

Comparison between linear and nonlinear regression models to study the Causes of Death in Preterm babies in Babylon

مقارنة بين نماذج الانحدار الخطية واللاخطية لدراسة اسباب وفيات الاطفال الخدج في بابل

أ.د. عدنان كريم نجم الدين
جامعة كربلاء/ كلية الادارة والاقتصاد/ قسم الاحصاء
ميثم عبد الوهاب صالح
بحث مستل

الخلاصة:

تهدف هذه الدراسة الى المقارنة بين نموذجي الانحدار الخطي واللاخطي لاختيار افضل نموذج يبين المتغيرات المؤثرة في المتغير المعتمد في تحليل الانحدار عن طريق تعيين المتغيرات ذات التأثير المعنوي، وقد تم تطبيق النماذج بعدة طرائق وهي طريقة الحذف الامامي وطريقة الحذف الخلفي وطريقة ادخال جميع المتغيرات والمفاضلة بين النماذج عن طريق معامل الارتباط ومعامل التحديد ومتوسط مربعات الخطأ وقيمة F . شملت بيانات الدراسة تسعة متغيرات توضيحية يعتقد بانها تؤثر في المتغير المعتمد وهو (وفاة الطفل)، جمعت البيانات من دائرة صحة بابل لمدة ثلاث سنوات واخذ مجتمع الدراسة بالكامل . وعند استخدام طريقة الحذف العكسي تبين ان استخدام نموذج الانحدار اللوغارتمي اعطى نتائج للمقاييس الاحصائية افضل من نتائج نموذج الانحدار المتعدد وتبين عن طريق التحليل ان المتغيرات (مدة الحمل ، وزن الطفل ، عمر الطفل ، نوع الولادة (طبيعية او قيصرية) ، ناتج الولاد (مفردة او توأم)) ذات تاثير مباشر في وفيات الاطفال الخدج.

Abstract:

The main goal of this study is to make a comparison between linear and nonlinear regression models in order to choose the best one, The models have been applied in several ways method, Forward, Backward and Enter, was applied to determine the significant explanatory variables affecting the independent variables preterm babies death . An additional statistical tests are utilized for this reason such as coefficient of determination. Goodness of fit test and mean error squares statistic. Babylon health directorate is the source of the data used in this study. One of the main findings of this study appears the logarithmic regression model with better significance and the explanatory variables pregnancy period infant, weight, infant age, type of delivery (caesarean or normal delivery) .

المقدمة

ان دراسة معدلات وفيات الاطفال الخدج تمثل انعكاسا للوضع الصحي العام للمجتمع ، وبخاصة صحة الطفل والام ويتضح ذلك عن طريق العناية بالحامل و الجنين لحين ولادته وفي السنة الاولى من عمره . ان لبيانات وفيات الاطفال الخدج اهمية كبيرة اذ انها تعكس المستوى الصحي في البلد ولا تقتصر على هذا فقط بل تعتبر مؤشرا رئيسيا هاما لمستويات الوفيات ولمدى التطور الاجتماعي والاقتصادي والثقافي في المجتمع ، ومعدل وفيات الخدج يعكس الواقع والصورة الحقيقية للمستوى الصحي والتنموي ويعكس مدى تظافر الجهود من اجل التقليل من نسبة الوفيات في هذه الفئة من العمر والتي تعتبر من اعلى الفئات في الوفيات . تحتل دراسة الوفيات مجالا خاصا في الابحاث الصحية والطبية لاثرها البالغ في النمو والتطور وانشاء جيل صحي لا يعاني من تشوهات وعقبات صحية . وتعد رسم الاتجاهات، ووضع السياسات وتصميم برامج الوقاية والتدخل في مجال الصحة العامة من اهم الخطط التنموية للبلد . وللتعرف اكثر على معنى الوفاة واسبابها نذكر بعض التعريف منها :

- الوفاة تعني الانتهاء التام لجميع مظاهر الحياة في أي وقت بعد حدوث الولادة الحية أي توقف الوظائف الحيوية بعد الولادة دون القدرة على الحياة وعليه فالوفاة هنا لا تشمل وفاة الأجنة

- سبب الوفاة والذي يعرف بأنه "المرض أو الإصابة التي بدأت به سلسلة الأحداث التي تؤدي مباشرة إلى الوفاة، أو الحادث والظروف الذي أدى إلى الإصابة المميتة . وبالاعتماد على التصنيف الدولي للأمراض في تحديد سبب الوفاة "(المركز الوطني للإحصاءات الصحية، 2003) [13]

- وحديث الولادة (new born) هو الطفل الذي يصل عمره لأقل من شهر واحد بعد الولادة. والخديج (Preterm): من وُلد مبكراً أي لمدة حمل أقل من 37 أسبوعاً. وتعد قصر مدة الحمل والأمراض التنفسية والتشوهات الخلقية وقلة الوزن من أهم الأسباب الكامنة لوفيات الأطفال الخدج

وبالرغم من أهمية المعلومات حول الوفيات وأسبابها فإن العديد من بلدان العالم مازالت تعاني إما من النقص في المعلومات الصحية أو النقص في جودة تلك المعلومات ،

مشكلة الدراسة :

نظراً لتزايد عدد الوفيات للاطفال الخدج في السنوات (2014,2015,2016) بات من الضروري دراسة هذه الظاهرة للوقوف على أهم الأسباب التي تؤدي الى نشوئها وتطورها وطرق المعالجة كونها تعد من المؤشرات السلبية على أداء المؤسسات الصحية في البلد .

هدف الدراسة:

تهدف هذه الدراسة الى اجراء مقارنة بين النماذج الخطية واللاخطية لدراسة اهم العوامل المؤثرة في وفيات الاطفال الخدج في محافظة بابل للمدة (2014- 2016) . وتتركز الاهمية التطبيقية لهذه الدراسة هي باختيار افضل نموذج انحدار عن طريق المقارنة بين النماذج المقدره بواسطة تطبيق المعايير والمقاييس الاحصائية المتعارف عليها. وفي ان النموذج المختار يقوم بتحديد اهم العوامل المؤثرة في حياة الاطفال الخدج

حدود الدراسة :

تم الحصول على بيانات الدراسة من سجلات شعبة الاحصاء في دائرة صحة بابل في الاعوام (2014-2016) واخذت بيانات الدراسة بالكامل وتشمل الوفيات الخدج في هذه السنوات الذين دخلوا وحدة العناية المركزة الخاصة بالاطفال الخدج .

المبحث الاول : الاطار النظري لمفهوم تحليل الانحدار وانواع النماذج

سنتناول في هذا الجانب الاطار النظري للنماذج الاحصائية المستخدمة في هذه الدراسة وهما نموذج الانحدار الخطي وانموذج الانحدار اللاخطي، بالاضافة الى طرائق تقدير المعالم وطرائق المفاضلة بين النماذج .

اولا : تحليل الانحدار Regression Analysis

عمد الباحثون والإحصائيون لمدة طويلة إلى استخدام تحليل الانحدار لدراسة أثر المتغيرات المستقلة (المفسرة) في المتغير التابع عن طريق بناء معادلة وذلك بغرض التفسير أو التقدير أو التحكم بقيمة المتغير التابع بدلالة متغير أو أكثر من هذه المتغيرات المستقلة وصولاً الى التنبؤ بمستقبل الظاهرة قيد الدراسة .

والانحدار هو العلاقة الرياضية المقاسة احصائياً بين مجموعة من المتغيرات المستقلة ومتغير تابع. وتمثل المتغيرات المستقلة المؤثرات والمتغير التابع الأستجابة

لغرض اعداد البحوث العلمية لا بد من الاخذ بنظر الاعتبار أن البحث عن الحقيقة ليس فيه اي مجال للتحيز أو عدم الدقة في التحليلات الإحصائية أو تفسيرها . ويعتبر تحليل البيانات وتفسير النتائج والتنبؤ اهم مرحلة في المشروع البحثي لأنها تعطي إجابات عن الأسئلة المطروحة . كما يتطلب ان يكون تحليل البيانات بطريقة متأنية ومدروسة ومن ثم تفسيرها اصف الى ذلك ضرورة تحليل البيانات بشكل صحيح وبصورة مفصلة اذ يمكن للأخريين الوصول الى نفس الاستنتاجات العلمية.

ويكتسب الانحدار [1] أهمية بالغة في تحليل العلاقات الدالية كونه يقدم اربع وظائف رئيسة وهي: الوصف (description) و السيطرة (control) و التنبؤ (prediction) .

ثانيا : نماذج الانحدار الخطية [2] Linear Regression Models

يختص تحليل الانحدار Regression Analysis بوصف ودراسة العلاقة بين المتغيرات بشكل انموذج. فاذا كانت العلاقة بين متغيرين وهما المتغير التابع (Y) و المتغير المستقل (X) فيسمى بانموذج الانحدار الخطي البسيط ، او يدرس أثر عدة متغيرات مستقلة في المتغير التابع ويسمى انموذج الانحدار الخطي المتعدد . والنموذج الخطي يشير الى الكيفية التي يكون فيها كل حد من حدود المعادلة وليس شكل دالة الانحدار :-

$$g(x_1, x_2, \dots, x_p) = B_0 + B_1X_{i1} + B_2X_{i2} + \dots + B_pX_{ip}$$

اي ان دالة الانحدار (g) خطية في المعالم اي ان المشتقات الجزئية لهذه الدالة هي :

$$\frac{\partial g(B_0, \dots, B_p)}{\partial B_i} \quad i= 1, 2, \dots, p$$

تكون خالية من المعالم B_0, B_1, \dots, B_p

1- انموذج الانحدار الخطي البسيط [11] Simple Linear Regression

يعبر عنه دائماً بمتغير معتمد يسمى (متغير الاستجابة) ومتصل خطياً بمتغير مستقل واحد ويسمى (المتغير التفسيري). ويعبر عن هذا النموذج رياضياً بالصيغة الآتية:

$$Y_i = B_0 + B_1 X_i + e_i \quad i = 1, 2, \dots, n$$

اذ ان :-

Y_i : المتغير المعتمد (متغير الاستجابة) dependent variable

X_i : المتغير المستقل (التفسيري) independent variable

B_0 : تمثل قيمة الحد الثابت intercept. وتساوي $E(Y_i)$ عندما $X=0$. وتمثل رياضياً المسافة العمودية لمحور Y بين نقطة الاصل ونقطة التقاطع لخط الانحدار مع محور Y ويشير الى معدل التأثير في Y من قبل كل المتغيرات التي استثنيت من النموذج "كاظم" 1990.

B_1 : ميل خط الانحدار Slope وهو يشير الى نسبة التغير في $E(Y_i)$ لكل تغير بمقدار وحدة واحدة في المتغير التفسيري X .

e_i : عنصر الخطأ ويمثل الاختلاف في Y_i الذي لا تفسره العلاقة الخطية. او الفرق بين قيمة Y الاصلية و \hat{Y} المقدر.

ويرجع وجود عنصر الخطأ الى عدة اسباب منها: [4], [7]

1- اهمال بعض المتغيرات المستقلة والتي لها تأثير بصورة مباشرة او غير مباشرة في المتغير المعتمد Y في النموذج.

2- اخطاء في صياغة النموذج المناسب.

3- اخطاء في تجميع البيانات.

يوفر نموذج الانحدار الخطي القيمة المتوقعة للمتغير العشوائي Y عندما ياخذ X قيمة محددة.

$$E(Y/X = x) = E(B_0 + B_1 X_i + e_i)$$

$$= B_0 + B_1 X_i$$

$$var(Y) = E[B_0 + B_1 X_i + e_i - (B_0 + B_1 X_i)]^2$$

$$E e_i^2 = \delta^2$$

فرضيات النموذج :-

1- هو دالة خطية في X (linear function) بالاضافة الى عنصر الخطأ العشوائي. وان Y يتوزع طبيعياً ويعتمد في توزيعه على توزيع عنصر الخطأ e_i .

$$Y \sim N(B_0 + B_1 X_i, \delta^2)$$

2- قيم X_i هي قيم ثابتة fixed (اي تكون نفسها في جميع العينات) ومستقلة عن عنصر الخطأ الذي تختلف قيمته من عينة الى اخرى.

3- قيم X_i قادرة على اظهار تأثيرها في تغير المتغير التابع Y_i اذ توجد على الاقل قيمة واحدة مختلفة عن باقي القيم.

4- المتغيرات المستقلة مقاسة بدون اخطاء.

5- عنصر الخطأ متغير عشوائي يتوزع طبيعياً بمتوسط (0) وتباين δ^2 يكون ثابت في كل مدة.

$$e_i \sim N(0, \delta^2)$$

$$E(e) = 0 \quad \text{and} \quad var(e_i) = E[e_i - E(e_i)]^2 = E(e^2) = \delta^2$$

وتعني هذه الفرضية ان الاخطاء لا تدخل في تفسير Y ، اي ان عنصر الخطأ e_i يأخذ قيم مختلفة عند كل قيمة لـ X_i ومجموع القيم يساوي صفر. وان تباينه حول وسطه الحسابي يبقى ثابتاً لكل قيمة من X_i وهو ما يسمى بـ (تجانس تباين الخطأ Homoscedasticity)

6- عدم وجود ارتباط ذاتي بين الاخطاء اي ان قيم عنصر الخطأ مستقلة بعضها عن بعض للملاحظات المختلفة. اي عدم وجود علاقة بين قيم e_i السابقة او اللاحقة، بمعنى ان التباينات المشتركة لاطء الملاحظات المختلفة تكون معدومة.

$$E(e_i e_j) = 0 \quad \text{for all } i \neq j$$

$$Cov(e_i, e_j) = 0$$

ان تباين الخطأ δ^2 قد يساوي او لا يساوي التباين حول نموذج الانحدار $\delta^2 x, y$ فاذا كان النموذج المفترض صحيح فان:

$$\delta^2 x, y = \delta^2 e$$

اما اذا كان غير ملائم (suffers from lack of fit) فانه لا يساويه ويقدر بـ (S^2) [31]

2- أنموذج الانحدار الخطي المتعدد [7] The multiple regression model

ويقترض هذا النموذج ان المتغير التابع Y هو دالة خطية لمتغيرين تفسيريين او لسلسلة من المتغيرات التفسيرية (المستقلة) (X_1, X_2, \dots, X_p) بالاضافة الى عنصر الخطأ وتكون صيغته كالاتي:

$$Y_i = B_0 + \sum_{j=1}^p B_j X_{ij} + e_i$$

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_p X_{ip} + e_i$$

$$i = 1, 2, \dots, n \quad , \quad j = 1, 2, \dots, p$$

n : يشير الى حجم العينة

p : يشير الى عدد المتغيرات المستقلة المأخوذة.

اذ ان :

Y : المتغير المعتمد dependent variable وهو متغير عشوائي .

Xi : المتغيرات المستقلة (المفسرة) independent Explanatory Variables وتكون قيم ثابتة fixed Values وغير عشوائية .

ei : عنصر الخطأ العشوائي او المتغير العشوائي(Random Variable) .

B₀ : الحد الثابت (الاعتراضي) intercept .

B₁, B₂, , B_p معاملات النموذج وتمثل المشتقات الجزئية لمتغير الاستجابة Y_i بالنسبة الى X_i. وتشير قيمها الى مقدار التغير الذي يحدث في المتغير التابع عندما تتغير قيمة المتغير التوضيحي وحدة واحدة .

فرضيات النموذج :-

1- Yi دالة خطية في Xi .

2- عدم وجود علاقة خطية تامة او شبه تامة بين اثنين او اكثر من المتغيرات المستقلة.

3- عنصر الخطأ يتوزع طبيعياً بمتوسط (0) وتباين δ^2 لجميع المشاهدات. $ei \sim N(0, \delta^2)$

4- عدم وجود ارتباط بين قيم عنصر الخطأ للمشاهدات المختلفة .

ثالثاً- نماذج الانحدار اللاخطية^[9] Nonlinear Regression Models

تعتبر النماذج الخطية اكثر واقعية وسهولة في تقديرها . ولكن هناك الكثير من الظواهر تسلك سلوك لاخطي اي ان النماذج تكون غير خطية في المعلمات ، لذا نقوم بتحويلها الى نماذج خطية وهنا يمكن الاشارة الى نوعين من النماذج اللاخطية (خطية جوهرياً ولا خطية جوهرياً)

(intrinsically linear and intrinsically nonlinear models) ولا بد الاشارة هنا الى ان الاولى يمكن تحويلها الى نماذج خطية بواسطة اجراء عمليات التحويل واخرى لا يمكن تحويلها الى نماذج خطية .

وتعتمد المشتقة في النموذج اللاخطي على معلمة من المعالم او اكثر ويمكن التعبير عنها بالشكل :

$$Y = f(X; \alpha, \beta, \gamma) + e$$

1- انحدار الدوال متعددة الحدود^{[7], [10]} polynomial regression

وهو حالة خاصة من انموذج الانحدار الخطي العام ويتكون من متغير مستقل واحد فقط ولكنه مضاعف ومتغير معتمد واحد ويمكن التعبير عنه بمعادلة الانحدار الآتية :

$$yi = B_0 + B_1x_i + B_2x_i^2 + B_3x_i^3 + \dots + B_kx_i^k + ei$$

i=1,2,...,n

$$Z_{mi} = X_i^m, \quad i=1,2,\dots,k$$

وللسهولة نعتمد التحويل الآتي :

$$yi = B_0 + B_1Z_{1i} + B_2Z_{2i} + \dots + B_KZ_{Ki} + ei$$

ويمكن تحديد رتبة النموذج (rank of the model) على اساس أعلى أس للمتغير المستقل فاذا كان الانموذج من الدرجة الاولى يسمى بانموذج الانحدار الخطي البسيط ، في حين اذا كان الانموذج من الدرجة الثانية (quadratic form) أي يأخذ الشكل المنحني او يكون الانموذج من الدرجة الثالثة (Cubic form) وهكذا . والغرض من زيادة القوى للمتغير المستقل هو للحصول على نموذج ملائم يمثل العلاقة الانحدارية بين x و y .

2- انموذج الانحدار اللوغارتمي^{[6], [8]} log Regression model

ويسمى ايضاً الانحدار المضاعف The multiplicative Regression ويضم عدة متغيرات مستقلة ويأخذ الشكل الآتي :

$$Y_i = B_0X_{1i}^{B_1}X_{2i}^{B_2} \dots X_{ki}^{B_k} ei \quad i=1,2,\dots,n$$

اذ ان (B₀, B₁, , B_k) تمثل معالم النموذج، وان (X₁, X₂, , X_k) هي المتغيرات التفسيرية . ولتحويله الى النموذج الخطي نأخذ اللوغارتم للطرفين فيكون :

$$\ln Y = \ln B_0 + B_1 \ln X_{1i} + B_2 \ln X_{2i} + \dots + B_K \ln X_{ki} + \ln ei$$

$$\ln e \sim N(0, I\delta^2)$$

e : هو الخطأ العشوائي المضاعف ويتوزع توزيع لوغارتمي

ونفرض ان :

$$\ln Y = y, \ln B_0 = B, \ln X_{1i} = Z_{1i}, \ln X_{2i} = Z_{2i}, \dots, \ln X_{ki} = Z_{ki}, \ln ei = v$$

ويكون النموذج الخطي كما يأتي :

$$y = B + B_1Z_{1i} + B_2Z_{2i} + \dots + B_KZ_{ki} + v$$

رابعاً - طرق تقدير معاملات النماذج Methods of estimation

يمكن تقدير العلاقة بين المتغيرات عن طريق تقدير معاملات الانحدار باستخدام الطرق الرياضية والاحصائية واختبار مدى معنويتها. وهناك عدة طرق لتقدير معالم النماذج واهمها 2 طريقة المربعات الصغرى الاعتيادية (OLS) وطريقة الامكان الاعظم Maximum Likelihood Estimator .

1- طريقة المربعات الصغرى الاعتيادية [12] Ordinary Least Square method (OLS)

لتقدير معالم الانحدار B بطريقة المربعات الصغرى تشير الى ان التقديرات \hat{B} يتم اختيارها ما يؤدي الى مجموع مربعات الانحرافات بين قيم المشاهدات Y_i والقيم المقدرة المتوقعة اقل ما يمكن وتكتب بالشكل :

$$e_i = Y_i - (B_0 + B_1 X_i)$$

نجد المقدرات \hat{B}_0 و \hat{B}_1 عن طريق أخذ المشتقات الجزئية لQ المعطاة في المعادلة السابقة ومساوات المشتقة الى الصفر.

$$\frac{\partial Q}{\partial B_0} = \sum_{i=1}^n -2[y_i - (\hat{B}_0 + \hat{B}_1 x_i)]$$

$$\frac{\partial Q}{\partial B_1} = \sum_{i=1}^n -2x_i[y_i - (\hat{B}_0 + \hat{B}_1 x_i)]$$

وعند تبسيط المعادلتين وباستخدام الحذف والتعويض نحصل على:

$$\hat{B}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

وهناك طرق عديدة للوصول الى قيم المعلمات المقدرة [3]

2- طريقة الامكان الاعظم [5] Maximum Likelihood Estimator

من البديهي ان تخمين قيم المعلمات وفق طريقة المربعات الصغرى OLS لا تتضمن افتراض حول شكل توزيع حد الخطأ العشوائي في النموذج. اما في طريقة الامكان الاعظم نكون بعض الافتراضات حول هذا التوزيع (وغالباً يتوزع طبيعياً) ، ومن ثم يتم تعظيم عينة المشاهدات التي تمثلها البيانات .

اذ ان $Y_i = Y_1, \dots, Y_n$ تمثل مشاهدات المتغير المعتمد Y لعينة حجمها n وان قيم المتغير المعتمد تكون مستقلة الواحدة عن الاخرى وعلى افتراض ان :

$$Y \sim N(BX, \delta^2) \quad ; \quad e \sim N(0, \delta^2)$$

اذ ان (BX) تمثل الوسط (μ)، وان (δ^2) تمثل التباين.

فانه يمكن التعبير عن دالة الكثافة الاحتمالية لكل مشاهدة كالآتي :

$$f(y_1, y_2, \dots, y_n) = f(y_i, \mu, \delta^2)$$

$$L(y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i, \mu, \delta^2)$$

$$L(y_1, y_2, \dots, y_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{\sum_{i=1}^n (y_i - (B_0 + B_1 X_i))^2}{2\sigma^2}}$$

$$\ln(L) = \frac{-n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{\sum (Y_i - (B_0 + B_1 X_i))^2}{2\sigma^2}$$

وبعد الاشتقاق الجزئي للمعلمات ومساوات المشتقة للصفر نحصل على المعادلتين :

$$\sum_{i=1}^n y_i = nB_0 + B_1 \sum x_i$$

$$\sum_{i=1}^n x_i y_i = B_0 \sum x_i + B_1 \sum x_i^2$$

وبحل المعادلتين نحصل على تقديرات المعلمات وهي مساوية للتقديرات التي حصلنا عليها في طريقة المربعات الصغرى الاعتيادية .

اختيار افضل نموذج انحدار^{[1],[2],[4]} Selecting the best regression model

يعتمد تحليل الانحدار على المتغيرات التوضيحية الموجودة في النموذج. ومن المفهوم في تحليل الانحدار أن المتغيرات التفسيرية المهمة فقط تظهر في النموذج والتي لها تأثير أو يعتقد أنها تؤثر في المتغير التابع (y) الذي يمكن التنبؤ به. كذلك ان عملية تحديد افضل نموذج عند وجود عدد كبير من المتغيرات المفسرة وفي وجود عينة كبيرة عملية صعبة، لذا لا بد من اختصار عدد المتغيرات المفسرة لان عملية ادخال جميع المتغيرات عملية مكلفة من ناحية الجهد والوقت والمال. ومن ناحية اخرى يجب ان لا يكون عدد المعالم قليل مما قد يؤدي الى تنبؤات غير واقعية ومتحيزة بمعنى ان لا تكون آلية الاختصار تؤثر على المعلومات المستحصلة، اذ نحصل على المعلومات نفسها كما لو استخدمنا جميع المتغيرات. والصعوبة تكمن في اختيار المتغيرات الداخلة والمتغيرات المستبعدة، بمعنى اننا نفاضل بين ما يضيفه المتغير المستقل في تفسير المتغير التابع وبين ما يضيفه من زيادة في التباين في حال ضعف تاثيره. لذلك هناك معايير خاصة للمفاضلة بين المتغيرات يجب معرفتها قبل الدخول في كيفية اختيار افضل نموذج ومنها:

أ- قيمة معامل التحديد R^2

$$R^2 = \frac{SSR(X_1, \dots, X_p)}{SST} = 1 - \frac{SSE(X_1, \dots, X_p)}{SST}$$

ويحسب كالاتي:

$$SST = \sum (Y_i - \bar{Y})^2; SSR = \sum (\hat{Y} - \bar{Y})^2; SSE = \sum (Y_i - \hat{Y})^2$$

اذ ان:

ونأخذ النموذج الذي له اعلى R^2

ب- قيمة متوسط مربعات الخطأ MSE Mean square error

$$MSE = \frac{SSE(X_1, \dots, X_p)}{n-p}$$

وتحسب كالاتي:

ونأخذ النموذج الذي له اقل MSE.

ج- قيمة اختبار F. F-Test

اختبار المعنوية الكلية للانحدار وتحسب كالاتي:

$$F_c = \frac{MSR}{MSE} \text{ or } F_c = \frac{R^2/p-1}{1-R^2/n-p}$$

في حال $F_c \leq F_t$ اذ ان F_c المحسوبة اقل او تساوي القيمة الجدولية سوف لا نرفض فرضية العدم أي ان المتغيرات (X_1, X_2, \dots, X_p) ليس لها تاثير معنوي مع بعضها في النموذج الانحدار.

اما في حالة $F_c > F_t$ سوف نرفض فرضية العدم (H_0) أي انه يوجد على الاقل متغير واحد له اهمية في تفسير المتغير المعتمد Y_i أي ان لها تاثير معنوي في النموذج الانحدار.

ويمكننا ايضا عن طريق اختبار F اختبار جزء من عدد معاملات الانحدار الكلية وسوف يستخدم بهذا اختبار F الجزئي Partial F test والغرض منه لمعرفة في حال اضافة المتغيرات (X_{k+1}, \dots, X_p) الى الانموذج الذي يحوي (X_1, X_2, \dots, X_k) هل له قيمة معنوية في التنبؤ.

$$F = \frac{MSR(x_{k+1}, x_{k+2}, \dots, x_p | x_1, \dots, x_k)}{MSE(x_1, x_2, \dots, x_p)}$$

كما يمكن استخدام هذا الاختبار لبيان تاثير كل متغير في النموذج الانحدار.

$$F = \frac{MSR(x_j | x_1 \dots x_{j-1} x_{j+1} \dots x_p)}{MSE(x_1, x_2, \dots, x_p)}$$

و غالباً ما يستخدم اختبار t (t-test) لهذا الغرض اذ ان :

$$t_c = \frac{\hat{B}_j}{S(\hat{B}_j)}$$

- قيمة إحصاءة مالو C_p : Mallow 's statistics

تعتمد قيمة هذا المقياس على متوسط مجموع مربعات الأخطاء لقيم البيانات الداخلة في النموذج الذي يضم كل المتغيرات (P) المؤثرة في النموذج وكذلك المتغيرات المراد معرفة تأثيرها m. وتحسب كالاتي:

$$C_p = \frac{SSE(X_1, \dots, X_p)}{MSe(X_1, \dots, X_m)} (n - 2p)$$

اذ ان:

p: عدد المعالم في النموذج ومن بينها B_0 .

n: عدد المشاهدات.

m: عدد المتغيرات التفسيرية

هـ – معيار المعلومات الذاتي (AIC) : Akaike,s Information Criterion

$$AIC = \ln(\hat{\delta}^2) + \frac{n+2p}{n}$$

اذ ان n عدد المشاهدات و p عدد المتغيرات المستقلة ويؤخذ النموذج الذي يعطي اقل قيمة للمعيار.

-معيار المعلومات الذاتي غير المتحيز (AIC_c) : Akaike,s Information Criterion

$$AIC_c = \ln(\hat{\delta}^2) + \frac{n+p}{n-p-2}$$

ويؤخذ النموذج الذي يعطي اقل قيمة للمعيار .

ز – معيار حنان – كوين (HQC) : Hannan–Quinn information criterion

$$HQC = -2L_{max} + 2p \ln(\ln(n))$$

اذ ان : p تمثل عدد المعلمات . و n تمثل عدد المشاهدات .

يكون النموذج الملائم الذي يقابل اقل قيمة للمعيار

وبعد ادخال جميع المتغيرات في النموذج نحاول اختصار عدد المتغيرات ليشمل النموذج المتغيرات المؤثرة والتي تكون ذات دلالة معنوية للحصول على النموذج الافضل وهناك طرائق متعددة لاختيار افضل نموذج (Picking the best model) منها :

1- طريقة الاختيار العكسي The backward elimination procedure

تتلخص هذه الطريقة بادخال جميع المتغيرات المفسرة (p) (مع الحد الثابت) ثم نقوم بحذف المتغيرات غير المعنوية بعد احتساب قيمة F_c لكل متغير او للنموذج (p-1) من المتغيرات ومقارنتها بالجدولية ونقوم بحذف المتغير الذي لديه اقل قيمة ل F_c . وبعد حذف المتغير يصاغ نموذج آخر بالمتغيرات المتبقية ونكرر العملية . اما اذا كانت قيمة F_c اكبر من الجدولية فلا نحذف المتغير لانه معنوي . بالاضافة الى احتساب قيمة إحصاءة مالو C_p و R^2 و MSE .

2- طريقة الاختيار الامامي Forward selection procedure

تبدأ معادلة الانحدار بهذه الطريقة بدون اي متغير توضيحي ، ثم نبدأ باختيار المتغيرات واحد تلو الآخر. ويضاف المتغير بالاعتماد على قيمة F_c ومقارنتها بالجدولية اذ يتم اضافة المتغير الذي له أعلى قيمة F_c . وليس بالضرورة ان تؤدي نتائج الحذف العكسي الى نفس نتائج الحذف الامامي .

3- طريقة الانحدار التدريجي The stepwise regression procedure

وهي عبارة عن دمج بين طريقتي الاختيار الامامي والاختيار العكسي . يتم في الخطوة الاولى اختيار المتغير الذي يملك أعلى F_c من بين باقي المتغيرات المفسرة و اضافته الى النموذج الانحدار . وفي الخطوة الثانية نقوم بحساب F_c الجزئية لباقي المتغيرات بوجود المتغير الذي تم اختياره اذا كانت F الجزئية اكبر من الجدولية نثبت المتغير في النموذج اما اذا كانت F الجزئية اقل من الجدولية لا ندخل المتغير ونتوقف عن الاختيار ، واختيار متغير آخر يملك أعلى F_c ليدخل مع المتغير الاول . والخطوة الثالثة نجري عملية الاختيار العكسي لاختبار المتغيرات المختارة بالاعتماد على قيمة F_c الجزئية .

المبحث الثاني : الجانب التطبيقي والمقارنة بين النتائج

تم تحليل البيانات بواسطة حزمة البرامج spss.v25، عن طريق تطبيق نموذج الانحدار المتعدد ونموذج الانحدار اللوغارتمي في البيانات للحصول على نموذج احصائي يمثل طبيعة هذه البيانات الحقيقية وذلك للتحقق من أدائها في الواقع العملي . ولدراسة العلاقة بين المتغيرات . اعتمدت بيانات من سجلات دائرة صحة بابل لوفيات الأطفال الخدج لمدة (2014- 2016) . تمثلت البيانات بكل من المتغير المعتمد (متغير الاستجابة) و فاة الطفل الخديج . بينما كانت المتغيرات التوضيحية تتمثل بمدة الحمل و وزن الطفل وعمر الطفل وعمر الام و نتائج الولادة (توأم) وسبب الرقود (نقص الوزن و قصر مدة الحمل P07، ذات الرئة الولادي P22 ، تشوهات خلقية Q00-Q99) ونوع الولادة (طبيعية)، غير المعقم (المعقم تعني خديج من داخل المستشفى والغير معقم تعني خديج من خارج المستشفى) كمتغيرات مستقلة تؤثر في حياة الطفل . والجدول الآتي يبين المتغيرات التوضيحية والمتغير المعتمد التي اعتمدت في التحليل .

جدول (1) المتغيرات والرمز المقابل لكل متغير

الرمز	المتغير	الرمز	المتغير
X6	عدد المتوفين للسبب (الولادة التوأم)	Y	الوفيات
X7	عدد المتوفين سبب الوفاة P07	X1	معدل مدة الحمل بـ(الاسابيع)
X8	عدد المتوفين سبب الوفاة P22	X2	معدل وزن الطفل بـ(كغم)
X9	عدد المتوفين سبب الوفاة Q00-Q99	X3	معدل عمر الطفل بـ(الايام)
X10	عدد المتوفين للسبب (غير معقم)	X4	معدل عمر الام بـ(السنوات)
		X5	عدد المتوفين لسبب (ولادة طبيعية)

ان تأثير المتغيرات التوضيحية في المتغير التابع يقاس عن طريق معاملات الانحدار B ، ولمعرفة فيما اذا كان للمتغيرات التوضيحية تأثير في المتغير التابع نفترض الفرضيات الآتية :-

فرضية العدم :- لا يوجد تأثير للمتغيرات التوضيحية في المتغير التابع

$$H_0: B_i = 0$$

الفرضية البديلة :- يوجد تأثير للمتغيرات التوضيحية في المتغير التابع

$$H_1: B_i \neq 0$$

$$i = 1,2,3, \dots$$

1- تطبيق نموذج الانحدار الخطي المتعدد . The multiple regression model

تم تطبيق نموذج الانحدار المتعدد في البيانات واختبار الانموذج بالطرائق الثلاث وهي طريقة ادخال جميع المتغيرات enter والانحدار التدريجي stepwise والحذف الخلفي backward وعن طريق دراسة معايير المفاضلة بين النماذج التي تم التطرق اليها في الجانب النظري. تم اختيار نموذج الانحدار بطريقة الحذف الخلفي لافضليته عن باقي الطرائق ونلاحظ ذلك عن طريق ارتفاع قيمة معامل التحديد ($R^2 = 0.971$) ومعامل التحديد المصحح ($R^2 Adj$) والبالغة (0.965) وكذلك نلاحظ انخفاض قيمة متوسط مربعات الخطأ ($MSE=7.934$) وكذلك انخفاض قيمة المعيار (AIC) اكاكي وقيمتة (3.46) وقيمة المعيار (H.Q) حنان كوين وقيمتة (23.249) اما قيمة F والبالغة (160.907) والتي تمثل معنوية الانموذج الكلي فانها مرتفعة في طريقة الحذف الخلفي عن مثيلاتها في الطرائق الاخرى وكما مبين في الجدول الاتي :

جدول (2) معايير اختيار الأنموذج الملائم للانحدار المتعدد بالطرائق الثلاث

Method	R	R ²	R ² Adj	AIC	CP	H.Q	MSE	F
Enter	0.987	0.973	0.963	3.736	23.65	23.689	8.374	91.719
stepwise	0.964	0.93	0.923	4.077	27.76	27.586	17.376	140.702
bakward	0.985	0.971	0.965	3.46	23.64	23.249	7.934	160.907

جدول (3) جدول المؤشرات الاحصائية للانحدار المتعدد

Model Summary^f

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.987 ^a	0.973	0.963	2.8938627	0.973	91.719	10	25	0.000	
2	.987 ^b	0.973	0.964	2.8381066	0.000	0.008	1	25	0.930	
3	.987 ^c	0.973	0.965	2.7893921	0.000	0.081	1	26	0.778	
4	.986 ^d	0.973	0.966	2.7564659	0.000	0.343	1	27	0.563	
5	.985 ^e	0.971	0.965	2.8168035	-0.002	2.283	1	28	0.142	2.712

من الجدول (3) (Model Summary) نلاحظ قيمة معامل الارتباط المتعدد (معامل التحديد) ($R^2 = 0.971$) والذي يمثل نسبة التغيرات التي تحدث في المتغير التابع والتي تشرحها المتغيرات المستقلة وبعبارة اخرى ان المتغيرات المستقلة تفسر ما قيمته (97.1%) من التغيرات التي تحدث في المتغير المعتمد والمتبقي (2.9) يعزى الى عوامل اخرى منها الخطأ ، وكذلك قيمة Adjusted R-Square وتساوي (0.965) وهي لتصحيح التحيز في المعامل السابق . وهناك ايضا قيمة الخطأ المعياري لانموذج الانحدار الذي تم استخلاصه وهو عبارة عن الانحراف المعياري Standrd Deviation للاخطاء وقيمته (2.8) ، و معامل Durbin-Watson لاختبار الارتباط الذاتي للبقايا Residuals وقيمته (2.712) اي انه لا يوجد مشكلة الارتباط الذاتي بين عناصر الخطأ العشوائي .

جدول (4) جدول التحليل التباين للانحدار المتعدد (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Sig.
5	Regression	7660.208	6	1276.701	160.907	.000 ^f
	Residual	230.097	29	7.934		
	Total	7890.306	35			

ومن جدول (4) (ANOVA) تحليل التباين نجد ان قيمة اختبار ($F = 160.907$) وبمعنوية ($\text{sig} = 0$) وهي اقل من 0.05 ومن ثم نرفض فرضية العدم القائلة بعدم معنوية انموذج الانحدار ، اي عدم رفض الفرضية البديلة وتعني ان انموذج الانحدار معنوي . وكذلك نلاحظ قيمة متوسط مربعات الخطأ (MSE) وقيمته (7.934) . وقد تم عرض قيم المعاملات (B) للمتغيرات في الجدول الاتي :

جدول (5) جدول المعاملات (Coefficients) والمعنوية ومعامل التضخم للانحدار المتعدد

Model		Unstandardized Coefficients		Partial correlation coefficient r_{Y,X_i}	T	Sig.	Collinearity Statistics	
		B	Std. Error				Tolerance	VIF
5	(Constant)	39.373	15.604		2.523	0.017		
	X4	-1.422	0.612	0.375	-2.323	0.027	0.859	1.164
	X6	0.230	0.113	0.915	2.039	0.051	0.161	6.219
	X7	0.742	0.118	0.8	6.303	0.000	0.356	2.806
	X8	0.964	0.105	0.843	9.159	0.000	0.288	3.474
	X9	0.799	0.140	0.51	5.722	0.000	0.738	1.354
	X10	0.444	0.075	0.435	5.890	0.000	0.810	1.234

ومن جدول (5) المعاملات (Coefficients) نلاحظ قيم معاملات الانحدار المتعدد ومعنويتها ومنها يمكن ملاحظة عدم معنوية المعاملات التي قيمة المعنوية اكبر من 0.05 ومنها المتغير (ناتج الولادة توأم) اما اذا كانت قيمة sig. اقل من 0.05 تعني ان للمتغير تأثيراً معنوياً في المتغير المعتمد. وتظهر تأثير المتغيرات (عمر الام ، وسبب الوفاة P07 والذي يمثل نقص الوزن وقلة مدة الحمل ، وسبب الوفاة P22 والذي يمثل عسر التنفس ، وسبب الوفاة Q00-Q99 والذي يمثل التشوهات الخلقية ، وغير المعقم والذي يمثل خديج من خارج المستشفى أو خديج منزلي) في سلوك المتغير المعتمد (الوفيات). نلاحظ من المعاملات انه في حالة زيادة عمر الام يؤدي الى انخفاض في عدد الوفيات وفي حالة زيادة سبب الوفاة P07 يؤدي الى زيادة في عدد الوفيات بمقدار 74.2% اما الزيادة في سبب الوفاة P22 يؤدي الى زيادة في عدد الوفيات بمقدار 96.4% وكذلك زيادة في سبب الوفاة Q00-Q99 سيؤدي الى زيادة في عدد الوفيات بمقدار 78% اما بالنسبة للعامل الخديج (غير المعقم) فان زيادته تؤدي الى زيادة في عدد الوفيات بمقدار 44.4% . وكذلك الجدول يبين قيم معامل التضخم Variance Inflation Factor (VIF) والذي يقوم بقياس الارتباط بين المتغيرات المستقلة (وان وجوده يؤدي الى زيادة او تضخم الاخطاء المعيارية لمعاملات الانحدار ولذا يقوم (VIF) بقياس التضخم في التباين للمعاملات المقدره لكل المتغيرات المستقلة في الانموذج واذا كانت قيمة معامل التضخم تفوق (10) هذا يعني ان المتغير يعاني من تضخم في تباين معاملته وانه مرتبط مع بقية المتغيرات المستقلة (اي حدوث مشكلة التعدد الخطي) وهذا يكون سببا كافيا لإهمال المتغير من التحليل أو استعمال طريقة اخرى .

وبذلك تكون معادلة الانحدار التقديرية :

$$\hat{Y} = 39.373 - 1.422(X_4) + 0.23(X_6) + 0.742(X_7) + 0.964(X_8) + 0.799(X_9) + 0.444X_{10}$$

$$t = (2.523) \quad (-2.323) \quad (2.039) \quad (6.303) \quad (9.159) \quad (5.722) \quad (5.89)$$

2- تطبيق انموذج الانحدار اللوغارتمي : Log Regression model

تم تطبيق انموذج الانحدار اللوغارتمي في البيانات واختبار الانموذج بالطرائق الثلاث وهي طريقة ادخال جميع المتغيرات enter والانحدار التدريجي stepwise والحذف الخلفي backward وعن طريق دراسة معايير المفاضلة بين النماذج التي تم التطرق اليها في الفصول السابقة تم اختيار الانموذج بطريقة Enter وذلك لارتفاع قيمة معامل التحديد ($R^2 = 0.959$) وارتفاع قيمة معامل التحديد المصحح ($R^2_{Adj} = 0.942$) عن باقي الطرائق وكذلك انخفاض قيمة متوسط مربعات الخطأ ($MSE=0.00092$) وانخفاض قيمة ($AIC=5.38$) عن مثيلاتها في الطرائق الاخرى وكما مبين في الجدول الاتي :

جدول (6) معايير اختيار الانموذج الملائم للانحدار اللوغارتمي بالطرائق الثلاث

Method	R	R^2	$R^2 Adj$	AIC	CP	MSE	F
Enter	0.979	0.959	0.942	5.38	4	0.00092	57.75
Stepwise	0.966	0.933	0.927	5.567	5	0.001125	148.6
Bakward	0.975	0.951	0.941	5.59	6	0.00093	93.91

جدول (7) جدول المؤشرات الاحصائية للانموذج اللوغارتمي

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson	
					R Square Change	F Change	df1	df2		Sig. F Change
1	.979 ^a	0.959	0.942	0.03003649	0.959	57.751	10	25	0.000	2.449

من الجدول (7) (Model Summary) نلاحظ قيمة معامل الارتباط المتعدد (معامل التحديد) ($R^2 = 0.959$) والذي يمثل نسبة التغيرات التي تحدث في المتغير التابع والتي تشرحها المتغيرات المستقلة وبعبارة اخرى ان المتغيرات المستقلة تفسر ما قيمته (95.9%) من التغيرات التي تحدث في المتغير المعتمد والمتبقي (4.1) يعزى الى عوامل اخرى منها الخطأ ، وكذلك قيمة (Adjusted R-Square) وتساوي (0.942) وهي لتصحيح التحيز في المعامل السابق ، وهناك ايضا قيمة الخطأ المعياري لانموذج الانحدار الذي تم استخلاصه وهو عبارة عن الانحراف المعياري (Standrd Deviation) للاخطاء وتساوي (0.03)، وهناك ايضا معامل (Durbin-Watson) اختبار الارتباط الذاتي للبقاوي وقيمته (2.45) ، اي انه لا يوجد مشكلة الارتباط الذاتي بين عناصر الخطأ العشوائي .

جدول (8) جدول تحليل التباين للانموذج اللوغارتمي (ANOVA)

Model	Sum of Squares	Df	Mean Square	F	Sig.	
1	Regression	0.521	10	0.052	57.751	.000 ^b
	Residual	0.023	25	0.001		
	Total	0.544	35			

ومن جدول (8) تحليل التباين (ANOVA) نجد ان قيمة اختبار ($F = 57.751$) وبمعنوية ($sig = 0$) وهي اقل من 0.05 ومن ثم نرفض فرضية عدم القائللة بعدم معنوية انموذج الانحدار ، اي عدم رفض الفرضية البديلة وتعني ان انموذج الانحدار معنوي . وقد تم عرض المعاملات للمتغيرات في الجدول الاتي :

جدول (9) جدول المعاملات (Coefficients) والمعنوية ومعاملات التضخم

Model	Unstandardized Coefficients		Partial correlation coefficient r_{Y,X_i}	T	Sig.	Collinearity Statistics		
	B	Std. Error				Tolerance	VIF	
1	(Constant)	0.501	1.350		0.371	0.714		
	X1	0.355	0.876	0.884	0.405	0.689	0.218	4.593
	X2	0.189	0.258	0.853	0.730	0.472	0.272	3.670
	X3	0.034	0.077	0.6	0.435	0.667	0.642	1.557
	X4	-0.462	0.406	0.5	-1.138	0.266	0.759	1.317
	X5	0.156	0.111	0.92	1.401	0.173	0.155	6.447
	X6	0.265	0.093	0.93	2.847	0.009	0.126	7.930
	X7	0.127	0.045	0.9	2.814	0.009	0.175	5.722
	X8	0.224	0.070	0.915	3.178	0.004	0.162	6.161
	X9	0.047	0.029	0.65	1.633	0.115	0.577	1.733
	X10	0.125	0.041	0.7	3.066	0.005	0.509	1.965

ومن جدول (9) (Coefficients) نلاحظ قيم معاملات الانحدار المتعدد ومعنويتها ومنها يمكن ملاحظة عدم معنوية المعاملات التي قيمتها المعنوية اكبر من 0.05 اما اذا كانت قيمة sig. اقل من 0.05 أي ان للمتغير تأثير معنوي في المتغير المعتمد. وتظهر تأثير المتغيرات (ولادة توأم ، سبب الوفاة P07 والذي يمثل نقص الوزن وقلة مدة الحمل ، وسبب الوفاة P22 والذي يمثل عسر التنفس ، وغير المعقم والذي يمثل خديج من خارج المستشفى او خديج منزلي) في سلوك المتغير التابع (الوفيات). ومن المعاملات نلاحظ في حال زيادة عدد الولادات التوأم فانها ستحصل زيادة في عدد الوفيات بمقدار 26.5% والزيادة الحاصلة في سبب الوفاة P07 ستؤدي الى زيادة في عدد الوفيات بمقدار 12.7% وكذلك الحال فان الزيادة في سبب الوفاة Q00-Q99 تؤدي الى زيادة في عدد الوفيات بمقدار 4.7% وايضا الزيادة في العامل (الخديج غير المعقم) تؤدي الى زيادة في عدد الوفيات بمقدار 12.5% . وكذلك من الجدول يمكن ملاحظة قيم معامل التضخم Variance Inflation Factor (VIF) والذي يقوم بقياس الارتباط بين المتغيرات المستقلة (وان وجوده يؤدي الى زيادة او تضخم الاخطاء المعيارية لمعاملات الانحدار ولذا يقوم (VIF) بقياس التضخم في التباين للمعاملات المقدره لكل المتغيرات المستقلة في الانموذج واذا كانت قيمة معامل التضخم تفوق (10) هذا يعني ان المتغير يعاني من تضخم في تباين معاملته وانه مرتبط مع بقية المتغيرات المستقلة (اي حدوث مشكلة التعدد الخطي) وهذا يكون سببا كافيا لإهمال المتغير من التحليل أو استعمال طريقة اخرى . وبذلك تكون معادلة الانحدار التقديرية :

$$\hat{Y} = 0.501 + 0.355X_1 + 0.189X_2 + 0.034X_3 - 0.462X_4 + 0.156X_5 + 0.265(X_6) + 0.127(X_7) + 0.224(X_8) + 0.047(X_9) + 0.125(X_{10})$$

$$t = (0.371) \quad (0.405) \quad (0.73) \quad (0.435) \quad (-1.138) \quad (1.401) \quad (2.847) \quad (2.814) \\ (3.178) \quad (1.633) \quad (3.066)$$

الاستنتاجات :

- عن طريق تحليل البيانات والتطبيق العملي تم التوصل الى الاستنتاجات الآتية :-
- (1) عن طريق تحليل البيانات الداخلة في انموذج الانحدار المتعدد نلاحظ من قيم (*sig.*) وقيم (*T*) معنوية العوامل المؤثرة في حياة الطفل الخديج وهي: (عمر الام ، ونقص الوزن ومدة الحمل، عسر التنفس، التشوهات الخلقية ، وغير المعقم) وان معاملات الارتباط الجزئي بين عدد حالات الوفاة وعدد من العوامل المؤثرة تراوحت بين ($r_{Y,X4}=0.374$) لعمر الام و (0.842) $(r_{Y,X10})$ لعسر التنفس .
 - (2) اما في حالة الانحدار اللوغارتمي نلاحظ من قيم (*sig.*) وقيم (*T*) معنوية العوامل المؤثرة في حياة الطفل الخديج وهي (ولادة توأم ، سبب الوفاة *P07*، وسبب الوفاة *P22*، وغير المعقم) وان ارتفاع معاملات الارتباط الجزئي تبين هذه العلاقة تراوحت بين ($r_{Y,X6} = 0.93$) ولادة توأم و ($r_{Y,X10} = 0.7$) غير معقم .
 - (3) اثبتت النتائج ان النموذجين المستعمله كانت معنوية اعتمادا على معنوية قيمة *F* الحسابية في جدول تحليل التباين ، وان انموذجي الانحدار(المتعدد و اللوغارتمي) لم تعطِ دلالة لبعض المتغيرات التي يعتقد ان لها تأثيراً معنوياً مثل (الولادة الطبيعية ، عمر الطفل ، وزن الطفل) ، اذ كانت قيمة (*t*) المحسوبة اقل من نظيراتها الجدولية عند مستوى معنوية 0.05 . وقد يعزى الى قصر المدة الزمنية التي توفرت عنها البيانات وبالبالغة (3 سنوات) .
 - (4) تشير قيمة معامل التحديد المصحح لانموذج الانحدار المتعدد ($\bar{R}^2 = \% 96.5$) على القدرة التفسيرية العالية للبيانات وقيمته اكبر من مقدار تفسيرالبيانات للانحدار اللوغارتمي والتي بلغت قيمته ($\bar{R}^2 = \% 94.2$) . ومن جهة ثانية اعطى الانموذج اعلى قيمة لمتوسط مربعات الخطأ اذ بلغت ($MSE=7.934$) ، في حين كانت قيمة متوسط مربعات الخطأ في الانموذج اللوغارتمي اقل وبلغت ($MSE=0.001$) .
 - (5) اختيار الانموذج اللوغارتمي لتحليل البيانات لكفاءته في تفسير الظاهرة .واعطائه اقل قيمة لمتوسط مربعات الخطأ ($Mse=0.001$)
 - (6) ساهمت المؤشرات الاحصائية المختارة للمقارنة بين النماذج بشكل جيد في اختيار الأنموذج المناسب .
 - (7) عن طريق تحليل البيانات المستخدمة تبين أن حياة الطفل الخديج مرتبطة بالعوامل (ولادة توأم، اسباب الوفاة: *P07* , *P22* , غير المعقم) أكثر من غيرها .

التوصيات :

- استنادا إلى ما جاء في الجوانب النظرية والتطبيقية للدراسة ،وفي ضوء الاستنتاجات التي تم التوصل إليها ،نوصي بما يلي:-
- (1) ضرورة استخدام الأنموذج اللوغارتمي في الحالات المشابهه لهذه البيانات .
 - (2) اعطاء اهمية في التداخلات الطبية والصحية لاسباب الوفاة وحسب مستوى معنوياتها كما ظهرت لنا في التحليل الاحصائي.
 - (3) 3-- اعتماد النتائج التي تم التوصل اليها في دراسة المستوى الصحي للمستشفيات وتوفير المستلزمات الواجبة للتقليل من هذه الظاهرة او الحد منها .
 - (4) ضرورة الاهتمام بتحديد سبب الوفاة للطفل الخديج ، لان دقة البيانات تؤثر في دقة النتائج التي نحصل عليها .وتحدد حسب التصنيف الدولي للأمراض .
 - (5) ضرورة ادخال متغيرات اخرى مثل(الحالة الصحية للام ونوع العقاقير المستخدمة من قبل الام اثناء مدة الحمل وقبلها ونوع الرضاعة وغيرها) مما يعتقد ان لها تأثير في وفيات الخدج .

المصادر

- 1 - التميمي، زهرة حسن عباس وآخرون ، 2014 ، تحليل الانحدار. مديرية دار الكتب للطباعة والنشر / جامعة البصرة .
- 2 - بري، عدنان ماجد عبد الرحمن، 2003 ، تحليل الانحدار الخطي، جامعة الملك سعود
- 3- نجم الدين ، عدنان كريم، 2017 ، محاضرات في القياس الاقتصادي ،
- 4- Alan Miller, 2002, Subset Selection in Regression, A CRC Press Company- Boca Raton London New York Washington, D.C. Second Edition.
- 5- Alvin. C. Rencher ,2002, Methods of multivariate analysis, Brigham Young University, Copyright by John Wiley & Sons, Inc, Second Edition
- 6- Bolch ,B.W, & Hung, C.J (1974) Multivariate statistical methods for Business and Economics prentice – Hall, Inc. New Jersey.
- 7- John.O.Rawlings & others, 1998, Applied regression analysis A research Tool North Carolina State University , USA , Second Edition
- 8- N.R.DRAPER & H.Smith ,(1981) Applied Regression Analysis , New York . printed in the United States . Second Edition
- 9- Samprit Chatterjee & Alis .Hadi ,2006 regression analysis by Example , by John Wiley & Sons, Inc . Fourth Edition.
- 10- shalabh - Linear regression analysis, Indian Institute of Technology Kanpur. PDF
- 11- S.R.SEARLE,(1971) Linear models , N.Y State College of Agriculture. New York-London. printed in the United States .
- 12- T.T. Soong ,2004 , Fundamentals of probability and statistics for engineers , State University of New York at Buffalo, Copyright John Wiley & Sons Ltd, England.
- 13- National Center for Health Statistics. (2003). NCHS definitions. Cause-of-Death. Available <https://www.cdc.gov/nchs/data/hestat/leadingdeaths03/leadingdeaths03.htm>