استعمال بعض الاساليب الاحصائية لدراسة امراض السرطان في العراق

أ.م.د. احمد ذياب احمد
ahmedthieb19@gmail.com
كلية الادارة والاقتصاد / جامعة بغداد

م.د. ابتسام كريم عبدالله
ebtisamsa@yahoo.com

أ.م. بيداء اسماعيل عبدالوهاب
baidaa_29@yahoo.com

الملخص: يعتبر السرطان من الامراض التي تصيب انواع مختلفة من اعضاء الجسم وتختلف اعراض هذا المرض من حالة الى اخرى ، ويعد سرطان الثدي الأكثر انتشاراً في العراق والعالم ومن الاسباب الرئيسية لوفاة الأناث ، وفي بحثنا تم تحليل بيانات مرض سرطان الثدي في العراق للاناث والذكور بالاعتماد على بعض الاساليب الاحصائية (التحليل العاملي وتحليل الانحدار اللوجستي) باستعمال البرنامج الاحصائي SPSS ، اذ بينت طريقة المركبات الرئيسية وجود سبعة مركبات رئيسية تؤثر على امراض السرطان وقد فسرت (64.688%) من التباين الكلي وتم بناء انموذج الانحدار اللوجستي بحالتين ، الاولى بالاعتماد على متغيرات مشتقه من التحليل العاملي من خلال اختيار المركبات التي تضم متغيرات تشبعاتها عاليه نسبيا والثانية بالاعتماد على كل المتغيرات.

الكلمات المفتاحية: التحليل العاملي، الانحدار اللوجستي، مرض سرطان الثدي.

"Using some Statistical Methods to Study Cancer Diseases in Iraq"

Ahmed Diyab Ahmed
Ibtisam Karem Abdula
Baydaa Ismaeel Abdulwahab

Abstract:
Cancer is a disease that most parts of the human body are infected with. It is also considered one of the leading causes of death for females in Iraq and all over the Globe. In this research we have analyzed a data of breast cancer for male and female by using some statistical methods (factor analysis and Logistic regression analysis) by using the SPSS packages. The principal components method has shown seven major compounds affecting cancer diseases, and it explains %64.688 of the total variations. The logistic regression model was constructed in two cases. The first case

that the variables depended on was derived from factor analysis through the selection of components, which includes relatively high loading while the second case all the variables were taken.

Keyword: Factor Analysis, Logistic Regression, Breast cancer disease.

1-**المقدمة:** يواجه الانسان العديد من المشاكل الصحية ومنها أمراض السرطان والذي يعد السبب الرئيسي للوفاة في دول العالم ومنها العراق ، وأن معدل الاصابة بهذا المرض يتزايد في الدول النامية وذلك لقلة السيطرة عليه مقارنة بالأمراض الأخرى ، ويعد مرض سرطان الثدي الأكثر انتشاراً وسبب رئيسي لوفاة الأناث في العالم اذ يشكل (23%) من المجموع الكلي لحالات السرطان و (14%) من الوفاة بالمرض. وفي العراق يعد مرض سرطان الثدي الأكثر شيوعاً اذ قدرمجلس السرطان العراقي عام (2008) نسبة هذا المرض حوالي (19.25) من مجموع امراض السرطان اذ بلغ عدد المصابين بهذا المرض من الرجال والنساء حوالي (2729). بصورة عامة يمكن القول ان فرص بقاء مرضى السرطان على قيد الحياة يكاد يكون ضئيل وهذا يعود الى اسباب (عوامل) كثيرة منها عمر الشخص أو الجنس أو الهرمونات والجينات أوالأصابة بأمراض سابقة، والسيطرة على هذه العوامل مهم للوقاية اوالحد من الاصابة بسرطان الثدي.

ولعلم الاحصاء بطرائقه وأساليبه التحليلية اهمية بوضع هذه العوامل بعلاقات ودراسة مدى تاثيرها على المرضى المصابين بسرطان الثدي لجعل هذه العوامل او المتغيرات واضحة للمرضى اولا وللاطباء لاعداد دراساتهم الطبية ومساعدة الحكومة في اعداد برامجها في مجالات التخطيط الصحي والرعاية الصحية لرفع المستوى الصحي للمجتمع والحد من انتشار الأمراض ثانيا. تناول هذا البحث اسلوبين احصائيين مختلفيين في التحليل، هما التحليل العاملي والتحليل باستعمال اسلوب الانحدار اللوجستي لدراسة مرض سرطان الثدي في العراق، وقد تبين انه لايوجد تاثير او علاقه بين الاسلوبين وان كل واحد منهم يستعمل لاغراض مختلفه، فالتحليل العاملي يعد اداة لاختزال عدد المتغيرات المدروسه ويركز على المتغيرات المستقله بشكل كامل في حين يدرس الانحدار اللوجستي العلاقه بين المتغيرات المستقله وتاثيرها على المتغير التابع ثنائي التفرع، وقد تم بناء انموذج انحدار لوجستي بالاعتماد على متغيرات مشتقه من التحليل العاملي من خلال اختيار المركبات التي تضم متغيرات تشبعاتها عاليه نسبيا، وكذلك بالاعتماد على كل متغيرات الدراسة ، ويمكن القول ان التحليل باستعمال الانحدار اللوجستي لايتاثر كثيرا سواء مع المتغيرات المستقله كاملة في الانموذج ام اختزلت باسلوب التحليل العاملي.

2-التحليل العاملي: (Factor Analysis)

هو اسلوب إحصائى يستعمل بيانات متعددة وتوجد علاقة ارتباط فيما بينها بدرجات مختلفة ، ويهدف اسلوب التحليل العاملي الى اختصار المتغيرات الى عدد اقل تسمى العوامل بحيث يكون لكل عامل من هذه العوامل دالة تربطه مع جزء من هذه المتغيرات او كلها ومن خلال هذه الدالة يمكن تفسير العوامل ، من هذا يتضح ان فكرة التحليل العاملي تستند على استخلاص مجموعة من العوامل ترتبط بالمتغيرات الاصلية بحيث تفسر هذه المتغيرات اكبر نسبة من التباين[3]، وهناك نوعان من التحليل العاملي هما[2] التحليل العاملي الإستكشافي (Exploratory Factor Analysis) ويستعمل في الحالات التي تكون فيها العلاقات بين المتغيرات والعوامل الكامنة غير معروفة ، والاخر هو التحليل العاملي التوكيدي (Confirmatory Factor Analysis) ويستعمل لاختبار الفرضيات المتعلقة بوجود أو عدم وجود علاقة بين المتغيرات والعوامل الكامنة وكذلك في تقييم قدرة انموذج العوامل على التعبير عن مجموعة البيانات الفعلية.

من الممكن استعمال التحليل العاملي لتحويل مجموعة المتغيرات المرتبطة الى مجموعة مستقلة ترتبط مع المجموعة الاولى بعلاقات خطية ، وان المعادلات التي تربط العوامل بالمتغيرات الاصلية تكون بالصيغة الاتية:

$$F_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$$
$$F_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2p}x_p$$
$$. \qquad \text{..... (1)}$$
$$.$$
$$F_q = a_{q1}x_1 + a_{q12}x_2 + ... + a_{qp}x_p$$

للتحليل العاملي طرائق مختلفة وسيتم توضيح طريقة واحدة هي طريقة المكونات الرئيسية

3-طريقة المكونات الرئيسية[3]: (Principal Components Method)

تعد طريقة المكونات الرئيسية من اكثر الطرائق استعمالا ودقة في نتائجها وتهدف الى تحويل المتغيرات الاصلية الى متغيرات جديدة بحيث يكون المتغير عبارة عن تركيب خطي (Linear Combination) من المتغيرات التي يتم دراستها وكل عامل يفسر اكبر ما يمكن من التباين وتؤدي الى تشبعات دقيقة ، ويتم حساب المكونات الرئيسية باستعمال مصفوفة التباين المشترك اذ تكون المتغيرات مقاسة بالانحرافات عن الوسط الحسابي وكذلك باستعمال مصفوفة الارتباطات للمتغيرات وفيها تكون المتغيرات معيارية عند اختلاف وحدات القياس

للمتغيرات ، وللحكم على قابلية مصفوفة الارتباط للتحليل العاملي الاستكشافي هناك معايير عدة منها[2]:

- اختبار بارتليت (Bartlett's Test): من خلاله يمكن معرفة طبيعة العلاقة بين المتغيرات وان مستوى الدلالة لهذه العلاقة تكون اقل من (0.50) للحكم على معنوية العلاقة بين المتغيرات.

- مقياس كايزر ماير اولكن ((Kaiser – Mayer – Olkin (KMO)): ويستعمل لقياس مدى كفاية عدد افراد العينة ومدى ملائمة مصفوفة الارتباط وتكون قيمته اكبر من (0.50).

4-تدوير العوامل[3]: (Factors Rotation)

عند استعمال التحليل العاملي لمصفوفة الارتباط ة فانه سيتم التوصل إلى مجموعة من العوامل والتي هي عبارة عن محاور متعامدة تمثل تشبعات المتغيرات واحداثياتها وتتحدد هذه المحاور بطريقة عشوائية ويختلف تحديد المحاور من طريقة عاملية لأخرى. اذا الهدف من تدوير العوامل هو للوصول الى تشكيلة من العوامل بحيث يمكن تفسيرها ، ويوجد نوعان من التدوير هما التدوير المتعامد (Orthogonat Rotation) وفيه تدار العوامل معا مع الاحتفاظ بالتعامد بينها، والتدوير المائل (Oblique Rotation) وفيه تدار المحاور دون الاحتفاظ بالتعامد وتترك لتتخذ الميل المناسب لها ، والعوامل المتعامدة غير المرتبطة معا هي معاملات الارتباط التي قيمها تساوى صفر، اما العوامل المائلة فهي عوامل بينها أرتباط اي متداخلة.

5-درجة الشيوع والتشبع: (Communality and Loading)

تعرف درجة شيوع المتغير (التشبعات) بأسهامات المتغير في كل العوامل وهو مجموع مربعات معاملات المتغير في العوامل المختلفة ، فمثلا درجة شيوع المتغير (j) يكون بالصيغة الاتية:

$$C_j = \sum_{i=1}^{m} a_{ij}^2 \qquad \text{..... (2)}$$

اذ ان المعامل (a_{ij}) يعرف بمعامل التشبع او معامل تحميل المتغير (i) على العامل (j) او هي ارتباط العامل بالمتغير، ان مجموع مربعات درجات التشبع لكل عامل تسمى الجذر الكامن (Eigen Values) والتي تعبر عن اهمية العامل في تفسير الاختلافات في المتغيرات ، وان مجموع الجذور الكامنة يعبر عن التباين الذي تم تفسيره من العوامل ، ويتم قبول العامل الذي قيمته اكبر من الواحد الصحيح.

6-**الانحدار اللوجستي[1,5]: (Logistic Regression)**

وهو طريقة نمذجة احصائية للبيانات الفئوية (categorical data) وفيه يكون متغير الاستجابة المتغير التابع من النوع الثنائي (Binary or Dichotomous) ويستعمل في البحوث الطبية وغيرها ، وفي انموذج الانحدار اللوجستي يتم استعمال احصاءة (Nagelkerke R^2) واحصاءة (Cox & Snell R^2) لتفسير العلاقة بين المتغير التابع والمتغيرات المستقلة ويحسبان وفق الصيغ الاتية[4]:

$$R^2_{Cox\&Snell} = 1 - \left[\frac{L_0}{L_1}\right]^{\frac{2}{n}} \quad \text{..... (3)}$$

$$R^2_{Nagelkerke} = \frac{R^2_{Cox\&Snell}}{1 - [L_0]^{\frac{2}{n}}} \quad \text{..... (4)}$$

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

L_1 : تمثل دالة الامكان الاعظم في حالة الانموذج المتضمن كل المتغيرات التوضيحية.

n: حجم العينة.

ومن الاختبارات الاحصائية المستعملة بأنموذج الانحدار اللوجستي (احصاءة Wald) وتتبع توزيع مربع كاي بدرجة حرية واحدة لكل معلمة وتحسب وفق الصيغة الاتية:

$$Wald = \left[\frac{\hat{B}_i}{S.E(\hat{B}_i)}\right]^2 \quad \text{..... (5)}$$

$\hat{B}_i$: تمثل المعلمة المقدرة ذات الرتبة (i).

$S.E\left(\hat{B}_i\right)$: يمثل الخطأ المعياري للمعلمة المقدرة ذات الرتبة (i).

اذا كانت القيمة الاحتمالية (p-value) ا قل من (0.05) دل ذلك على ان المعلمة معنوية ولا تساوي صفر في المجتمع الذي سحبت منه العينة.

اما الاختبار الاخر فهو احصاءة (Hosmer and Lemeshow) وتتبع توزيع مربع كاي لبيان تمثيل الانموذج للبيانات بشكل مناسب من عدمه ، فاذا كانت الاحصاءة اكبر من (0.05) عندئذ يكون الانموذج ممثلا للبيانات بشكل جيد.

7-**البيانات:** تم أخذ البيانات والمتعلقة بالمرضى الذكور والاناث من مستشفى الأمل للعلاج الأشعاعي في محافظة بغداد والتابعة لوزارة الصحة العراقية اذ تعد هذا المستشفى الأكثر تسجيلًا وعلاجاً لمرض سرطان الثدي في المحافظة ، وحددت العينة من المرضى المصابين

بسرطان الثدي بحجم (n = 70) وان متغيرات الدراسة موضحه كما في الجدول رقم (1) ادناه.

جدول رقم (1) يمثل متغيرات الدراسة

ت	المتغير	الوصف	ت	المتغير	الوصف
1	الجنس Sex	ذكر/انثى	41	ER	مستقبلات الاستروجين Positive/Negative
2	العمر Age	يمثل اعمار المرضى (Year)	51	PR	مستقبلات البروجستيرون Positive/Negative
3	الوزن Weight	يمثل اوزان المرضى (Kg)	61	HER-2	عامل النمو الجلدي البشري Positive/Negative
4	الطول Height	يمثل اطوال المرضى (c.m)	71	Hb	مستوى هموغلوبين الدم g/dL
5	BMI	مؤشر كتلة الجسم	81	WBC	مستوى كريات الدم البيضاء $X10^9/L$
6	catBMI		19	Blood Urea Nitrogen (BU)	مستوى نتروجين يوريا الدم mg/dL
7	السكن Residence	السكن للمرضى	20	Creatinine (Cr)	مستوى الكرياتين Mmol/L
8	الوظيفة Job	موظف/غيرموظف (اخرى)	12	شكل وحجم الكبد Liver disease	الكبد طبيعي /الكبد غير طبيعي
9	التدخين Smoke	مدخن/غير مدخن	22	شكل وحجم الكلية Kidney disease	الكلية طبيعية/الكلية غير طبيعية
10	الكحول Alcohol	يشرب/لايشرب	23	بداية الشكوى Beginning of complaint	الفترة التي بدا المريض يعاني فيها من بعض الاعراض Months
11	الحالة	متزوج/غير	24	Stage of C50	مرحلة الاصابة بمرض

	الاجتماعية Marital Status	متزوج			سرطان الثدي
12	القدرة على الانجاب Kids	لديه اطفال/ليس لديه اطفال	25	Nitrogen	النايتروجين
31	سجل عائلي بامراض السرطان Family history	متغير يمثل فيما اذا كانت هناك اصابات سابقة لدى عائلة المريض	26	اصابة سابقة بامراض السرطان Patients history	متغير يمثل فيما اذا كان المريض مصاب باحد امراض السرطان سابقا

وقد تم تحليل البيانات باسلوبين:

7-1- اسلوب التحليل العاملي: بعد ان تم ادخال المتغيرات في البرنامج الاحصائي (SPSS) وتحليلها تبين ان هناك ارتباطات عالية بين مؤشر كتلة الجسم (BMI) ووزن الجسم (kg) وكذلك بين مستوى نتروجين يوريا الدم (mg/dL) والنتروجين (Nitrogen) اثرت على وجود مشكلة الارتباط الذاتي وبذلك تم استبعاد هذه المتغيرات الاربعة للتخلص من هذه المشكلة اذ كانت قيمة محدد المصفوفة تساوي (0.001) وهي اكبر (0.0001) ، كما ان قيمة مقياس (KMO = 0.539) ومستوى معنوية اختبار بارتليت (sig = 0.000) وهذا يؤكد كفاية عدد افراد العينه ودقة التحليل من ان هذه المتغيرات ذات دلاله احصائيه على التوالي، ويمثل الجدول رقم (2) تباين المكونات والجذور الكامنة للمصفوفات اذ نلاحظ ان هناك سبعة عوامل جذورها اكبر من او تساوي الواحد الصحيح وان مجموع ما تفسره هذه المكونات يشكل (64.688) من التباين الكلي.

جدول رقم (2) تباين المكونات والجذور الكامنة للمصفوفات

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.240	15.431	15.431	3.240	15.431	15.431
2	2.637	12.558	27.988	2.637	12.558	27.988
3	1.921	9.149	37.137	1.921	9.149	37.137
4	1.748	8.323	45.461	1.748	8.323	45.461
5	1.542	7.344	52.804	1.542	7.344	52.804
6	1.386	6.601	59.406	1.386	6.601	59.406
7	1.109	5.283	64.688	1.109	5.283	64.688
8	.986	4.694	69.383			
9	.926	4.410	73.793			
10	.873	4.158	77.951			
11	.745	3.546	81.498			
12	.650	3.096	84.593			
13	.558	2.657	87.251			
14	.488	2.325	89.576			
15	.476	2.266	91.842			
16	.437	2.079	93.921			
17	.381	1.816	95.736			
18	.316	1.507	97.243			
19	.244	1.163	98.407			
20	.189	.900	99.307			
21	.146	.693	100.000			

يمثل الجدول رقم (3) مصفوفة تشبعات المكونات بعد التدوير (**Rotated Component Matrix**) وان المتغيرات الاكثر تشبعا في هذه العوامل هي (العامل الاول ان متغير (PR = 0.871) والعامل الثاني (Alcohol = 0.758) والعامل الثالث (Mmol/L = 0.703) والعامل الرابع (Marital Status = 0.854) والعامل الخامس (Job = 0.710) والعامل السادس (catBMI = 0.801) والعامل السابع (Family history with cancer = 0.810)) والمؤثرة على مرض سرطان الثدي.

جدول رقم (3) مصفوفة تشبعات المكونات بعد التدوير

	Component						
	1	2	3	4	5	6	7
Sex	-.101-	.645	.155	-.128-	.392	.293	-.196-
Years	.253	.615	.182	-.310-	-.233-	-.180-	.169
c.m	.314	.208	.098	.020	.281	.683	-.105-
catBMI	-.191-	-.062-	-.034-	-.030-	-.193-	.801	.019
Residence	-.017-	-.192-	-.477-	.135	.171	-.103-	-.080-
Job	-.163-	.221	-.041-	.120	.710	-.128-	.017
Smoke	-.056-	.649	.097	.010	.128	.164	.283
Alcohol	-.144-	.758	.046	.117	.176	-.121-	-.281-
Marital Status	.045	-.040-	.015	.854	-.013-	-.026-	.146
Kids	-.116-	-.031-	.036	.851	-.097-	.011	-.094-
Family history with cancer	-.040-	-.005-	.068	.030	.133	-.055-	.810
ER	.768	-.136-	-.118-	.113	-.090-	-.071-	.026
PR	.871	-.042-	-.024-	-.040-	-.119-	.068	.100
HER2	.853	.097	.104	-.044-	-.122-	.039	-.128-
g/dL	.260	.205	-.487-	.246	.042	.243	.137
x10^9/L	.045	.171	.621	.330	.248	-.066-	-.102-
Mmol/L	-.129-	.023	.703	.115	.318	.305	-.027-
Liver disease	.118	-.403-	.465	-.068-	.273	-.101-	-.298-
Kidney disease	-.019-	.024	.097	-.201-	.660	.095	.150
The beginning of the complaint in months	-.062-	.142	.661	.057	-.120-	-.070-	.302
Stage	-.608-	.092	.203	.234	-.305-	.108	.132

7-2- اسلوب تحليل الانحدار اللوجستي: تم تحليل البيانات باسلوبين ، **الاسلوب الاول** بالاعتماد على نتائج التحليل العاملي اذ يمثل (Patients history) المتغير المعتمد (تاريخ المريض مصاب او غير مصاب) والبقية تمثل المتغيرات المستقلة فالجدول رقم (4) يبين معنوية احصاءة اختبار مربع كاي وكفاءة الانموذج وجودته والتي تساوي ($x^2 = 48.988$).

جدول رقم (4)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	48.988	21	.001
	Block	48.988	21	.001
	Model	48.988	21	.001

يشير الجدول رقم (5) الى ان الانموذج جيد من خلال قيمة (Nagelkerke R^2 = 0.796) اما قيمة (Cox & Snell R^2) تعني ان احتمالية التغير في المتغير التابع كونه مصاب مسبقا بالسرطان قد فسرت من قبل الانموذج اللوجستي بمقدار (50%).

جدول رقم (5)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	21.069[a]	.503	.796

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

ومن اجل الكشف عن انحرافات الانموذج اللوجستي ومدى الجوده والتوافق بين القيم المشاهده والقيم الحقيقيه تم حساب احصاءة (Hosmer and Lemeshow) وكانت قيمة (sig=0.678>0.05) كما في الجدول رقم (6) وتعني عدم وجود فروق معنويه بين القيم المشاهده والقيم المتوقعه.

جدول رقم (6)

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.724	8	.678

اما الجدول رقم (7) فيشير الى ان النسبه المئويه للتصنيف الصحيح تساوي (91.4 %) وهذا يعني ان احتمال الخطأ الكلي يساوي (8.6 %) اي ان الانموذج يمثل البيانات بشكل جيد.

جدول رقم (7)

Classification Table [a]

Observed			Predicted		
			Patients history with cancer		Percentage Correct
			No	Yes	
Step 1	Patients history with cancer	No	53	3	94.6
		Yes	3	11	78.6
Overall Percentage					91.4

a. The cut value is .500

يبين الجدول رقم (8) تقدير معلمات الانموذج الامثل والخطا المعياري لكل معلمه واحصاءة (Wald) في العمود الثالث لاختبار معنوية المعلمات وتبين قيمة (sig < 0.05) ان جميع المعلمات (Bi =0,i=1,2,…,21) غير معنويه في المجتمع الذي سحبت منه العينه اما العمود (Exp(B)) فيمثل نسبة الارجحيه للمتغيرات المستقله اي احتمال وقوع الحدث الى احتمال

عدم وقوعه عندما يتغير المتغير المستقل المرتبط بالمعلمه B، وان معادلة انموذج الانحدار اللوجستي ستكون بالصيغة الاتية :

$$\hat{Y} = -5.330 + 29.721X_1 + 1.591X_2 - 0.719X_3 + 115.061X_4 - 2.442X_5 + \ldots + 34.189X_{19} - 0.498X_{20} + 7.484X_{21}$$

جدول رقم (8)

Variables in the Equation

Step 1[a]	B	S.E.	Wald	Df	Sig.	Exp(B)
Sex	29.721	21.252	1.956	1	.162	8.082E12
Age	1.591	1.355	1.378	1	.240	4.907
Height	-.719-	.526	1.872	1	.171	.487
catBMI	115.061	90.476	1.617	1	.203	9.343E49
Residence	-2.442-	1.641	2.216	1	.137	.087
Job	78.112	62.009	1.587	1	.208	8.386E33
Smoke	53.680	41.238	1.695	1	.193	2.056E23
Alcohol	-220.984-	17899.810	.000	1	.990	.000
Marital Status	-1.532-	3.011	.259	1	.611	.216
Kids	25.856	17.583	2.162	1	.141	1.695E11
Family history	-47.258-	38.533	1.504	1	.220	.000
ER	-27.677-	18.946	2.134	1	.144	.000
PR	-3.944-	6.755	.341	1	.559	.019
HER2	-26.254-	21.672	1.468	1	.226	.000
Hb	8.931	7.340	1.480	1	.224	7561.213
WBC	4.340	3.374	1.654	1	.198	76.685
Cr	-1.108-	.838	1.748	1	.186	.330
Liver disease	-93.129-	6905.607	.000	1	.989	.000
Kidney disease	-34.189-	27.027	1.600	1	.206	.000
Beginning of complaint	-.498-	.525	.900	1	.343	.608
Stage	7.494	8.330	.809	1	.368	1796.476
Constant	5.330	19185.055	.000	1	1.000	206.456

a. Variable(s) entered on step 1: Sex, Age, Height, catBMI, Residence, Job, Smoke, Alcohol, Marital Status, Kids, Family history, ER, PR, HER2, Hb, WBC, Cr, Liver disease, Kidney disease, beginning of complaint, Stage.

الاسلوب الثاني بالاعتماد على جميع المتغيرات(25) متغير مستقل بعض النظر عن النتائج التي تم الحصول عليها من التحليل العاملي اذ كانت احصاءة اختبار مربع كاي وكفاءة الانموذج وجودته تساوي (x^2 = 70.056) ، وان الانموذج جيد من خلال قيمة

(Nagelkerke R^2 = 1.000) اما قيمة (Cox & Snell R^2) فقد فسرت من قبل الانموذج اللوجستي بمقدار (63%) ، كما تم حساب احصاءة (Hosmer and Lemeshow) وكانت قيمة (sig=1.000>0.05) وتعني عدم وجود فروق معنويه بين القيم المشاهده والقيم المتوقعه ، وان النسبه المئويه للتصنيف الصحيح تساوي (100 %) اي ان الانموذج يمثل البيانات بشكل ممتاز، وان معادلة انموذج الانحدار اللوجستي هي:

$$\hat{Y} = -1224.005 - 37.055\ X_1 + 1.973X_2 - 4.513X_3 + 8.535X_4 + \ldots - 63.410X_{25}$$

8- الاستنتاجات:

1- التخلص من مشكلة الارتباط الذاتي وذلك باستبعاد المتغيرات ذات الارتباطات العالية (مؤشر كتلة الجسم (BMI) ووزن الجسم (kg) ومستوى نتروجين يوريا الدم (mg/dL) والنتروجين (Nitrogen)).

2- اكد مقياس (KMO) واختبار بارتليت دقة التحليل بالنسبه للتحليل العاملي.

3- اختزلت المتغيرات والتي تؤثر على مرض سرطان الثدي بطريقة المركبات الرئيسية الى سبعة عوامل وان المتغيرات الاكثر تشبعا هي (في العامل الاول متغير (PR) والعامل الثاني (Alcohol) والعامل الثالث (Mmol/L) والعامل الرابع (Marital Status) والعامل الخامس (Job) والعامل السادس (catBMI) والعامل السابع (Family history with cancer).

4- ان العلاقة بين المتغير التابع (Patients history) والمتغيرات المستقلة كانت قوية في الحالتين اي بالاعتماد على التحليل العاملي او بدونه ، وانه لاتوجد فروق معنويه بين القيم المشاهده والقيم المتوقعه.

5- كان انموذج الانحدار اللوجستي الذي تم تقدير معلماته انموذجا ممثلا للبيانات اذ اجتاز اختبار (Cox & Snell R2) واختبار (Hosmer and Lemeshow) ، وكان تصنيف المشاهدات للانموذج اللوجستي بشكل جيد.

9- المصادر:

1- سعيد، رشا عادل، (2015) ، "استخدام انموذج الانحدار اللوجستي في دراسة العوامل المساعدة على تشخيص حالات الاصابة بسرطان المثانة"، *مجلة العلوم الاقتصادية والادارية*، المجلد (21) ، العدد (83).

2- عباس، هديل خضير، (2017) ، "*التحليل الاحصائي لواقع المعاقين في العراق لسنة (2013) بالاعتماد على بيانات خارطة الفقر ووفيات الامهات*"، بحث دبلوم عالي، كلية الادارة والاقتصاد/ جامعة بغداد.

3- عبد الكريم، نور كمال، (2016)، "*العوامل المؤثرة على انتاج بعض المحاصيل الحقلية في العراق*"، بحث دبلوم عالي، كلية الادارة والاقتصاد/ جامعة بغداد.

4- قاسم، بهاء عبد الرزاق، (2011) ، "تحليل اثر بعض المتغيرات في الاصابة بمرض اللثة باستخدام انموذج الانحدار اللوجستي"، *مجلة العلوم الاقتصادية والادارية*، المجلد (7) ، العدد (27).

5- Burns, Robert and Burns, Richard, (2008), "*Business Research Methods and Statistics using SPSS" five extra advanced chapters*, chapter (24) Logistic Regression. **www.uk.sagepub.com/burns/website**