استخدام نموذج الانحدار اللوجستي في دراسة العوامل المساعدة على تشخيص حالات الإصابة بسرطان المثانة

م.م. رشا عادل سعيد/ جامعة بغداد / كلية الإدارة والاقتصاد / قسم الإحصاء

المستخلص :

يُعد مرض السرطان من أكبر المشكلات الصحية التي تواجه العالم ، وتحتل أورام المثانة تحديداً موقعاً خاصاً بين أكثر أنواع الأورام انتشاراً في الوطن العربي ولاسيما العراق ومصر[٢] . إن سرطان المثانة من الأمراض التي يمكن علاجها والشفاء منها إذا تم الكشف عنها في وقتٍ مبكر . لذا جاء هذا البحث بهدف دراسة العوامل المساعدة على تشخيص حالات الإصابة بهذا المرض مثل [عمر المريض ، جنس المريض ، الاعراض الرئيسة التي يشكو منها المريض وهي وجود البيلة الدموية ، الحرقة أو الألم أثناء التبول ، واضطرابات التبول] وتحديد أي تلك العوامل الأكثر تأثيراً في امكانية تشخيص حالات الإصابة بسرطان المثانة وذلك بإستخدام نمذجة إحصائية للبيانات متمثلة بإنموذج الإنحدار اللوجستي والاعتماد على عينة عشوائية قوامها (128) مريضاً . وبعد جمع وتحليل المعلومات المتوفرة عنهم إحصائياً خَلُص البحث الى أن عامل (البيلة الدموية) كان له الاثر الاكبر في امكانية تشخيص حالات الاصابة بهذا المرض ، يليه عامل (عمر المريض) . أما العوامل الأخرى (جنس المريض والحرقة أو الألم أثناء التبول واضطرابات التبول) فلم يكن لها تأثير معنوي في إمكانية تشخيص حالات الإصابة بهذا المرض . وفي نهاية البحث قُدِمَت بعض التوصيات المتعلقة بالموضوع .

المصطلحات الرئيسية للبحث/ الانحدار اللوجستي- سرطان المثانة.

١- المقدمة :

تُشير العديد من البحوث الطبية إلى أن هناك علاقة وثيقة بين الجوانب التشخيصية للأمراض وبين بعض الدراسات والتطبيقات الإحصائية المستخدمة في هذا المجال وبما يُساعد على سرعة التشخيص المبكر للأمراض الصعبة . لكن على الرغم من ذلك نجد ان اغلب الدراسات والتطبيقات الإحصائية على المستوى العربي عموماً والمستوى المحلي خصوصاً التي تناولت موضوع إستخدام التطبيق الإحصائي لإنموذج الانحدار اللوجستي في المجال الطبي قد ركزت على دراسة العوامل المؤثرة في الإصابة بالامراض المختلفة (أي دراسة عوامل الخطورة ومسببات الامراض المختلفة) ولم تتناول الجانب التشخيصي للامراض (وهذا ما يُمثل مشكلة البحث) . لذا جاء هذا البحث بهدف توضيح العلاقة بين العوامل المساعدة في تشخيص حالات الإصابة بمرض (سرطان المثانة) والحالات المرضية المشكوك بإصابتها بهذا المرض ، وتحديد أولوية تلك العوامل بإستخدام التطبيق الإحصائي لإنموذج الإنحدار اللوجستي .

٢- الدراسات السابقة :

نورد فيما يأتي بعض الدراسات الإحصائية السابقة التي تناولت موضوع التطبيق الإحصائي لإنموذج الانحدار اللوجستي في المجال الطبي على وجه الخصوص .

أ-دراسة الباحثة هبة ابراهيم صالح البياتي (عام ٢٠٠٥) الموسومة (تحليل المسار في إنموذج الانحدار اللوجستي مع تطبيق عملي) . تناولت هذه الدراسة دراسة التأثيرات المباشرة وغير المباشرة لمجموعة من العوامل المؤثرة على الإصابة بمرض فقر الدم لمجموعتين من المجاميع السكانية الاكثر عرضة للاصابة بهذا المرض هم الاشخاص الذين اعمارهم تقل عن (١٨ سنة) والنساء الحوامل[١].

ب- دراسة الباحثة امل حسن محمد ياسين (عام ٢٠٠٨) الموسومة (استخدام تحليل الانحدار اللوجستي لمعرفة مسببات الاصابة بسرطان الثدي لدى الاناث ، دراسة حالة المركز القومي للعلاج بالطب النووي بالخرطوم) . كانت تهدف هذه الدراسة الى التعرّف على العوامل المؤثرة في الاصابة بسرطان الثدي لدى (٢٠٠) امرأة من مراجعي المركز القومي للعلاج بالطب النووي في الخرطوم[١١].

ج- دراسة الباحثة انوار ادريس حسن عبد الماجد (عام ٢٠٠٩) الموسومة (استخدام الإنموذج اللوجستي المتعدد لتحديد العوامل المؤثرة في مرض الجلكوما) . ركزت هذه الدراسة على تحديد عوامل الخطر لمرض الجلكوما[٨].

د- دراسة المدرس المساعد بهاء عبد الرزاق قاسم (عام ٢٠١١) الموسومة (تحليل أثر بعض المتغيرات في الاصابة بمرض اللثة باستخدام إنموذج الانحدار اللوجستي) . كانت تهدف هذه الدراسة الى دراسة وتحليل بعض العوامل المؤثرة في الاصابة بمرض اللثة مثل الترسبات الكلسية وسوء التغذية وغيرها من العوامل[١٠].

هـ- دراسة الباحثين H. Yusuff , N. Mohamad , U.K. Ngah & A.S. Yahaya (عام ٢٠١٢) الموسومة (Breast Cancer Analysis Using Logistic Regression - تحليل مرض سرطان الثدي بإستخدام الانحدار اللوجستي) . كانت تهدف هذه الدراسة الى استخدام الانحدار اللوجستي كأداة مكملة لتصوير الثدي بالاشعة السينية (mammogram) من اجل تشخيص الاصابة بمرض سرطان الثدي[١٥].

و- دراسة الاستاذ الدكتور محمد حبيب الشاروط و الاستاذ المساعد اميرة جابر محيسن (عام ٢٠١٣) الموسومة (الانحدار اللوجستي لدراسة الاصابة بسرطان الثدي) . ركزت هذه الدراسة على استخدام اسلوب الانحدار اللوجستي كأداة مهمة في دراسة العوامل المؤثرة في الاصابة بسرطان الثدي[٣].

ز- دراسة الباحثة فاطمة احمد حسن (عام ٢٠١٤) الموسومة (استخدام الإنموذج اللوجستي لتحديد العوامل المؤثرة في الاصابة بمرض فقر الدم (الأنيميا) لدى الاطفال (٢٠٠٩م – ٢٠١٣م) . وتهدف هذه الدراسة الى دراسة وتحليل بعض العوامل المؤثرة في الاصابة بمرض فقر الدم (الأنيميا) لدى الاطفال[٦].

٣- مشكلة البحث :

تتمثل مشكلة البحث في شحة البحوث والدراسات الإحصائية التي تتناول موضوع دراسة العوامل المساعدة على تشخيص الامراض المختلفة بإستخدام إنموذج الانحدار اللوجستي .

٤- هدف البحث :

يهدف البحث الى استخدام إنموذج الانحدار اللوجستي لتوضيح العلاقة بين العوامل المؤثرة في تشخيص حالات الاصابة بسرطان المثانة مثل (عمر المريض ، جنس المريض ، الاعراض الرئيسة التي يشكو منها المريض) والحالات المرضية المشكوك بإصابتها بهذا المرض . وتحديد اولوية تلك العوامل بمعنى تحديد العوامل الاكثر تأثيراً في امكانية تشخيص حالات الاصابة بهذا المرض قبل اجراء التحليلات المختبرية المتعددة .

٥- منهجية البحث :

اعتمد البحث في تحقيق هدفه على المنهجيات الآتية :

أ- التعريف بشكل مبسط بمرض سرطان المثانة واعراضه الرئيسة.

ب- توضيح المفهوم الإحصائي لإنموذج الإنحدار اللوجستي والتركيز على خصائصه وكيفية تقدير معلماته.

ج- أُعتُمِدَ إنموذج الانحدار اللوجستي المتعدد واختباراته للوصول الى النتائج ، بإستخدام البرنامج الجاهز SPSS.

٦- سرطان المثانة واعراضه :

يُعرّف سرطان المثانة بأنه نمو خبيث غير عادي لخلايا المثانة ويمكن ان ينقسم هذا السرطان الى نوعين هما النوع السطحي المنشأ والنوع عميق المنشأ في عضلات المثانة[٤] .

اهم اعراض مرض سرطان المثانة ما يأتي[٥] :

أ- البيلة الدموية (بالإنجليزية Hematuria) وهي وجود الدم في البول.

ب- احساس بالحرقة او الالم اثنام التبول دون وجود اي مؤشر على عدوى او إلتهاب ضمن المسالك البولية.

ج- تغيير في عادات المثانة مثل الاضطرار الى التبول اكثر من المعتاد او الشعور القوي بالحاجة الى التبول دون خروج كمية بول كبيرة.

عند وصول السرطان الى مرحلة متقدمة اكثر تظهر اعراض اضافية[٥] :

- فقدان الوزن
- فقدان الشهية
- الحمى

❖ألم في العظام او المستقيم والشرج او منطقة الحوض.

والجدير بالذكر ان وجود دم في البول هو عرض مشترك مع كل من الامراض الاتية[٤] :

سرطان المثانة ، سرطان الكلية (ورم ويلمز) ، حصى الكلى ، سرطان الكلى ، إلتهاب المثانة في شهر العسل . اما الاحساس بالحرقة أوالالم اثناء التبول فهو عرض مشترك مع كل من الامراض الاتية[٤] :

سرطان المثانة ، إلتهاب المسالك البولية ، الاضطرابات الجسدية . كذلك ، اضطرابات التبول هو عرض مشترك مع كل من الامراض التالية[٤] :

سرطان المثانة ، سرطان المبيض ، تضخم البروستاتا الحميد، الورم النخاعي المتعدد ، إلتهاب المثانة في شهر العسل .

٧- عينة البحث ومتغيراته :

تم الاعتماد على عينة إحصائية حجمها (128) شخصاً (مريضاً) (89 ذكور و 39 إناث) تمثلت بمراجعي مستشفى الكاظمية التعليمي / قسم استشارية الجراحة البولية خلال عام (2006) كانوا جميعاً يشكون على الاقل من احد الاعراض الآتية :

[وجود دم في البول (بيلة دموية) ، حرقة أو ألم اثناء التبول ، اضطرابات التبول (تغيير في عادات المثانة)]

. أما متغيرات البحث فقد كانت كالآتي :

$Y=g(x)$: المتغير التابع وهو من النوع الثنائي (dichotomous) ويأخذ القيمة (1) اذا كان الشخص مصاب بسرطان المثانة والقيمة (0) اذا كان الشخص غير مصاب بسرطان المثانة اما المتغيرات المستقلة فهي :

X_1: جنس المريض وهو متغير ثنائي يأخذ القيم (ذكر =1 و انثى =0)

X_2: الفئة العمرية للمريض وتم التعبير عنها بالقيم (الفئة 40-49=1 ، والفئة 50-59= 2 ، والفئة 60-69 =3 ، والفئة 70-79=4 ، والفئة 80 فأكثر =5) .

X_3: وجود دم في البول (بيلة دموية) وهو متغير ثنائي يأخذ القيم (كلا=0 و نعم=1).

X_4: وجود حرقة أو ألم اثناء التبول وهو متغير ثنائي يأخذ القيم (كلا=0 و نعم=1).

X_5: وجود اضطرابات في التبول (تغيير في عادات المثانة) وهو متغير ثنائي يأخذ القيم (كلا=0 و نعم=1).

٨- مفهوم الانحدار اللوجستي :

يُعرّف الانحدار اللوجستي بانه طريقة احصائية لتحليل مجموعة البيانات (dataset) التي يتحدد فيها متغير الاستجابة من خلال متغير مستقل واحد او اكثر(one or more independent variable). حيث يكون متغير الاستجابة (المتغير التابع) من النوع الثنائي (binary or dichotomous)[١٦].

ويُعرّف الانحدار اللوجستي ايضاً بانه طريقة نمذجة احصائية للبيانات الفئوية (categorical data) . ويستخدم بشكل واسع في مجال البحوث الطبية والمالية والاعمال والهندسة والتسويق والاقتصاد والتأمين الصحي ... الخ[١٣] .

ان الانحدار اللوجستي يشبه الانحدار الخطي من حيث انه يهدف الى توضيح العلاقة بين المتغير التابع (متغير النتيجة) ومجموعة من المتغيرات المستقلة (المتغيرات التوضيحية او التفسيرية) لكنه يختلف عنه بان المتغير التابع في إنموذج الانحدار اللوجستي يجب ان يكون من النوع الثنائي وهذا الاختلاف ينعكس بدوره على افتراضات الانحدار اللوجستي[١٤] .

٩- خصائص الانحدار اللوجستي(١٢) :

أ- الانحدار اللوجستي لا يفترض وجود علاقة خطية بين المتغير التابع والمتغيرات المستقلة.

ب- المتغير التابع يجب ان يكون ثنائي التفرع (dichotomous) بحيث يحتوي على فئتين (مصابين وغير مصابين على سبيل المثال).

ج- الانحدار اللوجستي لايشترط ان تكون المتغيرات المستقلة من النوع المستمر ولا ان تتبع التوزيع الطبيعي (not normally distributed) ولا ان تكون العلاقة بين المتغير التابع و المتغيرات المستقلة خطية ولا يفترض تساوي التباين ضمن كل فئة. وهذا يجعل إنموذج الانحدار اللوجستي أكثر مرونة من بقية نماذج التنبؤ والتصنيف[7] .

د- يجب ان تكون الفئات محددة وشاملة بحيث ان كل مفردة تنتمي الى فئة واحدة فقط.

هـ- يجب ان يكون حجم العينة المستخدم في الانحدار اللوجستي اكبر من حجم العينة المستخدم في الانحدار الخطي لان معاملات إنموذج الانحدار اللوجستي يتم تقديرها باستخدام طريقة دالة الامكان الاعظم (maximum likelihood method) وهي طريقة تحتاج الى عينة كبيرة الحجم نسبياً .

١٠-تقدير معلمات إنموذج الانحدار اللوجستي :

يُبنى إنموذج الانحدار اللوجستي على فرض اساسي هو ان المتغير التابع (y) متغير الاستجابة الذي نهتم بدارسته هو متغير ثنائي يتبع توزيع (Bernoulli) يأخذ الرتبة (1) باحتمال (p) والرتبة (0) باحتمال (q=1−p) اي حدوث الاستجابة وعدم حدوثها[٩] .

يمكن التعبير عن إنموذج الانحدار اللوجستي في حالة إحتوائه على متغير مستقل واحد فقط بالصيغة الرياضية الآتية[١٤] :

$$\pi(x) = E(Y/x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \qquad .....(1)$$

حيث أن :

$E(Y/x) = \pi(x)$: يُمثل المتوسط المشروط (conditional mean) لمتغير الاستجابة (Y) عند قيمة معينة لـ(x).

β_0, β_1 : معلمات إنموذج الانحدار اللوجستي.

أما إذا كان إنموذج الإنحدار اللوجستي يحتوي على أكثر من متغير مستقل واحد وليكن (k) من المتغيرات المستقلة عندئذٍ يُعبر عنه بالصيغة الرياضية الآتية :

$$\pi(x) = E(Y/x) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}} \quad \ldots\ldots\ldots\ldots(2)$$

حيث أن :

$E(Y/x) = \pi(x)$: (تم تعريفها في المعادلة (١)).

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$: معلمات إنموذج الانحدار اللوجستي.

ولتحويل المعادلة (2) الى الشكل الخطي يُستخدم ما يسمى بتحويلة لـوجت(logit transformation) والتي يُعبر عنها بالمعادلة (3) ادناه[١٤] :

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad \ldots\ldots(3)$$

حيث أن :

$P(Y = 1/x) = \pi(x)$ عندما Y = 1 (حيث $P(Y = 1/x)$ يُمثل الاحتمال الشرطي (conditional probability) بأن تكون Y = 1 عند قيمة معينة لـ(x)).

$P(Y = 0/x) = 1 - \pi(x)$ عندما Y = 0 (حيث $P(Y = 0/x)$ يُمثل الاحتمال الشرطي (conditional probability) بأن تكون Y = 0 عند قيمة معينة لـ(x)).

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$: (تم تعريفها في المعادلة (٢)).

اما لتقدير معلمات المعادلة (3) فتُستخدم طريقة الامكان الاعظم (maximum likelihood method) وكما يأتي[14] :

بما انه $P(Y = 1/x) = \pi(x)$ عندما Y = 1

و $P(Y = 0/x) = 1 - \pi(x)$ عندما Y = 0

لذا عندما يكون لدينا عينة من المشاهدات المستقلة التي حجمها (n) (sample of (n) independent observations) فإن لكل زوج (x_i, y_i) ، i = 1 , 2 , ... , n ، حيث (y_i) تُمثل رتبة متغير الاستجابة الثنائي للمفردة (i) و (x_i) تُمثل قيمة المتغير المستقل للمفردة (i). فإنه :

$P(y_i = 1/x) = \pi(x_i)^{y_i}$ عندما تكون y_i = ١

و $P(y_i = 0/x) = [1 - \pi(x_i)]^{1-y_i}$ عندما تكون y_i= ٠

وعليه يمكن التعبير عن دالة الامكان (likelihood function) بالصيغة الرياضية الآتية[١٤] :

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad \ldots\ldots\ldots(4)$$

حيث أن :

$P(y_i = 1/x) = \pi(x_i)^{y_i}$ عندما y_i=1 (حيث $P(y_i = 1/x)$ يُمثل الاحتمال الشرطي (conditional probability) بأن تكون y_i 1= عند قيمة معينة لـ(x)).

$P(y_i = 0/x) = [1 - \pi(x_i)]^{1-y_i}$ عندما y_i=0 (حيث $P(y_i = 0/x)$ يُمثل الاحتمال الشرطي (conditional probability) بأن تكون y_i 0= عند قيمة معينة لـ(x)).

وبأخذ اللوغاريتم الطبيعي للطرفين نحصل على المعادلة (5) الاتية[١٤] :

$$L(\beta) = ln[l(\beta)] = \sum_{i=1}^{n} \{y_i ln[\pi(x_i)] + (1 - y_i) ln[1 - \pi(x_i)]\} \ldots\ldots\ldots\ldots(5)$$

(الرموز تم تعريفها سابقاً)

ثم يتم اشتقاق المعادلة (5) بالنسبة للمعلمات المراد تقديرها (β) وجعلها مساوية للصفر فينتج عدد من المعادلات التي لا يمكن حلها الا من خلال خوارزمية تكرارية[١٠] ، تُسمى خوارزمية المربعات الصغرى الموزونة التكرارية (Iteratively Weighted Least Squares Algorithm)[١٧] .

١١-تقييم القوة التفسيرية للإنموذج :

في إنموذج الانحدار اللوجستي يستعاض عن معامل التحديد (R^2) الذي يستخدم لمعرفة جودة معادلة الانحدار التقديرية في تفسير العلاقة بين المتغير التابع و المتغيرات المستقلة بالإحصائتين $R^2_{cox\ \&snell}$ و $R^2_{Nagelkerke}$ اللتين لهما نفس هدف معامل التحديد (R^2) المستخدم في الانحدار الخطي المتعدد. إلا ان $R^2_{cox\ \&\ snell}$ لا يمكن ان تصل قيمتها الى الواحد الصحيح في حين ان $R^2_{Nagelkerke}$ بامكانها ذلك[10] فحدود $R^2_{Nagelkerke}$ تمتد من (صفر الى الواحد الصحيح) مما يجعلها اكثر موثوقية من $R^2_{cox\ \&\ snell}$ ومن الطبيعي ان تكون قيمتها اعلى من $R^2_{cox\ \&\ snell}$ [12] .

تُحسب احصاءة $R^2_{cox\ \&\ snell}$ من الصيغة الرياضية الآتية[١٠] :

$$R^2_{cox\ \&snell} = 1 - \left[\frac{L_0}{L_1}\right]^{(2/n)} \quad \text{.....(6)}$$

حيث ان : L_0 : دالة الامكان الاعظم في حالة الإنموذج المتضمن الحد الثابت فقط.

L_1 : دالة الامكان الاعظم في حالة الإنموذج المتضمن لجميع المتغيرات التوضيحية.

n : حجم العينة.

اما احصاءة $R^2_{Nagelkerke}$ فتُحسب من الصيغة الرياضية الآتية[10] :

$$R^2_{Nagelkerke} = \frac{R^2_{cox\ \&\ snell}}{1-[L_0]^{\left(\frac{2}{n}\right)}} \quad \text{.....(7)}$$

حيث أن : $R^2_{cox\ \&snell}$: تُمثل الإحصاءة المحسوبة وفق الصيغة الرياضية (٦).

L_0 : دالة الامكان الاعظم في حالة الإنموذج المتضمن الحد الثابت فقط.

n : حجم العينة.

١٢-الإختبارات الإحصائية الخاصة بإنموذج الإنحدار اللوجستي :

أ-**إختبار** wald : تختبر إحصاءة wald التي تتبع توزيع مربع كاي (x^2) وبدرجة حرية df=1 لكل معلمة من معلمات إنموذج الانحدار اللوجستي المقدرة معنوية تلك المعلمات .

وتُحسب إحصاءة wald على وفق الصيغة الرياضية الآتية[9] :

$$wald = \left[\frac{\widehat{\beta_j}}{S.E.(\widehat{\beta_j})}\right]^2 \quad \text{.....(8)}$$

حيث أن : $\widehat{\beta_j}$: المعلمة المقدرة ذات الرتبة (j).

$S.E.(\widehat{\beta_j})$: الخطأ المعياري للمعلمة المقدرة ذات الرتبة (j).

أما فرضية الاختبار فهي :

$H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$ j= 1,2,.....,k

(أي إختبار معنوية كل معلمة من معلمات إنموذج الانحدار اللوجستي) . وهو اختبار من طرفين[9] . فإذا كانت القيمة الإحتمالية لإحصاءة wald (significance value) أقل من (0.05) عندئذٍ تُرفض فرضية العدم[12] . أي ان المعلمة (β_j) معنوية ولا تساوي صفراً في المجتمع الذي سُحبت منه العينة[9] .

ب- **إختبار** Homser & Lemeshow (H&L) : تُستخدم إحصاءة H&L التي تتبع توزيع مربع كاي (x^2) لمعرفة فيما إذا كان الإنموذج يُمثل البيانات بشكلٍ جيد أم لا (well-fitting model or not)[12] . من خلال تقييم الفرق بين القيم المشاهدة (observed values) والقيم المتوقعة (expected values)[10] . وإختبار الفرضية الآتية[18] :

H_0 : عدم وجود فرق معنوي بين القيم المشاهدة والقيم المتوقعة .

vs

H_1 : وجود فرق معنوي بين القيم المشاهدة والقيم المتوقعة .

فإذا كانت إحصاءة H&L أكبر من (0.05) عندئذٍ يكون الإنموذج ممثلاً للبيانات بشكلٍ جيد[12]. وبهذا نفشل في رفض فرضية العدم (أي نقبل فرضية العدم) القائلة بعدم وجود إختلافات بين القيم المشاهدة والقيم المتوقعة[18].

١٣-الجانب التطبيقي للبحث :

أ- خطوات الجانب التطبيقي للبحث :

لتحقيق هدف البحث أستُخدِمَت الحزمة البرمجية الجاهزة SPSS حيث أُدخلت البيانات الى الحاسوب ومن ثم النقر على Analyze >> Regression >> Binary Logistic و تم وضع متغير الإستجابة في حقل المتغير المعتمد ، ووضعت المتغيرات (جنس المريض ، عمر المريض ، البيلة الدموية ، الحرقة أو الألم اثناء التبول ، اضطرابات التبول) في حقل المتغيرات المستقلة و أُختيرت طريقة Enter ثم النقر على المفتاح ok فتم الحصول على تسعة جداول ، حيث بيّن الجدولين (١) و (٢) المعلومات الوصفية الملخصة لعينة البحث و رموز قيم المتغير التابع (متغير الاستجابة) كما هو واضح في الجدولين الاتيتين .

جدول (١) Case Processing Summary

Unweighted Cases[a]	N	Percent
Selected Cases Included in Analysis	128	100.0
Missing Cases	0	.0
Total	128	100.0
Unselected Cases	0	.0
Total	128	100.0

a. If weight is in effect, see classification table for the total number of cases.

يوضح الجدول (١) عدد الحالات المدخلة في التحليل وهو (١٢٨) وعدد الحالات المفقودة (البيانات المفقودة) وهو (صفر) وحجم العينة الكلي (١٢٨).

جدول (٢) Dependent Variable Encoding

Original Value	Internal Value
غير مصاب	0
مصاب	1

Block 0: Beginning Block

يوضح جدول (٢) قيم المتغير التابع الثنائي حيث عُبر عن الشخص المصاب بالرقم (١) وغير المصاب بالرقم (صفر).

اما الجداول (٣) و (٤) و (٥) و (٦) و (٧) و (٨) و (٩) فقد بيّنت نتائج الجانب التطبيقي للبحث .

ب- نتائج الجانب التطبيقي للبحث :

جدول (٣) Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients					
			Constant	جنس المريض	عمر المريض	بيلة دموية	حرقة أو ألم اثناء التبول	اضطرابات التبول
Step 1	1	78.695	-2.991	.398	.225	3.536	.209	.265
	2	70.529	-5.174	.868	.503	4.977	.431	.522
	3	69.521	-6.533	1.191	.709	5.616	.558	.627
	4	69.483	-6.844	1.263	.759	5.751	.582	.632
	5	69.483	-6.858	1.266	.761	5.757	.583	.631
	6	69.483	-6.858	1.266	.761	5.757	.583	.631

a. Method: Enter.
b. Constant is included in the model.
c. Initial -2 Log Likelihood: 176.945.
d. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

يوضح الجدول (٣) عدد الدورات التكرارية (Iterations) للحصول على اقل قيمة لسالب ضعف لوغارتيم دالة الامكان (-2 Log Likelihood) والتي تم الحصول عليها فعلاً في الدورة السادسة حيث بلغت قيمتها (69.483) علماً أن قيمتها الاولية في حالة الإنموذج المتضمن الحد الثابت فقط (Constant) كانت تساوي (176.945) كما هو مذكور في الفقرة (c) المذكورة انفا . و واضح من الفقرة (d) المذكورة انفا بانه قد تم التوقف عند الدورة السادسة لان التغير في قيم المعلمات المقدرة ($\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\beta_4}, \widehat{\beta_5}$) اصبح اقل من (0.001) وبهذا تُعد قيم ($\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\beta_4}, \widehat{\beta_5}$) التي تم الحصول عليها في الدورة السادسة هي افضل تقدير ممكن لمعلمات إنموذج الانحدار اللوجستي .

جدول (٤) Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	107.462	5	.000
	Block	107.462	5	.000
	Model	107.462	5	.000

يوضح جدول (٤) كفاءة الإنموذج ككل وجودته بإستخدام اختبار نسبة الامكان الاعظم (the likelihood ratio test) والذي يتبع توزيع مربع كاي (x^2) وبدرجة حرية (df) مساوية الى عدد المتغيرات المستقلة وهي (5). ويُحسب على وفق الصيغة الآتية :

$$x^2 = -2logL_0 - (-2logL_1)$$

حيث أن :

$-2logL_0$: سالب ضعف لوغارتيم دالة الامكان الاعظم في حالة الإنموذج المتضمن الحد الثابت فقط.

$-2logL_1$: سالب ضعف لوغارتيم دالة الامكان الاعظم في حالة الإنموذج المتضمن جميع المتغيرات المستقلة.

أي أن : x^2=176.945 - 69.483 = 107.462

ويتضح من الجدول (٤) ان قيمة (x^2) معنوية عند مستوى دلالة (α =0.001) لأن (sig.=0.000) مما يؤكد كفاءة ومعنوية الإنموذج .

جدول (٥) Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	69.483[a]	.568	.758

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

يوضح جدول (٥) قيم ($R^2_{cox\ \&\ snell}$) و ($R^2_{Nagelkerke}$) حيث كانت ($R^2_{cox\ \&\ snell}$ = 0.568) وهذا يشير الى أن (٥٦.٨%) من التغير في المتغير التابع يتم تفسيره من خلال إنموذج الانحدار اللوجستي ، اما ($R^2_{Nagelkerke} = 0.758$) وهذا يشير الى أن (٧٥.٨%) من التغير في المتغير التابع يتم تفسيره من خلال إنموذج الانحدار اللوجستي .

جدول (٦) Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.421	8	.308

يوضح جدول (٦) نتائج اختبار Hosmer and Lemeshow حيث يظهر ان إحصاءة H & L التي تتبع توزيع مربع كاي (x^2) تساوي (٩.٤٢١) وهي غير معنوية لان (sig.=0.308) وهي اكبر من (٠.٠٥) لذا نقبل فرضية العدم القائلة بعدم وجود فرق معنوي بين القيم المشاهدة والقيم المتوقعة وهذا ما تؤكده ايضاً نتائج الجدول (٧) .

جدول (٧) Contingency Table for Hosmer and Lemeshow Test

		الاستجابة = غير مصاب		الاستجابة = مصاب		
		Observed	Expected	Observed	Expected	Total
Step 1	1	7	6.885	0	.115	7
	2	13	12.596	0	.404	13
	3	10	9.491	0	.509	10
	4	8	9.357	2	.643	10
	5	12	10.763	0	1.237	12
	6	11	11.034	2	1.966	13
	7	3	5.032	14	11.968	17
	8	0	.602	6	5.398	6
	9	2	1.433	16	16.567	18
	10	2	.808	20	21.192	22

جدول (٨) Classification Table[a]

Observed			Predicted		
			الاستجابة		Percentage Correct
			مصاب غير	مصاب	
Step 1	الاستجابة	مصاب غير	63	5	92.6
		مصاب	6	54	90.0
	Overall Percentage				91.4

a. The cut value is .500

يوضح الجدول (٨) النسبة المئوية للتصنيف الصحيح حيث يتضح ان ($\frac{63}{63+5} * 100\% = 92.6\%$) من الاشخاص تم تصنيفهم بشكل صحيح ضمن فئة غير المصابين و ($\frac{54}{6+54} * 100\% = 90.0\%$) من الاشخاص تم تصنيفهم بشكل صحيح ضمن فئة المصابين وبشكل عام ($\frac{63+54}{128} * 100\% = 91.4\%$) من الاشخاص تم تصنيفهم بشكل صحيح الى الفئات التي ينتمون اليها وان (١١) شخص فقط أي ما يعادل ($\frac{11}{128} * 100\% = 8.6\%$) فقط تم تصنيفهم بشكل خاطئ وهي نسبة قليلة مما يدل على ان الإنموذج يمثل البيانات تمثيلاً جيداً .

جدول (٩) Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	جنس المريض	1.266	.887	2.037	1	.153	3.548	.623	20.196
	عمر المريض	.761	.376	4.094	1	.043	2.140	1.024	4.471
	بيلة دموية	5.757	1.627	12.515	1	.000	316.505	13.034	7685.665
	حرقة أو ألم اثناء التبول	.583	1.409	.172	1	.679	1.792	.113	28.355
	اضطرابات التبول	.631	1.523	.172	1	.678	1.880	.095	37.177
	Constant	-6.858	2.763	6.162	1	.013	.001		

a. Variable(s) entered on step 1: جنس المريض, عمر المريض, بيلة دموية, حرقة او الم اثناء التبول, اضطرابات التبول.

يوضح جدول (9) في عموده الأول القيم التقديرية للمعلمات ($\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\beta_4}, \widehat{\beta_5}$) أي أن معادلة إنموذج الانحدار اللوجستي التقديرية ستكون بالشكل الآتي :

$$\hat{Y} = \hat{g}(x) = -6.858 + 1.266\ X_1 + 0.761\ X_2 + 5.757\ X_3 + 0.583\ X_4 + 0.631\ X_5$$

كما يوضح الجدول (9) في عموده الثاني قيم الخطأ المعياري (S.E.) لكل معلمة من معلمات الإنموذج المقدرة وفي عموده الثالث والرابع والخامس تظهر قيم إحصاءة (wald) لكل معلمة ودرجات الحرية (df) ومعنوية تلك المعلمات (sig.) على التوالي . أما قيم عموده السادس ((Exp(B)) فتُمثل نسبة الأرجحية (odds ratio) حيث تُشير قيمة ((Exp(B)) الى مقدار التغير الحاصل في نسبة أرجحية وقوع الحدث (مصاب) عند حدوث تغير في قيمة المتغير المستقل المرتبط بالمعلمة (B) . والعمود السابع يُبين حدود الثقة لـ ((Exp(B)) .

إذا كانت قيمة ((Exp(B)) اكبر من (1) عندئذٍ تزداد نسبة أرجحية وقوع الحدث (مصاب) . أما إذا كانت قيمة ((Exp(B)) أقل من (1) عندئذٍ أي زيادة في قيمة المتغير المستقل (X) تؤدي الى إنخفاض نسبة أرجحية وقوع الحدث (مصاب) .

بالنسبة للمتغير الأول (X_1 جنس المريض) فإن :

$$Exp(B) = Exp(1.266) = e^{1.266} = 3.548$$

وهذا يعني أن أرجحية أن يكون الشخص مصاب بسرطان المثانة أعلى بمقدار (3.548) عند الرجال مما هو عند النساء .

بالنسبة للمتغير الثاني (X_2 عمر المريض) فإن :

$$Exp(B) = Exp(0.761) = e^{0.761} = 2.140$$

وهذا يعني أنه عند زيادة عمر المريض بمقدار وحدة واحدة فإن نسبة أرجحية أن يكون الشخص مصاب بسرطان المثانة ستكون أكبر بمقدار (2.140) مرة .

بالنسبة للمتغير الثالث (X_3 وجود دم في البول) فإن :

$$Exp(B) = Exp(5.757) = e^{5.757} = 316.505$$

وهذا يعني أنه عند تغير المتغير (X_3) من عدم وجود دم في البول (أي $X_3 = 0$) الى وجود دم في البول (أي $X_3 = 1$) فإن نسبة أرجحية أن يكون الشخص مصاب بسرطان المثانة ستكون أكبر بمقدار (316.505) مرة .

بالنسبة للمتغير الرابع (X_4 الحرقة أو الألم أثناء التبول) فإن :

$$Exp(B) = Exp(0.583) = e^{0.583} = 1.792$$

وهذا يعني أنه عند تغير المتغير (X_4) من عدم وجود حرقة أو ألم أثناء التبول (أي $X_4 = 0$) الى وجود حرقة أو ألم أثناء التبول (أي $X_4 = 1$) فإن نسبة أرجحية أن يكون الشخص مصاب بسرطان المثانة ستكون أكبر بمقدار (1.792) مرة .

بالنسبة للمتغير الخامس (X_5 إضطرابات التبول) فإن :

$$Exp(B) = Exp(0.631) = e^{0.631} = 1.880$$

وهذا يعني أنه عند تغير المتغير (X_5) من عدم وجود إضطرابات في التبول (أي $X_5 = 0$) الى وجود إضطرابات في التبول (أي $X_5 = 1$) فإن نسبة أرجحية أن يكون الشخص مصاب بسرطان المثانة ستكون أكبر بمقدار (1.880) مرة .

١٤-الإستنتاجات والتوصيات :

أ- الإستنتاجات :

١- العلاقة بين المتغير التابع والمتغيرات المستقلة كانت قوية إلى حدٍ ما حيث بلغت قيمة ($R_{Nagelkerke}$) (0.758) .

٢- إن إنموذج الإنحدار اللوجستي الذي تم تقدير معلماته وإختبار معنويتها كان إنموذجاً ممثلاً للبيانات بشكلٍ جيد (good fitting model) حيث أنه إجتاز كل إختبارات حُسن المطابقة بنجاح (إختبار نسبة الإمكان الأعظم Likelihood ratio test و إختبار H & L) .

٣- المتغير الثالث (X_3) الذي يُمثل وجود بيلة دموية (أي وجود دم في البول) له الأثر الأكبر في إمكانية تشخيص حالات الإصابة بسرطان المثانة حيث أظهر إختبار (wald) معنوية عالية للمعلمة ($\widehat{\beta_3}$) المرتبطة بالمتغير (X_3) فكانت (sig. = 0.000) .

٤- المتغير الثاني (X_2) الذي يُمثل عمر المريض يحتل المرتبة الثانية في إمكانية تشخيص حالات الإصابة بسرطان المثانة حيث أظهر إختبار (wald) معنوية المعلمة ($\widehat{\beta_2}$) المرتبطة بالمتغير (X_2) فكانت (sig. = 0.043) .

٥- إن المتغيرات (X_1 جنس المريض و X_4 الحرقة أو الألم اثناء التبول و X_5 إضطرابات التبول) لم يكن لها تأثير معنوي في إمكانية تشخيص حالات الإصابة بسرطان المثانة حيث أظهر إختبار (wald) عدم معنوية المعلمات ($\widehat{\beta_5}\widehat{\beta_1}$, $\widehat{\beta_4}$) المرتبطة بتلك المتغيرات فكانت قيم (sig.) لها على التوالي (0.678 , 0.679 , 0.153) .

ب- التوصيات :

أ-بعد أن أظهرت الدراسة الأثر الكبير لعامل (البيلة الدموية) في إمكانية تشخيص حالات الإصابة بسرطان المثانة نوصي بضرورة مراجعة الاشخاص الذين يشكون من وجود دم في البول الى المؤسسات الصحية المختصة بسرعة لتلقي العلاج المناسب .

ب- إجراء دراسات متقدمة لتقدير أعداد الإصابات الجديدة بمرض سرطان المثانة على المستوى المحلي (على مستوى العراق) لمعرفة إتجاهات إنتشار هذا المرض ووضع ستراتيجيات مناسبة لمواجهة مثل هذه الحالات المرضية مستقبلاً .

ت- توسيع إستخدام إنموذج الإنحدار اللوجستي في مجال الدراسات المختلفة (الطبية وغيرها) خصوصاً عند توفر عينة كبيرة الحجم وذلك لكفاءة ذلك الإنموذج وحصانته ومرونته .

١٥-المصادر :

أ- المصادر العربية :

١- البياتي ، هبة ابراهيم صالح ، (٢٠٠٥) ، "تحليل المسار في إنموذج الانحدار اللوجستي مع تطبيق عملي" ، رسالة ماجستير ، كلية الادارة والاقتصاد ، جامعة المستنصرية .

٢- السامرائي ، البروفسور الدكتور سمير احمد ، "اورام المثانة اسبابها وتشخيصها وعلاجها الحديث – Professor Al – Samarrai Medical Center . (دبي) Articles

www.profsam.net/arabic/Articles1360.htm

٣- الشاروط ، أ.د. محمد حبيب ، محيسن ، أ.م. اميرة جابر ، (٢٠١٣) ، "الانحدار اللوجستي لدراسة الاصابة بسرطان الثدي" . Cosit.gov.iq/images%5CA%5C3rd.pdf

٤- الصفحة الرئيسة لمستشفى الملك فيصل التخصصي ومركز الابحاث.

www.appserv.kfshrc.edu.org.sa/default/health

٥- الموسوعة الحرة (ويكيبيديا). www.wikipedia.org/wiki/سرطان_المثانة

٦- حسن ، فاطمة احمد ، (٢٠١٤) ، "استخدام الإنموذج اللوجستي لتحديد العوامل المؤثرة على الاصابة بمرض فقر الدم (الأنيميا) لدى الاطفال ٢٠٠٩م – ٢٠١٣م" ، رسالة ماجستير ، كلية العلوم ، جامعة السودان للعلوم والتكنولوجيا. Repository.sustech.ed/handle/123456789/9957

٧- عباس ، علي خضير ، (٢٠١٢) ، "استخدام الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية " ، مجلة جامعة كركوك للعلوم الادارية والاقتصادية ، المجلد الثاني ، العدد (٢) : صفحة ٢٣٤-٢٥٣ .

٨- عبد الماجد ، انوار ادريس حسن ، (٢٠٠٩) ، "استخدام الإنموذج اللوجستي المتعدد لتحديد العوامل المؤثرة في مرض الجلكوما" ، رسالة ماجستير ، كلية العلوم ، جامعة السودان للعلوم والتكنولوجيا .

Repository.sustech.ed/handle/123456789/2842

٩- غانم ، د.عدنان ، الجاعوني ، د.فريد خليل ، (٢٠١١) ، "استخدام تقنية الانحدار اللوجستي الثنائي الاستجابة في دراسة المحددات الاقتصادية والاجتماعية لكفاية دخل الاسرة (دراسة تطبيقية على عينة عشوائية من الاسر في محافظة دمشق)" ، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية ، المجلد السابع والعشرون ، العدد (١) : صفحة ١١٣-١٣٢ .

١٠-قاسم ، م.م.بهاء عبد الرزاق ، (٢٠١١) ، "تحليل اثر بعض المتغيرات في الاصابة بمرض اللثة باستخدام إنموذج الانحدار اللوجستي" ، مجلة العلوم الاقتصادية المجلد السابع ، العدد(٢٧) : صفحة ١٣٩-١٦٤ .

١١- ياسين ، امل حسن محمد ، (٢٠٠٨) ، "استخدام تحليل الانحدار اللوجستي لمعرفة مسببات الاصابة بسرطان الثدي لدى الاناث دراسة حالة المركز القومي للعلاج بالطب النووي بالخرطوم" ، رسالة ماجستير ، كلية العلوم ، جامعة السودان للعلوم والتكنولوجيا .

www.sustech.ed/sudannewar/thesis.php?collno=2&serial=6059

ب- المصادر الاجنبية :

12- Burns , Robert & Burns , Richard , (2008) ," Business Research Methods and Statistics using SPSS", Five extra advanced chapters , chapter 24 Logistic Regression : pp 568 -575.
www.uk.sagepub.com/burns/website
13- El-Habi , Abdalla & El-Jazzar , Majed , (2014) ," A comparative Study between Linear Discriminat Analysis and Multinomial Logistic regression " , An-Najah University Journal Research (Humanities) , Vol.28 , No.6 : pp 1526 - 1548.
14- Hosmer , David W. & Lemeshow , Stanley , (2013) , "Applied Logistic Regression" , 3rd edition , NewYork :wiley
http://ihmsi.org/
15- H. Yusuff , N. Mohamad , U.K. Ngah & A.S. Yahaya , (2012) , "Breast Cancer Analysis Using Logistic Regression" , IJRRAS ,Vol 10 , No.1 : pp 14 – 22.
www.arpapress.com/Volumes/Vol10Issue1/IJRRAS_10_1_02.pdf
16- Logistic regression – MedCalc
www.medcalc.org/manual/logistic_regression.php
17-Müller , Marlene , (2004) , " Generalized Linear Models " , Fraunhofer Institute for Industrial Mathematics (ITWM) , (Germany) .
www.marlenemueller.ed/publications/HandbookCS.pdf
18-Wuensch , Karl L. , (2014) ,"Binary Logistic Regression with SPSS".
www.care.ecu.edu/psyc/wunesch/MV/Multerg/logistic_SPSS.

Using the Logistic Regression Model in Studding the Assistant Factors to Diagnose Bladder Cancer

Abstract:

The cancer is one of the biggest health problems that facing the world . And the bladder cancer has a special place among the most spread cancers in Arab countries specially in Iraq and Egypt[(2)] . It is one of the diseases which can be treated and cured if it is diagnosed early . This research is aimed at studying the assistant factors that diagnose bladder cancer such as (patient's age , gender , and other major complains of hematuria , burning or pain during urination and micturition disorders) and then determine which factors are the most effective in the possibility of diagnosing this disease by using the statistical model (logistic regression model) and depending on a random sample of (128) patients . After collecting and analyzing concerned collected data , the research concluded that the factor of (hematuria) has been found the most effective in the possibility of diagnosing this disease , followed by the factor (patient's age) . Other factors (gender , burning or pain during urination and micturition disorders) were found not significant effective in the possibility of diagnosing this disease . The research ended with some recommendations in this concern .

Keywords: logistic regression ، bladder cancer .