



دانشگاه علوم پزشکی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد رشته آمار زیستی

عنوان:

مقایسه معیارهای مختلف شباهت تحلیل خوشه ای در تشخیص بیماری های کبدی

توسط: مونس بذرافشان

استاد راهنما: دکتر عباس بهرامپور

استاد مشاور: دکتر محمد جواد زاهدی

سال تحصیلی: ۱۳۹۰-۱۳۹۱



**Comparison of distance measures of cluster analysis in diagnosis
of liver diseases**

A Thesis
Presented to
The Graduate Studies

By

Moones Bazrafshan

In Partial Fulfillment
of the Requirements for the Degree
M.Sc.

Biostatistics

Kerman University of Medical Sciences

July 2012

چکیده

مقدمه و هدف: افتراق در تشخیص بیماری‌های کبدی غالباً بر اساس تعداد زیادی از تست‌های آزمایشگاهی صورت می‌گیرد، کبد عضوی پیچیده با عملکردهای زیادی است و متخصصین با ابوهی از داده‌ها جهت تشخیص این بیماری روبرو هستند، این پژوهش می‌تواند با ایجاد الگوهایی بهینه به متخصصین در افتراق و تشخیص بیماری‌های کبدی کمک کند. در این پژوهش جهت برآورد تشخیص بیماری‌های کبدی شامل سیروز کبدی، سرطان کبد، هپاتیت حاد، مزمن و کبد چرب از تحلیل خوشه‌ای استفاده شده است. هدف گروه‌بندی بیماران بر اساس متغیرهای آزمایشگاهی در بیماری‌های مربوطه است که برای گروه‌بندی تحلیل خوشه‌ای را بکار برده شده است. تحلیل خوشه‌ای یکی از روش‌های چندمتغیره است که گروه‌هایی از مشاهدات را که بصورت مشابه رفتار می‌کنند در گروه‌های همگن و متمایز از سایر گروه‌ها قرار می‌دهد. تحلیل خوشه‌ای شامل دو مرحله، تعیین فاصله و تعیین روش خوشه‌بندی است. در ابتدا بایستی نوع فاصله و همچنین روش خوشه‌بندی را تعیین کرد و سپس جهت تعیین مهمترین متغیرها در هر خوشه از تحلیل عاملی استفاده شده است.

مواد و روش‌ها: در این پژوهش ۴۶۵ بیمار مورد بررسی قرار گرفته‌اند، جهت ثبت مشاهدات از چک لیست استفاده شده و داده‌های گمشده بعضی از متغیرها با روش "الگوریتم برآورد داده‌ها گمشده" (EM)، برآورد شدند و با روش تحلیل خوشه‌ای و تحلیل عاملی مورد تجزیه و تحلیل قرار گرفتند. تحلیل خوشه‌ای با استفاده از نرم افزار SPSS و Stata و تحلیل عاملی با استفاده از نرم افزار Stata انجام گردید. اولین گام در تحلیل خوشه‌ای تعیین نوع اندازه‌گیری فاصله است زیرا خوشه‌بندی مشاهدات بر مبنای فاصله بین مشاهدات انجام می‌شود در این پایان نامه فاصله اقلیدسی، مربع اقلیدسی، چیشف، مینکوسکی و کسینوسی مورد بررسی قرار گرفته، سپس نوع روش خوشه‌بندی سلسله‌مراتبی (پیوند تکی، کامل، متوسط درون گروهی و بین گروهی، مرکزی، میانه و ward) و یا غیر سلسله‌مراتبی (k-میانه و k-میانگین) را بر اساس هفت حالت مختلف برای k، نیز مشخص کرده و هر دوی این روش‌ها مورد بررسی بوده است.

اهمیت هریک از متغیرها در تشخیص این بیماری‌ها با استفاده از تحلیل عاملی مشخص گردید، تحت مدل عاملی هر متغیر پاسخ بصورت ترکیبی خطی از متغیرهای عاملی مشترک غیر قابل مشاهده در می‌آید، در تحلیل عاملی با فرض داشتن تعداد معینی از عامل‌های غیر قابل مشاهده به دنبال یافتن ضرایب مربوط به ترکیبی خطی از پاسخ‌ها هستیم. در این روش در ابتدا در هر خوشه تحلیل عاملی را بصورت مجزا انجام دادیم و سپس بر اساس ضرایب مربوط به متغیرها، مهمترین آن‌ها در هر خوشه تعیین شد.

یافته ها: بر اساس نتایج ۱۰۵ حالت طبقه بندی (۲۱ روش با ۵ تابع فاصله) نشان داده شد که روش غیر سلسله مراتبی k-میانه با فاصله اقلیدسی و مربع اقلیدسی بهینه می باشد و با استفاده از روش تحلیل خوشه ای، خوشه هایی جهت غربال مطابق با بیماری های سیروز کبدی، سرطان کبد، هپاتیت حاد و مزمن و کبد چرب بدست آمد بنابر این جهت تشخیص اولیه بیماری های کبدی استفاده از نتایج به دست آمده از این روش توصیه می شود که در هر بک دامنه و شاخص های پارامترهای شیمیایی مورد نیاز جهت تشخیص اولیه مشخص گردیده است.

نتیجه گیری: در این مطالعه بیماری هایی از جمله کبد چرب و سیروز کبدی به خوبی افتراق و در دو خوشه متفاوت قرار گرفتند در صورتی که در مطالعات مشابه این افتراق به خوبی انجام نشده است و در نتیجه با توجه به این که تعداد افراد در بیماری های مختلف با هم برابر نیست و این روش به افراد اجازه جابه جاشدن از یک خوشه به خوشه دیگر را می دهد و این مزیت روش انجام شده در این پژوهش در مقایسه با سایر مقالات مشابه می باشد و می تواند افتراق بین بیماری ها را بهتر از سایر روش هایی که تا کنون انجام شده است تعیین نماید.

واژه های کلیدی: تحلیل خوشه ای، توابع فاصله، بیماری های کبدی، تحلیل عاملی

Abstract

Background and Objectives: Differentiation of liver disease is based on a number of laboratory tests. Liver is a complex organ, and specialist are faced with a large number of data for diagnosis, this research can be improved by creating templates to help specialist in the differentiation of liver disease. For primary diagnosing of liver diseases which are cirrhosis, cancer, fatty liver, chronic and acute hepatitis, cluster analysis is applied. Chemical parameters in each group are also determined.

Methods: A total of 465 patients with a proven diagnosis of liver disease were selected. Check list is used for collecting data. The data are analyzed by cluster analysis with different distance measures and factor analysis.

Results: We found out that the Euclidian and Square Euclidian in K-median method is optimal. Clusters according to these diseases were obtained, so this method for primary diagnosis is recommended.

Conclusion: In this study cirrhosis and fatty liver are well differentiated in two different clusters and this has not done well in previous studies. Since the frequencies are different in those diseases and this method allows people to move from one cluster to another, so the applied analysis is recommended.

Key words: Cluster analysis, Distance functions, Liver disease, Laboratory parameters

