بسم الله الرحمن الرحيم
جامعة السودان للعلوم والتكنولوجيا
كلية الدراسات العليا

تحديد العوامل المؤثرة في مرض السكري بإستخدام طرائق متعددة المتغيرات

Determination Of Factors Affecting the Diabetics by Using Multivariate Methods

(دراسة حالة ولاية شمال كردفان)

بحث لنيل درجة الدكتوراة في الاحصاء

إعداد الدارس: محمد بكري عبيد محمد

إشراف الدكتور: أحمد محمد عبدالله حمدي

أغسطس2015 ـــــــــــ ذو القعدة 1436

الفصل الأول

خطة البحث

1-0 : تمهيد

إن استخدام التحليل الإحصائي في البحوث العلمية للظواهر يهدف إلي التوصل إلي تفسير منطقي للظواهر الاجتماعية والطبيعية وتفسيره ببيانات عنها كدراستها وتحليلها ويعتبر التحليل الإحصائي جزاً هاماً من طرق البحث العلمي التي يكثر تطبيقها عند دراسة مشاكل الاقتصاد والعلوم الصحية والطبيعية والاجتماعية ... الخ . وأحدي طرق التحليل الإحصائي في مجال دراسة الظواهر المختلفة ، التحليل التمييزي أو التصنيف وهو احد أساليب التحليل الإحصائي متعددة المتغيرات التي تهتم بدراسة وتحليل المتغيرات لمجتمع أو عينة البحث .

استخدمت هذه الدراسة أسلوب التحليل التمييزي بهدف صياغة نموذج رياضي يمكن من خلاله تصنيف الأشخاص إلى أشخاص مصابين بمرض السكري وآخرين غير مصابين بمرض السكري وذلك نتيجة لما تمثله .

كذلك أسلوب التحليل العاملي هو أحد وسائل متعددة المتغيرات يهدف الى د راسة مجموعة من المتغيرات المرتبطة فيما بينها و تحويلها الى مجموعة اقل من العوامل غير المرتبطة ،و جاء الانتشار الواسع للتحليل العاملي ضمن الواقع العملي في الدرا ستي العلمية والانسانية بناءاً على ما يتمتع به هذا التحليل من خصائص ومميزات ساعدت على انتشاره ، فهو من ناحية يقلص عدد المتغيرات المدروسة الى عدد اقل من العوامل المؤثرة على الظاهرة المدروسة و من الناحية الاخرى يعد أحد الوسائل أو الحلول المتبعة في التخلص من مشكلة التعدد الخطي ، في الواقع العملي فان دراسة اي ظاهرة يتطلب استقراءاً للمتغيرات التي تؤثر فيها و قد تكون هذه المتغيرات ذات وحدات قياس مختلفة و هذا الاختلاف قد يقف في بعض الاحيان حائلاً دون الوصول الى النتائج المرجوة ،التحليل العاملي و لسبب اعتماده على مصفوفة الارتباط وليس على مصفوفة التباين المشترك فانه يخلصنا من مشكلة وحدات القياس للمتغيرات و الوصول الى وصف دقيق للعلاقات بين المتغيرات المدروسة.

انطلاقاً من هذه المفاهيم حول التحليل العاملي,والتحليل التميزي والتحليل العنقودي تمت د راسة و تطبيق هذه الوسيلة الاحصائية لتحديد مرض السكري والعوامل المؤثرة عليه في ولاية شمال كردفان من اجل الحصول على مؤشرات تساعدنا في وصف وتحديد اهم المتغيرات المؤثرة في الإصابة بهذا المرض و بالتالي الخروج باستنتاجات تساعد في التخفيف من وطأة هذه المشكلة .

1- 1 مشكلة الدراسة :

رغم أهمية التحليل أحادي المتغير وثنائي المتغير إلا ان قصور هذه الطرق عنما يتعلق الأمر بعدد كبير من المتغيرات أدى إلى اللجوء إلى نوع آخر من التحليل هو التحليل متعدد المتغيرات وهي تلك الأساليب الإحصائية التي تهتم بدراسة متغيرات متعددة أو مجموعة من المتغيرات في وقت واحد .

كما شهدت السنوات الأخيرة زيادة ملحوظة في أعداد المصابين بمرض السكري ، حتى أننا لا نكاد نرى بيت يخلو من مريض بالسكري ، خاصة وأن هذا المرض تصاحبه بعض المضاعفات الخطيرة على القلب والشرايين والأعصاب والكلى والعين . وهذا ما دفعنا للتفتيش عن الأسباب الرئيسية وراء مرض السكري ، وأهم العوامل التي تؤثر في الإصابة بهذا المرض واهم المضاعفات التى تنتج جراء الإصابة به وخاصة إننا يمكن إن نستخدم طريقة تحليل متعددة المتغيرات لإيجاد دالة تمييزية تساعد الأطباء في تشخيص المرض ويمكن الاعتماد عليها في تصنيف الأشخاص إلي مصابين أو غير مصابين بالمرض.كما يمكننا أيضاً إستخدام التحليل العاملي لتحديد اهم عوامل الإصابة بهذا المرض .

وعلى اساس ما تقدم فإنه يمكننا صياغة مشكلة الدراسة في التساؤلات التالية :

ماذا نعني بالتحليل متعدد المتغيرات ؟

وماهي أهم أساليبه ؟

وماهي أوجه التشابه والإختلاف بين هذه الأساليب ؟

وهل هناك أفضلية بين هذه الأساليب أم تعتبر مكملة لبعض ؟

وماذا نعني بالدالة التميزية وهل فعلاً لها القدرة في تصنيف الأشخاص إلى مصابين وغير مصابين بمرض السكري؟

وما هو التحليل العاملي وهل يشكل أداة فعالة بالنسبة للباحث الإحصائي ؟ واذا كان الأمر كذلك وما هي شروط إستخدام هذا الاسلوب لفهم طبيعة ومحتوى البيانات التي نحن بصدد دراستها .

وهل العناقيد المحددة منسجمة مع المعلومات السابقة للمشكلة ؟

وهل العناقيد المحددة متطابقة مع البيانات جيداً ؟

وهل توجد تجزئة جيدة لمجموعة البيانات ؟

وهل النتائج المستخرجة بطريقة ما أفضل من تلك المستخرجة بالطريقة العنقودية ؟

وما هو العدد الصحيح من العناقيد ؟

وما المقصود بمرض السكري ؟.....

وكيف أعرف أنني مصاب بمرض السكري أم لا ؟.....

ما الجديد في تشخيص هذا اللص الخطير ؟.....

ما السبب وراء الإصابة بمرض السكر ي؟.....

ما السر وراء زيادة انتشار هذا المرض في السنوات الأخيرة ؟

وما أحدث الأبحاث حول وسائل علاج مرض السكري ؟

وهل للغذاء دور في علاج داء السكري ؟

هذه الأسئلة وغيرها الكثير مما يدور في ذهن العديد من الباحثين و القراء ومرضى السكري ، سوف نجيب عليها بمشيئة الله تعالى في هذا البحث بشيء من التفصيل .

1-2 أهمية البحث

تستمد هذه الدراسة أهميتها من ما يلي :

1. قلة الدراسات في هذا المجال من الناحية النظرية والتطبيقية ، وهذا المجال لا يزال بحاجة إلى المزيد من الدراسات ، تسليط الضوء على هذا الموضوع من خلال مناقشته وإظهاره للقاري بصورة مبسطة مع مناقشة المشكلات التي تواجه الباحث أثناء استخدامه.
2. كما تتمثل في توضيح كيفية تحليل وتكوين الدوال التمييزية لبيانات متعددة المتغيرات باستخدام التحليل التمييزي ، وكذلك تسلط هذه الدراسة الضوء على جوانب التشابه والاختلاف بين استخدام التحليل التمييزي والتحليل العاملي والتحليل العنقودي ولفت أنظار الباحثين المتخصصين إلى الاهتمام بهذا الموضوع ذو الأهميمة بمكان في تحليل البيانات عديدة المتغيرات .
3. المعرفة والتخطيط الذي يستند إلى المعلومة الإحصائية والتحليل الإحصائي وذلك من خلال استخدام التحليل التميزي والتحليل العاملي و التحليل العنقودي والتي تتناول المتغيرات الآتية :

(العمر ، المهنة ، التدخين ، الكحول ،الرياضة ، ضغط الدم، تنسيب كتلة الجسم ، العيون ، القلب ، امراض المستوى التعليمي ، درجة القرابة، حالة التغذية ، امراض وراثية بامراض العصبى ، الإصابة الكلى ، الجهاز امراض اخري ، بامراض الإصابة الاطراف ، مضاعفات) فى بالرجل ، بتر باليد، جروح جروح العائلة ، فى اخرى والتي تساعد الأطباء في معرفة مدي الإصابة بمرض السكري

4. ندرة البحوث المختصة بمرض السكري في المكتبات السودانية في حدود علم الباحث .

5. أهمية استيعاب برامج التحليل الإحصائي وخاصة التحليل التميزي والتحليل العاملي والتحليل العنقودي ضمن مقررات العملي في مؤسسات التعليم العالي والبحث العلمي.
6. محاولة لبناء نموذج أو عدة نماذج يمكن بواسطتها القيام بعملية التشخيص الأولية مما يساعد الطبيب المعالج .

1- 3 أهداف البحث-:

1. توضيح مفهوم التحليل الإحصائي متعدد المتغيرات وعلى وجه الخصوص أسلوب التحليل التمييزي والتحليل العاملي والتحليل العنقودي ، وتحليل البيانات المتعددة المتغيرات ،حيث أن موضوع الدراسة الحالية لا يزال يحتاج للمزيد من الإيضاح للباحثين المهتمين بهذا المجال .
2. يهدف البحث الى تحليل و تحديد اهم العوامل المؤثرة على الإصابة بمرض السكري وأهم مضاعفات مرض السكري، بأستخدام اساليب متعددة المتغيرات وهي التحليل التميزي والتحليل العاملي والتحليل العنقودي (Factor & Discrimination & Cluster Analysis)

بغية الوصول الى مؤشرات احصائية تساعد المعنيين في ايجاد تفسيرات ملائمة لها و اسبابها و بالتالي الخروج باستنتاجات و توصيات تساعد في اِتخاذ القرار الملائم بشأنها من قبل أصحاب القرار.

3. سد النقص الذي تعاني منه مكتبتنا العربية في مجال متعددة المتغيرات .
4. محاولة إلغاء الضوء على أساليب وتقنيات التحليل متعدد المتغيرات ، كي تصبح ملائمة للبيانات والمعلومات المتوافرة عن ظاهرة بإعتبارها أساليب تحليل متقدمة.
5. الإبقاء على طبيعة البيانات الأصلية مع تبسيطها لعدد أقل يمكن إستخدامها في تحليلات تالية .
6. تحديد مدى وجود إختلافات معنوية بين المجموعتين (المصابين وغير المصابين بمرض السكري) بالنسبة لمجموعة من المتغيرات المستقلة .
7. تحديد المتغيرات المستقلة التي تساهم بأكبر قدر في تعظيم الفرق بين المجموعتين .
8. إتباع إجراء لتقسيم الأفراد إلى مجموعتين بناءاً على قيم المتغيرات المستقلة .
9. تصميم وتحديد عدد الأبعاد التي تميز بين المجموعتين والمكونة من المتغيرات المستقلة
10. تقسيم الأفراد إلى مجموعتين أو اكثر بناءاً على مدى تشابههم بالنسبة لبعض الخصائص .

11. لفت إنتباه الباحثين الإحصائيين إلى فاعلية هذه الاساليب المتقدمة في تحليل البيانات وإستخراج النتائج .
12. لفت إنتباه المؤسسات التعليمية وبالأخص كليات العلوم إلى هذه الاساليب حيث تعتمد بصورة أساسية على الرياضيات ومن ثم ضرورة الإهتمام بها.
13. ومن الأهداف الرئيسة للدراسة استخدام طرق التحليل العاملي والتميزي باعتبارهم أساليب إحصائية حديثة على بيانات ميدانية تم تجميعها باستبيان خاص.
14. ومن الأهداف الهامة إبراز أهمية التحليل الإحصائي متعدد المتغيرات في مجالات العلوم الطبية .
15. استخدام الدالة التمييزية بهدف صياغة نموذج رياضي يمكن من خلاله تصنيف الأشخاص إلى أشخاص مصابين بمرض السكري وآخرين غير مصابين بمرض السكري ذلك اعتماداً علي المتغيرات(أهم عوامل الإصابة بهذا المرض وأهم مضاعفاته)، ومعرفة مدي تأثير هذه المتغيرات على مستوي الإصابة بمرض السكري .
16. الإجابة عن الاسئلة العلمية للأسر عن مرض السكري ومضاعفاته ومعرفة الأمراض المترتبة عليه وأسبابها وكيفية علاجها .

1-4 فرضيات الدراسة :

يمكننا صياغة الفرضيات التالية :

1. المتغيرات المستخدمة كلها معنوية
2. معادلة الدالة التمييزية معنوية أي قادرة على تصنيف الأشخاص إلى أشخاص مصابين بمرض السكري وآخرين غير مصابين بمرض السكري
3. يمكن التنبؤ بالحالات المرضية باستخدام الدالة التميزية
4. نسبة الخطأ في عملية التمييز صغيرة جداً.
5. التحليل العاملي يمكننا من تخفيف مجموعة المتغيرات الى عدد قليل قابل للإدارة مع المحافظة على المعلومات الأصلية .
6. العناقيد المحددة متطابقة مع البيانات جيداً

1-5 منهجية البحث

تتمثل منهجية البحث في الأسلوب التحليلي والإستنتاجي والوصفي الذي يعتمد على تطبيق أهم أساليب متعددة المتغيرات ومنها التحليل التميزي والتحليل العاملي والتحليل

العنقودي للوصول إلى أهم المتغيرات التي تؤثر في حالات الاصابة بمرض السكري واهم المضاعفات التي يتركها هذا المرض.

واعتمد الباحث على بيانات الإستبيان حيث تم اخذ عينة عشوائية من الأشخاص المصابين بمرض السكري وغير المصابين بمستشفي الأبيض التعليمي ومستشفى الشرطة ومركز ود الياس لمرضى السكري وقد بلغ حجم العينة 216 مريض منهم 162 مصاب بالمرض و54 غير مصابين بالمرض.

وكذلك يمكننا تطبيق أسلوب التحليل التمييزي للوصول إلى دالة تمييزية والتي يمكن أن تستخدم في عملية التصنيف كل ذلك بإستخدام برنامج SPSS 16 وبرنامج MINTAB 16 .

1-6 حدود البحث

حدود البحث الزمنية : 2015

حدود البحث المكانية :مدينة الأُبيض بولاية شمال كردفان

1-7 الدراسات السابقة

1 -في عام 2012، قام الباحثين عبدالسلام الامين تاور ، عبدالماجد عبدالرحيم ، هيثم اسماعيل ، برداسة بعنوان دراسة احصائية لاستخدام متعدد المتغيرات لتجميع المستشفيات في ولاية الخرطوم (القطاع العام 2010) ومن اهم النتائج :

أ- بصورة عامة ومن خلال مقارنة التحليل الذي تم التوصل اليه بالطرق المختلفة للمستشفيات تبين ان هنالك عدم تماثل في المجموعات.

ب- عند التحليل بالطريقة (k) للمتوسطات تم تقسيم المفردات الي اربعة مجموعات وكانت المفردات غير المتجانسة للتوزيع ، بينما عند التحليل بالطريقة الهرمية تم تقسيم المفردات الي ثلاثة مجموعات وكانت المفردات اكثر تجانسا (اي ان التحليل بالطريقة الهرمية اكثر تجانسا من التحليل بطريقة (k) للمتوسطات بالنسبة للمستشفيات)

2 -في عام 2012 قام الباحث فيصل ناجي نامق بدراسة رعنوان ، دراسة تحليلية مقارنة للاعوام 2006، 2007, 2008, لتصنيف محافظات العراق وفقاً لإصابات مرض الكبد الفيروسي بإستخدام التحليل العنقودي، ومن أهم النتائج التي توصل إليها:

3 - أن تقارب وتباعد قيم المعاملات في جدول التقارب بمسافات كبيرة عن بعضها يحدد عدد العناقيد التي تتكون وذلك تبعاً لمقدار القفز من هنا نستنتج انه اذا كانت هذه المسافة قريبة من بعضها فلا توجد فرصة لتكوين العناقيد اذا حدثت قفزة كبيرة في قيم هذه المعاملات.

4 -ب- ان المراحل اللاحقة لتكوين عقدة جديدة في السنوات الثلاثة قيد الدراسة كانت متشابهة وانحصرت بين محافظتي (ذي قار وميسار) من هنا نستنتج بأن التجانس بعدد الاصابات في كلا المحافظتين متشابه وتعاني نفس المشاكل بأعداد المصابيين بهذا المرض.

ج- التصنيف بإستخدام طريقة التحليل العنقودي كانت قد صنفت محافظات (دهوك ، اربيل ، بغداد الكرخ ، بغداد الرصافة ، ذي قار) ضمن العنقود الاول لسنتي 2006، 2007 في حين صنفت محافظات (نينوي والمتني) ضمن العنقود الثاني ولثلاث سنوات 2006، 2007، 2008 ومن هذا نستنتج بأن التصنيف يتم وفقاً لعوامل مشتركة بين المحافظات.

5 في العام 2008 قام الباحث فيصل ناجي نامق بدراسة بعنوان اسلوب التحليل العنقودي لتصنيف الانفاق على السلع والخدمات الاساسية وفقاً للمستوي البيئي (حضر وريف) ، ومن أهم النتائج التي توصل إليها:

أقل فرق بين الحالات (السنوات) للسلسلة الزمنية قيد الدراسة بين 2002 وسنة 2007 حيث بلغ (0.231) وهذا يعني ان هاتين الفترتين كانتا أكثر تجانسا في الانفاق على المجاميع السلعية مقارنة بالفترات الزمنية الاخرى.

ب- الانفاق على المجاميع السلعية في الريف سجل أعلى علاقة بين المجموعة الثامنة (نفقات العناية الطبية والخدماتلاالصحية) والمجموعة التاسعة (السلع والخدمات المتنوعة) من هذا نستنتج بأن السلوك الاستهلاكي للريف يتجه نحو النواحي الكمالية.

ج- أقل فرق بين الحالات (السنوات) بالنسبة للحضر كان بين 2002 وسنة 2007 ايضا حيث بلغ (0.509) وهذا يشير إلي تحقق التجانس في الانفاق في هاتين الفترتين الزمنيتين.

6 في العام 2000 قدم الباحث (عبد المجيد أحمد محمد) بحثاً بعنوان (شروط ومعايير إستخدام التحليل العاملي – دراسة إحصائية تطبيقة – جامعة ام القرى – مكة المكرمة) كجزء من متطلبات نيل درجة الماجستير .

قام الباحث بدراة مدى صحة إستخدام التحليل العاملي من قبل الباحثين في كلية التربية بجامعة أم القرى ومقارنة هذا الإستخدام بإستخدام الطالبات في بعض كليات التربية التابعة لرئاسة تعليم البنات وذلك في دراسة إستطلاعية بناءً على نتائجها تم توضيح كيفية التأكد من الشروط والمعايير الصحيحة الواجب توفرها في البيانات عند إستخدام التحليل العاملي حيث تم عمل دراسة تطبيقية بإستخدام قائمة سمات الشخصية المبتكرة

الذي أعده (سيد خير الله ، 1990) على عينة إستطلاعية مكونة من 30 طالباً لحساب الصدق والثبات وعلى عينة مكونة من 100 طالب من طلاب الصف الأول الثانوي بثانوية جبل النور بمكة المكرمة ، وحللت البيانات عاملياً ، ويمكن تلخيص أهم النتائج كما يلي :

أ – لم يستخدم أسلوب التحليل العاملي بشكل مناسب من قبل باحثي كلية التربية بجامعة أم القرى وقد تركزت عيوب الإستخدام في عدم التحقق من الشروط الواجب توفرها في مصفوفة معاملات الإرتباط ونوع تدوير المحاور وعينة الدراسة .

ب – تم إستخراج أربعة عوامل تؤثر في الشخصية المبتكرة وهذه العوامل هي :

- عامل الثقة في النفس والتحمل .
- عامل الإستقلالية في التفكير والعمل .
- عامل خصوبة الخيال .
- عامل المثابرة .

7 –في العام 1999 م قدم (بدر محمد) بحثاً بعنوان (أسلوب التحليل العاملي : عرض منهجي نقدي لعينة من الدراسات العربية أستخدمت التحليل العاملي) بحث مقدم بندوة البحث العلمي في المجالات الإجتماعية في الوطن العربي .

وأهم النتائج التي توصل إليها :

أ – إختيار ثلاثة متغيرات على الأقل لكل عامل متوقع .

ب –يجب أن يكون العامل قابل لإعادة الإستخراج أو التكرار عبر عينات ومجتمعات وثقفات مختلفة .

ت –يجب ان تتسم النظرية التي تكمن وراء هذا المفهوم بالشمولية بحيث تحتوي على بيانات متعددة وشاملة عن السلوك الإنساني وتجميعها خلال سنوات عديدة .

ث –إستخدام طرق بحث عديدة مثل (الإستخبارات ، المقابلات المقننة ، المقاييس الفيزيولوجية ، الملاحظات .... الخ) لتقويم العامل وقياسه وذلك للوصول إلى مستوى التنظير .

8 -دراسة " أحمد عبد الخالق ، ابراهيم عبد الستار ، ميشيل بوديك "

(Abdel-Khalek, Ibrahin & Budek, 1986)

التى أجريت على عينة مصرية قوامها (316) من طلاب جامعة الاسكندرية بواقع (164) طالب و(152) طالبة ، طبق عليهم استخبار عوامل الشخصية لستة عشر (PF - 16) واستخبار أيزنك للشخصية (EPQ) . أجرى تحليل عاملي بطريقة المكونات الأساسية ، أعقبه تدوير متعامد للمحاور بالفاريماكس للمقياسين ، وقد كشف هذا الإجراء عن استخلاص ثلاثة عوامل (الانبساط ، العصابية ، الذهانية) مما يؤكد عمومية الابعاد الثلاثة الاساسية للشخصية . على حين عندما أجرى نفس التحليل العاملى على مقياس (16-PF) وقد أسفر التحليل العاملى على استخلاص عاملين (الانبساط والعصابية).

والرأى لدى كاتب هذه السطور أن هذه النتيجة تعد بوجه عام - مقبولة . ولكن هذه الدراسة تنقد بمايلى:

1 - لاستخبار عوامل الشخصية الستة عشر (16-PF) خواص سيكومترية معقدة ، وكذلك جوانب نقد شديد (انظر : أحمد عبد الخالق ، 1993) .

2 - انخفاض معاملات ثبات عدد من المقاييس الفرعية المستخدمة فى استخبار عوامل الشخصية الستة عشر ، ومثل هذه المتغيرات غير مرغوبة فى التحليل العاملي بوجه عام.

3 - يجب إجراء تحليل مائل للمقاييس .

4- صغر حجم العينة .

وتعد هذه الدراسة تكملة لسلسلة من البحوث التى أجراها " أحمد عبد الخالق " فى مجال الشخصية.

1-9: هيكلة البحث

يتكون هذا البحث من خمس فصول الأول منها يحتوي على خطة البحث وهي (التمهيد - المشكلة - الاهمية - الأهداف - الفرضيات - المنهجية - الدراسات السابقة - هيكل البحث)

والفصل الثاني يشتمل على (التحليل التميزي - التحليل العاملي - التحليل العنقودي).

أما الفصل الثالث يحتوي على (مرض السكري - الأعراض - المسببات - المضاعفات - العلاقة بالأمراض الاخرى)

بينما إشتمل الفصل الرابع على (التطبيق على البيانات بواسطة التحليل التميزي والعاملي والعنقودي) .

وأخيراً الفصل الخامس إحتوى على (النتائج والتوصيات)

الفصل الثاني

الجانب النظري

2-0 تمهيد

رغم أهمية التحليل أحادي المتغير وثنائي المتغير إلا ان قصور هذه الطرق عنما يتعلق الأمر بعدد كبير من المتغيرات أدى إلى اللجوء إلى نوع آخر من التحليل هو التحليل متعدد المتغيرات وهي تلك الأساليب الإحصائية التي تهتم بدراسة متغيرات متعددة أو مجموعة من المتغيرات في وقت واحد .

ومن أهمها:أسلوب التحليل التميزي وأسلوب التحليل العاملي و أسلوب التحليل العنقودي نتناولها في هذا الفصل بشئ من التفصيل كما يلي :

أولاً: التحليل التمييزي

2-1 : الدالة التمييزية Discriminate function

الدالة التمييزية "التحليل التمييزي" أو التصنيفي يهتم بكيفية التمييز بين مجموعتين أو أكثر من الأفراد أو الأشياء فالطبيب مثلاً قد يرغب في تصنيف الأطفال إلى أطفال مصابين وغير مصابين بناءا على بعض المعايير مثل الوزن ، الطول ، والصوديوم ، وهكذا ...

ويعتبر التحليل التمييزي والتصنيف من أساليب تحليل المتغيرات المتعددة التي تهتم بفصل مجموعات مختلفة من المفردات (المشاهدات) وبتوزيع المفردات (المشاهدات) الجديدة علي مجموعات سبق تعريفها .

ويعتبر التحليل التمييزي استكشافيا بطبيعته كوسيلة للفصل يستخدم التحليل التمييزي لاكتشاف أسباب الاختلاف المشاهدة عندما لا نستطيع فهم العلاقات السببية بدرجة كافية الدقة .

أما طرق التصنيف فإنها أقل استكشافية بمعني أنها تؤدي إلي قواعد معرفة تعريفاً دقيقاً بحيث يمكن استخدامها لتصنيف المشاهدة الجديدة ، ويتطلب التصنيف عادة تحديداً أكثر لبنية المشكلة وذلك الذي يتطلبه التمييز ، إن أول من أستخدم كلمة تمييز هو العالم الإحصائي فيشر عند أول معالجة حديثة لمشاكل الفصل والتمييز [1].

2-2 مفهوم الدالة التمييزية:

(1) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

عبارة عن تقنية تستخدم لبناء نموذج للتنبوء بتصنيف عضوية مشاهدة ما إلي مجموعة معينة بناء علي عدة متغيرات ، يمكن للدالة التمييزية تصنيف مجموعة من الأشخاص إلي مجموعتين رئيستين أشخاص (مصابين – غير مصابين) علي أساس ، تولد الدالة التمييزية من عينة معروف مسبقاً انتماء أي مفردة ، أول مشاهدة فيها إلي احدي المجموعات المراد التمييز بينها ، فتعمل الدالة على تصنيف مشاهدات أو مفردات معلومة المتغيرات مجهولة التصنيف .

وللدالة التمييزية عدة أهداف منها :

- تصنيف المشاهدات ضمن مجموعات مختلفة.
- تحديد ابسط طريقة للتمييز من المجموعات
- التحقق من الفرق داخل المجموعات وبين المجموعات
- إيجاد نسبة التباين في المتغيرات المستقلة في تصنيف المجموعات
- إبعاد المتغيرات التي ليس لها تأثير في نسب المجموعات

3-2 حساب الدالة التمييزية [2]:

إن المشكلة الإحصائية تكمن هنا في كيفية إ يجاد دالة تمييزية وفقاً للمعايير أو القياسات التي يمكن الحصول عليها من الأفراد الجدد (المجهولين) الانتماء إلي المجموعة الصحيحة وسيتم هنا تناول نموذج التشخيص التمييزي الخطي بين مجموعتين فقط .

خطوات حساب الدالة التمييزية

يتم حساب الدالة التمييزية حسب الخطوات الآتي :

أولا :

1. نوجد متوسط كل متغير في كل مجموعة ثم توجد الفرق بين متوسطي كل متغير من المجموعتين .

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1-2)$$

متوسطات المجموعة الأولي[3]

([2]) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

$$\bar{x}_i(1) = \begin{vmatrix} \bar{x}_{11} \\ \bar{x}_{21} \\ \bar{x}_{31} \\ . \\ . \\ . \\ \bar{x}_{n1} \end{vmatrix}$$

متوسطات المجموعة الثانية

$$\bar{x}_i(2) = \begin{vmatrix} \bar{x}_{12} \\ \bar{x}_{22} \\ \bar{x}_{32} \\ . \\ . \\ . \\ \bar{x}_{n2} \end{vmatrix}$$

الفرق ما بين متوسطات المجموعتين

$$d = \bar{x}_i\ (1) - \bar{x}_i\ (2) = \begin{vmatrix} \bar{x}_{11} - \bar{x}_{12} \\ \bar{x}_{21} - \bar{x}_{22} \\ : \\ : \\ \bar{x}_{n1} - \bar{x}_{n2} \end{vmatrix} = \begin{vmatrix} d_1 \\ d_2 \\ : \\ : \\ d_n \end{vmatrix} \quad \ldots\ldots\ldots(2-2)$$

متوسط المتغيرات في المجموعة الأولي $\equiv \bar{x}_i(1)$

متوسط المتغيرات في المجموعة الثانية $\equiv \bar{x}_i(2)$

الفرق بين المتوسطين $\equiv d$

المسافة بين كل متغيرين في المجموعتين $\equiv d_i$

2/ إيجاد التباين والتقارير المشترك وذلك من المجموعتين

([3]) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

$$sii = \sum x_i^2 - \frac{(\sum x_i)2}{n} \quad \text{.....(3-2)}$$

$$sij = \sum x_i x_j - \frac{\sum x_i x_j}{n} \quad \text{.....(4-2)}$$

sii هو عبارة عن مجموع المربعات للمتغير xi

sij هو عبارة عن مجموع حاصل ضرب للمتغيران xi ، xj ناقص $\frac{\sum x_i \sum xj}{n}$

Vii ***التباين المشترك للمتغير بين المجموعين***

$$vii = \frac{sii(1)+sii(2)}{n_1+n_2-2} \quad \text{.....(5-2)}$$

Vij **= التغاير المشترك**

$$vij = \frac{sij(1) + sij(2)}{n_1 + n_2 - 2} \quad \text{.....(6-2)}$$

$$i \neq j$$

مصفوفة التباين والتغاير المشترك

$$\mathbf{v} = \begin{vmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ : & : & : & : \\ v_{n1} & v_{n2} & \dots & v_{nn} \end{vmatrix} \quad \text{.....(7-2)}$$

وتمثل مصفوفة مربعة ومتماثلة قطرها الرئيسي يمثل التباينات وعناصرها الأخرى تمثل التغايرات المشتركة

3/ معادلة الدالة التميزية :

$$\hat{L} = \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \dots\dots\dots\dots + \hat{\alpha}_n x_n \quad \text{.....(8-2)}$$

$$\hat{a} = v^{-1} d = \begin{vmatrix} \hat{a}_1 \\ \hat{a}_2 \\ . \\ . \\ \hat{a}_n \end{vmatrix} \quad \text{.....(9-2)}$$

4-2 الأهمية النسبية للمتغيرات :

إن فائدة التحليل التمييزي هو أنة يمكن عمل مقارنة بين المتغيرات المستقلة من حيث أهميتها في عملية التمييز، ويتم حسابها بالصيغة الآتية:

$$\alpha_i^* = \hat{a}_i \sqrt{V_{ii}} \quad \text{.....(10-2)}$$

تمثل عناصر القطر الرئيسي في مصفوفة التباين والتغاير المشتركه V_{ii} حيث أن

وبمقارنة القيم الناتجة من حساب α_i^* فإن أكبر قيمة من جملة قيم المتغيرات هو أهم متغير له القدرة على عملية التمييز بين المجموعتين ويليها ثاني أكبر قيمة له القدرة على التمييز وهكذا .....

ويتم ترتيب المتغيرات المستقلة حسب الأهمية بصورة تنازلية علماً بان المقارنة يجب أن تكون بإهمال الإشارات السالبة .

2-5 اختبار الدالة التمييزية:

وهي خطوة ذات أهمية في التحليل التمييزي ، بعد حساب الدالة التمييزية ، يتم اختبار قدرة الدالة على التمييز ويمكن اختبارها باستخدام اختبار ***F*** عن طريق تكوين جدول تحليل التباين.

ثم تختبر الفرضية الآتية [4]:

H_0 : **الدالة ليس لها القدرة التمييز**

H_1 : **الدالة لها القدرة التمييز**

في هذا الاختيار يتم حساب مجموعة مربعات الخطأ من خلال كونه مربع المسافة بين المجموعتين .

$$SSE=D^2 \text{.....(11 - 2)}$$

$$D = \sqrt{\hat{a}_1 d_1 + \hat{a}_2 d_2 \dots\dots + \hat{a}_n d_n} \quad \text{.....(12-2)}$$

$$D^2 = \hat{a}_1 d_1 + \hat{a}_2 d_2 \dots\dots + \hat{a}_n d_n \quad \text{.....(13-2)}$$

ثم حساب مجموع مربعات بين المتغيرات "المعالجات " "SSB"

$$SSB = \frac{n_1 n_2}{n_1 + n_2 (n_1 + n_2 - 2)} * (D^2)^2 \quad \text{.....(14-2)}$$

أما مجموع المربعات الكلي يساوي

$$SST = SSB + SSE \quad \text{.....(15-2)}$$

بالنسبة لدرجات الحرية مجموع مربعات بين المتغيرات (المعالجات) هي عبارة عن عدد المتغيرات أو العوامل الداخلة في التمييز مطروح منها (1) .

(4) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

وبالنسبة لدرجة الحرية الكلية هي عبارة عن مجموع عدد المشاهدات في التجربة مطروح منها (1) .

أما درجات الحرية بالنسبة للخطأ فتأخذ الفرق بين درجة حرية المعالجات ومجموع المربعات الكلي ويتم تكوين جدول تحليل التباين ومنه نحسب قيمة F من خلال النسبة بين متوسط مجموع المربعات بين المتغيرات (المعالجات) ومتوسط مجموع مربعات داخل المتغيرات (الخطأ)[5].

جدول رقم (*3-1*)

جدول تحليل التباين

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	F
المعالجات	$K-1$	SSB	MSB	$F = \frac{MSB}{MSE}$
الخطأ	$n-k$	SSE	MSE	
الكلي	$n-1$	TSS		

وتقارن قيمة F المحسوبة مع الجدولية المستخرجة عند درجتي الحرية $F_{K-1,n-k,\alpha}$

فإذا كانت القيمة المحسوبة أكبر من الجدولية يعني أن الدالة التمييزية لها القدرة في عملية التمييز . والعكس صحيح .

وهناك أيضاً طريقة أخرى لاختبار قدرة الدالة على التمييز وهي اختبار t بواسطة المقارنة بين الأوساط الحسابية للقيم التمييزية

أولا نوجد القيم التمييزية لكل مفردة في كل مجموعة وذلك بتعويض قيم المتغيرات x_i المستقلة

حيث $i= 1,2,3,\ldots,K$

فمثلا القيمة التمييزية للمفردة الأولى في المجموعة الأولي كالأتي

$$L_{1(1)} = \hat{\alpha}_1 x_{11} + \hat{\alpha}_2 x_{21} + \ldots\ldots\ldots\ldots\ldots\ldots + \hat{\alpha}_K x_{K1}$$

وكذلك القيمة التميزية للمفردة الثانية :-

$$L_{2(1)} = \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_{22} + \ldots\ldots\ldots\ldots\ldots\ldots + \hat{\alpha}_k x_{K2}$$

وكذلك القيمة التمييزية لباقي مفردات المجموعتين.

ثانياً :

نحسب الوسط الحسابي للقيم التمييزية لكل مجموعة كالآتي :-

([5]) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

$$\bar{L}_{(1)} = \frac{\sum L_{i(1)}}{n_1} \quad \text{..... (16-2).}$$

حيث $i=1,2,3,\ldots\ldots, n_1$

$$\bar{l}_{(2)} = \frac{\sum L_{i(2)}}{n_2} \quad \text{.....(17-2).}$$

حيث $i=1,2,3,\ldots\ldots, n_2$

حيث $\bar{L}_{(1)}$ و $\bar{L}_{(2)}$ يمثلان الوسطان الحسابيان للقيم التمييزية في المجموعتين على التوالي .

ولاختبار هل أن الفرق بين المجموعتين معنوي أم لا ، نستخدم الاختبار الإحصائي (t) للمقارنة بين الأوساط الحسابية وذلك لبيان أهمية دالة التصنيف في تصنيف المجاميع ، ويمكن حساب الاختبار الإحصائي حسب الصيغة الآتية :-

$$t = \frac{\bar{L}_{(1)} - \bar{L}_{(2)}}{\sqrt{S^2 P\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{.....(18-3)}$$

ونقارن قيمة t المحسوبة مع الجدولية $t_{n_{1+}n_{2-2}, \frac{\propto}{2}}$

ثم نختبر الفرضية الآتية

$H_0 : \bar{L}_{(1)} = \bar{L}_{(2)}$

$H_1 : \bar{L}_{(1)} \neq \bar{L}_{(2)}$

فإذا قبلت H_0 فذلك يعني أن نمط القيم التمييزية في المجموعتين متشابهة وهذا يعني عدم قدرة الدالة التمييزية على التمييز ، أما إذا رفض H_0 وقبلت H_1 هذا يعني قدرة الدالة التمييزية للتمييز .

6-2 نقطة الفصل Cut Points :

عندما يراد تصنيف المفردات الجديدة فإننا نحتاج إلى نقطة فاصلة تفصل بين المجموعتين بحيث يمكن خلال مقارنة القيمة التميزية بين المفردة الجديدة في النقطة الفاصلة من خلال ذلك معرفة إلى إي مدى من المجموعتين تنتمي المفردة الجديدة هذه النقطة الفاصلة عبارة عن متوسط المتوسطين للقيم التميزية للمجموعتين .

$$\bar{\bar{L}} = \frac{\bar{L}(1) + \bar{L}(2)}{2} \quad \text{.....(19-2)}$$

$\bar{\bar{L}}$ يمثل متوسط المتوسطين للقيم التمييزية للمجموعتين

$\bar{L}(1)$ يمثل متوسط القيم التمييزية للمجموعة الأولى

$\bar{L}(2)$ تمثل متوسط القيم التمييزية للمجموعة الثانية

وهنالك حالتين :

- **الحالـة الأولـي :** إذا كانت $\bar{L}(1) > \bar{L}(2)$ فاذا تبين ان المفردة الجديدة اقل من نقطة الفاصلة فانها تنتمى الى المجموعة الأولي . واذا تبين ان النقطة الفاصلة اقل من المفردة الجديدة فانها تنتمي الى المجموعة الثانية .
- **الحالـة الثانيـة :** إذا كانت $\bar{L}(1) < \bar{L}(2)$ فاذا تبين ان المفردة الجديدة اكبر من النقطة الفاصلة فانها تنتمى الى المجموعة الثانية ، واذا كانت المفردة الجديدة اقل من النقطة الفاصلة فإنها تنتمي إلى المجموعة الأولي .

2-7 نسبة الخطأ :

عند التمييز المفردات الجديدة بين مجموعتين يترتب عليها نوعين من الخطأ :

أولاً : الخطأ الظاهري : وهو يمثل احتمال التصنيف الخاطئ .

n_1 : هي حجم العينة الأولي

n_{12}: عدد المفردات التى تنتمى بالاصل للمجموعة الأولي وتم تصنيفها للمجموعة الثانية ويتم حسابها كالآتي :

$$P_{12} = \frac{n_{12}}{n_1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20-2)$$

n_{21}: عدد المفردات التى تنتمى بالاصل للمجموعة الثانية وتم تصنيفها للمجموعة الأولي ويتم حسابها كالآتي :

$$P_{21} = \frac{n_{21}}{n_2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21-2)$$

n_2: هي حجم العينة الثانية

ثانياً : الخطأ الحقيقي : يعتبر أهم من الخطأ الظاهري ويتم إيجاده باستخدام جدول احتمالات للتوزيع الطبيعي القياسي بالصيغة .

$$P_{12} = P_{12} = \emptyset\left[-\frac{\sqrt{D^2}}{2}\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots (22-2)$$

حيث :

D^2 : تمثل مجموع مربعات الخطأ .

وبعدايجاد القيمة بين القوسين ، يتم ايجاد الاحتمال المقابل لها من جدول التوزيع الطبيعي ، فكلما كان الاختمال صغير (اى قريب من الصفر) دل على قوة الدالة التمييزية والعكس كلما كان الاحتمال اكبر دل على ضعف الدالة التمييزية.

8-2 حساب دالة التمييز باستخدام تحليل الانحدار المتعدد [6]:

في هذا الموضوع سيتم إيجاد دالة التمييز من خلال تحليل الانحدار والمشكلة في استخدام تحليل الانحدار هو عدم وجود المتغير التابع بين المتغيرات المستقلة هي نفس المتغيرات أو العوامل الموجودة في المجموعتين . ويتم إيجاد نموذج الانحدار الخطي المتعدد لبيانات لمجموعتين معاً بمعنى آخر يتم إدراج البيانات إحدى المجموعتين فوق أو تحت بيانات المجموعة الأخرى بحيث يكون هناك متغير واحد للمجموعتين وعدد مشاهداته $n_1 + n_2$.

أما المتغير التابع فهو متغير نوعي وذلك من خلال إعطاء (1) للمجموعة التي تتمتع بالصفة المدروسة بينما يتم إعطاء القيم " صفر " للمجموعة التي لا تتمتع بالصفة المدروسة .

مثلاً $Y=1$ إذا كان المريض مصاب و $Y=0$ إذا كان المريض غير مصاب ، وبعد تكوين بيانات المتغيرين " المتغير التابع والمستقل" يتم تقدير نموذج انحدار الخطى المتغير سوء بالقيمة الفعلي أو بانحرافات المتغيرات عن وسطها الحسابي الهدف هو الحصول على معاملات الانحدار

$$B = \begin{vmatrix} B_1 \\ B_2 \\ B_3 \\ \vdots \\ \vdots \\ B_n \end{vmatrix} \quad \text{.....(23-2)}$$

والتي ستكون لها علاقة مع قيمة $\hat{a}$ بالتناظر نلاحظ هي :

$$\hat{a}_i = \frac{\hat{B}_i}{c} \quad \text{.....(24-2)}$$

حيث :

(6) ريتشار جونسون ، دين وشرن، ـ التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية ـ دار المريخ للنشر ـ الرياض المملكة العربية

2-9 خطوات تطبيق التحليل التمييزي:

يمكن تلخيص الخطوات التي يجب على الباحث إتباعها عند تطبيق التحليل التميزي في الرسم البياني التالي :

شكل () : يبين خطوات خوارزمية التحليل التمييزي

ثانياً :التحليل العاملي

2-9مفهوم التحليل العاملي factor –analysis

هو أسلوب أحصائي يستهدف تفسير معاملات الارتباطات الموجبة – التي لها دلالة أحصائية –بين المتغيرات ، و بمعنى آخر فإن التحليل العاملي عملية رياضية تستهدف تبسيط الارتباطات بين مختلف المتغيرات الداخلة في التحليل وصولاً الى العوامل المشتركة التي تصف العلاقة بين هذه المتغيرات و تفسيرها . و يعد التحليل العاملي منهجاً إحصائياً لتحليل بيانات متعددة إرتبطت فيما بينها بدرجات مختلفة من الارتباط في صورة تصنيفات مستقلة قائمة على أسس نوعية للتصنيف ، و يتولى الباحث فحص هذه الأسس و أستكشاف ما بينها من خصائص مشتركة وفقاً للإطار النظري و المنطق العلمي الذي بدأ به . يبدأ التحليل العاملي بحساب الأرتباطات بين المتغيرات للحصول على مصفوفة من الأرتباطات بين هذه المتغيرات لدى عينة ما ، ثم بعد ذلك تحليل هذه المصفوفة الأرتباطية تحليلاً عاملياً للوصول الى أقل قدر ممكن من المحاور أو العوامل للتعبير عن أكبر قدر من التباين بين هذه المتغيرات ، و أن التحليل العاملي عبارة عن نموذج رياضي (**Mathematical Model**) يوضح العلاقة بين مجموعة كبيرة من المتغيرات بدلالة عدد قليل من العوامل الاساسية.

ان اعتماد معاملات الارتباط لتفسير العلاقة بين المتغيرات ينطوى على صعوبة كبيرة خاصة عندما يكون هناك عدد كبير من المتغيرات لذلك فالتحليل العاملي يقدم وسيلة لتلخيص هذه العلاقات لتسهيل تفسيرها ، و ان هذا التحليل لا يتطلب وضع اية فروض حول طبيعة المتغيرات او المشاهدات قيد الد ا رسة وهذا هو سبب مرونته حيث يمكن استخدامه على اوسع نطاق في الدراسات العلمية لتحليل عدد كبير من المتغيرات وارجاعها الى عدد اقل من العوامل المهمة التي تشكل من المتغيرات الاصلية بحيث تفسر معظم الاختلافات الموجودة في البيانات المحصول عليها وتعطى بعد ذلك النموذج الملائم الذي يمثل المشكلة المدروسة و بوصف اقتصادي و موضوعي للظواهر متعددة المتغيرات.

2- 10 شروط إستخدام التحليل العاملي

يتطلب الإستخدام الجيد لأي أسلوب إحصائي بعض الشروط التي يجب أن تتوفر في البيانات المطلوب تحليلها. وفيما يلي شروط إستخدام التحليل العاملي :

1. أن تكون المتغيرات موزعة توزيعاً طبيعاً .
2. وجود علاقة خطية بين المتغيرات .
3. يجب ألا يكون حجم العينة صغيراً
4. يجب تجنب إستخدام متغيرات غير مستقلة (متداخلة) .

2-11 نموذج التحليل العاملي factor analysis model

ان نموذج التحليل العاملي يفسر (p) من المتغيرات لعينة حجمها (n) على اساس دالة خطية تتكون من (p) من متوسطات المتغيرات و(m) من العوامل المشتركة و(p) من العوامل الوحيدة (Unique Factors) لكل متغير بحيث m<p والنموذج الخطي شكله كالاتي :

$$X_{(p\times1)} = \mu_{(p\times1)} + A_{(p\times m)}\ F_{(m\times1)} + U_{(p\times1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

حيث ان-:

X = يمثل المتجه العشوائي للمتغيرات المشاهدة:

μ= يمثل متجه اوساط المتغيرات

A= مصفوفة تحميلات العوامل Loading Factors من المتغيرات

F = المتجه العشوائي للعوامل المشتركة common Factors التي تم اختيارها من (p) المتغيرات.

U = الموجه العشوائي للعوامل الوحيدةunique factors للمتغيرات .

اذا كانت وحدات قياس المتغيرات مختلفة يستخدم القيمة المعيارية (Standard Value) في تحليل مصفوفة الارتباط وذلك بتحويل المتغيرات الى متغيرات قياسية(لها وسط حسابي واحد وتباين واحد)

اي ان:

1- متجه اوساط المتغيرات سيكون متجه صفري اي ان0= μ =EX

2- متجه تباين المتغيرات سيكون متجه احادي اي ان 1 =VAR(X)

ففي هذه الحالة يكون شكل نموذج التحليل العاملي كالاتي-:

$$X_{(Px1)} = A_{(pxm)}\ F_{(mx1)} + U_{(px1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

ان موجهي متوسطات كل من العوامل المشتركة والوحيدة هما موجبان صفريان تبعا لإفتراض ان موجه اوساط المتغيرات صفري ايضا اي ان

اما مصفوفة التغاير (Covariance Matrix) لكل من U,F (إفتراض كونها مستقلة) هي

$$E(U) = 0_{p\times 1}$$
$$E(F) = 0_{m\times 1}$$
$$COV(U,F) = E(U\acute{F}) = 0_{p\times m}$$
$$E(F\acute{F}) = \emptyset_{m\times m}$$
$$E(U\acute{U}) = \varphi_{p\times p}$$

حيث أن

Φ = مصفوفة التغاير لـF

φ = المصفوفة القطرية للتغاير *U*

وان مصفوفة التغاير لـ *X* هي

حيث أن $\sum$ هي مصفوفة موجبة متماثلة ومن الرتبة p

وان النموذج الخطي المعياري للتحليل العاملي لتفسير قيمة المفردة i للمتغير j لـ m من العوامل

Z ji = a j 1 F 1 i + a j 2 F2 i +..... + a jm F mi+Uji.....(3)

حيث :

Z ji تمثل القيمة المعيارية للمشاهدة i بالنسبة للمتغير j

a j 1 , a j 2 , ..., , a jm تمثل تحميلات العوامل (Loading Factors)

F 1 I, F2 I ,... F mi, تمثل القيمة المعيارية للمفردة i للعامل المشترك المحدد

Uji تمثل القيمة المعيارية للمفردة i للعامل الوحيد الخاص بالمتغير j

12-2 الفرضيات الاساسية للتحليل العاملي

تستند التحليل العاملي الى فرضيتين اساسيتين هما-:

الفرضية الاولى :وجود ارتباط بين مجموعة من المتغيرات التي يعرف أحيانا بـ(inter-correlation) و ان هذه الارتباطات ناتجة عن وجود عوامل مشتركة تؤثر فيها وان مقدار هذه الارتباطات تعود الى واقع تلك العوامل حيث يسعى التحليل العاملي الى تفسير الارتباطات بين المتغيرات بعوامل تكون اقل من المتغيرات المستخدمة ، ومن هذه العوامل يمكن تمثيل القيمة المعيارية وذلك في حالة افتراض وجود (m) من العوامل وكما توضحه المعادلة التالية :

S ji = a j 1 Z 1 i + a j 2 Z2 i +..... + a jm Z mi.....(4)

حيث:

S_{ji} تمثل القيمة المعيارية للقياس i بالنسبة للمتغير j

a_{jm} تمثل تحميل(تشبع) العامل m بالنسبة للمتغير j

Z_{mi} تمثل القيمة المعيارية للمشاهدة i بالنسبة للعامل m

الفرضية الثانية:

ان التحليل العاملي يفترض أيضاً وجود ارتباط بين المتغيرات (j , k) بحيث يمكن حسابه على طبيعة وتأثير تحميلات (تشبعات) العوامل المشتركة.

ويمكن أيجاد قيم الارتباط بين تلك المتغيرات كالآتي:

$$R = A A' \quad \ldots\ldots\ldots\ldots\ldots (5)$$

حيث ان

R: تمثل مصفوفة الارتباط

A: مصفوفة تحميلات العوامل

2- 13 طرق التحليل العاملي :

تحدد الطرق الحسابية المستخدمة فى التحليل العاملي كثيرا ، فهناك الطريقة القطرية ، والطريقة المركزية ، والطريقة المركزية بإستخدام متوسط الارتباطات ، وطريقة المكونات الاساسية ، ونوجزها فيما يلى :

1 - الطريقة القطرية Diagonal method : وتعد الطريقة القطرية من الطرق المباشرة والسهلة فى التحليل العاملي ، ويمكن استخدامها إذا كان لدينا عدد قليل من المتغيرات وتؤدى إلى استخلاص أكبر عدد ممكن من العوامل وتتطلب هذه الطريقة معرفة سابقة ودقيقة بقيم شيوع المتغيرات ، وبدون هذه المعرفة لايمكن استخدامها . وتستمد الطريقة القطرية اسمها من كونها تقوم على استخدام القيم القطرية فى المصفوفة الارتباطية مباشرة . وتبدأ الطريقة القطرية باستخلاص هذه القيمة بكاملها فى العامل الأول ، وبذلك يكون جذر هذه القيمة هو تشبع المتغير الأول على العامل الأول ، ويطلق عليه اسم التشبع القطري وهكذا

2 - الطريقة المركزية Centroid method : كانت الطريقة المركزية " لثرستون " أكثر طرق التحليل العاملي استخداما وشيوعا إلى عهد قريب نظرا لسهولة حسابها فضلا عن استخلاص عدد قليل من العوامل العامة . غير أن هذه الطريقة تفتقر إلى عدد من المزايا الهامة ، أهمها أنها لا تستخلص الا قدرا محدودا من التباين الارتباطي ، تتحدد قيم

الشيوع فى المصفوفة الإرتباطية وفق تقديرات غير دقيقة حيث تستخدم أقصى ارتباط بين المتغير وأى متغير فى المصفوفة وهو اجراء يؤدى إلى خفض رتبة المصفوفة.

3 – الطريقة المركزية باستخدام متوسط الارتباطات Averoid method : لا تختلف هذه الطريقة عن الطريقة المركزية المعتادة إلا فى استخدامها تقدير الشيوع عبارة عن متوسط ارتباطات المتغير ببقية المتغيرات فى المصفوفة ثم حساب العوامل بعد وضع المتوسط الخاص بارتباطات كل متغير فى خليته القطرية ولهذا السبب يطلق على هذا الاسلوب اسم الطريقة المركزية بإستخدام المتوسطات . غير أن هذه الطريقة لا توفر نفس الدقة التى تجدها فى الطريقة المركزية التامة ، إذ تؤدي إلى خفض محدود فى نسبة التباين التى تعبر عنها العوامل الناتجة. غير أن هذه الطريقة تبدو مفيدة فى حالة وجود عدد كبير من المتغيرات دون توفر وسائل آلية لاجراء العمليات الحسابية .

4 – طريقة المكونات الأساسية Principal componants : تعد طريقة المكونات الأساسية التى وضعها "هويتلنج" Hottelling عام 1933 من أكثر طرق التحليل العاملي دقة وشيوعاً في بحوث الشخصية ، ولهذه الطريقة مزايا عدة منها أنها تؤدي إلى تشبعات دقيقة . وكذلك " فإن كل عامل يستخرج أقصى كمية من التباين (أى أن مجموع مربعات تشبعات العامل تصل إلى أقصى درجة بالنسبة لكل عامل) ، وتؤدى إلى أقل قدر ممكن من البواقي ، كما أن المصفوفة الارتباطية تختزل إلى أقل عدد من العوامل المتعامدة (غير المرتبطة).

ولم تلق طريقة المكونات الأساسية فى البداية قبولا كبيرا بين الباحثين نظرا لحاجتها إلى وقت حسابات طويل لإنمامها ولذا كان من المستحيل استخدامها يدويا فى حالة المصفوفات الكبيرة ، ولكن بعد الاعتماد على الآلات الحاسبة الالكترونية ذا السرعة الفائقة والدقة الشديدة وطاقة التخزين الكبيرة ، اصبحت هذه الطريقة الآن من بين أكثر الطرق شيوعا نظرا لدقة نتائجها بالمقارنة ببقية الطرق .

لذلك اعتمدت طريقة العامل الرئيسي لتحليل بيانات عينة البحث، لذا سوف يكون تركيزنا على هذه الطريقة و تطبيقاتها في هذه الدراسة .

طريقة العامل الرئيس (Principle Factor Method)

هي تطبيق لطريقة المكونات (Principal Factor Method) ان طريقة العامل الرئيس الاساسية

(Principle Component Method) ولكن بإستخدام مصفوفة الإرتباط المختزلة (Reduced Correlation Matrix)

حيث ان طريقة المكونات الاساسية تعد من اكثر طرق التحليل العاملي دقةً و شيوعاً في البحث ولهذه الطريقةمزايا عدة منها انها تؤدي الى تشبعات دقيقة و تؤدي الى اقل قدر ممكن من البواقي

كما ان المصفوفة الارتباطية تختزل الى اقل عدد من العوامل المتعامدة (غير المرتبطة)،يمكن كتابة

المعادلة رقم (3) بصيغة المصفوفات:

$$Z(p\times 1) = A(p\times m)F(m\times 1) + U(P\times 1).....(6)$$

لاجل تقدير مصفوفة تحميلات العوامل (A) : تبع الخطوات الاتية :

1 - نحسب مصفوفة معاملات الارتباط R (Correlation Coefficient Matrix) من القيم المعيارية (Standard Value) للمتغيرات ذات وحدات قياس مختلفة أما اذا كانت المتغيرات لها نفس وحدات القياس نستخدم مصفوفة التباين (التغاير) **Matrix Variance–Covariance**

حيث ان:

$$r_{ij} = \frac{\sum_{k=1}^{n} Z_{ik}Z_{jk}}{n-1}$$

$$r_{ii} = r_{jj} = 1$$

2 - نحسب مربع معامل الارتباط المتعدد لكل متغير مع بقية المتغيرات كتقدير أولي لقيم الشيوع لتحل محل الأحاد كعناصر قطرية ،أي أستبدال r_{11} بـ $R_{1,rest}$ و r_{22} بـ $R_{2,rest}$

هكذا نحصل على مصفوفة الأرتباط المختزلة Reduced Correlation Matrix ترمز ب Rr من مصفوفة الارتباط المختزلة نستخرج القيم المميزة (Eigen Values) بموجب المعادلة المميزة (Characteristics Equation) الأتية :

$$|Rr - \lambda I| = 0 .....(7)$$

حيث نحصل على قيم عينية بعدد المتغيرات (p)

3- نختار االقيم المميزة التي تزيد قيمتها على الواحد ، حيث يمثل عددها عدد العوامل التي نستخدمها في التحليل

4- نستخرج المتجه العيني Eigen Vector–a المرافق لكل قيمة مميزة مختارة و ابتداء بأكبر

قيمة و ذلك بموجب نظام المعادلات الآتية:

$$|Rr - \lambda I|a = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

حيث ان القيمة المميزة هي مقدار مساهمة العامل في مجموع قيم الشيوع ان المتجه المميز المرافق لاكبر قيمة مميزة يمثل تحميلات العامل الاول المقدرة، و المتجه المميزة المرافق الثاني أكبر

قيمة مميزة يمثل تحميلات العامل الثاني المقدرة و هكذا ، و بهذه الطريقة نحصل على مصفوفة تحميلات العوامل المقدرة الاولي أي أن:

5- نستخرج قيم الشيوع من A_1 وكما يأتي

$$h_1{}^2 = a_{11}{}^2 + a_{12}{}^2 + \ldots\ldots\ldots + a_{1m}{}^2$$
$$\vdots$$
$$h_p{}^2 = a_{p1}{}^2 + a_{p2}{}^2 + \ldots\ldots\ldots + a_{pm}{}^2$$

توضع هذه القيم في مصفوفة الارتباط المختزلة كعناصر قطرية Rr .

6- للحصول على مصفوفة تحميلات العوامل المقدرة الثانية A_2 تعاد الخطوات (3 و 4 و 5 و 6) وهكذا نستمر الى أن يكون الفروق بين $h_j{}^2$ لمصفوفتين متتاليتين صغيرة جداً و تكون هذه المصفوفة هي مصفوفة العوامل المقدرة الاخيرة (A) و التي تكون الحل الاولي Initial Solution

2- 14 التركيب البسيط و تدوير المحاور Simple Structure and Rotation of Axes

يؤدي التحليل العاملي لمصفوفة ارتباطية ، بأية طريقة من الطرق العاملية إلى استخلاص عوامل معينة ، وهذه العوامل ، بمعنى آخر ، عبارة عن محاور متعامدة تمثل تشبعات المتغيرات إحداثياتها ، وهى تتحدد بطريقة عشوائية ، ويختلف هذا التحديد للمحاور من طريقة عاملية

لأخرى ، فهل يمكننا قبول العوامل الناتجة فى تحليلاتنا على أنها الصورة النهائية التى تلخص لنا العلاقات الارتباطية المتعددة وبصورة مقبولة إحصائيا .

تعد هذه الصورة غير مقبولة بوجه عام من وجهة نظر علماء الإحصاء هنا يقوم الباحث بإجراء جديد على هذه العوامل أو المحاور يهدف أساسا إلى إعادة تحديد مواضعها ، بهدف الوصول بها إلى قدر من الثبات والاتساق وحتى يتسنى لنا تفسيرها ، واضعا فى اعتباره أن الخطوات الحسابية لاستخلاص العوامل إنما تقوم على التعامل مع إرتباطات بين متغيرات فى صورة كمية لا تتضمن ما تعنيه هذه المتغيرات أو مضمون الارتباطات ، حيث يقوم الباحث بإجراء تعديل فى مواضع المحاور التى توصل إليها ليكسب هذه المحاور معناها الواضح .

وهنا نوعان من التدوير تبعا للزاوية التى تفصل بين المحاور المرجعية وهما التدوير المتعامد Orthogonal Rotation والتدوير المائل Oblique Rotation ففى التدوير المتعامد تدار العوامل معا (اثنين منها مثلا) مع الاحتفاظ بالتعامد بينها . أما التدوير المائل ففيه تدار المحاور دون احتفاظ بالتعامد ، فتترك لتتخذ الميل الملائم لها .

والعوامل المتعامدة غير مرتبطة معا ، أى أن معاملات الارتباط بينها تساوى صفرا ، إذ تصنف العوامل الاختبارات أو المتغيرات إلى فئات غير مرتبطة ، وهكذا يصبح التقسيم جادا غير متداخل. أما العوامل المائلة فهي عوامل بينها ارتباط أي أنها عوامل متداخلة ، ويفضل بعض المحللين العامليين استخراج عوامل متعامدة غير مرتبطة ، فى حين يهتم آخرون باستخلاص المائلة . ويهدف تدوير المحاور إلى تحقيق ما يسميه " ثيرستون " البناء البسيط .

تسمى العوامل الناتجة عن استخدام إحدى الطرق الحسابية للتحليل العاملي بالعوامل المباشرة ، وهى تمثل الحل الرياضى ، وهذا الحل واحد فقط من حلول كثيرة ممكنة كما أسلفنا ، وكذلك فإنه فى أحوال غير قليلة يصعب تفسير مثل هذه العوامل المباشرة ، فيكون الهدف إذن هو أن تحول هذه العوامل الى وضع يمكن الباحث من تفسيرها إحصائياً وتزيد كذلك من بساطتها ومعنوية ارتباط العوامل بمتغيرات القياس الأصلية . ويرى " ثيرستون " أنه يصعب تفسير العوامل إلا بعد تدوير المحاور وتبسيط كل " عمود " بقدر الامكان ، ويكون ذلك بتحويل نمط التشبعات إلى ما يسميه بالبناء البسيط ، ويرى أن الأخير يضمن وصول التحليل إلى نتيجة ثابتة تكون عواملها قابلة للتكرار من دراسة إلى أخرى .

وتتعدد الطرق العملية للتدوير وكان إسلوب الكوارتيماكس Quartimax الذى اقترحه " كارول " Carroll هو أول الأساليب التحليلية التى ظهرت فى سنة 1953 في محاولة لتقديم حل رياضي للبناء البسيط ، ثم تناولت بعد ذلك عدة طرق رياضية لعل أشهرها طريقة الفاريماكس Varimax الذى قدمه " كايزر " Kaiser فى سنة 1958 ، وتتقبل طريقة " الفاريماكس " فكرة

البناء البسيط مع الاحتفاظ بالتعامد بين العوامل ، ويميل أغلب الباحثون لاستخدام طريقة الفاريماكس لكايزر والتى تؤدى إلى أفضل الحلول التى تستوفى خصائص البناء البسيط .

وكما نجد عدد من الأساليب التحليلية لحساب العوامل المتعامدة ، يوجد أيضا عدد آخر من الأساليب المعروفة لحساب العوامل المائلة ، بعضها حلول قائمة على العوامل المباشرة والبعض الآخر يبدأ من الحل المتعامد . ومن الطرق المعروفة فى مجال التدوير المائل : طرق الكوارتيمن Quartimin و الأوبليمين Oblimin " لكارول " Carroll والـ Covarimin " لكايزر " والـ Binornamin " لكايزر وديكمان " Dickman والـ Promax " لهندريكسون ووايت " Hendrickson and White و غيرها . عندما نقوم بتدوير متعامد لمصفوفة عاملية فإننا نصل إلى نتيجة واحدة هي مصفوفة العوامل بعد التدوير وحيث تكون التشبعات على العوامل هى نفسها – أيضا – الارتباطات بين المتغيرات والعوامل . غير أن هذا الأمر يختلف فى حالة التدوير المائل .

فعندما تصبح العوامل مائلة يتحدد معنى التشبعات بإعتبارها إحداثيات المحاور العاملية Primary Factors بينما توجد لدينا إحداثيات المتجهات المرجعية Refexence vactors والتى تعبر عن الارتباطات بين المتغيرات والعامل ، ومثل هذا التمييز بين المتجهات المرجعية والعوامل الأولية يؤدي إلى خروجنا من التدوير المائل بمصفوفتين ، الأولى : هى مصفوفة النمط العاملي Factorial pattern أو نمط العوامل الأولية وقيم عواملها هى تشبعات المتغيرات على العوامل . والثانية : هى مصفوفة البناء العاملي Factorial Structure وقيم عواملها هى معاملات الارتباط بين المتغيرات والعوامل .

والواقع أن الميزة الأساسية التى تكمن فى التدوير المائل هى أنه الخطوة اللازمة للتقدم نحو التحليلات العاملية من الرتبة العليا أو الثانية . فطالما لدينا عوامل مائلة (مترابطة) فنستطيع أن نحسب مصفوفة معاملات الارتباط بين عواملها ثم نقوم بإجراء تحليل هذه المصفوفة الارتباطية لكى نستخرج عوامل من الرتبة الثانية أو الراقية التى تستخرج من التحليل العاملى للارتباط بين العوامل ، وتفسر هذه المصفوفات بطريقة تفسير العوامل الأولية ذاتها ، فيما عدا – بطبيعة الحال – أن المتغيرات هنا هى العوامل من التحليل العاملي ذى الرتبة الأولى أو الدنيا . وإذا ما كان هناك عديد من العوامل ذات الرتبة الثانية وأديرت تدويرا مائلا ، نتج أيضا مصفوفة ارتباطات بين هذه العوامل ذات الرتبة الثانية ، وهذه المصفوفة الارتباطية الأخيرة يمكن أن تحلل أيضا وتؤدى إلى العوامل ذات الرتبة الثالثة ، ويمكن أن تستمر العملية طالما أمكن إنتاج مصفوفة ارتباطية بالتدوير ، وتتوقف التحليلات ذات الرتبة الراقية حتى يحدث أن يستخرج عامل واحد فقط أو عوامل غير مرتبطة.

إما عن المقارنة بين النوعين بين العوامل (ذات الرتبة الأولى والثانية) أن الفروق التى قد تكون ضئيلة بين نتائج التدويرين المتعامد والمائل ليست هى موضوع اهتمام الباحثين بقدر ما يتجه اهتمامهم نحو إضفاء منطق الترابط أو التعامد بين العوامل ، ويعتقد عدد كبير من الباحثين أن التدوير المائل يعد أكثر كفاءة فى إبراز معالم البناء البسيط .

ويهدف التحليل العاملي الى اعطاء صورة واضحة عن طبيعة العلاقة بين المتغيرات من خلال ابراز العوامل الكامنة و ا رء هذه العلاقات و حيث ان تفسير النتائج المستخلصة يعتبر هدفاً اساسياً فان مصفوفة العوامل التي يعتمد عليها هذا التفسير لابد أن تكون معاملاتها سهلة التفسير و ذات دلالة معنوية و من هنا تأتي أهمية التدوير للحصول على التركيب البسيط لمصفوفة العوامل المستخلصة A ففي عام 1947 وضع Thurston المعايير الخاصة للتركيب البسيط Simple Structure التي ينبغي أن تتصف بها مصفوفة العوامل و هي :

1 - يضم كل صف من صفوف مصفوفة العوامل A قيمة صفرية واحدة على الاقل

2 -يضم كل عمود من أعمدة مصفوفة العوامل على P من القيم الصفرية على الاقل

15-2 مصفوفة العوامل المدورة

البحث هذا في اعتمد قد و المدورة العوامل مصفوفة على للحصول طرق عدة هناك تعديل هي و المتعامد للتدوير (Quartimax) للتركيب قرباً أكثر طريقة وتعبر على تعتمد و شيوعاً المتعامد التدوير طرق أكثر من انها اذ ، البسيطVarimax لطريقة ان اذ $S_p{}^2$ تباين تحميلات خلال من العوامل تبسيط

$$S_p{}^2 = (1/n)\ \Sigma_{j=1}^{n}(a_{p1}{}^2)^2 - (1/n^2)\ (\Sigma_{j=1}^{n} a_{p1}{}^2)^2 \ldots\ldots\ldots (9)$$

التباين يكون عندما و المدورة المصفوفة فيp للعمود jالصف عنصر حيث أن هو أعظم ما يمكن يكون للعامل قابلية للتفسير و التبسيط على أساس أن تحميلاته تتجه حول الصفر والواحد فمقياس اعلى تبسيط لمصفوفة العامل الكامل Complete Factor Matrix يعرف بأنه تعظيم لمجموع التبسيطات للعوامل المفردة فاذا جمعت المعادلة (12) لكل العوامل فان:

$$S^2 = \Sigma_{p=1}^{m} S_p{}^2 = (1/n)\ \Sigma_{p=1}^{m}\Sigma_{j=1}^{n}(a_{p1}{}^2)^2 - (1/n^2)\ \Sigma_{p=1}^{m}(\Sigma_{j=1}^{n} a_{p1}{}^2)^2 \quad (10)$$

حيث ان a_{jp} تمثل قيمة تشبع j بالعامل p أن n هو عدد المتغيرات ان التعظيم في المعادلة (13يسمى مقياس تعظيم التباين العام ل Kaiser (Row varimax criterion) وبترجيح المعادلة (13) بدلالة قيم الشيوع $h_j{}^2$ فان معيار الإقتران لتعظيم التباين هو جعل قيمة (V)اعظم مايمكن للحصول على أفضل تحميلات للعوامل و كما يأتي:

$$V = \sum_{p=1}^{m} S_p{}^2 = (n)\ \sum_{p=1}^{m}\sum_{j=1}^{n}(a_{jp}/h_j)^4 - \left(1/n^2\right)\ \sum_{p=1}^{m}(\sum_{j=1}^{n} a_{p1}{}^2/h_j)^2 \ldots(11)$$

وقد اطلق على المعادلة (11) مقياس تعظيم التباين الطبيعي (Normal Varimax) من قبل Kaiser

16-2 عدد العوامل

محكات تحديد عدد العوامل المستخرجة :

تعد مشكلة تقدير عدد العوامل التى يتعين إنتاجها فى الدراسة العاملية من المشكلات التى تؤرق الباحثين ، ذلك أن إمكان استخلاص عوامل من المصفوفة الارتباطية إلى الحد الذى تصبح فيه آخر مصفوفة بواقى صفرية من الأمور الممكنة وحيث يمكن استخلاص عدد من العوامل يساوى عدد المتغيرات التى بدأنا بها .

أنه من الممكن – نظريا وحسابيا – أن يستمر استخراج عدد من العوامل مساوٍ لعدد المتغيرات . ويتنازع المحلل العاملي فى هذه الحال مطلبان قد يكونان متعارضين وهما :

1 – أن يستخرج أقل عدد من العوامل وفى هذا تحقيق لمنطق الطريقة وواحد من الأهداف الهامة للتحليل العاملى من حيث هو منهج علمي ينحو نحو الايجاز والدقة وتفسير الكثرة بالقلة ، وهذا هو مبدأ الاختزال .

2 – ألا يهمل جزءا من التباين الجوهرى الذى يكشف عن الفروق الفردية ، وقد يكون هذا الجزء الذى تركه هاماً فى تفسير الظاهرة موضع البحث وهذا هو مبدأ الكثرة أو التعدد.

أنه لا توجد حتى الآن قاعدة رياضية مقبولة من قبل الجميع للتوقف على استخلاص العوامل ، وإن كان هناك عدد من المحكات التى يمكن استخدامها لهذا الغرض ، والواقع أنها تؤدى فى الغالب إلى نتائج متقاربة ومن أهم هذه المحكات الآتى :

(1) محك تيكر : Tuker's Criterion

وهو كما يظهر من اسمه يقوم أساسا على استخدام معامل فاي ويعتمد على مبدأ أنه إذا لم يكن هناك تناقض واضح فى حجم قيم البواقي من مصفوفة إلى أخرى تليها (بعد استخلاص عامل آخر) فإن العوامل العامة الجوهرية فى المصفوفة الارتباطية تكون قد استخلصت بالفعل وما يتبقى ليس إلا بواقي لا أهمية لها .

(2) محك همفري : Huamphrey Criterion

بينما كانت الطريقة السابقة تعتمد على حجم التباين فى مصفوفة البواقي ومدى تناقصه تدريجيا بعد كل عامل مستخلص فان محك " همفري " تقوم على أساس آخر مختلف تماما فهي من ناحية تعتمد على حجم العينة الأصلية التى حسبت الارتباطات بين متغيراتها وتعتمد ثانيا على فكرة أن تشبعين فقط (وليس ثلاثة) كافيين لتقرير وجود عامل عام وعلى ذلك نكتفي هذه

القاعدة باستخدام مؤشرات عاملية عبارة عن أعلى تشبعين لمتغيرين بالاضافة إلى حساب الخطأ المعياري لمعامل ارتباط صفري للمقارنة بينهما كمؤشر للتوقف أو الاستمرار فى استخلاص عوامل جديدة .

(3) محك كومب : Coomb Criterien

ومنطق هذا الأسلوب يعتمد على تناول نمط البواقي في المصفوفة أكثر من اعتماده على حجمها أو دلالاتها حيث يفترض أنه فى حالة وجود عوامل ذات دلالة مرتفعة لم تستخلص بعد وليس مجرد تباين خطأ فى المصفوفة فعلينا أن لا نتوقع قيم سالبة أكثر فى مصفوفة البواقي بعد العكس مما يتوقع بحكم الصدفة فى مصفوفة ناتجة عن ارتباطات ايجابية .

(4) محك كايزر : Kaiser Criterion

محك كايزر محك رياضي فى طبيعته اقترحه " جوتمان " (Guttman،1954)فى فترة سابقة ومنطلق هذا المحك يعتمد على حجم التباين الذى يعبر عنه العامل ، فلكي يكون العامل بمثابة فئة تصنيفية فلابد أن يكون تباينه أو جذره الكامن أكبر أو مساوٍ على الأقل لحجم التباين الأصلى للمتغير ، وبما أننا لا نستطيع نظريا استخلاص كل تباين المتغير فى عامل واحد فإن حصولنا على عامل جذره الكامن لا يقل عن واحد صحيح لابد أن يكون مصدر تباينه أكثر من متغير وبالتالي يكون عاملا معبرا عن تباين مشترك بين متغيرات متعددة.

وعلى ذلك فان هذا المحك يتطلب مراجعة الجذر الكامن للعوامل الناتجة وعلى أن تقبل العوامل التى يزيد جذرها الكامن عن الواحد الصحيح وتعد عوامل عامة . ويبدو هذا الأسلوب صالحا ومناسبا على وجه الخصوص لطريقة كالمكونات الأساسية " لهوتلينج ".

أن العوامل الدالة فى هذه الطريقة هى العوامل التى يساوى أو يزيد جذرها الكامن على واحد صحيح ، أى أن التباين الذى يستوعبه كل عامل (مجموع مربعات التشبعات على كل عامل) 1.0، بشرط أن يكون قد وضع فى الخلايا القطرية واحد صحيح . ومن حسن الطالع أن هذه الطريقة تعطى نتائج متقاربة تماما مع عدد العوامل المستخرجة عادة ، بالإضافة إلى سهولة حساب هذا المعيار وهو شائع الاستخدام . ويذكر " وايت وزملائه " (White, et al, 1969) أن هذا المعيار تتطابق نتائجه مع معايير أخرى ، ويزكون استخدامه على أساس " أنه من غير المعقول أن نقبل عوامل لا تستوعب تباينا أكبر مما هو متو فر في المتغيرات الأصلية ذاتها " ، أى أن العامل الذى يقل الجذر الكامن له عن واحد صحيح يشير إلى قدر ضئيل من التباين فى المتغيرات الأصلية ذاتها لذا فمن الأجدر استبعاده لعدم دلالته . وسوف نعتمد على هذا المعيار فى الدراسة الحالية.

(5) محك كاتل : Cattell Criterion

ويذكر " صفوت فرج " (1991: 245) أن خطوات استخلاص العوامل من المصفوفة الارتباطية تؤدي إلى إنتاج العوامل الأكثر عمومية أولا فى كل الأساليب العاملية بلا استثناء ، ثم تبدأ العوامل الخاصة أو التباين النوعى فى الظهور ، وفى طريقة كالمكونات الأساسية لا تفرق بين عوامل عامة وأخرى غير عامة يفترض أيضا أن حجم التباين النوعي الذى يتسرب إلى العوامل الناتجة يتزايد فى العوامل الأخيرة ويبدأ فى فرض صورة تقلل من أهمية المصفوفة العاملية ويتطلب الأمر فى هذه الحالة تحديد العدد الأمثل من العوامل قبل أن يؤدى ظهور التباينات الخاصة إلى أحداث خلل فى مصفوفة العوامل ، ويقترح " كاتل " هنا محكا بسيطا يطلق عليه اسم البقايا المبعثرة Scree test وذلك بأن تقوم برسم محورين متعامدين ، أفقي نضع عليه عدد العوامل فى تحليلنا (الذى انتج فيه عددا كبيرا من العوامل) ويقسم المحور الرأسي وفقا لوحدات منتظمة معبرة عن الجذر الكامن المستخلص للعوامل المختلفة.

وسنلاحظ بعد إتمام رصد عواملنا وجذورها الكامنة ، أن حجم الجذر يتناقص بشكل كبير فى العوامل الأولى إلى أن يصل إلى نقطة معينة هى غالبا حول جذر كامن واحد صحيح ثم يبدأ حجم الجذر فى التناقص بصورة ضئيلة بحيث يستوى فيها الخط البياني مع الخط الأفقي .

وإذا افترضنا أن النقطة التى سنتوقف لديها فى قبولنا للعوامل هى عند العامل الرابع على سبيل المثال فإن الفرق لمن يكون كبيرا فى الواقع بين ما يقدمه محك " كاتل " وبين ما يقدمه محك " كايزر" الذي يتطلب التوقف عند العامل الثالث هذا على سبيل المثال .

وتتبقى لطريقة " كايزر " ميزتها فى هذه الحالة فى كونها لا تتطلب استخلاص عدد كبير من العوامل ثم رصدها فى الشكل البياني للتعرف على نقطة توقف التناقص واستواء الخط ، حيث يمكن حساب الجذر الكامن لكل عامل بطريقة كايزر قبل استخلاص العامل التالي مما يوفر جهدا لا مبرر له .

ان عدد العوامل التي يمكن استخراجها لتصغير الارتباط بين المتغيرات تحدد بعددالجذور المميزة (القيم العينية) التي قيمتها تزيد عن الواحد الصحيح ، ففي عام 1960 قدم Kaiser هذه الطريقة و التي تعد من ابسط الطرق لتحديد العوامل المعنوية.

ثالثاً : التحليل العنقودي

2-18 مفهوم التحليل العنقودي

إن العنقدة أو التحليل العنقودي هي طريقة أنموذجية لتجميع نقاط البيانات (العناصر) ضمن محيط التصنيف غير الموجَّه . فيتم تقسيم مجموعة من البيانات الى عدد من المجاميع الجزئية أو العناقيد بالإعتماد على تشابه العناصر ، حيث تملك العناصر داخل العنقود الواحد درجة عالية من التشابه بينما تملك العناصر المنتمية الى عناقيد أخرى ومختلفة درجة عالية من عدم التشابه . وتتم عملية تصنيف العناصر الى عناقيد على أساس المقاييس الموضوعة على هذه العناصر ، فمثلاً إذا توفرت لدينا مجموعة من البيانات $[X=(x_1,....,x_n)]$ التي توضع في موجه خواص العناصر المنتمي الى فضاء العينة (R^n) فإن الهدف هو تجزئة أو تحديد المجاميع الجزئية أو العناقيد لعناصر متشابهه على أساس مجموعة متجهات خواص العناصر $[X_1=x_1,.....,X_n=x_n]$. تتضمن العنقدة عدة خطوات محددة منها تحديد المسافة الملائمة بين العناصر بالإعتماد على خواص وصفات ملائمة، ومن ثم يجب إختيار وتطبيق خوارزمية العنقدة المناسبة ويعتمد إختيار وتصنيف خوارزميات العنقدة على وفق الآتي :

- نوعية البيانات الداخلة الى الطريقة .
- معيار العنقدة من خلال تعريف مقياس التشابه بين العناصر .
- تحديد النظريات والمفاهيم الأساسية التي تبين على ماذا تعتمد أساليب التحليل العنقودي (مثل النظرية الضبابية والإحصاءات) .

عند إجراء عملية العنقدة يتم البحث عن التجزئة المثلى (غير العادية) non-trivial partition لـ (n) من العناصر الى (k) من المجاميع الجزئية حيث أن المعيار المحدد للتجزئة يكون هو أمثل معيار . ولإيجاد أفضل تجزئة يتطلب ذلك فحص كل تجزئة ممكنة ، حيث تمثل عدد التجزئيات المثلى لـ(n) من العناصر الى (k) من العناقيد بالآتي :

$$N(n,k)=\frac{1}{K!}\sum_{i=1}^{K}(-1)^{K-i}\binom{K}{i}.i^n \qquad \ldots(1.2)$$

2-19 التصنيف والتحليل العنقودي Classification and Cluster Analysis

تقسم عملية التصنيف الى نوعين أساسيين :

1- التصنيف غير الموجّه (التعلم غير الموجّه) Unsupervised Classification

عملية تصنيف البيانات الى مجاميع جزئية غير معلومة وغير محددة مسبقاً وتتم عملية تشخيصها وإكتشافها من خلال طبيعة البيانات حيث تكون العناصر داخل المجموعة المحددة متشابه فيما بينها [31] ، مثال على ذلك التحليل العنقودي (Cluster Analysis) ، إكتشاف تصنيف الفئة (Class Classification) وتمييز الأنموذج غير الموجّه (Unsupervised Pattern Recognition) .

2- التصنيف الموجّه (التعلم الموجّه) Supervised Classification

في هذا النوع من التصنيف يتم التحديد المسبق (Predefined) للمجاميع الجزئية للبيانات المحددة مسبقاً حيث تكون العناصر داخل المجموعة الواحدة بالنسبة للمجاميع المعروفة الأخرى غير معّرفة ولابدّ من تحليلها . ويتركز هدف هذا النوع على فهم أساس التصنيف بالإعتماد على مجموعة العناصر المميزة (label objects) أو المعلومات المسبقة (prior information) حول المجموعات الجزئية ، حيث تستخدم هذه المميزات أو المعلومات بعد ذلك لتصنيف خواص مشاهدات العناصر [31]، [38] . مثال على هذه تصنيفات التحليل المميز (Discriminant Analysis) ، تنبؤ قيمة الحالة (الفئة) Class Prediction وتمييز النموذج الموجّه (Supervised Pattern Recognition) .

في معظم الحالات تكون العنقدة (Clustering) أكثر صعوبة وشمولية من التصنيف (Classification) . وبصورة عامة إن المفاهيم التي تخص التصنيف تخص التحليل العنقودي مع إضافة مجموعة من النقاط [31] ، [38] :

i) عدم وجود مجموعة مجموعة تدريبية (Learning Set) للمشاهدات المميزة وكذلك عدم توفر المعلومة المسبقة أو إنها تكون غير كافية .

ii) عدد مجاميع الفئات غير معلومة .

iii) يعتمد التصنيف على إختيار خواص العناصر المناسبة ومقياس المسافة معاً .

iv) قد تكون الأهداف غير واضحة (إيجاد بعض العناقيد المهمة ذات الأهتمام الخاص في البيانات) .

v) أغلب الطرائق المستخدمة أو المتعامل معها تكون صعبة حسابياً عند ايجاد الحلول المثلى، لذلك يتم إستخدام الحلول التقريبية بدلاً عنها وتصبح هي النتائج المعّول عليها .

20-2 تجزئة العناقيد Cluster Partition

تهدف أساليب التحليل العنقودي بصورة أساسية الى عملية تجزئة وتصنيف مجموعة من البيانات الى مجاميع جزئية متجانسة (عناقيد) . وإن إمكانيات التصنيف هذه تكون وفقاً لمجاميع جزئية حادة (hard) أو ضبابية (fuzzy) . وتعتمد طرائق العنقدة الواضحة على نظرية المجموعة التقليدية (classical set theory) وتتضمن فكرة هذه الطرائق على إنتماء أو عدم إنتماء العنصر الى العنقود . بينما تسمح طرائق العنقدة الضبابية للعناصر بالإنتماء الى عدة عناقيد في وقت واحد مع درجات مختلفة من العضوية . في معظم الحالات تكون العنقدة الضبابية أكثر مرونة من العنقدة الحادة بحيث لاتنتمي العناصر بالكامل الى عنقود واحد ولكنها تحدد درجات العضوية بين 0 و 1 . وتتمثل تشكيلة مصفوفة التجزئة $(U=(u_{ij}))$ (partition matrix) , لـ k من العناقيد $(i=1,...,k)$ و n من العناصر $(j=1,.....,n)$ بالشكل الآتي :

$$U=(u_{ij})=\begin{bmatrix} u_{11} & u_{12} & . & . & . & u_{1n} \\ u_{21} & u_{22} & . & . & . & u_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ u_{k1} & u_{k2} & . & . & . & u_{kn} \end{bmatrix} \qquad \ldots(2.2)$$

2-3-1 التحليل العنقودي الحاد او الواضح ***Hard or Crisp Cluster Analysis (HCA)*** *(28) ،(35)، (56)*

تتضمن عملية العنقدة الحادة تجزئة مجموعة البيانات (X) الى (K) من العناقيد بإستخدام نظرية المجاميع التقليدية (classical sets) . وتعرف العناقيد (S_i)على إنها عائلة من المجاميع الجزئية ((P(X)) بحيث $\{S_i | 1 \le i \le K \subset P(X)\}$ والتي تتصف بالخصائص الآتية :

1. $\sum_{i=1}^{K} S_i = X$.

2. $S_i \cap S_j = \phi$ ؛ لكل $1 \le i \ne j \le K$.

3. $\phi \subset S_i \subset X, \quad 1 \le i \le K$.

حيث أن X تمثل مجموعة البيانات وتمثل K عدد العناقيد .

تبين هذه الخصائص أن المجاميع الجزئية S_i تضم كل عناصر البيانات X والتي يجب أن تكون غير مرتبطة مع بعضها ولا تكون فارغة أو تحوي جميع البيانات في عنقود واحد . وتمثل الدوال العضوية (membership function) بالآتي :

(a $\sum_{i=1}^{K} u_i = 1$.

(b $i = 1,....,K; 0 < u_i < 1$.

حيث أن u_i تمثل درجة عضوية العناصر ، أما مصفوفة $U = [u_{ij}]$ ذات البعد (n x k) فتمثل التجزئة الحادة إذا وفقط إذا تحققت الشروط الآتية :

(i $u_{ij} \in (0,1)$ لكل $1 \le j \le n; 1 \le i \le k$.

(ii $\sum_{i=1}^{K} u_{ij} = 1$ لكل $1 \le j \le n$.

(iii $0 < \sum_{j=1}^{n} u_{ij} < n$ لكل $1 \le i \le k$.

حيث أن n تمثل عدد العناصر .

2.3.2 التحليل العنقودي الضبابي Fuzzy Cluster Analysis [28]

[56]، [83]

إن مفهوم عدم التأكد (Uncertian) قد مر بمرحلتين في معالجته ، تتمثل المرحلة الأولى بالنظريات التقليدية وتتمثل المرحلة الثانية بالنظريات الحديثة . فعندما تكون نقاط البيانات (العناصر) موزعة ضمن عناقيد مفصولة جيداً ، عندئذ تتم عملية التصنيف بصورة واضحة وبدون أي مشكلة . في معظم الأحيان لايمكن تجزئة العناصر في مجموعة البيانات الى عناقيد مفصولة جيداً ، عندئذ سوف يكون هناك إعتباطية في تعيين العناصر الى العناقيد الخاصة . فمثلاً إذا كان هناك عنصر (x_j) يقع قرب حدَّي عنقودين لكن بنحو قليل يقترب الى أحد العنقودين (s_i) فيكون من الملائم تحديد الوزن (u_{ij}) الذي يعين درجة إنتماء (x_j) الى العنقود (s_i) .

في معظم الأحيان يتم إستخدام الأساليب الأحتمالية مثل النماذج المختلطة (mixture models) $(u_{ij} = p_{ij})$ لإحتمالية إنتماء العنصر (x_j) الى العنقود (s_i) ولكن عندما تكون هناك صعوبة في تحديد الأنموذج الأحصائي المناسب فيكون من الملائم توفر أساليب عنقودية غير إحتمالية بحيث تقدم الصفات الضبابية نفسها .

تعتمد أساليب العنقدة الضبابية (Fuzzy Clustering Methods) على النظريات الحديثة (modern theory) وتتضمن هذه النظريات :

1. نظرية المجموعات الضبابية Fuzzy Set Theory .
2. نظرية الأمكانية Possibilistic Theory .

قدم (Zadah) عام 1965 نظرية المجموعة الضبابية كطريقة للتعامل مع الغموض (imprecision) وعدم التأكد (uncertainty) . حيث تسمح نظرية المجموعات الضبابية بإنتماء العناصر الى عدة عناقيد في وقت واحد أي الإنتماء جزئياً الى المجموعة مع درجات مختلفة من العضوية . تتميز المجموعة الضبابية A في فضاء العينة R^n بدالة العضوية (membership function) $u_A(x)$ والتي تربط كل عنصر $x \in R^n$ بعدد حقيقي $u_A(x)$ في الفترة المغلقة $[0,1]$ وهذه الدالة تمثل درجة إنتماء العنصر x الى المجموعة الضبابية A ، حيث تأخذ هذه النظرية جميع القيم المحصورة بين الصفر والواحد كدرجة إنتماء للعنصر في المجموعة وليس كما هو الحال في

المجموعة الأعتيادية التي فيها العنصر أما ينتمي الى المجموعة ويأخذ العدد (واحد) أو لاينتمي الى المجموعة فيأخذ العدد (صفر) وعلى النحو الآتي :

$$u(x)=\begin{cases} x^2/\left(x^2+1\right) & ;\ x\geq 0 \\ 0 & ;x<0 \end{cases} \qquad \ldots(3.2)$$

وهنا يمكن إعتبار المجموعة الأعتيادية (الواضحة) ضمن هذا التعريف بأنها المجموعة الضبابية التي تكون لدالة إنتمائها قيمتين 0 و 1 . ويعتمد قياس عدم الوضوح في نظرية المجموعة الضبابية على مقياس الأنتروبي ويتصف هذا المقياس بالخصائص الآتية :

إذا كان U يمثل المجموعة الكلية وأن e يمثل صنف غير خالي(unempty classified) لجميع المجموعات الجزئية الى U فإن قياس الضبابية u في (U,e) هو دالة $u:A\longrightarrow[0,1]$ التي تحقق الشروط الآتية :

1. $u(\phi)=0\ ;\ u(A)=1$.
2. إذا كانت كل من A و B تنتمي الى e وكانت $A\subseteq B$ فان $u(A)\leq u(B)$ لكل $x\in R$.
3. لأية متسلسلة متزايدة $A_1\subset A_2\subset\ldots.$ في e واذا كان $\bigcup_{i=1}^{\infty}A_i\in e$ فإن $\lim_{i\to\infty}u(A_i)=u\left(\bigcup_{i=1}^{\infty}A_i\right)$.
4. لأية متسلسلة متناقصة $A_1\supset A_2\supset\ldots.$ في e واذا كان $\bigcap_{i=1}^{\infty}A_i\in e$ فان $\lim_{i\to\infty}u(A_i)=u\left(\bigcap_{j=1}^{\infty}A_i\right)$.

تعرف مجموعة العناقيد الضبابية $(S_1,S_2,\ldots,S_k)$ بأنها مجموعة جزئية من كل المجاميع الجزئية الضبابية الممكنة لـ X . بمعنى أن أوزان درجة العضوية (u_{ij}) تمتلك نقاط تحديد بين 0 و 1 لكل عنصر x_j وكل عنقود S_i . يتم فرض شروط معينة على مصفوفة التجزئة $U=(u_{ij})$ وعلى النحو الآتي:

1. مجموع كل الأوزان للعنصر المعطى تساوي واحد .

$$\sum_{i=1}^{k} u_{ij} = 1 \qquad ; j = 1,.....,n \qquad \ldots(4.2)$$

2. يحوي كل عنقود (المجموعة الجزئية الضبابية) على الأقل عنصر واحد ، لكن ليس لكل العناصر.

$$0 < \sum_{j=1}^{n} u_{ij} < n \qquad ; i = 1,2,.....,k \qquad \ldots(5.2)$$

$$u_{ij} \in [0,1] \; ; \forall i = 1,...,k \; ; j = 1,....,n \qquad \ldots.(6.2) \qquad .3$$

بمعنى آخر أن كل x_j ممكن أن تنتمي الى أكثر من عنقود واحد مع درجة إنتماء كسرية بين 0 و 1 . والقيم التي تكون قريبة من الواحد تؤشر درجة عالية من العضوية أي درجة إنتماء قوية ، بينما تشير القيم القريبة من الصفر الى درجة عضوية منخفضة . ويؤكد الشرطان (1) و (2) على أن التجزئة الضبابية (fuzzy partition) والعناقيد تشكل ما يطلق عليها بتجزئة بسيدو الضبابية (fuzzy psuedo partition) .

أما نظرية الأمكانية (Possibility theory) فتمثل شكل جديد من نظرية المعلومات التي ترتبط لكن بإستقلالية بالمجموعات الضبابية (fuzzy sets) والنظرية الأحتمالية (probability theory) معاً .

تقنياً يمثل التوزيع الأمكاني (possibility distribution) المجموعة الضبابية الطبيعية (على الاقل درجة عضوية واحدة تساوي واحد) ، فمثلاً أن كل الأعداد الضبابية تمثل توزيعات إمكانية ، وبالتالي فإن الفرق بين التجزئة الأمكانية (Possibility Partition) والتجزئة الضبابية (fuzzy Partition) هو أن مجموع درجات العضوية للعناصر ليس من الضروري أن تساوي 1.

2-21 طرائق التحليل العنقودي Cluster Analysis Methods

إن عملية تصنيف عناصر البيانات ووضعها ضمن تشكيلة معينة من العناقيد تتم وفق إسلوبين، يتمثل الإسلوب الأول إستخدام طرائق العنقدة الحادة (Hard Clustering) والأسلوب الآخر يمثل إستخدام طرائق العنقدة الضبابية (Fuzzy Clustering) وكما هو موضح في المخطط (1.2) الآتي الذي يبين الطرائق الرئيسية للتحليل العنقودي مع الطرائق التي تتفرع منها .

1.4.2 طرائق التحليل العنقودي الحاد Hard Cluster Analysis Methods [35] (HCA) [38]،[80]

1.1.4.2 التحليل العنقودي الهرمي Hierarchical Cluster Analysis

تهدف العنقدة الهرمية الى الحصول على الشكل الهرمي للعناقيد المسماة بالمخطط الشجري (Dendrogram) الذي يبين عملية إرتباط العناقيد بعضها مع بعض من خلال سلسلة متداخلة من التجزيئات بإدماج العناقيد الصغيرة بصورة متكررة الى عناقيد أكبر (طرائق التجميع) أو بفصل عناقيد كبيرة الى عناقيد أصغر (طرائق التقسيم) ، كما هو مبين في الشكل رقم (2.2) والشكل رقم (3.2) .

تتشابه الأساليب الهرمية مع أساليب علم التصنيف (Taxonomy) المستخدمة ضمن مجالات العلوم البيولوجية مثل أقسام العالم الطبيعي (النبات ، الحيوان والجماد) (Kingdom) ، شعبة تصنيف النباتات (Phylum) ، الجينات (Gen) ، ...الخ . في بعض الاحيان يطلق على عناصر الربط الشجري بـ(operational taxonomic units) .

على الرغم من أن معظم الاساليب الهرمية تتضمن إرتباط عنقودين أو فصل عنقود الى عنقودين ، فإن بعض الأساليب الهرمية تربط أكثر من عنقودين في خطوة واحدة أو تفصل عنقود الى أكثر من عنقودين جزئيين .

ولاتفرض الأساليب الهرمية أي عدد خاص من العناقيد وبذلك يمكن الحصول على العدد المقبول من العناقيد بقطع المخطط الشجري عند مستوى معين [79] . ولايكون المخطط الشجري للعنقدة الهرمية وحيد ، حيث أن هناك $2^{(n-1)}$ من المخططات الشجرية المختلفة التي يمكن إنجازها .

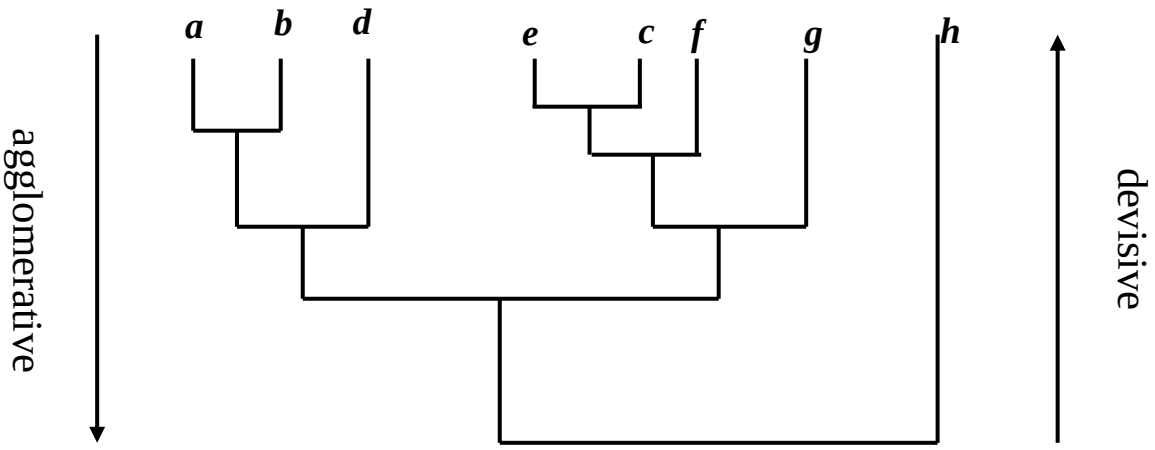
يمكن تلخيص أساليب العنقدة الهرمية بالخطوات الآتية :

1. إيجاد التشابه أو عدم التشابه بين كل زوج من العناصر في مجموعة البيانات من خلال حساب مقياس المسافة .
2. تجميع العناصر الى مخطط شجري (Dendrogram) من العناقيد الهرمية بربط عناقيد مشكلة جيداً .

3. بالنتيجة يتم الحصول على التجزئة النهائية بإختيار مستوى قطع مناسب للشجرة العنقودية الهرمية عند مستوى (q) مثلاً في المتسلسلة $k = n - q - 1$ (تمثل n مجموعة بيانات العينة وتمثل k عدد العناقيد) .

هناك أسلوبان أساسيان لإنشاء العنقدة الهرمية:

1. **أسلوب العنقدة الهرمية التجميعية .**
2. **أسلوب العنقدة الهرمية التقسيمية .**



شكل رقم (2.2) : المخطط الشجري الهرمي التجميعي والتقسيمي للعناصر (a,b,c,d,e,f,g,h) ضمن مجموعة من العناقيد(*) .

شكل (3.2) : يبين خطوات العنقدة الهرمية التجميعية

طرائق العنقدة الهرمية التجميعية أو قوانين الربط (Bottom–Up)

Methods of Agglomerative Hierarchical Clustering or Linkage Rules[35],[38],[80]

تبدأ طرائق العنقدة الهرمية التجميعية بفرض إنتماء كل مشاهدة الى عنقود مفرد (n من المجاميع ذات حجم 1) ، وعند كل خطوة يتم إدماج أقرب زوج من العناقيد حتى الحصول على عنقود واحد فقط يحوي جميع عناصر البيانات . ويمكن تلخيص الخطوات الرئيسة للعنقدة الهرمية التجميعية بالآتي :

لتكن n تمثل مجموعة عناصر البيانات التي سيتم عنقدتها (x_i) الى S من العناقيد .

1. نفرض أن $j=1,...,n; x_j=\{x_j\}, S^*=n$ ، يتم إنشاء مجموعة العناقيد S بتحديد كل عنصر الى عنقوده الخاص $(S=x_1,x_2,...,x_{n-1},x_n)$ وفي حالة أن $\overset{*}{S}\le S$ فيتم التوقف .
2. حساب دالة الأدماج (merging function) الممثلة بإستخدام مقياس التشابه أو المسافة بين كل زوج من العناصر في S لإيجاد أقرب عنقودين .
3. تحديث مقياس التشابه أو المسافة لتعكس مدى التقارب بين العنقود الجديد والعناقيد الأصلية .
4. الرجوع الى الخطوة رقم (2) حتى الحصول على عنقود واحد فقط .

إن نقطة الخلاف بين طرائق العنقدة الهرمية التجميعية تتمثل في تحديد دالة هدف الأدماج (objective function merging) المتمثلة (بدالة المسافة او التشابه) لدمج عدة عناقيد معاً. ولتحديد المسافة بين هذه العناقيد هناك حاجة الى قانون تجميع أو ربط لتحديد موقع ربط عنقودين متشابهين معاً . توجد عدة إمكانات للقيام بعملية الربط هذه ، مثلاً يتم ربط عنقودين عندما يكون هناك أقرب عنصرين أو إستخدام أقرب جوار (nearest neighbor) عبر العناقيد لتحديد المسافات بين العناقيد أو إستخدام أبعد جوار (farthest neighbor) لتحديد **المسافات بين العناقيد والتي يمكن توضيحها على النحو الآتي :**

طريقة الربط المفرد (أقرب جوار)

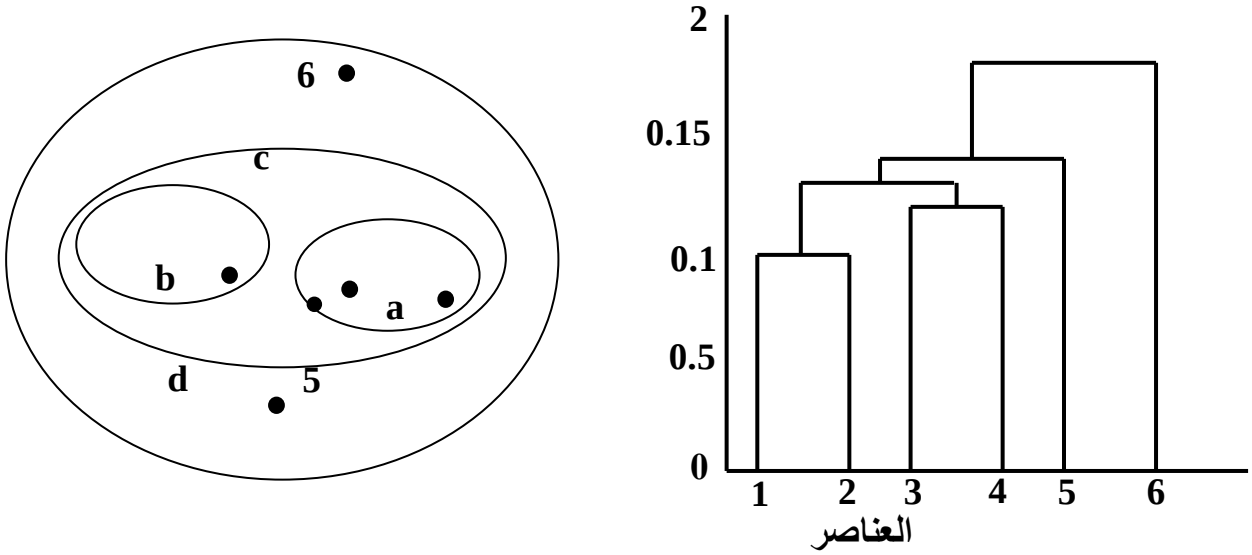
Single Linkage (Nearest Neighbor) Method

في هذه الطريقة يتم تحديد المسافة بين العناقيد بأصغر مسافة (أعظم تشابه) بين أي عنصرين من العناقيد (أقرب جوار) في العناقيد المختلفة ، ويطلق على هذا الأسلوب بطريقة الربط المفرد لأنها تبدأ مع كل النقاط كعناقيد مفردة ومن ثم يتم إضافة الترابطات الأقوى بين النقاط لتجميع العناصر وتشكيل العناقيد ، وتقود العناقيد الناتجة الى سلسلة طويلة من الترابطات كما في الشكل (4.2) .

وتمثل صيغة الربط المفرد بالشكل الآتي :

$$d_{\min}(S_i, S_j) = \min_{\substack{x_i \in S_i \\ x_j \in S_j}} \|x_i - x_j\| \qquad \ldots(7.2)$$

حيث أن $S_i = \{x_1, x_2, ..., x_n\}$ يمثل العنقود الأول و $S_j = \{x_1, x_2, ..., x_m\}$ يمثل العنقود الثاني.



(i) Single Link Clustering **(ii) Single Link Dendrogram**

شكل (4.2) : يبين (i) عنقدة الربط المفرد لمجموعة من العناصر (1,2,3,4,5,6) ضمن العناقيد (a,b,c,d,e) ،

(ii) المخطط الشجري لعنقدة الربط المفرد

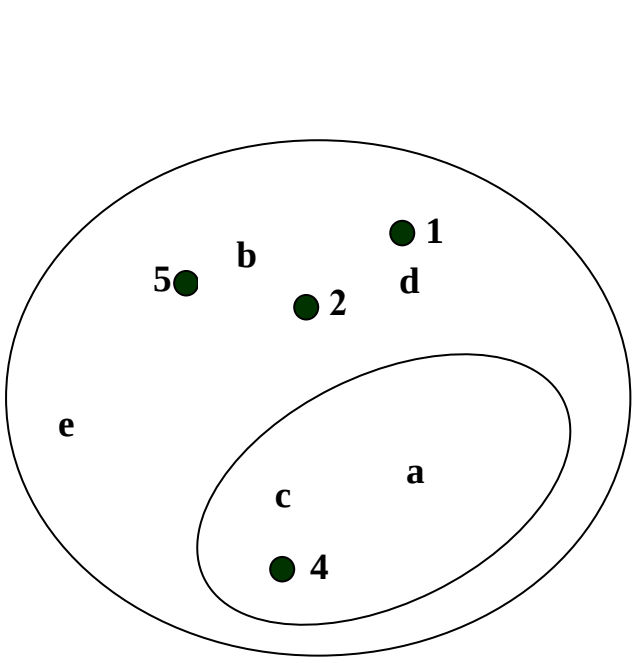
طريقة الربط التام (أبعد جوار)

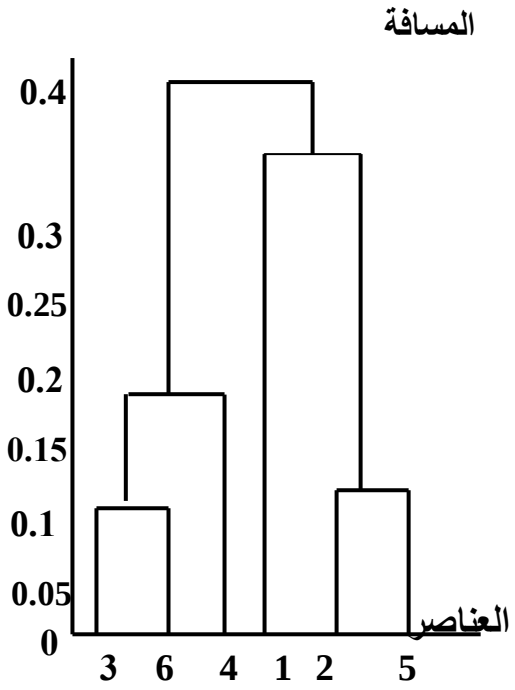
Complete Linkage (Furthest Neighbor) Method

في هذه الطريقة يتم تحديد المسافات بين العناقيد بأكبر مسافة بين أي عنصرين ضمن العناقيد المختلفة (أبعد جوار) . يطلق على هذا الإسلوب بطريقة الربط التام لأنها تبدأ مع كل العناصر كعناقيد مفردة ومن ثم يتم إضافة أقوى إرتباطات بين العناصر ، ولاتمثل مجموعة العناصر عنقود إلا بربط جميع العناصر ربطاً تاماً وتشكيل الكتل (clumps) الواضحة كما هو مبين في الشكل (5.2) . وتمثل صيغة الربط التام بالشكل الآتي :

$$d_{\max}(S_i, S_j) = \max_{\substack{x_i \in S_i \\ x_j \in S_j}} \|x_i - x_j\| \qquad \ldots(8.2)$$

حيث أن $S_i = \{x_1, x_2, \ldots, x_n\}$ ويمثل العنقود الأول و $S_j = \{x_1, x_2, \ldots, x_m\}$ يمثل العنقود الثاني .





(i) Complete Link Clustering **(ii) Complete Link Dendrogram**

شكل (5.2) : يبين (i) عنقدة الربط التام لمجموعة من العناصر (1,2,3,4,5,6) ضمن العناقيد (a,b,c,d,e) ،

(ii) المخطط الشجري لعنقدة الربط التام .

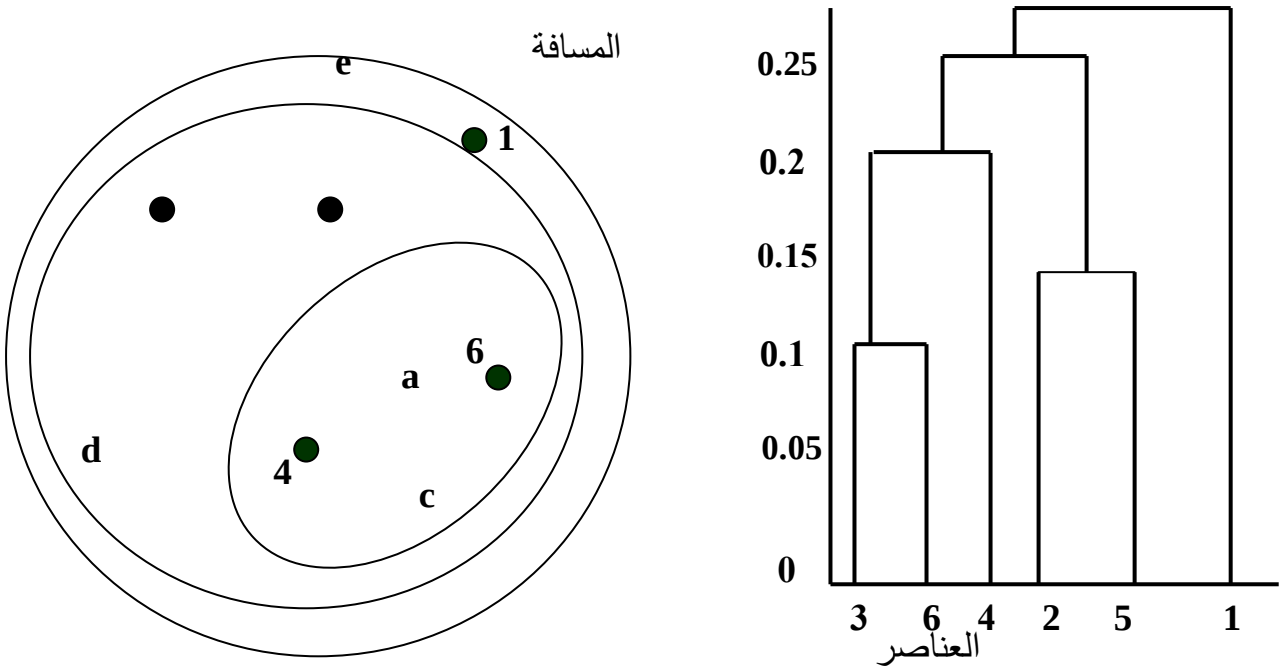
طريقة معدل الربط أو معدل زوج المجموعة

Average Linkage or Pair-Group Average Method

يتم تحديد المسافة بين عنقودين بإستخدام معدل التقاربات (المسافة) الزوجية بين كل أزواج العناصر في العناقيد المختلفة . ويمثل هذا إسلوب متوسط بين طريقتي ربط Min و Max ويتم التعبير عن ذلك بالمعادلة الآتية :

$$\text{Proximity}(s_1, s_2) = \frac{\sum proxinity(x_1, x_2)}{Size(s_1) * Size(s_2)} \quad \text{...(9.2)}$$

حيث أن s_1 و s_2 يمثلان العنقودين الأول والثاني ، ويمكن تمثيل عنقدة معدل الربط بالمخطط الآتي :



(i) Group Average Clustering **(ii) Group Average Dendrogram**

شكل (6.2) : يبين (i) عنقدة معدل المجاميع لمجموعة من العناصر (1,2,3,4,5,6) ضمن العناقيد (a,b,c,d,e) ،

(ii) المخطط الشجري لعنقدة معدل المجاميع .

تقسم طريقة معدل الربط الى نوعين أساسيين :

طريقة معدل زوج المجموعة غير الموزون

Unweighted Pair-Group Average Method

في هذه الطريقة تحسب المسافة بين عنقودين كمعدل مسافة بين كل أزواج العناصر ضمن عنقودين مختلفين . وتكون هذه الطريقة ذات كفاءة عالية عندما تشكل العناصر كتل واضحة طبيعية. وتمثل صيغة المعدل غير الموزون بالآتي :

$$d_{ave.}(S_i, S_j) = \frac{1}{n_i n_j} \sum_{x_i \in S_i} \sum_{x_j \in S_j} |x_i - x_j| \qquad \ldots(10.2)$$

حيث أن S_i و S_j تمثلان العنقودين ith و jth ،وتمثل n_i و n_j عدد العناصر في العنقودين S_i و S_j

طريقة معدل زوج المجموعة الموزون

Weighted Pair-Group Average Method

في طريقة معدل الربط الموزون تعطى المجموعتين أوزان متساوية بغض النظر عن عدد المشاهدات في كل مجموعة عند تحديد المجموعة المجمعة لذلك يتم إستخدم الطريقة هذه عندما تصبح أحجام العناقيد غير متساوية .

طريقة العنقدة المركزية او مركز متوسط زوج المجموعة

Centroid Clustering or Pair-Group Centroid Method

في هذه الطريقة تحسب المسافة بين عنقودين بالإعتماد على مراكز العناقيد وهناك إسلوبين مستخدمين لطريقة العنقدة المركزية هما :

طريقة مركز متوسط زوج المجموعة غير الموزونة

Unweighted Pair-Group Centroid Method

يمثل المركز المتوسط (Centeroid) للعنقود نقطة المعدل في فضاء متعدد الأبعاد . إذ يمثل مركز الثقل للعنقود الخاص وفي هذه الطريقة تحدد المسافة بين عنقودين كإختلاف بين مركزين كما في الصيغة الآتية :

$$d_{mean}(S_i, S_j) = |m_i - m_j| \qquad \ldots(11.2)$$

حيث أن $m_i = \frac{1}{n}\sum x_i$ و $m_j = \frac{1}{m}\sum x_j$ ويمثلان متوسطي العنقودين S_i و S_j بالتعاقب.

في حالة إستخدام الوسيط بدلاً من مراكز العنقود فإن الصيغة تمثل :

$$d_{med}(S_i, S_j) = |med_i - med_j| \qquad \ldots(12.2)$$

حيث أن med_i و med_j يمثلان وسيطي العنقودين S_i و S_j بالتعاقب .

طريقة مركز متوسط زوج المجموعة الموزونة

Weighted Pair-Group Centroid Method

تتشابه هذه الطريقة مع الطريقة السابقة ماعدا أنه هناك أوزان تؤخذ بنظر الأعتبار عند إختلاف أحجام العناقيد (بمعنى عدد العناصر المحتواة فيها) .

طريقة وورد Ward's Method

تفضل هذه الطريقة على بقية الطرائق الهرمية التجميعية السابقة التي أُقترحت من قبل (Ward)[79] عام 1963 والتي يطلق عليها في بعض الأحيان بطريقة أصغر تباين (Minimum Variance Method) لأنها تستخدم إسلوب تحليل التباين لحساب المسافات بين العناقيد ، والمعطاة حسب الصيغة الآتية :

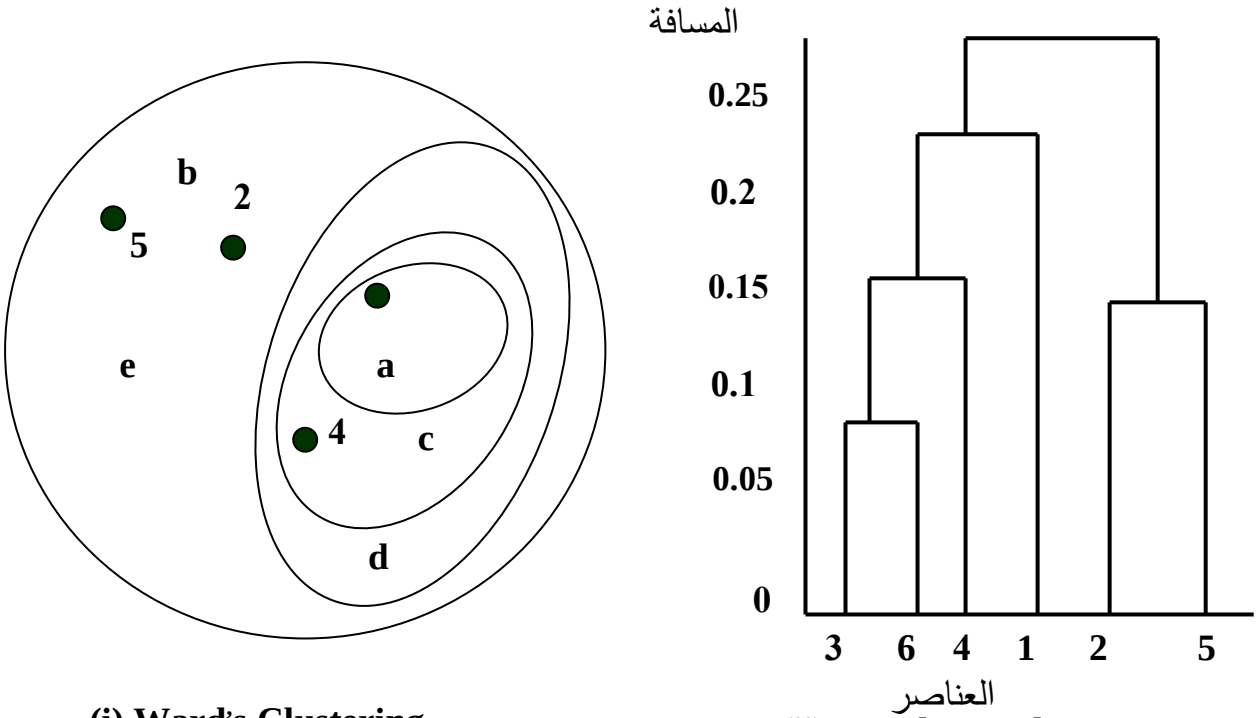
$$d_{ward}(S_i, S_j) = n.m.d_{ij}^2/(n+m) \qquad \ldots\ldots(13.2)$$

حيث أن d_{ij}^2 تمثل المسافة بين (العنقود i) و (العنقود j) المعرفة في طريقة العنقدة المركزية غير الموزونة (معادلة 11.2) و m,n يمثلان عدد العناصر في العنقودين i , j بالتعاقب . وتحاول هذه الطريقة العمل على حساب مجموع مربعات الخطأ بين كل زوج من العناصر ومن ثم العمل على ربط الزوجين الذي يعطي أصغر مجموع مربعات خطأ (SSE) . ويتم حساب (SSE) لأي عنقود حسب الصيغة الآتية :

$$SSE = \sum_{j=1}^{n}\left[\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2\right] \qquad \ldots(14.2)$$

حيث أن n تمثل عدد العناصر الكلية وتمثل n_j عدد العناصر في العنقود jth .

يعرف التقارب بين عنقودين بطريقة (Ward)على أنه الزيادة في مربع الخطأ(Square Error) الناتجة عند إدماج عنقودين ، لذلك تستخدم هذه الطريقة نفس دالة الهدف المستخدمة في عنقدة C– Means . وبإجراء بعض الخطوات البسيطة نلاحظ أن هذه الطريقة تشابه طريقة معدل زوج المجموعة عندما يكون التقارب بين نقطتين محدد ليمثل مربع المسافة بينهما كما في الشكل (7.2) .



(i) Ward's Clustering **(ii) Ward's Dendrogram**

شكل (7.2) : يبين (i) عنقدة Ward لمجموعة من العناصر (1,2,3,4,5,6) ضمن العناقيد (a,b,c,d,e) ،

(ii) المخطط الشجري لعنقدة Ward .

الصيغة التكرارية لـ (Lance–William)

Lance and William's Recurrence Formula

طـور كـلُ مـن (Lance و William)[57] عـام (1967) الصـيغة التكراريـة لطرائـق العنقـدة الهرمية كحالة خاصة من خلال إشتقاق هذه الطرائق بصورة مباشرة من معادلة خاصة ممثلة بالصيغة الآتية :

$$d_{S(ij)} = \alpha_i d_{Si} + \alpha_j d_{Sj} + \beta\, d_{ij} + \gamma \left| d_{Si} - d_{Sj} \right| \qquad \ldots(15.2)$$

حيـث أن d_{ij} تمثـل المسـافة (عـدم التشـابه) بـين العنقـود (i) والعنقـود (j) . وتمثـل $d_{S(ij)}$ المسـافة (عـدم التشـابه) بـين العنقـود (S) والعنقـود الجديـد النـاتج مـن إرتبـاط i و j . ويمثـل كـل مـن $\gamma, \beta, \alpha_j, \alpha_i$ المعالم التي تحدد طبيعة التعنقد بالإعتماد على طريقة التحليل العنقودي الخاصة .

تسمح الصيغة (15.2) عند كل مستوى جديد من العنقدة الهرمية بحساب عدم التشابه بين العنقود الجديد الناتج وبقية العناقيد الحالية المتداولة .

جدول (1.2) : قيم $\alpha_i\ \alpha_j$ ،$\beta\ \gamma$ لطرائق العنقدة الهرمية التجميعية وقيم n التي تمثل عدد المشاهدات في المجموعة i،j،k بالتعاقب[*] .

Clustering Method	α_i	α_j	β	γ
Single Linkage	1/2	1/2	0	1/2
Complete Linkage	1/2	1/2	0	1/2
Average Linkage	$n_i/(n_i+n_j)$	$n_j/(n_i+n_j)$	0	0
Weighte Average Linka	1/2	1/2	0	0
Centroid Linkage	$n_i/(n_i+n_j)$	$n_j/(n_i+n_j)$	$-\alpha_i\alpha_j$	0
Median Linkage	1/2	1/2	−1/4	0
Ward's Linkage	$\frac{n_i+n_S}{n_i+n_j+n_S}$	$\frac{n_j+n_S}{n_i+n_j+n_S}$	$\frac{-n_S}{n_i+n_j+n_}$	0

مرادفات الطرائق الهرمية التجميعية

Synonyms Methods of Hierarchical Agglomerative

هناك عدة مرادفات (synonyms) أخرى لطرائق العنقدة الهرمية التجميعية تم وضعها من قبل الباحثين بمسميات مختلفة ، حيث أن إختلاف هدف الدراسة ومجال تطبيق التحليل العنقودي الهرمي التجميعي أدى الى تنوع مسميات هذه الطرائق المختلفة . وقدم كلُ من (Blashfield and Aldenderfer) [19] عام 1978 جدول من المصطلحات المقابلة لطرائق التحليل العنقودي الهرمي التجميعي كما هو مبين في الجدول رقم (2.2) الآتي :

جدول (2.2) : يبين مرادفات الطرائق العنقودية الهرمية التجميعية[*] .

Single Linkage	Weighted–Average Linkage
–Nearest–Neighbor Method	–Weighted Pair–Group Method using Arithmetic Average.
–Minimum Method	–Weighted Group Average Method.
–Hierarchical Analysis	**Centroid Linkage**
–Space–Contracting Method	–Unweighted Centroid Method.
–Elementary Linkage Analysis	–Unweighted Pair–Group Centroid Method.
–Connectedness Method	–UPGMC.
Complete Linkage	–Nearest–Centroid Sorting.
–Furthest–Neighbor Method.	**Median Linkage**
–Maximum Method.	–Gower's Method.
–Compact Method.	–Weighted Centeroid Method.
–Space–Distarting Method.	–Weighted Pair–Group Centroid Method.
–Space–diloing space–dilaing method.	–Weighted Pair Method.
–Rank Order Typal Analysis.	–Weighted Group Method.
–Diameter Analysis.	**Ward's Method**
Average Linkage	–Minimum–Variance Method.
–Arithmetic Average Clustering.	–Error–Sum of Squares Method.
–Unweighted Pair–Group Method using Arithmetic Average.	–Hierarchical Grouping to Minimize tr(w).
–Unweighted Clustering.	–HGROUP.
–Group Average Method.	
– Unweighted Group Mean.	
–Unweihted Pair–Group Method.	

طريقة العنقدة الهرمية التقسيمية (top–down)

Divisive Hierarichal Method [38]

تبدأ العنقدة الهرمية التقسيمية بإنتماء كل المشاهدات الى عنقود واحد بعد ذلك يتم فصل هذا العنقود الى عنقودين والإستمرار بهذه العملية حتى يتم تثبيت كل المشاهدات في مجموعتها المنفصلة . ويتم فحص العناقيد عند كل مستوى إذ تحاول العنقدة التقسيمية إيجاد أفضل عملية فصل بإفتراض $\left\lfloor 2^{(n-1)} - 1 \right\rfloor$ من المقارنات وتزداد هذه الأعداد أُسياً مع (n) ، وبذلك هناك حاجة لإتخاذ قرار حول أي من العناقيد يجب أن تفصل عند كل خطوة .

تتمثل الفكرة أن العنصر الذي لايفي بشروط العنقود المتواجد فيه يتم فصله عن العنقود ، حيث يكون العنصر الذي لايفي بالشروط متباعد ومختلف عن كل العناصر الأخرى داخل العنقود . وبفرض معدل المسافة بين العنصر j والعنقود S_j يمثل بالآتي:

$$D_{j,S_i} = n_i^{-1} \sum_{x_h \in S_j} d(x_i, x_h) \qquad \ldots(16.2)$$

فإن العنصر الذي لايفي بالغرض يفصل خارجاً ويشكل عنقود جديد . ويطابق هذا فصل العنقود ذات أكبر قطر (largest diameter) ، بمعنى الذي يحوي أكثر زوج مختلف من العناصر .

بعد ذلك عندما يقترب العنصر (h) من عنقود آخر يتم نقله بإستخدام معدل المسافة لكل العناصر في العنقود ، والإستمرار بذلك حتى الوصول الى عناصر غير قابلة للإنتقال .

تستمر الخوارزمية بفصل العنقود ذات أكبر قطر، حتى الحصول على (n) من العناقيد ويمكن تلخيص خطوات خوارزمية العنقدة التقسيمية على النحو الآتي :

1. إختيار العنقود الذي يحوي أكثر زوج مختلف من العناصر والذي يمثل العنقود ذات أكبر قطر .
2. ضمن هذا العنقود يتم إيجاد العنصر ذات أكبر معدل مسافة عن بقية العناصر . بعد ذلك يتم إستبعاد العنصر من العنقود بالسماح له بتشكيل عنقود مفرد جديد .
3. للعنصر h في العنقود الذي تم فصله يتم حساب معدل المسافة بينه وبين العنقود الحالي ومعدل المسافة بين العنصر والعنقود الجديد . في حالة أن المسافة الى العنقود الجديد أقل من المسافة

للعنقود الحالي ، يتم نقل العنصر h الى العنقود الجديد وتستمر الحلقة (Loop) هذه عبر كل العناصر في العنقود .

4. في حالة عدم وجود عناصر قابلة للإنتقال ولكن العدد الحالي للعناقيد أكبر من k يتم الذهاب الى الخطوة (1) ، عدا ذلك يتم التوقف .

تكون أساليب التقسيم أقل شيوعاً وإستخداماً من الأساليب التجميعية بسبب صعوبة حساباتها عند إفتراض كل التقسيمات الممكنة لمجموعتين وسوف يتم تقديم طريقة واحدة حول هذا الأسلوب. لكن بالإمكان دراسة طرائق أخرى تعتمد على الرسومات البيانيه التقاربية مثل طريقة أقرب جوار (Nearest Neighbor) أو أبعد جوار (Furthest Neighbor) بطرائق الرسم البياني .

طريقة التقسيم المبسط (أدنى إمتداد شجري)

Simple Divisive Method (Minimum Spanning Tree (MST) [40], [49]

يطلق على خوارزمية التقسيم المبسط بأدنى إمتداد شجري بسبب إعطاء أصغر توسع شجري ممتد لكل التكتلات التي تعطى من عنقود واحد فقط . حيث تمثل (MST) مجموعة من الحافات (الحدود) edges التي تربط كل العناصر معاً ضمن المجموعة الواحدة مع أصغر مجموع ممكن من أطوال قيم الحافات ، وتمثل قيم الحافات المستخدمة هنا مقياس المسافة المكافئ من نقطة الى نقطة والذي يشابه قوانين أقرب جوار (Nearest Neighbour) .

هناك طرائق مختلفة لتقسيم العناقيد بصورة متتابعة ، إحدى هذه الطرائق هي العمل على إستبعاد أطول حد (largest edge) بين العناصر وإنشاء عنقودين ومن ثم إستبعاد أطول حد حتى الحصول على العدد المقبول من العناقيد .

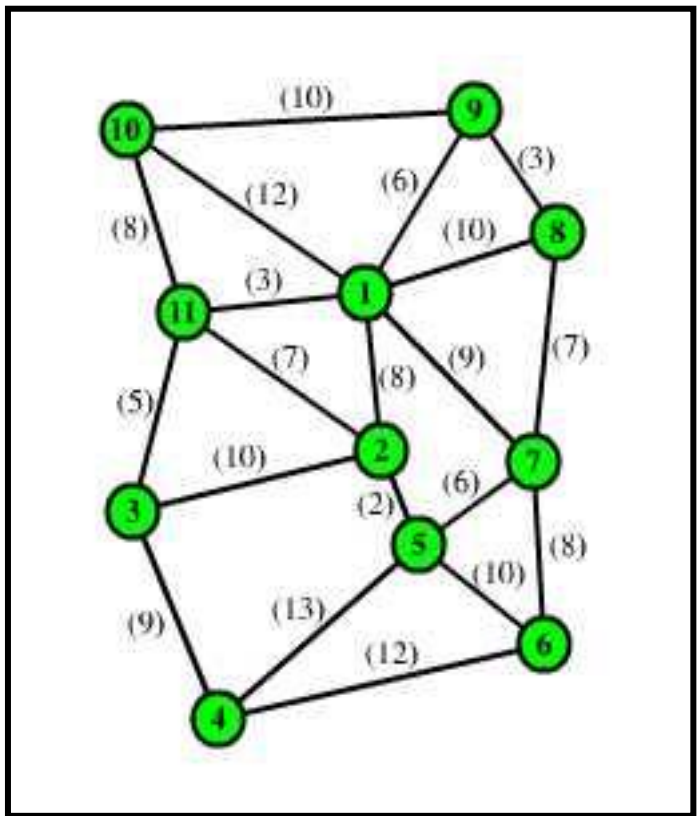
ولإختيار الحد يتم مقارنة طول الحد الى أطوال الحدود الاخرى عند تكتلاتها . حيث يتم تقارب الحد عندما يكون طوله (L) أكبر من (L`) ، حيث أن L` تمثل معدل طول كل الحدود الأخرى العرضية عند تكتلاتها . إن هذا الأسلوب ذو فائدة لتحديد العناقيد التي تملك مسافات مختلفة مفصولة واحدة عن الأخرى .

الإستخدام الآخر لإسلوب أدنى إمتداد شجري هو إيجاد عناقيد كثيفة محددة ضمن مجموعة من نقاط البيانات المبعثرة . ويتم عمل ذلك بإزالة كل الحدود التي تكون أطوالها أكبر من بعض أطوال

الحدود المعاد تحديدها ثانية من أجل إستخلاص العناقيد التي تكون أطوالها أقرب من الطول المحدد لكل عنقود .

ويمكن تلخيص خطوات طريقة MST على النحو الآتي :

1. حساب أدنى إمتداد شجري للرسم البياني التقاربي المبين في الشكل (8.2) من خلال إختيار أي تكتل بصورة عشوائية ووضعه في المجموعة المحددة (مثلا S_1) وجعل التكتلات الأخرى في المجموعة (S_2) .
2. إنشاء عنقود جديد بقطع الترابط المقابل لأكبر مسافة من خلال إختيار أصغر طول حدي L(l,j) من بين جميع حدود المجموعة S_1 وحدود المجموعة S_2 ثم العمل على إضافة هذا الحد الى الأمتداد الشجري ، بعد ذلك يتم إضافة التكتل j (node j) الى المجموعة S_1 وحذفها من المجموعة S_2 .
3. إعادة الخطوة (2) لحين الحصول على عناقيد مفردة فقط .



شكل (8.2) : يبين الإمتداد الشجري لطريقة MST [*] .

طرائق العنقدة غير الهرمية (العنقدة التجزيئية)

Nonhierarchical Clustering (Partitional Clustering) Methods

تنشأ عملية التصنيف بطريقة العنقدة غير الهرمية بتجزئة مجموعة البيانات الى مجاميع جزئية وذلك بتحديد فئة من المجاميع غير المتداخلة التي تحتوي على علاقات غير هرمية فيما بينها.

لاتملك طرائق العنقدة غير الهرمية أشكال تشبه الشجرة (Tree) حيث تتشكل العناقيد بإتباع إسلوب العنقدة التجريئية (Partitional Clustering) للحصول على تجزئة مفردة لمجموعة من العناصر الى عناقيد بالإعتماد على الأمثلية التكرارية (iterative optimization) للدالة المعيارية (دالة الهدف) التي تعكس (القبول) agreement أو التوافق بين عناصر البيانات وعملية تجزئتها .

تعتمد طرائق التجزئة التي تستخدم مربع الخطأ (squared error) الى تحديد عدد العناقيد (K) وتمثيل كل عنقود بنموذج أولي (Prototype) ومن ثم العمل على تصغير دالة الهدف التي تمثل مجموع كل عناصر البيانات لمربع المسافة بين العنصر والنموذج الأولي للعنقود المحدد لها . وغالباً ما تمثل النماذج الأولية هذه مراكز العناقيد (cluster centers) كما في خوارزمية (K-Means) [15] ، طريقة (K-Medoid) [53] ، وطريقة (ISODATA) [15] حيث تمثل النموذج الأولي للعنقود بالمركز (central) .

طريقة عنقدة C من المتوسطات الحادة أو الواضحة

Hard or Crisp C-Means Clustering Method

تختلف طريقة العنقدة هذه عن الطرائق الهرمية ، حيث تتوافق طريقة (K-Means) المقترحة من قبل (Ball and Hall) [15] بطريقة (C-Means) والتي يطلق عليها اختصاراً بـ (HKM) أو (HCM) . حسابياً تشابه هذه الطريقة إسلوب تحليل التباين(ANOVA) analysis of variance ولكن بصورة عكسية . حيث تبدأ الطريقة مع K من العناقيد العشوائية ومن ثم تحريك (نقل) العناصر بين هذه العناقيد مع العمل على تصغير الإختلافات داخل العناقيد وتعظيم الإختلافات بين العناقيد وهذا يشبه عمل (ANOVA) ولكن بالعكس ، حيث أن إختبار المعنوية في (ANOVA) يحسب

الاختلاف بين المجموعات ضد الاختلاف داخل المجموعة عند إختبار معنوية الفرضية حيث تختلف المتوسطات ضمن المجموعة بعضها عن بعض .

تبدأ خوارزمية (HCM) بتحديد عدد ثابت من العناقيد وإختيار مراكز العناقيد الأولية عشوائياً. من ثم يتم تعيين كل البيانات الى أقرب مركز عنقودي بدرجة عضوية مساوية الى (1) وعضوية (0) الى العناقيد الأخرى ، بعد ذلك يتم إختيار مقياس المسافة الملائم .

تعتمد طريقة (HCM) على تصنيف عناصر مصفوفة البيانات (X) ذات البعد (nxp) الى (K) من العناقيد لتصغير مجموع مربعات المسافات بين البيانات $(x_j ;\ j=1,2,....,n)$ ومراكز العناقيد $(c_i ; i=1,2,.....,k)$ والتي تعمل بصورة غير مباشرة على تصغير مجموع المربعات داخل العنقود المتمثلة بدالة الهدف $[J(U,C)]$ (Objective Function) :

$$Min\ J(U,C:X)=\sum_{i=1}^{k}\sum_{j=1}^{n}(u_{ij})\ \|x_j-c_i\|^2 \qquad \ldots(17.2)$$

حيث أن

X: تمثل مصفوفة البيانات الأصلية ذات البعد (nxp) ، وتمثل n عدد العناصر وتمثل p عدد الخواص .

C: مصفوفة مراكز العناقيد c_i ذات البعد (kxp) ، حيث تمثل k عدد العناقيد .

U: مصفوفة التجزئة (partition matrix) أو مصفوفة العضوية (membership matrix) للتجزئة الحادة ذات البعد (kxn) والتي تمثل درجة عضوية (x_j) الى العنقود (ith) والتي تتحدد بالتوليد العشوائي المنتظم (uniform distribution) وعلى النحو الآتي :

$$U=u_{ij}=\begin{cases}1 & if\ \ x\in k\\ 0 & if\ \ x\notin k\end{cases} \qquad \ldots(18.2)$$

إن مصفوفة التجزئة تمتلك شروط معينة هي :

1. $u_{ij}\in\{0,1\}\ ;i=1,2,.....,k\ ,j=1,2,.....,n$.
2. $0<\sum_{j=1}^{n}u_{ij}<n$.

وتمثل $\|x_j - c_i\|$ المسافة بين عناصر البيانات x_j ومراكز العناقيد c_i .

يمكن تلخيص خطوات خوارزمية C–Means على النحو الآتي :

بفرض أن $X = (x_{1p}, x_{2p}, ..., x_{np})$ تمثل مصفوفة مجموعة من البيانات التي يتم تصنيفها الى k من العناقيد وتمثل هذه العناقيد موجهات (nx1) ممكن إختيارها عشوائياً من عناصر البيانات الكلية أو تمثل أول العينات k أو تمثل العينات المستندة على بعض المعلومات السابقة . ويتم تثبيت معيار التوقف $\in$ (termination tolerance) لطريقة العنقدة مثلاً 0.01 و 0.001 وهكذا .

1. توليد مصفوفة التجزئة $U = \lfloor u_{ij} \rfloor$ لكل $(j = 1,...., n \ , i = 1,2,....., k)$ توليداً قياسياً منتظماً .
2. إختيار مراكز العناقيد $(C_1, C_2, ..., C_k)$ حيث :
   i) يحدد كل عنقود بواسطة المركز المتوسط (Centroid) .
   ii) تكون عدد المراكز مساوية الى عدد العناقيد النهائية .

   ويمكن حساب مراكز العناقيد حسب المعادلة الآتية :

$$c_i = \frac{\sum_{j=1}^{n} u_{ij} . x_j}{\sum_{j=1}^{n} u_{ij}} \quad ; \ 1 \le i \le k \qquad ...(19.2)$$

3. تعيين كل عناصر البيانات الى أقرب مركز متوسط (أقرب جوار) بإستخدام مقياس المسافة الملائم ، مثلاً في حالة إستخدام المسافة الأقليدية فإن عناصر x_j للعنقود c_i تحدد على أساس أن :

$$|x_j - c_r| < |x_j - c_i| \qquad \begin{matrix} for\ all\ \ i, r = 1,2,....,k \\ j = 1,2,...,n \ ; r \ne i \end{matrix} \qquad ...(20.2)$$

4. إعادة حساب المراكز لكل عنقود ، حيث يتم تحديد مراكز العناقيد الجديدة بحساب معدل عناصر البيانات في كل عنقود كما في الصيغة الآتية :

$$c_i = \frac{\sum_{j=1}^{n_i} x_j}{n_i} \quad ; \ x_j \in A_j \qquad ...(21.2)$$

حيث أن A_j تمثل مجموع العناصر في العنقود ith وتمثل n_i عدد العناصر في A_j .

5. إعادة تحديد أعضاء العنقود $U = \lfloor u_{ij} \rfloor$ لتصغير مربع الخطأ بين عناصر البيانات ومراكز العناقيد الحالية :

$$u_{ij} = \begin{cases} 1 & ; \ if \ \ d(X_j, C_i) = \min_{1 \le r \le k} (X_j, C_r) \\ 0 & ; otw. \end{cases} \qquad \ldots(22.2)$$

في بعض الأحيان قد تنشأ مشكلة الحصول على عناقيد فارغة (empty cluster) إذا لم يتم إنتماء أي من العناصر الى هذه العناقيد مما يتطلب إتخاذ خطوات لإختيار مراكز بديلة . إحدى هذه الخطوات هي إختيار العنصر الذي يقع بعيداً عن المركز الحالي ليمثل المركز الجديد ، أما الخطوة الأخرى فتمثل إختيار المركز البديل من العنقود الذي يملك أعلى (SSE) . وفي حالة تواجد أكثر من عنقود فارغ فيتم إعادة هذه الخطوات عدة مرات .

6. إعادة تكرار الخطوات السابقة حتى الوصول الى التقارب المتمثل بحالة عدم التغير في مراكز العناقيد بمعنى إذا تحقق :

$$C_i^{(l+1)} = C_i^{(l)} \qquad ; \ \forall i = 1,.....,k \qquad \ldots(23.2)$$

أو أن الخوارزمية تتوقف في حالة أن التغيرات في مصفوفة التجزئة تكون قليلة جداً أي عندما :

$$\prod_{j=1}^{n} \max \left\| U^{(l)} - U^{(l-1)} \right\| < \in \qquad \ldots(24.2)$$

حيث أن l تمثل عدد التكرارات ويمثل $\in$ معيار التوقف .

طريقة عنقدة (ISODATA) ISODATA Clustering Method

تعرف طريقة (ISODATA)[38] بطريقة تحليل بيانات التنظيم الذاتي التكراري (Iterative A Organizing Data Analysis Techniques) ، وهي طريقة أخرى من طرائق التحليل العنقودي غير الهرمي المقترحة من قبل (Ball–Hall)[15] عام 1965 لحل مشكلة العنقدة عندما تكون عدد العناقيد غير معلومة . وتتضمن طريقة (ISODATA) مجموعة القواعد لدمج وفصل العناقيد المستخدمة من أجل الحصول على العنقدة النهائية .

وتلخص طريقة عنقدة (ISODATA) بالخطوات الآتية :

1. إختيار K من نقاط البيانات (العناصر) المفصولة جيداً كمراكز تقريبية أولية .

2. تشكيل العناقيد K بتصنيف كل عنصر من عناصر البيانات (n) الى أحد العناقيد k (بتعيين كل العناصر الى أقرب مركز عنقودي) بإستخدام مقياس أقل مسافة ، مثلاً يتم تعيين العنصر jth الى العنقود S_a في حالة أن :

$$d_{aj} = \left[\left(x_j - c_a \right)' \left(x_j - c_a \right) \right]^{\frac{1}{2}} = \min_{b=1,....,k} \left[\left(x_j - c_b \right)' \left(x_j - c_b \right) \right]^{\frac{1}{2}} \quad ...(25.2)$$
$$; j = 1,...., n \; ; i = 1,....., k$$

حيث أن x_j يمثل العنصر j في مجموعة البيانات n وتمثل c_a و c_b مركزي العنقودين a و b بالتعاقب .

3. إعادة حساب مراكز كل العناقيد

$$c_i = \frac{\sum_{j=1}^{n_i} x_j}{n_i} \quad ...(26.2)$$

حيث أن n_i تمثل عدد عناصر العنقود ith .

4. الإستمرار بالعملية حتى الحصول على ثبات في مراكز العناقيد

$$\prod_{j=1}^{n} \max \left| c^l - c^{l-1} \right| < \varepsilon \quad ...(27.2)$$

حيث أن l تمثل عدد التكرارات .

5. تتم عملية إزالة العناقيد الكثيرة ذات العناصر القليلة جدا بدمجها معاً عملية الدمج (Merging) . أما في حالة تواجد عناقيد قليلة فيتم إستخدام عملية الفصل (Split) ، وتتم عملية فصل وإدماج العناقيد بالإعتماد على معايير محددة مثل :

i) للعنقود s_r يتم حساب معدل المسافة $\left(\bar{d}_r \right)$ بين عناصر بيانات داخل العناقيد ومراكزها:

$$\bar{d}_r = \frac{1}{n_r} \sum_{x_r \in s_r} \| x_r - c_r \| \quad ...(28.2)$$

حيث أن n_r تمثل عدد العناصر في العنقود r وتمثل x_r عناصر العنقود وتمثل c_r متوسط العنقود r .

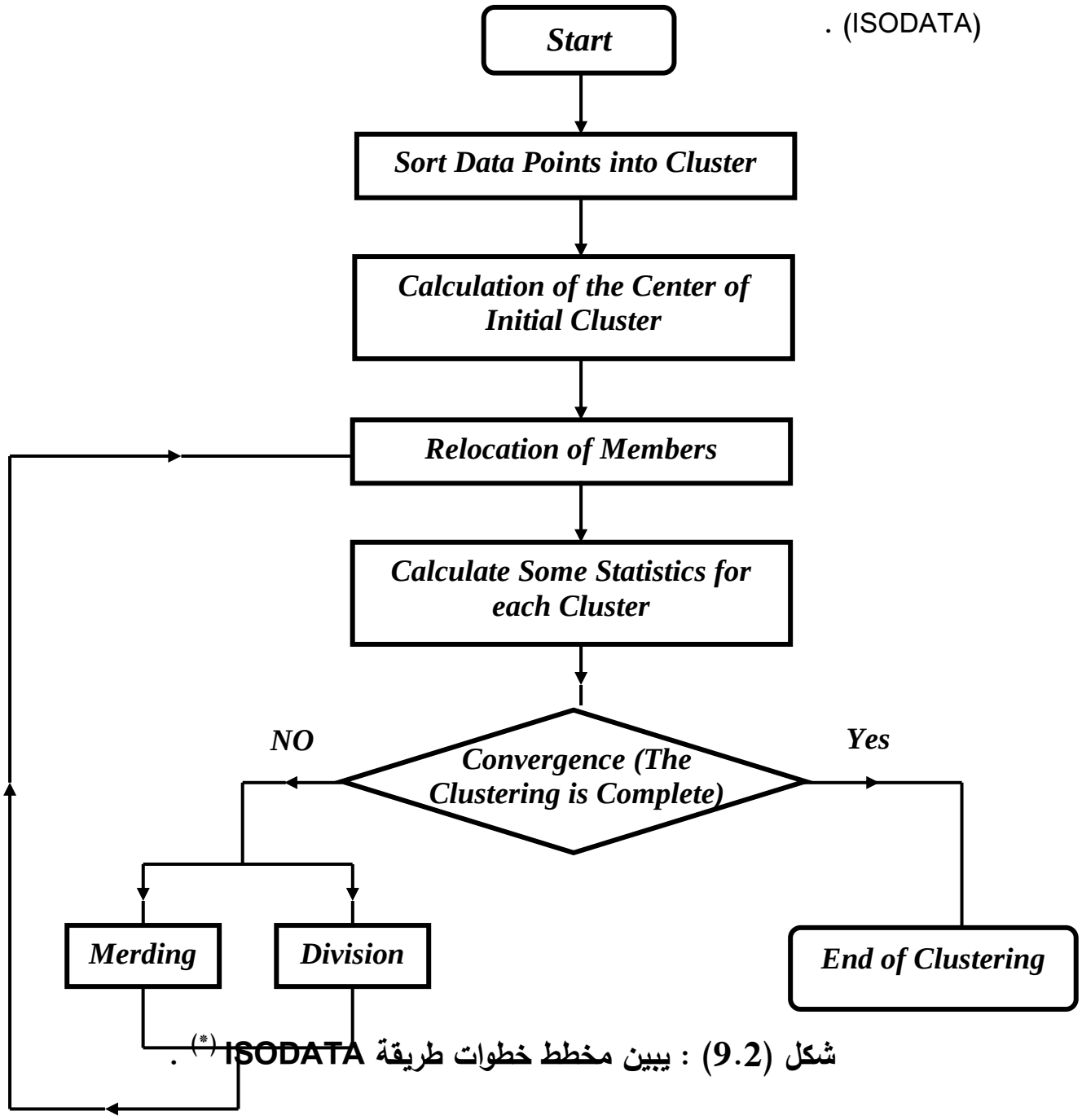
ii) لكل عنقود S_r يتم حساب كمية الانتشار σ_r (مصفوفة التباين المشترك ذات البعد (pxp)) حسب الصيغة الآتية :

$$\sigma_r = \frac{1}{n_r - 1}\sum_{S=1}^{n_r}(x_{rs} - c_r)(x_{rs} - c_r)' \qquad \ldots(29.2)$$

iii) معدل المسافة بين العناقيد والمعرفة حسب الصيغة الآتية :

$$\bar{d} = \frac{1}{n}\sum_{r=1}^{n_l} n_r \bar{d}_r \qquad \ldots(30.2)$$

وعند إجراء عملية الفصل يتم إختيار العنقود الذي يملك أكبر إنحراف معياري لفصله الى عنقودين . أما عند إجراء عملية الإدماج فيتم إدماج العنقودين ذات أقرب مركزين أو إدماج العنقودين اللذين ينتجان أصغر زيادة في (SSE)عند إدماجهما . ويبين الشكل (8.2) مخطط خطوات طريقة (ISODATA) .



شكل (9.2) : يبين مخطط خطوات طريقة ISODATA (*) .

طريقة عنقدة K–Medoids K–Medoids Clustering Method

تعمل طريقة (K–Medoid) المقترحة من قبل (Kaufman–Rousseeuw)[53] عام 1987 على إيجاد مجموعة غير متداخلة من العناقيد من خلال تجزئة فضاء المسافة الى K من العناقيد ، بحيث يملك كل عنقود عناصر التمثيل أو العناصر النموذجية التي يتم إختيارها وتعيينها مركزياً من مجموعة البيانات ، والتي يطلق عليها بـ(Medoids) . وأكثر طريقة شائعة هي طريقة التجزئة حول Medoid (Partitioning Around Medoids Algorithm) أو ما يطلق عليها إختصاراً (PAM) وتستخدم عندما تكون كمية البيانات كبيرة جداً .

تتمثل دالة هدف طريقة (K–Medoid) بتجزئة مجموعة البيانات X الى K من العناقيد ، وهي تشبه طريقة (K–Means) من حيث إستخدام المدخلات والمخرجات الأساسية ، والإختلاف الرئيسي بينهما يتمثل في حساب مراكز العناقيد ، حيث يمثل مركز العنقود الجديد أقرب نقطة بيانات (عنصر) الى متوسط نقاط العناقيد . ويمكن تلخيص خطوات طريقة (K–Medoid) على النحو الآتي :

1. لمجموعة البيانات X يتم إختيار K من العناصر النموذجية الأولية (medoids) في K من العناقيد .
2. حساب المسافات بين العناصر في العناقيد المختارة (medoids) .

$$d_{ij} = \left[\left(x_j - v_i \right)' \left(x_j - v_i \right) \right]^{\frac{1}{2}} \quad ;1 \le i \le K \quad ;1 \le j \le n \qquad \ldots(31.2)$$

حيث أن x_j تمثل عدد عناصر n وتمثل v_i الـ medoids .

3. إختيار العناصر المنتمية الى العنقود الواحد التي تملك أقل مسافة .
4. حساب مراكز العناقيد الجديدة .

$$v_i^* = \frac{\sum_{j=1}^{n_i} x_j}{n_i} \qquad \ldots(32.2)$$

حيث أن n_i تمثل عدد العناصر داخل العنقود .

5. إختيار أقرب نقطة بيانات لتمثل مركز العنقود حسب الصيغة الآتية :

$$d_{ij}^* = \left[\left(x_j - v_i^* \right)' \left(x_j - v_i^* \right) \right]^{\frac{1}{2}} \quad ; \; j = 1,\ldots,n ; i = 1,\ldots,k \ldots(33.2)$$

حيث أن v_i^* تمثل الـ (medoid) المستخرجة من المعادلة (33.2) بعد ذلك يتم حساب

$$x_i^* = \arg\min_i \left(d_{ij}^*\right);\ v_i^{(1)} = x_i^* \qquad \ldots(34.2)$$

6. الإستمرار بالعملية حتى الحصول على الثبات :

$$\prod_{j=1}^{n} \max \left| v^{(l)} - v^{(l-1)} \right| \neq 0 \qquad \ldots(35.2)$$

حيث أن l تمثل عدد التكرارات .

طريقة متوسطات – K الأنشطارية Bisecting K–Means Method

إن طريقة متوسطات K الانشطارية [74] هي إمتداد لطريقة متوسطات K الأساسية (K–Means Method) والتي تعتمد على الفكرة الآتية :

للحصول على العناقيد K يتم فصل مجموعة عناصر البيانات الكلية (X) الى مجموعتين جزئيتين (عنقودين) ، من ثم يتم إختيار أحد هذين العنقودين والعمل على فصله الى عنقودين آخرين والأستمرار بهذه العملية حتى الحصول على K من العناقيد . ويمكن إيجاز خطوات طريقة متوسطات K الانشطارية على النحو الآتي:

1. إنشاء عدد ملائم من العناقيد لتحوي العنقود الذي سوف يضم جميع عناصر البيانات .
2. تكرار العملية وإختيار عنقود من مجموعة العناقيد المنشئة .
3. العمل على فصل العنقود المحدد بإستخدام طريقة متوسطات C الحادة الأساسية (HCM) .
4. إضافة العنقودين الناتجين من عملية الفصل والتي تملك أصغر مجموع مربعات للخطأ (SSE) الى العناقيد المنشئة .
5. الإستمرار بخطوات العملية حتى إحتواء العناقيد المحددة على K من العناقيد .

هناك عدة أساليب مختلفة لإختيار العنقود الذي سوف يتم فصله (شطره) في الخطوة رقم (2) ، مثلاً يمكن إختيار أكبر عنقود عند كل خطوة أو إختيار العنقود ذات أقل تشابه كلي أو إستخدام الإسلوب المعتمد على الحجم والتشابه الكلي معاً .

2.4.2 طرائق التحليل العنقودي الضبابي Fuzzy Cluster Analysis Method

1.2.4.2 طريقة متوسطات C – الضبابية Fuzzy C–Means Algorithm (FCM)

تمثل طريقة متوسطات–C الضبابية طريقة تصنيف غير موجهة (Unsupervised Classification) بإنتمائها الى صنف العنقدة التجزيئية (Clustering Partition) ، إذ أنها أُشتقت من طريقة متوسطات –C الحادة غير الهرمية ، حيث وسع (Dunn)[32] طريقة (HCM) بتوسيع مفهوم التجزئة الأعتيادية للعناصر الى التجزئة الضبابية وذلك بالسماح لعناصر البيانات بالإنتماء الى جميع العناقيد مع درجات عضوية تقع ضمن الفترة [0،1] .

تهدف طريقة عنقدة متوسطات –C الضبابية (FCM) الى تصغير دالة الهدف المسماة دالة متوسطات C والمعرفة حسب الصيغة الآتية :

$$J_2(X:U,C)=\sum_{i=1}^{k}\sum_{j=1}^{n}(u_{ij})^2\ \|x_j-c_i\|^2 \qquad \ldots(36.2)$$

حيث أن X : تمثل مصفوفة مجموعة البيانات ذات البعد (nxp) ، حيث تشير n الى عدد العناصر وتشير P الى الأبعاد (عدد خواص البيانات number of features) .

وتمثل C : مصفوفة $C=[c_1,c_2,....,c_k]$; $c_i\in R^n$ ذات البعد (kxp) والتي تشير الى مراكز العناقيد (center of cluster) .

وتمثل U : المجموعة الشاملة لكل التجزيئات الممكنة لمجموعة البيانات X وتبين درجة إنتماء عضوية العناصر الى العناقيد وتقع تحت ثلاثة شروط :

1. $\sum_{i=1}^{k}u_{ij}=1$ لكل $j=1,2,....,n$.
2. $u_{ij}\in[0,1]$ ؛ لكل $j=1,2,....,n$ ؛ $i=1,2,...,k$.
3. $0<\sum_{j=1}^{n}u_{ij}<n$ لكل $i=1,2,...,k$.

وتمثل $\|x_j-c_i\|^2$: المسافة الأقليدية المربعة المحسوبة بين مجموعة البيانات x_j ومراكز العناقيد c_i والممثلة بالصيغة الآتية :

$$d_{ij}^2 = \|x_j - c_i\|_A^2 - (x_j - c_i)' A (x_j - c_i) \qquad \ldots(37.2)$$

حيث أن A تمثل مصفوفة أحادية (Identity Matrix) لذلك يمكن كتابة المسافة الأقليدية المربعة على النحو الآتي :

$$d_{ij}^2 = \|x_j - c_i\|^2 = (x_j - c_i)'(x_j - c_i) \qquad \ldots(38.2)$$

بعد ذلك عمم (Bezdek)[17] عام 1980 دالة هدف Dunn $[J_2(U,C)]$ الى عدد غير محدد من دوال الهدف بمعنى $[J_m(U,C)]$ حيث ان $1 < m < \infty$ والذي يطلق عليه بمعامل التضبيب والذي يوجه عضوية مشاركة العناصر في العناقيد الضبابية أي يحدد ضبابية العنقود ، أيضاً يطلق عليه في بعض الاحيان بمركب الأس الموزون (weighting exponent) . عندما (m=1) فإن $\{u_{ij} = (0,1)\}$ أي أن مصفوفة $(U = u_{ij})$ تمثل تجزئة حادة ، وعندما $(m = \infty)$ فإن $\{u_{ij} = 1/|C|\}$ أي أن درجة العضوية متساوية لكل العناقيد وأفضل إختيارعندما (m=2) .

تمثل دالة الهدف الجديدة لطريقة (C–Means) أو (K–Means) والتي يطلق عليها في بعض الأحيان بطريقة (ISODATA) الضبابية (Fuzzy ISODATA Method) نفس شروط تجزئة (C–Mean) الضبابية في المعادلة (36.2) وعلى النحو الآتي :

$$J_m(X : U, C) = \sum_{i=1}^{k} \sum_{j=1}^{n} (u_{ij})^m \|x_j - c_i\|^2 \qquad \ldots(39.2)$$

ويمكن تلخيص خطوات طريقة (FCM) على النحو الآتي :

1. نفرض أن $X = (x_1, \ldots, x_n)$ تمثل مجموعة عناصر البيانات ، يتم إختيار عدد العناقيد والتي يرمز لها (k) ؛ $2 \le K \le n$ ، وإختيار قيمة عنصر الأس الموزون $(m > 1)$ ومن ثم إختيار معيار التوقف (termination tolerance) والذي يمثل معيار إنتهاء الخوارزمية $(\varepsilon > 0)$.
2. تكوين مصفوفة التجزئة الأولية $U^{(0)}$ وهي قيم عشوائية متمثلة بمصفوفة ذات بعد (nxk) تولد بتوزيع قياسي منتظم مع الأخذ بنظر الاعتبار الشروط الثلاثة المقيدة بها :
   - $\sum_{i=1}^{k} u_{ij} = 1$.
   - $u_{ij} \in [0,1]$.

- $0 < \sum_{j=1}^{n} u_{ij} < n$ ، لكل $i = 1,....,k$ ؛ $j = 1,....,n$.

3. حساب النماذج الأولية (Prototypes) الممثلة بمراكز العناقيد وفقاً للمعادلة الآتية :

$$c_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m} \quad ; \forall\ i = 1,...., k \qquad \ldots(40.3)$$

حيث أن u_{ij} تمثل عناصر مصفوفة العضوية و x_j يمثل موجه عناصر البيانات ويمثل m معامل التضبيب .

4. حساب مقياس المسافة بين مجموعة البيانات X ومراكز العناقيد C والممثلة بالمسافة الأقليدية القياسية الآتية :

$$d_{ij} = \left[(x_j - c_i)' (x_j - c_i) \right]^{\frac{1}{2}} \ ; \ i = 1,2,...., k \ ; \ j = 1,2,...., n \qquad \ldots(41.2)$$

5. تحديث مصفوفة التجزئة الضبابية $U = (u_{ij})$ وفقاً للمعادلة الآتية :

$$u_{ij} = \begin{cases} \dfrac{1}{\sum_{r=1}^{k} \left[\dfrac{\|x_j - c_i\|}{\|x_j - c_r\|} \right]^{\frac{1}{(m-1)}}} ; & if \ \ d_{ij} > 0 \ for \ i = 1,..., k \ ; \ j = 1,...., n \\ 0 \qquad\qquad\qquad ; & otw. \end{cases} \qquad \ldots(42.2)$$

أما في حالة أن $d_{ij} = 0$ فان قيمة u_{ij} تمثل على النحو الآتي:

$$u_{ij} = \begin{cases} 1 & ; \ for \, Some \ i \in I \subseteq \{1,2,.....,k\} \\ 0 & ; \ for \ i \notin I \end{cases} \qquad \ldots(43.2)$$

6. إختبار التقارب إذا كان معيار التوقف يحقق المتباينة الآتية :

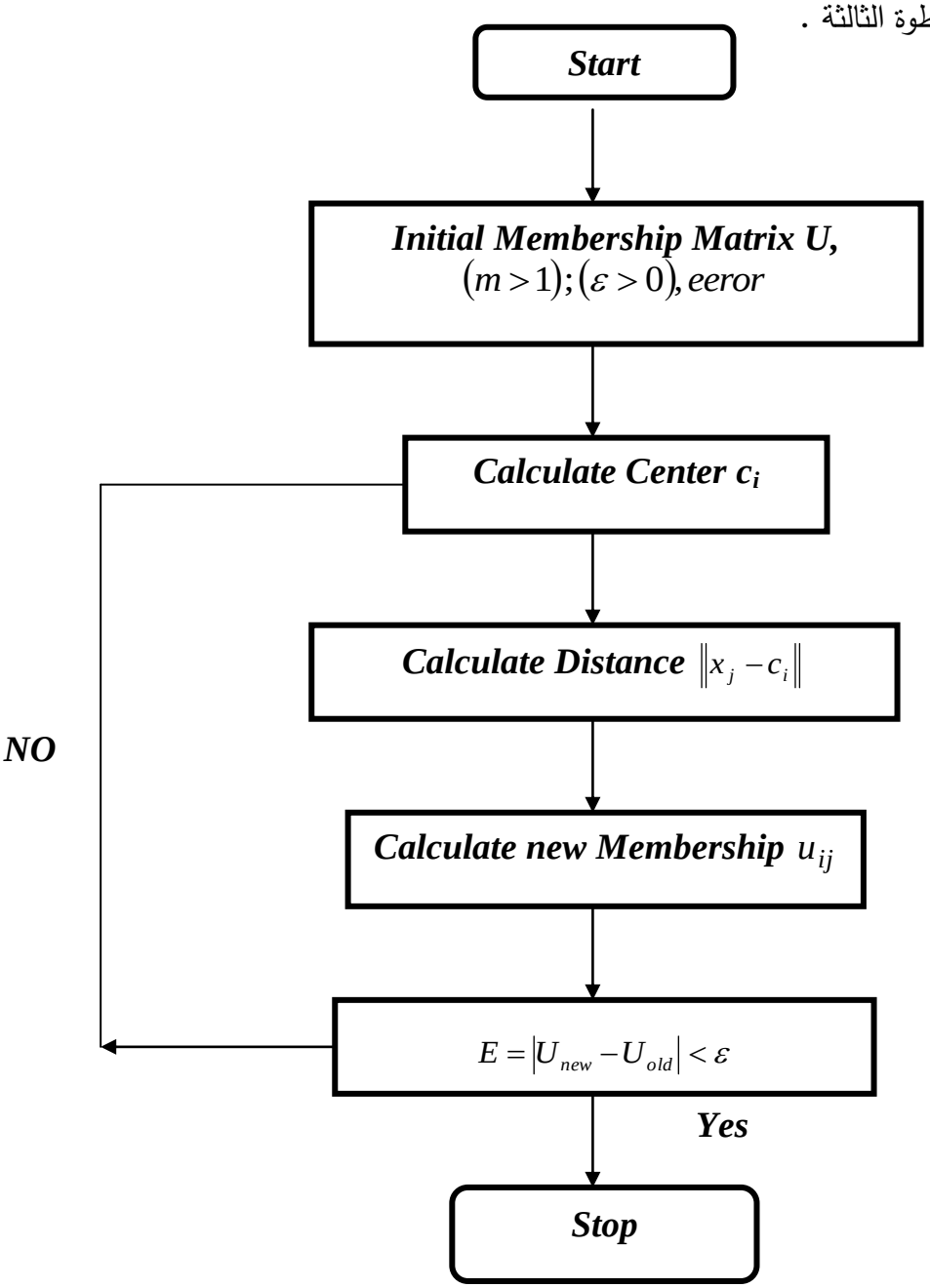
$$\left\| U^l - U^{l-1} \right\| < \varepsilon \qquad \ldots(44.2)$$

Or

$$\left\| C^{l+1} - C^l \right\| < \varepsilon \qquad \ldots(45.2)$$

حيث أن l تمثل التكرار و ε يمثل معيار التوقف $(\varepsilon > 0)$.

أو يمكن وقف تكرار الطريقة عند تصغير دالة الهدف والوصول الى حالة الثبات . عدا ذلك يتم العودة الى الخطوة الثالثة .



شكل (10.2) : يبين مخطط خطوات خوارزمية (FCM)

طريقة Gustafson – Kessel الضبابية

Fuzzy Gustafson – Kessel Method (GK)

تشبه طريقة عنقدة (GK) الضبابية الى حد ما طريقة (FCM) ، الإ أنه في طريقة (FCM) تكون العناقيد ذات شكل كروي(spherical cluster) ولاتسمح للعنقود بتغيير شكله بالأعتماد على البيانات لذلك إقترح كلُ من (Gustafson–Kessel)[41] طريقة تمثل إمتداد لطريقة (FCM) بحيث يمكن تكييف العنقود الى أشكال مفرطة البيضوية (hyperellipsoid shape) من أجل تكييف أشكال مختلفة من البيانات وقد تم إستخدام مصفوفة التغاير(covariance matrix) لإستخراج الخصائص البيضوية للعناقيد .

تستخدم طريقة (GK) مسافة مهلنوبس المبينة في المعادلة الآتية :

$$d_{ij}^2 = (x_j - c_i) A_i (x_j - c_i)' \qquad \ldots(46.2)$$

حيث تمثل d_{ij} المسافة وتمثل c_i مركز العنقود i وتمثل x_j موجه البيانات j ، ولخوارزمية (GK) يمكن تعريف A_i حسب المعادلة الآتية :

$$A_i = P_i . \det(F_i)^{\frac{1}{n}} F_i^{-1} \qquad \ldots(47.2)$$

حيث أن n تمثل عدد البيانات وتمثل P_i حجم العنقود ith الذي يجب إمتلاكه ، وعادة يساوي واحد وتمثل F_i^{-1} معكوس مصفوفة التغاير F_i للعنقود i .

تعمل دالة الهدف(objective function) لطريقة (GK) على تصغير المسافات من مراكز العناقيد الى نقاط البيانات ويتم التحكم بالمسافات المعطاة في المعادلة (46.2) بمصفوفة مسافة مهلنوبس التي تتوافق مع مقياس خواص البيانات عند ذلك يصبح حجم العناقيد أصغر ، لذلك تكون هناك بعض القيود الضرورية على حجم العناقيد وعلى A_i . وتعرف دالة الهدف طريقة (GK) على النحو الآتي :

$$J(X;U,V,A) = \sum_{i=1}^{k} \sum_{j=1}^{n} (u_{ij})^m \, d_{ij}^2 \qquad \ldots(48.2)$$

حيث أن

X : تمثل مصفوفة مجموعة البيانات ذات البعد (nxp) .

U : تمثل مصفوفة التجزئة التي تعمل على تصنيف وتجزئة مجموعة البيانات x_j الى العناقيد c_i بعضويات مختلفة مع التقيد بالشروط الثلاثة الآنفة الذكر في طريقة (FCM) .

A : تمثل المسافة الأقليدية المبينة في الصيغة (41.2) .

d_{ij}^2 : تمثل مصفوفة مسافة مهلنوبس المبينة في المعادلة (46.2) .

تحدد المتجهات الآنية (eigenvectors) والقيم الآنية (eigenvalues) لمصفوفة (A) شكل وحجم وموقع العنقود ith . وتتشكل العناقيد ذات الأشكال مفرطة البيضوية (hyperellipsiod clusters) بتوجيه المتجهات الآنية والحصول على الأطوال والمقاييس الناتجة من جذر n'th للقيم الآنية وعندما تكون كل المتجهات الآنية متعامدة بعضها لبعض ، فيتم الحصول على الشكل التقريبي لحجم العنقود من جذر n'th لقيمها الذاتية . ومن الممكن حساب محددة المصفوفة A من ناتج القيم الآنية في A ، لذلك تستخدم $\sqrt[n]{\det(A_i)}$ لتقدير حجم العنقود ith .

ويمكن تلخيص خطوات طريقة Gustafson–Kessel (GK) على النحو الآتي :

1. بفرض مجموعة البيانات $x_j = \{x_1,....,x_n\}$ ، يتم إختيار عدد العناقيد k $(2 \le k \le n)$ وإختيار معامل التضبيب (الأس الموزون) m (m>1) ومن ثم تحديد معيار التوقف $(\varepsilon > 0)$.
2. توليد مصفوفة التجزئة $U = (u_{ij})$ بصورة عشوائية توليداً قياسياً منتظماً مع الأخذ بنظر الأعتبار الشروط الثلاثة السابقة .
3. حساب مراكز العناقيد (النماذج الأولية) Cluster Center (Prototype) وفقاً للصيغة الآتية:

$$c_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m} \quad ; i = 1,....,k \qquad \ldots(49.2)$$

4. حساب مصفوفات تغاير العناقيد حسب الصيغة الآتية :

$$F_i = \frac{\sum_{j=1}^{n} (u_{ij})^m (x_j - c_i)(x_j - c_i)'}{\sum_{j=1}^{n} (u_{ij})^m} \quad ; i = 1,....,k \qquad \ldots(50.2)$$

5. حساب مسافة مهلنوبس على النحو الآتي :

$$d_{ij} = \left\{ (x_j - c_i)' \left[P_i \det(F_i)^{\frac{1}{n}} F_i^{-1} \right] (x_j - c_i) \right\}^{\wedge(1/2)} \qquad \ldots(51.2)$$

6. تحديث مصفوفة التجزئة $U = \lfloor u_{ij} \rfloor$ وفقاً للصيغة الآتية :

$$u_{ij} = \frac{(d_{ij})^{\frac{1}{m-1}}}{\sum_{r=1}^{k} [(d_{rj})]^{\frac{1}{m-1}}} \quad ; i = 1,....,k ; j = 1,....,n \qquad \ldots(52.2)$$

7. يتم الإستمرار بتكرار الخطوات السابقة حتى الحصول على :

$$E = \left\| U^l - U^{l-1} \right\| < \varepsilon \qquad \ldots(53.2)$$

حيث أن l تمثل عدد التكرارات .

طريقة Gath–Geva الضبابية Fuzzy Gath–Geva Method

تجمع طريقة (Gath–Geva) الضبابية (GG) المقترحة من قبل كلُ من (Gath–Geva)[37] عام 1989 بين طريقة متوسطات C – الضبابية (FCM) وتقديرات الأمكان الأعظم الضبابية Fuzzy Maximum Liklehood Estimate (FMLE) . وتعمل طريقة (GG) على تصغير المسافات المربعة الموزونة (sum of weighted squared distances) بين نقاط البيانات $(x_j ; j = 1,....,n)$ ومراكز العناقيد $(c_i ; i = 1,....,k)$.

$$J(X;U,C) = \sum_{i=1}^{k} \sum_{j=1}^{n} (u_{ij})^m (d_{ij})^2 \qquad \ldots(54.2)$$

حيث أن

X : تمثل مصفوفة البيانات ذات البعد (nxp) .

U : تمثل مصفوفة التجزئة والتي تمثل الاحتمالية اللاحقة (posterior probability) عند إختيار العنقود ith .

C: تمثل مصفوفة مراكز العناقيد ، ويمثل m معامل التضبيب $(m>1)$.

d_{ij} : تمثل المسافة المحسوبة في خوارزمية (GG) والممثلة بالصيغة الآتية :

$$d_{ij}^{\,2}=\frac{\sqrt{\det(F_{wi})}}{P_i}\exp\left[\frac{1}{2}(x_j-c_i)'F_{w_i}^{-1}(x_j-c_i)\right] \qquad \ldots(55.2)$$

أما F_{wi} تمثل مصفوفة التغاير الضبابية المحسوبة وفقاً للصيغة الآتية :

$$F_{wi}=\frac{\sum_{j=1}^{n}(u_{ij})^w(x_j-c_i)(x_j-c_i)'}{\sum_{j=1}^{n}(u_{ij})^w};\quad 1\le i\le k \qquad \ldots(56.2)$$

بحيث أن w المركب الأُسي (w>1 وعلى الأغلب يساوي 2) ، بينما في خوارزمية (FMLE) الأصلية (w=1) . والفرق بين مصفوفة (F_i) في خوارزمية (GK) ومصفوفة (F_{wi}) المعرفة أعلاه هو أن F_{wi} لاتتضمن الأس الموزون (m) ولكن تتضمن بدلاً عنه (w) لتصبح التجزئة أكثر ضبابية لكي تكافئ الجانب الأسي (Exponential Term) لمقياس المسافة . وبالتالي التمكن من إكتشاف عناقيد ذات أشكال وأحجام وكثافات مختلفة .

وتمثل (Pi) الإحتمالية السابقة (Prior Probebility) لإختيار العنقود i والمعطاة حسب الصيغة الآتية :

$$P_i=\frac{1}{n}\sum_{j=1}^{n}u_{ij} \qquad \ldots..(57.2)$$

حيث أن n تمثل عدد العناصر الكلية في مجموعة البيانات .

ويمكن تلخيص خطوات طريقة (GG) الضبابية على النحو الآتي :

1. بفرض مجموعة البيانات $(x_j;\ j=1,\ldots..,n)$ ، يتم إختيار الأس الموزون (m>1) ومعيار التوقف $(\varepsilon>0)$.
2. توليد مصفوفة التجزئة $U=\lfloor u_{ij}\rfloor$.
3. حساب مراكز العناقيد وفقاً للصيغة الآتية :

$$c_i=\frac{\sum_{j=1}^{n}(u_{ij})^w x_j}{\sum_{j=1}^{n}(u_{ij})^w};\quad i=1,2,\ldots..,k \qquad \ldots(58.2)$$

4. حساب مقياس المسافة (d_{ij}) وفقا للصيغة (55.2) .

5. تحديث مصفوفة التجزئة حسب المعادلة الآتية :

$$u_{ij} = \frac{\left(\frac{1}{d_{ij}}\right)^{\frac{1}{m-1}}}{\sum_{r=1}^{k}\left(\frac{1}{d_{rj}}\right)^{\frac{1}{m-1}}} \quad ; \ 1 \le i \le k, 1 \le j \le n \qquad \ldots(59.2)$$

6. الإستمرار بتكرار الطريقة حتى الحصول على

$$E = \left\| U^{l} - U^{l-1} \right\| < \varepsilon \qquad \ldots(60.2)$$

حيث أن l تمثل عدد التكرارات .

2- 22 الطريقة المقترحة (طريقة مسافة الأرتباط)

من أجل الحصول على الحل الأمثل لمشكلة التغير الكبير بأشكال العناقيد والكثافات المبينة في الطرائق الضبابية الثلاثة السابقة الذكر تم تطوير إستراتيجة جديدة للعنقدة الضبابية من خلال إيجاد مدى قوة الترابط بين عناصر البيانات (x_j ; j=1,.....,n) وبين مركز العناقيد(c_i ;i=1,....,k) المصنفة اليها . وتتم هذه العملية بالإعتماد على إستخدام مسافه الأرتباط (Correlation Distance) بدلاً من المسافه الأقليدية المربعة في طريقة (FCM) ومسافة مهلنوبس في طريقة (GK) والمسافة المختلطة في طريقة (GG) وتعمل الطريقة المقترحة على تصغير المسافات المربعة الموزونة بين عناصر البيانات ومراكز العناقيد من خلال تصغير داله الهدف (Objection Function) قدر الأمكان وبأقل تكرارات ممكنة :

$$MIN\ J(X;U,C) = \sum_{i=1}^{k}\sum_{j=1}^{n}(u_{ij})^{m}(d_{ij})^{2} \qquad \ldots..(61.2)$$

حيث أن X تمثل مصفوفة البيانات ذات البعد (n x p) . وتمثل U مصفوفة التجزئة ، ويمثل (m>1) m معامل التضبيب ، وتمثل (d_{ij}) مسافة الارتباط بين عناصرالبيانات المستخدمة (x_j) ومراكز العناقيد (c_i) المصنفه اليها .

ويمكن تلخيص خطوات الطريقة المقترحة كما هو مبين في الشكل (11.2) وعلى النحو الآتي :

1. تحديد مصفوفة البيانات X ذات البعد (n x p) المتضمنة العناصر (x_j ; j=1,.....,n) وإختيار الأس الموزون (معامل التضبيب) m>1 وكذلك تحديد معيار التوقف $(\varepsilon \succ 0)$.
2. توليد مصفوفة التجزئة U=(u_{ij}) ، مع الأخذ بنظر الأعتبار الشروط الثلاث التي ذكرت في الصيغ (4.2) و(5.2) و(6.2) .
3. حساب مراكز العناقيد وفقاً للصيغة الآتية :

$$c_I = \frac{\sum_{j=1}^{n} (u_{ij})^m . x_j}{\sum_{j=1}^{n} (u_{ij})^m} \qquad ;i = 1,\ldots\ldots\ldots,k \qquad \ldots\ldots(62.2)$$

4- حساب مقياس (Cosine) والذي يمثل الارتباط بين عناصر البيانات (x_j) ومراكز العناقيد (c_i) وفقاً للصيغة الآتية :

$$\delta_{ij} = \frac{\sum_{d=1}^{p} (x_{jd} - \bar{x}_j)(c_{id} - \bar{c}_i)}{\sqrt{\sum_{d=1}^{p} (x_{jd} - \bar{x}_j)^2 . \sum_{d=1}^{p} (c_{id} - \bar{c}_i)^2}} \qquad \ldots\ldots(63.2)$$

حيث أن $\bar{c}_i, \bar{x}_j$ تمثلان المركز المتوسط لعناصر البيانات ومراكز العناقيد وبالتعاقب والممثلان بالصيغتين الآتيتين :

$$\bar{x}_j = \frac{\sum_{d=1}^{p} x_{jd}}{p} \qquad \ldots\ldots\ldots(64.2)$$

$$\bar{c}_i = \frac{\sum_{d=1}^{p} c_{id}}{p} \qquad \ldots\ldots\ldots\ldots(65.2)$$

ومن ثم يتم حساب مسافة الأرتباط ($d_{ij} = 1 - \delta_{ij}$) .

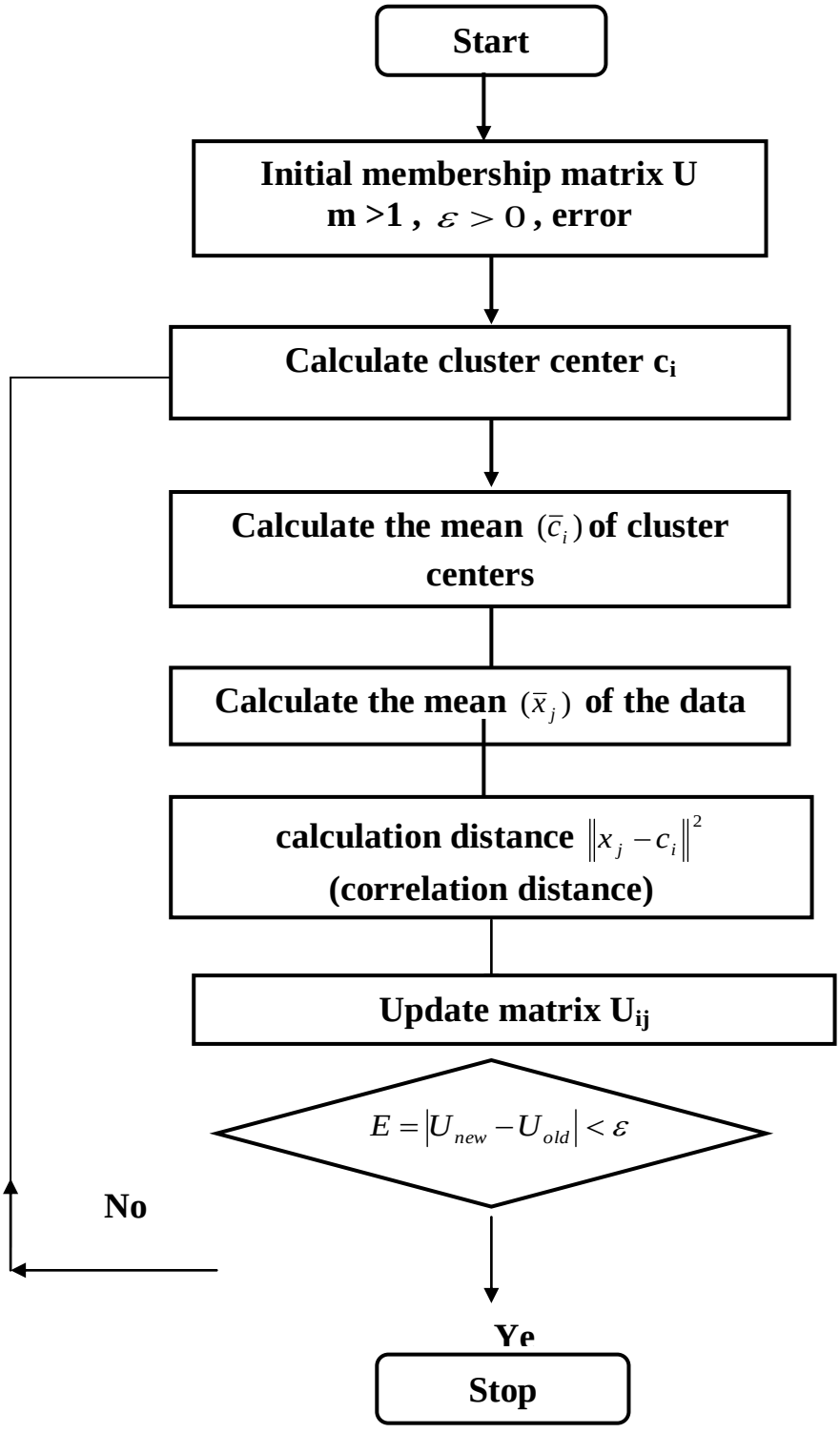
5- تحديث مصفوفة التجزئة (u_{ij}^*) حسب المعادلة الآتية :

$$u_{ij}^* = \frac{(\frac{1}{d_{ij}})^{\frac{1}{m-1}}}{\sum_{r=1}^{k} (\frac{1}{d_{rj}})^{\frac{1}{m-1}}} \qquad \begin{matrix} ;i = 1,\ldots\ldots,k \\ j = 1,\ldots\ldots,n \end{matrix} \qquad \ldots.(66.2)$$

6- الإستمرار بتكرار الطريقة حتى الحصول على أصغر دالة هدف أو الحصول على

$$E = \left\| U^l - U^{l-1} \right\| < \varepsilon \qquad \ldots\ldots(67.2)$$

حيث أن l تمثل عدد التكرارات .



شكل (11.2) : يبين خطوات خوارزمية الطريقة المقترحة

2- 23 طرائق صحة العنقدة Clustering Validation Methods

تعتبر عملية إختبار تشابه العناصر والنقاط داخل العنقود ومن ثم التنبؤ ومعرفة العدد الأمثل للعناقيد من مشاكل التصنيف غير الموجه الرئيسية حيث أن معظم طرائق العنقدة تتطلب الإجابة على أسئلة مطروحة عند تطبيق طرائق العنقدة مثل :

- ❖ هل العناقيد المحددة منسجمة مع المعلومات السابقة للمشكلة ؟
- ❖ هل العناقيد المحددة متطابقة مع البيانات جيداً ؟
- ❖ هل توجد تجزئة جيدة لمجموعة البيانات ؟
- ❖ هل النتائج المستخرجة بطريقة ما أفضل من تلك المستخرجة بالطريقة الأخرى ؟
- ❖ ما هو العدد الصحيح من العناقيد ؟

وللإجابة على هذه الأسئلة وحل تلك المشاكل تم إيجاد طرائق قانونية (Validity Methods) تحدد شكل التجزئة العنقودية للوصول الى الشكل العنقودي الأمثل ، وقد أطلق عليها بطرائق صحة العنقدة أو صحة نتائج العنقدة (Clustering Validation Methods) .

تعتبر طرائق صحة العنقدة طرائق مهمة في تحليل العنقدة حيث تعمل على إثبات صحة نتائج العنقدة في أغلب التطبيقات من خلال مقارنة كفاءة النتائج لطريقة عنقدة معينة بالنسبة الى باقي نتائج طرائق العنقدة الأخرى أو المقارنة مع نفس الطريقة بإستخدام قيم معالم مختلفة .

تم إقتراح بعض من طرائق صحة إختيار العدد الصحيح والمثالي للعناقيد من قبل الباحثين خلال السنوات الماضية والتي يمكن تصنيفها الى واحدة من ثلاثة أصناف :

- ❖ تعتمد بعض طرائق صحة وقانونية العنقود على إستخدام الشكل الحاد(Hard Shape) لمجموعة البيانات المستخدمة في خوارزميات العنقدة لقياس صحة نتائج تجزئة العنقدة .
- ❖ في حالة طرائق العنقدة الضبابية فإن بعض الطرائق القانونية تستخدم نتائج مصفوفة درجات العضوية الضبابية (fuzzy membership degrees) لحساب صحة نتائج العنقدة .
- ❖ أما الصنف الثالث من الطرائق القانونية فيعمل على إستخدام شكل مجموعة البيانات الى جانب إستخدام درجات العضوية الضبابية .

1.6.2 مقاييس صحة العنقدة للتحليل العنقودي الحاد

Clustering Validation for Hard Cluster Analysis

1.1.6.2 مقياس صحة العنقدة لـ Dunn (D)

يهدف هذا المقياس المقترح من قبل (Dunn)[33] عام 1974 الى تشخيص مجموعة العناقيد المدمجة والمفصولة جيداً (Compact and Well Separated Cluster) والذي يمكن تعريفه بالشكل الآتي :

لأي عملية تجزئة للبيانات لتكن $u \leftrightarrow s : s_1 u \ldots\ldots s_i u \ldots\ldots s_k$ حيث ان s_i تشير الى العنقود ith عند التجزئة U في R^n ، يعرف القطر Δ *diameter* للمجموعة الجزئية s_r (العنقود rth) بالآتي:

$$\Delta(s_r) = \max_{x_i, x_j \in s_r} \left| d(x_i, x_j) \right| \qquad \ldots(68.2)$$

حيث أن $d(x_i, x_j)$ تمثل المسافة بين عناصر المجموعة S_r .

وأن المسافة بين المجموعتين الجزئيتين s_i و s_j تعرف كالآتي :

$$\delta(s_i, s_j) = \min_{x_i \in s_i, x_j \in s_j} \left| d(x_i, x_j) \right| \qquad \ldots(69.2)$$

بعد عملية تجزئة العناقيد يستخدم مقياس (Dunn) والذي يعرف بالصيغة التالية :

$$D = \min_{i=1,\ldots.k} \left[\min_{\substack{j=1,\ldots.k \\ i \neq j}} \left\{ \frac{\delta(s_i, s_j)}{\max_{r=1,\ldots.k} (\Delta(s_r))} \right\} \right] \qquad \ldots(70.2)$$

حيث أن k تمثل عدد العناقيد .

إن الهدف الأساسي من هذا المقياس هو تعظيم المسافة بين العناقيد inter cluster $(\delta(s_i, s_j))$ وتصغير المسافة بين عناصر العنقود intra cluster $(\Delta(s_r))$. لذلك فإن القيم الكبيرة لمقايس (D) تدل على وجود عناقيد جيدة ، وبالتالي فإن عدد العناقيد التي تعظم المقياس (D) تعتبر أمثل عدد للعناقيد .

2.1.6.2 مقياس صحة العنقدة لـ Calinski–Harabasz (CH)

يعرف مقياس (CH) المقترح من قبل (Harabasz and Calinski)[20] عام 1974 لـn من العناصر و k من العناقيد بالصيغة الآتية :

$$CH = \frac{[trace\ B/(k-1)]}{[trace\ W/(n-k)]} \qquad \ldots(71.2)$$

حيث تمثل B و W مصفوفات الإنتشار بين وداخل العنقود وبصورة متعاقبة .

ويعرف أثر مصفوفة إنتشار العناصر بين العنقود B بالصيغة الآتية :

$$trace\ B = \sum_{i=1}^{k} n_i \ \|c_i - c\| \qquad \ldots(72.2)$$

حيث أن n_i تمثل عدد العناصر في العنقود ith وتمثل c_i مركز العنقود ith وتمثل c مركز مجموعة البيانات الكلية .

أما أثر مصفوفة إنتشار العناصر داخل العنقود W فيمثل على النحو الآتي:

$$trace\ W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \|x_j - c_i\| \qquad \ldots(73.2)$$

ينتج من المعادلتين أعلاه أن مقياس (CH) يمكن كتابته حسب الصيغة الآتية :

$$CH = \left[\frac{\sum_{i=1}^{k} n_i \ \|c_i - c\|^2}{k-1} \right] \Big/ \left[\frac{\sum_{j=1}^{k} \sum_{j=1}^{n_i} \|x_j - c_i\|}{n-k} \right]$$

وتعكس القيم الكبيرة للمقياس القانوني (CH) التجزئة الصحيحة للبيانات وبالتالي تعكس العدد الصحيح للعناقيد .

3.1.6.2 مقياس صحة العنقدة لـ Davies–Bouldin (DB)

كما في مقياس صحة العنقدة لـ (Dunn) يهدف مقياس (Davies–Bouldin)[25] في عام 1979 الى تحديد مجموعة العناقيد المدمجة والمفصولة جيداً ولكن يختلف عنه في أنه يفرض إستخدام حالة معدل الخطأ (Average Error Case) لكل عنقود .

إن مقياس (DB) هو دالة لقياس نسبة إنتشار العناصر داخل العنقود (within cluster scatter) الى الفصل بين العناقيد (between cluster separation) . ويحسب إنتشار العناصر (s_i) داخل العنقود ith بالصيغة الآتية :

$$v_i = \frac{1}{|n_i|}\sum_{x \in s_i}\left[\|X - c_i\|\right] \qquad \ldots(74.2)$$

حيث أن s_i تمثل العنقود ith و c_i تمثل مركز العنقود ith و n_i تمثل عدد العناصر في العنقود ith . أما المسافة d_{ij} بين العنقودين s_i و s_j فتعرف بالشكل الآتي:

$$d_{ij} = \|c_i - c_j\| = \sum_{r=1}^{p}|c_{ri} - c_{rj}| \qquad \ldots(75.2)$$

من ثم يتم تحديد مقياس التشابه R_{ij} بين العناقيد s_i و s_j بالإعتماد على إنتشار العنقود s_i عند تحقق الشروط الآتية :

1. $R_{ij} \geq 0$.
2. $R_{ij} = R_{ji}$.
3. اذا $S_i = 0$ و $S_j = 0$ فان $R_{ij} = 0$.
4. اذا $S_j > S_r$ و $d_{ij} = d_{ir}$ فان $R_{ij} > R_{ir}$.
5. اذا $S_j = S_r$ و $d_{ij} < d_{ir}$ فان $R_{ij} < R_{ir}$.

تبين الشروط أعلاه ان R_{ij} تكون متماثلة وغير سالبة ويمكن حسابها حسب الصيغة الآتية :

$$R_{ij} = (v_i + v_j)/d_{ij} \qquad \ldots(76.2)$$

وبالتالي فإن مقياس (DB) يمكن حسابه حسب المعادلة التالية :

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i \qquad \ldots(77.2)$$

حيث أن k تمثل عدد العناقيد وأن :

$$R_i = \max_{\substack{j=1,\ldots..k \\ j \neq i}} R_{ij} \qquad ; i = 1,2,\ldots..k \qquad \ldots(78.2)$$

وللحصول على نتائج تحليل عنقودي جيدة وعدد أمثل من العناقيد يتم العمل على تصغير قيمة (DB) قدر الأمكان .

مقياس صحة العنقدة الظلي Silhouette Validity Measure

يحسب المقياس القانوني المقترح من قبل (Rousseeuw)[68] عام 1987 العرض الظُلي (Silhouette Width) لكل عنصر، ومعدل العرض الظُلي لكل عنقود والمعدل الكلي للعرض الظُلي لمجاميع البيانات . يستخدم هذا المقياس لحساب صحة أو قانونية العنقدة ولمقارنة العدد الأمثل من العناقيد المختارة من خلال تمثيل كل عنقود بما يسمى بالظِّلية التي تعتمد على مقارنة إدماج وفصل العناقيد والذي يمثل مؤشر أو دليل ثقة لعضوية العنصر jth في العنقود المحدد s_i .

يمكن حساب مقياس العرض الظُلي $SR(j)$ للعنصر jth في العنقود ith حسب الصيغة الآتية :

$$SR(j) = \frac{b(j) - a(i)}{\max\left[a(j), b(j)\right]} \qquad \ldots(79.2)$$

حيث أن $a(j)$ تمثل معدل المسافة بين العنصر jth وجميع العناصر الاخرى في العنقود s_i، وتمثل $b(j)$ أصغر معدل مسافة بين العنصر jth وكل العناصر الأخرى المعنقدة في $(i \neq r\,; r = 1,\ldots,k)s_r$.

يلاحظ من الصيغة أعلاه بأن قيمة $SR(j)$ تكون بين 1 و 1− ؛ $(-1 \le S(j) \le 1)$ ففي حالة إقتراب قيمة $S(j)$ من الواحد فيستنتج من هذا أن العناصر تكون معنقدة جيداً وأنها معينة الى العنقود الملائم . وعندما تقترب قيمة $SR(j)$ من الصفر فيبين هذا أن العنصر يقع الى حد ما بين عنقودين ومن الممكن تعيينه الى أقرب عنقود . أما في حالة إقتراب قيمة $SR(j)$ من 1− فيعني هذا أن العنصر غير مصنف (Misclassified) جيداً .

وتدل القيم الكبيرة لمقياس $SR(j)$ على العنقدة الجيدة وعلى وجود عدد مثالي من العناقيد .

من جانب آخر يمكن حساب المقياس الظُلي (SR_i) الذي يصف الخصائص المختلفة لتغير العناصر ويبين العناصر المعزولة للعنقود s_i حسب الصيغة الآتية :

$$SR_i = \frac{1}{r}\sum_{i=1}^{r} SR(j) \qquad \ldots(80.2)$$

حيث أن r تمثل عدد العينات في (SR_i).

كذلك يمكن إستخدام قيمة (Silhouette) لتحديد الشكل الكروي للعناصر وإستخدامها كمقياس قانوني مؤثر لتقدير عدد العناقيد الملائمة للتجزئة U حسب الصيغة الآتية:

$$SR_u = \frac{1}{k}\sum_{i=1} SR_i \qquad \ldots(81.2)$$

حيث أن SR_i تمثل القيمة الناتجة من المعادلة (80.2) و k تمثل عدد العناقيد ، وتعكس القيم الكبيرة لـ SR_u أمثل تجزئة .

عند تطبيق طرائق صحة وقانونية تجزئة مجموعة البيانات الى عناقيد يتم إستخدام أنواع مختلفة من مقاييس التشابه وعدم التشابه لقياس الأختلاف أو المسافة بين العناقيد وداخل العنقود الواحد . تحتوي طرائق قانونية العنقود على توافيق مختلفة من مقاييس المسافة المتمثلة بستة مسافات بين العناقيد distances intercluster $(\delta_j ; 1 \le j \le 6)$ ، وثلاثة مسافات بين عناصر العنقود distances intracluster $(\Delta_i ; 1 \le i \le 3)$. فمثلاً عند تطبيق المقياس القانوني لـ (Dunn) فإن المؤشر القانوني (D_{13}) يشير الى إعتماده المسافة بين العناقيد (δ_1) ، والمسافة بين عناصر العنقود (Δ_3) وأن الصيغة الرياضية لهذه المسافات (intercluster) و (intracluster) تعرف كما في الجدولين (3.2) و (4.2) وعلى التوالي [12] و[13] :

جدول (3.2) : يبين مسافات Intercluster المستخدمة في المقاييس القانونية

$\delta_1(S,T) = \min_{x\in S, y\in T}[d(x,y)]$
$\delta_2(S,T) = \max_{x\in S, y\in T}[d(x,y)]$
$\delta_3(S,T) = \frac{1}{\lvert S\rvert \lvert T\rvert}\sum_{x\in S, y\in T} d(x,y)$
$\delta_4(S,T) = d(c_s, c_t) \qquad where: c_s = \frac{1}{\lvert S\rvert}\sum_{x\in S} x \;;\; c_t = \frac{1}{\lvert T\rvert}\sum_{y\in T} y$
$\delta_5(S,T) = \frac{1}{\lvert S\rvert + \lvert T\rvert}\left[\sum_{x\in S} d(x,c_t) + \sum_{y\in T} d(y,s_s)\right]$
$\delta_6(S,T) = \max[\delta(S,T), \delta(T,S)] \;\; where\; \delta(S,T) = \max_{x\in S}\left[\min_{y\in T}(d(x,y))\right]$ $\delta(S,T) = \max_{y\in T}\left[\min_{x\in S}(d(x,y))\right]$

حيث أن T,S تمثل العناقيد الناتجة من عملية التجزئة U ، وتمثل d(x,y) المسافة بين عنصرين x و y المنتمية الى العناقيد T,S بالتعاقب ، وتمثل $|S|$ و $|T|$ عدد العناصر المتضمنة داخل العناقيد T,S وبالتعاقب .

جدول (4.2) : يبين مسافات Intracluster المستخدمة في المقاييس القانونية

$\Delta_1(S) = \max_{x,y \in S}[d(x,y)]$
$\Delta_2(S) = \frac{1}{\|S\|.(\|S\|-1)} \sum_{\substack{x,y \in S \\ x \neq y}} d(x,y)$
$\Delta_3(S) = 2\left[\frac{\sum_{x \in S} d(x,\bar{c})}{\|S\|}\right] \quad where \quad \bar{c} = \frac{1}{\|S\|}\sum_{x \in S} x$

.

حيث أن S تمثل العنقود الناتج من عملية التجزئة (U) وتمثل $d(x,y)$ المسافة بين أي عنصرين x و y المنتمية الى S ، وتمثل $|S|$ عدد العناصر داخل العنقود S .

5.1.6.2 مقاييس صحة العنقدة لـ CD , SPR , RS , RMSSDT

تعرف مقاييس صحة العنقدة الأربعة (CD , SPR , RS , RMSSDT) والتي تستخدم لتحديد عدد العناقيد الصحيحة في مجموعة البيانات بمقاييس (Sharma)[*] عام 1996 . ويتم تطبيق هذه المقاييس الأربعة عند كل خطوة من خطوات طرائق العنقدة الهرمية .

يعرف مقياس (RMSSDT) بالجذر التربيعي لمتوسط مربع الانحراف المعياري (Root Mean Square Standard Deviation) للعنقود الجديد الذي يضم خواص العناصر المستخدمة في عملية العنقدة . ويقيس هذا المؤشر مدى تجانس العناقيد عند كل خطوة من خطوات طرائق العنقدة عند تشكيل مجاميع متجانسة ، ويحسب مقياس (RMSSDT) حسب الصيغة الآتية :

$$RMSSDT = \left[\sum_{\substack{i=1,\ldots,k \\ d=1,\ldots,p}} \sum_{j=1,\ldots,n_{id}} \left(x_j - \bar{x}_d\right)^2 \Big/ \sum_{\substack{i=1,\ldots,k \\ d=1,\ldots,p}} (n_{id}-1)\right]^{\frac{1}{2}} \qquad \ldots(82.2)$$

حيث أن K تمثل عدد العناقيد ، وتمثل p عدد المتغيرات (أبعاد البيانات) وتمثل n_{id} عدد قيم البيانات للبعد d التي تنتمي الى العنقود i ، وتمثل $\bar{x}_j$ متوسط قيم البيانات للبعد d . وتعكس القيم الصغيرة لمقياس (RMSSDT) مدى تجانس العناقيد المصنفة .

أما مقياس صحة العنقدة الثاني RS (R–Squared) فيمثل النسبة بين مجموع المربعات بين المجاميع (SSb) الى مجموع المربعات الكلي (SSt) . ويقيس درجة التجانس بين المجاميع وتقع قيمته بين الصفر والواحد . فعندما تساوي قيمة (RS) صفراً فيبين ذلك عدم وجود إختلافات بين المجاميع . وعندما تساوي قيمة (RS) واحد فيؤشر ذلك وجود إختلافات معنوية بين المجاميع .

يعرف مقياس (RS) حسابياً حسب الصيغة التالية :

$$RS = \frac{\left\{\sum_{d=1,\ldots,p}\left[\sum_{j=1}^{n_d}\left(x_j-\bar{x}_d\right)^2\right]\right\}-\left\{\sum_{\substack{i=1,\ldots,k\\ d=1,\ldots,p}}\left[\sum_{j=1}^{n_{id}}\left(x_j-\bar{x}_d\right)^2\right]\right\}}{\sum_{d=1,\ldots,p}\left[\sum_{j=1}^{n_d}\left(x_j-\bar{x}_d\right)^2\right]} \qquad \ldots(83.2)$$

حيث أن k تمثل عدد العناقيد ، وتمثل d عدد المتغيرات (أبعاد البيانات) ، وتمثل n_d عدد قيم البيانات للبعد d بينما تمثل n_{id} عدد قيم البيانات للبعد d التي تنتمي الى العنقود i ، وتمثل $\bar{x}_d$ متوسط قيم البيانات للبعد d .

أما مقياس صحة العنقدة الثالث (SPR) للعنقود الجديد فيمثل الفرق بين مجموع المربعات داخل المجموعة (SS_w) للعنقود الأول وبين قيم (SS_w) للعنقود المرتبط للحصول على العنقود الجديد مقسوماً على مجموع مربعات مجموع البيانات الكلية (SS_t) . ويقيس هذا المقياس مقدار فقدان التجانس بعد عملية دمج عنقودين لخطوة مفردة عند تطبيق الطريقة . فإذا كانت قيمة المؤشر تساوي صفر فإن العنقود الجديد يمثل ناتج حاصل دمج عنقودين متجانسين بصورة تامة ، أما إذا كانت قيمته عالية فإن ناتج العنقود الجديد متأتي من إدماج عنقودين غير متجانسين .

أما مقياس صحة العنقدة الرابع (CD) فيمثل المسافة بين عنقودين (Distance Betwee Two Cluster) مدمجين عند خطوة معينة . ويتم قياس هذه المسافة عند كل مرة بالإعتماد على العناصر الممثلة المختارة للعنقدة الهرمية التي أُنجزت . وفي حالة العنقدة الهرمية المركزية (centroid

(CD) يمثل hierardical clustering) فإن تشكيل العناصر الممثلة تمثل مركز كل عنقود لذلك المسافة بين مراكز العناقيد . وفي حالة إستخدام عنقدة الربط المفرد (Single Linkage Clustering) فإن (CD) يقيس أصغر مسافة بين كل أزواج العناصر . أما في حالة الربط التام (Complete Linkage) فإن (CD) يمثل أكبر مسافة بين كل أزواج العناصر وهكذا .

عند تطبيق المقاييس أعلاه في حالة العنقدة غير الهرمية (مثل K–Means) فمن الممكن إستخدام بعضها لحساب صحة نتائج العنقدة ومنها مقياس (RMSSTD) ومقياس (RS) . وتتمثل الفكرة بتكرار عملية العنقدة عدة مرات بأعداد مختلفة من العناقيد في كل مرة ، ومن ثم يتم رسم التخطيط البياني المتعاقب لنتائج المقاييس القانونية لخوارزميات العنقدة ، بعد ذلك يتم البحث عن أكبر قفزة (greater jump) لقيم هذه المقاييس من أعلى عدد من العناقيد الى أصغرها من خلال الرسم التخطيطي .

2.6.2 مقاييس مقياس صحة العنقدة للتحليل العنقودي الضبابي

Clustering Validation Fuzzy Cluster Analysis

في هذا المبحث يتم تقديم مقاييس صحة العنقدة الملائمة للتحليل العنقودي الضبابي . فمن المعلوم أن هدف الطرائق العنقودية الضبابية هو الحصول على موجهات من البيانات المصنفة التي تظهر درجة عالية من العضوية في العنقود الواحد . لذلك تعرف العنقدة الضبابية بمصفوفة $U = (u_{ij})$ حيث أن u_{ij} تشير الى درجة عضوية العنصر x_j في العنقود ith . لذلك نهج الباحثين الى إستخدام نوعين من المقاييس القانونية ، يستخدم النوع الأول قيم العضوية (membership values) فقط لبيانات التجزئة الضبابية ، ويستخدم النوع الثاني كلاً من مصفوفة العضوية U ومجموعة البيانات الأصلية نفسها .

1.2.6.2 المقاييس صحة العنقدة لـ Bezdek

من مقاييس صحة العنقدة المتضمنة إستخدام قيم العضوية فقط ، المقياس القانوني المسمى معامل التجزئة (partition coefficient) المقترح من قبل (Bezdek)[17] عام 1980 والذي يعرف بالصيغة الآتية:

$$PC = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n}u_{ij}^2 \qquad \ldots(84.2)$$

حيث أن u_{ij} تمثل درجة عضوية العنصر jth في العنقود ith لـ(n) من العناصر و(k) من العناقيد

يقيس مقياس صحة العنقدة لـ (PC) كمية التداخل (OverLap) بين العناقيد ومقدار تقارب العناصر الى مراكز عناقيدها . وتقع مدى قيم المؤشر (PC) ضمن الفترة $\left[\frac{1}{k}, 1\right]$ ، فعند إقتراب قيمة (PC) من الواحد فالنتيجة الحصول على عنقدة واضحة (Crisp Clustering) ، وفي حالة تساوي كل قيم العضوية الى التجزئة الضبابية عند تصنيف البيانات الى العناقيد بمعنى $\left(u_{ij} = \frac{1}{k}\right)$ فإن قيمة (PC) تحصل على أوطئ قيمة لها وتؤشر الحصول على عنقدة غير جيدة ، أما أفضل إختيار لعدد العناقيد وأفضل تجزئة للبيانات فيتم الحصول عليه عند :

$$optimal(PC) = \max_{k} \left\{ \max_{\Omega k}(PC) \right\}; k = 2,...,n-1 \qquad \ldots(85.2)$$

المقياس الآخر لهذا النوع من مقاييس صحة العنقدة التي تعتمد على قيم العضوية هو معامل انتروبي للتجزئة Partition Entropy Coefficient (PE) أو ما يسمى بمقياس انتروبي للتصنيف (Classification Entropy Measure) المقترح من قبل (Bezdek)[17] والمعرف بالشكل الآتي:

$$PE = -\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n} u_{ij}.Ln(u_{ij}) \qquad \ldots(86.2)$$

تحسب قيمة مقياس صحة العنقدة لـ (PE) ضمن المدى $[0, Ln(k)]$. وعندما تساوي قيمة (PE) صفراً فإن نتيجة التجزئة تمثل عنقدة حادة (Hard Clustering) ، وعندما تساوي ((Ln (k)) فيدل ذلك على ضعف نتيجة العنقدة . ويتم الحصول على أفضل قيمة للعنقدة وأمثل عدد من العناقيد عندما تساوي قيمة (PE) الصيغة الآتية:

$$optimal(PE) = \min_{k} \left\{ \min_{\Omega k} PE \right\} \quad , k = 2,....,n-1 \qquad \ldots(87.2)$$

أما المقياس القانوني الضبابي الآخر المعدل من قبل (Bezdek)[65] هو مقياس انتروبي للتجزئة المعدلة Normalised Partition Entropy (NCF) الممثل بالصيغة الآتية :

$$NCF = \frac{PE}{Ln\ k} \qquad \ldots(88.2)$$

حيث أن (PE) يمثل مقياس انتروبي للتجزئة المبين في الصيغة (86.2) وتمثل ((Ln (k)) لوغاريتم عدد العناقيد .

2.2.6.2 مقياس صحة العنقدة لـ Fukuyama–Sugeno (FS)

المقياس القانوني الآخر الذي يعتمد على قيم العضوية (u_{ij}) ومجموعة البيانات (X) هو المقياس القانوني (FS) المقترح من قبل (Fukuyama–Sugeno)[*] عام 1989 والمعرف حسب الصيغة الآتية :

$$FS = \sum_{i=1}^{k}\sum_{j=1}^{n} u_{ij}^{m} \left[\left\| x_j - c_i \right\|^2 - \left\| c_i - c \right\|^2 \right] \qquad \ldots(89.2)$$

حيث أن c_i تمثل مركز العنقود ith. وتمثل c مركز مجموعة البيانات X وتمثل m المعامل الضبابي

يقيس الحد الأول من الصيغة (89.2) إدماج العناقيد ، أما الحد الثاني فيقيس المسافة الممثلة في العنقود. وللحصول على عناقيد مدمجة ومفصولة جيداً يتم إختيار القيم الصغيرة لمقياس (FS) .

3.2.6.2 مقاييس صحة العنقدة لـ Gath–Geva

أقترح هذا المقياس من قبل (Gath–Geva)[37] عام 1989 لتحديد صحة وقانونية العناقيد بالإعتماد على مؤشرات الحجم المفرط (hyper volume) والكثافة (density) . حيث عرف (Gath–Geva) المقياس القانوني لحجم العنقود ith في التجزئة الضبابية على النحو الآتي:

$$FHV = \sum_{i=1}^{k} \sqrt{\det\ (F_i)} \qquad \ldots(\ 90.2)$$

حيث أن (F_i) det تمثل محددة مصفوفة التباين المشترك الضبابية (F_i) للعنقود ith والمعرفة حسب الصيغة الآتية :

$$F_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} \left[(x_j - c_i)(x_j - c_i)' \right]}{\sum_{j=1}^{n} u_{ij}^{m}} \quad ; i = 1,\ldots\ldots,k \qquad \ldots(91.2)$$

حيث أن u_{ij} تمثل قيم العضوية ويمثل x_j عناصر البيانات وتمثل c_i مركز العنقود ith أما m فتمثل المعامل الضبابي .

بعد ذلك عرف مقياس الحجم المفرط الضبابي الكلي (Total Fuzzy Hyper Volume)بالصيغة الآتية :

$$TFHV = \sum_{i=1}^{k} FHV \qquad \ldots(92.2)$$

حيث أن k تمثل عدد العناقيد وتمثل (FHV) صيغة المعادلة (90.2) وتؤشر القيم الصغيرة لمقياس (TFHV) على وجود عناقيد مدمجة جيداً .

والنوع الآخر من هذه المؤشرات هو مقياس معدل كثافة التجزئة (average partition density) المقترحة من قبل (Gath–Geva) أيضاً والتي يمكن تعريفها حسب الصيغة الآتية :

$$PA = \frac{1}{k}\sum_{i=1}^{k}\frac{S_i}{FHV} \qquad \ldots(93.2)$$

حيث أن $S_i = \sum_{x_i \in X} u_{ij}$

حيث أن x_i تمثل مجموعة نقاط البيانات التي تقع ضمن المنطقة المعاد تصنيفها حول مركز العنقود (c_i) والتي تعرف بمجموع الاعضاء المركزية للعنقود ith .

والمؤشر الآخر المختلف المقترح هو مقياس كثافة التجزئة (Partition Density) والمعرف على النحو الآتي [37] :

$$PD = \frac{S}{TFHV} \qquad \ldots(94.2)$$

حيث أن (TFHV) هي الصيغة المبنية في المعادلة (92.2) وأن :

$$S = \sum_{i=1}^{k} S_i \qquad \ldots(95.2)$$

4.2.6.2 مقياس صحة العنقدة لـ Beni–(XB)

إقترح كل من (Xie و Beni)[82] عام 1991 مؤشراً قانونياً يتضمن إستخدام قيم العضوية u_{ij} مع مجموعة البيانات x_j لقياس صحة قانونية العنقدة وعدد العناقيد . يطلق أيضاً على هذا المقياس بالدالة القانونية للفصل والإدماج (Compactness and Separation Validity Function) ويمثل النسبة بين إدماج وفصل مجموعة البيانات بطريقة التجزئة الضبابية ، حيث يعمل على جعل كل العناصر الداخلة قريبة قدر الأمكان من مراكز عناقيدها وجعل مراكز العناقيد المفصولة بعيدة عن بعضها .

لتكن المسافة بين البيانات x_j ومركز العنقود c_i الموزون بقيم العضوية الضبابية u_{ij} للعنصر jth الى العنقود ith يمثل الانحراف المعياري الضبابي الآتي :

$$d_{ij} = u_{ij}^{m} \left\| x_j - c_i \right\| \qquad \ldots(96.2)$$

ويعرف الانفصال للتجزئة الضبابية بأنه أصغر مسافة بين مراكز العناقيد أي :

$$d_{\min} = \min \left\| c_i - c_r \right\| \qquad \ldots(97.2)$$

حيث أن c_i و c_r تمثل مركزي العناقيد ith و rth على التوالي .

من أعلاه يعرف مقياس (XB) حسب الصيغة الآتية :

$$XB = \frac{d_{ij}}{n\, d_{\min}} \qquad \ldots(98.2)$$

وتعكس القيم الصغيرة لمقياس (XB) أمثل تجزئة قانونية وبالتالي الحصول على عناقيد مدمجة ومفصولة جيداً . ومن الواضح أن مقياس (XB) يتناقص بصورة رتيبة عند إزدياد أعداد العناقيد وإقترابها الى n . وتتحدد قيمة (XB) بين $[2, C_{\max}]$ حيث أن $C_{\max}$ تمثل نقطة البداية للشكل الرتيب للتناقص . كما أن قيمة (XB) تعتمد على قيم الضبابية m فإذا كانت $m \to \infty$ فإن $XB \to \infty$.

5.2.6.2 مقياس الانفصال لصحة العنقدة (S)

إقترح كل من (Bensaid، Hall، Bezdek، Clarke، Silbiger، Arrington، Murtagh) [16] عام 1996 مؤشراً قانونياً لقياس صحة وقانونية نتائج العنقدة المستخرجة من الطرائق العنقودية وكذلك إيجاد العدد الأمثل من العناقيد من خلال نسبة الإختلافات الكلية ضمن العناقيد (total variation within clusters) وفصل العناقيد (separation of clusters) وحسب الصيغة الآتية :

$$S = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}(u_{ij})^2 \left\|x_j - c_i\right\|}{n.\min_{i,j}\left\|x_j - c_i\right\|} \qquad \ldots\ldots(99.2)$$

حيث أن u_{ij} تمثل عناصر مصفوفة التجزئة ، وتمثل $\left\|x_j - c_i\right\|$ المسافة بين عناصر البيانات x_j ومراكز العناقيد c_i ، وتمثل n عدد عناصر البيانات اما $\min_{ij}\left\|x_j - c_i\right\|$ فتمثل أصغر مسافة بين عناصر البيانات ومراكز العناقيد .

وللحصول على أمثل عدد من العناقيد يتم العمل على تصغير قيمة (s) .

6.2.6.2 مقياس صحة العنقدة لـ Halkidi

من المقاييس القانونية الحديثة المقترحة هو مقياس (SD) المقترح من قبل (Halkidi)[42] عام 2001 والذي يعتمد على مفهوم معدل إنتشار العناقيد (Average Scattering for Clusters) والانفصال الكلي بين العناقيد (Total Separation Between Clsuters) .

يعرف معدل إنتشار العناقيد حسب الصيغة الآتية :

$$Scat\,(k) = \frac{1}{k}\sum_{i=1}^{k}\left\|\sigma(c_i)\right\| / \left\|\sigma(X)\right\| \qquad (100.2)$$

حيث أن k تمثل عدد العناقيد ، و $\sigma(c_i)$ تمثل إنتشار مراكز العناقيد ، وتمثل $\sigma(X)$ إنتشار مجموعة البيانات .

ويعرف الأنتشار (الانفصال) الكلي بين العناقيد حسب الصيغة الآتية :

$$Dis(k) = \frac{D_{\max}}{D_{\min}}\sum_{S=1}^{k}\left[\sum_{Z=1}^{k}\left\|c_S - c_Z\right\|\right]^{-1} \qquad \ldots(101.2)$$

حيث أن $D_{\max} = \max\left(\left\|c_i - c_r\right\|\right)$ لكل $(i, r \in 1,2,\ldots..,k)$ تمثل أكبر مسافة بين مراكز العناقيد ، وتمثل $D_{\min} = \min\left(\left\|c_i - c_r\right\|\right)$ لكل $(i, r \in 1,2,\ldots..,k)$ أصغر مسافة بين مراكز العناقيد .

ومن المعادلتين (100.2) و (101.2) يعرف المقياس القانوني (SD) على النحو الآتي :

$$SD(k) = a\,.Scat\,(k) + Dis\,(k) \qquad \ldots(102.2)$$

حيث أن a تمثل المعامل الموزون (weighting factor) والمساوي الى $Dis\,(C_{\max})$ حيث $C_{\max}$ تمثل أكبر عدد من العناقيد الداخلة ، ويستخدم المعامل a الحدين (Scat)و (Dis) بطريقة متوازنة .

تعكس قيم (SD) الناتجة العدد المثالي من العناقيد بالإعتماد على زيادة قيمة Scat(k) وتقليل قيمة Dis(k) .

29-2 مقارنة بين طرئق تحليل متعددة المتغيرات:

يمكن عمل مقارنة توضح الفروق الأساسية بين طرائق متعددة المتغيرات وهي التحليل التمييزي والتحليل العاملي والتحليل العنقودي كما يلي :

- يختلف أسلوب التحليل العاملي عن أساليب التحليل الإعتمادية والتي منها التحليل التمييزي في أن أساليب الإعتمادية تحدد بوضوح المتغيرات التابعة والمتغيرات المستقلة ، بينما يتم أخذ كل المتغيرات في الإعتبار وفي نفس الوقت بالنسبة للتحليل العاملي حيث ترتبط كل المتغيرات . ويتم تحديد العوامل بحيث تعظم قدرتها على شرح البيانات الأصلية ككل وليس بغرض التنبؤ بالمتغير التابع كما هو الحال بالنسبة لطرق التحليل الإعتمادية .
- يمكن النظر إلى كل المتغيرات في التحليل العاملي على أنها متغيرات تابعة بمعنى أنها تعتمد على بعض الأبعاد المستترة والتي تتكون من كل المتغيرات الأخرى ، وبالتالي يتم التنبؤ بكل متغير بواسطة المتغيرات الأخرى .كذلك يمكن النظر إلى كل عامل على أنه متغير تابع حيث يعتمد على كل المتغيرات الأخرى .

  وبالتالي يمكن القول أن الغرض الأساسي من التحليل العاملي هو البحث عن البناء بينما الهدف من التحليلات الإعتمادية هو التنبؤ .
- كذلك يكمن الفرق بين التحليل العاملي والتحليل العنقودي في أن الأول مبني على الإرتباط الداخلي بين المستجوبين . بينما تتكون الجاميع أو العناقيد في الثاني على أساس التشابه المبني على المسافة بين القيم التي يحددها المستجوب على المتغيرات التى يشملها التحليل .

الفصل الثالث

مرض السكري

3-0 تمهيد:

المرض المزمن (او الازمان)هو المصطلح الي يدل علي الثبات والاستمرار في ظهور مرض من الامراض وفي هذه الحاله يصبح مرض مزمن والمرض المزمن او الامراض المزمنة هي التي تتسم بخاصيه الازمان اي الثبات والاستمراريه وهذا يعني انها ليست تلك التي تظهر وتختفي مع تناول العلاج وهي التي تظهر وتبقي والتي لم تكتشف بعد وسيله للشفاء منها بصوره نهائيه ،وبما ان خيار علاج مثل حالات الامراض المزمنه غير ممكن فان الخيار الثاني المتوفر هوفي مراقبتها والسيطره عليها وهذا هو أقل ما يمكن فعله حتي لا تزداد الحاله سوء .ومرض السكر من النوع الثاني هو واحد من الامراض المزمنه وسرطان الجلد ،ومرض هشاشه العظام هو كذلك حاله مزمنه .

والمرض المزمن يمكن ان يصيب اي من وظائف الجسم او انظمته الفسيولوجيه .

وتفيد تقارير منظمه الصحه العالميه انه من الممكن الوقايه من حوالي 80% من امراض السكر من النوع الثاني والقلب والشرايين والسرطان بالمقارنه بمعدلات الاصابه العالميه

3-1 مرض السكر :-

الاسم العلمي هو (سكروز) ونستعمله للتحليل ،والسكروز يتحول من الجسم الي الجلكوز ام الجلكوز فهوموجود في دم كل واحد منا واهميته لنا مثل اهميه البنزين للسياره ،فيحرقه الجسم ليحصل علي الطاقه التي بها نؤدى انشطتنا .وليس معني ذلك ان الزياده في تناول السكر تؤدى الى الزياده الجلكوز في الدم وبالتالي الي زياده طاقتنا ونشاطنا فزياده تناول السكر او النشويات او الدهون...الخ تؤدى الي زياده الجلكوز في الدم ولكن نسبته محدوده والباقي قد يحوله الجسم الي دهون تترسب في اجزاء الجسم المختلفه فتزداز الدهون وتؤدي الي ما نسميه السمنه.

ماهي هذه النسبه وماذا لو ذادت؟

نسبه الجلكوز في الدم يجب ان لا تقل عن 80 ملجم جلكوز في كل 100 مل دم وان لا تزيد عن 120 ملجم هذا اذا تم التحليل للشخص الصائم (لم يتناول الافطار صباحا).وجسمنا دائما يعمل علي ان تكون النسبه 80\120 ثابته. وذلك بان تقدم البنكرياس بافراز هرمون الانسولين المسئول عن حرق السكر في الجسم.هرمون الانسولين يفرز في دم الانسان بواسطه خلايا في البنكرياس تسمي (جزر لانجر هانز) .وهي تفرز الانسولين المسؤل عن تنظيم نسبه السكر في الدم بحيث لا تتعدي 120 ملجم جلكوز لكل 100 مل دم واي زياده للجلكوز عن تلك النسبه

تحفز اقران الانسولين سيجعل خلايا الجسم تحرق الجلكوز وتحوله الي طاقه والمتبقي يتحول بواسطه الكبد الي دهون ويخزنها فيزداد وزن الجسم وهو ما يعرف بالسمنه ايضا.

هل هنالك مخاطر اخري غير السمنه من تناول الدهون والسكر؟مثلا الحلويات.

ان الافراط في ما تم ذكره له مخاطر عديده مثل ترسب دهون ضاره علي الكبد وعلي جدران الاوعيه الدمويه وهو ما يؤدي في النهايه الي تصلب الشرايين.اما اذا فشل البنكرياس في افراز كميات كافي من الانسولين سيؤدي ذلك الي ارتفاعنسبه الجلكوز في الدم .الذي اذا زاد عن 120ملجم لكل 100 مل دم سيكون ذلك سكر مرتفع النسبه اما اذا زاد عن 180 ملجم لكل 100 مللي دم فسيظهر السكر في بول المريض وهو ما يعرف بمرض (البول السكري) لان الكليتين تسمحان بمرور الجلكوزمع البول اذا تعدي تركيزه في الدم 180 ملجم لكل 100مل دم .وعندها نقول ان الشخص مريض بالسكر.وبالتالي يحتاج الي تخفيض تناول السكريات والنشويات والدهون الي ادني حد.واذا فشل هذا الرجيم فيكون لابديل عن العلاج بالادويه

3- 2 انواع مرض السكري

مرض السكري نوعان :-

النوع الاول (المعتمد علي الانسولين):

هو عدم المقدره التامه للخلايا الموجوده بالبنكرياس علي افراز الانسولين وهذا غالبا ما يصيب الشخص منذ طفولته وتلعب العوامل ،وتلعب العوامل الوراثيه دورا كبرا اذ ان معظم المصابين بهذا النوع لديهم اباء واجداد يعانون من هذا المرض .والعلاج الحالي لهذا المرض هو اعطاء الانسولين من الخارج في صوره حقن تحت الجلد ولذلك يسمي (المعتمد علي الانسولين) وتجري الان محاولات كثيره لتوفير بديل لحقن الانسولين مثل لبوس شرجي لاعطاء وزرع خلايا تحت الجلد تفرز الانسولين.

النوع الثاني (الغير معتمده علي الانسولين) :

هو امكانيه تنشيط الخلايا للبنكرياس من جديد لافراز كميات كافيه من الانسولين تكفي لضبط نسبه الجلكوز في الدم. وبالتالي لا نحتاج الي الانسولين الخارجي لذلك يسمي السكر الغير معتمد علي الانسولين .

وهذا النوع عادا ما يصيب البالغين وكبار السن وله اسباب عديده مثل تناول الحلوي والمشروبات المحتويه علي سكر بكثره .او عدم مقدره البنكرياس علي افراز كميات كافيه من الانسولين مثل الهيدروكربتزون ومشتقاته.

وهذا النوع من مرض السكر يمكن علاجه بالاقراص المنشطه لخلايا البنكرياس لافراز هرمون الانسولين . كما ان مرض السكري يريد ان يتبع نظام غذائي معين يخلو من البروتينات والدهون

ويقلل من النشويات والسكريات ويكون غذائه غنيا نالخضروات والفاكهه ويجب عليه تناول فايتمين(ب) المركب لتجنب التهاب الاعصاب خاصا اعصاب الارجل والعينين واليدين .
واهمال مرض السكر والاكل العشوائى دون اكترا ث يؤدي الي متاعب كثيره مثل التهاب الاعصاب بما فيه اعصاب العيون وزياده شديده في جلكوز الدم قد يؤدي الي الدخول في غيبوبه سكريه.

3-3 أعراض مرض السكر:

1- زياده التعرق وشحوب الوجه وبروده الاطراف .
2- تكرار التبول خاصه اثناء الليل مع زياده البول عن المعتاد .
3- العطش الزائد والشعور بالجوع وزياده الشهيه
4- الخمول والضعف العام والهزيان
5- غثيان ودوار اضطرابات الرؤيه وتكرا حدوث التخمجات

3-4 مضاعفات مرض السكر :

ان ارتفاع مستوي السكر في الدم لفترات طويله يؤدي الي مضاعفات صحيه مثل :-
1- اعصاب العين باعتلال الشبكيه والمياه البيضاء
2- اصابه الكليتين بالفشل الكلوي
3- اصابه القلب والاوعيه الدمويه بالزبحه الصدريه وارتفاع ضغط الدم والكلسترول
4- اصابه القدم بالقرح المزمنه (القدم السكري)
5- الضعف الجنسي
6- إلتهاب اعصاب الارجل واليدين والعينين

لذلك يجب الحفاظ علي مستوي السكر بالدم تجنبا لحدوث مثل هذه المضاعفات كفقدان البصر.

ماذا يحدث اذا زادت او إنخفضت نسبه السكر بالدم :-

غيبوبه زياده نسبه السكر بالدم

السبب:-

1- زياده السكر في الدم
2- نقص هرمون الانسولين

الاعراض :-

1- الجلد جاف ودافئ
2- الوجه محتقن
3- التنفس ريع وعميق

4- ضغط الدم منخفض والنبض ضعيف

5- الاحساس بالعطش

6- يتبول المريض بكميات كبيره

7- واهم ما يميز غيوبه السكر قبل حدوثها واثناء الغيبوبه رائحه الاستون (التفاح الفاسد من فم المصاب).

وهذه الرائحه تظهر نتيجه لتحول الجلكوز الزائد في الدم في النهايه الي استون الذي يعبر حاجز المخ ويسبب الغيبوبه

العلاج :-

النقل الي المستشفي فورا لعمل اللازم واخذ الانسولين .

غيبوبه نقص نسبه السكر في الدم (صدمه الانسولين):-

1- انخفاض مسبه السكر في الدم نتيجه تعاطي كميات كبيره من الانسولين

2- المريض لم ياكل بعد تناول الانسولين او تاخر ميعاد الاكل (وذلك من المهم جدا) تعاطي المريض طعامه فورا بعد تعاطي الانسولين او قرص خفض السكر في الدم.

3- بذل مجهود عضلي شديد ادي الي انخفاض نسبه السكر في الدم.

الاعراض :-

1- الجلد باهت وشاحب وعرق بارد

2- النبض سريع وضعيف

3- شعور بالجوع والضعف العام قبل حدوث الاغماء

4- هزيان ورعشه وعدم اتزان قبل حدوث الغماء

5- قد يموت المريض فورا اذا لم يعالج بسرعه

العلاج :-

1- يعطي المريض اي شئ به سكر

2- اما اذا كان فاقد الوعي يعطي محلول الجلكوز

الحلات التي يتم علاجها بالانسولين:-

1- مرض السكر من النوع الاول

2- ظهور الاستون او حدوث غيبوبه السكر والذي يحدث في حاله زياده نسبه السكر

3- خلال الحمل

4- مرض السكر من النوع الثاني.في حاله فقد الاستجابه للاقراص المخفضه للسكر عن طريق الفم.

الحالات التي يتم علاجها بالاقراص المخفضه للسكر في الدم :-

تستعمل هذه الاقراص من مرض السكر من النوع الثاني ولكنها لا تستعمل في مرض السكر من النوع لاول .لان هذه الاقراص تعتمد في عملها علي وجود الاستولين الطبيعي الذي يفرز من البنكرياس.وهي معدومه اصلا من النوع الاول

3-5 كيفيه علاج مرض السكر :-

تضافرت جهود الجميع افرادا وجماعات ومؤسسات وحكومات للقضاء علي الداء العضال والحد من انتشاره وقد كانت افضل الطرق واكثرها فاعليه في محاربه هذا المرض ما يقوم علي أسس عامه هي :-

1- المعالجه عن طريق الغذاء فقط .

2- المعالجه بالغذاء والانسولين.

3- المعالجه بالغذاء والاقراص المخفضه لسكر الدم.

4- التثقيف العام والشامل للمريض ومن حوله المجتمع .

5- انشاء العيادات المتخصصه الدائمه لرعايه مرض السكري، والغذاء المتوازن افضل الاساليين لتجنب الامراض والعلل كلها والحميه الغذائيه تشكل الساس لمعالجه السكري . واذا ارتبطت هذه الحميه ببرامج لزياده النشاط الفيزيائي والرياضه البدنيه فهو المطلوب . وحينها يساعد ذلك المعالجه بالادويه من هنا ينصح مريض السكري يحمله

3-6 ارشادات غذائيه منها :-

1- الابتعاد عن الاغذيه المحلاه بالسكر ،او تناولها بكميات معتدله .

2- الحد من تناول الدهنيات من اصل حيواني ونباتي كاللحوم والبيض والسمن والزيت لتجنب البدانه واضرار الكوليسترول .

3- الاكثار من تناول الاغذيه الغنيه بالالياف مثل :الخبز الاسمد- الحبوب بانواعها- شوربه الشعير البقوليات مثل فول الصويا والبوزولياء الخضراء الفاصوليا البيضاء علي ان تؤكل مع قشورها غضافه الي الخضروات والفواكه الطازجه ذات القشور مثل التفاح والخوخ تناول حبه واحده او حبتين . والبندوره والبطاطا .

4- التقليل من الملح في الطعام .

5- الابتعاد عن الكحول .

6- تناول خمسه وجبات يوميا مع مراعاه التقليل في كميه الوجبه

3-7 علاقه مرض السكر بامراض القلب

مرض السكري لايتوقف تاثيره السلبي عند عرض معين او عند جهاز محدد بل ان تاثير المرض في حاله عدم التحكم في مستوي السكر في الدم يمتد الي جميع اعضاء الجسم صحه بلا استثناء وبلا انتقاء فيشمل القلب وقدرته والكبد وفعاليته ناهيك عن البصر الذي قد يضعفه

والعظم الذي يؤهنه .اضافه الي الجلد الذي قد يفقد نضارته والكلي التي تقلل بسببه فعاليتها كما اننا لو تحدثنا عن علاقه مرض السكر بامراض القلب لوجدنا ان مرض السكري لا يقتصر تاثيره علي القلب فقط او علي الشرايين فقط بل تاثيره يصل الي الجهاز الدوري باكمله

3-8 الفئات المعرضه للاصابه بالسكري

كل الناس معرضون للاصابه بالسكري الا ان بعض الفئات تكون اكثر عرضه للاصابه بسبب العوامل الاتيه :-

- اصابه احدي الاقارب بداء السكري
- السمنه وبشكل خاص بمنطقه البطن
- بلوغ سن 45 عاما فاكثر
- نمط الحياه الفاتر
- اصاه سابقه بالسكري اثناء الحمل
- اصابه سابقه باختلال نسبه السكر بالدم

3-9 مضاعفات داء السكري ايضا

اذا ظلت نسبه السكر مرتفعه لمده طويله فانها تؤدي مضاعفات حاده واخري مزمنه قد تسبب الوفاه او يكون لها تاثير شديد علي صحه المرض وجوده حيالهم .

وتشمل مضاعفات السكري ما يلي :-

- مضاعفات القلب والجهاز الدوري الدموي .اختلاف وظائف عضله القلب والجهاز الدوري الدموي . ومضاعفات النظر فقدان البصر بسبب تاثير السكري علي العين خاصه شبكيه العين . مضاعفات الجهاز العصبي ، يؤثر السكري علي الجهازالعصبي المركزي (المخ والنخاع الشوكي)والجهاز العصبي الطرفي.

مضاعفات السكر علي الكلي . قد تؤدي الي الفشل التام في وظائف الكلي

3-10 اعتلال الكلي السكري ؟

اعتلال الكلي السكري هو مرض مزمن يتفاقم تدريجيا ويحدث عند حوالي ثلث مرضي السكري بنوعيه الاول والثاني . ويبلغ عدد المصابين به في العالم حوالي 60 مليون شخص . يؤدي الارتفاع المزمن بنسبه السكر بالدم الي تلف الاوعيه بالجسم وعندما تصاب الاوعيه الدقيقه بالكلي ،فان وظيفه تصفيه الدم من الشوائب تختل وينتج عن ذلك تراكم الفضلات والشوائب في الدم تدريجيا . ويصاحب ذلك فقدان البروتين في البول اذا استمر تدهور وظائف الكلي التام ،بحيث تتوقف الكلي عن العمل تماما وهو ما يعرف بالمرحله النهائيه لمرض الكلي .وقد تؤدي هذه الحاله للوفاه

عوامل خطوره الاصابه باعتلال السكري ؟

- التدخين
- ارتفاع ضغط الدم
- ارتفاع نسبه السكر بالدم
- اصابه احدي الاقارب باعتلال الكلي السكري

اعراض الاصابه باعتلال الكلي السكري :-

قد لا يشعر المصاب باعتلال الكلي السكري بايه اعراض علي الاطلاق فهي غالبا لا تظهر الا في المراحل المتقدمه من فشل الكلي . ومع ذلك فإن بعض العلامات المرضيه قد تدل علي اعتلال الكلي في مراحل الاوليه مثل :-

- ظهور البروتين في البول
- ارتفاع ضغطالدم
- تدهور وظائف الكلي
- حدوث تورم وآلا بالساقين
- كثره التبول بالليل
- انخفاض احتياج المصاب من الانسولين او الاقراص الغمويه الخافضه لسكر الدم
- الغثيان والقي

طرق علاج فشل الكلي السكري :-

علاج الفشل التام في وظائف الكلي يمكن في احدي الطريقتين التاليتين :

- الغسيل الدموي
- زراعه الكلي

3-11 تاثير مرض السكري علي العين :-

يتسبب السكر سنويا في حدوث 12% من نسبه الاصابه بالعمي وذلك نتيجه مضاعفات هذا المرض علي العين بشكل مباشر هذا مع العلم بان مضاعفات المرض تختلف حسب طول مده الاصابه بالمرض .

تأثير السكري علي الشبكيه :-

يؤثر مرض السكري علي جميع اجزاء العين ومن اهم تاثيراته تلك المضاعفات علي شبكه العين والتي تؤدي الي فقدان البصر .

كيفيه شعور الشخص بتاثر السكري علي العين .

يبدأ الشخص المصاب بالسكري في الشعور بتغير حده النظر خلال اليوم الواحد من تحسن الرؤيه الي تدني ثم الي التحسن حسب نسبه السكر في الدم وذلك بسبب تاثير عدسه العين بالتغير في نسبه السكر بالدم .

وكما يشعر الشخص خلال الشهور المتتاليه بتدهور يدريجي وذلك نتيجه حدوث غرتشاح ونزيف في العين وظهور اوعيه دمويه جديده غير طبيعيه وقد يؤدي الي نزيف بالجسم الزجاجي وإنفصال شبكي واحيانا يتسبب بالاصابه بالمياه الزرقاء بالعين (الجلووكوما) وقد ينتهي الامر بفقد البصر .

يستطيع الطبيب متابعه العين بعدنالعمليه للشخص المصاب بالسكري بالكشف الدوري والتصوير بصيغه (الفلوروسين) لتحديد مدي تأثير السكري علي شرايين وانسجه العين .

التأثيرات الاخري للسكري :-

يسبب السكري تغيرا بعدسه العين يؤدي الي اعتامها وهو ما يعرف بالنزول الابيض وعلاج مثل هذه الحالات باجراء عمليات جراحيه وزرع عدسات بالعين وهذه العمليات تعيد الابصار .

يسبب السكري زياده ضغط العين والذي يعالج باجراء جراحه اوباشعه الليزر او القطرات المخفضه لضغط العين والعلاج لا يعيد البصر الذي فقد ولكن يحافظ علي الباقي من البصر .

2-13 منطقة الدراسة: ولاية شمال كردفان

تقع ولاية شمال كردفان بين خطى عرض 12-17 وخطى طول من 17- 32 درجة شرقاً . وتمتاز بمناخ شبه صحراوي في مساحة وقدرها 110.852 كيلو م م وتمثل 54.5 من مساحة كردفان الكبرى ويقدر سكانها 3188310 نسمة

وتتكون من ثمانية محليات هي : شيكان – الرهد – أم روابة – أم دم حاج أحمد – بارا – غرب بارا – جبرة – سودري .الحرفة الرئيسية الزراعة والرعي والتجارة .

أما بالنسبة للجانب الصحي فتحتوي الولاية على 28 مستشفى و 151 مركز صحي 357 وحدة صحية .

الفصل الرابع

الجانب التطبيقي

4-0 تمهيد :

يتضمن هذا الفصل الجانب التطبيقي او العملي على بيانات مرضى السكري حيث نتطرق إلى بيانات البحث والمتغيرات والتحليل الإحصائي بإستخدام التحليل التميزي والتحليل العاملي والتحليل العنقودي بإعتبارهم من أساليت تحليل متعددة المتغيرات المتقدمة .

4-1 بيانات البحث :

اعتمد الباحث على بيانات الإستبيان حيث تم اخذ عينة عشوائية من الأشخاص المصابين بمرض السكري وغير المصابين بمستشفي الأبيض التعليمي ومستشفى الشرطة ومركز ود الياس لمرضى السكري وقد بلغ حجم العينة 216 مريض منهم 162 مصاب بالمرض و54 غير مصاب بمرض السكري.

4-2 متغيرات البحث :

وقد تضمن الإستبيان المحاور التالية :

بيانات تعريفية عن العمر ، الجنس ، الحالة الاجتماعية ، المستوى التعليمي .

معلومات خاصة بمرض السكري الأسباب و الاعراض والمضاعفات والعلاقة ببعض الامراض المصاحبة لمرض السكري

الرقم	رمز المتغير	اسم المتغير	مستويات المتغير الوصفية	مستويات المتغير الرقمية
1	X1	العمر		
2	X2	الجنس	ذكر أنثى	1 2
3	X3	الحالة الاجتماعية	عازب متزوج مطلق أرمل	1 2 3 4

4	X4	المستوى التعليمي	أمي أساس ثانوي جامعي فوق الجامعة	1 2 3 4 5
5	X5	المهنة	طالب موظف عامل أعمال حرة بدون عمل	1 2 3 4 5
6	X6	مرض السكري	نعم لا	1 2
7	X7	نوع الإصابة	عرضية وراثية صدمة	1 2 3
8	X8	نوع مرض السكري	النوع الاول النوع الثاني	1 2
9	X9	هل انت مصاب بامراض اخري	نعم لا	1 2
10	X10	امراض القلب	نعم لا	1 2
11	X11	امراض العيون	نعم لا	1 2
12	X12	الكلى	نعم لا	1 2

13	X13	امراض الجهاز العصبي	نعم لا	1 2
14	X14	مدة المرض	أقل من 10 سنوات 10- 25 سنة أكثر من 25 سنة	1 2 3
15	X15	التاريخ العائلي	نعم لا	1 2
16	X16	ماهى درجة القرابة	الدرجة الاولى الدرجة الثانية	1 2
17	X17	هل توجد امراض وراثية اخرى فى العائلة	نعم لا	1 2
18	X18	التدخين	نعم لا	1 2
19	X19	الكحول	نعم لا	1 2
20	X20	حالة التغذية	منظم جيد غير منظم عدم المعرفة رفض	1 2 3 4
21	X21	عندما قمت بالفحص هل كنت	صائم فاطر	1 2
22		الرياضة	نعم لا	1 2

	X22			
23	ضغط الدم X23	طبيعي عالي		1 2
24	هل توجد مضاعفات X24	نعم لا		1 2
25	نوع المضاعفات X25	جرح باليد جرح بالرجل فتور بالجسم		1 2 3
26	كيف كانت نتيجة الفحص الاولى X26	يوجد سكري لا يوجد سكري		1 2
27	هل يوجد انخفاض فى السكرى X27	نعم لا		1 2
28	مدة الذهاب للفحص X28	دورياً أسبوعياً شهرياً		1 2 3
29	ارتفاع الدهون X29	نعم لا		1 2
30	هل الدهون تسبب لك مضاعفات X30	نعم لا		1 2
31	ادوية منظمة للسكر X31	نعم لا		1 2
32	بتر فى الاطراف	نعم		1

	X32		لا	2
33	X33	ممارسة الحياة الزوجية بدون مشاكل	نعم لا	1 2
34	X34	الاعراض	الصداع كثرة التبول زيادة في نبطات القلب	1 2 3
35	X35	نوع العلاج	أنوسلين داونيل فورميت حمية	1 2 3 4
36	X36	مقدار الجرعة	حبة حبتين 20 وحدة صباح ومساء 40-20 وحدة	1 2 3 4
37	X37	تنسيب كتلة الجسم	ضعيف مناسب أكثر من المطلوب سمين وزن زائد	1 2 3 4 5

قيمة الوسيط لكل عبارة

الرقم	رمز المتغير	اسم المتغير	مستويات المتغير الوصفية	مستويات المتغير الرقمية	قيمة الوسيط
1	X1	العمر			50
2	X2	الجنس	ذكر أنثى	1 2	2
3	X3	الحالة الاجتماعية	عازب متزوج مطلق أرمل	1 2 3 4	2
4	X4	المستوى التعليمي	أمي أساس ثانوي جامعي فوق الجامعة	1 2 3 4 5	3
5	X5	المهنة	طالب موظف عامل أعمال حرة بدون عمل	1 2 3 4 5	4
6	X6	مرض السكري	نعم لا	1 2	1
7	X7	نوع الإصابة	عرضية وراثية صدمة	1 2 3	2
8	X8	نوع مرض السكري	النوع الاول النوع الثاني	1 2	2
9	X9	هل انت مصاب بامراض اخري	نعم لا	1 2	1
10	X10	امراض القلب	نعم لا	1 2	2
11	X11	امراض العيون	نعم لا	1 2	2
12	X12	الكلى	نعم لا	1 2	2
13	X13	امراض الجهاز العصبي	نعم لا	1 2	2

14	مدة المرض	أقل من 10 سنوات 10- 25 سنة أكثر من 25 سنة	1 2 3	X14	1
15	التاريخ العائلي	نعم لا	1 2	X15	1
16	ماهى درجة القرابة	الدرجة الاولى الدرجة الثانية	1 2	X16	2
17	هل توجد امراض وراثية اخرى فى العائلة	نعم لا	1 2	X17	2
18	التدخين	نعم لا	1 2	X18	2
19	الكحول	نعم لا	1 2	X19	2
20	حالة التغذية	منظم جيد غير منظم عدم المعرفة رفض	1 2 3 4	X20	1
21	عندما قمت بالفحص هل كنت	صائم فاطر	1 2	X21	1
22	الرياضة	نعم لا	1 2	X22	1
23	ضغط الدم	طبيعي عالي	1 2	X23	1
24	هل توجد مضاعفات	نعم لا	1 2	X24	1
25	نوع المضاعفات	جرح باليد جرح بالرجل فتور بالجسم	1 2 3	X25	3
26	كيف كانت نتيجة الفحص الاولى	يوجد سكري لا يوجد سكري	1 2	X26	1
27	هل يوجد انخفاض فى السكرى	نعم لا	1 2	X27	1
28	مدة الذهاب للفحص	دورياً أسبوعياً شهرياً	1 2 3	X28	3
29	ارتفاع الدهون	نعم لا	1 2	X29	1
30	هل الدهون تسبب لك مضاعفات	نعم لا	1 2	X30	2
31	ادوية منظمة للسكر	نعم لا	1 2	X31	2

32	X32	بتر فى الاطراف	نعم لا	1 2	2
33	X33	ممارسة الحياة الزوجية بدون مشاكل	نعم لا	1 2	1
34	X34	الاعراض	الصداع كثرة التبول زيادة في نبطات القلب	1 2 3	2
35	X35	نوع العلاج	أنوسلين داونيل فورميت حمية	1 2 3 4	2
36	X36	مقدار الجرعة	حبة حبتين 20 وحدة صباح ومساء 40-20 وحدة	1 2 3 4	
37	X37	تنسيب كتلة الجسم	ضعيف مناسب أكثر من المطلوب سمين وزن زائد	1 2 3 4 5	4

4-3 التحليل الإحصائي

4-4 الإحصاءات الوصفية Descriptive Statistics

جدول رقم (4-1)

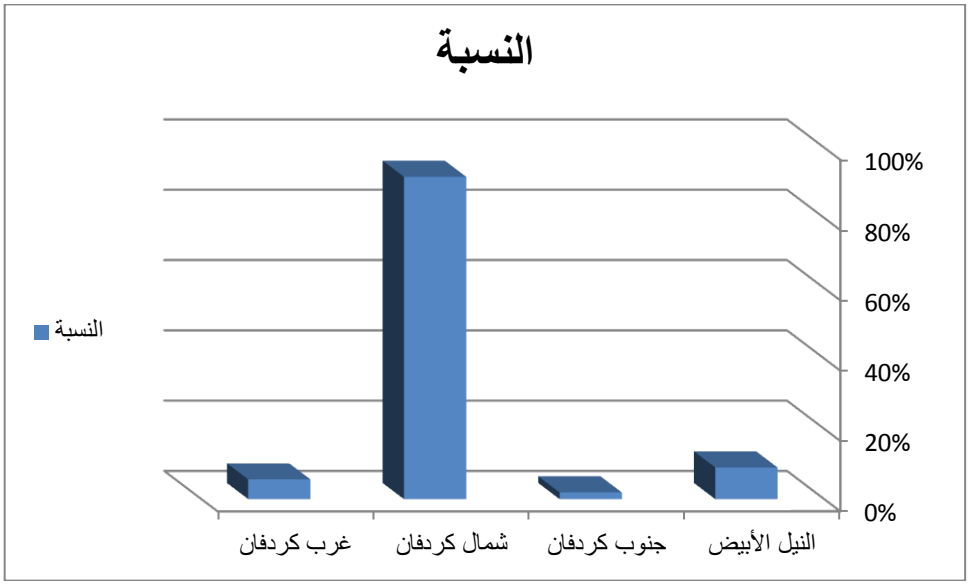
التوزيع التكراري البسيط لمكان الإقامة

الولاية	العدد	النسبة
النيل الأبيض	2	9%
جنوب كردفان	4	1.9%
شمال كردفان	198	91.7%
غرب كردفان	12	5.6%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم 4-1

التوزيع التكراري البسيط لمكان الإقامة



من الجدول رقم(1) والشكل رقم(1) نلاحظ ان 92.1% من الحالات المبحوثة من شمال كردفان و5.1%من غرب كردفان و1.9%من جنوب كردفان و0.9% من النيل الابيض ونستنتج من ذلك ان الغالبية العظمى من شمال كردفان يعزى ذلك الى قرب الموقع الجغرافى للمرضى القادمين من شمال كردفان .

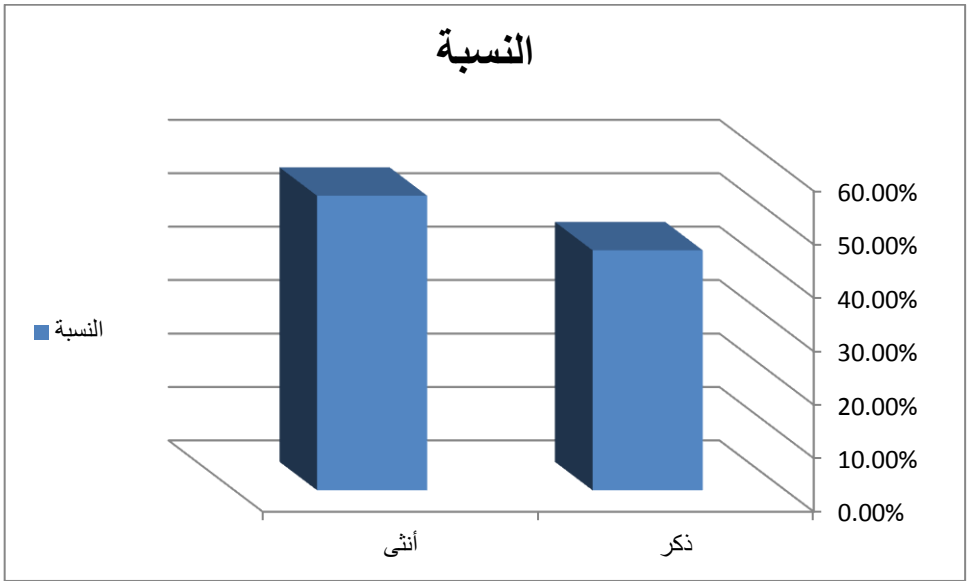
جدول رقم (4-2)

التوزيع التكراري البسيط لمتغيرالنوع

النوع	العدد	النسبة
ذكر	97	44.9%
أنثى	119	55.1%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-2)



من الجدول رقم2 و الشكل رقم2 نلاحظ ان نسبة الاصابة بالسكر فى اوساط الاناث وهى55.1% اكبر من نسبة الاصابة بالسكر فى اوساط الذكوروهى 44.9%.

جدول رقم (4-3)

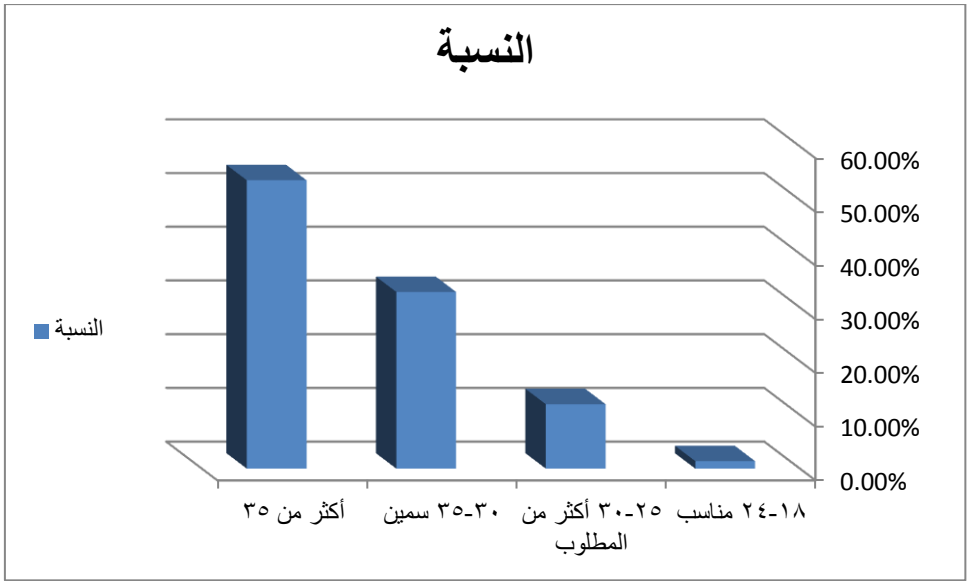
التوزيع التكراري البسيط لتنسيب كتلة الجسم (BMI)

تنسيب الكتلة	العدد	النسبة
18-24 مناسب	3	1.4%
25-30 أكثر من المطلوب	26	12.0%
30-35 سمين	71	32.9%
أكثر من 35	116	53.7%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-3)

التوزيع التكراري البسيط لتنسيب كتلة الجسم (BMI)



الجدول رقم3 والشكل رقم3 ان نسبة الوزن الزائدهى 53.70% وهى اكبر نسبة وان BMI السمين تصل نسبتة الى 32.87% ونسبة12.04% هى اكثر من المطلوب و1.39% هى مناسب

جدول رقم (4-4)

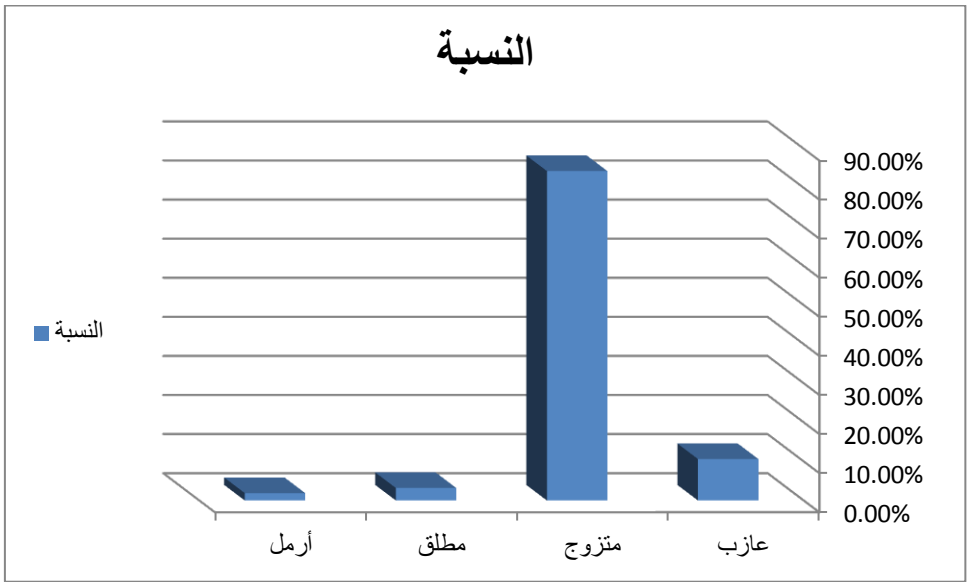
جدول التوزيع التكراري البسيط لمتغير الحالة الاجتماعية

الحالة الإجتماعية	العدد	النسبة
عازب	23	10.6%
متزوج	182	84.3%
مطلق	7	3.2%
أرمل	4	1.9%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-4)

التوزيع التكراري البسيط **للحاله الاجتماعيه**



نلاحظ من الجدول رقم4 والشكل رقم 4 ان اكبر نسبه للمتزوجين وهي 84.3% ونسبة غير المتزوجين هى 10.6% ونسبة الاصابه بالسكري فى وسط المطلقين والارامل هى 3.2%و 1.9%.

جدول رقم (4-5)

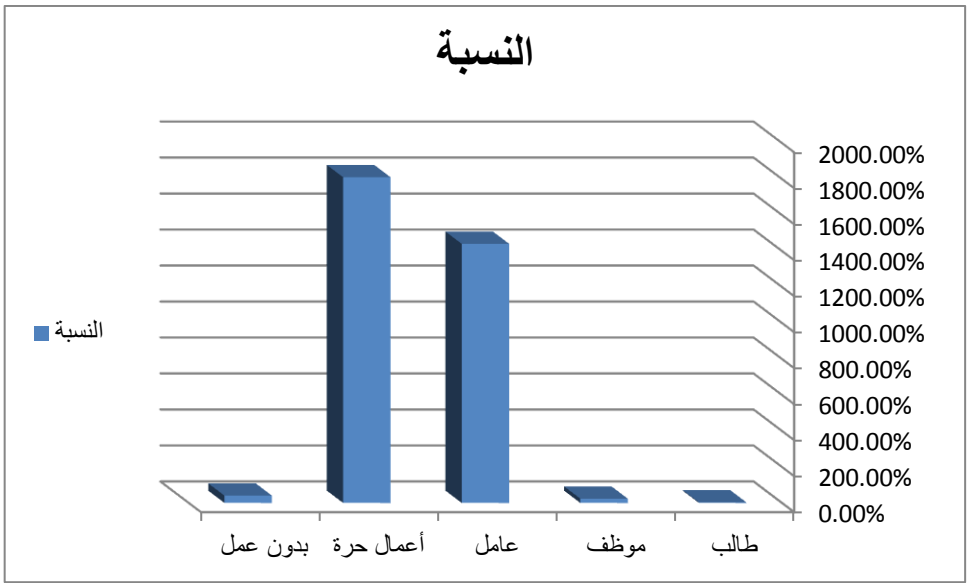
التوزيع التكراري البسيط لمتغير المهنه

المهنة	العدد	النسبة
طالب	7	3.2%
موظف	51	23.6%
عامل	31	14.4
أعمال حرة	39	18.1
بدون عمل	88	40.7%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-5)

التوزيع التكراري البسيط **لمتغير المهنه**



الجدول رقم5 والشكل رقم5 يوضح ان نسبة السكرى في وسط العاملين هى 14.4% وفى وسط الموظفين 23.6% وفى وسط الطلاب 3.2% وفي وسط الاعمال الحرة18.1% بينما نجد اكبر فى وسط بدون عمل 40.7%

جدول رقم (4-6)

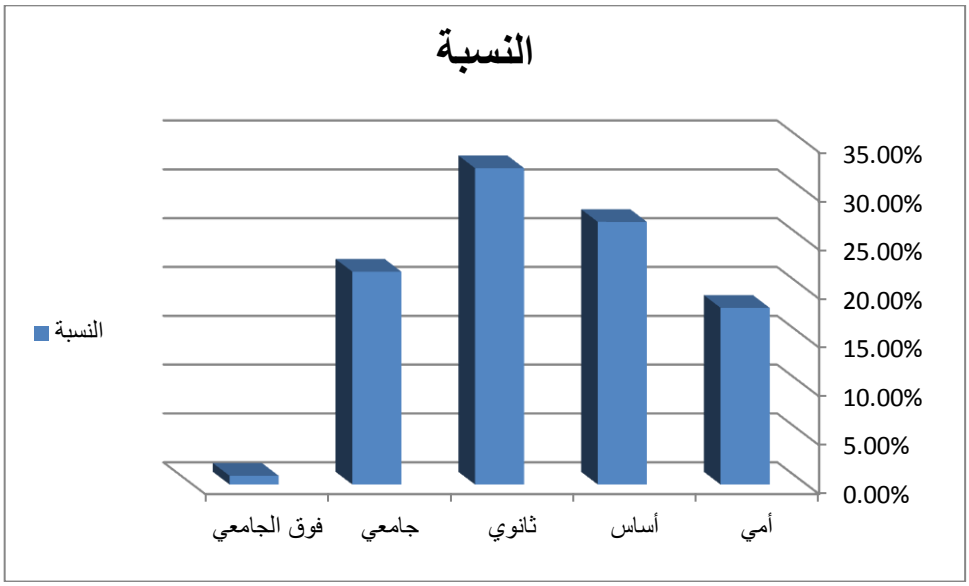
التوزيع التكراري البسيط **لمتغير المستوي التعليمي**

التعليمي المستوي	العدد	النسبة
أمي	39	18.1%
أساس	58	26.9%
ثانوي	70	32.4%
جامعي	47	21.8%
فوق الجامعي	2	0.9%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-6)

التوزيع التكراري البسيط لمتغير المستوي التعليمي



الجدول رقم 6 والشكل رقم 6 يوضح نسبة الاصابه بالسكري في وسط الذين تعلم الى مرحلة الثانوي32.45% وهى اكبر نسبة وفى وسط الذين تعلموا ومرحلة الاساس بلغت 26.9% والاميين كانت 18.1% ووسط الجامعيين وفوق الجامعى بلغت 21.8% و0.9%.

جدول رقم (4-7)

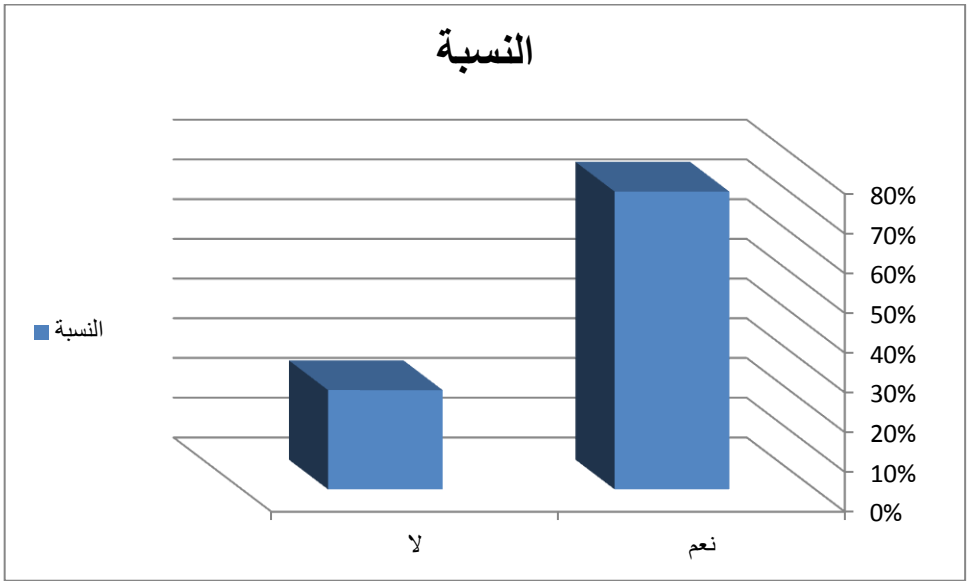
التوزيع التكراري البسيط **لمتغير للإصابة بالسكرى**

الإصابة بالسكري	العدد	النسبة
نعم	162	%75
لا	54	%25
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

التوزيع التكراري البسيط **لمتغير للإصابة بالسكرى**

شكل رقم (4-7)



الجدول رقم 7 والشكل رقم7 يبين ان الاصابة بالسكرى بلغت 75%وهم الاشخاص المصابين و25%هم غير المصابين.

جدول رقم (4-8)

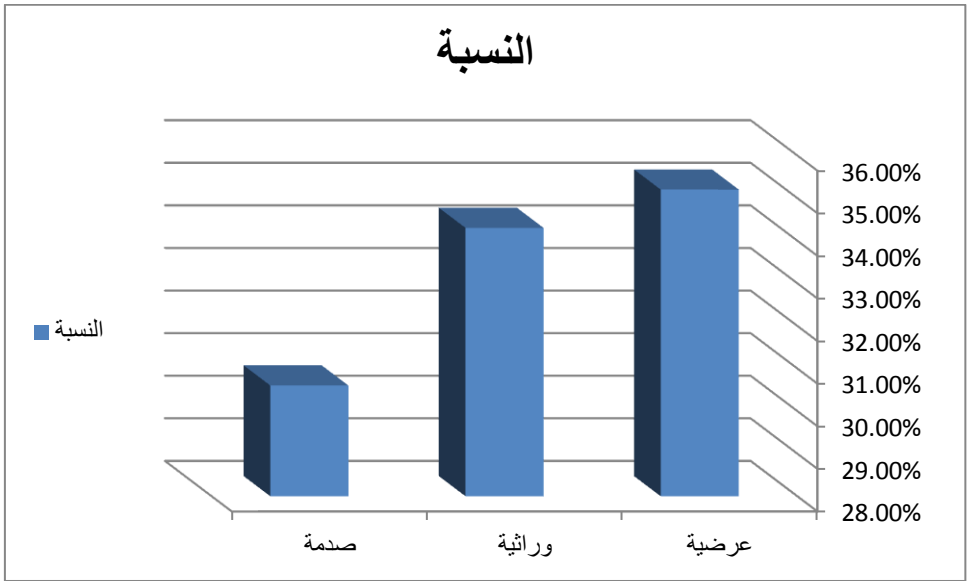
التوزيع التكراري البسيط **لمتغير نوع الاصابة**

نوع الإصابة	العدد	النسبة
عرضية	76	%35.2
وراثية	74	%34.3
صدمة	66	30.6%
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-8)

التوزيع التكراري البسيط **لمتغير نوع الاصابة**



الجدول رقم8 والشكل رقم 8 يوضح ان نوع الاصابة بالسكر اكثرها كانت اعراض بنسبة 35.2% وتاتى بعدها الوراثة بنسبة 34.3% ثم الصدمة بنسبة 30.6%.

جدول رقم (4-9)

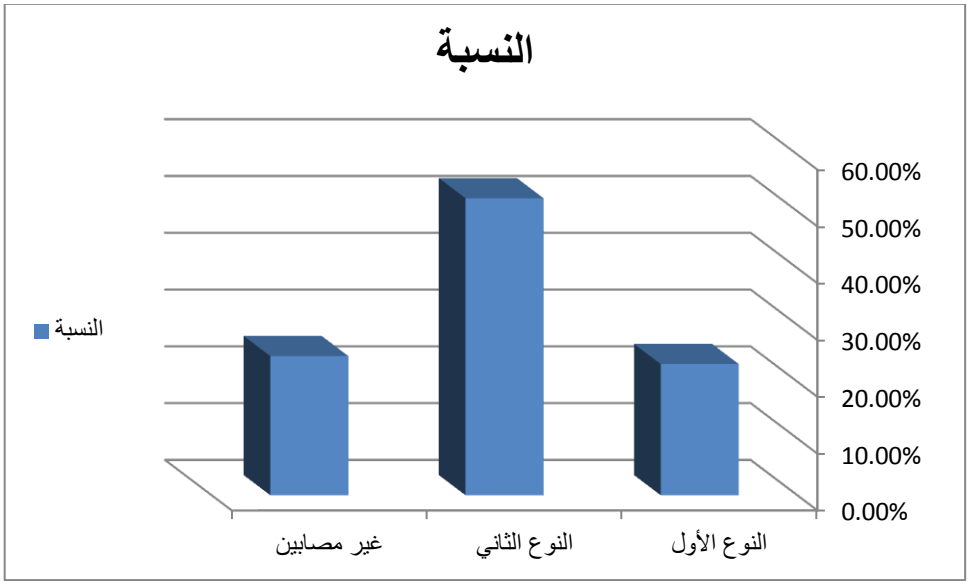
التوزيع التكراري البسيط **لمتغير نوع مرض السكر**

نوع السكري	العدد	النسبة
النوع الأول	50	%23.1
النوع الثاني	113	%52.3
غير مصابين	53	%24.5
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-9)

التوزيع التكراري البسيط **لمتغير نوع مرض السكري**



الجدول رقم 9 والشكل رقم 9 يوضح ان الاصابة بالسكر من النوع الثانى اعلى من النوع الاول حيث بلغت نسبتة فى النوع الثانى52.3% والاول 23.1%.

جدول رقم (4-10)

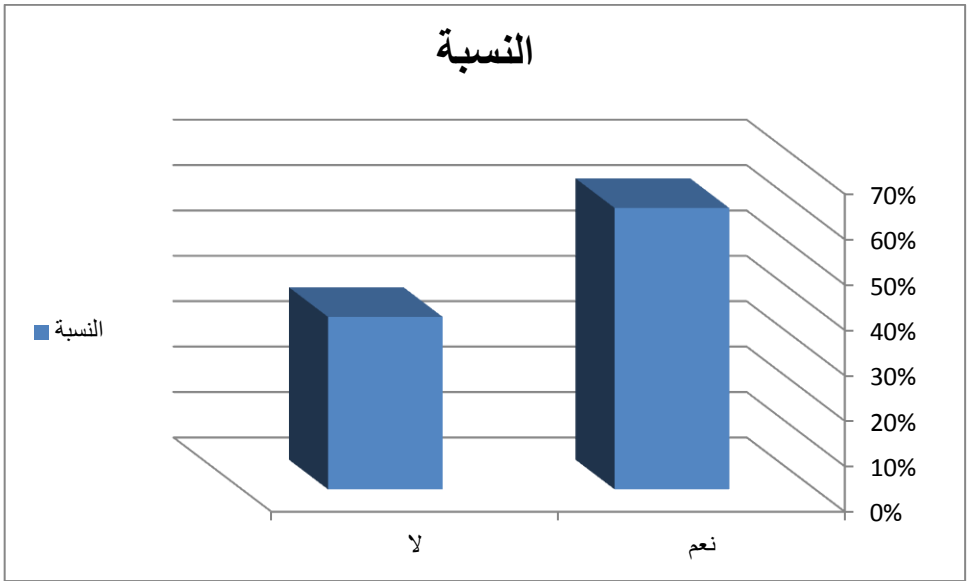
التوزيع التكراري البسيط **لمتغير الإصابة بامراض اخرى**

الإصابة بأمراض أخري	العدد	النسبة
نعم	134	%62
لا	82	%38
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-10)

التوزيع التكراري البسيط **لمتغير الإصابة بامراض اخرى**



نلاحظ من خلال الجدول رقم10 والشكل رقم10 حيث بلغت الاصابة بالامراض الاخرى نسبت المصابين كانت 62% وغير المصابين 38%.

جدول رقم (4-11)

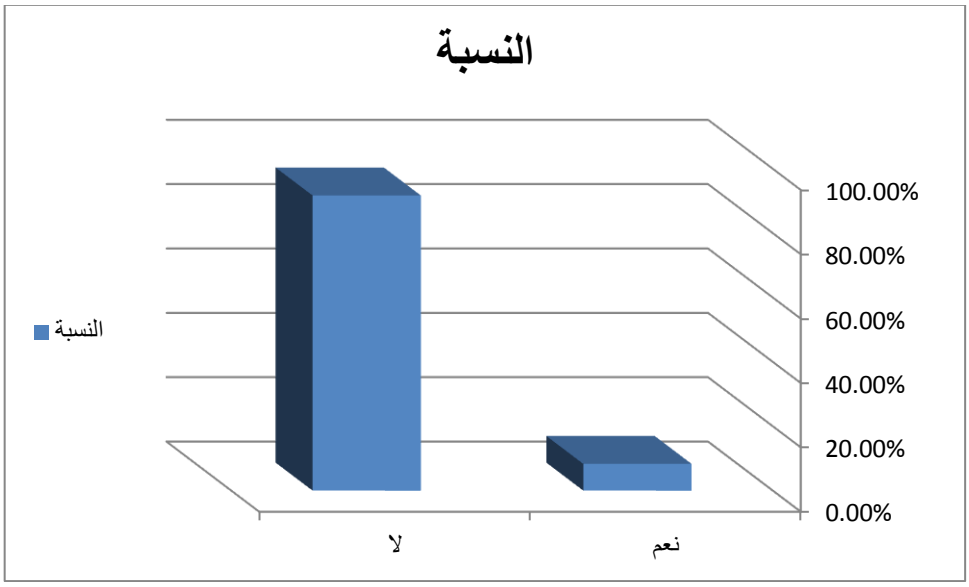
التوزيع التكراري البسيط **لمتغير امراض القلب**

أمراض القلب	العدد	النسبة
نعم	18	%8.3
لا	198	%91.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-11)

التوزيع التكراري البسيط **لمتغير امراض القلب**



الجدول رقم 11 والشكل رقم11 يوضح نسبة الاشخاص المصابين بامراض القلب وكانت 8.3% وغير المصابين 91.7%.

جدول رقم (4-12)

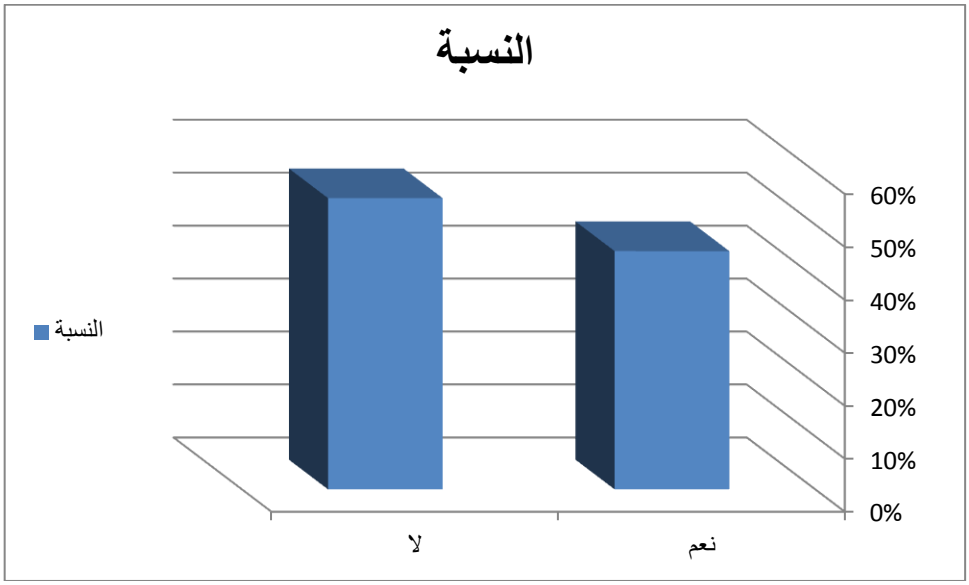
التوزيع التكراري البسيط لمتغير امراض العيون

أمراض العيون	العدد	النسبة
نعم	99	%45
لا	117	%55
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-12)

التوزيع التكراري البسيط **لمتغير امراض العيون**



الجدول رقم12 والشكل رقم 12 يبين ان نسبة الاشخاص المصابين بمرض العيون كانت 45.8% وغير المصابين 53.7%

جدول رقم (4-13)

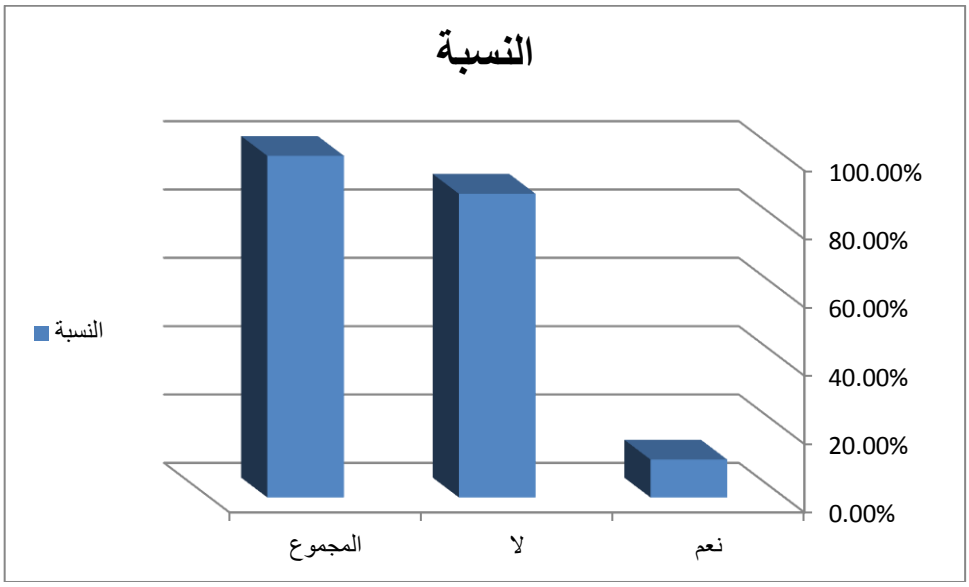
التوزيع التكراري البسيط **لمتغير امراض الكلى**

أمراض الكلى	العدد	النسبة
نعم	24	%11.1
لا	192	%88.9
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-13)

التوزيع التكراري البسيط **لمتغير امراض الكلى**



نلاحظ من الجدول رقم13 والشكل رقم 13 نجد ان الاشخاص المصابين بامراض الكلى كانت نسبتهم 11.1% و غير المصابين 88.9%.

جدول رقم (4-14)

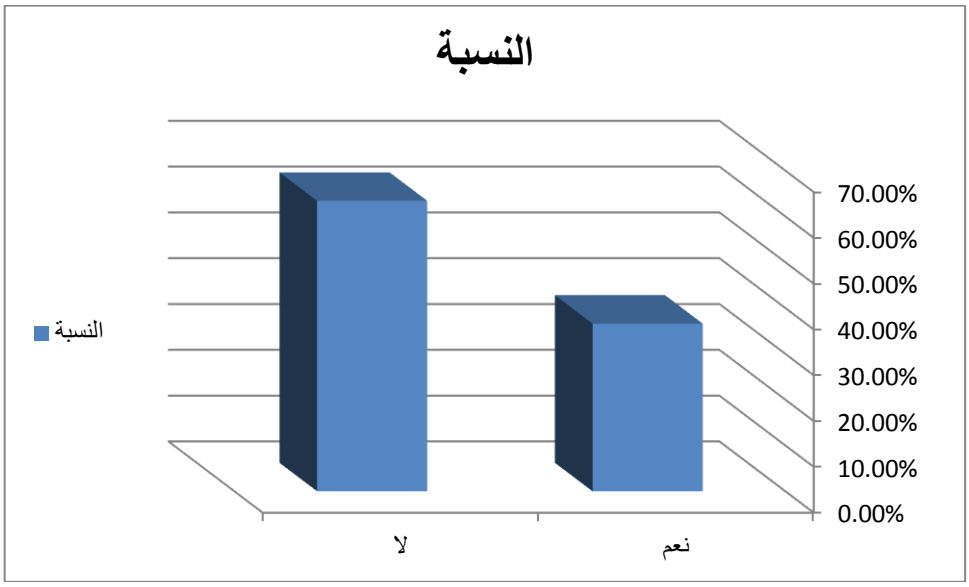
التوزيع التكراري البسيط **لمتغير** الجهاز العصبى

أمراض الجهاز العصبي	العدد	النسبة
نعم	79	%36.6
لا	137	%63.4
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-14)

التوزيع التكراري البسيط **لمتغير** الجهاز العصبى



نلاحظ من الجدول رقم14 والشكل رقم 14 انسبة المصابين هى36.57% وغير المصابين 63.43%.

جدول رقم (4-15)

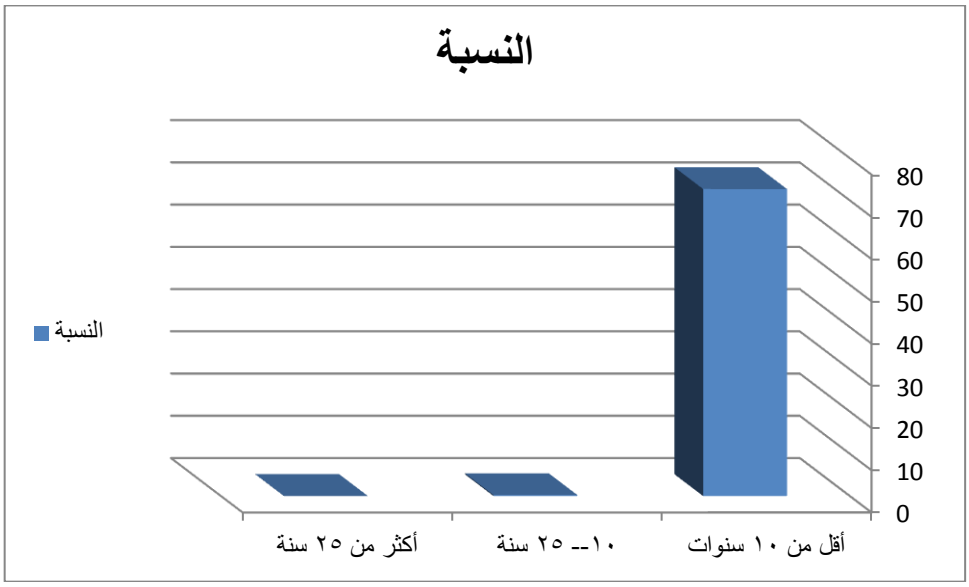
التوزيع التكراري البسيط لمتغير مدة المرض

مدة المرض	العدد	النسبة
أقل من 10 سنوات	120	72.7
25 --10سنة	40	%24.2
أكثر من 25 سنة	5	%3
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-15)

التوزيع التكراري البسيط **لمتغير مدة المرض**



نلاحظ من الجدول رقم 15 الشكل رقم15 يبين ان مدة المرض للمصابين بالسكر اكثرها كانت اقل من 10سنة بنسبة55.6% و18.5%فى المدة من 10-25سنة و2.3%فى المدة اكثرمن 25سنة.

جدول رقم (4-16)

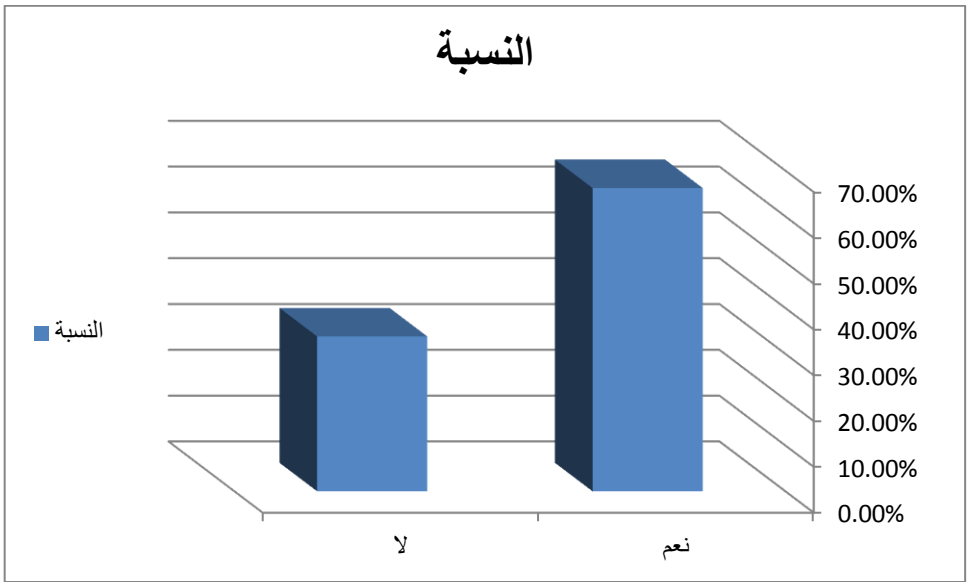
التوزيع التكراري البسيط **لمتغير التاريخ العائلي للاضابة بالسكري**

التاريخ العائلي	العدد	النسبة
نعم	143	%66.2
لا	73	%33.8
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-16)

التوزيع التكراري البسيط **لمتغير التاريخ العائلي للاضابة بالسكري**



نلاحظ من الجدول رقم 16 الشكل رقم16 والجدول ان نسبة الاصابة بالسكر وسط العائلة بلغت 66.2% وغير المصابين33.8%.

جدول رقم (4-17)

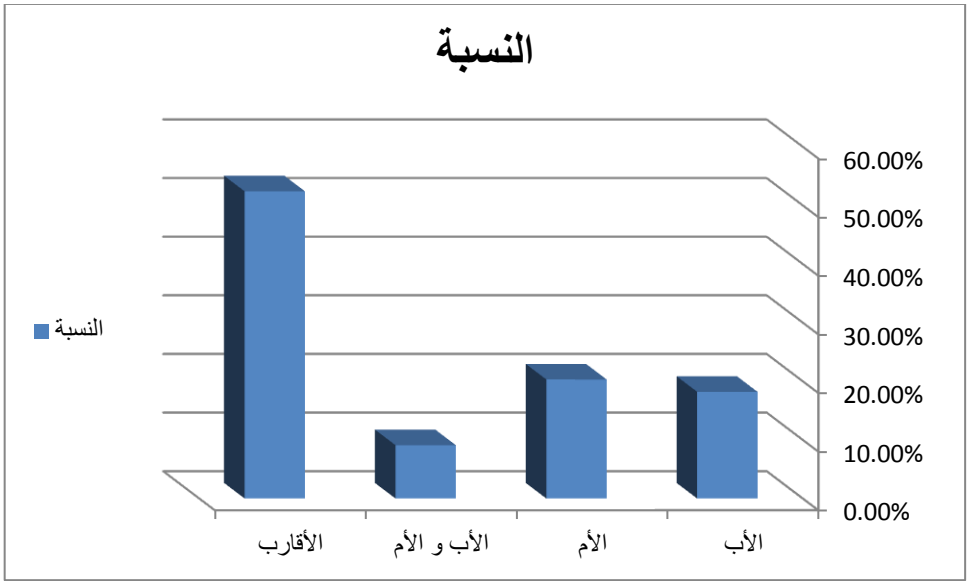
التوزيع التكراري البسيط لمتغير درجة القرابة

درجة القرابة	العدد	النسبة
الأب	26	%18.2
الأم	29	%20.3
الأب و الأم	13	%9.1
الأقارب	75	%52.4
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-17)

التوزيع التكراري البسيط **لمتغير درجة القرابة**



نلاحظ من الجدول رقم17 والشكل رقم 17 أن درجة القرابة كانت اعلى نسبة حيث بلغت 34.7% فى وسط الاقارب حيث كانت القرابة من جهة الام بنسبة 13.4%والاب بـ12% والاب والام بلغت 6%.

جدول رقم (4-18)

التوزيع التكراري البسيط لمتغير امراض وراثية فى العائلة

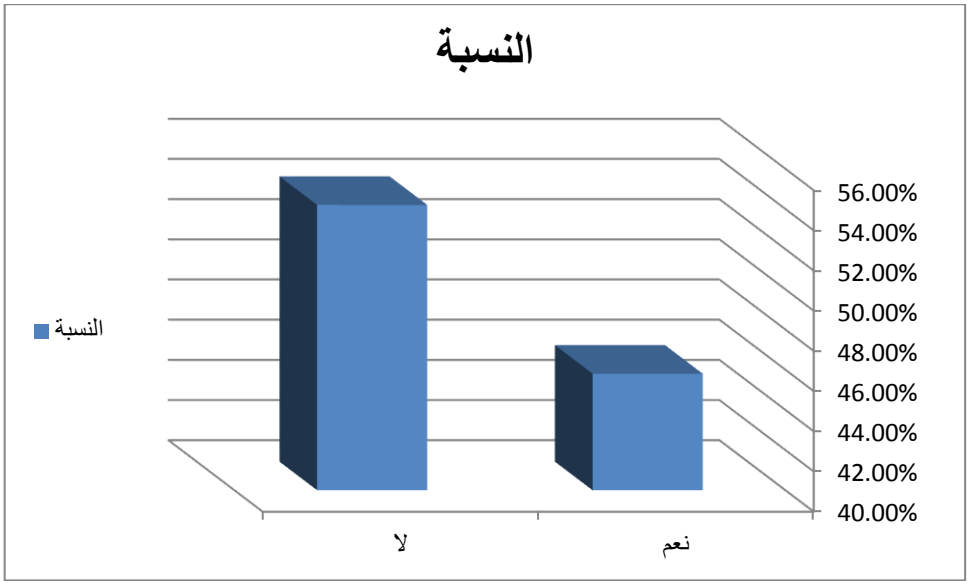
امراض وراثية	العدد	النسبة

نعم	99	%45.8
لا	117	%54.2
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-18)

التوزيع التكراري البسيط **لمتغير درجة القرابة**



نلاحظ من الجدول رقم18 والشكل رقم18 حيث بلغت نسبة الامراض الوراثية التى وجدت كانت بنسبة45.8% والتى لم توجد بنسبة 54.2%.

جدول رقم (4-19)

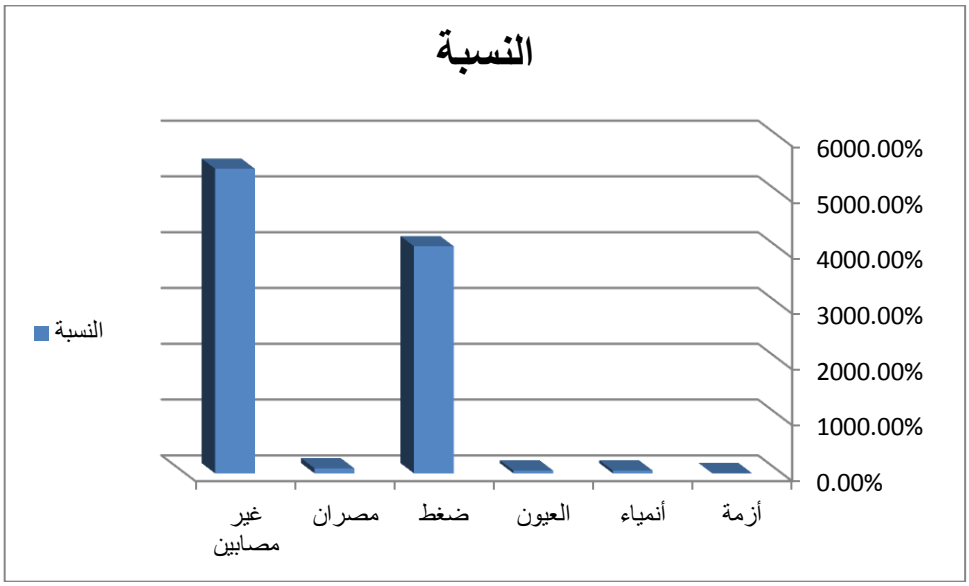
اذا كانت الإجابة نعم اذكرها

	العدد	النسبة

أزمة	6	%2.8
أنمياء	1	0.5
العيون	1	0.5
ضغط	88	40.7
مصران	2	0.9
غير مصابين	117	54.6
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-19)



الجدول رقم 19 والشكل رقم19 يوضح ان نسبة الاصابة بالضغط كانت اعلى نسبةوهى 40.7%والازمة بلغت 2.8% والمصران 0.9%والعيون والانمياء بنسبة 0.5% على التوالى.

جدول رقم (4-20)

التوزيع التكراري البسيط **لمتغير التدخين**

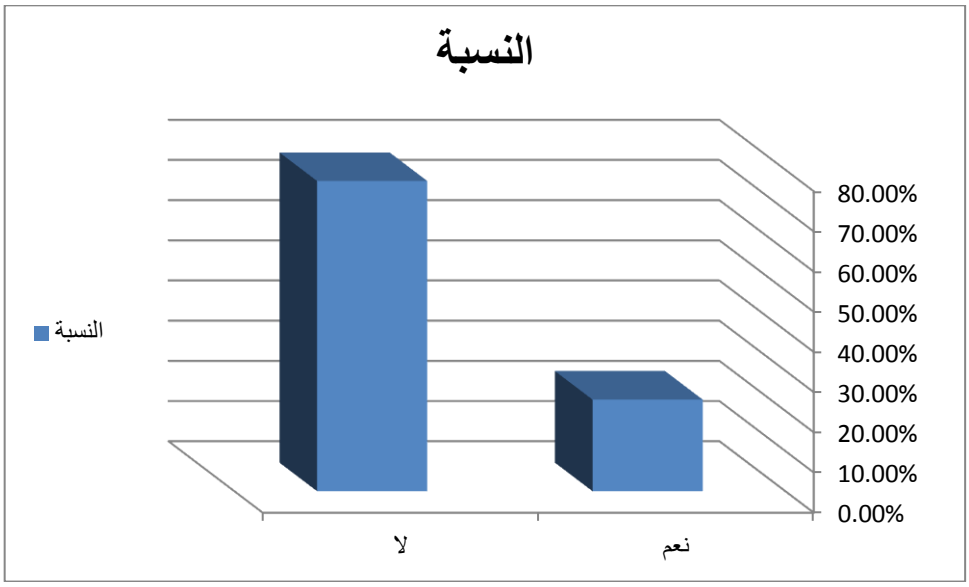
التدخين	العدد	النسبة

نعم	49	%22.8
لا	166	%77.2
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-20)

التوزيع التكراري البسيط **لمتغير التدخين**



نلاحظ من الجدول رقم 20 والشكل رقم 20 حيث بلغت نسبة التدخين فى وسط المصابين بالسكري22.7% وغير المدخنين بـ76.9%

جدول رقم (4-21)

التوزيع التكراري البسيط لمتغير تناول الكحول باستمرار

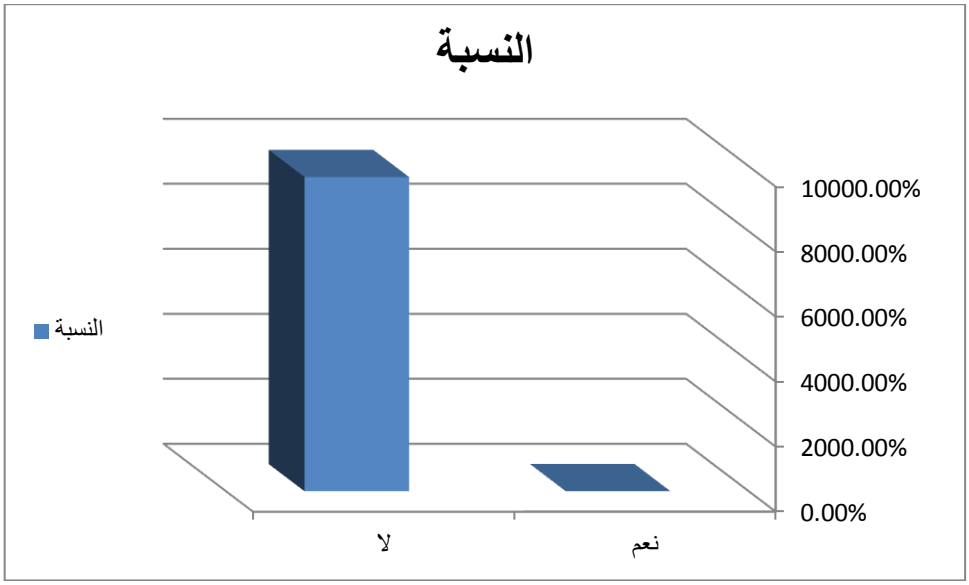
الكحول	العدد	النسبة

نعم	7	%3.3
لا	208	96.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-21)

التوزيع التكراري البسيط **لمتغير تناول الكحول باستمرار**



نلاحظ من الجدول رقم21 الشكل رقم21 يوضح نسبة تناول الكحول وسط المصابين بالسكرى كانت النسبة 3.2% والذين لا يتناولون بـ96.3%

جدول رقم (4-22)

التوزيع التكراري البسيط **لمتغير الرياضة**

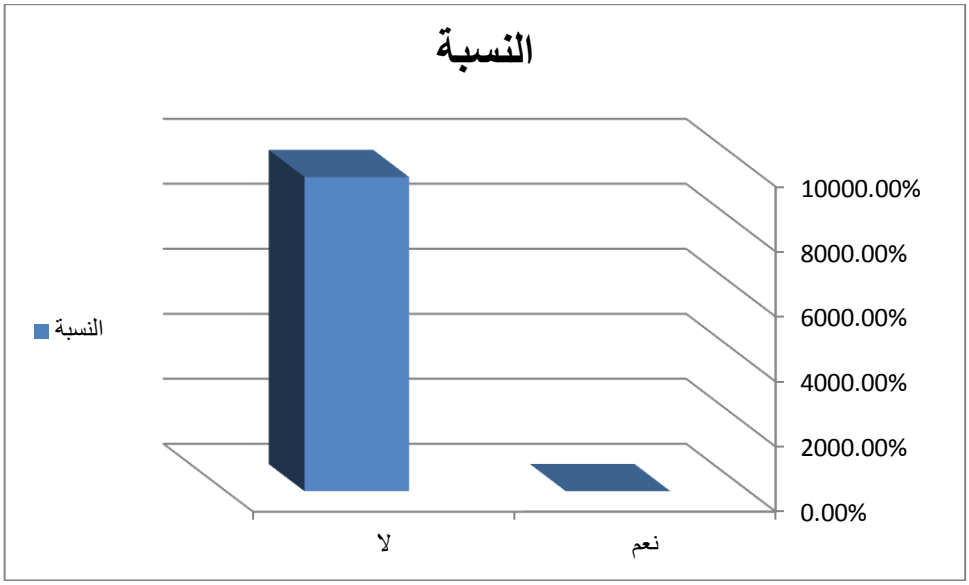
الكحول	العدد	النسبة

نعم	7	%3.3
لا	208	96.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-22)

التوزيع التكراري البسيط **لمتغير الرياضة**



الجدول رقم 22 والشكل رقم22 يبين ان الاشخاص المصابين بالسكرى ويمارسون الرياضة كانت نسبتهم 63%و الذين لا يمارسون الرياضة كانت نسبتهم 37%

جدول رقم (4-23)

التوزيع التكراري البسيط لمتغير حالة التغذية

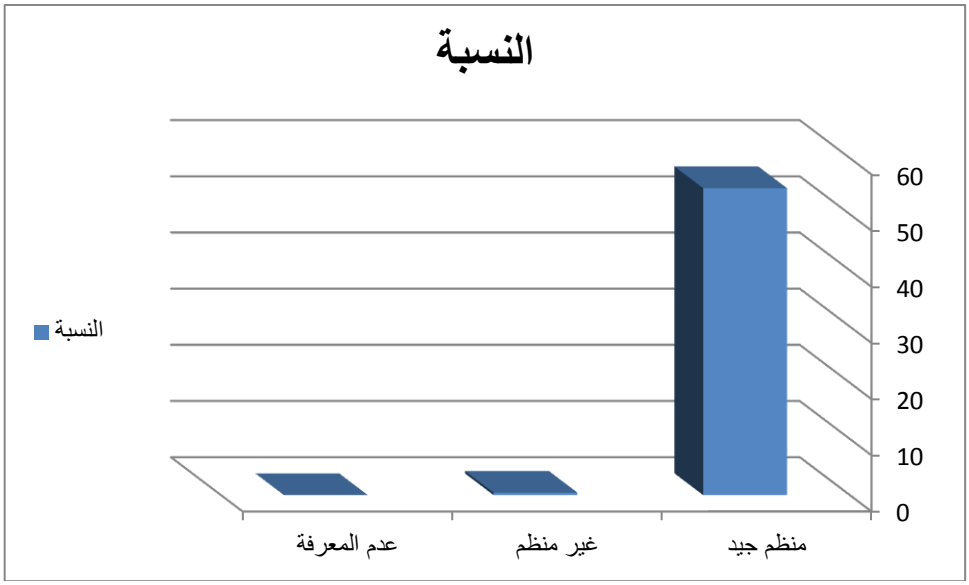
التغذية	العدد	النسبة

منظم جيد	118	54.65
غير منظم	97	%44.9
عدم المعرفة	1	%0.5
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-23)

التوزيع التكراري البسيط **لمتغير حالة التغذية**



الجدول رقم 23 والشكل رقم23 يبين حالة التغذية وسط المصابين بالسكرى حيث كان التنظيم الجيد اعلى نسبة بلغ 54.6% وغير المنظم بلغ 44.9% وعدم المعرفه 0.5%

جدول رقم (4-24)

التوزيع التكراري البسيط لمتغير الصيام

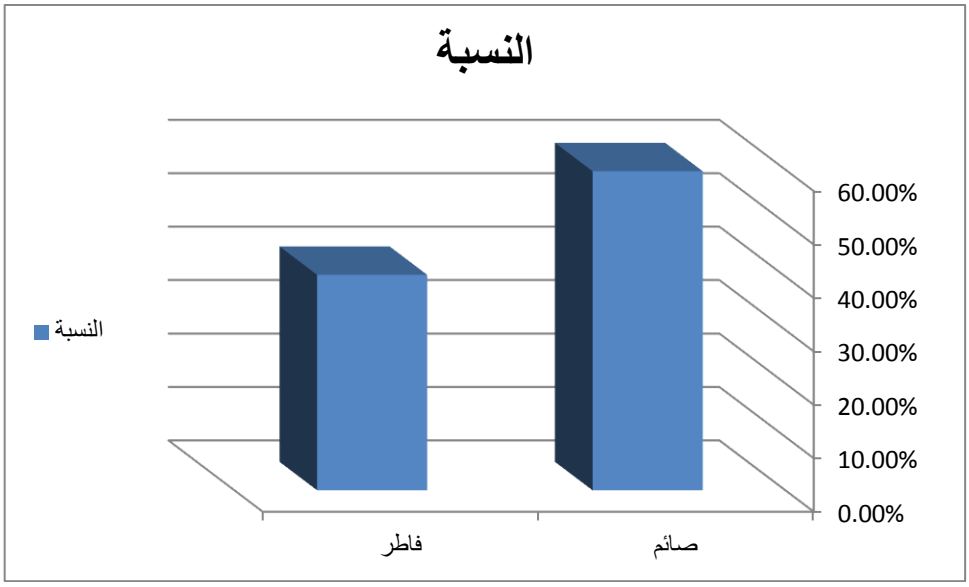
الصيام	العدد	النسبة

صائم	129	%59.7
فاطر	87	%40.3
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-24)

التوزيع التكراري البسيط **لمتغير الصيام**



الجدول رقم 24 والشكل رقم 24 يبين ان الاشخاص الذين قدموا للفحص وهم صائمون كانت نسبتهم 59.7% والفاطرون بنسبة40.3%

جدول رقم (4-25)

التوزيع التكراري البسيط لمتغير **ضغط الدم**

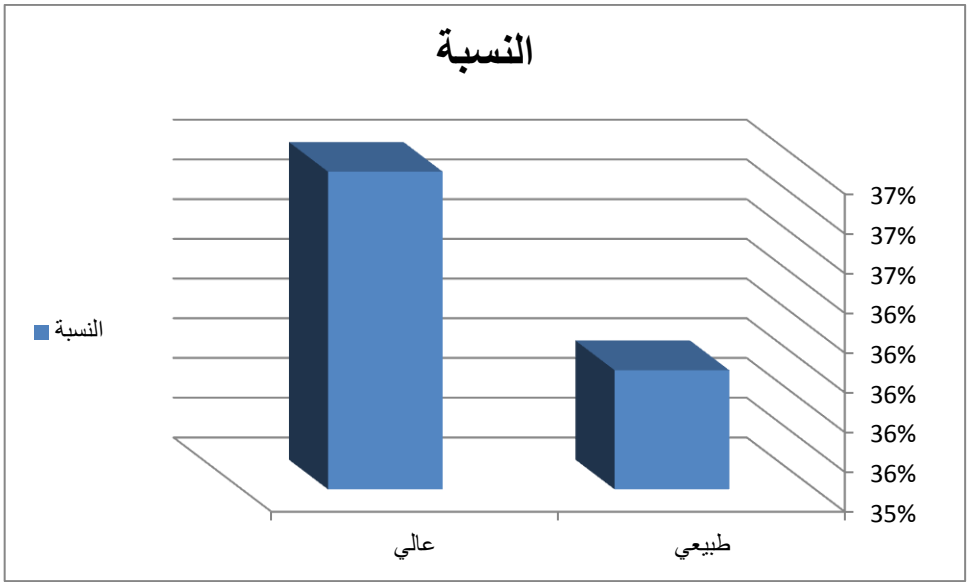
ضغط الدم	العدد	النسبة

طبيعي	136	%36
عالي	80	%37
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-25)

التوزيع التكراري البسيط لمتغير **ضغط الدم**



الجدول رقم 25 والشكل رقم 25 يوضح ضغط الدم للمرضى وكانت الذين ضغطهم طبيعى 63% وعالى بلغت 37%

جدول رقم (4-26)

التوزيع التكراري البسيط لمتغير المضاعفات

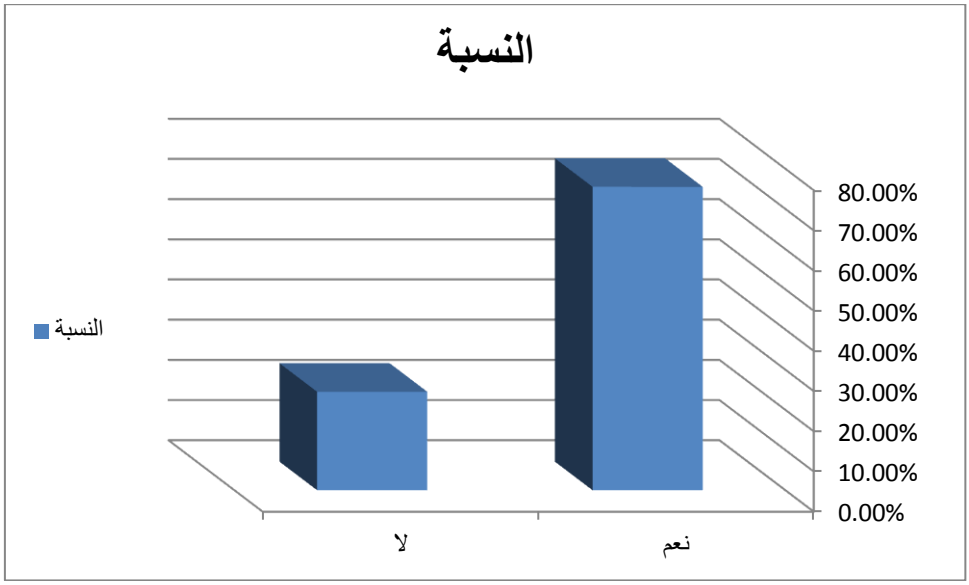
المضاعفات	العدد	النسبة

نعم	163	%75.5
لا	53	%24.5
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-26)

التوزيع التكراري البسيط لمتغير **المضاعفات**



الجدول رقم 26 والشكل رقم 26هذا يوضح نوع المضاعفات وكانت اكثرها فتور بالجسم بنسبة62.5% وجرح بالرجل بلغ 8.3% وجرح باليد بـ4.6%.

جدول رقم (4-27)

التوزيع التكراري البسيط لمتغير نتيجة الفحص الاولى

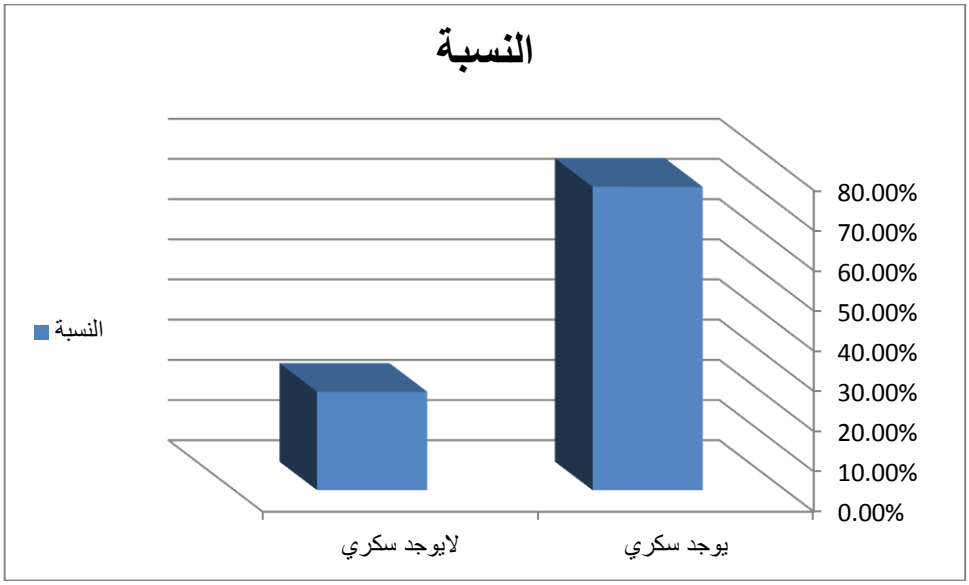
نتجة الفحص الاولى	العدد	النسبة

يوجد سكري	163	%75.5
لايوجد سكري	53	%24.5
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-27)

التوزيع التكراري البسيط لمتغير **نتيجة الفحص الاولى**



الجدول رقم 27 والشكل رقم 27 يبين نتيجة الفحص الاولى وبلغ عدد المصابين 75% وغير المصابين 24.

جدول رقم (4-28)

التوزيع التكراري البسيط لمتغير **مدة الذهاب للفحص**

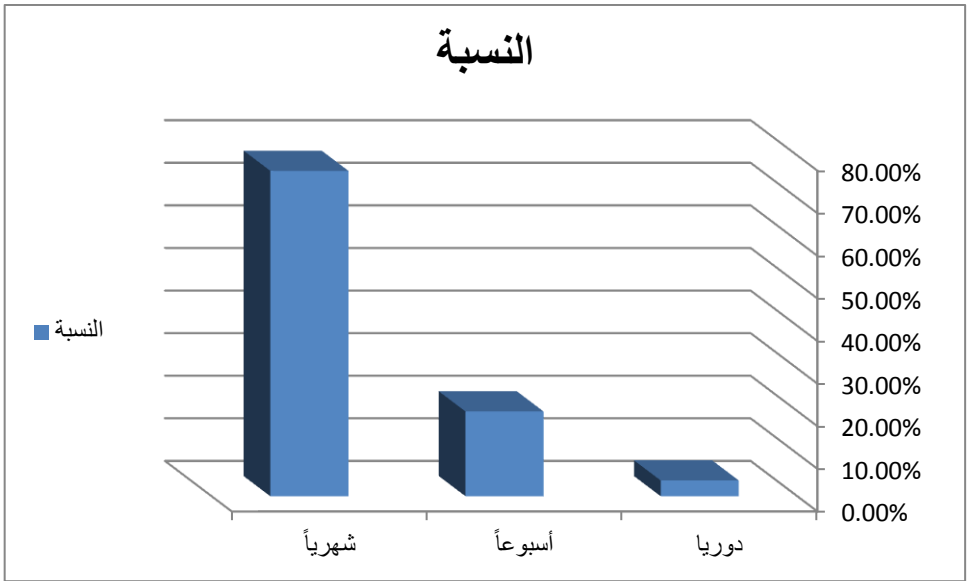
مدة الذهاب للفحص	العدد	النسبة

دوريا	8	%3.7
أسبوعاً	43	%19.9
شهرياً	165	%76.4
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-28)

التوزيع التكراري البسيط لمتغير **مدة الذهاب للفحص**



من الجدول رقم 28 والشكل رقم 28 نلاحظ ان مدة الذهاب للفحص شهريا اكثرها نسبة حيث كانت 76.4%اما اسبوعيا كانت 19.9% واخيرا دوريا بلغت 3.7%.

جدول رقم (4-29)

التوزيع التكراري البسيط لمتغير **ارتفاع الدهون الثلاثة**

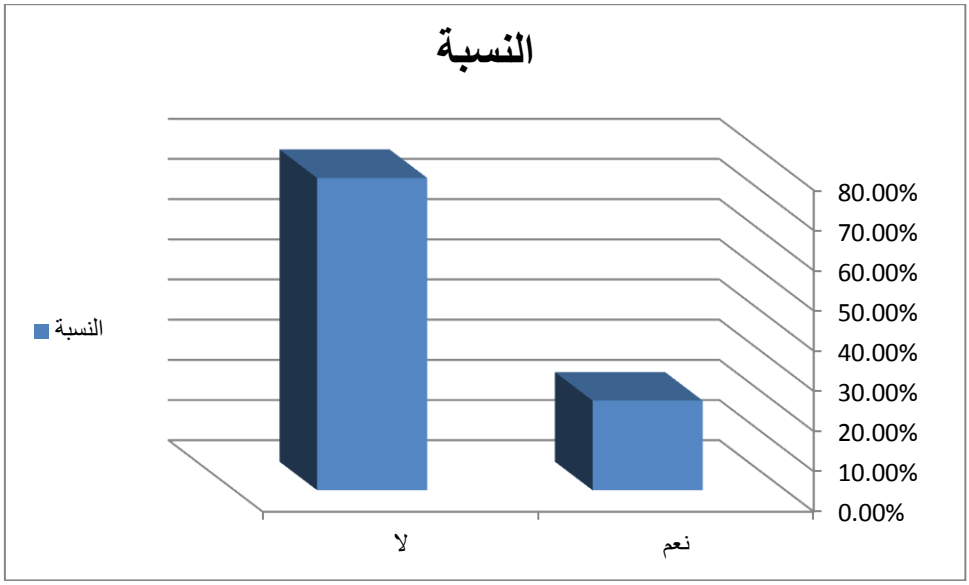
ارتفاع الدهون	العدد	النسبة

نعم	48	%22.3
لا	167	%77.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-29)

التوزيع التكراري البسيط لمتغير **ارتفاع الدهون الثلاثة**



نلاحظ من الجدول رقم 29والشكل رقم 29 ان ارتفاع الدهون الثلاثة بلغ وسط المصابين 22.% وغيرهم بـ77.3%.

جدول رقم (4-30)

التوزيع التكراري البسيط لمتغير الدهون تسبب لك مضاعفات

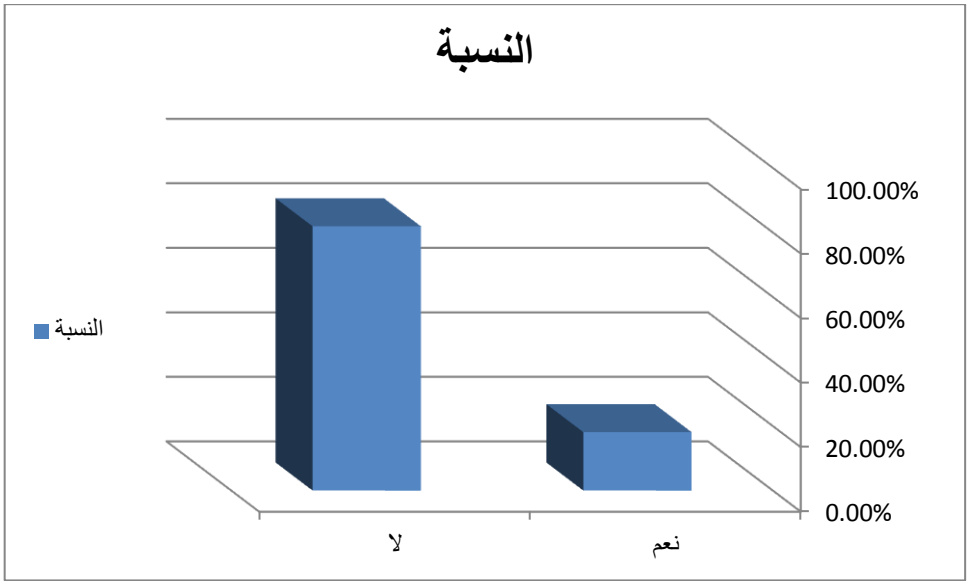
هل الدهون تسبب مضاعفات	العدد	النسبة

نعم	39	%18.1
لا	177	%81.9
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-30)

التوزيع التكراري البسيط لمتغير **الدهون تسبب لك مضاعفات**



الجدول رقم 30 والشكل رقم 30 يبين ان الاشخاص الذين تسبب لهم الدهون بمضاعفات بلغ عددهم 18.1% والذين لاتسبب لهم بـــ81.9%

جدول رقم (4-31)

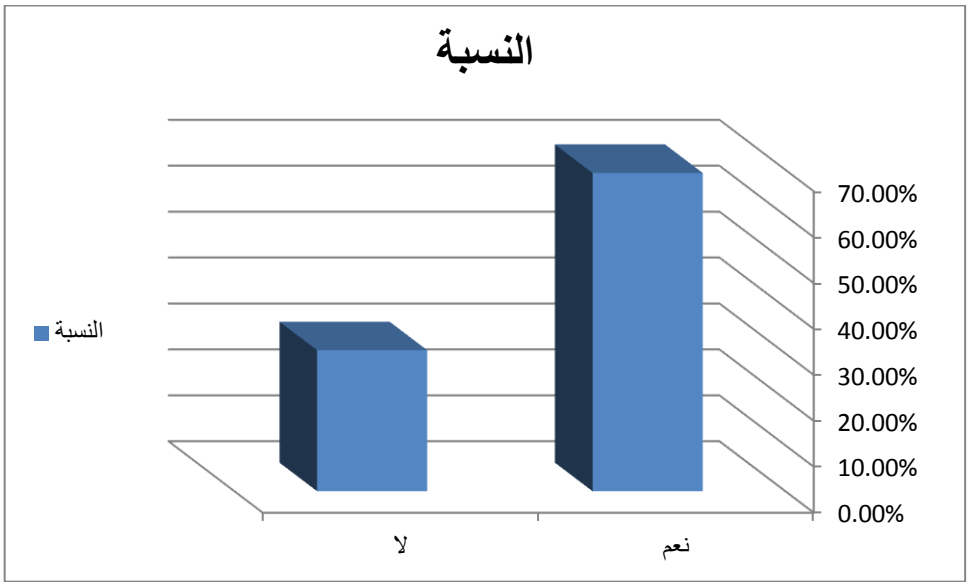
التوزيع التكراري البسيط لمتغير إنخفاض فى السكر

إنخفاض السكر	العدد	النسبة
نعم	113	%69.3
لا	50	%30.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-31)

التوزيع التكراري البسيط لمتغير **إنخفاض فى السكري**



نلاحظ من الجدول رقم31 والشكل رقم 31 إن نسبة انخفاض السكرى كانت بنسبة52.3% والتى لم تنخفض بـــــ23.1%.

جدول رقم (4-32)

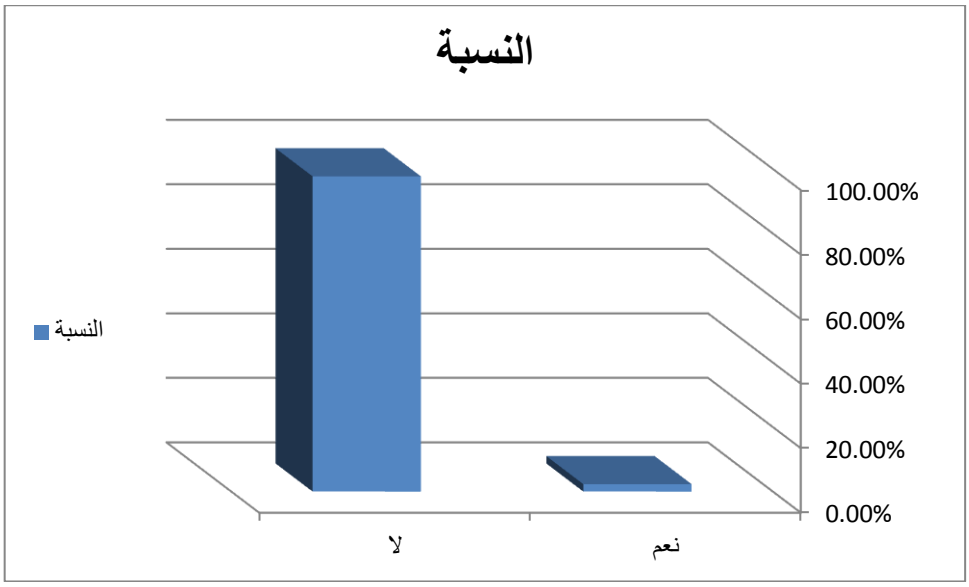
التوزيع التكراري البسيط لمتغير استخدام ادوية منظمة للسكري

أدوية منظمة للسكر	العدد	النسبة
نعم	5	%2.3
لا	211	%97.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-32)

التوزيع التكراري البسيط لمتغير استخدام ادوية منظمة للسكري



نلاحظ من الجدول رقم32 والشكل رقم32 ان نسبة استخدام ادوية منظمة للسكر كانت 2.3% و97.% لم يستخدمو ا ادوية منظمة للسكر.

جدول رقم (4-33)

التوزيع التكراري البسيط لمتغير **بتر الاطراف**

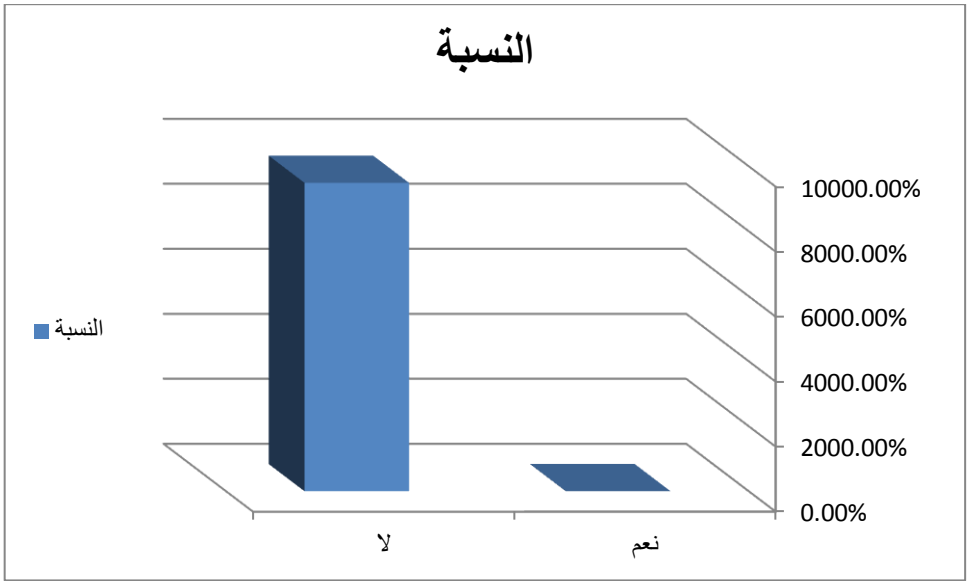
بتر الاطراف	العدد	النسبة

نعم	11	%5.1
لا	205	94.9
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-33)

التوزيع التكراري البسيط لمتغير **بتر الاطراف**



الجدول رقم 33 والشكل رقم33 يوضح ان البتر فى الاطراف كانت نسبتهم 5.1% و94.9% لايوجد لديهم بتـــــر.

جدول رقم (4-34)

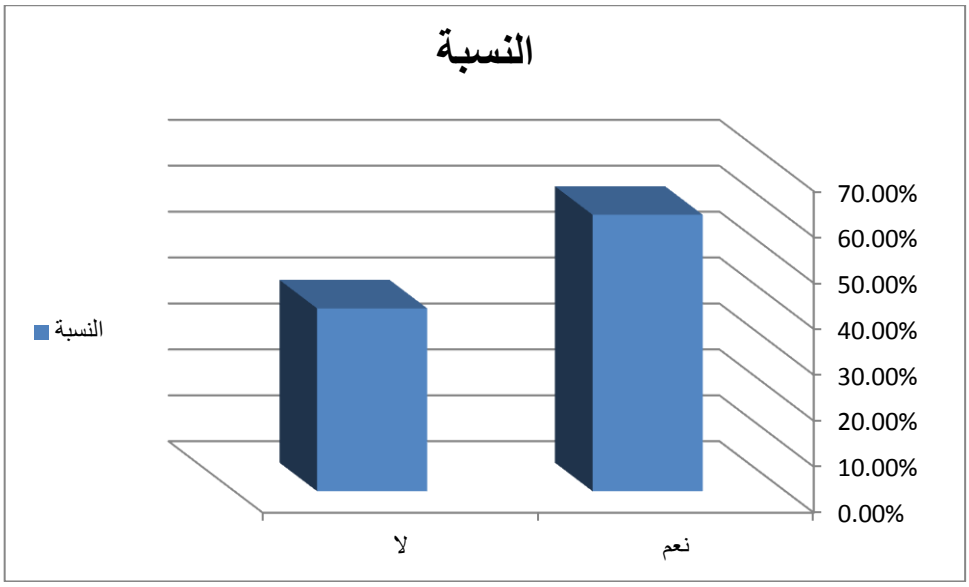
التوزيع التكراري البسيط لمتغير الضعف الجنسي

الضعف الجنسي	العدد	النسبة
نعم	130	%60.2
لا	86	%39.8
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-34)

التوزيع التكراري البسيط لمتغير الضعف الجنسي



نلاحظ من الجدول رقم 34 الشكل رقم 34 يوضح ان نسبة 60%يمكنهم ممارسة الحياة الزوجية بدون مشاكل ونسبة39.8% لايمكنهم ممارسة الحياة الزوجية

جدول رقم (4-35)

التوزيع التكراري البسيط لمتغير **الاعراض**

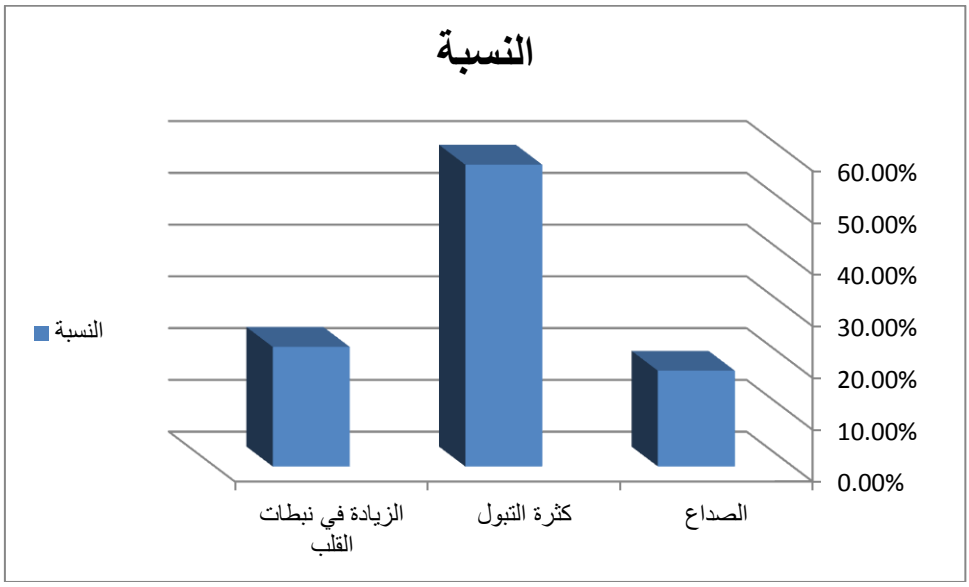
الأعراض	العدد	النسبة

الصداع	40	%18.5
كثرة التبول	126	%58.3
الزيادة في نبطات القلب	50	%23.1
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-35)

التوزيع التكراري البسيط لمتغير **الاعراض**



نلاحظ من الجدول رقم 35 الشكل رقم 35 يوضح ان اكثر الاعراض مسببة لمرض السكر هى كثرة التبول بنسبة58.3% وذيادة نبضات القلب بـ23.1% والصـــــــداع بـــ 18%

جدول رقم (4-36)

التوزيع التكراري البسيط لمتغير نوع العلاج

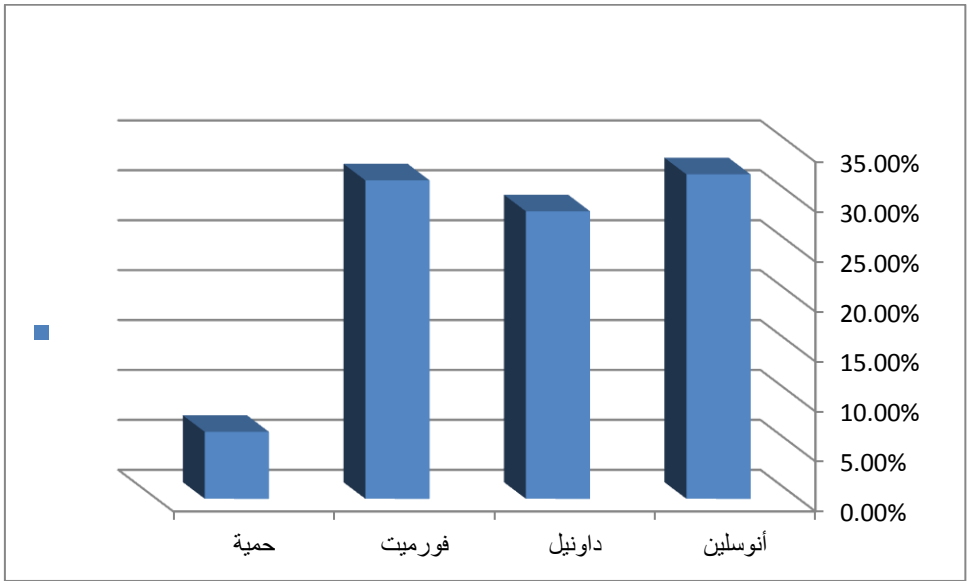
الأعراض	العدد	النسبة

أنوسلين	53	%32.5
داونيل	47	%28.8
فورميت	52	%31.9
حمية	11	%6.7
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-36)

التوزيع التكراري البسيط لمتغير **نوع العلاج**



نلاحظ من الجدول رقم 36 والشكل رقم36 ان اكثر العلاج استخداما هو الانسولين حيث بلغ 24.5%والفورميت بنسبة 24.1% والدوانيل بـ21.8% والحمية5.1%.

جدول رقم (4-37)

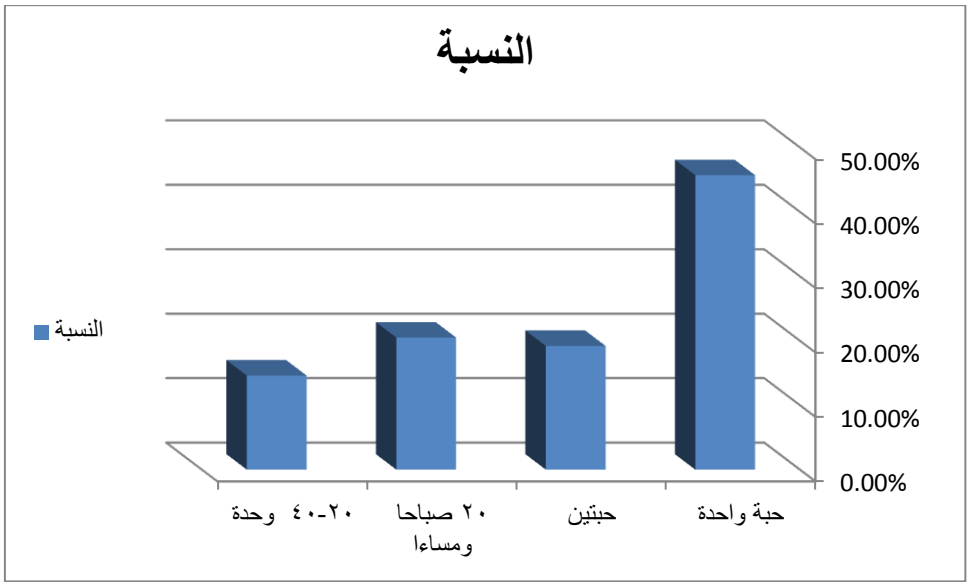
التوزيع التكراري البسيط لمتغير مقدار الجرعة

الأعراض	العدد	النسبة
حبة واحدة	69	%45.7
حبتين	29	%19.2
20 صباحا ومساءا	31	%20.5
20-40 وحدة	22	%14.6
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-37)

التوزيع التكراري البسيط لمتغير **مقدار الجرعة**



نلاحظ من الجدول رقم 37 الشكل رقم 37 يوضح مقدار الجرعة وكانت اكثرها تناولا هى حبة واحدتبنسبة 45.7% واقلها بنسبة 14.6%

جدول رقم (4-38)

التوزيع التكراري البسيط لمتغير العمر

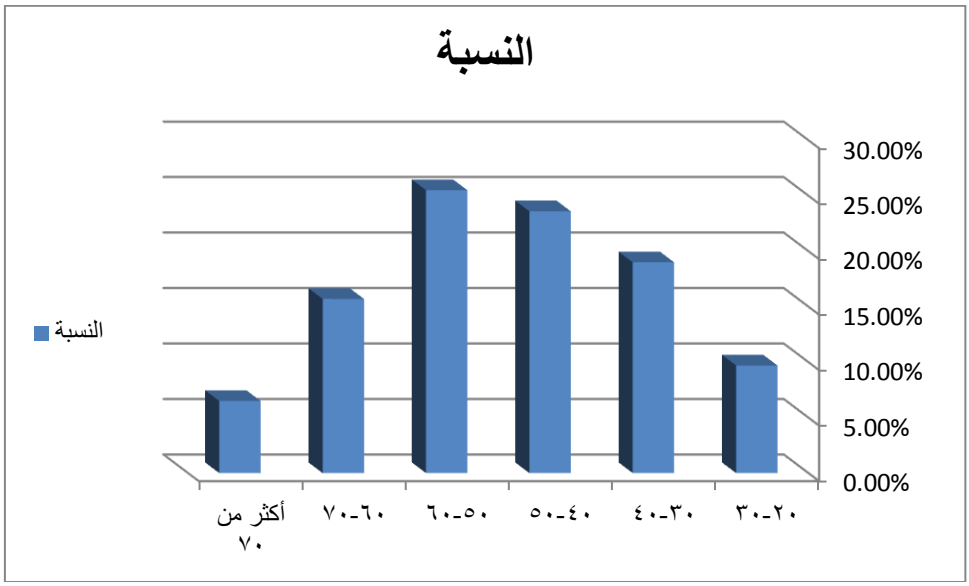
العمر	العدد	النسبة
20-30	21	%9.7

40-30	41	%19
50-40	51	%23.6
60-50	55	%25.5
70-60	34	%15.7
أكثر من 70	14	%6.5
المجموع	216	100%

المصدر : إعداد الباحث من بيانات الدراسة

شكل رقم (4-38)

التوزيع التكراري البسيط لمتغير **العمر**



نلاحظ من الجدول رقم 38 والشكل رقم 38 ان اكثر الفئات العمرية مصابة بالسكرى هى من عمر 50-60 بنسبة 25.5% وبعدها من40-50 بنسبة 23.6% ثم من 30-40 بنسبة 19% ثم من 60-70 بنسبة 15.7% ومن 20-30 بنسبة 9.7% واخيرا من 70 فاكثر بنسبة6.5% .

4-5 التطبيق الاول التحليل الإحصائي بإستخدام التحليل التمييزي

4-5-1 التحليل الإحصائي بإستخدام التحليل التمييزي لأهم عوامل الإصابة بمرض السكري:

1 وصف البيانات[7]:

في هذا الجزء نقوم بتطبيق أسلوب التحليل التمييزي على بيانات عينة البحث وحجمها 216 منهم (162) مصابين بمرض السكري و (54) غير مصابين بهدف الوصول إلى الدالة التمييزية ومن ثم استخدامها في التصنيف بناء على تحقيق الفرضيات .

2 متغيرات البحث :-

وتشمل متغيرات البحث على متغير تابع (y) وهو متغير ثنائي يأخذ قيمتين (0,1) (1) عندما يكون الشخص مصاب ، (0) غير مصاب ، ويشمل كذلك على المتغيرات المستقلة الآتية :

(x_1 النوع ، x_2 العمر ، x_3 التاريخ العائلي ، x_4 التدخين ، x_5 الكحول ، x_6 الرياضة، x_7 الضغط، x_8 تنسيب كتلة الجسم، x_9 أمراض القلب، x_{10} حالة التغذية)

وتم استخدام برنامج " **SPSS**" لتحليل البيانات الإحصائية ولإيجاد الدالة التمييزي

3 إيجاد الدالة التمييزية

أولاً : حساب متوسط المتغيرات في كل مجموعة

المجموعة الأولى **المجموعة الثانية :**

$$\overline{x}_i\,(1) = \frac{\sum_{i=1}^{n_i} x_i}{n_1} \qquad \overline{x}_i\,(2) = \frac{\sum_{i=1}^{n_i} x_i}{n_2}$$

جدول رقم (4-39)

متوسطات المتغيرات في كل مجموعة

المجموعة المصابة(**الأولى**)	المجموعة غير المصابة (**الثانية**)

(7) المصدر: إعداد الباحث من بيانات الدراسة2014

المتغير	المتوسط	المتغير	المتوسط
$\bar{x}_1(1)$	1.57	$\bar{x}_1(2)$	***1.43***
$\bar{x}_2(1)$	53.02	$\bar{x}_2(2)$	***39.7***
$\bar{x}_3(1)$	1.28	$\bar{x}_3(2)$	***1.52***
$\bar{x}_4(1)$	1.84	$\bar{x}_4(2)$	***1.57***
$\bar{x}_5(1)$	1.86	$\bar{x}_5(2)$	***1.57***
$\bar{x}_6(1)$	1.46	$\bar{x}_6(2)$	***1.30***
$\bar{x}_7(1)$	1.33	$\bar{x}_7(2)$	***1.17***
$\bar{x}_8(1)$	4.06	$\bar{x}_{82}(2)$	***3.72***
$\bar{x}_9(1)$	1.77	$\bar{x}_9(2)$	***1.91***
$\bar{x}_{10}(1)$	1.36	$\bar{x}_{10}(2)$	***1.74***

المصدر : إعداد الباحث من بيانات الدراسة

ثانياً : إيجاد الفرق بين متوسط كل متغير في المجموعتين[8]

$$d = \bar{x}_i\ (1) - \bar{x}_i\ (2) = \begin{vmatrix} \bar{x}_{11} - \bar{x}_{12} \\ \bar{x}_{21} - \bar{x}_{22} \\ \vdots \\ \vdots \\ \bar{x}_{n1} - \bar{x}_{n2} \end{vmatrix} = \begin{vmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_n \end{vmatrix}$$

ثالثاً : إيجاد مصفوفة التباينات والتغايرات المشتركة بين المجموعتين ويتم إيجادها باستخدام المعادلات (5) ، (6) كالآتي :

1. مصفوفة التباينات والتغايرات المشتركة لمجموعة المصابين

(8) المصدر: إعداد الباحث من بيانات الدراسة

	.243	−.812	−.004	.083	.072	−.015	−.027	.010	−.003	.243
	−.812	171.931	.864	.835	.661	1.109	.006	.620	.570	−.976
.	−.004	.864	.202	.001	.002	.001	.002	−.016	.016	.004
V =	.083	.835	.001	.136	.120	.037	−.015	−.003	.012	−.016
	.072	.661	.002	.120	.123	.035	−.015	.014	.004	−.004
	−.015	1.109	.001	.037	.035	.250	−.016	−.001	−.003	.017
	−.027	.006	.002	−.015	−.015	−.016	.221	−.056	.003	−.002
	.010	.620	−.016	−.003	.014	−.001	−.056	.922	−.006	−.045
	−.003	.570	.016	.012	.004	−.003	.003	−.006	.181	−.001
	−.018	−.976	.004	−.016	−.004	.017	−.002	−.045	−.001	.245

2. مصفوفة التباينات والتغايرات المشتركة لمجموعة غير المصابين :

.249	−.607	.001	.166	.166	−.015	.003	.026	−.016	−.057
−.607	143.194	−.542	−.940	−.940	.920	1.673	3.709	−.066	−.984
.001	−.542	.254	−.001	−.001	.051	.025	−.042	.011	−.014
.166	−.940	−.001	.249	.249	−.022	−.003	.012	−.021	−.075
.166	−.940	−.001	.249	.249	−.022	−.003	.012	−.021	−.075
−.015	.920	.051	−.022	−.022	.212	−.031	.140	−.010	−.035
.003	1.673	.025	−.003	−.003	−.031	.142	.028	.016	−.031
.026	3.709	−.042	.012	.012	.140	.028	.808	−.064	−.092
−.016	−.066	.011	−.021	−.021	−.010	.016	−.064	.086	−.024
−.057	−.984	−.014	−.075	−.075	−.035	−.031	−.092	−.024	.196

3. مصفوفة التباينات والتغايرات المدمجة للمجموعتين:

	.249	−.340	−.010	.112	.104	−.010	−.015	.025	−.011	−.040
	−.340	197.44	−.090	1.060	.975	1.475	.820	2.215	.054	−1.918
	−.010	−.090	.225	−.011	−.011	.006	.000	−.037	.021	.016
	.112	1.060	−.011	.176	.165	.031	−.004	.017	−.004	−.049
V=	.104	.975	−.011	.165	.168	.030	−.004	.031	−.010	−.041
	−.010	1.475	.006	.031	.030	.245	−.015	.044	−.009	−.008
	−.015	.820	.000	−.004	−.004	−.015	.206	−.025	.002	−.021

.025	2.215	−.037	.017	.031	.044	−.025	.911	−.029	−.080
−.011	.054	.021	−.004	−.010	−.009	.002	−.029	.160	.003
−.040	−1.918	.016	−.049	−.041	−.008	−.021	−.080	.003	.259

رابعاً: إيجاد معادلة الدالة التمييزية

$$\hat{L} = \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \ldots\ldots\ldots\ldots\ldots + \hat{\alpha}_n x_n$$

$$\hat{\alpha} = v^{-1} d = \begin{vmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ . \\ . \\ \hat{\alpha}_n \end{vmatrix}$$

وبإجراء عمليات ضرب متجه الفروقات في معكوس مصفوفة التباينات والتغايرات المشتركة نحصل علي قيم $\hat{\alpha}$ وباستخدام برنامج spss تم الحصول علي الآتي :

$$\hat{\alpha} = \begin{vmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\alpha}_5 \\ \hat{\alpha}_6 \\ \hat{\alpha}_7 \\ \hat{\alpha}_8 \\ \hat{\alpha}_9 \\ \hat{\alpha}_{10} \end{vmatrix} = \begin{vmatrix} 0.231 \\ 0.051 \\ -0.873 \\ -1.852 \\ 2.768 \\ 0.246 \\ 0.554 \\ 0.059 \\ -0.602 \\ -0.835 \end{vmatrix}$$

∴ معادلة الدالة التمييزية هي :

$$\hat{L}_i = 0.231x_1 + 0.051x_2 - 0.873x_3 - 1.852x_4 + 2.768x_5 + 0.246x_6 + 0.554x_7 + 0.059x_8 - 0.602x_9 - 0.835\ x_{10}$$

4 الأهمية النسبية للمتغيرات[9] :

وبعد أن توصلنا لدالة التمييز يمكن بيان أهمية كل متغير مقارنة بالمتغيرات الأخرى وباستخدام المعادلة **(9)** لترتيب المتغيرات حسب الأهمية كالآتي :

(9) المصدر: إعداد الباحث من بيانات الدراسة

$$\boldsymbol{\alpha_i^* = \hat{a}_i \sqrt{V_{ii}}}$$

$\boldsymbol{\alpha_1^*} = \hat{a}_1 \sqrt{V_{11}} =$ $= \boldsymbol{0}.115$

$\boldsymbol{\alpha_2^*} = \hat{a}_2 \sqrt{V_{22}} =$ $= 0.716$

$\boldsymbol{\alpha_3^*} = \hat{a}_3 \sqrt{V_{33}} =$ $= -0.414$

$\boldsymbol{\alpha_4^*} = \hat{a}_4 \sqrt{V_{44}} =$ $= -0.777$

$\boldsymbol{\alpha_5^*} = \hat{a}_5 \sqrt{V_{55}} =$ $= 1.134$

$\boldsymbol{\alpha_6^*} = \hat{a}_6 \sqrt{V_{66}} =$ $= 0.122$

$\boldsymbol{\alpha_7^*} = \hat{a}_7 \sqrt{V_{77}} =$ $= 0.252$

$\boldsymbol{\alpha_8^*} = \hat{a}_8 \sqrt{V_{88}} =$ $= 0.056$

$\boldsymbol{\alpha_9^*} = \hat{a}_9 \sqrt{V_{99}} =$ $= -0.241$

$\boldsymbol{\alpha_{10}^*} = \hat{a}_9 \sqrt{V_{1010}} =$ $= -0.425$

كما بينا من قبل أن أكبر قيمة (بإهمال الإشارة السالبة مع الترتيب التنازلي) تعني أن المتغير المقابل لها هو أهم متغير له القدرة على التمييز بين المجموعتين ، وثاني أكبر قيمة تقابل ثاني أهم متغير له القدرة على التمييز بين المجموعتين وهكذا .

جدول رقم (4-40)

يوضح ترتيب المتغيرات حسب الأهمية (بإهمال الإشارة السالبة مع الترتيب التنازلي)

رمز المتغير	اسم المتغير	قيمة أهمية المتغير
x_5	الكحول	1.134
x_4	التدخين	−0.777
x_2	العمر	0.716
x_{10}	حالة التغذية	−0.425
x_3	التاريخ العائلي	−0.414
x_7	ضغط الدم	0.252
x_9	أمراض القلب	−0.241
x_6	الرياضة	0.122
x_1	النوع	0.115
x_8	تنسيب كتلة الجسم	0.056

المصدر : إعداد الباحث من بيانات الدراسة

وبالنظر إلي النتائج السابقة ل α_i^* نجد ان اكثر المتغيرات أهمية هو المتغير x_5 والذي يمثل الكحول والذي له اكبر تاثير في الأصابة بمرض السكري ثم يليه من حيث الأهمية x_4 والذي يمثل التدخين ثم يليه x_2 والذي يمثل العمر ثم يلية x_{10} والذي يمثل حالة التغذية ثم يلية x_3 والذي يمثل التاريخ العائلي ثم يليه x_7 والذي يمثل ضغط الدم ثم يليه x_9 والذي يمثل أمراض القلب ثم يليه x_6 والذي يمثل الرياضة وأخيرا x_8 والذي يمثل تنسيب كتلة الجسم .

جدول رقم (4-41)

يوضح إختبار المتغيرات المختلفة

Tests of Equality of Group Means

	Wilks' Lambda	F	df 1	df 2	Sig.
النوع	.979	4.602	1	214	.033
العمر	.831	43.564	1	214	.000
التاريخ العائلي	.951	10.924	1	214	.001
التدخين	.925	17.431	1	214	.000
الكحول	.910	21.216	1	214	.000
الرياضة	.979	4.671	1	214	.032
ضغط الدم	.976	5.172	1	214	.024
BMI	.977	5.033	1	214	.026
امراض القلب	.976	5.196	1	214	.024
حاله التغذية	.897	24.635	1	214	.000

المصدر : إعداد الباحث من بيانات الدراسة

هذا الجدول يبين ما إذا كانت هناك فروق معنوية إحصائياً بين المتغيرات المستقلة كل علي حده في كل من المجموعتين لمتغير التصنيف التابع (y) ، ويتضح من الجدول أن الفروق بين متوسطات المجموعتين لجميع المتغيرات المستقلة معنوية إحصائياً.

نلاحظ أن قيم" **sig**" للمتغيرات الآتية (النوع ،العمر ،التاريخ العائلي للإصابة ، التدخين ، الكحول ،الرياضة ، ضغط الدم، تنسيب كتلة الجسم ، أمراض القلب ، حالة التغذية) أقل من **0.05** أي أن المتغيرات جميعها معنوية وهذا يعني أن لها تأثير كبير في الإصابة بمرض السكري[10] .

5 -اختبار قدرة الدالة على التمييز وذلك باستخدام اختبار F عن طريق تكوين جدول تحليل التباين [11]:

(10) المصدر: إعداد الباحث من بيانات الدراسة
(11) المصدر: إعداد الباحث من بيانات الدراسة

وذلك باختبار الفرضية الآتية :

H_0 : الدالة ليس لها القدرة التمييز

H_1 : الدالة لها القدرة على التمييز

في هذا الاختبار نقوم بحساب مجموعة مربعات داخل المتغيرات (الخطأ)

(*SSE*) من خلال كونه مربع المسافة بين المجموعتين

$$SSE=D^2$$

حيث :

$$D = \sqrt{\hat{a}_1 d_1 + \hat{a}_2 d_2 \dots\dots + \hat{a}_n d_n}$$

$$SSE= D^2 = \hat{a}_1 d_1 + \hat{a}_2 d_2 \dots\dots + \hat{a}_n d_n$$

$$SSE = (0.231*0.16) + (0.051*13.32) - (.873*0.27) +(1.852*0.29) +$$

$$(2.768*0.16) + (0.246*0.16) + (0.554*0.16) + (0.059*0.34) +(0.602*0.14) +$$

$$(0.835*0.24)$$

$$SSE= D^2 = 1.78$$

ثم نحسب مجموع المربعات بين التغيرات " المعالجات " *SSB* "

$$SSB = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} * (D^2)$$

$$SSB = \frac{162 * 54}{(162 + 54)(162 + 54 - 2)} * (1.78^2)$$

$$SSB = 0.6$$

ثم نحسب مجموع المربعات الكلية "**SST**"

$$SST = SSB + SSE$$

$$SST = 0.6 + 1.78 = 2.38$$

ويتكون جدول تحليل التباين كما يلي:

جدول رقم (4-42)

جدول تحليل التباين

S.O.V	d.F	S.S	M.S	F

SSB	*9*	*0.6*	*0.067*	*3.350*
SSE	*90*	*1.78*	*0.02*	
SST	*99*	*2.38*		

المصدر : إعداد الباحث من بيانات الدراسة

وتحسب قيمة F الجدولية كالآتي :

$$F_{K-1,n-k,0.05} = F_{9,206,0.05} = 1.88$$

بما أن F المحسوبة أكبر من الجدولية نرفض H_o إذن الدالة لها القدرة على التمييز .

جدول رقم (4-43)

يوضح أيضا اختبار معنوية الدالة التمييزية

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	0.668	84.205	10	.000

المصدر : إعداد الباحث من بيانات الدراسة

ومن الجدول أعلاه نري أن قيمة (0 .000) = sig وهي اقل من (0.05) وتعني معنوية الدالة التمييزية وإنها تمتلك القدرة على التمييز بين المجموعتين.

6 -نقطة الفصل CUTOFF POINT [12]:

بعد أن توصلنا إلى أن الدالة التمييزية تملك القدرة على تمييز الأشخاص المصابين وغير المصابين بمرض السكري عليه نقوم بحساب نقطة الفصل بالقانون التالي:

$$\bar{\bar{L}} = \frac{\bar{L}(1) + \bar{L}(2)}{2}$$

ثانياً :

إيجاد متوسط *L* في كل مجموعة

$$\bar{L}(1) = \frac{358.76}{162} = 2.215$$

$$\bar{L}(2) = \frac{155.79}{54} = 2.885$$

(12) المصدر: إعداد الباحث من بيانات الدراسة

ثالثاً: إيجاد النقطة الفاصلة

$$\bar{\bar{L}} = \frac{\bar{L}(1)+\bar{L}(2)}{2} = \frac{2.215+2.8852}{2} = 2.55$$

بعد ذلك تتم المقارنة

وهنالك حالتين [13]:

- **الحالة الأولي** : إذا كانت $\bar{L}(1) > \bar{L}(2)$ فاذا تبين ان المفردة الجديدة أكبر من نقطة الفاصلة فانها تنتمى الى المجموعة الأولي . واذا تبين ان النقطة الفاصلة اقل من المفردة الجديدة فانها تنتمي الى المجموعة الثانية .
- **الحالة الثانية** : إذا كانت $\bar{L}(1) < \bar{L}(2)$ فاذا تبين ان المفردة الجديدة اكبر من النقطة الفاصلة فإنها تنتمي إلى المجموعة الثانية ، وإذا كانت المفردة الجديدة اقل من النقطة الفاصلة فإنها تنتمي إلى المجموعة الأولي .

وبما أن

$$\bar{L}(1) < \bar{L}(2)$$

$$2.215 < 2.885$$

وهنا تنطبق الحالة ال التي تقول إنه إذا كانت المفردة الجديدة المراد تصنيفها اقل من نقطة الفاصلة فانها تنتمى الى المجموعة الأولي . واذا تبين ان المفردة الجديدة اكبر من النقطة الفاصلة فانها تنتمي الى المجموعة الثانية .

ملاحظة :

أو يمكن مقارنة المفردة الجديدة المراد تصنيفها مع اي من متوسط المجموعتين وتنتمي إلى اقرب واحد من المجموعتين :

متوسط المجموعة الأولى = *2.215*

متوسط المجموعة الثانية = *2.885*

فمثلاً نأخذ من عينة البحث بيانات المريض رقم (*20*) والتي تنتمي إلى مجموعة المصابين وكانت البيانات كالأتي :-

$x_1=1.0$, $x_2=44$, $x_3=1$, $x_4=1$, $x_5=1$, $x_6=1$, $x_7=1$, $x_8=4$, $x_9=2$, $x_{10}=2$

وبتعويض هذه القيم في دالة التمييز

(13) المصدر: إعداد الباحث من بيانات الدراسة

$$\widehat{L_i} = 0.231x_1 + 0.051x_2 - 0.873x_3 - 1.852x_4 + 2.768x_5 + 0.246x_6 + 0.554x_7 + 0.059x_8 - 0.602x_9 - 0.835\ x_{10}$$

نجدها تساوي *0.68*

وهي اقل من النقطة الفاصلة فانها تنتمى الى المجموعة الأولي إذن الشخص مصاب بمرض السكري .

7 نسبة الخطأ[14]:

أولا : الخطأ الظاهري

الخطأ الظاهري للمجموعة الأولي

$$P_{12} = \frac{n_{12}}{n_1}$$

الخطأ الظاهري للمجموعة الثانية

$$P_{21} = \frac{n_{21}}{n_2}$$

حيث:

n_{12} : عدد المفردات التى تنتمى بالاصل للمجموعة الأولي وتم تصنيفها للمجموعة الثانية (تصنيف خاطىء).

n_{21} : عدد المفردات التى تنتمى بالاصل للمجموعة الثانية وتم تصنيفها للمجموعة الاولي (تصنيف خاطئ).

n_1 : هي حجم العينة الأولي.

n_1 : هي حجم العينة الثانية .

$$P_{12} = \frac{n_{12}}{n_1} = \frac{5}{162} = 0.0309$$

$$P_{21} = \frac{n_{21}}{n_2} = \frac{1}{54} = 0.0185$$

(14) المصدر: إعداد الباحث من بيانات الدراسة

وبذلك يتضح أن في عملية التصنيف نسبة الخطأ صغيرة جداً اعتمادا على الخطأ الظاهري.

الخطأ الحقيقي[15]

يعتبر أهم من الخطأ الظاهري ويتم إيجاده باستخدام جدول احتمالات للتوزيع الطبيعي القياسي بالصيغة .

$$P_{12} = P_{12} = \emptyset\left[-\frac{\sqrt{D^2}}{2}\right]$$

$$P_{12} = P_{12} = \emptyset\left[-\frac{\sqrt{\mathbf{1.78}}}{2}\right]$$

$$= \emptyset[-1.334]$$

من جدول (z) السالب

$$P_{12} = P_{12} = \mathbf{0.0179}$$

وهذه القيمة صغيرة جداً ويعني أن نسبة الخطأ الناتج من عملية التمييز صغيرة .

4-5-2 التحليل الإحصائي بإستخدام التحليل التميزي لأهم مضاعفات مرض السكري :

1 وصف البيانات[16]:

(15) المصدر: إعداد الباحث من بيانات الدراسة

(16) المصدر: إعداد الباحث من بيانات الدراسة2014

في هذا الجزء نقوم بتطبيق أسلوب التحليل التمييزي على بيانات عينة البحث وحجمها 216 منهم (162) مصابين بمرض السكري و (54) غير مصابين بهدف الوصول إلى الدالة التمييزية ومن ثم استخدامها في التصنيف بناء على تحقيق الفرضيات .

2 متغيرات البحث :-

وتشمل متغيرات البحث على متغير تابع (y) وهو متغير ثنائي يأخذ قيمتين (0,1) (1) عندما يكون الشخص مصاب ، (0) غير مصاب ، ويشمل كذلك على المتغيرات المستقلة الآتية :

(x_1 امراض القلب ، x_2 امراض العيون ، x_3 هل انت مصاب بامراض اخري ، x_4 امراض الكلى ، x_5 الجهاز العصبى ، x_6 هل توجد امراض وراثية اخرى فى العائلة ، x_7 هل توجد جروح باليد، x_8 هل توجد جروح بالرجل ، x_9 بتر فى الاطراف ، x_{10} هل توجد مضاعفات)

وتم استخدام برنامج " **SPSS**" لتحليل البيانات الإحصائية ولإيجاد الدالة التمييزي

3 إيجاد الدالة التمييزية

أولاً : حساب متوسط المتغيرات في كل مجموعة

المجموعة الأولى **المجموعة الثانية :**

$$\bar{x}_i(1) = \frac{\sum_{i=1}^{n_i} x_i}{n_1} \qquad \bar{x}_i(2) = \frac{\sum_{i=1}^{n_i} x_i}{n_2}$$

جدول رقم (4-44)

متوسطات المتغيرات في كل مجموعة

المجموعة المصابة (الأولى)		المجموعة غير المصابة (الثانية)	
المتغير	المتوسط	المتغير	المتوسط

1.9074	$\bar{x}_1(2)$	1.7702	$\bar{x}_1(1)$
1.7963	$\bar{x}_2(2)$	1.4534	$\bar{x}_2(1)$
1.5000	$\bar{x}_3(2)$	1.3354	$\bar{x}_3(1)$
1.7037	$\bar{x}_4(2)$	1.8323	$\bar{x}_4(1)$
1.6667	$\bar{x}_5(2)$	1.8137	$\bar{x}_5(1)$
1.6667	$\bar{x}_6(2)$	1.4969	$\bar{x}_6(1)$
2.0000	$\bar{x}_7(2)$	1.9379	$\bar{x}_7(1)$
2.0000	$\bar{x}_8(2)$	1.8882	$\bar{x}_8(1)$
1.7222	$\bar{x}_9(2)$	1.5466	$\bar{x}_9(1)$
1.3704	$\bar{x}_{10}(2)$	1.205	$\bar{x}_{10}(1)$

المصدر : إعداد الباحث من بيانات الدراسة

ثانياً : إيجاد الفرق بين متوسط كل متغير في المجموعتين[17]

$$d = \bar{x}_i\,(1) - \bar{x}_i\,(2) = \begin{vmatrix} \bar{x}_{11} - \bar{x}_{12} \\ \bar{x}_{21} - \bar{x}_{22} \\ \vdots \\ \vdots \\ \bar{x}_{n1} - \bar{x}_{n2} \end{vmatrix} = \begin{vmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_n \end{vmatrix}$$

$$\begin{vmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \\ d_9 \\ d_{10} \end{vmatrix} = \begin{vmatrix} -0.137 \\ -0.343 \\ -0.165 \\ 0.129 \\ 0.147 \\ -0.170 \\ -0.062 \\ -0.112 \\ -0.004 \\ -0.176 \end{vmatrix} =$$

(17) المصدر: إعداد الباحث من بيانات الدراسة

ثالثاً : إيجاد مصفوفة التباينات والتغايرات المشتركة بين المجموعتين ويتم إيجادها باستخدام المعادلات (5) ، (6) كالآتي :

3. مصفوفة التباينات والتغايرات المشتركة لمجموعة المصابين

-.003-	.020	.012	-.002-	.021	.007	.005	.009	-.001-	.178
-.976	.013	-.006-	.007	-.009-	.055	.035	.014	.166	.249
.004	.025	-.022-	.006	-.010-	.070	.050	-.006-	.224	.166
-.016	.010	.011	.006	-.004-	.003	.006	.140	-.006-	.014
-.004	.032	.009	.004	-.005-	.049	.153	.006	.050	.035
.017	.023	.027	.012	-.019-	.252	.049	.003	.070	.055
-.002	.013	.009	.000	.059	-.019-	-.005-	-.004-	-.010-	-.009-
-.045	-.003-	.020	.012	-.002-	.021	.007	.005	.009	-.001-
-.001	.013	-.006-	.007	-.009-	.055	.035	.014	.166	.249
.245	.025	-.022-	.006	-.010-	.070	.050	-.006-	.224	.166

4. مصفوفة التباينات والتغايرات المشتركة لمجموعة غير المصابين :

.016	-.007-	.000	.000	.006	.006	-.028-	.028	.000	.086
-.984	.058	-.001-	.000	.000	.063	.006	-.005-	.104	.165
-.014	.057	-.009-	.000	.000	.075	.094	.000	.255	.104
-.075	.017	.010	.000	.000	-.006-	-.006-	.212	.000	-.005-
-.075	.031	-.038-	.000	.000	.000	.226	-.006-	.094	.006

.063	.075	-.006-	.000	.226	.000	.000	.019	.013	-.035
.000	.000	.000	.000	.000	.000	.000	.000	.000	-.031
.000	.000	.000	.000	.000	.000	.000	.000	.000	-.092
-.001-	-.009-	.010	-.038-	.019	.000	.000	.204	-.027-	-.024
.058	.057	.017	.031	.013	.000	.000	-.027-	.238	.196

3. مصفوفة التباينات والتغايرات المدمجة للمجموعتين:

V=

.158	.008	.018	-.007-	.003	.022	.000	.012	.018	.006
.008	.250	.160	.001	.018	.067	-.003-	.013	.007	.035
.018	.160	.236	-.008-	.056	.076	-.006-	.008	-.013-	.038
-.007-	.001	-.008-	.161	.007	-.004-	-.005-	.002	.007	.007
.003	.018	.056	.007	.174	.032	-.006-	-8.694E-5	-.008-	.027
.022	.067	.076	-.004-	.032	.250	-.012-	.013	.030	.025
.000	-.003-	-.006-	-.005-	-.006-	-.012-	.045	.001	.009	.012
.012	.013	.008	.002	-8.694E-5	.013	.001	.077	.022	.021
.018	.007	-.013-	.007	-.008-	.030	.009	.022	.243	-.006-
.006	.035	.038	.007	.027	.025	.012	.021	-.006-	.187

رابعاً: إيجاد معادلة الدالة التمييزية

$$\hat{L} = \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \ldots\ldots\ldots\ldots\ldots + \hat{\alpha}_n x_n$$

$$\alpha = V^{-1} d$$

$$v^{-1}d = \begin{vmatrix} \hat{a}_1 \\ \hat{a}_2 \\ . \\ . \\ \hat{a}_n \end{vmatrix}$$

وبـإجراء عمليـات ضـرب متجـه الفروقـات فـي معكـوس مصـفوفة التباينـات والتغـايرات المشـتركة نحصل علي قيم $\hat{a}$ وباستخدام برنامج spss تم الحصول علي الآتي :

$$\hat{a} = \begin{vmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{a}_9 \\ \hat{a}_{10} \end{vmatrix} = \begin{vmatrix} -0.62554 \\ -1.36439 \\ 0.25169 \\ 0.76928 \\ 0.98887 \\ -0.33275 \\ -1.01844 \\ -0.78119 \\ -0.49811 \\ -0.65019 \end{vmatrix}$$

∴ معادلة الدالة التمييزية هي :

$$\hat{L}_i = -0.62554x_1 - 1.36439\ x_2 + 0.25169x_3 + 0.76928x_4 + 0.98887x_5 - 0.33275x_6 - 1.01844x_7 - 0.78119x_8 - 0.49811x_9 - 0.65019x_{10}$$

4 الأهمية النسبية للمتغيرات[18] **:**

وبعد أن توصلنا لدالة التمييز يمكن بيان أهمية كل متغير مقارنة بالمتغيرات الأخرى وباستخدام المعادلة **(9)** لترتيب المتغيرات حسب الأهمية كالآتي :

$$\alpha_i^* = \hat{a}_i \sqrt{V_{ii}}$$

$$\alpha_1^* = \hat{a}_1 \sqrt{V_{11}} = \quad = -0.099$$

$$\alpha_2^* = \hat{a}_2 \sqrt{V_{22}} = \quad = -0.341$$

$$\alpha_3^* = \hat{a}_3 \sqrt{V_{33}} = \quad 0.059$$

(18) المصدر: إعداد الباحث من بيانات الدراسة

$$\alpha_4^* = \hat{a}_4 \sqrt{V_{44}} = \quad = 0.124$$

$$\alpha_5^* = \hat{a}_5 \sqrt{V_{55}} = \quad = 0.172$$

$$\alpha_6^* = \hat{a}_6 \sqrt{V_{66}} = \quad -0.083$$

$$\alpha_7^* = \hat{a}_7 \sqrt{V_{77}} = \quad = -0.046$$

$$\alpha_8^* = \hat{a}_8 \sqrt{V_{88}} = \quad = -0.060$$

$$\alpha_9^* = \hat{a}_9 \sqrt{V_{99}} = \quad = -0.121$$

$$\alpha_{10}^* = \hat{a}_9 \sqrt{V_{1010}} = \quad = -0.122$$

كما بينا من قبل أن أكبر قيمة (بإهمال الإشارة السالبة مع الترتيب التنازلي) تعني أن المتغير المقابل لها هو أهم متغير له القدرة على التمييز بين المجموعتين ، وثاني أكبر قيمة تقابل ثاني أهم متغير له القدرة على التمييز بين المجموعتين وهكذا .

جدول رقم (4-45)

يوضح ترتيب المتغيرات حسب الأهمية (بإهمال الإشارة السالبة مع الترتيب التنازلي)

رمز المتغير	اسم المتغير	قيمة أهمية المتغير
	امراض العيون	-0.341
	الجهاز العصبى	0.172
	امراض الكلى	0.124
	هل توجد مضاعفات	-0.122
	بتر فى الاطراف	-0.121

	امراض القلب	−0.099
	هل توجد امراض وراثية اخرى فى العائلة	−0.083
	هل توجد جروح بالرجل	−0.060
	هل انت مصاب بامراض اخري	0.059
	هل توجد جروح باليد	−0.046

المصدر : إعداد الباحث من بيانات الدراسة

وبالنظر إلي النتائج السابقة ل α_i^* نجد ان اكثر المتغيرات أهمية هو المتغير x_2 والذي يمثل امراض العيون والذي له اكبر تاثير في الأصابة بمرض السكري ثم يليه من حيث الأهمية x_5 والذي يمثل الجهاز العصبى ثم يليه x_4 والذي يمثل امراض الكلى ثم يلية x_{10} والذي يمثل هل توجد مضاعفات ثم يليه x_9 والذي يمثل بتر فى الاطراف ثم يلية x_1 والذي يمثل امراض القلب ثم يليه x_6 والذي يمثل هل توجد امراض وراثية اخرى فى العائلة ثم يليه x_8 والذي يمثل هل توجد جروح بالرجل ثم يليه x_3 والذي يمثل هل انت مصاب بامراض اخري وأخيرا x_7 والذي يمثل هل توجد جروح باليد

جدول رقم (4−46)

يوضح إختبار المتغيرات المختلفة

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
امراض القلب	.977	4.910	1	213	.028

امراض العيون	.911	20.810	1	213	.000
هل انت مصاب بامراض اخري	.978	4.725	1	213	.031
امراض الكلى	.981	4.223	1	213	.041
الجهاز العصبى	.977	5.112	1	213	.025
هل توجد امراض وراثية اخرى فى العائلة	.978	4.751	1	213	.030
هل توجد جروح باليد	.984	3.543	1	213	.050
هل توجد جروح بالرجل	.969	6.734	1	213	.010
بتر فى الاطراف	.976	5.237	1	213	.023
هل توجد مضاعفات	.972	6.069	1	213	.015

المصدر : إعداد الباحث من بيانات الدراسة

هذا الجدول يبين ما إذا كانت هناك فروق معنوية إحصائياً بين المتغيرات المستقلة كل علي حده في كل من المجموعتين لمتغير التصنيف التابع (y) ، ويتضح من الجدول أن الفروق بين متوسطات المجموعتين لجميع المتغيرات المستقلة معنوية إحصائياً.

نلاحظ أن قيم" **sig**" للمتغيرات الآتية (x_1 امراض القلب ، x_2 امراض العيون ، x_3 هل انت مصاب بامراض اخري ، x_4 امراض الكلى ، x_5 الجهاز العصبى ، x_6 هل توجد امراض وراثية اخرى فى العائلة ، x_7 هل توجد جروح باليد، x_8 هل توجد جروح بالرجل ، x_9 بتر فى الاطراف ، x_{10} هل توجد مضاعفات) أقل من **0.05** أي أن المتغيرات جميعها معنوية وهذا يعني أن لها تأثير كبير في الإصابة بمرض السكري[19] .

5 -اختبار قدرة الدالة على التمييز وذلك باستخدام اختبار F عن طريق تكوين جدول تحليل التباين [20]:

وذلك باختبار الفرضية الآتية :

H_0 : الدالة ليس لها القدرة التمييز

H_1 : الدالة لها القدرة على التمييز

في هذا الاختبار نقوم بحساب مجموعة مربعات داخل المتغيرات (الخطأ)

(19) المصدر: إعداد الباحث من بيانات الدراسة

(20) المصدر: إعداد الباحث من بيانات الدراسة

من خلال كونه مربع المسافة بين المجموعتين (*SSE*)

$$SSE=D^2$$

حيث :

$$\widehat{L_\iota} = -0.62554x_1-1.36439\ x_2+0.25169x_3+0.76928x_4 +0.98887x_5-0.33275x_6-1.01844x_7-0.78119\ x_8-0.49811x_9- 0.65019x_{10}$$

$$D = \sqrt{\hat{a}_1\, d_1 + \hat{a}_2 d_2 \ldots\ldots + \hat{a}_n\, d_n}$$

$$SSE= D^2 = \hat{a}_1\, d_1 + \hat{a}_2 d_2 \ldots\ldots + \hat{a}_n\, d_n$$

$$SSE = (-0.62554 * 0.16) + (-1.36439 * 13.32) - (0.25169 * 0.27) -$$

$$(0.76928 * 0.29) + (2.768 * 0.16) + (0.246 * 0.16) + (0.554 * 0.16) + (0.059 * 0.34)$$

$$+(0.602 * 0.14) + (835 * 0.24)$$

$$SSE= D^2 = 1.145$$

ثم نحسب مجموع المربعات بين التغيرات " المعالجات " *SSB* "

ثم نحسب مجموع المربعات الكلية "**SST**"

$$\mathbf{SST} = \mathbf{SSB} + \mathbf{SSE}$$

$$\mathbf{SST} = 1.145 + 0.248 = 1.393$$

ويتكون جدول تحليل التباين كما يلي:

جدول رقم (4-47)

جدول تحليل التباين

S.O.V	d.F	S.S	M.S	F
SSB	*9*	*1.145*	*0.127*	

SSE	206	*0.248*	*0.0012*	105.833
SST	215	*1.393*		

المصدر : إعداد الباحث من بيانات الدراسة

وتحسب قيمة F الجدولية كالآتي :

$$F_{K-1,n-k,0.05} = F_{9,206,0.05} = 1.88$$

بما أن F المحسوبة أكبر من الجدولية نرفض H_o إذن الدالة لها القدرة على التمييز .

جدول رقم (4-48)

يوضح أيضا اختبار معنوية الدالة التمييزية

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	0.780	51.681	10	.000

المصدر : إعداد الباحث من بيانات الدراسة

ومن الجدول أعلاه نري أن قيمة (0 .000) = sig وهي اقل من (0.05) وتعني معنوية الدالة التمييزية وإنها تمتلك القدرة على التمييز بين المجموعتين.

6 -نقطة الفصل CUTOFF POINT [21]:

بعد أن توصلنا إلى أن الدالة التمييزية تملك القدرة على تمييز الأشخاص المصابين وغير المصابين بمرض السكري عليه نقوم بحساب نقطة الفصل بالقانون التالي:

ثانياً :

إيجاد متوسط L في كل مجموعة

$$\bar{L}(1) = \frac{-818.59}{162} = -5.0530$$

$$\bar{L}(2) = \frac{-330.12}{54} = -6.2286$$

ثالثاً: إيجاد النقطة الفاصلة

$$\bar{\bar{L}} = \frac{\bar{L}(1)+\bar{L}(2)}{2} = \frac{-5.0530+-6.2286}{2} = -5.6408$$

(21) المصدر: إعداد الباحث من بيانات الدراسة

بعد ذلك تتم المقارنة

وهنالك حالتين [22]:

- **الحالـة الأولـي :** إذا كانت $\bar{L}(1) > \bar{L}(2)$ فاذا تبين ان المفردة الجديدة أكبر من نقطة الفاصلة فانها تنتمى الى المجموعة الأولي . واذا تبين ان المفردة الجديدة اقل من النقطة الفاصلة فانها تنتمي الى المجموعة الثانية .
- **الحالـة الثانيـة :** إذا كانت $\bar{L}(1) < \bar{L}(2)$ فاذا تبين ان المفردة الجديدة اكبر من النقطة الفاصلة فإنها تنتمي إلى المجموعة الثانية ، وإذا كانت المفردة الجديدة اقل من النقطة الفاصلة فإنها تنتمي إلى المجموعة الأولي .

وبما أن

وهنا تنطبق الحالة الأولى التي تقول إنه إذا كانت المفردة الجديدة المراد تصنيفها اقل من نقطة الفاصلة فانها تنتمى الى المجموعة الأولي . واذا تبين ان المفردة الجديدة اكبر من النقطة الفاصلة فانها تنتمي الى المجموعة الثانية .

ملاحظة :

أو يمكن مقارنة المفردة الجديدة المراد تصنيفها مع اي من متوسط المجموعتين وتنتمي إلى اقرب واحد من المجموعتين :

متوسط المجموعة الأولى = -5.0530

متوسط المجموعة الثانية = -6.2286

فمثلاً نأخذ من عينة البحث بيانات المريض رقم (68) والتي تنتمي إلى مجموعة المصابين وكانت البيانات كالأتي :-

$x_1=1.0$, $x_2=2$, $x_3=2$, $x_4=2$, $x_5=2$, $x_6=2$, $x_7=2$, $x_8=2$, $x_9=2$, $x_{10}=1$

وبتعويض هذه القيم في دالة التمييز

$$\widehat{L_i}=-0.62554x_1-1.36439x_2+0.25169x_3+0.76928x_4+0.98887x_5-0.33275x_6-1.01844x_7-0.78119\,x_8-0.49811x_9-0.65019x_{10}$$

(22) المصدر: إعداد الباحث من بيانات الدراسة

نجدها تساوي 5.25−

وهي أكبر من النقطة الفاصلة فانها تنتمى الى المجموعة الاولى إذن الشخص مصاب بمرض السكري .

7 نسبة الخطأ[23]:

أولا : الخطأ الظاهري

الخطأ الظاهري للمجموعة الأولي

$$P_{12} = \frac{n_{12}}{n_1}$$

الخطأ الظاهري للمجموعة الثانية

$$P_{21} = \frac{n_{21}}{n_2}$$

حيث:

n_{12} : عدد المفردات التى تنتمى بالاصل للمجموعة الأولي وتم تصنيفها للمجموعة الثانية (تصنيف خاطىء).

n_{21} : عدد المفردات التى تنتمى بالاصل للمجموعة الثانية وتم تصنيفها للمجموعة الاولي (تصنيف خاطئ).

n_1 : هي حجم العينة الأولي.

n_1 : هي حجم العينة الثانية .

$$\boldsymbol{P_{12} = \frac{n_{12}}{n_1}} = \frac{5}{162} = 0.0309$$

$$\boldsymbol{P_{21} = \frac{n_{21}}{n_2}} = \frac{1}{54} = 0.0185$$

وبذلك يتضح أن في عملية التصنيف نسبة الخطأ صغيرة جداً اعتمادا على الخطأ الظاهري.

الخطأ الحقيقي[24]

يعتبر أهم من الخطأ الظاهري ويتم إيجاده باستخدام جدول احتمالات للتوزيع الطبيعي القياسي بالصيغة .

(23) المصدر: إعداد الباحث من بيانات الدراسة
(24) المصدر: إعداد الباحث من بيانات الدراسة

$$P_{12} = P_{12} = \emptyset\left[-\frac{\sqrt{D^2}}{2}\right]$$

$$P_{12} = P_{12} = \emptyset\left[-\frac{\sqrt{\mathbf{17.689}}}{2}\right]$$

$$=\emptyset[-2.103]$$

من جدول (z) السالب

$$P_{12} = P_{12} = \mathbf{0.0179}$$

وهذه القيمة صغيرة جداً ويعني أن نسبة الخطأ الناتج من عملية التمييز صغيرة .

4-6 التحليل الإحصائي بإستخدام التحليل العاملي لأهم عوامل الإصابة بمرض السكري :

التأكد من عدم وجود مشكلة الإزدواج الخطي Multi-collinearity

ويتم الحكم على وجود أو عدم وجود مشكلة الإزدواج الخطي من خلال إيجاد محدد مصفوفة الإرتباط فإذا كانت قيمة المحدد أكبر من 0.00001 هنا نحكم بعد وجود مشكلة الإزدواج الخطي بين المتغيرات أما إذا كانت قيمته أقل من 0.00001 فهذا يعني وجود مشكلة الإزدواج الخطي بين المتغيرات ، وفي حالة وجود مشكلة الإزدواج الخطي نقوم بإستبعاد المتغيرات التي لها معامل إرتباط أكبر من 0.80 بغض النظر عن الإشارة .

كما نلاحظ أيضا أن محدد المصفوفة يساوي 0.019 وهو أكبر من 0.00001 وهذا يعني عدم وجود مشكلة الإزدواج الخطي بين المتغيرات .

تم تنفيذ عملية التحليل العاملي للمره الأولى first run على بيانات الدراسة الميدانية بإستخدام برنامج SPSS 16 وتم الحصول على مصفوفة الدوران الموضحة في الجدول الاتي. بينت المصفوفة أن عملية التحليل العاملي الإبتدائية قد قسمت المتغيرات المستقلة الى4 عوامل

جدول رقم (4-49)

KMO and Bartlett's Test

0.582	Kaiser –Meyer–Olkin (KMO)
547.348	Bartlett,s Test of Spherecity
45	Df
0.000	Sig

المصدر : إعداد الباحث من بيانات الدراسة

KMO Test

من خلال هذا الإختبار يتم الحكم على مدى كفاية حجم العينة . وبصفة عامة تتراوح قيمة إحصائية KMO بين الصفر والواحد الصحيح . وكلما إقتربت قيمته من الواحد الصحيح دل ذلك على زيادة الإعتمادية Reliability للعوامل التي نحصل عليها من التحليل ، والعكس صحيح ويشير هنا صاحب الإختبار (Kaiser 1974) أي أن الحد الأدنى المقبول لهذا الإحصائي هي 0.50 حتى يمكن الحكم بكفاية حجم العينة . أما في حالة أن تكون قيمته أقل من ذلك فإنه يتعين زيادة حجم العينة .

ومن النتائج أعلاه في الجدول السابق ، نجد ان قيمة ال KMO تساوي 0.582 أي أكبر من الحد الأدنى الذي إشترطه Kaiser إذا يمكننا ان نحكم بكفاية حجم العينة في التحليل الحالي .

Bartlett,s Test

الهدف من هذا الإختبار هو تحديد ما إذا كانت مصفوفة الإرتباط Correlation Matrix هي مصفوفة الوحدة Identity Matrix أم لا .

ومن النتائج نجد أن قيمة الـ Sig تساوي الصفر أي أقل من مستوى المعنوية 5% إذا نرفض فرض العدم ونقبل الفرض البديل أي أن مصفوفة الإرتباط ليست مصفوفة الوحدة .

جدول رقم (4–50)

معاملات الشيوع Communalities

العوامل (المتغيرات) Factors	قيم الشيوع Communalities
العمر	0.77
المهنة	0.70

التدخين	0.95
الكحول	0.95
التغذية	0.64
الرياضة	0.69
ضغط الدم	0.73
BMI	0.72
درجة القرابة	0.68
المستوى التعليمي	0.84

الطريقة المستخدمة : تحليل المكونات الرئيسية

المصدر : إعداد الباحث من بيانات الدراسة

القيم القاعدية ونسبة تباين العامل من التباين الكلي والتباين المتجمع للعوامل _المتغيرات الخاصة بظاهرة مرض السكري.

جدول رقم(51-4) يوضح المكونات الرئيسة

التباين الكلي المفسر Total Variance Explained

اسم المتغير	Eigen Values الجزور المميزة	Variance of % نسبة تباين	Cumulative % التباين المتجمع
العمر	4.111	41.102	41.102
المهنة	1.286	12.859	53.961
التدخين	1.226	12.262	66.223
الكحول	1.055	10.546	76.769
التغذية			

الرياضة			
ضغط الدم			
BMI			
درجة القرابة			
المستوى التعليمي			

المصدر : إعداد الباحث من بيانات الدراسة الطريقة المستخدمة : تحليل المكونات الرئيسية .

جدول رقم(4-52)

مصفوفة المكونات للعوامل المدورة Rotated Component Matrix[a]

اسم المتغير	1	2	3	*4*
العمر	0.812			
المهنة	0.785			
التدخين	-0.677			
الكحول	0.666			
التغذية	0.543	-0.521		
الرياضة		0.720		
ضغط الدم			-0.782	
BMI		-0.444	0.578	0.405
درجة القرابة	0.383			0.694

المستوى التعليمي			0.493	-0.573

الطريقة المستخدمة : تحليل المكونات الرئيسية

المصدر : إعداد الباحث من بيانات الدراسة

من خلال جدول نتائج التحليل العاملي نلاحظ ظهورخمسة عوامل رئيسية تؤثر في مرض السكري

وتمثل % 66.68 من اجمالي التباين وبعد د ا رسة متغيرات هذه العوامل وجد ان المتغيرين الحالة الاجتماعية ونوع الاصابة (ليس لهما اي دور سلبي او ايجابي في) تفسير نتائج عوامل الد ا رسة كما وبما ان التحليل العاملي يستخدم لتقليص البيانات و استخلاص العوامل و المتغيرات المهمة التي تكون مؤثرة و فعالة على بيانات الد ا رسة فقد قمنا باستبعاد هذين المتغيرين ثم اعادة العملية على عشرة متغيرات فقط فظهرلنا وجود اربعة عوامل رئيسية تؤثر في الظاهرة المدروسة و تشارك 76.77% من اجمالي التباين و

هذا ان دل على شيء فانما يدل على جودة المعالجة المتبعة من خلال استبعاد المتغيرات غير المعنوية و كالآتي:

العامل الاول

وجد أن هذا العامل يشكل 41.102% من اجمالي التباين و هذه النسبة تدل على مدى أهمية هذا

العامل في تفسير العلاقة بين المتغيرات المدروسة اما متغيرات هذا العامل فهي:

1-العمر وبمقدار تشبع مساوي 0.812

2-المهنة وبمقدار تشبع مساوي 0.785

3-التدخين وبمقدار تشبع مساوي 0.677

4-الكحول وبمقدار تشبع مساوي 0.666

5- التغذية وبمقدار تشبع مساوي 0.543

العامل الثاني

ياتي هذا العامل بالرتبة الثانية في الاهمية ويفسر 20.206% من اجمالي التباين ، و يحتوي هذا العامل على متغير و احد يساهم بشكل أساسي في تكوين العامل وهو الرياضة بتشبع مقداره 0.720 وهذا يدل على مدى أهمية الرياضة :

العامل الثالث

يفسر هذا العامل 12.262% من اجمالي التباين و متغيرات هذا العامل هي:

1-ضغط الدم بتشبع مقداره 0.782

2-تنسيب كتلة الجسم (**BMI**) بتشبع مقدار 0.578

العامل الرابع

يأتي هذا العامل في المرتبة الأخيرة من حيث أهميته في تفسير العلاقة بين المتغيرات حيث يمثل (10.546%) من أجمالي التباين و متغيرات هذا العامل هي:

1-درجة القرابة بتشبع مقداره 0.694

2-المستوى التعليمي بتشبع مقدار 0.573

4-7 التحليل الإحصائي بإستخدام التحليل العاملي لأهم مضاعفات مرض السكري

التأكد من عدم وجود مشكلة الإزدواج الخطي Multi-collinearity

ويتم الحكم على وجود أو عدم وجود مشكلة الإزدواج الخطي من إيجاد محدد مصفوفة الإرتباط فإذا كانت قيمة المحدد أكبر من 0.00001 هنا نحكم بعد وجود مشكلة الإزدواج الخطي بين المتغيرات أما إذا كانت قيمته أقل من 0.00001 فهذا يعني وجود مشكلة الإزدواج الخطي بين المتغيرات ، وفي حالة وجود مشكلة الإزدواج الخطي نقوم بإستبعاد المتغيرات التي لها معامل إرتباط أكبر من 0.80 بغض النظر عن الإشارة .

كما نلاحظ أيضا أن محدد المصفوفة يساوي 0.257 وهو أكبر من 0.00001 وهذا يعني عدم وجود مشكلة الإزدواج الخطي بين المتغيرات .

تم تنفيذ عملية التحليل العاملي للمره الأولى first run على بيانات الدراسة الميدانية بإستخدام برنامج SPSS 16 وتم الحصول على مصفوفة الدوران الموضحة في الجدول الاتي. بينت المصفوفة أن عملية التحليل العاملي الإبتدائية قد قسمت المتغيرات المستقلة الى6 عوامل

جدول رقم (4-53)

KMO and Bartlett's Test

0.543	Kaiser –Meyer–Olkin (KMO)
284.435	Bartlett,s Test of Sphericity
55	Df
0.000	Sig

المصدر : إعداد الباحث من بيانات الدراسة

KMO Test

من خلال هذا الإختبار يتم الحكم على مدى كفاية حجم العينة . وبصفة عامة تتراوح قيمة إحصائية KMO بين الصفر والواحد الصحيح . وكلما إقتربت قيمته من الواحد الصحيح دل ذلك على زيادة الإعتمادية Reliability للعوامل التي نحصل عليها من التحليل ، والعكس صحيح ويشير هنا صاحب الإختبار (Kaiser 1974) أي أن الحد الأدنى المقبول لهذا الإحصائي هي 0.50 حتى يمكن الحكم بكفاية حجم العينة . أما في حالة أن تكون قيمته أقل من ذلك فإنه يتعين زيادة حجم العينة .

ومن النتائج أعلاه في الجدول السابق ، نجد ان قيمة ال KMO تساوي 0.543 أي أكبر من الحد الأدنى الذي إشترطه Kaiser إذا يمكننا ان نحكم بكفاية حجم العينة في التحليل الحالي .

Bartlett,s Test

الهدف من هذا الإختبار هو تحديد ما إذا كانت مصفوفة الإرتباط Correlation Matrix هي مصفوفة الوحدة Identity Matrix أم لا .

ومن النتائج نجد أن قيمة الـ Sig تساوي الصفر أي أقل من مستوى المعنوية 5% إذا نرفض فرض العدم ونقبل الفرض البديل أي أن مصفوفة الإرتباط ليست مصفوفة الوحدة .

جدول رقم (4–54)

قيم التشبعات للعوامل

العوامل (المتغيرات) Factors	قيم الشيوع Communalities
هل انت مصاب بامراض اخري	0.77
امراض العيون	0.82
هل يوجد فتور بالجسم	0.58

هل توجد جروح بالرجل	0.81
ممارسة الحياة الزوجية بدون مشاكل	0.77
هل توجد جروح باليد	0.66
بتر فى الاطراف	0.48
هل توجد امراض وراثية اخرى فى العائلة	0.62
امراض الكلى	0.90
امراض القلب	0.72

الطريقة المستخدمة : تحليل المكونات الرئيسية

المصدر : إعداد الباحث من بيانات الدراسة

القيم القاعدية ونسبة تباين العامل من التباين الكلي والتباين المتجمع للعوامل_ المتغيرات الخاصة بظاهرة مرض السكري.

جدول رقم (4-55)

يوضح المكونات الرئيسة

Total Variance Explained

اسم المتغير	Eigen Values الجزور المميزة	Variance of % نسبة تباين	Cumulative % التباين المتجمع
هل انت مصاب بامراض اخري	1.925	17.497	41.102
امراض العيون	1.565	14.230	31.727
هل يوجد فتور بالجسم	1.099	9.990	41.717
هل توجد جروح بالرجل	1.088	9.894	51.611
ممارسة الحياة الزوجية بدون مشاكل	1.088	9.894	61.505
هل توجد جروح باليد	1.046	9.510	71.014

بتر فى الاطراف			79.329
هل توجد امراض وراثية اخرى فى العائلة			86.762
امراض الكلى			93.007
امراض القلب			97.420
بتر فى الاطراف			100.000

الطريقة المستخدمة : تحليل المكونات الرئيسية . **المصدر : إعداد الباحث من بيانات الدراسة**

جدول رقم(4-56)

يوضح المتغيرات المؤثرة في كل عامل

العوامل المدورة Rotated Component Matrix[a]

اسم المتغير	1	2	3	4	5	6
هل انت مصاب بامراض اخري	.888					
امراض العيون	.868					
هل يوجد فتور بالجسم		.762				
هل توجد جروح بالرجل		-.737				
ممارسة الحياة الزوجية بدون مشاكل		.546			-.442	
هل توجد جروح باليد			.940			
بتر فى الاطراف				.795		
هل توجد امراض وراثية اخرى فى العائلة	.463			.518		

امراض الكلى					.845	
امراض القلب						.874
بتر فى الاطراف						.417

المصدر : إعداد الباحث من بيانات الدراسة .الطريقة المستخدمة : تحليل المكونات الرئيسية .

من خلال جدول نتائج التحليل العاملي نلاحظ ظهورستة عوامل رئيسية تعتبر من مضاعفات مرض السكري

وتمثل 63.4% من اجمالي التباين وبعد د راسة متغيرات هذه العوامل وجد ان المتغيرات الثلاثة إرتفاع الدهون ، هل إرتفاع الدهون يسبب مضاعفات ، هل توجد مضاعفات (ليس لهم اي دور سلبي او ايجابي في) تفسير نتائج عوامل الد ا رسة كما وبما ان التحليل العاملي يستخدم لتقليص البيانات و استخلاص العوامل و المتغيرات المهمة التي تكون مؤثرة و فعالة على بيانات الد ا رسة فقد قمنا باستبعاد هذه المتغيرات ثم اعادة العملية على إحدى عشرة متغيرات فقط فظهرلنا وجود ستة عوامل رئيسية تؤثر في الظاهرة المدروسة و تشارك 71.014% من اجمالي التباين وهذا ان دل على شيء فانما يدل على جودة المعالجة المتبعة من خلال استبعاد المتغيرات غيرالمعنوية و كالآتي

العامل الاول

وجد أن هذا العامل يشكل 17.497% من اجمالي التباين و هذه النسبة تدل على مدى أهمية هذا

العامل في تفسير العلاقة بين المتغيرات المدروسة اما متغيرات هذا العامل فهي

1. هل انت مصاب بامراض اخري وبمقدار تشبع مساوي 0.888
2. امراض العيون وبمقدار تشبع مساوي 0.868

العامل الثاني

ياتي هذا العامل بالرتبة الثانية في الاهمية ويفسر 14.230% من اجمالي التباين ، و يحتوي هذا العامل على ثلاثة متغيرات وهي :

1. هل يوجد فتور بالجسم وبمقدار تشبع مساوي 0.762
2. هل توجد جروح بالرجل وبمقدار تشبع مساوي −0.737
3. ممارسة الحياة الزوجية بدون مشاكل وبمقدار تشبع مساوي 0.546

العامل الثالث

ياتي هذا العامل بالرتبة الثالثة في الاهمية ويفسر 9.990% من اجمالي التباين ، و يحتوي هذا العامل على متغير و احد يساهم بشكل أساسي في تكوين العامل وهو هل توجد جروح باليد بتشبع مقداره 0.940 وهذا يدل على مدى تأثير الجروح باليد

العامل الرابع

يفسر هذا العامل 9.8940% من اجمالي التباين و متغيرات هذا العامل هي:

1. بتر فى الاطراف بتشبع مقداره 0.795
2. هل توجد امراض وراثية اخرى فى العائلة بتشبع مقداره 0.518

العامل الخامس

ياتي هذا العامل بالرتبة الخامسة في الاهمية ويفسر 9.990% من اجمالي التباين ، و يحتوي هذا العامل على متغير و احد يساهم بشكل أساسي في تكوين العامل وهو أمراض الكلى بتشبع مقداره 0.845 وهذا يدل على مدى تأثير أمراض الكلى .

العامل السادس

يفسر هذا العامل 9.510% من اجمالي التباين و متغيرات هذا العامل هي:

1. امراض القلب بتشبع مقداره 0.874
2. الجهاز العصبى بتشبع مقداره 0.417

4-8 التحليل الإحصائي بإستخدام التحليل العنقودي لأهم عوامل الإصابة بمرض السكري :

مصفوفة القرابة بين عوامل الإصابة بمرض السكري

Proximity Matrix

جدول رقم (4-57)

Proximity Matrix

Case	Matrix File Input									
	المهنه	العمر	القرابة	التدخين	الكحول	الرياضة	ضغط الدم	BMIIIII	المستوي التعليمي	التغذية
المهنه	.000	.090	.997	.998	.998	.997	.997	.999	.998	.997
العمر	.090	.000	.000	.020	.021	.006	.001	.117	.047	.005
القرابة	.997	.000	.000	1.000	1.000	1.000	1.000	.996	.999	1.000
التدخين	.998	.020	1.000	.000	1.000	1.000	1.000	.997	.999	1.000
الكحول	.998	.021	1.000	1.000	.000	1.000	1.000	.997	.999	1.000
الرياضة	.997	.006	1.000	1.000	1.000	.000	1.000	.996	.999	1.000
ضغط الدم	.997	.001	1.000	1.000	1.000	1.000	.000	.996	.999	1.000

BMIIIIII	.999	.117	.996	.997	.997	.996	.996	.000	.998	.996
المستوي التعليمي	.998	.047	.999	.999	.999	.999	.999	.998	.000	.999
التغذية	.997	.005	1.000	1.000	1.000	1.000	1.000	.996	.999	.000

المصدر : إعداد الباحث من بيانات الدراسة

الجدول أعلاه يعطي مصفوفة القرابة ذات البعد (n*n) التي تستخدم كل تشابه (أو عدم تشابه) زوجي بين العناصر فمثلاً إذا كان لدينا x_i و x_j يمثلان العنصرين ith و jth بالتعاقب فإن عناصر مصفوفة القرابة (P_{ij}) ستمثل التشابه (S_{ij}) أو عدم التشابه (D_{ij}) بين x_i و x_j .

وتم إستخدام المسافة الإقليدية لحساب المسافة بين العناصر وفق الصيغة التالية :

$$dE(x_j, x_j) = \sqrt{\sum_{k=1}^{P}(x_{ik} - x_{jk})^2}$$

لكل $i, j \epsilon n$

هذه المصفوفة متماثلة بحيث ان القيم اعلى القطر الرئيسي تتطابق مع القيم أسفل هذا القطر . وبما ان المصفوفة مسافات فإن المقياس يعبر عن عدم التشابه او التقارب بين الوحدات ، وكلما كانت المسافة كبيرة دلت على وجود تشابه قليل بين الوحدتين .

جدول رقم (4-58)

خطوات التجميع

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	5	1.000	0	0	2
2	4	6	1.000	1	0	5
3	3	7	1.000	0	0	4
4	3	10	1.000	3	0	5
5	3	4	1.000	4	2	6
6	3	9	.999	5	0	8
7	1	8	.999	0	0	8
8	1	3	.997	7	6	9
9	1	2	.034	8	0	0

المصدر : إعداد الباحث من بيانات الدراسة

الجدول() يعطي خطوات التجميع ، فيمكن تحديد المفردات او المجموعات التي تم ربطها معاً في كل خطوة من خطوات التحليل . ففي الخطوة الاولى تم تجميع المفردة 4 مع المفردة 5 وكذلك في الخطوة الثانية تم تجميع المفردة 4 مع المفردة 6 والمسافة في الخطوتين متساوية مما يعني دمجهم معاً وتظهر المسافة في الجدول أعلاه في العمود المسمى بـ Coefficients والقيمة الحقيقية للمعاملات تعتمد على مقياس المسافة وطريقة الربط المتوسط Average Linkage Method. ففي هذه الطريقة يتم تعريف المسافة بين كل عنقودين كمتوسط المسافة بين كل زوج من المفردات ، ويتم دمج المفردات الأقرب ، وفي كل مرحلة يتم إدماج مفردات لتكون مجموعات جديدة ، واخيراً يتم إدماج المجموعات لتكون مجموعة وحيدة تضم كل المفردات .

في أغلب الأحيان تكون متابعة خطوات التجميع عن طريق الشجرة لأنها اكثر سهولة إلا ان هذا الشكل لم يوضح مقدار التجانس في المجموعات التي تم ضمها ، فقيمة المعامل الصغيرة تبين ان المجموعة متجانسة بينما القيمة الكبيرة تبين ان التجانس بين المجموعات اقل . فمثلاً التجانس بين المفردة 1 و 2 اكثر تجانساً من 1 و3 لن قيمة المعامل بين 1 و2 تساوي 0.034 وهي اقل من قيمة المعامل بين 1و3 وتساوي 0.997 لذلك يتم ضم 1 و2 بدلاً من 1 و 3 وهكذا .

أعضاء المجموعات

جدول رقم (4-59)

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
المهنه	1	1	1
العمر	2	2	2
القرابة	3	3	1
التدخين	3	3	1
الكحول	3	3	1
الرياضة	3	3	1
ضغط الدم	3	3	1
BMI	4	1	1
المستوي التعليمي	3	3	1

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
المهنه	1	1	1
العمر	2	2	2
القرابة	3	3	1
التدخين	3	3	1
الكحول	3	3	1
الرياضة	3	3	1
ضغط الدم	3	3	1
BMI	4	1	1
المستوي التعليمي	3	3	1
التغذية	3	3	1

المصدر : إعداد الباحث من بيانات الدراسة

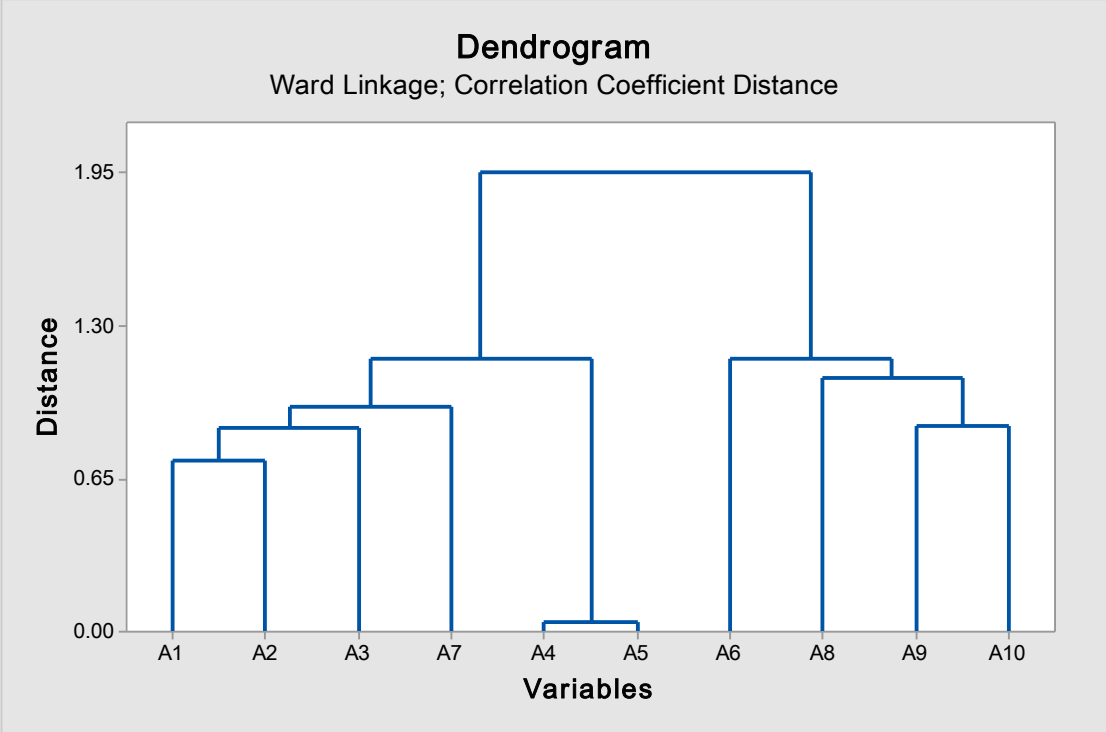
الجدول رقم () يوضح أعضاء المجموعات Cluster Membership من المفردات بإستخدام مدى الحلول Range Of Solution في برنامج التحليل وفيه تم تحديد عدد المجوعات من 2 إلى 4 . لتجميع أعضاء المجوعات إستخدام الاساليب التجميعية المتفرعة من الاساليب المتدرجة والتي تبداء بسلسلة من الإندماجات المتتالية من n من الوحدات والتي تتحول إلى مجموعات ، وتقوم هذه الاساليب بدمج المفردات أو مجموعة من المفردات الأكثر قرباً او تشابهاً .

نتائج التوزيع على مجوعات تظهر في الأعمدة الثلاث المسماة 2Clusters و 3Clusters و 4 Clusters .

ففي حالة التوزيع إلى مجموعتين فإن المجموعة الأولى تضم المهنة والقرابة و التدخين والكحول و الرياضة و ضغط الدم و BMI و المستوى التعليمي والتغذية أما المجموعة الثانية تضم العمر فقط . وفي حالة التوزيع إلى ثلاث مجموعات فإن المجموعة الاولى تضم المهنة و BMI وتضم المجموعة الثانية العمر فقط ، بينما تضم المجموعة الثالثة والقرابة و التدخين والكحول و الرياضة و ضغط الدم و المستوى التعليمي والتغذية .

اما في حالة التوزيع إلى أربعة مجموعات فإن المجموعة الأولى تضم المهنة فقط وتضم المجموعة الثانية العمر فقط وتضم المجموعة الثالثة القرابة و التدخين والكحول و الرياضة و ضغط الدم و المستوى التعليمي والتغذية ، بينما المجموعة الرابعة تضم BMI فقط

شكل رقم (4-39)

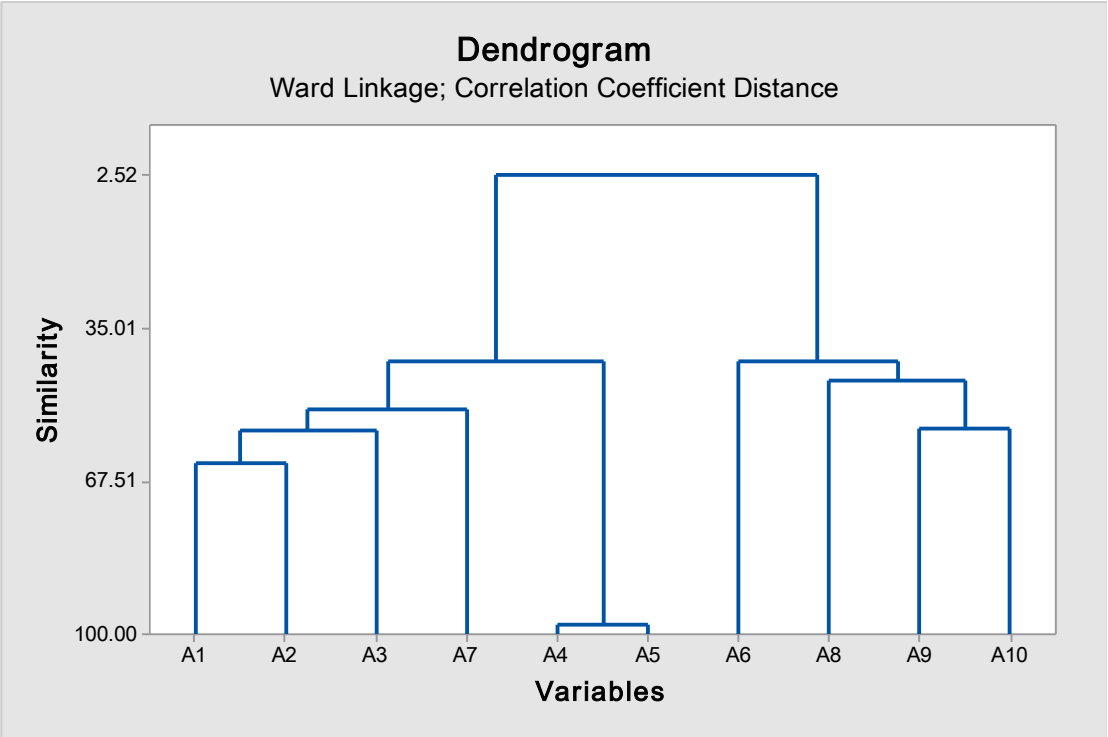


حيث:

A1 يمثل العمر ، A2 يمثل المهنة ، A3 يمثل درجة القرابة ، A4 يمثل التدخين ، A5 يمثل الكحول ، A6 يمثل حالة التغذية ، A7 يمثل الرياضة ، A8 يمثل ضغط الدم ، A9 يمثل المستوى التعليمي ، A10 يمثل تنسيب كتلة الجسم.

شكل رقم (4-40)

رسم الشجرة



حيث:

A1 يمثل العمر ، A2 يمثل المهنة ، A3 يمثل درجة القرابة ، A4 يمثل التدخين ، A5 يمثل الكحول ، A6 يمثل حالة التغذية ، A7 يمثل الرياضة ، A8 يمثل ضغط الدم ، A9 يمثل المستوى التعليمي ، A10 يمثل تنسيب كتلة الجسم.

شكل الشجرة Dendrogram الذي يصور خطوات التجميع والذي تم عرضه في الجدول () وأُستخدمت فيه طريقة الربط المتوسط ، كما يبين الشكل المراحل المختلفة لدمج مفردات المجموعات بإستخدام الطرق المختلفة لأساليب العنقدة المتدرجة . كما يمكّن من تحديد المفردات أو المجموعات التي تم ربطها معاً في كل خطوة من خطوات التحليل .

كما يوضح الشكل Dendrogram كيفية تكوين المجموعات ، ويقدم مقياساً لربط المسافات بغرض التجميع Linkage Distance .

بناءاً على ما تقدم فإنه يمكننا التوصل إلى أنه يمكن توزيع عوامل الإصابة بمرض السكري إلى أربعة مجموعات كما يلي :

المجموعة الأولى تضم المهنة فقط وتضم **المجموعة الثانية** العمر فقط وتضم **المجموعة الثالثة** القرابة و التدخين والكحول و الرياضة و ضغط الدم و المستوى التعليمي والتغذية ، بينما **المجموعة الرابعة** تضم تنسيب كتلة الجسم (BMI) فقط

4-9 التحليل الإحصائي بإستخدام التحليل العنقودي لأهم مضاعفات بمرض السكري :

جدول رقم (4-60)

Proximity Matrix

مصفوفة القرابة بين مضاعفات مرض السكري

Case	Matrix File Input									
	الإصابة بامراض اخري	امراض القلب	امراض العيون	امراض الكلى	الجهاز العصبى	الإصابة بأمراض وراثية	جروح باليد	جروح بالرجل	المضاعفات	بتر الاطراف
الإصابة بامراض اخري	.000	116.000	41.000	127.000	98.000	77.000	1.340E2	126.000	78.000	118.000
امراض القلب	116.000	.000	99.000	71.000	70.000	93.000	48.000	48.000	138.000	88.000
امراض العيون	41.000	99.000	.000	102.000	95.000	78.000	1.010E2	95.000	97.000	103.000
امراض الكلى	127.000	71.000	102.000	.000	69.000	104.000	51.000	53.000	137.000	93.000
الجهاز العصبى	98.000	70.000	95.000	69.000	.000	89.000	56.000	58.000	126.000	100.000
الإصابة بأمراض وراثية	77.000	93.000	78.000	104.000	89.000	.000	1.050E2	95.000	101.000	93.000
جروح باليد	134.000	48.000	101.000	51.000	56.000	105.000	.000	26.000	152.000	86.000
جروح بالرجل	126.000	48.000	95.000	53.000	58.000	95.000	26.000	.000	144.000	82.000
المضاعفات	78.000	138.000	97.000	137.000	126.000	101.000	1.520E2	144.000	.000	120.000
بتر الاطراف	118.000	88.000	103.000	93.000	100.000	93.000	86.000	82.000	120.000	.000

المصدر : إعداد الباحث من بيانات الدراسة

الجدول أعلاه يعطي مصفوفة القرابة ذات البعد (n*n) التي تستخدم كل تشابه (أو عدم تشابه) زوجي بين العناصر فمثلاً إذا كان لدينا x_i و x_j يمثلان العنصرين ith و jth بالتعاقب فإن عناصر مصفوفة القرابة (P_{ij}) ستمثل التشابه (S_{ij}) أو عدم التشابه (D_{ij}) بين x_i و x_j .

وتم إستخدام المسافة الإقليدية لحساب المسافة بين العناصر وفق الصيغة التالية :

$$dE(x_j, x_j) = \sqrt{\sum_{k=1}^{P}(x_{ik} - x_{jk})^2}$$

لكل $i, j \epsilon n$

هذه المصفوفة متماثلة بحيث ان القيم اعلى القطر الرئيسي تتطابق مع القيم أسفل هذا القطر .

وبما ان المصفوفة مسافات فإن المقياس يعبر عن عدم التشابه او التقارب بين الوحدات ، وكلما كانت المسافة كبيرة دلت على وجود تشابه قليل بين الوحدتين

جدول رقم (4-61)

خطوات التجميع

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	8	26.000	0	0	3
2	1	3	41.000	0	0	6
3	2	7	48.000	0	1	4
4	2	4	58.333	3	0	5
5	2	5	63.250	4	0	7
6	1	6	77.500	2	0	8
7	2	10	89.800	5	0	9
8	1	9	92.000	6	0	9
9	1	2	112.917	8	7	0

المصدر : إعداد الباحث من بيانات الدراسة

الجدول() يعطي خطوات التجميع ، فيمكن تحديد المفردات او المجموعات التي تم ربطها معاً في كل خطوة من خطوات التحليل . ففي الخطوة الاولى تم تجميع جروح باليد (المفردة 7) مع جروح بالرجل (المفردة 8) وكذلك في الخطوة وتم إختيار هذا الزوج في هذه الخطوة لان المسافة بينها أصغر من المسافة بين أي زوج آخر وتظهر المسافة في الجدول أعلاه في العمود المسمى بـ Coefficients والقيمة الحقيقية للمعاملات تعتمد على مقياس المسافة وطريقة الربط المتوسط Average Linkage Method .ففي هذه الطريقة يتم تعريف المسافة بين كل عنقودين كمتوسط المسافة بين كل زوج من المفردات ، ويتم دمج المفردات الأقرب ، وفي كل مرحلة يتم إدماج مفردات لتكون مجموعات جديدة ، واخيراً يتم إدماج المجموعات لتكون مجموعة وحيدة تضم كل المفردات .

في أغلب الأحيان تكون متابعة خطوات التجميع عن طريق الشجرة لأنها اكثر سهولة إلا ان هذا الشكل لم يوضح مقدار التجانس في المجموعات التي تم ضمها ، فقيمة المعامل الصغيرة تبين ان المجموعة متجانسة بينما القيمة الكبيرة تبين ان التجانس بين المجموعات اقل . فمثلاً التجانس بين جروح باليد (المفردة 7) مع جروح بالرجل (المفردة 8) اكثر تجانساً من جروح باليد (المفردة 7) و3 أمراض القلب (المفردة2) لان قيمة المعامل بين (المفردة 7) و(المفردة 8) تساوي 26 وهي اقل من قيمة المعامل بين (المفردة 7) و(المفردة2) وتساوي 41 لذلك يتم ضم 7 و8 بدلاً من 7 و 2 وهكذا .

جدول رقم (4-62)

أعضاء المجموعات

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
الإصابة بامراض اخري	1	1	1
امراض القلب	2	2	2
امراض العيون	1	1	1
امراض الكلى	2	2	2
الجهاز العصبى	2	2	2
الإصابة بأمراض وراثية	1	1	1
جروح باليد	2	2	2
جروح بالرجل	2	2	2
المضاعفات	3	3	1
بتر الاطراف	4	2	2

المصدر : إعداد الباحث من بيانات الدراسة

الجدول رقم () يوضح أعضاء المجموعات Cluster Membership من المفردات بإستخدام مدى الحلول Range Of Solution في برنامج التحليل وفيه تم تحديد عدد المجوعات من 2 إلى 4 . لتجميع أعضاء المجوعات إستخدام الاساليب التجميعية المتفرعة من الاساليب المتدرجة والتي تبداء بسلسلة من الإندماجات المتتالية من n من الوحدات والتي تتحول إلى مجموعات ، وتقوم هذه الاساليب بدمج المفردات أو مجموعة من المفردات الأكثر قرباً او تشابهاً .

نتائج التوزيع على مجوعات تظهر في الأعمدة الثلاث المسماة 2Clusters و 3Clusters و 4 Clusters .

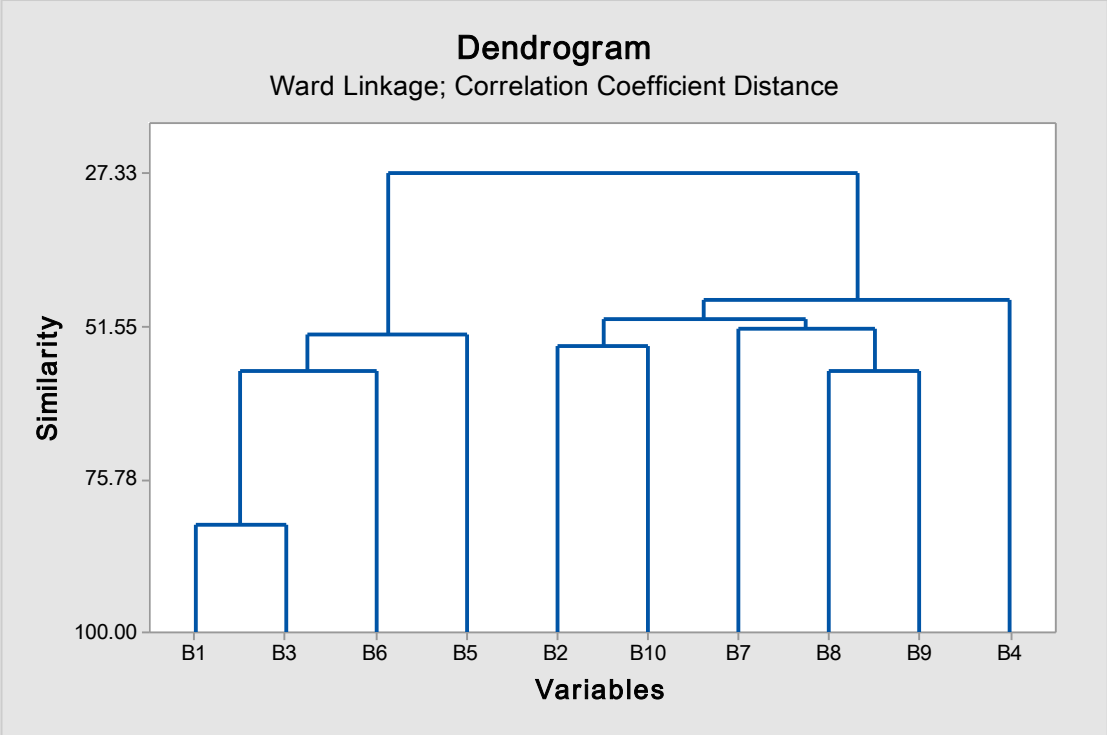
ففي حالة التوزيع إلى مجموعتين فإن المجموعة الأولى تضم الإصابة بأمراض اخرى و أمراض العيون الإصابة بأمراض وراثية و المضاعفات أما المجموعة الثانية أمراض القلب و امراض الكلى و أمراض الجهاز العصبي وجروح باليد وجروح بالرجل وبتر الأطراف. وفي حالة التوزيع إلى ثلاث مجموعات فإن المجموعة الاولى الإصابة بأمراض اخرى و أمراض العيون الإصابة بأمراض وراثية وتضم المجموعة الثانية أمراض القلب و امراض الكلى و أمراض الجهاز

العصبي وجروح باليد وجروح بالرجل وبتر الأطراف، بينما تضم المجموعة الثالثة المضاعفات فقط .

اما في حالة التوزيع إلى أربعة مجموعات فإن المجموعة الأولى تضم الإصابة بأمراض اخرى و أمراض العيون الإصابة بأمراض وراثية وتضم المجموعة الثانية أمراض القلب و امراض الكلى و أمراض الجهاز العصبي وجروح باليد وجروح بالرجل وتضم المجموعة الثالثة المضاعفات فقط ، بينما المجموعة الرابعة تضم بتر الأطراف فقط

شكل رقم (4-41)

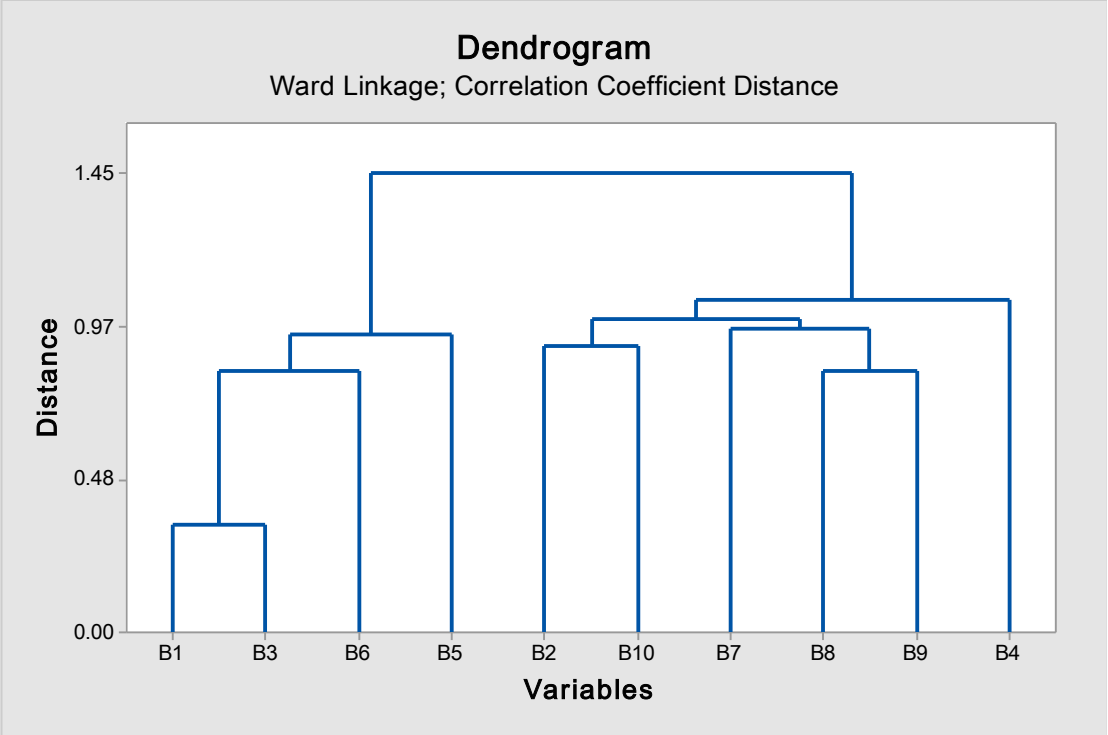
رسم الشجرة



حيث:

B1 يمثل هل انت مصاب بامراض اخرى ، B2 يمثل امراض القلب ، B3 يمثل امراض العيون، B4 يمثل امراض الكلى ، B5 يمثل امراض الجهاز العصبي ، B6 يمثل هل توجد امراض وراثية اخرى ، B7 يمثل جروح باليد ، B8 يمثل جروح بالرجل، B9 يمثل هل توجد مضاعفات ، B10 يمثل بتر في الاطراف.

شكل رقم (4-42)



حيث:

B1 يمثل هل انت مصاب بامراض اخرى ، B2 يمثل امراض القلب ، B3 يمثل امراض العيون، B4 يمثل امراض الكلى ، B5 يمثل امراض الجهاز العصبي ، B6 يمثل هل توجد امراض وراثية اخرى ، B7 يمثل جروح باليد ، B8 يمثل جروح بالرجل، B9 يمثل هل توجد مضاعفات ، B10 يمثل بتر في الاطراف.

شكل الشجرة Dendrogram الذي يصور خطوات التجميع والذي تم عرضه في الجدول () وأُستخدمت فيه طريقة الربط المتوسط ، كما يبين الشكل المراحل المختلفة لدمج مفردات المجموعات بإستخدام الطرق المختلفة لأساليب العنقدة المتدرجة . كما يمكّن من تحديد المفردات أو المجموعات التي تم ربطها معاً في كل خطوة من خطوات التحليل .

كما يوضح الشكل Dendrogram كيفية تكوين المجموعات ، ويقدم مقياساً لربط المسافات بغرض التجميع Linkage Distance .

بناءاً على ما تقدم فإنه يمكننا التوصل إلى أنه يمكن توزيع عوامل الإصابة بمرض السكري إلى أربعة مجموعات كما يلي :

المجموعة الأولى تضم الإصابة بأمراض اخرى و أمراض العيون الإصابة بأمراض وراثية وتضم **المجموعة الثانية** أمراض القلب و امراض الكلى و أمراض الجهاز العصبي وجروح باليد وجروح بالرجل وتضم **المجموعة الثالثة** المضاعفات فقط ، بينما **المجموعة الرابعة** تضم بتر الأطراف فقط

الفصل الخامس

النتائج والتوصيات

5-0 تمهيد :

بعد تطبيق أساليب تحليل المتغيرات المتعددة والتي منها أسلوب التحليل التميزي وأسلوب التحليل العاملي وأسلوب التحليل العنقودي علي بيانات مرضى السكري بولاية شمال كردفان توصلت الدراسة إلى عدة نتائج وتوصيات بالنسبة لعوامل الإصابة بمرض السكري ومضاعفاته أهمها :

5-1 النتائج :

أولاً:نتائج التحليل التميزي لأهم عوامل الإصابة بمرض السكري بولاية شمال كردفان :-

1. معادلة الدالة التمييزية معنوية أي هي قادرة على تصنيف الأشخاص إلي مصابين وغير مصابين بمرض السكري بناءاً على عوامل الإصابة بمرض السكري في ولاية شمال كردفان.
2. هناك تأثير معنوي من قبل متغيرات الدراسة (النوع ،العمر ،التاريخ العائلي للإصابة ، التدخين ، الكحول ،الرياضة ، ضغط الدم، تنسيب كتلة الجسم ، أمراض القلب ، حالة التغذية) في مستوي الإصابة بمرض السكري.
3. أهم المتغيرات وأكثرها تأثيراً على الإصابة بمرض السكري هي الكحول يليه التدخين ثم يليه العمر ثم يليه حالة التغذية ثم يلية التاريخ العائلي للإصابة واقلها تأثيراً ضغط الدم ،أمراض القلب ، الرياضة ، النوع ، تنسيب كتلة الجسم .
4. يمكن التنبؤ بالحالات المرضية للأشخاص من ناحية الإصابة بمرض السكري بناءاً على مضاعفات لمرض السكري في ولاية شمال كردفان عن طريق استخدام النقطة الفاصلة.
5. نسبة الخطأ في عملية التمييز صغيرة جداً وهذا يدل علي جودة الدالة التمييزية في عملية التصنيف وهو يتفق مع ما تم التوصل إلية من مقدرة الدالة على التمييز بناءاً على عوامل الإصابة بمرض السكري في ولاية شمال كردفان.

ثانياً:نتائج التحليل التميزي لأهم مضاعفات مرض السكري بولاية شمال كردفان :-

1. معادلة الدالة التمييزية معنوية أي هي قادرة على تصنيف الأشخاص إلي مصابين وغير مصابين بمرض السكري بناءاً على المضاعفات لمرض السكري في ولاية شمال كردفان.
2. هناك تأثير معنوي من قبل متغيرات (x_1 امراض القلب ، x_2 امراض العيون ، x_3 هل انت مصاب بامراض اخري ، x_4 امراض الكلى ، x_5 الجهاز العصبى ، x_6 هل توجد امراض وراثية اخرى فى العائلة ، x_7 هل توجد جروح باليد، x_8 هل توجد جروح بالرجل ، x_9 بتر فى الاطراف ، x_{10} هل توجد مضاعفات) في مستوي الإصابة بمضاعفات مرض السكري.
3. أهم المتغيرات وأكثرها تأثيراً على الإصابة بمرض السكري هي الكحول يليه التدخين ثم يليه العمر ثم يليه حالة التغذية ثم يلية التاريخ العائلي للإصابة واقلها تأثيراً ضغط الدم ،أمراض القلب ، الرياضة ، النوع ، تنسيب كتلة الجسم .
4. أهم المتغيرات وأكثرها تأثيراً على الإصابة بمضاعفات مرض السكري هو x_2 والذي يمثل امراض العيون والذي له اكبر تاثير في الأصابة بمرض السكري ثم يليه من حيث الأهمية x_5 والذي يمثل الجهاز العصبى ثم يليه x_4 والذي يمثل امراض الكلى ثم يلية x_{10} والذي يمثل هل توجد مضاعفات ثم يلية x_1 والذي يمثل امراض القلب ثم يليه x_9 والذي يمثل بتر فى الاطراف ثم يليه x_8 والذي يمثل هل توجد جروح بالرجل ثم يليه x_7 والذي يمثل هل توجد جروح باليد وأخيرا x_3 والذي يمثل هل انت مصاب بامراض اخري.
5. يمكن التنبؤ بالحالات المرضية للأشخاص من ناحية الإصابة بمرض السكري بناءاً على مضاعفات مرض السكري في ولاية شمال كردفان عن طريق استخدام النقطة الفاصلة.
6. نسبة الخطأ في عملية التمييز صغيرة جداً وهذا يدل علي جودة الدالة التمييزية في عملية التصنيف وهو يتفق مع ما تم التوصل إلية من مقدرة الدالة على التمييز بناءاً على مضاعفات مرض السكري في ولاية شمال كردفان.

ثالثاً :نتائج التحليل العاملي لأهم عوامل الإصابة بمرض السكري بولاية شمال كردفان :

7. **أظهرت طريقة المكونات الرئيسية أن هنالك أربعة عوامل إصابة بمرض السكري بولاية شمال كردفان .**

العامل الأول كان الجزر المميز له (Eigen Value = 4.111) ويفسر مانسبته % 41.102 من إجمالي التباين الكلي التباين لحالات مرض السكري . وقد إرتبط بخمسة متغيرات هي العمر والمهنة والتدخين والكحول والتغذية .

في حين أن العامل الثاني كان الجزر المميز له (Eigen Value =1.286) ويفسر مانسبته % 12.859 من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغير واحد فقط وهو الرياضية وذلك لأهميتها لما تمثلة من وقاية من الإصابة بمرض السكري .

أما العامل الثالث كان الجزر المميز له (Eigen Value = 1.226) ويفسر مانسبته 12.262 % من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغيرين هما ضغط الدم وتنسيب كتلة الجسم .

وأخيراً العامل الرابع كان الجزر المميز له (Eigen Value = 1.055) ويفسر مانسبته 10.546% من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغيرين أيضاً هما درجة القرابة والمستوى التعليمي.

8. كذلك العوامل الأربعة فسرت مجتمعة 76.769% من إجمالي التباين الكلي للمتغيرات الأصلية ، وهي نسبة عالية توضح أن الغالبية العظمى من المعلومات التي دخلت التحليل تضمنت في عملية التفسير .

9. كما نجد ان أقلة قيمة لمعاملات الشيوع لجميع المتغيرات الداخلة في التحليل هي 0.64 وهي أكبر من 0.5 مما يعني ان جزءاً كبيراً من البيانات المتعلقة بهذه المتغيرات قد ضمنت في العوامل التي تم أشتقاقها .

10. نجد أن العامل الأول يفسر % 41.102 من اجمالي التباين لحالات مرض السكري لذلك يعتبر اهم العوامل المشتقة .

11. من خلال جدول نتائج التحليل العاملي نلاحظ ظهورخمسة عوامل رئيسية تؤثر في مرض السكري وتمثل % 66.68 من اجمالي التباين وبعد د ا رسة متغيرات هذه العوامل وجد ان المتغيرين الحالة الاجتماعية ونوع الاصابة (ليس لهما اي دور سلبي او ايجابي في) تفسير نتائج عوامل الد ا رسة كما وبما ان التحليل العاملي يستخدم

لتقليص البيانات و استخلاص العوامل و المتغيرات المهمة التي تكون مؤثرة و فعالة على بيانات الد ا رسة فقد قمنا باستبعاد هذين المتغيرين ثم اعادة العملية على عشرة متغيرات فقط فظهرلنا وجود اربعة عوامل رئيسية تؤثر في الظاهرة المدروسة و تشارك 76.77%من اجمالي التباين وهذا ان دل على شيء فانما يدل على جودة المعالجة المتبعة من خلال استبعاد المتغيرات غيرالمعنوية

رابعاً: نتائج التحليل العاملي لأهم مضاعفات مرض السكري بولاية شمال كردفان –:

1. **أظهرت طريقة المكونات الرئيسية أن هنالك ستة عوامل تمثل أهم مضاعفات مرض السكري بولاية شمال كردفان .**

العامل الأول كان الجزر المميز له (Eigen Value =1.925) ويفسر مانسبته % 17.497 من إجمالي التباين الكلي لحالات مرض السكري . وقد إرتبط بمتغيرين هما الإصابة بإمراض أخرى وأمراض العيون .

في حين أن العامل الثاني كان الجزر المميز له (Eigen Value =1.565) ويفسر مانسبته % 14.23 من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط وقد إرتبط بثلاث متغيرات هي فتور بالجسم وجروح بالرجل والضعف الجنسي .

و العامل الثالث كان الجزر المميز له (Eigen Value = 1.099) ويفسر مانسبته % 9.99 من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغير واحد فقط وهو جروح بالرجل .

و العامل الرابع كان الجزر المميز له (Eigen Value = 1.088) ويفسر مانسبته 9.894% من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغيرين أيضاً هما بتر الأطراف والامراض الوراثية.

أما العامل الخامس كان الجزر المميز له (Eigen Value = 1.088) ويفسر مانسبته 9.894% من إجمالي التباين الكلي التباين لحالات مرض السكري. وقد إرتبط بمتغير واحد فقط وهو أمراض الكلى .

وأخيراً العامل السادس كان الجزر المميز له (Eigen Value =1.046) ويفسر مانسبته % 9.51 من إجمالي التباين الكلي التباين لحالات مرض السكري . وقد إرتبط بمتغير واحد فقط وهو أمراض القلب .

2. كذلك العوامل الستة فسرت مجتمعة 71.014% من إجمالي التباين الكلي للمتغيرات الأصلية ، وهي نسبة عالية توضح أن الغالبية العظمى من المعلومات التي دخلت التحليل ضمنت في عملية التفسير .

3. إن أكبر قيمة للتشبعات التي توضح مدى إلتصاق المتغيرات بالعوامل المشتقة كانت 0.888 من خلال متغير الإصابة بأمراض أخرى .

4. نجد أن العامل الأول يفسر % 17.497 من اجمالي التباين لحالات مرض السكري لذلك يعتبر اهم العوامل المشتقة .

5. من خلال جدول نتائج التحليل العاملي نلاحظ ظهور ستة عوامل رئيسية تعتبر من مضاعفات مرض السكري وتمثل 63.4% من اجمالي التباين وبعد د راسة متغيرات هذه العوامل وجد ان المتغيرات الثلاثة إرتفاع الدهون ، هل إرتفاع الدهون يسبب مضاعفات ، هل توجد مضاعفات (ليس لهم اي دور سلبي او ايجابي في) تفسير نتائج عوامل الد ا رسة كما وبما ان التحليل العاملي يستخدم لتقليص البيانات و استخلاص العوامل و المتغيرات المهمة التي تكون مؤثرة و فعالة على بيانات الد ا رسة فقد قمنا باستبعاد هذه المتغيرات ثم اعادة العملية على إحدى عشرة متغيرات فقط فظهرلنا وجود ستة عوامل رئيسية تؤثر في الظاهرة المدروسة و تشارك 71.014% من اجمالي التباين وهذا ان دل على شيء فانما يدل على جودة المعالجة المتبعة من خلال استبعاد المتغيرات غيرالمعنوية .

خامساً : نتائج التحليل العنقودي لأهم عوامل الإصابة بمرض السكري بولاية شمال كردفان :

أما فيما يتعلق بالتحليل العنقودي فيمكن تصنيف عوامل الإصابة بمرض السكري بولاية شمال كردفان وإعتماداً على عشرة متغيرات ، تم التوصل إلى أربعة عناقيد فإن العنقود او المجموعة

الأولى تضم المهنة فقط وتضم المجموعة الثانية العمر فقط وتضم المجموعة الثالثة القرابة و التدخين والكحول و الرياضة و ضغط الدم و المستوى التعليمي والتغذية ، بينما المجموعة الرابعة تضم تنسيب كتلة الجسم (BMI) فقط

سادساً :نتائج التحليل العنقودي لأهم مضاعفات مرض السكري بولاية شمال كردفان

أما فيما يتعلق بالتحليل العنقودي فيمكن تصنيف مضاعفات مرض السكري بولاية شمال كردفان وإعتماداً على عشرة متغيرات ، تم التوصل إلى أربعة عناقيد فإن العنقود او المجموعة الأولى تضم الإصابة بأمراض اخرى و أمراض العيون الإصابة بأمراض وراثية وتضم المجموعة الثانية أمراض القلب و امراض الكلى و أمراض الجهاز العصبي وجروح باليد وجروح بالرجل وتضم المجموعة الثالثة المضاعفات فقط ، بينما المجموعة الرابعة تضم بتر الأطراف فقط.

التوصيات :

1. يعتبر التحليل التمييزي والتحليل العاملي اسلوبين من أساليب تحليل المتغيرات المتعددة يكمل إحداهما الأخر لذلك نوصي بإستخدامهما معاً لكي تكتمل الصورة .

2. كما نجد أن التحليل التمييزي والتحليل العنقودي متشابهين من ناحية التصنيف لذلك نوصي بإستخدامهما في حالة المقارنة .
3. كذلك لوحظ من خلال الدراسات السابقة ان الخطاء في عملية التصنيف في حالة إستخدام التحليل التميزي دائماً صغير جدا مقارنة مع أساليب التصنيف الأخرى لذا نوصي بإستخدام التحليل التميزي إذا كان الغرض من الدراسة التصنيف او التمييز بين مجموعتين او أكثر
4. عمل برامج تثقيفية لمرضى السكري بخصوص كيفية التغذية الصحية .
5. عمل برامج تتضمن مسوحات للاكتشاف المبكر للحالات وسرعة علاجها.
6. نوصي بتنوع في الوجبات خلال اليوم ، لأن ذلك يوفر التوازن الغذائي .

المراجع

أولاً: المراجع العربية

1. أحمد ، عبد المجيد (2000) (شروط ومعايير إستخدام التحليل العاملي – دراسة إحصائية تطبيقية) رسالة ماجستير – جامعة أم القرى – مكة المكرمة .
2. باهي ، وآخرون (2002) (التحليل العاملي ، النظرية والتطبيق) المطبعة المركزية للكتاب – القاهرة .
3. الجبوري ،عابد (2000) (التحليل متعدد المتغيرات) – كتب مطبعة الصحافة – بغداد .
4. الحاج ، أزهري عبد الله (2002) (التحليل الإحصائي الشركة العالمية للطباعة والنشر ، السودان) .
5. زكريا ، عبد الجبار (1977) (المدخل إلى التحليل العاملي) مديرية دار الكتب للطباعة والنشر – الجامعة المستنصرية – العراق .
6. عكاشة ، محمود (2002) (التحليل العاملي والعنقودي بإستخدام SPSS) جامعة الأزهر – غزة .
7. فرج ، صفوت (1991) (التحليل العاملي في العلوم السلوكية) دار الفكر العربي – الطبعة الثانية – القاهرة .
8. محمد ، وفاء (أسلوب التحليل العاملي : عرض منهجي نقدي لعينة من الدراسات العربية أستخدمت التحليل العاملي) المجلس الأعلى لرعاية الفنون والآداب والعلوم الإجتماعية – وزارة التعليم العالي – الجمهورية العربية السورية .
9. نامق ، فيصل ناجي ، دراسة تحليلية مقارنة للاعوام (2006 , 2007 , 2008) لتصنيف محافظات العراق وفقاً للاصابات مرض الكبد الفيروسي بإستخدام التحليل العنقودي , 2012 – بغداد.
10. وشرن ، جونسون ريتشارد ، التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية، ترجمة عزام عبدالمرضي ، 1998، دار المريخ.

ثانياً : المراجع الأجنبية :

11. Agresti, A. (2002). Categorical data analysis. New York: Wiley.
12. An darson , T.W .(1984) An Introduction to Multivaiate Analysis Methods and Application , John Wiley & sons New York .U.S.A
13. Anderson, T.W. (1984). Introduction to multivariate statistics. New York: Wiley.
14. Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum.
15. Bentler, P.M. (2004). EQS structural equations program manual. Encino, CA: Multivariate Software.
16. C. R. Rao, The use and interpretation of principal component analysis in applied research, Sankhy¯a, A26 (1964) 329-358.
17. concepts, applications, and programming. Thousand Oaks, CA: Sage Publications.
18. Ferguson, C. (2007), Univariate and Multivariate Statistical Methodologies for Lake Ecosystem Modelling, PhD thesis, University of Glasgow.
19. Ferguson, G.A., and Takane, Y. (1989). *Statistical analysis in psychology and education*. 6th ed. New York: McGraw Hill.
20. Field, Andy P., (2005). Discovering statistics using spss (2nd edition), London, Sage. (Chapter 15).
21. Harman , H.H.(1976) " Modern Factor Analysis " the University Of Chicago press .
22. Hwang, H., and Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. Boca Raton, FL: Chapman and Hall/CRC Press.Inc.

23. J. Richard , (1975) (Aprimer of Multivaiate Statistics) Academic Press Inc- New York .
24. Johnson, R.A. & Wichern, D.W. (2002). Applied multivariate statistical analysis. Upper Saddle River, NJ: Prentice Hall.
25. Jolliffe, I.T. (1972). Discarding variables in a principal component analysis, I:Artificial data. Applied Statistics, 21, 160–173.
26. Jolliffe, I.T. (2002). Principal component analysis (second edition). New York, NY: Springer-Verlag, Inc.
27. Kaiser , H. F (1985) the Varimax Criterion For Analysis Rotation in Factor Analytic .
28. Kaiser, H.F. (1960). The application of electronic computer to factor analysis Educational and Psychological Measurement, 20, 141–151.
29. Krzanowski (1988)
30. L. L Thurstone , (1931) (Multiple Factor analysis) University of Chicago .
31. Little, R.J.A. & Rubin, D.B. (2002). Statistical analysis with missing data. New York: Wiley
32. M. A. Hunter, Y. Takane, Constrained principal component analysis: Various applications,Journal of Educational and Behavioral Statistics, 27 (2002) 41-81.
33. Mardia, K., Kent, J. and Bibby, J. (1979), *Multivariate Analysis*, Academic Press, Lon-don.
34. Morrison, D.F. (1976). Multivariate statistical methods. San Francisco, CA: McGraw Hill.

35. Multivariate analysis is by nature a more advanced topic. All of these books have merit. At present, I prefer Krzanowski and the
36. Olson, C.L. (1976). On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 83, 579–586.
37. Principles of Multivariate Analysis: A User's Perspective, Oxford
38. R.J.Rummel , T.Duane, C.Robert , J.Erin , (1999) (Evaluating the Use of Exploratory Factor analysis Psychological Researches) American Psychological Association .
39. R.L. Gorsuch ,(1974) (Factor analysis) U.B Sounders Company Philadelphia.
40. R.Mac callum, (2004) (Factor analysis class notes) Ohio State University .
41. R.Mac callum, (2004) (Factor analysis class notes) Ohio State University .
42. Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models. Thousand Oaks, CA: Sage.
43. Raykov, T. (2001). Testing multivariable covariance structure and means hypotheses via structural equation modeling. Structural Equation Modeling, 2, 224–257.
44. Rencher, A.C. (1995). An introduction to multivariate statistics. New York: Wiley.
45. Rencher, A.C. (1998). Multivariate statistical inference and applications. New York: Wiley.
46. Roussas, G.G. (1997). A course in mathematical statistics. New York: Academic Press.
47. S. James press , (1972) (Applied Multivaiate analysis) Holt, Rinehart and Winston . Inc- New York .

48. S. Sharma (1996) (Applied Multivaiate Techniques) Jonhn Wiley and son , Inc. U.S.A .
49. Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.
50. Statistical Society, Series B, 26, 211–246.
51. Takane, Y . (2013). *Constrained principal component analysis and related techniques*. Boca Raton, FL: Chapman and Hall/CRC Press
52. Takane, Y. (1980). *Multidimensional scaling*. Tokyo: University of Tokyo Press, (in Japanese).
53. Takane, Y. (1995). *Constrained principal component analysis*. Tokyo: Asakurashoten, (in Japanese).
54. Takane, Y., and Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive
55. Takane, Y., and Hunter, M. A. (2011). New family of constrained principal component analysis
56. Takane, Y., and Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. Psyhometrika, 56, 97-120.
57. Tatsuoka, M.M. (1971). Multivariate analysis. New York: Macmillan.
58. Tatsuoka, M.M. (1988). Multivariate analysis. New York: Macmillan.

theory. Appl Algebr Eng Comm, 12, 391-419.
59. Thurstone, L.L. (1947). Multiple-factor analysis. Chicago, IL: University of Chicago.
60. Timm, N.H. (2002). Applied multivariate analysis. New York: Springer.
61. W.Hardle , L . Simar , (2007) (Applied Multivaiate Statistical analysis)
62. Y. Takane, M. A. Hunter, Constrained principal component analysis: A comprehensive theory

Applicable Algebra in Engineering,

Communication, andComputing, 12 (2001) 391-

419.

63. Y. Takane, T. Shibayama, Principal component analysis with external information on both subjects and variables, Psychometrika, 56 (1991) 97-120.

64. Yanai, H., Takeuchi, K., and Takane, Y . (2011). *Projection matrices, generalized inverse matrices, and singular value decomposition*. New York: Springer.

الملاحق

Tests of Equality of Group Means

	Wilks' Lambda	F	df 1	df 2	Sig.
النوع	.979	4.602	1	214	.033
العمر	.831	43.564	1	214	.000
التاريخ العائلي	.951	10.924	1	214	.001
التدخين	.925	17.431	1	214	.000
الكحول	.910	21.216	1	214	.000
الرياضة	.979	4.671	1	214	.032
ضغط الدم	.976	5.172	1	214	.024
BMI	.977	5.033	1	214	.026
امراض القلب	.976	5.196	1	214	.024
حاله التغذية	.897	24.635	1	214	.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
امراض القلب	.977	4.910	1	213	.028
امراض العيون	.911	20.810	1	213	.000
هل انت مصاب بامراض اخري	.978	4.725	1	213	.031
امراض الكلى	.981	4.223	1	213	.041
الجهاز العصبى	.977	5.112	1	213	.025
هل توجد امراض وراثية اخرى فى العائلة	.978	4.751	1	213	.030
هل توجد جروح باليد	.984	3.543	1	213	.050
هل توجد جروح بالرجل	.969	6.734	1	213	.010
بتر فى الاطراف	.976	5.237	1	213	.023
هل توجد مضاعفات	.972	6.069	1	213	.015

مصفوفة القرابة لعوامل الإصابة

Proximity Matrix

Case	Matrix File Input									
	المهنه	العمر	القرابة	التدخين	الكحول	الرياضة	ضغط الدم	BMIIIIII	المستوي التعليمي	التغذية
المهنه	.000	.090	.997	.998	.998	.997	.997	.999	.998	.997
العمر	.090	.000	.000	.020	.021	.006	.001	.117	.047	.005
القرابة	.997	.000	.000	1.000	1.000	1.000	1.000	.996	.999	1.000
التدخين	.998	.020	1.000	.000	1.000	1.000	1.000	.997	.999	1.000
الكحول	.998	.021	1.000	1.000	.000	1.000	1.000	.997	.999	1.000
الرياضة	.997	.006	1.000	1.000	1.000	.000	1.000	.996	.999	1.000
ضغط الدم	.997	.001	1.000	1.000	1.000	1.000	.000	.996	.999	1.000
BMIIIIII	.999	.117	.996	.997	.997	.996	.996	.000	.998	.996
المستوي التعليمي	.998	.047	.999	.999	.999	.999	.999	.998	.000	.999
التغذية	.997	.005	1.000	1.000	1.000	1.000	1.000	.996	.999	.000

المصدر : إعداد الباحث من بيانات الدراسة

خطوات التجميع لعوامل الإصابة

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	5	1.000	0	0	2
2	4	6	1.000	1	0	5
3	3	7	1.000	0	0	4
4	3	10	1.000	3	0	5
5	3	4	1.000	4	2	6
6	3	9	.999	5	0	8
7	1	8	.999	0	0	8
8	1	3	.997	7	6	9
9	1	2	.034	8	0	0

المصدر : إعداد الباحث من بيانات الدراسة

أعضاء المجموعات لعوامل الإصابة

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
المهنه	1	1	1
العمر	2	2	2
القرابة	3	3	1
التدخين	3	3	1
الكحول	3	3	1
الرياضة	3	3	1
ضغط الدم	3	3	1
BMI	4	1	1
المستوي التعليمي	3	3	1
التغذية	3	3	1

المصدر : إعداد الباحث من بيانات الدراسة

Proximity Matrix

مصفوفة القرابة بين مضاعفات مرض السكري

Case	Matrix File Input									
	الإصابة بامراض اخري	امراض القلب	امراض العيون	امراض الكلى	الجهاز العصبى	الإصابة بأمراض وراثية	جروح باليد	جروح بالرجل	المضاعفات	بتر الاطراف
الإصابة بامراض اخري	.000	116.000	41.000	127.000	98.000	77.000	1.340E2	126.000	78.000	118.000
امراض القلب	116.000	.000	99.000	71.000	70.000	93.000	48.000	48.000	138.000	88.000
امراض العيون	41.000	99.000	.000	102.000	95.000	78.000	1.010E2	95.000	97.000	103.000
امراض الكلى	127.000	71.000	102.000	.000	69.000	104.000	51.000	53.000	137.000	93.000
الجهاز العصبى	98.000	70.000	95.000	69.000	.000	89.000	56.000	58.000	126.000	100.000
الإصابة بأمراض وراثية	77.000	93.000	78.000	104.000	89.000	.000	1.050E2	95.000	101.000	93.000
جروح باليد	134.000	48.000	101.000	51.000	56.000	105.000	.000	26.000	152.000	86.000
جروح بالرجل	126.000	48.000	95.000	53.000	58.000	95.000	26.000	.000	144.000	82.000
المضاعفات	78.000	138.000	97.000	137.000	126.000	101.000	1.520E2	144.000	.000	120.000
بتر الاطراف	118.000	88.000	103.000	93.000	100.000	93.000	86.000	82.000	120.000	.000

المصدر : إعداد الباحث من بيانات الدراسة

خطوات التجميع لمضاعفات مرض السكري

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	8	26.000	0	0	3
2	1	3	41.000	0	0	6
3	2	7	48.000	0	1	4
4	2	4	58.333	3	0	5
5	2	5	63.250	4	0	7
6	1	6	77.500	2	0	8
7	2	10	89.800	5	0	9
8	1	9	92.000	6	0	9
9	1	2	112.917	8	7	0

المصدر : إعداد الباحث من بيانات الدراسة

أعضاء المجموعات لمضاعفات مرض السكري

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
الإصابة بامراض اخري	1	1	1
امراض القلب	2	2	2
امراض العيون	1	1	1
امراض الكلى	2	2	2
الجهاز العصبى	2	2	2
الإصابة بأمراض وراثية	1	1	1
جروح باليد	2	2	2
جروح بالرجل	2	2	2
المضاعفات	3	3	1
بتر الاطراف	4	2	2

المصدر : إعداد الباحث من بيانات الدراسة