هذه العلاقة.

خلصت الدراسة إلى نموذج إحصائي لتقدير احتمال الإصابة بالفشل الكلوي استناداً إلى العوامل المحددة، مؤكدةً على أهمية الجمع بين التحليل العنقودي والانحدار اللوجستي لتعزيز الفهم الطبي للأمراض المزمنة ودعم القرارات العلاجية.

Ministry of Higher Education
And Scientific Research
University of Mosul
College of Computer Science and Mathematics
Department of Statistics and Informatics

Application Cluster Analysis and Logistic Regression to Study the Incidence of Renal Failure

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 Higher Diploma in Statistics

By

Suhaib Bashar Hameed Ahmed

Supervised by

Dr. Mahmood Mohammed Taher Jader

2025 A.D. 1446 A.H.

Abstract

The study addressed an integrated strategy that combines cluster analysis and logistic regression to identify the main factors affecting kidney failure. The study relied on data from 142 patients collected from Ibn Sina Hospital.

The study demonstrated the importance of using cluster analysis to group variables and improve the accuracy of statistical models. The independent variables that were analyzed included (age, blood urea and creatinine levels, potassium, albumin, and calcium), and their impact was tested using logistic regression models. The study concluded by identifying the variables with the greatest impact on the likelihood of kidney failure.

The variables were grouped using cluster analysis into four clusters, each cluster representing a set of variables that were analyzed using logistic regression. The results showed that some variables had a clear impact on the likelihood of infection, and logistic regression models were used to explain and estimate this relationship.

The study concluded with a statistical model to estimate the likelihood of kidney failure based on the identified factors, emphasizing the importance of combining cluster analysis and logistic regression to enhance medical understanding of chronic diseases and support therapeutic decisions.