

0-1: تمهيد

مع تطور ونقدم أساليب التنبؤ أصبح هناك العديد من الطرق والتحليل الإحصائية التي تساعد على دراسة المتغيرات التي لم يتم دراستها من قبل، ومن هذه الطرق التحليل التمييزي و الانحدار اللوجستي وهما يهدفان لمعرفة أكثر المتغيرات تأثيرا على المتغير المصنف أو المعتمد وتصنيف الحالات الجديدة اعتمادا علي البيانات السابقة.

وانه لمن المناسب من هذه النقطة أن نستخدم أسلوب (التحليل التمييزي و الانحدار اللوجستي) في إحدى المجالات الطبية وذلك نتيجة لما تتميز به متغيرات هذا المجال من ضخامة في عددها مما يجعلها ملائمة للدراسة .

1-1 : مشكلة البحث

تكمن مشكلة الدراسة إلي المعرفة و الوقوف على بعض العوامل وأهمية دورها اتجاه مرض سرطان الدم، و ذلك من خلال الأسلوب التمييزي و الانحدار اللوجستي و استخدامهما في عملية تصنيف المتغيرات المؤدية للإصابة بمرض سرطان الدم.

2-1:أهمية البحث

إن قلة البحوث والدراسات التي تناولت الموضوع بصورة تطبيقية ومعقدة زادت من أهمية هذا البحث ، حيث لازال هذا المجال بحاجة إلي المزيد من الدراسات التي تمكن من الوقوف على أكبر قدر من العوامل ذات التأثير على سرطان الدم مما يساعد على رسم وتخطيط السياسات المناسبة في مجال محاربة هذا المرض اللعين. يمكن أن تسهم نتائج الدراسة في تقديم مجموعة من الإجراءات التي يمكن تطبيقها .إضافة إلى لفت أنظار الباحثين و المصابين بسرطان الدم إلى أهمية وضرورة استخدام التحليل التمييزي والانحدار اللوجستي و مدى كفاءة كل منهما في التصنيف .

1-3: حدود البحث

الحدود المكانية : تم جمع بيانات 102 مريض جميعهم مصابين بسرطان الدم بأنواعه المختلفة من المركز القومي للعلاج بالأشعة والطب النووي (الذرة) بالخرطوم .

الحدود الزمانية :الذين تم تسجيلهم كمصابين بسرطان الدم في الفترة من يناير 2012م-ديسمبر 2015م.

1-4: أهداف البحث

- تسليط الضوء على المتغيرات التي تؤدي الى الاصابه بسرطان الدم .
- إبراز أهمية كل من التحليلين (التمييزي و الانحدار اللوجستي).
- إثبات مدي كفاءة كل من التحليل التمييزي و الانحدار اللوجستي .

1-5: فرضيات البحث

هنالك تأثير من قبل المتغيرات (النوع ،العمر،الحالة الاجتماعية،الوظيفة،كريات الدم البيضاء،كريات الدم الحمراء)في تحديد نوع الإصابة بمرض سرطان الدم(حاد،مزمن).
المقارنة بين أسلوب الدالة التمييزية و الانحدار اللوجستي و معرفة أي من الطريقتين تؤدي إلي نتائج أدق في التنبؤ أكثر من استخدام الطريقة الأخرى .

1-6: منهجية البحث

في هذا البحث تم استخدام الأسلوب الوصفي عن طريق وصف متغيرات الدراسة واستخدام المنهج التحليلي عن طريق تحليل البيانات وتكوين الدالة التمييزية والانحدار اللوجستي عبر إدخال المتغيرات المؤثرة و تنبؤاتها الخاصة و معالجتها للحصول علي أفضل نتائج . كما تمت المقارنة بين أسلوب الدالة التمييزية و الانحدار اللوجستي من حيث دقة التصنيف لمعرفة أفضل طريقة وتم استخدام برنامج التحليل الإحصائي (SPSS) في عملية التحليل.

7-1: هيكلية البحث

يحتوي هذا البحث على خمس فصول وقد تناول الفصل الأول مقدمة البحث والتي اشتملت على مشكلة وأهمية وأهداف البحث واحتوت أيضا علي فروض وحدود ومنهجية البحث وأيضا البحوث والدراسات السابقة ، أما الفصل الثاني يحتوي علي معلومات عن سرطان الدم ، بينما يحتوي الفصل الثالث علي أسلوب الدالة التمييزية و كيفية حسابها ثم نموذج الانحدار اللوجستي ، و في الفصل الرابع تم تناول الجانب التطبيقي ويتكون من أربع مباحث المبحث الاول يتضمن وصف كافي لمفردات عينة الدراسة ،والمبحث الثاني يتضمن بناء الدوال واستخدامها في عملية التصنيف ،والمبحث الثالث يتضمن بناء نموذج الانحدار اللوجستي،والمبحث الرابع يتضمن مقارنة بين الانحدار اللوجستي والدالة التمييزية. وأخيرا الفصل الخامس الذي يشتمل علي النتائج والتوصيات.

8-1: البحوث و الدراسات السابقة

فيما يلي بعض البحوث والدراسات السابقة التي تم فيها استخدام أسلوب الدالة التمييزية و الانحدار اللوجستي :

1. دراسة ظافر رمضان 1998م بعنوان التحليل التمييزي وأمراض أسنان الأطفال ، تناولت الدراسة التحليل التمييزي وتم استخدامه للحصول على دالة فعلية يمكن بها التمييز بين الأطفال صغار السن من حيث إصابتهم بتسوس الأسنان العادي والتسوس المتفشي ذلك أن التسوس المتفشي يعني إصابة أكثر من 70% من أسنان الطفل بتسوس أو بضعف في مادة السن ومن خلال الدالة التمييزية التي يتم الحصول عليها يمكن التمييز بين هاتين الحالتين ثم تقديم العلاج المناسب لكل حالة.

2. في عام 2008 قام الباحثة أمل حسن محمد يس بإعداد دراسة بعنوان استخدام تحليل الانحدار اللوجستي لمعرفة مسببات الاصابه بسرطان الثدي لدي الإناث : دراسة حاله المركز القومي للعلاج بالاشعه والطب النووي قامت هذه الدراسة

باستخدام اسلوب الانحدار اللوجستي اخذي عينه عشوائية بسيطة من مجتمع للمرضي المترددين علي المركز القومي للعلاج بالطب النووي بالخرطوم وغير المصابين بسرطان الثدي خارج هذا النطاق وقد تم تحديد حجم العينة وهو 200 فرد 100 مصابين و100 غير مصابين .

تهدف هذه الدراسة إلي معرفه أكثر المتغيرات تأثيرا علي الاصابه بسرطان الثدي لدي الإناث تم جمع البيانات بواسطة الاستبيان ومن ثم تم استخدام اسلوب الانحدار اللوجستي في تحليل تلك البيانات .

3. تقدمت عفاء هاشم عبد اللطيف محمد برسالة بعنوان : تطبيق التحليل التمييزي لتصنيف الإصابة بسرطان الثدي لدى الإناث حسب مراحلها المتعددة في عام 2011م لكلية الدراسات العليا بجامعة السودان للعلوم و التكنولوجيا لنيل درجة الدكتوراه PH.D.

0-2 : تمهيد⁽⁵⁾

الدم عبارة عن نسيج سائل من أشكال النسيج الضام وهو يجري في الأوعية الدموية حيث يصل إلى جميع أجزاء الجسم ، وتسمى حركته في الجسم بالدورة الدموية ، وعلى هذه الحركة تتوقف الحياة . ويبلغ حجمه في الإنسان البالغ من 6 إلى 5 لترات (أي حوالي 6-8 % من وزن الجسم تقريبا) ويتكون سائل الدم من جزئيين أساسيين هما : البلازما (معظمها ماء) والخلايا التي تنقسم إلى ثلاثة أنواع : الحمراء والبيضاء والصفائح الدموية ، ومن المعلوم أن للدم وظائف متعددة وما يهم هنا هو وظائف خلايا الدم البيضاء التي تعتبر خط الدفاع الأول عن جسم الإنسان حيث تقوم بالتهاجم الأجسام الغريبة والجراثيم بالإضافة للخلايا الليمفاوية التي تقوم بإفراز الأجسام المضادة التي تهاجم الميكروبات والجراثيم وبذلك تحمي الجسم من أخطارها ومن الأمراض التي تسببها .

2-1: سرطان الدم(5)

سرطان الدم هو عبارة عن تكاثر غير محكوم لخلايا الدم ، وعادة ما يكون هذا التكاثر لخلايا الدم البيضاء ، حيث أن عددها في الدم يزيد عن الحد الطبيعي وهو من 5000 إلى 10000 خلية لكل ملم ، ففي سرطان الدم يكون عدد الخلايا من 15000 إلى 30000 خلية لكل ملم ، وقد يصل هذا التكاثر إلى أكثر من 100000 خلية لكل ملم أو حتى أن يكون أقل من 5000 خلية لكل ملم (أي أقل من العدد الطبيعي) .

ويعود اختلاف هذا العدد إلى مرحلة سرطان الدم ومدى تقدمه في الجسم ، فخلايا كريات الدم البيضاء غير الطبيعية تنتج في نخاع العظم وهو مصنع خلايا كرات الدم جميعها في الجسم ، فيؤدي تكاثرها في نخاع العظم إلى خروجها في الدم ، ففي تكاثرها في المراحل الأولى يؤدي ذلك إلى تقليل عدد خلايا كرات الدم المصنعة وقبل أن تخرج كرات الدم البيضاء إلى خارج الدم مما يؤدي إلى وجود نقص مرحلي في كرات الدم البيضاء في الدم في بداية المرض ، ولكن سرعان ما يصدر نخاع الجسم هذه الخلايا السرطانية إلى الدم ليزيد بذلك عدد كريات الدم البيضاء في الدم ، وهذه الزيادة عادة تكون زيادة مضطربة .

2-1-1: الأسباب المؤدية للميل إلى نشوء السرطان بصورة عامة(5)

1. الاختلالات الوراثية .
2. التعرض للإشعاع .
3. العلاج الإشعاعي والكيميائي .
4. المواد الكيميائية .
5. أمراض الدم المؤدية لسرطان الدم .
6. الفيروسات

3-2-2: الأغذية المسببة للسرطان(5)

1. الأغذية التي تتعرض لدرجات حرارة عالية أو للنار المباشر بالذات اللحوم المشوية والمقلية .

2. اللحوم المملحة والمعالجات بالكيماويات من أجل حفظها أو المحافظة على ثبات لونها .
 3. الأغذية التي تم رشها بالمبيدات الحشرية ولم تترك لعدة أيام قبل جلبها للسوق وتناولها .
 4. زيادة الدهون المشبعة في الغذاء والتي أكدت الأبحاث علاقتها بسرطان الثدي والبروستاتا وبطانة الرحم .
 5. السمنة وربما لعلاقتها بتناول الدهون الكثيرة ولعا علاقة بسرطان القولون وبطانة الرحم بذات لدى النساء المنقطع لديهن الطمث ، والمثانة والمريء والبنكرياس والكلى .
 6. بعد الملونات الغذائية والمواد الحافظة .
- 2-2-3: أعراض المصاب بسرطان الدم⁽⁵⁾**

- من آثار هذه الزيادة أن يصاب المريض بنقص الهيموجلوبين وبالتالي يفقر في الدم فيبدأ يعاني من تعب من أي مجهود يقوم به والغثيان وكثرة الميل للثوم والكسل .
 - مع استمرار المرض يعاني المريض من نقص في الصفائح الدموية فيصبح دمه يتأخر في التجلط ويكون عدة كدمات في الجلد نتيجة لعدم قدرة الجسم على وقف الجلطات .
 - نزيف متكرر من الأنف أو اللثة وزيادة كرات الدم البيضاء لدرجة كبيرة في الدم تمنع كرات الدم الحمراء الحاملة للأوكسجين من الوصول لخلايا الجسم مما يصيب المريض بجلطات صغيرة مختلفة مع آلام في البطن والصدر . كما أن هذه الزيادة تضعف من مقاومة الجسم للأمراض المختلفة فتجد أنه غالباً ما يأتي المصاب بعدة التهابات بكتيرية أو فيروسية قبل أن يشخص بمثل هذا المرض .
- 2-2-4: أنواع سرطان الدم⁽⁵⁾**

هناك عدة أنواع من سرطان الدم تصل إلى أكثر من عشرين نوعاً ، لكل نوع منها صفاته وعلاجه الخاص ولكن هناك نوعين رئيسيين ينضم تحتها عدة أنواع أخرى وهما سرطان الدم الحاد وسرطان الدم المزمن .

1. سرطان الدم الحاد : إذ يتم تكاثر خلايا الدم بشكل سريع جداً وتكون قاصرة على القيام بالوظائف الخاصة بها .
 2. سرطان الدم المزمن : إذ تكون حالة المريض سيئة للغاية والخلايا السرطانية تم انتشارها في جميع أرجاء الدم ويقضي على الجسم ببطء شديد .
 3. سرطان الدم الليمفاوي : وذلك يحدث عندما يصل الضرر إلى خلايا الدم البيضاء الناتجة عن النسيج الليمفاوي .
 4. سرطان الدم نخاعي : عندما تتضرر خلايا الدم البيضاء التي تنشأ من نخاع العظام .
- فسرطان الدم قد يصيب عدة أنواع من خلايا كرات الدم البيضاء أو الحمراء أو الصفائح الدموية فلكل نوع من هذه الأنواع علاجه الخاص به. فبعض هذه الأنواع أصبح بالإمكان العلاج الشافي منها وبعض الأنواع علاجها أصعب . كما أن بعض الأنواع قد يحتاج لعملية زراعة نخاع

العظم لشفائها تماما . ويساعدنا على هذا التشخيص الدقيق لمثل هذه الحالات وذلك بعمل صبغات خاصة تصبغ كل خلية من نوع معين صبغة مختلفة ومن مجموع هذه الصبغات نستطيع التعرف على نوع هذه الخلايا السرطانية .

5-2-2: تشخيص سرطان الدم⁽⁵⁾

يشخص سرطان الدم بعمل فحوصات الدم الروتينية في مختبر به كفاءات عالية وذلك لقراءة شريحة الدم ورؤية إن كان بها خلايا غير طبيعية . فإن وجدت بها خلايا غير طبيعية ومنها نلجأ إلى أخذ عينة من نخاع العظم وأفضل مكان لأخذ العينة هو عظمة الحوض حيث أنها مكان آمن من أية مشاكل . ومن دراسة خلايا نخاع العظم يقدر إن كان هناك خلايا سرطانية أم لا . فإن وجدت فإن الخلايا السرطانية تصبغ بعدة صبغات خاصة ونتيجة هذه الصبغات الخاصة تحدد ما هو نوع يشخص سرطان الدم بالتحديد . ويؤخذ عينة لدراسة كروموسومات سرطان الدم فكثير من العلاجات تعتمد وبشدة على نوع التغيرات الكروموسومية التي تسبب سرطان الدم ، وثم يبا المريض العلاج المناسب لهذا النوع .

6-2-2: علاج سرطان الدم⁽⁵⁾

سرطان الدم الحاد علاجه مختلف عن سرطان الدم المزمن فكل منهما علاجه الخاص مع تقدم الطب واكتشاف العلماء ، التغيرات الجينية في كل سرطان أصبح من الممكن إصلاح هذه التغيرات الجينية ، فقد أصبح لبعض الأنواع الآن علاجاً شافياً بأدوية قد لا يكون علاج كيميائي بيولوجي يؤثر فقط إلى درجة كبيرة على الخلايا السرطانية دون أي تأثير على باقي خلايا الجسم . ونذكر على سبيل المثال أن نوع من أنواع سرطان الدم الحاد ويسمى (م3) نستطيع أن نشفيه بإعطاء كميات وجرعات عالية من فيتامين أ .

ونوع آخر من سرطان الدم المزمن يسمى CML (CHRONIC MYEL OID LEUKEMIA) أصبح يمكن علاجه بطريقة فعالة بإعطاء المريض من مثبطة لإنزيم خاص ويعرف تجارياً بدواء "جليفك" ، هذين النوعين بالذات تحقق لهما نجاحاً كبيراً في العشر سنوات الأخيرة ، كما أن التطور لزم معظم باقي الأنواع . وذلك في عملية زراعة نخاع العظم لإبداله بدم المريض الفاسد . وكثير من هذه العمليات ناجحة الآن . إن سرطان الدم في المستقبل القريب سوف يصبح بالإمكان معالجة معظم أنواعه بفعالية كبرى .

7-2-2: الأغذية المضادة للسرطان⁽⁵⁾

عموماً أثبتت الدراسات أن تناول الخضار والفواكه الطازجة يقلل وبشكل واضح فرص الإصابة بالسرطان والأغذية التالية إذا تم تناولها بشكل متوازن مفيدة للوقاية من السرطان لكن لم يثبت بعد هل زيادة تناولها تفيد أكثر أم لا وهي :

1. الأغذية من الفصيلة الصليبية : منها البروكلي والزهرة والكرنب وجميع أنواع الملفوف وكذلك الخردل ، هذه المواد تحتوي على أحد مركبات الكبريت .

2. الثوم والبصل : لاحتوائها على مركبات الكبريت المضادة للسرطان الذي ينشط إفراز بعض الإنزيمات .
 3. الموالح : لوجود عناصر غذائية مضادة مثل الفلافينويد وفيتامين ج التي تعمل كمضاد للأكسدة ومنع السرطان .
 4. الصويا : يحتوي كمية كبيرة من مادة الإيزوفلافونز التي توقف بعض النشاط الهرموني داخل الخلايا ، وللصويا دور كبير في الوقاية من سرطان الثدي والبروستاتا على وجه الخصوص . ويمكن تناوله على هيئة فول الصويا أو يضاف زيتته للغذاء أو بتناول حليب الصويا .
 5. الطماطم : غنية بمادة ليكربين وهي مادة مضادة قوية للأكسدة وهي المسببة للون الأحمر ، ووجدت الأبحاث أن الطماطم المطبوخة تقلل من سرطان المعدة والقولون والبروستاتا .
 6. الفواكه والخضروات الطازجة بصورة عامة فقد أثبتت الأبحاث أهميتها في تقليل خطر الإصابة بأنواع كثيرة من السرطانات .
- 2-2-8: مستقبل المصاب بسرطان الدم(5)**

لمعرفة مستقبل المصاب بسرطان الدم يجب معرفة نوع سرطان الدم ، فإن كان من الأنواع التي فيها تطور كبير فإن معظم هؤلاء المرضى يشفون . أما إن كان من الأنواع الأقل تطوراً في العلاج فإن أمورهم ستكون مرهونة بتقدمهم في العلاج وأثر العلاج عليهم . كما أنه يجب التنويه على أن سرطان الدم الذي يصيب الأطفال خاصة له علاقة كبيرة في السن الذي يصاب به الطفل فكلما كان الطفل صغيراً (ولكن لا يكون أقل من سنة في العمر) كلما زاد ذلك من فرص نجاته من سرطان الدم .

أي أن للحكم على مدى نجاح كل فرد مصاب بسرطان الدم فالتقدم مرهون بعدة عوامل منها نوع سرطان الدم بالتحديد ، نوع التغيرات الكروموسومية في سرطان الدم ، عمر المريض ، جنس المريض (إن كان ذكراً أو أنثى) ، نسبة الخلايا السرطانية في الدم ، ووجود تراكمات من سرطان الدم خارج الدم (في الأنسجة أم لا) .

كما يجب عدم إهمال الفحوصات الدورية والتي من شأنها أن تكشف أموراً مبكرة قبل أن يبدأ الجسد بالإحساس بالأعراض من وجود مثل هذه الأمراض والذي قد يكون متقدماً في ذلك الحين .

3-0: تمهيد :

تصنيف الظواهر يعتبر من اهم التطبيقات الاحصائية التي تستخدم في التنبؤ ويتم فيها التنبؤ بوجود او عدم وجود الحدث اعتمادا علي بيانات حقيقيه مصنفه تحتوي علي عدد من المتغيرات التي تؤثر علي المتغير المصنف .

تجدر الاشارة الي انه توجد العديد من الطرق الاحصائية التي قد تستخدم في التصنيف وسيتم تناول بعض الطرق التي تستخدم في التصنيف (التحليل التمييزي و الانحدار اللوجستي) .

3-1: التحليل التمييزي⁽⁴⁾

التحليل المميز هو أسلوب إحصائي لتحليل البيانات متعددة المتغيرات، يصنف مفردات العينة إلى مجموعتين أو أكثر . الدالة المميزة هي عبارة عن تركيب خطي للمتغيرات المستقلة ،والدالة المميزة هي التي تقوم بعملية التصنيف (التمييز). وعملية التصنيف تأتي بعد تكوين الدالة المميزة ، حيث يعتمد عليها في تصنيف المفردات الجديدة لإحدى المجموعات قيد الدرس بأقل خطأ تصنيف ممكن .

التمييز الخطي هو إحدى حالات التمييز التي يتطلب فيها شرط تساوي التباينات للمجموعات المبحوثة ،فهناك تمييز خطي في حالة مجموعتين، وتمييز خطي في حالة أكثر من مجموعتين أما التمييز غير الخطي فيستخدم في حالة عدم تساوي التباينات

3-1-1 : الدالة المميزة Discriminant Function⁽⁴⁾

عبارة عن تقنية تستخدم لبناء نموذج للتنبؤ بتصنيف عضوية مشاهدة ما إلى مجموعة معينة بناء على عدة متغيرات ، يمكن للدالة المميزة تصنيف مجموعة من الأشخاص إلى مجموعتين رئيسيتين أشخاص (مصابين - غير مصابين) على أساس تباينات متغيرات متعددة . وتولد الدالة المميزة من عينة معروفة مسبقا انتماء أي مفردة أو مشاهدة فيها إلى إحدى المجموعات المراد التمييز بينها ، فتعمل الدالة على تصنيف مشاهدات أو مفردات جديدة معلومة المتغيرات مجهولة التصنيف ، وللدالة المميزة عدة أهداف منها:

- تصنيف المشاهدات ضمن مجموعات مختلفة
- التحقق من الفرق داخل المجموعات وبين المجموعات

- تحديد أبسط طريقة للتمييز بين المجموعات
 - إيجاد نسبة التباين في المتغيرات المستقلة في تصنيف المجموعات
 - إبعاد المتغيرات التي ليس لها تأثير في تصنيف المجموعات
- تستخدم الدالة المميزة الخطية عندما تكون المجتمعات المبحوثة ذات توزيع طبيعي متعدد المتغيرات بمتجهات متوسطات مختلفة ومصفوفات تباين وتباين مشترك متساوية .

3-1-2: الدالة المميزة الخطية في حالة وجود مجموعتين

(The Linear Discriminant Function –For Two Group)

إن المشكلة الإحصائية هنا تكمن في كيفية إيجاد دالة مميزة وفقا للمعايير أو القياسات التي يمكن الحصول عليها من الأفراد الجدد (مجهولي الانتماء) إلى المجموعة الصحيحة وسيتم هنا تناول نموذج التشخيص المميز الخطي بين مجموعتين .

وستكون خطوات حساب الدالة المميزة في حالة مجموعتين كالاتي :

أولا : إيجاد متوسط كل عامل (متغير مستقل) في كل مجموعة وكالاتي :

المجموعة الأولى:

$$\bar{X}_{1(1)} = \sum_{i=1}^n \frac{x_{1i(1)}}{n_{1(1)}} \dots \dots (3 - 1)$$

$$\bar{X}_{2(1)} = \sum_{i=1}^n \frac{x_{2i(1)}}{n_{2(1)}} \dots \dots (3 - 2)$$

$$\bar{X}_{n(1)} = \sum_{i=1}^n \frac{x_{ni(1)}}{n_{n(1)}} \dots \dots (3 - 3)$$

المجموعة الثانية:

$$\bar{X}_{1(2)} = \sum_{i=1}^n \frac{x_{1i(2)}}{n_{1(2)}} \dots \dots (3 - 4)$$

$$\bar{X}_{2(2)} = \sum_{i=1}^n \frac{x_{2i(2)}}{n_{2(2)}} \dots \dots (3 - 5)$$

$$\bar{X}_{n(2)} = \sum_{i=1}^n \frac{x_{ni(2)}}{n_{n(2)}} \dots \dots (3 - 6)$$

ثانيا: إيجاد الفرق (المسافة) بين متوسطي كل متغيرين في المجموعتين:

$$d_1 = \bar{x}_{1(1)} - \bar{x}_{1(2)} \dots \dots (3 - 7)$$

$$d_2 = \bar{x}_{2(1)} - \bar{x}_{2(2)} \dots \dots (3 - 8)$$

$$d_n = \bar{x}_{n(1)} - \bar{x}_{n(2)} \dots \dots (3 - 9)$$

ثالثا: إيجاد مصفوفة التباينات والتغايرات المشتركة :

بالنسبة للتباينات هي عبارة عن التباينات المجمع بين كل متغير في مجموعة ونفس المتغير في المجموعة الثانية :

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} \cdots & v_{1n} \\ v_{21} & v_{22} \cdots & v_{2n} \\ \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} \cdots & v_{nn} \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} \cdots & v_{1n} \\ v_{21} & v_{22} \cdots & v_{2n} \\ \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} \cdots & v_{nn} \end{pmatrix} \dots\dots\dots(3-10)$$

$$V_{pooled} = \frac{S^2_{i(1)} + S^2_{i(2)}}{n_{i(1)} + n_{i(2)} - 2} \dots\dots\dots (3 - 11)$$

$$S^2_i = \sum_{i=1}^n X^2_i - \frac{(\sum_{i=1}^n X_i)^2}{n} \dots\dots\dots (3 - 12)$$

أما بالنسبة للتغايرات المشتركة فيتم حسابها كالاتي :

$$V_{ij} = \frac{S^2_{ij(1)} + S^2_{ij(2)}}{n_{(1)} + n_{(2)} - 2} \dots\dots\dots (3 - 13)$$

$$s_{ij} = \sum_{i=1}^n x_i x_j - \frac{(\sum_{i=1}^n X_i \sum_{j=1}^n x_j)}{n} \dots\dots\dots (3 - 14)$$

رابعا : حساب الدالة المميزة وذلك كالاتي :

$$L = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n \dots\dots\dots(3-15)$$

حيث أن :

$$\alpha = \mathcal{V}^{-1} \mathcal{D} \dots\dots\dots (3 - 16)$$

وأن :

$$\mathcal{D} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \dots\dots\dots (3-17)$$

3-1-3: الدالة المميزة الخطية في حالة وجود أكثر من مجموعتين:

(The Linear Discriminant Function – More Than Two Group)

نفرض أن لدينا K من المجموعات ،و كل مجموعة لها n مشاهدة وكل مشاهدة تتضمن p من المتغيرات.

n_i هو حجم العينة المسحوبة من المجموعة i .

$$n = \sum_{i=1}^k n_i$$

نفرض أن T تمثل مصفوفة التباين والتغاير المشترك الكلي.

$$T = \sum_{t=1}^k \sum_{j=1}^n (X_{ij} - \bar{X})(X_{ij} - \bar{X})' \dots \dots (3-18)$$

$$T = \begin{bmatrix} S_{11T} & S_{12T} & \dots & S_{1pT} \\ S_{21T} & S_{22T} & \dots & S_{2pT} \\ \vdots & \vdots & & \vdots \\ S_{p1T} & S_{p2T} & \dots & S_{ppT} \end{bmatrix} \dots \dots (3-19)$$

ونفرض ان W_t تمثل مصفوفة التباين والتغاير المشترك للمجموعة i.

$$W_t = \sum_{i=1}^{n_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' \dots \dots (3-20)$$

وان مصفوفة التباين والتغاير المشترك داخل المجاميع تساوي W:

$$W = W_1 + W_2 + \dots + W_k \dots \dots (3-21)$$

$$W = \begin{bmatrix} S_{11W} & S_{12W} & \dots & S_{1pW} \\ S_{21W} & S_{22W} & \dots & S_{2pW} \\ \vdots & \vdots & & \vdots \\ S_{p1W} & S_{p2W} & \dots & S_{ppW} \end{bmatrix} \dots \dots (3-22)$$

وان مصفوفة التباين والتغاير المشترك بين المجاميع هي:

$$B = T - W \dots \dots (3-23)$$

$$B = \begin{bmatrix} S_{11B} & S_{12B} & \dots & S_{1PB} \\ S_{21B} & S_{22B} & \dots & S_{2PB} \\ \vdots & \vdots & & \vdots \\ S_{P1B} & S_{P2B} & \dots & S_{PPB} \end{bmatrix} \dots \dots (3 - 24)$$

الهدف هو إيجاد التراكيب الخطية والتي هي :

$$Y = [Y_1, Y_2, \dots, Y_r] \dots \dots (3 - 25)$$

والتي تعظم مقياس التمييز عن طريق تعظيم λ لكل b .

$$\lambda = \frac{\text{Between group}}{\text{With in group}} \dots \dots (3 - 26)$$

$$\lambda = \frac{b' B b}{b' W b} \dots \dots (3 - 27)$$

ولجعل λ اعظم مايمكن نأخذ المشتقة الجزئية بالنسبة ل b :

$$\frac{d\lambda}{db} = \frac{[2(b' W b * B b - b' B b * W b)]}{(b' W b)^2}$$

$$\frac{d\lambda}{db} = 0$$

$$(b' W b) B b - (b' B b) W b = 0$$

بقسمة الطرفين على $b' W b$ وبالتعويض عن λ بما يساويها نحصل على :

$$B b - \lambda W b = 0$$

$$(B - \lambda W) b = 0$$

$$(W^{-1} B - \lambda I) B = 0$$

نوجد قيم λ ، اكبر قيمة الى λ هي اكبر جذر مميز لمصفوفة $W^{-1} B$ والذي يقابل اكبر متجه

مميز b_1 .

$$b_1 = (b_{11}, b_{12}, \dots, b_{1p}) \dots \dots (3 - 28)$$

يمثل مقياس تمييز للدالة Y_1 والتي تساوي:

b_1 الأولى

$$Y_1 = b_{11}X_1 + b_{12}X_2 + \dots + b_{1p}X_p \dots \dots (3 - 29)$$

والذي يقابل ثاني اكبر متجه مميز وثاني اكبر جذر مميز لمصفوفة $W^{-1}B$ هو λ_2 والذي يقابل ثاني اكبر متجه مميز b_2 والذي يمثل مقياس التمييز للدالة الثانية والتي تساوي:

$$Y_2 = b_{21}X_1 + b_{22}X_2 + \dots + b_{2p}X_p \dots \dots (3 - 30)$$

من الضروري أن تكون Y_1 غير مرتبطة مع Y_2 .

Y_3 تمتلك ثالث اكبر متجه مميز .

$$Y_3 = b_{31}X_1 + b_{32}X_2 + \dots + b_{3p}X_p \dots \dots (3 - 31)$$

Y_3 غير مرتبطة مع Y_1 و Y_2 .

وهكذا نستمر إلى Y_r والتي تكون غير مرتبطة مع Y_1, Y_2, \dots, Y_{r-1} ويطلق على الدوال

(Y_1, Y_2, \dots, Y_r) الدوال الخطية المميزة والتي يمكن التعبير عنها في شكل مصفوفة :

$$\underline{y} = \underline{b}x$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \dots & b_{rp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \dots \dots (3-32)$$

يحدد r بعدد الدوال المميزة بالاعتماد على رتبة المصفوفة المركبة $W^{-1}B$.

ان رتبة المصفوفة $W_{p \times p}$ يساوي P وان رتبة $W^{-1}B = W$ وان رتبة مصفوفة B يكون اصغر

من $(P, K-1)$ وعادة يكون $(K-1)$ اصغر من (P) وبهذا تكون رتبة $W^{-1}B$:

$$\text{rank}(W^{-1}B) = \min(k-1, P)$$

اي يكون عدد الدوال المميزة ل k من المجموعات و P من المتغيرات هو :

$$\text{No of discriminant function} = \min(P, k-1) .$$

وعملية التصنيف تكون عن طريق تعويض قيم المتغيرات الخاصة بأي مشاهدة يراد تصنيفها في

جميع الدوال المميزة ، وتصنف إلى الدالة المميزة لأكبر مقدار .

3-1-4: اختبار معنوية الدالة الخطية المميزة في حالة وجود أكثر من مجموعتين:

عندما يراد التمييز بين أكثر من مجموعتين تكون إحصاءات الاختبار كالأتي :

1- اختبار ويلكس لأمبدا :

يحسب هذا المقياس وفقاً للصيغة التالية :

$$A = \frac{|W|}{|T|} \dots \dots (3 - 33)$$

حيث تمثل كل من :

T : مصفوفة التباين و التغاير المشترك الكلي للمجموعات .

W : مصفوفة التباين و التغاير المشترك داخل للمجموعات .

تتراوح قيمة A بين الصفر و الواحد ، فإذا كانت قريبة او مساوية للواحد فإن ذلك يشير الى ان متوسطات المجموعات متساوية لذلك لا يوجد تمييز بين المجموعات ، و هذا يعني ان الدالة المميزة المحسوبة غير قادرة على التمييز . أما إذا كانت قيمته قريبة من الصفر فإن ذلك يدل على قوة التمييز ، أي أن الدالة قادرة على التمييز بدرجة عالية .

2- مقياس χ^2 :

الفرضية المراد اختبارها هنا هي :

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$$

و تكون صيغته الرياضية كالأتي :

$$\chi^2 = -N \text{Log}(A)$$

يعد هذا المقياس أكثر دقة من مقياس (A) ، و يكون توزيعه مقارباً الى توزيع $\chi^2_{p(k-1),\alpha}$ و

قد طورت صيغته من قبل Barttlete .

و يأخذ الشكل الآتي :

$$\chi^2 = - \left[N - 1 - \frac{1}{2}(P + k) \right] \text{Log}(A) \dots \dots (3 - 34)$$

و تستخرج قيمة χ^2 الجدولية اعتماداً على درجة حرية و مستوى معنوية $\chi^2_{p(k-1),\alpha}$ بعد

ذلك تتم مقارنة قيمة مربع كاي المحسوبة مع القيمة الجدولية ، فإذا كانت القيمة المحسوبة اقل من

القيمة الجدولية يؤدي هذا الى قبول فرضية العدم ، اي انه لا يوجد اختلاف بين متوسطات المجموعات و في هذه الحالة فإنه لا يوجد تمييز بين المجموعات . اما اذا كانت القيمة المحسوبة اكبر من القيمة الجدولية يؤدي هذا إلى رفض فرضية العدم و قبول الفرضية البديلة و هذا يعني أن هنالك تمييز بين المجموعات و بدرجة عالية .

3- مقياس F :

الفرضية المراد اختبارها هنا هي :

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$$

و بالنظر لصعوبة الحصول على القيمة الجدولية لعدم توفر الجداول الخاصة بمعيار ولكس ، تم وضع معيار بديل آخر من قبل Rao ، و هذا المعيار هو مقياس F و صيغته الرياضية هي :

$$F_{Rao} = \frac{1 - A^{\frac{1}{5}}}{A^{\frac{1}{5}}} * \frac{ms - 2\lambda}{P(k - 1)} \dots \dots \dots (3 - 35)$$

بدرجات حرية $df_1 = P(k - 1)$ ، $df_2 = ms - 2\lambda$ اي :

$$F_{P(k-1),ms-2\lambda,\alpha}$$

حيث أن :

$$m = N - \frac{1}{2}(P + k) \dots \dots \dots (3 - 36)$$

$$S = \left[\frac{P^2(1-k)^2 - 4}{(1-k)^2 + P^2 - 5} \right]^{\frac{1}{2}} \dots \dots \dots (3 - 37)$$

$$\lambda = \frac{P(1 - k) - 2}{4} \dots \dots (3 - 38)$$

فإذا كانت :

$$F_{Rao} > F_{P(k-1),ms-2\lambda,\alpha}$$

ترفض فرضية العدم و تقبل الفرضية البديلة أي انه يوجد تمييز بين المجموعات .

3-2: الانحدار اللوجستي

3-2-1 : مفهوم نموذج الانحدار اللوجستي

يعرف النموذج اللوجستي على أنه أحد نماذج الانحدار التي تكون فيها العلاقة بين المتغير التابع (y)

والمتغير المستقل (x) غير خطية ، وغالباً ما تأخذ دالة الاستجابة الشكل (S) . كما يتميز النموذج اللوجستي في كونه أكثر مرونة من نماذج الانحدار التقليدية . حيث يمكن افتراض علاقة معينة تربط بين المتغير التابع أو المتغيرات المستقلة الأخرى ومن خلاله نستطيع مباشرة تقدير احتمال حدوث حدث ما . فضلاً عن سهولة تحويله إلى الشكل الخطي البسيط باستخدام ما يعرف بتحويل اللوجت .

3-2-2: النموذج اللوجستي الثنائي: ¹

أفرض أن (Y) متغير تابع يأخذ القيمة (واحد) إذا حدث حدث معين والقيمة (صفر) إذا لم يحدث أي عندما يكون للمتغير التابع قيمتين فقط وأن (x) متغير كمي أو نوعي يسمى النموذج بنموذج الانحدار اللوجستي الثنائي *Binary Logistic Regression Model* . إذا رسمنا المنحني الذي يمثل العلاقة بين (y) كمتغير تابع و(x) كمتغير مستقل فسندج إن الدالة ستأخذ شكل منحنى (*S-Shape curve*)

من أن قيم ($E(Y)$) تنحصر بين (الصفر والواحد) وبالتالي تتراكم بين هاتين القيمتين كما يعتمد شكل الدالة التزايد والتناقصي علي إشارة (β_j) والمتغير التابع في النموذج المقدر هو متغير برنولي يأخذ احدي القيمتين (1 و 0) .

دالة هذا المنحني هي الدالة اللوجستية لهذا فعندما يكون (Y) متغير ثنائي القيمة نفرض أن :

$$E (y / x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} \dots\dots\dots(3-39)$$

وتسمى المعادلة السابقة بالدالة اللوجستية

الثنائية أو النموذج اللوجستي الثنائي *Binary Logistic Regression Model* .

ومن خصائص الدالة اللوجستية أنه يمكن تحويلها لدالة خطية بتحويله مناسبة . فإذا وضعنا

$$E(y/x) = \pi(x) \dots\dots\dots(3-40)$$

و استخدمنا التحويله

¹ريتشارد جونسون ، دينوشن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية دار المريخ للنشر .

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) \dots\dots\dots(3-41)$$

$\pi(x)$ هي إحتمال وجود الظاهرة بينما)

$(1 - \pi(x))$ هي إحتمال لانعدام الظاهرة حيث نجد أن :

$$g(x) = \beta_0 + \beta_1 x \dots\dots\dots(3-42)$$

وتسمى $g(x)$ اللوجت *Logit* وترجع أهمية النموذج اللوجستي لكون المقدار (e^{β_1})

يعطى ما يسمى بنسبة الخطر *odd ratio* لتوضيح فكرة نسبة الخطر أفرض أن (Y) يأخذ القيمة (واحد) إذا كان الشخص مصاب بمرض معين و القيمة (صفر) إذا لم يكن مصاباً. كذلك أفرض إن (X) يأخذ القيمة (واحد) إذا كان الشخص تنطبق عليه صفة معينة والقيمة (صفر) إذا لم ينطبق عليه الصفة .

المخاطرة *odds* أن يكون الشخص مصاباً من الأشخاص الذين عندهم $X = 1$ أي تعرف كالأتي :

$$Odds(1) = \frac{\text{عدد الذين عندهم المرض وسط الذين يحملون الصفة}}{\text{عدد الذين ليس عندهم المرض وسط الذين يحملون الصفة}}$$

كما أن مخاطر *odds* أن يكون الشخص مصاباً بين الذين ليس عندهم الصفة أي $X = 0$ تعرف كالأتي :

$$Odds(0) = \frac{\text{عدد الذين عندهم المرض وسط الذين لا يحملون الصفة}}{\text{عدد الذين ليس عندهم المرض وسط الذين لا يحملون الصفة}}$$

وتعرف نسبة الخطر *odd ratio* بأنها النسبة بين الاثنتين أي :

$$OR = \frac{Odds(1)}{Odds(0)}$$

تمثل هذه النسبة مخاطر الإصابة بالمرض للأشخاص الذين عندهم الصفة مقارنة بتلك التي للأشخاص الذين ليس عندهم المرض. ويمكننا تحليل الانحدار اللوجستي مباشرة من معرفة نسبة الخطر.

3-2-3: النموذج اللوجستي المتعدد¹

إذا كان لدينا متغير تابع (Y) ثنائي يأخذ القيم (صفر وواحد) و P من المتغيرات المستقلة.

$$X = (X_1, X_2, X_3, \dots, X_p) \quad \dots\dots\dots(3-43)$$

يمكن تعميم النموذج في (1-2) ليأخذ الشكل :

$$E(y/x) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}} \quad \dots\dots\dots(3-44)$$

وفي هذه الحالة يسمى بالنموذج اللوجستي المتعدد كما في المعادلة (3-6) وتأخذ دالة اللوجت الشكل التالي :

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p \quad \dots\dots\dots(3-45)$$

إذا كان أحد المتغيرات المستقلة نوعي أو اسمي وله K من المستويات عندئذ سيكون له $K - 1$ من الحالات نرسم لها بالرمز D_{ji} ولدينا $i = 1, 2, 3, \dots, K-1$ حينها تصبح المعادلة :

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \sum_{i=1}^{Kj-1} \beta_{ji} D_{ji} + \beta_p X_p \quad \dots\dots\dots(3-46)$$

ومن النموذج (3-8) نجد أن قيمة واحدة فقط من D_{ji} تأخذ القيمة (واحد) وذلك عندما يتصف المتغير المستقل بصفة معينة ، بينما باقي القيم تأخذ القيمة (صفر) دلالة علي عدم وجود الصفة للمتغير المستقل.

3-2-4 : تقدير معاملات الانحدار اللوجستي المتعدد

(1)ريتشارد جونسون ، دينوشن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من وجهة التطبيقية - دار المريخ للنشر.

الطريقة التي تستخدم في تقدير معالم النموذج اللوجستي هي طريقة الترجيح الأعظم *Maximum Likelihood* وذلك لأن تباين حد الخطأ في النموذج اللوجستي غير ثابت وبالتالي باستخدام طريقة المربعات الصغرى التي غالباً ما تستخدم لتقدير معالم نموذج الانحدار الخطي تعتبر طريقة غير دقيقة لأنها قائمة علي فكرة ثابت تباين الخطأ حيث تتضمن هذه الطريقة تصغير مجموع المربعات لانحرافات النقاط الفعلية عن خط التوفيق إلي حد ممكن .

ولتوضيح تطبيق هذه الطريقة نلاحظ أولاً أن الشكل دالة الترجيح الأعظم للزوج المرتب X_i, y_i هو : (3-47)
 $\pi(x_i)^{y_i} [1 - \pi(x_i)^{1-y_i}] \dots \dots \dots$
 بافتراض أن المشاهدات مستقلة تكون الدالة لعينة حجمها n :

$$L(\beta) = \prod_{i=1}^n \left(\pi(x_i)^{y_i} [1 - \pi(x_i)^{1-y_i}] \right) \dots \dots \dots (3-48)$$

ونختار قيم β_j لكل $j = 1, 2, \dots, P$ التي تجعل معالم النموذج اللوجستي المتعدد نجد هنالك طرق أخرى للتقدير منها :

- طريقة المربعات الصغرى المرجحة غير التكرارية *Non- iteratively Weighted Least Squares*
- طريقة الدالة التمييزية *Discriminant Function*

3-2-5 : اختبار معنوية معاملات النموذج¹

من الطرق لتحديد ما إذا كانت المعاملات معنوية أم لا في الانحدار اللوجستي تتم من خلال اختبار نسبة الإمكان الأعظم ، وترتكز على إيجاد إمكانية حدوث نتائج العينة التي بين أيدينا إذا ما عرفنا تقديراً لمعاملات النموذج ، ويسمى احتمال حدوث النتائج المشاهدة إذا ما أعطينا القيم المقدرة لمعاملات النموذج بالإمكان (L) *(Likelihood)* ، حيث أن قيم الإمكان دائماً قيم صغيرة (أقل من الواحد) وبالتالي نستخدم (-2) مضروبة في $(-2 \ln L)$

(1) ريتشارد جونسون ، دينوشن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من وجهة التطبيقية - دار المريخ للنشر .

(كمقياس لمدى ملائمة (توافق) النموذج للبيانات ، فالنموذج الجيد هو ذلك الذي له قيمة إمكان كبيرة لنتائج المشاهدة .
 عليه فهو ذلك الذي له قيمة ل $(-2 \ln L)$ صغيرة (إذا كان النموذج مثالياً فإن قيمة الإمكان
 تكون مساوية للواحد وقيمة $(-2 \ln L)$ مساوية للصفر) .
 بفرض أن لدينا n مشاهدة لكل متغير فإن الانحراف (D) لنموذج معرف كالأتي :

$$D = -2 \log \left[\frac{\text{LikeLihood Of The Fitting model}}{\text{LikeLihood Of The Saturated model}} \right] \dots\dots(3-49)$$

حيث *Saturated model* (النموذج المشبع) هو ذلك النموذج الذي يحتوي على n من
 المعلمات بحيث أن هذا النموذج يتنبأ بالقيم الحقيقية للملاحظات مجموعة البيانات ، أي أنه لا يوجد
 خطأاً للتقدير، وبمعنى آخر القيم المنتبأ بها من خلال هذا النموذج تساوي القيم المشاهدة
 فعلاً في مجموعة البيانات .

الانحراف (D) يعطي طريقة لمقارنة دالة الإمكان للنموذج الذي تم تقديره أو (إحتمال الحصول على
 القيم المشاهدة للبيانات من خلال النموذج المقدر) مع النموذج المشبع .
 سوف نستخدم الاحصاء (D) في تقويم دالة المتغير أو المتغيرات المستقلة من خلال مقارنة
 الإحصاء (D) للنموذج الذي يتضمن المتغير أو المتغيرات المستقلة موضع التقويم مع قيمة
 الإحصاء (D) للنموذج بدون تلك المتغيرات ، بحيث يكون أي مقدار من التغير في قيمة (D) بين
 النموذجين عائداً لتضمن ذلك المتغير أو المتغيرات المستقلة في النموذج . وباستخدام الانحراف
 (D) الذي سبق وصفه يمكن اختبار الفرضيات :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{at least one of the above B's} \neq 0$$

حيث نستخدم إحصاء جديدة للفرق بين قيمتي (D) لأي نموذجين مختلفين وسيتم الرمز له بالرمز
 (G)

$G=D$ (for the model containing B_0 ,or the reduced model)- D (for the full model)

$$G = -2\ln \left[\frac{\text{LikeLihood Of The reduced model}}{\text{LikeLihood Of The Saturated model}} \right] + 2\ln \left[\frac{\text{LikeLihood Of The full model}}{\text{LikeLihood Of The Saturated model}} \right]$$

$$G = -2\ln \left[\frac{\text{LikeLihood Of The reduced model}}{\text{LikeLihood Of The full model}} \right]$$

$$G = -2\ln (\text{Likelihood of the reduced model}) + 2\ln (\text{Likelihood of the full model})$$

وبلاحظ أن الاحصاء G تتبع توزيع كاي بدرجة حرية تساوي عدد المعلمات المقدرة ناقص واحد (عدد المعلمات - 1). وفي الحالة التي يكون فيها النموذج يحتوي على متغير مستقل واحد فإنه يكون لدينا دالة الإمكان للنموذج مع المتغير المستقل ، ودالة الإمكان للنموذج بدون المتغير المستقل وحيث أن الاحتمال $y=0$ سوف يكون مساوياً لمتوسط عدد القيم التي تساوي صفر في العينة ، في حين أن الاحتمال $y=1$ سوف يكون مساوياً لمتوسط عدد القيم التي تساوي واحد في العينة لذلك فإن :

$$L(B_0) = Y^{-\sum_{i=1}^n y_i} (1-y)^{\sum_{i=1}^n (1-y_i)} \dots\dots\dots(3-50)$$

وبمعلومية دالة الإمكان الأعظم يمكن كتابة معادلة إحصاء الاختبار (G) كالآتي :

$$G = -2 \ln \left[\frac{\left[\frac{\sum_{i=1}^n y_i}{n} \right]^{\sum_{i=1}^n y_i} \left[\frac{\sum_{i=1}^n (1-y_i)}{n} \right]^{\sum_{i=1}^n (1-y_i)}}{\prod_{i=1}^n \hat{\pi} (1-\hat{\pi})^{(1-y_i)}} \right] \dots\dots\dots(3-51)$$

حيث G تتبع توزيع X^2 بدرجة حرية واحدة . ولاختبار معنوية (B_i) يكون فرضية العدم والفرضية البديلة كالآتي:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

ولهذا الاختبار يكون القرار في حالة $G > X^2(1 - \alpha; 1)$ ، وهو رفض H_0 وهذا يعني أن العامل (B_i) معنوي ويعني آخر إذا كان $G \leq X^2(1 - \alpha; 1)$ (فيجب علينا عدم رفض H_0) .
هنالك عدة اختبارات أخرى لاختبار جودة النموذج :

أولاً :اختبار *Lemshow - Hosmer* لجودة المطابقة :-¹

لتقييم تناسب النموذج المقدر اعتمد اختبار *Hosmer-Lemshow* لجودة المطابقة على تقسيم البيانات إلى مجموعات (g) على قيم الاحتمالات المتوقعة .

وقد أوصى *Hosmer-Lemshow* بأن $g=10$ بمعنى أن يتم توزيع مشاهدات العينة بعد ترتيبها تصاعدياً حسب القيم المتوقعة باحتمالات على عشرة مجموعات ، بحيث يكون عدد الحالات في كل مجموعة هو $(n/10)$ ، بحيث توضع في المجموعة الأولى الحالات ذات أقل قيم للاحتمالات المتوقعة وتوضع في المجموعة الأخيرة الحالات ذات القيم الأعلى للاحتمالات المتوقعة للحالات وهكذا بقية المجموعات بالترتيب فإنه يتم جمع القيم المشاهدة والمتوقعة للحالات وفقاً لقيمتي ($y=1, y=0$) وذلك لكل فئة من مجموعات التصنيف .

وبعد ذلك يتم حساب إحصاء *Hosmer - Lemshow* لجودة المطابقة والتي يرمز له بالرمز \hat{C} حيث :

$$\hat{C} = \sum_{k=1}^g \left[\frac{\left[\sum y_{ik} - n_k \bar{\pi}_k(x_i) \right]^2}{n_k \bar{\pi}_k(x_i) [1 - \bar{\pi}_k(x_i)]} \right] \dots \dots \dots (3-52)$$

حيث :

(1)ريتشارد جونسون ، دينوشن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية - دار المريخ للنشر.

$n_k \equiv$ العدد الكلي للمشاهدات في المجموعة k

$\bar{\pi}_k(x_i) \equiv$ متوسط الاحتمالات المتوقعة للمجموعة k

والإحصاء \hat{C} تتبع توزيع مربع كاي (χ^2) بدرجة حرية (g-2) ويكون القرار في حالة

$\hat{C} > \chi^2(1-\alpha; g-2)$ وهو رفض H_0 بينما إذا كان $\hat{C} \leq \chi^2(1-\alpha; g-2)$ فإنه نقبل

فرضية العدم حيث يكون شكل الفرضية كالآتي:

H_0 : النموذج مطابق للبيانات

H_1 : النموذج غير مطابق للبيانات

ويرى *Hosmer-Lemshow* بأن إيجابية إحصاءه \hat{C} لجودة المطابقة هي أنها تعطي قيمة واحدة بسيطة وقابلة للتفسير والتي يمكن استخدامها لتقويم جودة المطابقة ، كما أن جدول التحليل عند فحص التكرارات المشاهدة والمتوقعة في الفئات المختلفة ، قد يشير إلى بعض المناطق أو الفئات التي لا يعمل فيها النموذج بشكل مرضي مقارنة ببقية الفئات الأخرى ، أما السلبية الوحيدة فهي أن عملية تجميع الحالات في مجموعات تصنيفية تجعلنا نفقد بعض الانحرافات عن المطابقة والعائدة على بعض الحالات الفردية . وهناك افتراض مهم في اختبار *Hosmer - Lemshow* لجودة المطابقة باعتباره أحد تطبيقات اختبار مربع كاي وهو أن تكون العينة كبيرة بشكل كافي لكي تتيح لعدد التكرارات المتوقعة في جميع الخلايا العشرين من خلايا جدول التحليل أن تكون أكثر من خمسة تكرارات .

ثانياً : اختبار $Wald$ ¹:

يستخدم اختبار *Wald* لاختبار معامل الانحدار ($\beta_k = 0$) وأن إحصاء *Wald* في هذه

(1)ريتشارد جونسون ، دينوشن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من وجهة التطبيقية - دار المريخ للنشر.

الحالة تساوي النسبة بين معامل الانحدار و خطئه المعياري أي أن $W = \frac{\hat{\beta}_k}{S.E \hat{\beta}_k}$ وهذه

الإحصاءة تتبع التوزيع الطبيعي القياسي Z حيث يكون شكل الفرضية كالاتي:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

وبذلك يكون القرار إذا كان $|W| \leq Z(1-\alpha/2)$ فإنه نقبل فرضية العدم H_0 القائل بأن معامل الانحدار يساوي صفر ، وذلك يعني أن معامل الانحدار لذلك المتغير المستقل X ليس له تأثير في التنبؤ بقيمة y مما يعني إمكانية حذفه من النموذج لأن ليس له دلالة إحصائية ، بينما إذا كان $|W| > Z(1-\alpha/2)$ في هذه الحالة نرفض فرضية العدم ونقبل الفرضية البديلة ، وبذلك فإن ذلك المتغير المستقل سيكون له تأثير في التنبؤ بقيمة المتغير التابع y .

هذا الاختبار يمكن أن يكتب بأسلوب آخر ، لأن مربع التغير العشوائي الطبيعي سيؤدي إلى المتغير العشوائي χ^2 بدرجة حرية واحدة فإن إحصاءة Wald في هذه الحالة ستكون مساوية لمربع النسبة بين قيمة معامل الانحدار و خطئه المعياري إي أن:

$$W^2 > \chi^2(1-\alpha;1)$$

حيث :

$$W^2 = \left[\frac{\hat{\beta}_k}{S.E \hat{\beta}_k} \right]^2 \dots\dots\dots(3-53)$$

بموجب هذا التغير يكون القرار إذا كانت $W^2 > \chi^2(1-\alpha;1)$ فإنه نرفض فرضية العدم القائل أن معامل الانحدار يساوي صفر ، بذلك فإن ذلك المتغير المستقل ليس له تأثير في التنبؤ بقيمة المتغير التابع y .

يجب ملاحظة أن إحصاءة Wald تعاني قصور شديداً إذا ما كانت القيمة المطابقة لمعامل الانحدار كبيرة وبالتالي تكون قيمة الخطأ المعياري له كبيرة جداً مما ينتج عنه قيمة صغيرة جداً للإحصاءة ،

الأمر الذي يؤدي إلى عدم رفض فرضية العدم في حين ينبغي رفض هذه الفرضية وهذا يؤدي إلى ارتكاب خطأ من النوع الثاني .

3-2-6 : تقدير فترة الثقة ¹

يمكن إنشاء فترة للمعلمة ($\hat{\beta}_j$) حيث تأخذ الفترة بدرجة ثقة الشكل :

$$\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \hat{SE} \left(\hat{\beta}_j \right) \dots \dots \dots (3-54)$$

ولثابت الانحدار :

$$\hat{B}_0 \pm Z_{1-\frac{\alpha}{2}} \hat{SE} \left(\hat{\beta}_0 \right) \dots \dots \dots (3-55)$$

كذلك فإن فترة الثقة لنسبة الخطر للمتغير Z تكون :

$$\exp \left[\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \hat{SE} \left(\hat{\beta}_j \right) \right] \dots \dots \dots (3-56)$$

(1)ريتشارد جونسون ، دينوشن -تعريب عبد الرحمن حامد عزام -التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية - دار المريخ للنشر .

0-4: تمهيد

في هذا الفصل سوف يتم تطبيق نوعين من التحليل هما التحليل المميز والانحدار اللوجستي بحيث يتم التعرف علي مجتمع الدراسة والعينة التي تم اختيارها وكيفية اختيار العينة من المجتمع (أي استخدام حجم العينة) لكي يتم الحصول علي نتائج ذات درجة عالية من الدقة يمكن تعميمها علي المجتمع ، وقبل الشروع في عملية التحليل يتم وصف مفردات العينة حسب المتغيرات المستخدمة في الدراسة . وبعد ذلك يتم الشروع في تطبيق التحليل المميز وتحليل الانحدار اللوجستي حسب الخطوات التي تم ذكرها في الفصل السابق بالتفصيل .

وننوه هنا إلي أنه سوف يستخدم برنامج التحليل الإحصائي (spss) .

1-4: عينة البحث

يعتبر مجتمع البحث هو المجموعة الكلية من العناصر التي يسعى الباحث ان يعمم عليها النتائج ذات العلاقة بمشكلة البحث ويمثل مجتمع البحث المرضى المصابين بسرطان الدم (مزمن ، حاد) والذين لهم سجلات بمستشفى الذرة.

عينة البحث هي عينة عشوائية بسيطة تتضمن نوعين من المتغيرات الثنائية والمتصلة لمجموعتين مختلفة من مرض سرطان الدم والتي تمثل المتغير المعتمد .

تتكون عينة البحث من 102 مريض ، وسحبت هذه العينة بطريقة عشوائية.

2-4 : المتغيرات المستخدمة في الدراسة

تم الاعتماد على عدة متغيرات جمعت من كل مفردة (مريض) من مفردات العينة ، و

لقد تم تقسيم متغيرات الدراسة إلى نوعين :

- المتغير المعتمد و الذي يمثل نوع المرض (سرطان الدم الحاد CLL) ،(سرطان الدم المزمن CML) .
- المتغيرات المستقلة و هي موضحة في الجدول الآتي :

جدول رقم (4-1) يبين المتغيرات المستخدمة في الدراسة

المتغير	ما يمثله المتغير
X_1	النوع
X_2	العمر
X_3	الحالة الاجتماعية
X_4	الوظيفة
X_5	كريات الدم البيضاء (WBC)
X_6	كريات الدم الحمراء (RBC)

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

3-4: وصف متغيرات عينة الدراسة

في هذا المبحث سوف يتم وصف متغيرات الدراسة .

1-3-4 : وصف متغيرات عينة الدراسة حسب المرض

كان توزيع مفردات العينة حسب المجموعات ونسبة كل مجموعة علي النحو التالي:

جدول رقم (4-2) يبين توزيع مفردات العينة حسب المرض بنوعيه

المجموعات	أسماء المجموعات	عدد الحالات	الاحتمال	النسبة
1	سرطان الدم الحاد	69	0.68	%68
2	سرطان الدم المزمن	33	0.32	%32
المجموع		102	1.00	%100

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

شكل رقم (4-1) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين



المصدر : إعداد الباحث باستخدام برنامج (Excel)، 2016

من الجدول رقم (4-2) والشكل رقم (4-1) يتبين أن غالبية أفراد عينة الدراسة كان نوع المرض لديهم CLL حيث بلغ عددهم (69) فردا وبنسبة (68%)، فيما بلغ عدد اللذين نوع المرض لديهم CML (33) فردا وبنسبة (32%) .

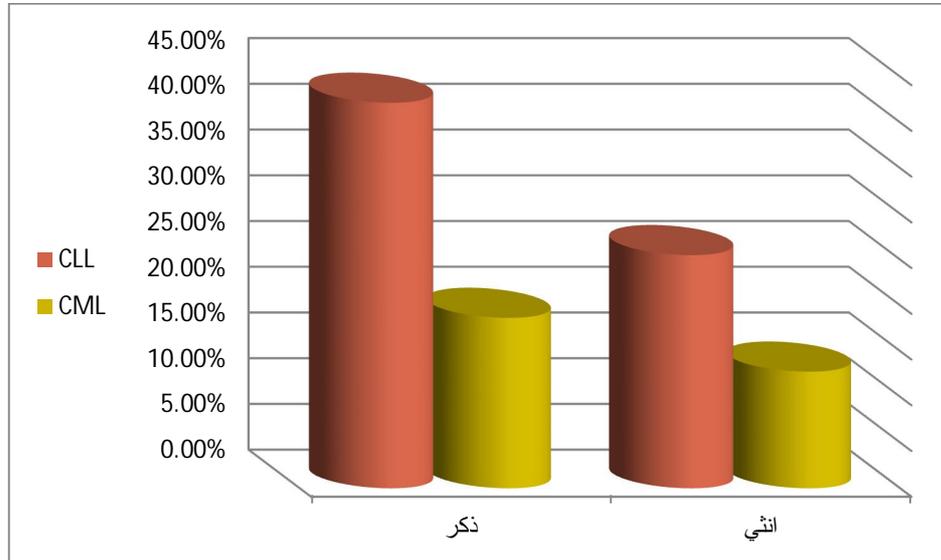
4-3-2: وصف متغيرات الدراسة حسب النوع

جدول رقم (4-3) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير النوع

نوع المرض				النوع
سرطان الدم المزمن		سرطان الدم الحاد		
النسبة	التكرار	النسبة	التكرار	
18.63%	19	42.15%	43	ذكر
12.75%	13	25.49%	26	أنثى
32.4%	32	67.6%	69	المجموع

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

شكل رقم (2-4) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير النوع



المصدر : إعداد الباحث باستخدام برنامج (Excel)، 2016،

من الجدول رقم (3-4) والشكل رقم (2-4) يتبين أن غالبية أفراد عينة الدراسة مصابين بسرطان الدم الحاد حيث بلغ عدد الإناث (26) فردا من العدد الكلي لأفراد عينة الدراسة بنسبة (25.49%) وبلغ عدد الذكور (34) فردا من العدد الكلي لأفراد عينة الدراسة بنسبة (42.14%)، وبلغ عدد الإناث المصابين بسرطان الدم المزمن (13) فردا من العدد الكلي للعينة بنسبة (12.75%) وبلغ عدد الذكور (19) فردا من العينة بنسبة (18.63%)

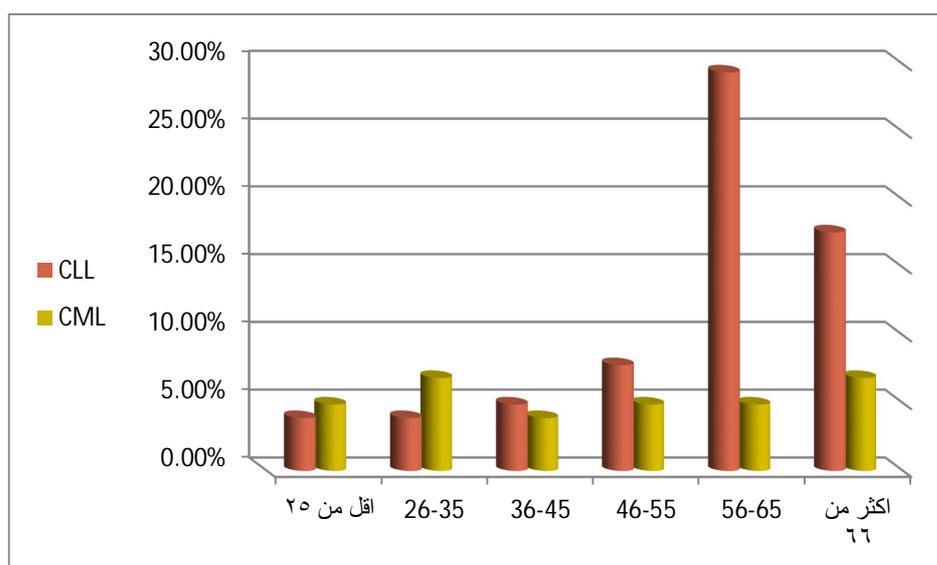
4-3-3: وصف متغيرات عينة الدراسة حسب متغير العمر

جدول رقم (4-4) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير العمر

نوع المرض				الفئات العمرية
سرطان الدم المزمن		سرطان الدم الحاد		
النسبة	التكرار	نسبة	التكرار	
4.9%	5	3.9%	4	اقل من 25 سنة
6.86%	7	3.9%	4	26-35 سنة
3.9%	4	4.9%	5	36-45 سنة
4.9%	5	7.8%	8	46-55 سنة
4.9%	5	29.4%	30	56-65 سنة
6.86%	7	17.6%	18	66 سنة فأكثر
32.4%	33	67.6%	69	المجموع

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

شكل رقم (4-3) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير العمر



المصدر : إعداد الباحث باستخدام برنامج (Excel) ، 2016،

من الجدول رقم (4-4) والشكل رقم (4-3) يتبين أن اعلي نسبة للإصابة بسرطان الدم الحاد تكون في الفئة العمرية (56-65) حيث بلغ عددهم (30) فردا بنسبة (29.4%)، و اقل نسبة تكون في الفئة العمرية (اقل من 25) و (26-35) حيث بلغ عددهم (4) فردا بنسبة (3.9%)، وان اعلي نسبة للإصابة بسرطان الدم المزمن تكون في الفئة العمرية (أكثر من 66) و (26-35) حيث بلغ عددهم (7) بنسبة (6.67%)، وان اقل نسبة تكون في الفئة العمرية (36-45) حيث بلغ عددهم (4) بنسبة (3.9%).

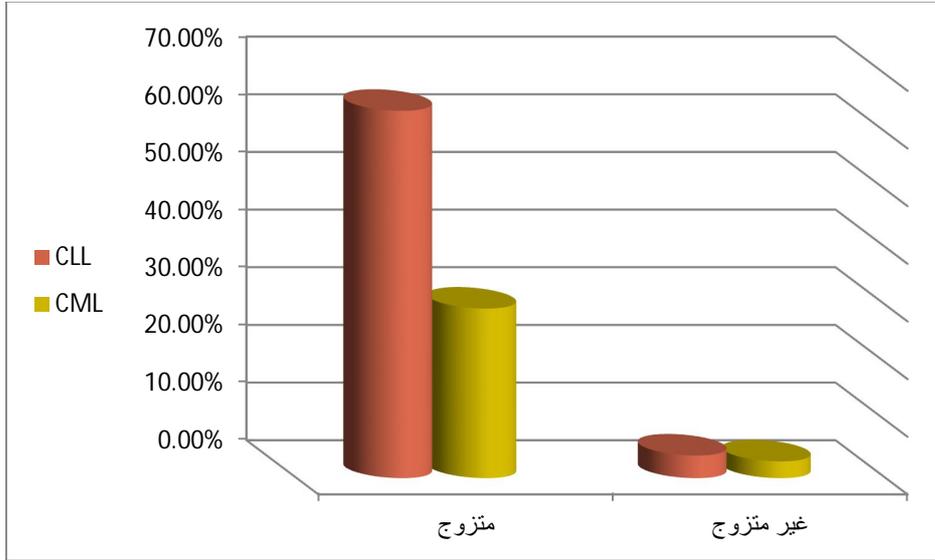
4-3-4: وصف متغيرات عينة الدراسة حسب متغير الحالة الاجتماعية

جدول رقم (4-5) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير الحالة الاجتماعية

نوع المرض				النوع
سرطان الدم المزمن		سرطان الدم الحاد		
النسبة	التكرار	النسبة	التكرار	
29.4%	30	63.7%	65	متزوج
2.9%	3	3.92%	4	غير متزوج
32.4%	33	67.6%	69	المجموع

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

الشكل رقم (4-4) يبين التوزيع التكراري لإفراد عينة الدراسة للمجموعتين حسب متغير الحالة الاجتماعية



المصدر : إعداد الباحث باستخدام برنامج (Excel)، 2016

من الجدول رقم (4-5) والشكل رقم (4-4) يتبين أن غالبية أفراد عينة الدراسة المتزوجين كان نوع المرض لديهم سرطان دم حاد حيث بلغ عددهم (65) فردا بنسبة (63.7%)، فيما بلغ عدد اللذين نوع المرض لديهم سرطان دم مزمن (30) فردا بنسبة (29.4%)، أما أفراد عينة الدراسة غير المتزوجين الذين كان نوع المرض لديهم سرطان دم حاد بلغ عددهم (4) فردا بنسبة (3.92%)، فيما بلغ عدد اللذين نوع المرض لديهم سرطان دم مزمن (3) فردا بنسبة (2.9%).

5-3-4: وصف متغيرات عينة الدراسة حسب متغير الوظيفة

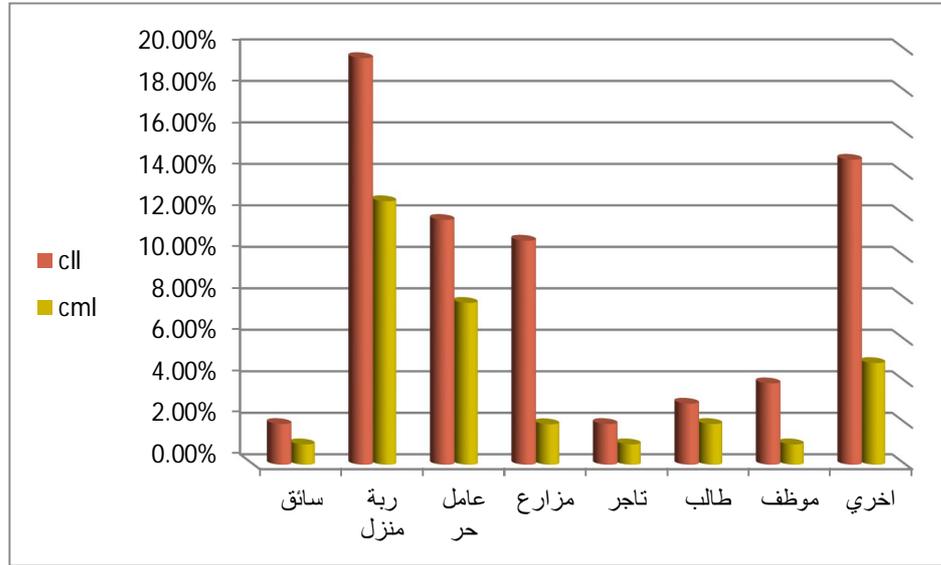
جدول رقم (4-6) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير الوظيفة

نوع المرض				الوظيفة
سرطان الدم المزمن		سرطان الدم الحاد		
النسبة	التكرار	النسبة	التكرار	
0.98%	1	1.96%	2	سائق
12.7%	13	19.6%	20	ربة منزل
7.8%	8	11.8%	12	عامل حر

1.96%	2	10.8%	11	مزارع
0.98%	1	1.96%	2	تاجر
1.96%	2	2.94 %	3	طالب
0.98%	1	3.92 %	4	موظف
4.90%	5	14.70%	15	اخرى
32.4%	33	67.6%	69	مجموع

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

الشكل رقم (4-5) يبين التوزيع التكراري لإفراد عينة الدراسة للمجموعتين حسب متغير الوظيفة



المصدر : إعداد الباحث باستخدام برنامج (Excel)، 2016

من الجدول رقم (4-6) والشكل رقم (4-5) يتبين أن غالبية أفراد عينة الدراسة الذين كانت وظيفتهم ربة منزل لديهم نسبة عالية بالإصابة بمرض سرطان الدم الحاد حيث بلغ عددهم (20) فردا بنسبة (19.6%)، وكذلك سرطان الدم المزمن حيث بلغ عددهم (13) فردا بنسبة (12.7%)، أما أفراد عينة الدراسة الذين كانت وظيفتهم تاجر هم أقل عرضة للإصابة بسرطان الدم الحاد حيث بلغ عددهم (2) فردا بنسبة (1.96%)، وكذلك سرطان الدم المزمن حيث بلغ عددهم (1) فردا بنسبة (0.98%).

4-3-6: وصف متغيرات عينة الدراسة حسب متغير كريات الدم البيضاء

جدول رقم (4-7) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير كريات

الدم البيضاء

المفردات	كريات الدم البيضاء
اقل قيمة	800.0
اعلي قيمة	448000.0
المتوسط	62548.300
الانحراف المعياري	90912.1096

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)

من الجدول رقم (4-7) بلغ متوسط كريات الدم البيضاء (62548.300) وانحراف معياري (90912.1096) وكما أن اعلي قيمة (448000.0) واقل قيمة (800.0).

7-3-4: وصف متغيرات عينة الدراسة حسب متغير كريات الدم الحمراء

جدول رقم (4-8) يبين التوزيع التكراري لأفراد عينة الدراسة للمجموعتين حسب متغير كريات الدم الحمراء

المفردات	كريات الدم الحمراء
اقل قيمة	16.0
اعلي قيمة	36300000.0
المتوسط	3950858.000
الانحراف المعياري	3.3919E6

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

من الجدول رقم (4-8) بلغ متوسط كريات الدم البيضاء (3950858.000) وانحراف معياري (3.3919E6) وكما أن اعلي قيمة (Maximum) واقل قيمة (16.0).

4-4: التحليل التمييزي

يتم في هذا الجزء من التحليل تحديد معنوية المتغيرات المستقلة ، اختبار معنوية الدوال المميزة الخطية عن طريق مقياس (ويلكس لامبدا) ، و أخيراً حساب احتمالات التصنيف الصحيح.

3-4-1: استخراج المؤشرات الأولية أي المتوسطات الحسابية والانحرافات المعيارية

جدول (4-9) يبين المتوسطات الحسابية لأفراد عينة الدراسة لمجموعتان

نوع المرض		متوسطات المتغيرات
سرطان الدم المزمن	سرطان الدم الحاد	
1.70	1.38	\bar{X}_1
3.58	4.61	\bar{X}_2
1.09	1.06	\bar{X}_3
3.34	4.32	\bar{X}_4
52756.006	67231.570	\bar{X}_5
4452197.455	3711086.957	\bar{X}_6

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

جدول (4-10) يبين الانحرافات المعيارية لأفراد عينة الدراسة لمجموعتان

نوع المرض		الانحرافات المعيارية
سرطان الدم المزمن	سرطان الدم الحاد	

1.741	0.488	S_1
1.803	1.396	S_2
0.92	0.235	S_3
2.268	2.404	S_4
87963.6121	92552.4780	S_5
5.8614E6	860303.2430	S_6

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

4-4-1: تحديد معنوية المتغيرات المستقلة

نعتمد في تحديد معنوية المتغيرات ذات التأثير المعنوي على مقياس ويلكس لامبدا و تقريب (F) ، حيث يعتبر أفضل متغير ذو تأثير معنوي الذي يحتوي على اكبر قيمة لـ (F) و اقل قيمة لـ (ويلكس لامبدا) .

جدول (4-11) يوضح قيم ويلكس لامبدا و قيم F :

Tests Of Equality Of Group Means

Variables	Wilks' Lambda	F
X_1	0.980	2.022
X_2	0.912	9.700
X_3	0.996	0.373
X_4	0.986	1.401
X_5	0.994	0.564
X_6	0.989	1.066

المصدر : إعداد الباحث باستخدام spss16.0، 2016

من الجدول رقم (4-11) نلاحظ إن المتغير (X_2) و الذي يمثل العمر هو أهم المتغيرات لأنه يحتوي على أكبر قيمة (F) و أقل قيمة (ويلكس لامبدا) و هما (9.700) و (0.912) على التوالي ، يليه في الأهمية المتغير (X_1) و الذي يمثل النوع حيث بلغت قيمة (F) (2.022) و بلغت قيمة (ويلكس لامبدا) (0.980) و هكذا ، حيث نلاحظ أن المتغير (X_3) (الحالة الاجتماعية) ليس له أي تأثير لأن قيمة (F) له أقل ما يمكن و كذلك قيمة (ويلكس لامبدا) أكبر ما يمكن .

2-4-4: تكوين الدوال المميزة الخطية

بعد تحديد المتغيرات ذات التأثير المعنوي واستخراج المصفوفات ،تستخرج مصفوفة المعاملات.

(12-4) يوضح معاملات الدوال المميزة الخطية

المتغيرات	دالة
	1
X1	-0.220
X2	0.976
X3	0.239
X4	0.331
X5	0.022
X6	-0.305

المصدر : إعداد الباحث باستخدام برنامج (spss16.0) 2016

وبعد ذلك نقوم بكتابة الدالة المميزة الخطية كالآتي :

$$Y_1 = -0.220X_1 + 0.976X_2 + 0.239X_3 + 0.331X_4 + 0.022X_5 - 0.305X_6$$

3-4-4:تكوين دوال التصنيف

نستخرج معاملات دوال التصنيف وكذلك الثابت وتكون النتائج كالآتي :

جدول (4-13) يوضح معاملات دوال التصنيف الخطية

نوع المرض		المتغيرات
سرطان الدم المزمن	سرطان الدم الحاد	
2.500	2.328	X1
3.922	4.447	X2
27.697	28.477	X3
.842	.958	X4
4.615E-9	2.043E-7	X5
3.081E-7	2.333E-7	X6
-27.188	-30.085	الثابت

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

وبعد استخراج معاملات دوال التصنيف نقوم بكتابة الدوال التصنيفية الاثنان وكالاتي :

$$Y'_1 = -30.085 + 2.328X_1 + 4.447X_2 + 28.477X_3 + 0.958X_4 + 2.043E - 7X_5 + 2.333E - 7X_6$$

$$Y'_1 = -27.188 + 2.500X_1 + 3.922X_2 + 27.697X_3 + 0.842X_4 + 4.615E - 9X_5 + 3.081E - 7X_6$$

وبعد استخراج الدوال أعلاه نستخرج الاحتمالات اللاحقة حيث يكون لكل قيمة يراد تصنيفها (2) احتمالات لاحقة وكل قيمة توضح إحتمال وقوع المفردة في المجموعة المقابلة ، حيث تصنف للمجموعة التي تقابل اكبر قيمة إحتمال لاحق وفي هذه الطريقة نحتاج أيضا إلي استخراج مقادير التمييز بالتعويض في الدوال أعلاه.

4-4-4 : حساب احتمالات التصنيف الصحيح

أما بخصوص التصنيف سوف يتم توضيح النسب المئوية للتصنيف الصحيح ، وقد جاءت بالشكل التالي :

جدول (4-14) يوضح نتائج واحتمالات التصنيف

نوع المرض	سرطان الدم الحاد	سرطان الدم المزمن
-----------	------------------	-------------------

النسبة	العدد	النسبة	العدد	
15.9%	11	84.1%	58	سرطان الدم الحاد
54.5%	18	45.5%	15	سرطان الدم المزمن

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016

ومن الجدول السابق يتبين أن نتائج التصنيف كانت كالآتي :

- تم تصنيف (58) من الأشخاص المصابين بسرطان الدم الحاد بشكل صحيح حيث بلغت نسبة التصنيف الصحيح (84.1%)
- تم تصنيف (18) من الأشخاص المصابين بسرطان الدم المزمن بشكل صحيح حيث بلغت نسبة التصنيف الصحيح (54.5%)
- كما بلغت نسبة التصنيف الصحيحة الكلية في المجموعتان (74.5%) ،حيث تم تصنيف (76) فرد بشكل صحيح

5-4: الانحدار اللوجستي

تم في هذا الجزء من التحليل تقدير نموذج الانحدار اللوجستي بطريقة (stepwise) للبيانات كما تم اختبار مدي مطابقة وجودة النموذج للبيانات و تم اختبار معنوية المتغيرات المستقلة و حساب نسبة الخطر في المتغيرات المستقلة و حساب مصفوفة التصنيف و حساب التنبؤات لاستخدامها لاحقاً.

1-5-4: تقدير نموذج الانحدار اللوجستي

نموذج الانحدار اللوجستي بطريقة (Enter) تقوم علي إدخال جميع المتغيرات ولا تعتمد علي حسب معنويتها في النموذج ومن ثم إعطاء النموذج الممكن .

الجدول (15-4) يوضح اختبار مقدرات النموذج

variable	B	S.E.	Wald	df	Sig.	Exp(B)
X1	.152	.238	.408	1	.523	1.164
X2	-.481	.166	8.419	1	.004	.618
X3	-.656	.947	.480	1	.488	.519
X4	-.124	.108	1.324	1	.250	.883
X5	.000	.000	.013	1	.909	1.000
X6	.000	.000	1.007	1	.316	1.000
Constant	1.975	1.624	1.477	1	.224	7.203

المصدر : إعداد الباحث باستخدام برنلمج (sps16.0)، 2016،

الجدول (15-4) يوضح المتغيرات التي تم إدخالها للنموذج و المقدرات الخاصة بها كما يوضح أن المتغيرات (X1، X3، X4، X5، X6) غير معنوية وذلك لان قيم ال Sig كانت أكبر من 0.05 وان المتغير X2 معنوي لان قيمة ال Sig اقل من 0.05.

$$Y = 1.975 + 0.152X_1 - 0.481X_2 - 0.656X_3 - 0.124X_4$$

2-5-4: توفيق النموذج

تم تطبيق اختبار (Hosmer and Lemeshow) لمطابقة البيانات للنموذج في هذا الاختبار تنص فرضية العدم علي ان النموذج مطابق للبيانات و الفرضية البديلة علي ان النموذج غير مطابق للبيانات .

الجدول (16-4) يوضح نتائج اختبار Hosmer and Lemesh

Step	Chi-square	df	Sig.
1	10.645	8	.223

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016،

من الجدول (4-16) القيمة الاحتمالية لمربع كاي تساوي (10.645) وهي أكبر من مستوى المعنوية (0.05). وبالتالي نقبل فرضية عدم الفائلة بأن التكرارات المتوقعة من النموذج لا تختلف عن الفعلية أي أن النموذج جيد أو مطابق للبيانات

4-5-3: حساب احتمالات التصنيف

الجدول (4-17) خطأ التصنيف

سرطان الدم المزمن		سرطان الدم الحاد		نوع المرض
النسبة	العدد	النسبة	العدد	
10.1%	7	89.9%	62	سرطان الدم الحاد
27.3%	9	72.7%	24	سرطان الدم المزمن

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)، 2016،

ومن الجدول أعلاه يتبين أن نتائج التصنيف كانت كالاتي :

- تم تصنيف (62) من الأشخاص المصابين بسرطان الدم الحاد بشكل صحيح حيث بلغت نسبة التصنيف الصحيح (89.9%)
- تم تصنيف (9) من الأشخاص المصابين بسرطان الدم المزمن بشكل صحيح حيث بلغت نسبة التصنيف الصحيح (27.3%)
- كما بلغت نسبة التصنيف الصحيحة الكلية في المجموعتان (69.6%) ،حيث تم تصنيف (71) فرد بشكل صحيح

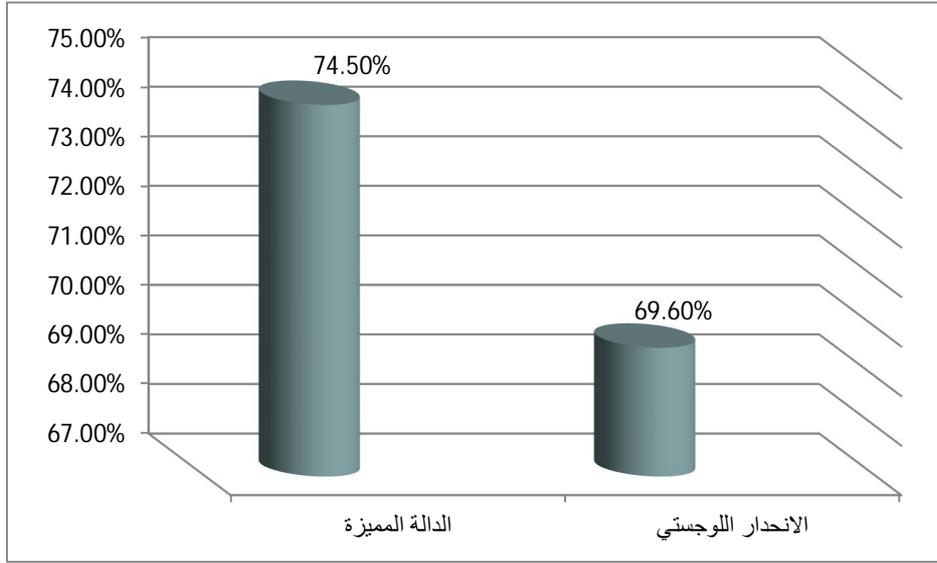
4-6 المقارنة :

من الملاحظ ان هنالك تفاوت بين الطرق في التصنيف لذلك تمت المقارنة بين الطرق الوارد ذكرها لاختيار الطريقة الافضل في التصنيف ، الجدول(4-18) يوضح نسب التصنيف الصحيحة الكلية للطريقتان.

الجدول(4-18) مقارنة بين طرق التصنيف

النسبة الكلية الصحيحة	الطريقة
74.5%	الدالة المميزة
69.6%	الانحدار اللوجستي

المصدر : إعداد الباحث باستخدام برنامج (spss16.0)



المصدر : إعداد الباحث باستخدام برنامج (Excel)، 2016

يتضح من الجدول (4-18) و الشكل (4-6) أن نسبة تصنيف الدالة المميزة اعلي من نسبة تصنيف الانحدار اللوجستي والتي بلغت (74.5%) .

5-1: النتائج:

بالاعتماد علي الجانب النظري والتطبيقي في هذا البحث تم التوصل للنتائج التالية :

1. من الجدول رقم (4-11) أن متغير العمر أهم متغير في تحديد التأثير المعنوي من بين المتغيرات الكلية ونلاحظ أن المتغير الأقل تأثير هو x_3 والذي يمثل الحالة الاجتماعية .
2. أن قدرة نموذج الدالة المميزة على تصنيف الأشخاص تصنيفاً صحيحاً تصل إلى (74.5%) بنسبة خطأ (25.5%).
3. أن قدرة نموذج الانحدار اللوجستي على تصنيف الأشخاص تصنيفاً صحيحاً تصل إلى (69.6%) بنسبة خطأ (30.4%).
4. إن قدرة الدالة المميزة في التصنيف المرض أفضل من طريقة الانحدار اللوجستي .

5-2: التوصيات:

من خلال نتائج البحث فإن هنالك بعض التوصيات التي يجب الأخذ بها للوصول إلي تصنيف أفضل في دراسة العوامل المؤثرة علي الإصابة بمرض سرطان الدم

1. اعتماد الدالة المميزة كأداة دقيقة للتصنيف و عمل حزم متخصصة فيها .
2. تطبيق الدالة المميزة التي تم التوصل إليها و استخدامها في التصنيف حتى يتمكن الطبيب من المعرفة المبكرة بالإصابة قبل أن يتطور و يتحول إلي مرحلة متقدمة .
3. آثار التصنيف الخاطئ إن و جدت يجب أن تؤخذ في الحسبان لان تصنيف فرد في مرض معين غير المرض المصاب به سوف تكون له آثار في عملية العلاج .
4. زيادة دقة التصنيف بإدخال متغيرات أخرى يُعتقد إنها تؤثر في الإصابة بمرض سرطان الدم وذلك بعد أخذ آراء الأطباء المتخصصين في المرض .
5. التثقيف و الإرشاد المستمر حتى يتمكن المرضي من معرفة كل العوامل التي تؤثر عليهم .
6. ضرورة وجود قاعدة بيانات شاملة لجميع المتغيرات الصحية المتعلقة بالإصابة بمرض سرطان الدم و تكون في متناول الباحثين علي أن يتم تحديث البيانات بصورة مستمرة .

المراجع والمصادر

أولاً : المراجع العربية

1. ريتشارد جونسون ، دينوشرن - تعريب عبد الرحمن حامد عزام - التحليل الإحصائي للمتغيرات المتعددة من الواجهة التطبيقية - دار المريخ للنشر.
2. أسامة الربيع أمين - التحليل الإحصائي باستخدام برنامج ال SPSS - كلية التجارة جامعة المنوفية - الطبعة الثانية 2007.
3. عثمان ، هالة فضل ، (2009) "استخدام الدالة التمييزية في تمييز الإصابة بمشكلات في القدمين لدي مرضي السكري " ، بحث تكميلي لنيل درجة الماجستير في الإحصاء التطبيقي بكلية العلوم _ جامعة السودان للعلوم والتكنولوجيا .
4. عبد اللطيف ، عفراء هاشم ، (2011) "تطبيق التحليل التمييزي لتصنيف بسرطان الثدي لدي الإناث حسب مراحلته المتعددة " رسالة دكتوراه بكلية العلوم _ جامعة السودان للعلوم والتكنولوجيا .
5. تقرير طبي من المركز القومي للعلاج بالأشعة والطب النووي (الذرة) بالخرطوم .

ثانيا: المراجع الانجليزية

1-Andreson J.A,(1969) , "Discrimination between K Population with Constraints on the Probabilities of Misclassification" ,oyal Statistic , Soc ,Vo1.31 ,No.1.

الملاحق البيانات المستخدمة في الدراسة

نوع المرض	النوع	العمر	الحالة الاجتماعية	الوظيفة	كريات الدم البيضاء	كريات الدم الحمراء
CLL	أنثي	56-65	متزوج	ربة منزل	5300.0	4890000.0
CLL	أنثي	56_65	متزوج	ربه منزل	6300.0	4890000.0
CML	ذكر	26_35	متزوج	عامل حر	16500.0	3170000.0
CML	ذكر	36-45	متزوج	مزارع	246400.0	3620000.0
CML	أنثي	66 فأكثر	متزوج	ربه منزل	7000.0	4090000.0
CLL	أنثي	66 فأكثر	متزوج	ربه منزل	117000.0	3766250.0
CLL	أنثي	36-45	متزوج	أخري	28000.0	2930000.0
CLL	ذكر	66 فأكثر	متزوج	سائق	30800.0	3766250.0
CLL	أنثي	56_65	متزوج	أخري	15000.0	1980000.0
CLL	ذكر	46-55	متزوج	تاجر	21500.0	3090000.0
CLL	أنثي	56_65	متزوج	ربه منزل	3300.0	4710000.0
CLL	ذكر	56_65	متزوج	مزارع	46400.0	2770000.0
CLL	ذكر	66 فأكثر	متزوج	مزارع	131200.0	2400000.0
CML	ذكر	46-55	متزوج	تاجر	25200.0	3380000.0
CML	أنثي	26_35	متزوج	أخري	47000.0	3766250.0
CLL	ذكر	46-55	متزوج	مزارع	44600.0	3580000.0
CLL	ذكر	56_65	متزوج	عامل حر	192000.0	4880000.0
CLL	ذكر	56_65	متزوج	مزارع	87200.0	3290000.0
CLL	أنثي	56_65	متزوج	عامل حر	38900.0	2970000.0
CLL	أنثي	56_65	متزوج	ربه منزل	14700.0	2400000.0
CML	أنثي	26_35	متزوج	ربه منزل	6000.0	2820000.0
CLL	ذكر	56_65	متزوج	أخري	102500.0	2970000.0
CLL	ذكر	66 فأكثر	متزوج	عامل حر	31000.0	3400000.0
CLL	أنثي	اقل من 25	غير متزوج	طالب	3500.0	4070000.0
CLL	ذكر	56_65	متزوج	موظف	19700.0	4000000.0
CLL	أنثي	26_35	متزوج	ربه منزل	800.0	4980000.0
CLL	ذكر	36-45	متزوج	أخري	61900.0	4200000.0
CLL	أنثي	56_65	متزوج	ربه منزل	63500.0	4390000.0
CLL	أنثي	56_65	متزوج	ربه منزل	19000.0	4230000.0
CLL	أنثي	46-55	غير متزوج	ربه منزل	264200.0	3930000.0
CLL	ذكر	66 فأكثر	متزوج	أخري	200730.0	2570000.0
CLL	ذكر	36-45	متزوج	موظف	900.0	2780000.0
CLL	ذكر	56_65	متزوج	عامل حر	46100.0	4600000.0
CLL	ذكر	66 فأكثر	متزوج	عامل حر	6700.0	3500000.0
CML	أنثي	66 فأكثر	متزوج	ربه منزل	113000.0	3.63E7
CML	ذكر	36-45	متزوج	عامل حر	175600.0	1730000.0
CML	ذكر	اقل من 25	غير متزوج	طالب	2000.0	3530000.0
CML	ذكر	46-55	متزوج	مزارع	222500.0	1430000.0
CLL	ذكر	اقل من 25	متزوج	أخري	4200.0	4800000.0
CLL	أنثي	56_65	متزوج	أخري	15100.0	4360000.0

4620000.0	2800.0	عامل حر	متزوج	46-55	ذكر	CLL
3760000.0	189600.0	مزارع	متزوج	46-55	ذكر	CLL
4000000.0	62548.3	عامل حر	متزوج	66 فأكثر	ذكر	CLL
3766250.0	448000.0	مزارع	متزوج	66 فأكثر	ذكر	CML
3766250.0	112600.0	ربه منزل	متزوج	66 فأكثر	أنثي	CLL
3280000.0	23100.0	ربه منزل	متزوج	26_35	أنثي	CLL
4680000.0	2100.0	ربه منزل	متزوج	56_65	أنثي	CLL
3010000.0	18900.0	أخري	متزوج	56_65	ذكر	CLL
16.0	15900.0	ربه منزل	متزوج	66 فأكثر	ذكر	CML
3900000.0	14400.0	عامل حر	متزوج	36-45	ذكر	CML
3400000.0	12600.0	ربه منزل	متزوج	اقل من 25	أنثي	CML
2200000.0	82100.0	ربه منزل	متزوج	66 فأكثر	أنثي	CML
4840000.0	5800.0	أخري	متزوج	56_65	ذكر	CLL
4200000.0	19300.0	ربه منزل	متزوج	66 فأكثر	أنثي	CML
3250000.0	18900.0	طالب	غير متزوج	اقل من 25	ذكر	CML
4400000.0	25500.0	موظف	متزوج	56_65	ذكر	CLL
3780000.0	40900.0	عامل حر	متزوج	56_65	ذكر	CLL
4080000.0	73200.0	ربه منزل	متزوج	اقل من 25	أنثي	CML
3410000.0	1600.0	عامل حر	متزوج	36-45	ذكر	CML
2680000.0	2600.0	عامل حر	متزوج	66 فأكثر	ذكر	CLL
2710000.0	28600.0	عامل حر	متزوج	46-55	ذكر	CLL
3650000.0	68100.0	مزارع	متزوج	56_65	ذكر	CLL
3370000.0	276000.0	عامل حر	متزوج	56_65	ذكر	CLL
3520000.0	2400.0	سائق	متزوج	46-55	ذكر	CML
3980000.0	132100.0	مزارع	متزوج	56_65	ذكر	CLL
3520000.0	2800.0	موظف	غير متزوج	اقل من 25	ذكر	CML
3570000.0	5500.0	أخري	متزوج	66 فأكثر	ذكر	CLL
4530000.0	7000.0	تاجر	متزوج	66 فأكثر	ذكر	CLL
4660000.0	183000.0	أخري	متزوج	56_65	ذكر	CLL
3000000.0	448000.0	ربه منزل	متزوج	56_65	أنثي	CLL
3400000.0	28500.0	أخري	غير متزوج	46-55	ذكر	CLL
3580000.0	5400.0	مزارع	متزوج	66 فأكثر	ذكر	CLL
3750000.0	33600.0	مزارع	متزوج	56_65	ذكر	CLL
3460000.0	26300.0	ربه منزل	متزوج	56_65	أنثي	CML
5080000.0	4500.0	ربه منزل	متزوج	36-45	أنثي	CLL
3766250.0	30100.0	طالب	غير متزوج	اقل من 25	ذكر	CLL
2920000.0	42600.0	ربه منزل	متزوج	66 فأكثر	أنثي	CLL
4700000.0	31100.0	ربه منزل	متزوج	26_35	أنثي	CLL
1560000.0	242000.0	مزارع	متزوج	56_65	ذكر	CLL
3880000.0	40800.0	ربه منزل	متزوج	56_65	أنثي	CML
8500000.0	32100.0	عامل حر	متزوج	26_35	ذكر	CML
2290000.0	62548.3	أخري	متزوج	26_35	ذكر	CML
4630000.0	45500.0	عامل حر	متزوج	66 فأكثر	ذكر	CLL
3060000.0	50800.0	ربه منزل	متزوج	26_35	أنثي	CLL
4020000.0	7300.0	عامل حر	متزوج	26_35	ذكر	CML

1240000.0	26800.0	أخري	متزوج	66 فأكثر	ذكر	CLL
3280000.0	5300.0	عامل حر	متزوج	36-45	ذكر	CML
4400000.0	800.0	عامل حر	متزوج	56_65	ذكر	CML
3630000.0	49900.0	طالب	غير متزوج	اقل من 25	ذكر	CLL
2780000.0	12300.0	ربه منزل	متزوج	56_65	أنثي	CLL
5060000.0	14200.0	ربه منزل	متزوج	46-55	أنثي	CLL
4000000.0	36500.0	موظف	متزوج	56_65	أنثي	CLL
3790000.0	387600.0	أخري	متزوج	66 فأكثر	ذكر	CLL
3070000.0	42300.0	أخري	متزوج	56_65	ذكر	CML
4350000.0	43300.0	أخري	متزوج	46-55	ذكر	CML
3170000.0	14800.0	ربه منزل	متزوج	56_65	أنثي	CML
4460000.0	6400.0	أخري	متزوج	66 فأكثر	ذكر	CLL
3380000.0	54500.0	ربه منزل	متزوج	66 فأكثر	ذكر	CLL
4550000.0	130500.0	ربه منزل	متزوج	26_35	أنثي	CML
3230000.0	3200.0	ربه منزل	متزوج	66 فأكثر	ذكر	CML
3260000.0	25700.0	أخري	متزوج	46-55	ذكر	CML
4280000.0	33500.0	سائق	متزوج	56_65	ذكر	CLL

المصدر: مركز العلاج بالأشعة-2016 م.