جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة كربلاء
كلية الإدارة والاقتصاد
قسم الإحصاء

اختيار افضل إنموذج احصائي لتحليل وفيات الاطفال الخدج(دراسة تطبيقية)

رسالة مقدمة الى

مجلس كلية الإدارة والاقتصاد في جامعة كربلاء وهي جزء من

متطلبات نيل درجة ماجستير في علوم الإحصاء

من قبل
ميثم عبد الوهاب صالح

بإشراف
الأُستاذ الدكتور
عدنان كريم نجم الدين

1439هـ كربلاء 2018 م

بسم الله الرحمن الرحيم

ن ۚ وَالْقَلَمِ وَمَا يَسْطُرُونَ (1)

مَا أَنتَ بِنِعْمَةِ رَبِّكَ بِمَجْنُونٍ (2)

وَإِنَّ لَكَ لَأَجْرًا غَيْرَ مَمْنُونٍ (3)

صدق الله العلي العظيم

(سورة القلم)

الإهداء

الى

متمم النعمة ومكارم الاخلاق الى من أرسله الله رحمة للعالمين ..

رسول الله محمد (ﷺ) وأهل بيته الاطهار .

.....عزي وعنوان الصبر روح والدي طيب الله ثراه.....

.....طوق الحنان واكف الدعاء والدتي.....

.....شموع الدرب اخوتي الشهداء.....

.....سعادتي وبلسم الروح زوجتي و اولادي.....

.....كل احبتي وخاصتي ومن وقف بجانبي عونا.....

ميثم

الشكر والتقدير

الحمد لله رب العالمين والصلاة و السلام على سيدنا محمد صلى الله عليه و سلم وعلى آله الطيبين الطاهرين. وبعد :

من أسدى إليك معروفاً فحقه الشكر، ومن قدم لك خيراً فحقه الثناء، ومن واصل العطاء استحق الامتنان،أقول هذا لأن ثمة من يستحق شكراً عظيماً، وثناءً جميلاً، وتقديراً ظاهراً ..

لا يسعني بعد أن انتهيت من كتابة هذه الرسالة إلا أن أتقدم بشكري وتقديري الى الاستاذ الدكتور عدنان كريم نجم الدين لاشرافه على الرسالة و لما بذله من جهود مخلصة ومتابعة مستمرة في مدة اشرافه وإعداد الرسالة.

كذلك أوجه شكري وتقديري للأساتذة الأفاضل رئيس وأعضاء لجنة المناقشة لقبولهم مناقشة رسالتي وما أبدوه من توصيات ونصائح وملاحظات علمية من شأنها دعم هذه الرسالة وتقويمها ووضعها في مسارها الصحيح وإظهارها بالشكل المناسب.

وأتقدم بجزيل الشكر والتقدير إلى اساتذتي الاجلاء الاستاذ الدكتور عواد الخالدي والاستاذ الدكتور عبدالحسين الطائي والدكتور جاسم ناصر والدكتورة شروق السباح لاخلاصهم ووفائهم وحرصهم وجهودهم طوال مدة الدراسة .

ومن واجب الاعتراف بالجميل أن أتقدم بشكري وتقديري إلى الاستاذ الدكتور محمد حبيب الشاروط لمساعدتي عن طريق جهود حثيثة ونصائح قيمة لإتمام هذه الرسالة .

ولا يفوتني أن أتقدم بالشكر والتقدير إلى جميع زملائي في الدراسات العليا.

وختاماً أتقدم بالشكر والتقدير إلى كل من مد يد العون....ذكره قلبي ولم يذكره قلمي .

ومن الله التوفيق.....

ميثم

الم□تخلص

يعد مستوى وفيات الاطفال الخدج من المؤشرات المهمة لقياس مدى تقدم الخدمات الصحية والوعي الصحي في أي مجتمع , وقد تم اختيار هذا الموضوع نظراً لقلة الدراسات التي بحثت في هذا المجال وبخاصة في محافظة بابل وعلى مستوى العراق عموما .

اعتمد الباحث على البيانات التي تم الحصول عليها من دائرة صحة بابل والتي اشتملت على مجاميع شهرية لعدد المتوفين من الاطفال الخدج للمدة من (2014–2016) وكذلك اشتملت على بيانات لكل طفل داخل وحدة الخدج وحالة خروجه منها (حياً ام ميتاً) .

من هنا جاءت فرضية البحث لبيان مدى وجود تاثير معنوي لعدد من العوامل مثل (مدة الحمل , وزن الطفل ,عمر الطفل, عمر الام , نوع الولادة ,الخديج المنزلي) على المتغير التابع (عدد وفيات الخدج), وذلك بهدف توجيه ذوي الاختصاص والمعنيين بالصحة العامة لوضع الحلول اللازمة لمعالجة هذه الظاهرة ووضع التدابير لمواجهتها او السيطرة عليها مستقبلاً عن طريق رسم السياسات الطبيه والصحية التي تسهم في الحد من حالات الوفاة , وكذلك إدخال برامج لصحة الأم والوليد ، وضمان استمرارية الرعاية التي تشمل خدمات مرحلة ما قبل الولادة ، واعتماد طاقم مؤهل اثناء الولادة ، والمتابعة اثناء الشهر الأول من الحياة

ويهدف البحث الى بناء انموذج احصائي مناسب لطبيعة البيانات نستطيع عن طريقه معرفة مدى تاثير تلك العوامل في وفيات الاطفال الخدج حسب معنوية كل متغير, و تحليل هذه الظاهرة وبيان آثارها السلبية . لذلك بدأت بدراسة نماذج مختلفة تتلائم وطبيعة البيانات مثل انموذج الانحدار المتعدد وانموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي لدراسة البيانات الكمية وانموذج الانحدار اللوجستي وتحليل الدالة التمييزية لدراسة البيانات وصفية الاستجابة .

و تبين من خلال البحث وجود علاقة معنوية بمستويات مختلفة بين المتغيرات التفسيرية والمتغير التابع . كما تبين ملائمة النماذج الاحصائية وذلك بالاعتماد على عدد من المؤشرات والاختبارات الاحصائية .

وقد خلص البحث الى عدد من التوصيات اهمها استعمال انموذج الانحدار اللوغارتمي في حال البيانات المشابه واستعمال انموذج الانحدار اللوجستي في حال بيانات المتغير الوصفي ثنائي الاستجابة . كما نوصي بضرورة الاهتمام بمستوى العناية الطبية للاطفال الخدج والاهتمام بالعوامل المؤثرة في وفيات الاطفال ووضع الحلول المناسبة لها .

قائمة المحتويات

التسلسل	العنوان	الصفحة
	الاهداء	أ
	شكر وتقدير	ب
	مستخلص	ت
	قائمة المحتويات	ث - ج
	قائمة الاشكال	ج
	قائمة الجداول	ح
الفصل الاول/ منهجية الدراسة		**14-1**
1-1	المقدمة	3-2
2-1	مشكلة الدراسة	3
3-1	اهداف الدراسة	3
4-1	اهمية الدراسة	3
6-1	بعض الدراسات السابقة	5-3
7-1	مفهوم الخديج ووفيات الخدج	7-5
8-1	بعض الاحصائيات العالمية	9-7
الفصل الثاني/ الجانب النظري		**36-11**
1-2	تمهيد	11
2-2	مفهوم تحليل الانحدار	12-11
3-2	نماذج الانحدار الخطية	12
1-3-2	انموذج الانحدار الخطي البسيط	13

2-3-2	انموذج الانحدار الخطي المتعدد	14-13
4-2	نماذج الانحدار اللاخطية	15-14
1-4-2	انحدار الدوال متعدد الحدود	15
2-4-2	انموذج الانحدار اللوغارتمي	15
3-4-2	انموذج الانحدار الأ□ي	17-16
5-2	نماذج الانحدار المتغيرات التابعة النوعية	20-17
1-5-2	الانموذج اللوج□تي	19-17
2-5-2	تحليل الدالة التمييزية	20-19
6-2	مشاكل الانحدار	26-21
1-6-2	مشكلة التعدد الخطي	23-21
2-6-2	مشكلة الارتباط الذاتي	24-23
3-6-2	مشكلة عدم تجانس تباين حد الخطأ	26-24
7-2	طرائق تقدير معاملات النموذج	33-26
1-7-2	طريقة المربعات الصغرى الاعتيادية OLS	30-26
2-7-2	طريقة المربعات الصغرى الموزونة WLS	32-30
3-7-2	طريقة تقديرالامكان الاعظم MLE	33-32
8-2	اختيار افضل انموذج انحدار	36-33
	الفصل الثالث / الجانب التطبيقي	54-38
1-3	تمهيد	38
2-3	ماهية البيانات التي اعتمدت في الدرا□ة	39-38
3-3	تحليل البيانات التي اعتمدت في الدرا□ة	39
4-3	التطبيق العملي الاول: البيانات الكمية (التجميعية) panel data	46-39
1-4-3	تطبيق انموذج الانحدار الخطي المتعدد	41-39
2-4-3	تطبيق انموذج الانحدار اللوغارتمي	44-42
3-4-3	تطبيق انموذج الانحدار الا□ي	45-44
6-3	التطبيق العملي الثاني: البيانات الوصفية descriptive data	54-46
1-6-3	تطبيق انموذج الانحدار اللوج□تي	50-46
2-6-3	تطبيق تحليل الدالة التمييزية	54-51
	الفصل الرابع / النتائج والتوصيات	58-56
1-4	الا□ستنتاجات	57-56
2-4	التوصيات	58-57
	المصادر والمراجع	63-60
	الملاحق	77-65

قائمة الأشكال

رقم الشكل	اسم الشكل	الصفحة
1-1	عدد وفيات الخدج حسب اسباب الوفاة (2014 – 2015)	6
2-1	عدد الاطفال المتوفين حسب مدة الحمل للسنوات (2014 - 2016)	7
3-1	عدد الوفيات حسب وزن الطفل للسنوات (2014-2016)	8
4-1	عدد الوفيات حسب عمر الطفل للسنوات (2014-2016)	8
5-1	عدد الاطفال الخدج وعدد الاطفال المتوفين منهم للسنوات(2014- 2016)	9
6-1	التوزيع العالمي لوفيات الاطفال، بحسب السبب (2012)	9
1-2	مقدرات طريقة المربعات الصغرى الاعتياديةOLS	27

قائمة الجداول

رقم الجدول	اسم الجدول	الصفحة
1-3	المتغيرات والرمز المقابل لكل متغير	39
2-3	معايير اختيار الأنموذج الملائم للانحدار المتعدد بالطرق الثلاث	40
3-3	جدول المؤشرات الاحصائية للانحدار المتعدد	40
4-3	جدول التحليل التباين للانحدار المتعدد	40
5-3	جدول المعاملات والمعنوية ومعامل التضخم للانحدار المتعدد	41
6-3	معايير اختيار الانموذج الملائم للانحدار اللوغارتمي بالطرق الثلاث	42
7-3	جدول المؤشرات الاحصائية للانموذج اللوغارتمي	42
8-3	جدول تحليل التباين للانموذج اللوغارتمي	43
9-3	جدول المعاملات والمعنوية ومعامل التضخم للنموذج اللوغارتمي	43
10-3	المؤشرات الاحصائية للانموذج الأسي	44
11-3	جدول تحليل التباين	44
12-3	معلمات الانموذج الأسي	45
13-3	المتغيرات والرمز المقابل لكل متغير	46
14-3	ملخص البيانات التي دخلت في التحليل اللوجستي	46
15-3	معايير اختيار الانموذج الملائم للانحدار اللوجستي بالطرق الثلاث	47
16-3	ختبار الدلالة الاحصائية للانموذج اللوجستي	47
17-3	اختبار هوزمر- ليمشو	47

3-18	تصنيف البيانات	48
3-19	المتغيرات الداخلة في المعادلة	49
3-20	ملخص البيانات التي دخلت في التحليل التمييزي	51
3-21	بيانات المتغيرات المستقلة واحصاءات المجاميع	51
3-22	تحليل التباين والمعنوية للمتغيرات	52
3-23	الاحصاءات المتعلقة بدالة التمييز	52
3-24	مصفوفة الارتباطات	53
3-25	معاملات الدالة التمييزية المعيارية	53
3-26	النقاط الفاصلة	54
3-27	نتائج دقة التصنيف	54

الفصل الأول

منهجية الدراسة

الفصل الأول

منهجية البحث

1.1 المقدمة: Introduction :

ان تطوير وتوسيع واقع الخدمات الصحية يعد ركناً اساسياً لخطط التنمية الاقتصادية والاجتماعية والتي تهدف الى انقاذ المجتمع من الاوبئة والامراض والاهتمام بتطوير الموارد البشرية. كما انها تاتي استجابة لكثير من المشكلات الاجتماعية والبيئية التي يواجهها الفرد.
وإن المعالجة الطبية الفعالة للكثير من الامراض من شأنها ان ترفد المجتمع بعناصر منتجة تسهم في تطور الناتج القومي والحركة التنموية في البلد ، فضلا عن توفير افضل الخدمات للافراد كافة. وتحتل دراسة الوفيات لاسيما وفيات الاطفال مكانة خاصة في مجال الابحاث الصحية والسكانية اذ انها تعكس الصورة الحقيقية للمستوى الصحي العام والجانب التنموي لكونها تمثل الجانب السلبي للنمو السكاني .

ونظراً للتقدم العلمي والتكنولوجي الذي حصل في الكثير من المجالات البحثية وبالأخص في مجال الدراسات والبحوث الاحصائية حتى اصبح الإحصاء علماً له اثره الواضح في خدمة الانسان وتطويره في شتى الميادين وذلك عن طريق استعماله في الاستدلال على سائر العلوم وتطويرها، و بدأت تطبيقات الاحصاء ومجالات استعمالاته تنمو وتتقدم عملياً في كل حقول المعرفة وبشكل واضح بوصفه من العلوم التطبيقية الصرفة ذات المجال التطبيقي الواسع والتي تساعد على إتخاذ القرارات المبنية على أسس علمية صحيحة. وبات واضحاً استعماله في المؤسسات الصحية لتحليل البيانات الخاصة بالامراض والوفيات والعوامل المسببة وتصميم نماذج الوفيات.

ومن الامور البديهية ان الطريقة الاحصائية لاتكون كاملة الا اذا تحققت جميع مراحلها وخطواتها بشكل صحيح ودقيق ومن اهم هذه المراحل هي اختيار الانموذج الامثل الذي يمثل الظاهرة قيد البحث تمثيلا صحيحا ومن ثم يتم تقدير معاملات الانموذج واختبارها وتحليلها واخيراً اعتمادها لاغراض التحليل والتنبؤ . لذا فقد تم تقسيم الرسالة إلى أربعة فصول، خصص الاول منها للمقدمة وهدف البحث والاستعراض المرجعي لاهم البحوث والدراسات المشابهه وتناول هذا

الفصل كذلك شرحاً موجزاً للمصطلحات المهمة المستخدمة في البحث وتضمن الفصل في مبحثه الاخير نبذة عن وفياة الاطفال الخدج وفكرة عن تعريف مفهوم تحليل الانحدار.

بينما تضمن الفصل الثاني الجانب النظري للدراسة , اذ تم تعريف انواع نماذج الانحدار الخطية واللاخطية ونماذج المتغير التابع ثنائي الاستجابة , وبعض مشاكل الانحدار وطرائق تقدير معلمات النماذج وطرائق اختيار الانموذج الافضل لتمثيل الظاهرة محل البحث .

اما الفصل الثالث فقد شمل الجانب التطبيقي من البحث وقد تضمن تطبيق عملي على البرنامج الاحصائي spss.v25 لتحليل البيانات والمقارنة بين النماذج حسب طريقة ادخال المتغيرات التوضيحية واختيار الانموذج الافضل عن طريق معايير المقارنة بين النماذج .

واما الفصل الرابع والاخير فانه تضمن الاستنتاجات الرئيسة لهذا البحث عن طريق النتائج التي توصل اليها الجانب التطبيقي، فضلاً عن التوصيات التي افضى اليها البحث .

2.1 مشكلة البحث : The problem of the Search:

نظراً لتزايد عدد الوفيات للاطفال الخدج في السنوات (2016,2015,2014) بات من الضروري دراسة هذه الظاهرة لمعرفة اهم العوامل المؤثرة في وفاة الطفل الخديج والحد من تطورها وطرائق المعالجة كونها تعد من المؤشرات السلبية على اداء المؤسسات الصحية في البلد .

3.1 هدف البحث: The aim of the Search :

تهدف هذه الرسالة الى :

1. اختيار افضل انموذج انحدار عن طريق اجراء المقارنة بين النماذج الاحصائية المقدرة اعتماداً على المقاييس والمعايير والمؤشرات الاحصائية ذات العلاقة .
2. تحليل اهم العوامل المؤثرة في وفيات الاطفال الخدج في محافظة بابل .

4.1 أهمية البحث : research importanc :

الأطفال هم مستقبل البلدان والاساس في بناء المجتمع وهم القاعدة التي ينطلق منها لبناء مستقبله السليم، فلابد من ان يكونوا بصحة وسلامة لاداء الدور المنوط بهم في المستقبل، لذا تكمن أهمية البحث في إنه يتعرض للمشاكل الصحية التي يعاني منها الاطفال الخدج حديثي الولادة في محافظة بابل والتي تؤدي الى ارتفاع نسب الوفاة ومحاولة وضع الحلول المستقبلية التي تقلل من اثر هذه الظاهرة وامكانية اعمامها على عموم المحافظات وذلك في حال تشابه الظروف المحيطة .

5.1 الدراسات السابقة: previous studies :

اعتمدت الدراسات[20] والبحوث التي تناولت موضوع وفيات الاطفال الخدج على طريقة (Brass) لتقدير معدلات الوفيات العمرية (Age Specific Deaths Rates) للأطفال وتعتمد على البيانات المتعلقة بالنساء في سن ألإنجاب وعدد المواليد الأحياء وعدد المتوفين والباقين على قيد الحياة لسنة محددة . وقد تم اعداد دراسة عن هذا الموضوع عام (1992) من لدن الباحثين الامريكيين (Ronald Lee & Lawrence Carter) وذلك اعتماداً على اسلوب جديد في التقدير والتنبؤ بمعدلات الوفيات العمرية والذي عرف بعد ذلك باسميهما (Lee-Carter Method)،والتي تعمل على دراسة العوامل ألمؤثرة في ألوفاة وتقوم على ربط ألأنموذج ألإحصائي بطرائق ألسلاسل ألزمنية (Time Series) لاستخراج معدلات ألوفيات ألتنبؤية. وتعتمد هذه الطريقة على البيانات التاريخية المتعلقة بالاتجاهات العمرية المستخدمة في تقدير الأنموذج الاحصائي الامثل. لقد قام عدد من الدارسين والباحثين في المجالين الصحي والاحصائي بدراسات متعددة للتوصل الى اسباب الوفاة للاطفال ووضع انموذج احصائي ملائم ومنها :

1) دراسة التلباني [12] . قسم الاحصاء/ جامعة الازهر– غزة , 2004 والتي تناولت (استعمال نموذجي الانحدار اللوجستي وانحدار كوكس لدراسة العوامل المؤثرة في وفيات الاطفال الرضع في فلسطين), وتوصلت هذه الدراسة انه على الرغم من اختلاف النموذجين الا انهما توصلا الى نفس المتغيرات المؤثرة في الظاهرة نفسها , كما توصلت الدراسة الى ان اهم عامل يؤثر في وفيات الاطفال الخدج هو جنس المولود وناتج الولادة (مفرد او توأم) , وقد

اوصت بان المستوى التعليمي للام والحالة التغذوية للام والطفل اهمية كبيرة في تفسير حالة البقاء على قيد الحياة للطفل الرضيع.

2) وفي عام 2005 قامت الهاشمي[24] بـ(دراسة احصائية لوفيات الاطفال الرضع لمحافظة نينوى للمدة 1987–2004) باستعمال النماذج التنبؤية والمقارنة بين هذه النماذج لخمسة مؤشرات وهي : متوسط مربعات الخطأ (MSE), ومتوسط البواقي المطلقة (MAE), ومتوسط مطلق نسب الاخطاء (MAPE), ومتوسط البواقي (ME) و متوسط نسب البواقي المئوية (MPE), وانتقاء الأنموذج الأفضل في ضوء هذه المؤشرات التي تتباين عن بعضها البعض.

3) اما دراسة حسين [13] ,(2006) فقد تناولت (وفيات الاطفال تحت تاثير العوامل الطبيعية والبشرية في محافظة ديالى) اذ استعمل الباحث معامل الارتباط لبيرسون للفصل بين العوامل المؤثرة وغير المؤثرة واوصى في دراسته الى ضرورة الاهتمام بالجانب الصحي للام الحامل وعدم الاقتراب من المناطق التي تعرضت لضربات عسكرية ملوثة وتوجيه الامهات بعدم استعمال العقاقير دون استشارة الطبيب وعدم تعرضهن للاجهاد .

4) وفي عام 2008 قامت مديرية الرعاية الصحية الأولية[21] في وزارة الصحة السورية بالتعاون مع منظمة اليونسيف – المكتب المركزي للاحصاء باعداد (دراسة أسباب وفيات الأطفال دون الخمس سنوات) واعتمدت الدراسة على معدلات الوفيات وبينت تاثير عامل التشوهات الخلقية في وفيات الاطفال الخدج ,وقد اوصت بتطبيق برنامج الفحص الطبي ما قبل الزواج و إعطاء حامض الفوليك للنساء الراغبات في الحمل بغية الوقاية من تشوهات الأنبوب العصبي .

5) وفي العام نفسه قدمت (باناجيوتا كيتسانتاس)(Panagiota Kitsantas)[53] دراسة (الاسباب الكامنة لوفيات الاطفال الخدج Underlying and Multiple Causes of Death in Preterm Infants) وباستعمال المنحنى اللوجستي وتحليل البيانات باستعمال برنامج spss اذ أظهرت الدراسة ان قصر مدة الحمل وامراض الجهاز التنفسي للاطفال من اهم الاسباب المؤدية الى وفاة الاطفال الخدج .كما اوصت الدراسة احتمال الوفاة تزداد (لا سيما بالنسبة لأولئك الذين ولدوا بين(24–30 أسبوع) والذين لديهم مشاكل في الجهاز التنفسي .

6) اما السباح [15] ,(2009) فقد تناولت (بناء أنموذج انحدار لوجستي معدل لحياة الأطفال الخدج في محافظة كربلاء) وتبين لديها أن حياة الطفل الخديج مرتبطة بالعوامل (مدة الحمل ,وزن الطفل ,نوع المستشفى , سبب الرقود, عمر الأم) أكثر من غيرها . كما اوصت باعتماد النتائج في دراسة واقع المستشفيات ومحاولة تعزيزها بالمستلزمات الضرورية لتوفير العناية الصحية للمواطنين .

7) وفي العام نفسه قامت البير[11] بدراسة (تقييم الواقع الصحي للطفولة في العراق بالعمر (0- 4) سنة والخدمات المتاحة لهم ومقترحات تطويرها) واستعملت في الدراسة المنهاج الوصفي التحليلي , واكدت ان نسبة كبيرة من وفيات الاطفال هي بسبب الامراض التنفسية ونقص التغذية والاسهال . كما اكدت على نقص مراكز الرعاية الصحية الاولية التي تعنى بالام والطفل .

8) كما توصل الباحثان Hanady J. Mahmood ,*Saad J. Sulaiman[51] في دراستهما عام 2014 (Assessment Of Factors Causing Mortality Rate Of Neonate In Al-Batool Teaching Hospital In Mosul City) الى تقييم أسباب وفيات الأطفال حديثي الولادة خلال عام واحد 2010 وتوصلا الى 38% من الوفيات تحدث في الايام الاولى من عمر الطفل. وكانت الاكثر شيوعاً لاسباب الوفاة هي متلازمة الضائقة التنفسية بنسبة (42%) يتبعها (19%) الولادة المبكرة, والتشوهات الخلقية بنسبة (14%) وكذلك نقصان الوزن بنسبة 11% .

9) اما الدراسة التي اعدت من لدن زيد [14] ,(2014) والتي تناولت (اثر العوامل الديموغرافية والاقتصادية والاجتماعية في وفيات الاطفال الرضع في الاراضي الفلسطينية من واقع مسح الاسر الفلسطينية) فقد استعملت اسلوب التحليل الوصفي وقد استنتجت ان هناك علاقة عكسية تربط بين عمر الام وصلة القرابة بين الزوجين وبين متوسط وفيات الاطفال واضافت اثر المستوى التعليمي والرضاعة الطبيعية ووزن الطفل في وفيات الاطفال.

أما في هذه الرسالة سيتم تطبيق انموذج الانحدار المتعدد وانموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي على البيانات ذات المتغير المعتمد الكمي , و تطبيق انموذج الانحدار اللوجستي والدالة التمييزية على البيانات ذات المتغير المعتمد الوصفي .

6.1 مفهوم الخديج ووفيات الخدج[27]: Preterm Concept the Death of Preterm

تعد ظاهرة وفيات الاطفال خسارة في وقت وجهود وثروة البلدان فضلاً عن آثارها النفسية والاقتصادية في المجتمع وان دراسة هذه الظاهرة والوقوف على تحليل اسبابها المختلفة ومحاولة وضع المعالجات تعطي رافداً للتقليل من نسبها والحفاظ على الثروة البشرية اضافة الى عدم هدر الجهد الطبي والخدمات الصحية والاموال . ولمعرفة ماهية هذه المشكلة لابد من معرفة اهمية تقديرات وفيات الاطفال الخدج والتعريف بمعنى حديث الولادة ومعنى الخديج وفهم أسباب هذه الظاهرة .

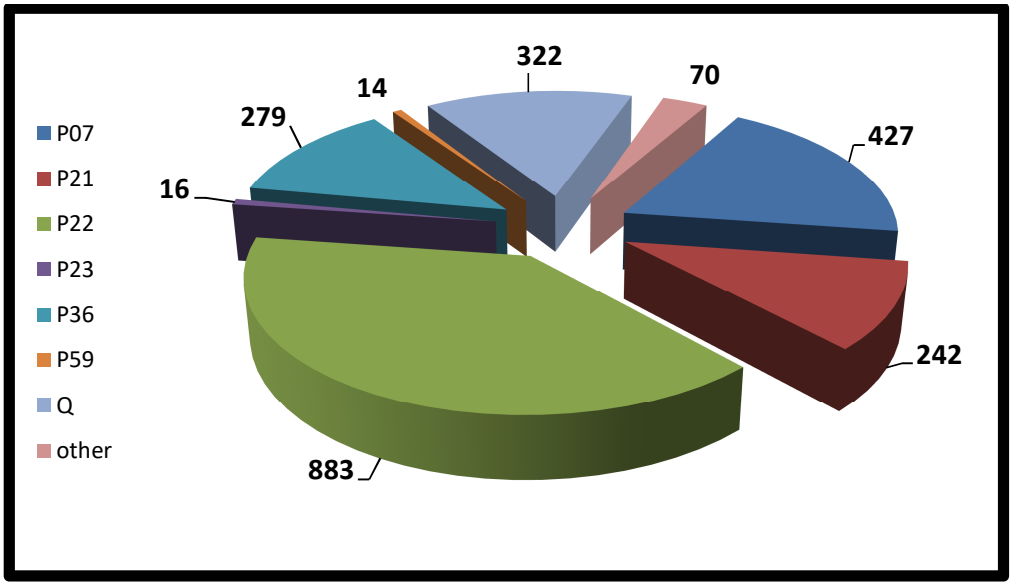
تستمد تقديرات وفيات الأطفال حديثي الولادة او الخدج اهميتها من أمرين هما رصد وتقييم البرامج الصحية الراهنة وصياغة السياسات وتوزيع الاولويات للبرامج المستقبلية . كما ينظر اليها كمؤشرات أساسية للوضع الاجتماعي والاقتصادي ونوعية الحياة ومستوى المعيشة العام في المجتمع .

وحديث الولادة (new born) هو الطفل الذي يصل عمره لأقل من شهر واحد بعد الولادة. **أما الخديج (Preterm)** هومَن وُلد مبكراً أي لمدة حمل أقل من 37 أسبوعا.

وفقاً لدراسة أجرتها منظمة الأمم المتحدة للطفولة (يونيسيف) وُجد أن معظم أسباب وفيات الأطفال حديثي الولادة يعود إلى تشوهات خلقية بنسبة 2.27 % و الولادات المتعددة بنسبة 26 % و عدم النضوج بنسبة 7.21 %، ومرض الأم بنسبة 7.6 % وأوضاع معينة للوليد 4.6 % والاختناق بنسبة 9.4 %. وأوضحت الدراسة أن أكثر التشوهات الخلقية شيوعًا، كانت للأمراض القلبية الخلقية والتشوهات الخلقية المتعددة، التي شكلت نسبتها 8.25% و 1.19 % على الترتيب .

وتشير البيانات التي تم الحصول عليها من دائرة صحة بابل ان عدد المتوفين للمدة (2014-2016) في المحافظة بلغ (1993) وكانت بنسبة 39.19% من اسباب الوفاة تعود لعسر التنفس (P22) , ثم بنسبة 18.95% لاسباب نقص الوزن ونقص مدة الحمل (P07), ومن ثم التشوهات الخلقية (Q00-Q99) بنسبة 14.29%, وتسمم الدم (P36) بنسبة 12.38% , وبنسبة 10.74% من اسباب الوفاة هي للاختناق الولادي (P21), وكان اليرقان

الولادي(P59) وذات الرئة الولادي (P23) واسباب اخرى بنسبة 4.45% .والشكل (1-1) يوضح ذلك .

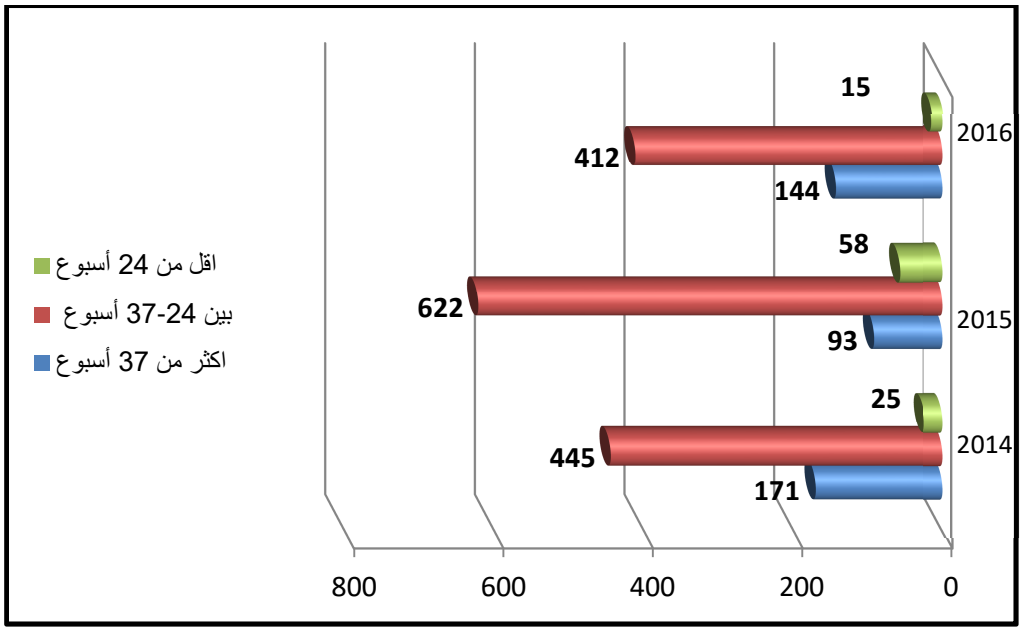


شكل (1-1)

عدد وفيات الخدج حسب اسباب الوفاة في محافظة بابل للسنوات (2014 – 2016)

والولادة المبكرة (Premature birth) هي اكبر سبب لوفيات المواليد، وهي ثاني أكبر سبب لوفاة الأطفال دون سن الخامسة ، اذ أكثر من مليون طفل يموت سنوياً بسبب مضاعفات الولادة المبكرة، ومعظمها في العالم النامي.

وعلى الصعيد العالمي فان أكثر من 1 من كل 10 حالات الحمل تنتهي بالولادة المبكرة. وان الأطفال المولودين مبكرا في أقل من 23 أسبوعا تكون فرصة بقائهم على قيد الحياة قريبة من الصفر، وفي 23 أسبوعا 15%, وفي 24 أسبوعا 55% وترتفع هذه النسبة الى 80% للاطفال المولودين بعمر 25 أسبوعا . والشكل (1-2) الآتي يوضح عدد الوفيات للاطفال الخدج في محافظة بابل حسب مدة الحمل للسنوات (2014-2016)



شكل (1-2)

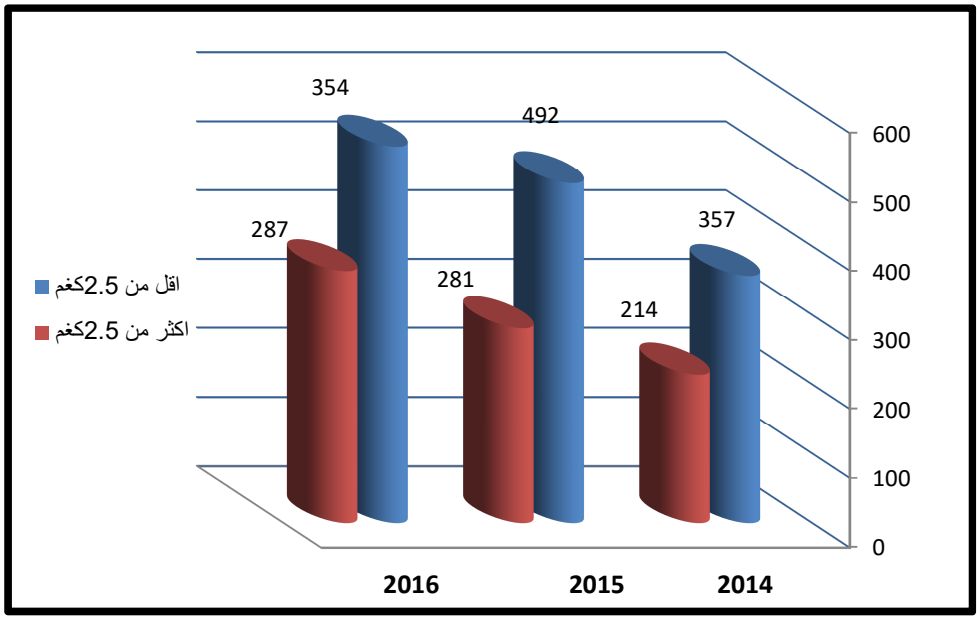
عدد الاطفال المتوفين حسب مدة الحمل للسنوات (2014- 2016)

نلاحظ ان عدد الوفيات ترتفع لمدة الحمل بين(24-37) اسبوعاً اذ تبلغ 73.37% بينما تبلغ نسبة الوفيات لمدة الحمل اكثر من 37 اسبوعاً 22.28% والنسبة الاقل لمدة الحمل الاقل من 24 اسبوعاً اذ تبلغ 4.35% لندرة ولادتهم احياء.

7.1 بعض الاحصاءات العالمية [8] :

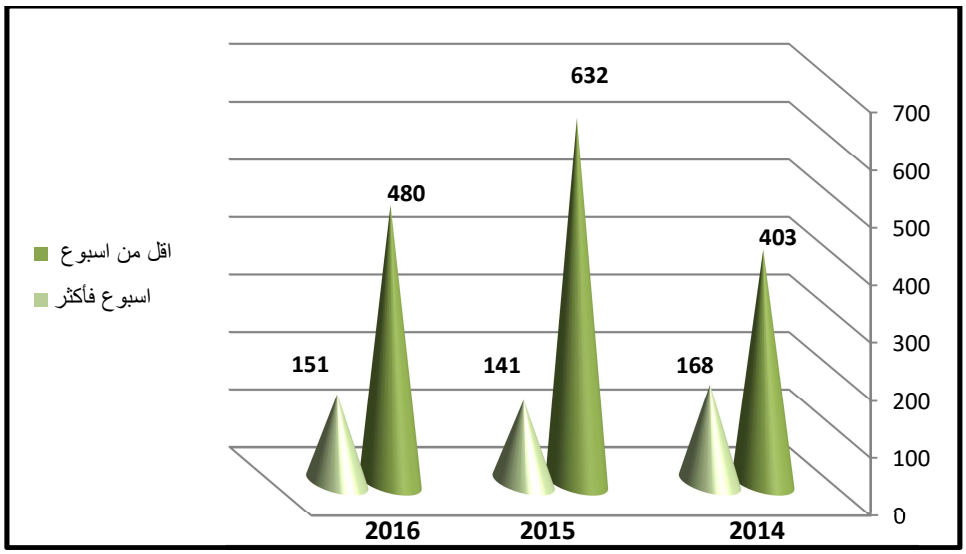
ويتضح من التحليل الذي اعد في منتصف المدة بين عامي 2000 و 2015 وجود بوادر مشجعة للتقدم في مجال الصحة ؛ كما يشير إلى عدم تحقق أي تقدم في مجال صحة الأمهات والاطفال, ويقدر عدد الأطفال الذين يعانون من نقص الوزن بنحو 112 مليون طفل، ولا يزال نقص التغذية يمثل سبباً رئيساً لما يزيد عن ثلث وفيات الأطفال.وفي عام 2007 قدر عدد وفيات الأطفال بنحو 9 ملايين وفاة، ومن البديهي ان معدلات وفيات الأطفال يعتمد اعتماداً كبيرا على معالجة مشكلة وفيات المواليد . وتشير التقديرات أن 37 % من وفيات الأطفال دون الخامسة على صعيد العالم تحدث في الشهر الأول من عمرهم، ومعظمها في الأسبوع الأول منه .

وما يخص موضوع البحث نلاحظ زيادة عدد وفيات الاطفال الذين تقل اوزانهم عن 2.5 كغم وكذلك ارتفاع الوفياة للاطفال الذين تقل اعمارهم عن اسبوع واحد . لاحظ الشكل (1-3) والشكل (1-4)



شكل (1-3)

عدد الوفيات حسب وزن الطفل للسنوات (2014 ـ 2016)

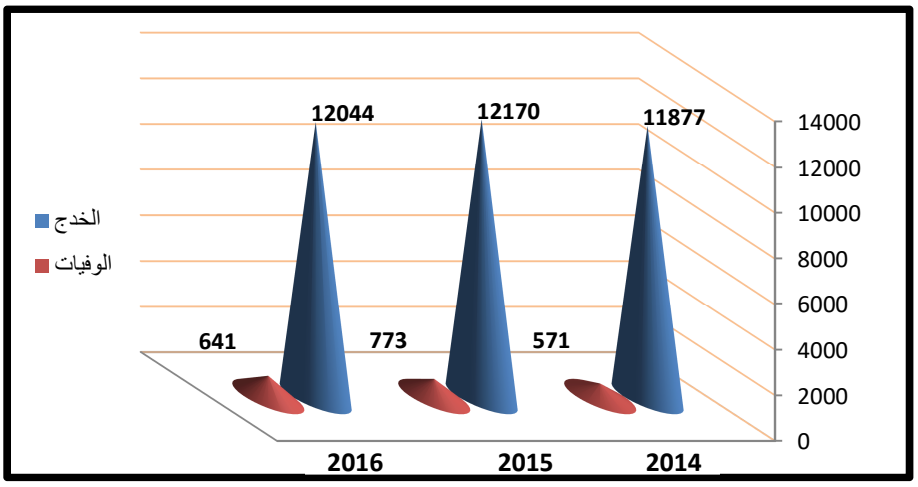


شكل (1-4)

عدد الوفيات حسب عمر الطفل للسنوات (2014 ـ 2016)

نود الاشارة الى وفيات الاطفال دون الخامسة والذي من ضمنهم الاطفال الخدج وعلى صعيد الدول العربية فان معدل وفيات الاطفال دون الخامسة لكل (1000) مولود حي هي 8 في الامارات و 10 في قطر و 10 في البحرين و 11 في الكويت و12 في عمان و17 في سوريا و 18 في ليبيا و 20 في الاردن و 21 في تونس و 25 في السعودية و 29 في لبنان و 34 في المغرب و 36 في مصر و45 في العراق و 73 في اليمن .

اما تقرير منظمة اطفال الحروب [28] (وور تشايلد) الفصلي لشهر آيار 2013 عن العراق فيشير ان معدل وفيات الاطفال الرضع دون السنة الواحدة من العمر يصل الى 100 طفل يومياً استنادا الى احصاءات مؤسسة (ايرين الخيرية) التابعة للامم المتحدة في تموز 2011 . والشكل (1-5) يبين عدد الاطفال الداخلين الى وحدات الخدج في مستشفيات محافظة بابل وعدد المتوفين منهم .

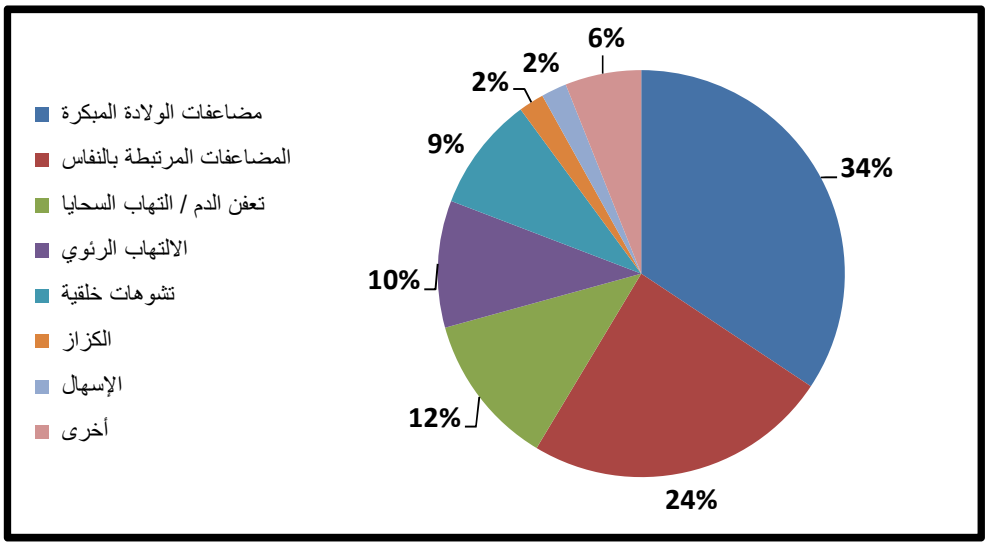


شكل (1-5)

عدد الاطفال الخدج وعدد الاطفال المتوفين منهم للسنوات (2014 - 2016)

نلاحظ ارتفاع عدد وفيات الاطفال خلال السنوات اذ بلغ عام 2014 عدد الوفيات 571 وفي عام 2015 بلغ 773 وفي عام 2016 بلغ 641 .

كما نشرت منظمة انقاذ الطفولة [48] عام 2012 التي تعمل باكثر من 120 دولة تقريراً اوضحت فيه نسب وفيات الاطفال حسب اسبابها والمبينة في الشكل الآتي :



شكل (1-6)

التوزيع العالمي لوفيات الاطفال، بحسب السبب (2012)

الفصل الثاني

انموذج الانحدار الخطي واللاخطي

الفصل الثاني
الاطار النظري لنماذج الانحدار الخطية و اللاخطية
Theoretical Frame for linear & Nonlinear Regression Models

1.2 تمهيد :

يمكن تقسيم نماذج الانحدار حسب شكل العلاقة بين المتغيرات الى النماذج الخطية والنماذج اللاخطية , كما يمكن تقسيم النماذج حسب عدد المتغيرات التفسيرية الى انموذج بسيط وانموذج متعدد . وهناك تصنيف آخر حسب مستوى القياس للمتغيرات ويعتمد هذا التصنيف على نوع بيانات المتغير التابع فان كانت تلك البيانات كمية فان النماذج تصنف الى نماذج الانحدار الشائعة , وان كانت وصفية تصنف الى نماذج الانحدار متعددة الاستجابة , وسنتناول في هذا الفصل دراسة أهم هذه النماذج وعرض بعض المشاكل التي تواجه فروض الانحدار وطرائق التقدير الشائعة لتحديد قيم معاملات الانحدار المجهولة كما سيتم التطرق الى اساليب اختيار افضل انموذج انحدار اعتماداً على المعايير والاختبارات الاحصائية ذات العلاقة .

2.2 مفهوم تحليل الانحدار : Regression Analysis :

لغرض اعداد البحوث العلمية[36] لا بد من الأخذ بنظر الاعتبار أن البحث عن الحقيقة ليس فيه أي مجال للتحيز أو عدم الدقة في التحليلات الإحصائية أو تفسيرها . ويعد تحليل البيانات وتفسير النتائج والتنبؤ أهم مرحلة في المشروع البحثي لأنها تعطي إجابات عن الأسئلة المطروحة .كما يتطلب ان يكون تحليل البيانات بطريقة متأنية ومدروسة ومن ثم تفسيرها أضف الى ذلك ضرورة تحليل البيانات بشكل صحيح وبصورة مفصلة بحيث يمكن للآخرين الوصول الى الاستنتاجات العلمية نفسها .

تم تطوير تحليل الانحدار[38] لاول مرة من لدن (Sir Francis Galton) في الجزء الاخير من القرن التاسع عشر وقد درس Galton العلاقة بين أطوال الآباء والأمهات وأطوال الأبناء , وأشار الى ان أطوال الأبناء من كل من الآباء ذات صفة طوال القامة والآباء ذات صفة قصار القامة يميلون الى "العودة او التراجع" الى متوسط الطول, وأُعد هذا الاتجاه هو

إنحدار نحو الوسط , وقد وضع Galton وصفاً رياضياً لميل الانحدار كان أساساً لتطوير نماذج الانحدار واستمر استخدامها لوصف العلاقات الاحصائية بين p من المتغيرات.

ويعرف الانحدار على انه علاقة دالية تربط بين متغيرين أو اكثر بواسطة صيغة رياضية فاذا كان X يدل على المتغير التفسيري وY يدل على المتغير التابع فالعلاقة الدالية بينهما يعبر عنها كما يأتي :

$$Y= f (X)$$

وقد اصبح تحليل الانحدار[43] واحداً من الادوات الاحصائية الاكثر استعمالاً لتحليل البيانات متعددة العوامل فضلاً عن انه يوفر مفهوما وبطرائق بسيطة لطبيعة العلاقات الدالية بين المتغيرات .

ويعرف تحليل الانحدار على انه مجموعة من التقنيات التحليلية التي تدرس العلاقات المتبادلة بين مجموعة معينة من المتغيرات .

كما يفسر الانحدار على انه دراسة التبعية (dependence study)[44] كونه يجيب عن الاسئلة الخاصة باعتماد متغير الأستجابة على واحد او أكثر من المتغيرات التفسيرية بما في ذلك التنبؤ بالقيم المستقبلية لمتغير الاستجابة وتقدير تأثير القيم المفسرة في قيم الاستجابة .

في حين ذهب (الراوي ,خاشع محمود) [4] لتفسير الانحدار على انه وسيلة إحصائية يستعمل لتحليل العلاقة بين متغير مستقل واحد أو أكثر ومتغير تابع .

أما (راو, بوتلري) [3] فقد عرف الانحدار بانه العلاقة السببية بين متغير معتمد (دالياً) ومتغير واحد او اكثر من المتغيرات المفسرة.

وبناءً على ماتقدم ذكره من تعاريف متعددة لمفهوم الانحدار يمكن أن نضع تعريفاً مبسطاً للانحدار الخطي البسيط كونه وصف لعلاقة بين متغيرين أحدهما تابع والآخر مستقل.

ويكتسب الانحدار[2] أهمية بالغة في تحليل العلاقات الدالية كونه يقدم أربعة وظائف رئيسة هي:

1) **الوصف (description):** إذ يمكن وصف وتلخيص البيانات بمعادلة واحدة او اكثر لوصف طبيعة العلاقة بين تلك المتغيرات .
2) **السيطرة (control):** السيطرة على نتائج المتغير التابع عند تغير قيم المتغيرات المفسرة.

3) **التنبؤ (prediction):** تقدير قيم متغير الاستجابة والتنبؤ بالمقادير التي يمكن ان يحصل عليها في للمستقبل.

4) **تقدير المعلمات (Estimation of parameters):** للاستدلال على أهمية وقوة العلاقة بين المتغيرات وإتجاه العلاقة .

3.2 نماذج الانحدار الخطية [1] : Linear Regression Models :

يختص تحليل الانحدار Regression Analysis بوصف ودراسة العلاقة بين المتغيرات على شكل انموذج يفسر هذه العلاقة . فاذا كانت العلاقة بين متغيرين وهما المتغير التابع (Y) و المتغير التفسيري (X) فيسمى بانموذج الانحدار الخطي البسيط , أو تكون العلاقة بين عدة متغيرات مستقلة مع المتغير التابع ويسمى انموذج الانحدار الخطي المتعدد . والانموذج الخطي يشير الى الكيفية التي يكون فيها كل حد من حدود المعادلة وليس شكل دالة الانحدار :

$$g(x_1, x_2, \dots, x_p) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + e_i$$

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

$$\frac{\partial g(\beta_0,\dots,\beta_p)}{\partial B_i} = 0 \qquad i=0, 1,2,\dots,p$$

تكون خالية من المعالم $\beta_0, \beta_1, \dots \beta_p$

1.3.2 انموذج الانحدار الخطي البسيط[41] :

Simple Linear Regression Model :

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

$$Y_i = \beta_0 + \beta_1 X_i + e_i \qquad i = 1,2,\dots,n \qquad \dots\dots\dots\dots (1\text{-}2)$$

اذ ان :

Y_i : المتغير المعتمد (متغير الاستجابة) dependent variable

Xi : المتغير المستقل (التفسيري) independent variable

B_0 : تمثل قيمة الحد الثابت intercept ..

B_1 : تمثل ميل خط الانحدار Slope

ei : عنصر الخطأ .

2.3.2 أنموذج الانحدار الخطي المتعدد [37]:

The multiple regression model :

ويفترض هذا الانموذج ان المتغير التابع Y هو دالة خطية لمتغيرين تفسيريين او لسلسلة من المتغيرات التفسيرية (x_1, x_2 , ….., x_p) فضلا عن عنصر الخطأ وتكون صيغته كالآتي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots .. + \beta_p X_{ip} + e_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2\text{-}2)$$

i = 1,2,…..,n , j = 1,2,….,p

n: يشير الى حجم العينة

p : يشير الى عدد المتغيرات التفسيرية المأخوذة .

اذ ان :

Y_i: المتغير المعتمد dependent variable وهو متغير عشوائي .

X_i: المتغيرات المفسرة independent Explanatory Variables وتكون قيم ثابتة fixed Values وغير عشوائية .

β_0 : الحد الثابت (الاعتراضي) intercept .

$\beta_1, \beta_2, \ldots .. \beta_p$ معلمات الانموذج وتمثل المشتقات الجزئية لمتغير الاستجابة Y_i بالنسبة الى X_i .

فرضيات الانموذج :

1) Y_i دالة خطية في X_i .
2) عدم وجود علاقة خطية تامة أو شبه تامة بين إثنين أو اكثر من المتغيرات التفسيرية.
3) عنصر الخطأ يتوزع طبيعياً بمتوسط (0) وتباين σ^2 لجميع المشاهدات. $e_i \sim N(0, \sigma^2)$
4) عدم وجود إرتباط بين قيم عنصر الخطأ للمشاهدات المختلفة .

وإذ أن لدينا n من المشاهدات يكون لدينا n من المعادلات التي يمكن التعبير عنها بواسطة المصفوفات :

$$\underline{Y} = \underline{\beta}X + \underline{e} \quad \text{..... (3-2)}$$
$$Y \sim N(\underline{\beta}X\ , \sigma^2 I)$$

اذ ان :

$\underline{Y}$: متجه من الدرجة $n \times 1$; X مصفوفة $[n \times (P+1)]$

$\underline{\beta}$: متجه $(p+1) \times 1$; e : متجه $n \times 1$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad ; X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} ; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} ; ei = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

ويحتوي كل عمود من أعمدة المصفوفة X على قيم لمتغير مستقل واحد ، B_0 لديه مضاعف ثابت 1 لجميع المشاهدات , ومن ثم فان متجه العمود 1 هو العمود الاول للمصفوفة X والصفوف هي المعاملات المقابلة للمعلمات (β) والتي تعطي التوقع $E(Y_i)$ ، وان ضرب الصف الاول من المصفوفة X بالعمود β وإضافة العنصر الاول من e يعطينا انموذج المشاهدة الاولى:

$$Yi = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots\cdots + \beta_p X_{ip} + e_i \quad \text{.....(4-2)}$$

كما ان عناصر المتجه $\underline{Y}$ ومتجه الاخطاء $\underline{e_i}$ هي متغيرات عشوائية .

والمتجه $\underline{\beta}$ هو متجه ثوابت غير معروفة يراد تقديرها اعتماداً على البيانات المتوفرة، وان كل عنصر في عمود β_j هو معامل الانحدار الجزئي الذي يعكس التغيير في المتغير التابع لكل وحدة تغيير في jth المتغير التفسيري .

4.2 نماذج الانحدار اللاخطية : Nonlinear Regression Models [43]:

تعد النماذج الخطية اكثر واقعية وسهولة في تقديرها . ولكن هناك الكثير من الظواهر تسلك سلوكا لاخطيا أي ان النماذج تكون غير خطية في المعلمات , لذا نقوم بتحويلها الى نماذج خطية وهنا يمكن الاشارة الى نوعين من النماذج اللاخطية (خطية بحد ذاته ولا خطية بحد ذاته) (intrinsically linear and intrinsically nonlinear models) الأولى يمكن

تحويلها الى نماذج خطية بواسطة اجراء عمليات التحويل والأخرى لا يمكن تحويلها الى نماذج خطية .

ونقصد بالانموذج اللاخطي هو ان المشتقه فيه تعتمد على معلمة من المعالم أو اكثر ويمكن التعبير عنها بالشكل :

$$Y = f(X;\alpha,\beta,\gamma) + e \quad \text{.....(5-2)}$$

فمثلاً الانموذج الآتي :

$$y = f(x,\theta) + e = \theta_1 e^{\theta_2 x} + e$$

فان مشتقات دالة التوقع للمعلمة θ_1 و θ_2 :

$$\frac{\partial f(x,\theta)}{\theta_1} = e^{\theta_2 x} \quad ; \quad \frac{\partial f(x,\theta)}{\theta_2} = \theta_1 x\, e^{\theta_2 x}$$

وبما ان المشتقات هي دالة للمعلمات المجهولة فان الانموذج غير خطي.

ومثال على النماذج التي لا يمكن تحويلها الى نماذج خطية $Y = \alpha + \beta\gamma^X$ المنحنى الأسي المعدل أو الانموذج

$$Y = \alpha_1 e^{\theta_1 X} + \alpha_2 e^{\theta_2 X}$$

و تقدر معالم النماذج غير القابلة للتحويل باستعمال سلسلة تايلر Taylor Series وعن طريق اخذ المشتقات الجزئية .وسنهتم في هذا البحث بالنماذج القابلة للتحويل الى نماذج خطية و تقدير معالمها بالطرائق المعروفة .

1.4.2 انحدار الدوال متعددة الحدود : [37],[56] 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 :

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_k x_i^k + e_i \quad \text{..... (6-2)}$$

i=1,2,…,n

وللسهولة نعتمد التحويل الاتي : $Z_{mi} = X_i^m$, i= 1,2,…,k

$$yi = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \cdots + \beta_K Z_{Ki} + ei \quad \text{..... (7-2)}$$

ويمكن تحديد رتبة الانموذج (rank of the model) على اساس أعلى أس للمتغير التفسيري 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 أو يكون الانموذج من الدرجة الثالثة (Cubic form) وهكذا .والغرض من زيادة القوى للمتغير التفسيري هو للحصول على انموذج ملائم يمثل العلاقة الانحدارية بين X و Y .

اما بشكل المصفوفات فيكون على النحو الآتي :

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix} ; X=\begin{bmatrix}1 & x_1 & x_1^2 & \cdots & x_1^k\\1 & x_2 & x_2^2 & \cdots & x_2^k\\\vdots & \vdots & \vdots & \ddots & \vdots\\1 & x_n & x_n^2 & \cdots & x_n^k\end{bmatrix} ; \beta=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_p\end{bmatrix} ; ei=\begin{bmatrix}e_1\\e_2\\\vdots\\e_n\end{bmatrix}$$

2.4.2 انموذج الانحدار اللوغارتمي :[39],[32] log Regression model :

هنا يجب أن تكون قيم المشاهدات الاصلية موجبة ، لان اللوغاريتم للقيم السالبة غير ممكن . كما يفترض ان حد الخطأ يتبع توزيعاً لوغاريتمياً طبيعياً . والصيغة العامة للانموذج تأخذ الشكل الآتي:

$$Y_i=\beta_0 X_{1i}^{\beta 1}X_{2i}^{\beta 2}\ldots\ldots X_{ki}^{\beta k}e_i \qquad i=1,2,\ldots,n \qquad \ldots\ldots\ldots (8\text{-}2)$$

اذ ان :$(\beta_0,\beta_1,\ldots\ldots\beta_K)$ تمثل معالم الانموذج , وان $(X_1,X_2,\ldots,X_K)$ هي المتغيرات التفسيرية . ولتحويله الى الانموذج الخطي نأخذ اللوغارتم للطرفين فيكون:

$$lnY=ln\beta_0+\beta_1 lnX_{1i}+\beta_2 lnX_{2i}+\ldots\ldots\ldots+\beta_K lnX_{ki}+lne_i \quad \ldots\ldots (9\text{-}2)$$

e : هو الخطأ العشوائي المضاعف ويتوزع توزيعاً لوغارتمياً طبيعياً .

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

وبافتراض ان :

$$lnY=y\,,ln\beta_0=\beta\,,lnX_{1i}=Z_{1i}\,,lnX_{2i}=Z_{2i}\ldots\ldots,lnX_{ki}=Z_{ki}\,,lnei=v$$

فيكون الانموذج الخطي كما ياتي :

$$y=\beta+\beta_1 Z_{1i}+\beta_2 Z_{2i}+\ldots\ldots\ldots+\beta_K Z_{ki}+v \qquad \ldots\ldots\ldots (10\text{-}2)$$

3.4.2 انموذج الانحدار الأُسي: [55],[57],[33]

Exponential Regression model :

هو احد النماذج اللاخطية ويختلف شكل الانموذج حسب عدد معالمه فيكون بمعلمة واحدة أو معلمتين أو اكثر وكذلك يختلف من حيث عنصر الخطأ (مضاف ومضروب) ويستعمل في كثير من الأحيان عندما تتعلق الدراسة بـ(الاستجابة) مقابل الزمن .

1- الانموذج الأسي بمعلمة واحدة وياخذ الصيغة الاتية [25] :

$$Y_i = X_i^{\beta} + e_i \quad \ldots\ldots\ldots (11\text{-}2)$$

اذ أن الخطأ العشوائي ei يتوزع توزيعاً طبيعياً بوسط صفر وتباين معلوم σ^2 لكل i=1,2,…,n وان β تمثل معلمة الانموذج المراد تقديرها .

2- الانموذج الأسي بمعلمتين وياخذ الصيغة الأتية :

$$Yi = \beta_0 e^{\beta 1 Xi} + ei \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12\text{-}2)$$

ei: يمثل الخطأ العشوائي يتوزع طبيعياً بوسط صفر وتباين معلوم σ^2.

$ei \sim N(0, \sigma^2)$

والشكل الاخر للانموذج الأسي بمعلمتين هو معادلة النمو الاسي Exponential growth equation وتاخذ الشكل الاتي :

$$Y = \beta_0 e^{\beta 1 X} e \quad \ldots\ldots\ldots (13\text{-}2)$$

e: هو عنصر الخطأ بتوقع=1 .

β_0: هو مقدار البدء بمقدار التوقع عندما X=0 .

نلاحظ في هذه الصيغة اننا نضرب بعنصر الخطأ (e) بدلاً من اضافته كما في النموذج الخطي لذا فانه لا يمكن ان يكون سالباً او صفراً حتى نتمكن من أخذ اللوغاريتم عند التحويل الخطي . كما ان قيمة (Y) لا يمكن ان تكون قيمة سالبة او صفر. في حين X غير محدده بهذا الشرط . لذلك هناك نوعان أساسيان من افتراضات الخطأ مضروب Multiplicative model, ومضافة Additive model. وفي النموذج المضروب، يفترض أن يكون توزيع الخطأ لوغاريتمياً .في حين يكون توزيع الخطأ المضاف توزيعاً طبيعياً

3- الانموذج الأسي في حال اكثر من معلمتين وياخذ عدة صيغ منها :

A - $Yi = EXP(\beta_0 . \beta_1 X_1 . \beta_2 X_2 . \cdots\cdots . \beta_P X_P) + e_i$

B - $Yi = EXP(\beta_0 . \beta_1 X_1 . \beta_2 X_2 \cdots\cdots \beta_P X_P) . e_i$

C - $Yi = \beta_0 + \beta_1 EXP(\beta_2 X_1 + \beta_3 X_2 + \cdots\cdots + \beta_{P+1} X_P) + e_i$

وسوف نستعمل الصيغة الاتية في موضوع الدراسة :

D - $Yi = \beta_0 EXP(\beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_P X_P) + e_i \quad \ldots\ldots\ldots (14\text{-}2)$

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

Y_i: المتغير التابع .

β_i: التغير النسبي في Y المقترن بتغير وحدة واحدة من X .

e_i: هو عنصر الخطأ بتوقع=0

X_i: هو المتغير التفسيري (غالباً ما يكون مقياساً للوقت واحياناً الجرعة) .

β_0: هو مقدار البدء بمقدار التوقع عندما Xi=0 .

ولتحويل النموذج الأسي الى النموذج الخطي نأخذ اللوغارتم لكلا الطرفين

$$\ln(Y) = \ln(\beta_0) + (\beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_P X_i) + e_i$$

واذا اعتبرنا ان : $A = \ln(\beta_0)$; $y^* = lnY$ فيكون الانموذج الأسي بالصيغة الخطية على الشكل الاتي :

$$y^* = A + (\beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_P X_i) + e_i \quad \ldots\ldots\ldots (15\text{-}2)$$

5.2 نماذج الانحدار المتغيرات التابعة النوعية :

Regression models of descriptive dependent variable

يستعمل تحليل الانحدار لدراسة تاثير متغير او مجموعة من المتغيرات التفسيرية في متغير تابع غالبا ما يكون متغيراً كمياً عن طريق بناء معادلة للتقدير او التنبؤ او التفسير . ولكن هناك بعض الظواهر تكون المتغيرات التابعة نوعية وليست كمية والتي لا يتحقق معها فرضيات الانحدار المعروفة ومن ثم فان تقديرها بطريقة المربعات الصغرى لا يمتاز بالكفاءة أي ان القيمة المتوقعة (الاحتمالية) للمتغير التابع قد لا تكون بين الصفر والواحد .لذلك تم اللجوء الى نماذج الاحتمال الخطي في الانحدار The Linear Probability of model وهو الذي يتخذ شكل المعادلة الاتية :

$$P = E(Y = 1/X)$$
$$= a + \beta X$$

وهو نموذج يحاول بيان أثر المتغير الكمي التفسيري في المتغير التابع النوعي . والمتغير التابع هنا هو متغير احتمالي يتضمن قيمتي الصفر والواحد .ومن اجل الحصول على نتائج تقدير منطقية للدوال ذات المتغيرات التابعة النوعية لابد من اعتماد النماذج المناسبة ومنها انموذج الانحدار اللوجستي ثنائي الاستجابة وانموذج الدالة التمييزية .

1.5.2الانموذج اللوجستي[34] : Logistic Model:

الانحدار الخطي غالبا ما يعمل بشكل جيد عندما يكون متغير الاستجابة كمياً , أما الحالة التي يأخذ متغير الاستجابة على احد اثنين من القيم المحتملة (1,0) فهي غالبا ناجمة عن استجابة نوعية . يستعمل تحليل الإنحدار اللوجستي في الدراسات الوبائية والطبية والذي عن طريقه يتم تحديد المتغيرات التفسيرية الكمية والنوعية التي تؤثر في إحتمال حدوث المتغير الناتج . ولا يشترط في هذا الانموذج أن تكون المتغيرات التفسيرية الداخلة في الانموذج تتبع التوزيع الطبيعي ولا أن تكون العلاقة خطية بينها وبين المتغير التابع .

يستند الانموذج الى دالة الاحتمال اللوجستي ويتم تحديده بصورة عامة :

$$E(Y) = \frac{\exp(\beta_0+\beta_1 X)}{1+\exp(\beta_0+\beta_1 X)} \quad \text{..... (16-2)}$$

ويمكن ان تؤول :

$$E(Y) = \frac{1}{1+\exp[-(\beta_0+\beta_1 X)]} \quad \text{..... (17-2)}$$

Y_i: وهو متغير عشوائي ثنائي الاستجابة binary response يأخذ القيم اما(0) او (1) اذ يكون توزيعه الاحتمالي كالاتي :

$$Yi \sim \beta(1, P)$$

إذ أن :

β_0: الحد الثابت .

β_1 : الميل (slop) وهو تعبير عن احتمال لوغارتمي ناتج عن زيادة وحدة واحدة في X .

هذا يعني ان نسبة الرجحان e^{β_1} تتغير عند زيادة X بمقدار وحدة واحدة .

X : المتغير المفسر .

e : هو اساس اللوغاريتم الطبيعي ≈ 2.718

ويمكن تحويل الانموذج اللوجستي الى دالة خطية :

$$Z_i = lin(\frac{pi}{1-pi})$$

$$Z_i = \beta_0 + \beta x \ \ or \ Z_i = \beta_0 + \sum \beta_i X_i \quad \text{.. ..... (18-2)}$$

وفي حالة وجود اكثر من متغير تفسيري يكون:

$$lin\left(\frac{p_i}{1-pi}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} \quad \text{..... (19-2)}$$

i=1,2,…,n ; j=1,2,…,k

نلاحظ ان اللوغارتم الطبيعي لنسبة الرجحان هو الدالة الخطية للمتغير التفسيري.

ويمكن التعبير عن دالة التوزيع الاحتمالي بالشكل الآتي[29] :

$$fi(Y_i) = {p_i}^{Y_i}(1 - p_i)^{1-Y_i} \quad ;Y_i = 0.1 \quad , i=1,\ldots,n$$

وبما ان المشاهدات Yi مستقلة عن بعضها فان دالة الاحتمال المشترك هي :

$$g(Y1, \ldots., Yn) = \prod_{i=1}^{n} fi(Yi) \Rightarrow = \prod_{i=1}^{n} {p_i}^{Yi}(1 - p_i)^{1-Yi}$$

عادة في انموذج الانحدار اللوجستي يمتد لاكثر من متغير تفسيري للحصول على وصف كافٍ.

ببساطة نستبدل $(\beta_0 + \beta_1 X)$ بـ $(\beta_0+\beta_1 X_1+\ldots..+\beta_{P-1} X_{P-1})$ ولتبسيط الصيغة نستبدلها بالمصفوفة والمتجهات الاتية :

$$\beta_{p\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} ; X_{p\times 1} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_{p-1} \end{bmatrix} ; X_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{i.p-1} \end{bmatrix} ;$$

$$X'\beta = \beta_0 + \beta_1 X_1 + \cdots\cdots + \beta_{p-1} X_{p-1}$$

$$Xi'\beta = \beta_0 + \beta_1 X_{i1} + \cdots\cdots + \beta_{p-1} X_{i.p-1}$$

$$E(Y_i) = \frac{\exp(X'\beta)}{1+\exp(Xi'\beta)} \qquad \ldots\ldots\ldots\ldots (20\text{-}2)$$

$$\ln(Y_i) = \ln \prod_{i=1}^{n} {p_i}^{Yi}(1 - p_i)^{1-Yi}$$

$$= \sum_{i=1}^{n} [Y_i \ln pi + (1 - Y_i) \ln(1 - p_i)]$$

$$= \sum_{i=1}^{n} \left[Y_i \ln\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} \ln(1 - p_i)\right]$$

اذ ان $E(Y_i) = p_i$ للمتغير الثنائي

$$1 - p_i = [\ 1+ \exp(\beta_0+\beta_1 X_i)]^{-1}$$

$$\ln\left(\frac{pi}{1-pi}\right) = \beta_0 + \beta_1 X_i$$

ويمكن التعبير عنها بالشكل :

$$\ln L(\beta_0, \beta_1) = \sum_{i=1}^{n} Yi(\beta_0 + \beta_1 X_i) - \sum_{i=1}^{n} \ln[\ 1 + \exp(\beta_0 + \beta_1 X_i)]$$

وهذه هي دالة الاحتمال للمعلمات المراد تقديرها .

ومن خصائص عنصر الخطأ :

1. حد الخطأ لا يتوزع توزيعاً طبيعياً للمتغير ثنائي الاستجابة وان:

$$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta} Xi)$$

ممكن ان تاخذ قيمتين فقط وهي :

1- If Y_i=1 : e_i=1- $\beta_0 - \beta X_i$

2- If Y_i=0 : e_i= - $\beta_0 - \beta X_i$

2. تباين حد الخطا غير ثابت : حد الخطا(ei) ليس له تباينات متساوية في حالة المتغير ثنائي الاستجابة :

$$\sigma^2(Yi) = E[Yi - E(Yi)]^2$$
$$= (1\text{-pi})^2\text{pi} + (0\text{-pi})^2\ (1\text{-pi})$$

ان تباين e_i هو تباين Y_i نفسه بسبب e_i=$Y_i - p_i$ و pi ثابت

$$\sigma^2(e_i) = p_i(1 - p_i) = \big(E(Y_i)\big)(1 - E(Y_i)$$

$\sigma^2(e_i)$ تعتمد على X_i لذا فان تباينات عنصر الخطأ تختلف باختلاف قيم X_i.

3. القيود على الدالة ثنائية الاستجابة : بما ان دالة الاستجابة تمثل احتمالات المتغير المفسر عندما يكون قيمة متغير الاستجابة هو 0,1 يجب ان تكون متوسط الاجابة مقيدة على النحو الآتي:

$$0 \leq E(Y_i) = p_i \leq 1$$

2.5.2 تحليل الدالة التمييزية : [35],[16] Discriminant Function Analysis :

يعد التحليل التمييزي أحد الطرائق المتبعة في تحليل متعدد المتغيرات وهو يبين تاثير مجموعة من العوامل (المتغيرات التفسيرية) على مجموعتين او اكثر مختلفة فيما بينها بمعنى تتميز مجموعة عن الاخرى . ويعمل على تصنيف المفردات حسب انتمائها الى المجموعة وكذلك التعرف على المتغيرات التي ساهمت في التصنيف .

لنفرض ان لدينا البيانات (X: X_1 , X_2 , …., X_P) أخذت لمجتمعين A_1 , A_2. بالمعالم $(\mu_1, \sum)$ و $(\mu_2, \sum)$ على الترتيب اي ان المجتمعين يختلفان في الوسط ولهما مصفوفة التباين المشترك نفسها $(\sum)$. فالدالة الخطية Y يمكن ان تكتب بالصيغة الاتية :

$$Y = a'X = a_1X_1 + a_2X_2 + \cdots \ldots + a_pX_p \qquad \ldots\ldots. (21\text{-}2)$$

اذ ان Y تمثل تركيبة خطية للمتغيرات التوضيحية . اذ ان المعاملات تقدر لتعطي أفضل تمييز بين المجموعتين وذلك بجعل نسبة التباين بين المجموعات الى التباين داخل المجموعات اكبر ما يمكن .فاذا رمزنا الى نسبة التباين بالرمز R فيكون :

$$R = \frac{\boldsymbol{Betwween\ Groups}}{\boldsymbol{Within\ Groups}}$$

وان مجاميع المعاملات الخطية كما ياتي :

$$a' = (\mu_1 - \mu_2)'\sum^{-1}$$

فيكون المركب الخطي Y بالشكل :

$$Y = a'X = (\mu_1 - \mu_2)'\sum^{-1}X \quad \ldots\ldots (22\text{-}2)$$

وهذه تسمى دالة فشر التمييزية الخطية Fisher's Linear discriminant function والتي تعمل على تعظيم الفرق بين متوسطي المجموعتين . ثم نجد النقطة الوسطية midpoint او(نقطة التصنيف) The classification point للحصول على تصنيف للمشاهدات عن طريق ايجاد المتوسط للمجتمعين وكالآتي :

بما ان الوسط الحسابي للمجتمعين A_1 , A_2 على الدالة Y هو :

$$\mu_{1Y} = a'\mu_1 \quad ; \quad \mu_{2Y} = a'\mu_2$$

فان النقطة الوسطية بين متوسطي المجتمعين تكون

$$m = 1 / 2(\mu_{1Y} + \mu_{2Y})$$
$$= 1/2\,(a'\mu_1 + a'\mu_2)$$

لذلك فان التصنيف سيكون بمجرد تحديد النقطة الوسطية وعليه فان :

if $Y_i = a'X_i \geq m$ فان X_i تصنف للمجتمع الاول A_1

if $Y_i = a'X_i < m$ فان X_i تصنف للمجتمع الثاني A_2

او بعبارة اخرى

if $Y_i = a'X_i - m \geq 0$ فان X_i تصنف للمجتمع الاول A_1

if $Y_i = a'X_i - m < 0$ فان X_i تصنف للمجتمع الثاني A_2

المعاملات التمييزية المعيارية[9]: Standardization Discriminant coefficients

تتمثل المعاملات التمييزية المعيارية بقيم (a) الظاهرة في المعادلة الآتية :-

$$Y^* = a_1^* x_1 + a_2^* x_2 + \ldots\ldots\ldots + a_p^* x_p \quad \ldots\ldots\ldots\ldots\ldots (23\text{-}2)$$

إذ إن :

Y^* : القيمة التمييزية المعيارية.

x_p : المتغيرات التمييزية المعيارية.

a_p : المعاملات التمييزية المعيارية.

p: عدد المتغيرات التمييزية المعيارية المكونة للمعادلة التمييزية,ويساوي(عدد المجموعات−1).

وتُستعمل المعادلة التمييزية المعيارية في تحديد اهمية المتغيرات في تكوين دالة التمييز, إذ إن المتغيرات التي تكون القيمة المطلقة لمعاملها كبيرة تسهم بشكل كبير في تكوين المعادلة التمييزية , وتعني اشارة المعامل التمييزي المعياري ان مساهمة النسبة في التمييز هي مساهمة موجبة او سالبة .

ويتم ايضاً باستعمال المعادلة التمييزية المعيارية تحديد الحد الفاصل بين المعاملات التمييزية بين المجاميع , اذ يمثل الحد الفاصل الوسط الحسابي للمعلامات التمييزية المعيارية.

6.2 مشاكل الانحدار : Regression Problem :

تعتمد نماذج الانحدار على افتراضات تتعلق ببناء النموذج , إذ انه في حالة عدم توفر تلك الافتراضات فأن الانموذج سيعاني جملة من المشكلات والتي تجعل عملية التقدير خاطئة او في بعض الاحيان غير ممكنة , وتعاني كثيراً من نماذج الانحدار عدد من المشاكل واهمها مشكلة التعدد الخطي والارتباط الذاتي وعدم تجانس التباين , وتكمن المشكلة في أن الكثير من الباحثين يقومون بتقدير النماذج دون فحص للبيانات بوجود مشكلة من عدمها او فحصها والإبقاء عليها دون حلها ومن ثم يؤدي الى تقديرات غير دقيقة .

1.6.2 مشكلة التعدد الخطي: [52],[50] Multicollinearity Problem :

وهي وجود علاقة ارتباط خطي تام بين متغيرين تفسيريين او بين جميع المتغيرات أي $(r_{x1,x2} = 1)$ والتي تؤدي الى زيادة في البواقي (الأخطاء) كما تسبب عدم إمكانية استعمال طريقة المربعات الصغرى في التقدير اذ ان $(X'\mathrm{X} = 0)$ ومن ثم لا يمكن ايجاد معكوس المصفوفة $(X'\mathrm{X})$ وان X مصفوفة ابعادها $(n \times p)$ لمشاهدات المتغيرات التفسيرية, ويطلق على هذه لحالة بالتعدد الخطي التام (Exact Multicollinearity) أما التعدد الخطي غير

التام (In exact Multicollinearity) فتحدث عندما وجود علاقة ارتباط خطي غير تام بين المتغيرات . اما كيفية التعامل مع البيانات التي تظهر فيها مشكلة التعدد الخطي فهناك اقتراحات منها :

في البيانات الرقمية يمكن حل المشكلة باضافة بعض البيانات أو تغيير مواصفات الانموذج (حذف بعض المتغيرات أو تحويل المتغيرات) أو إعادة النظر في الفرضيات أو عن طريق اختيار طريقة لتقدير المعالم مناسبة او باستعمال إنحدار الحرف أو إنحدار المركبات الرئيسة .

أما في البيانات الفئوية هناك طريقة واحدة لحل هذه المشكلة وهي التحليل الفئوي المركبات الرئيسة (CATPCA) Categorical Principal Component Analysis التي تستعمل للبيانات الأسمية والرتبية والرقمية والتي قد لا تكون لها علاقة خطية مع بعضها البعض. في هذه الطريقة يتم تقليص البيانات التي تحول تسميات الفئة الى قيم رقمية كما في الانحدار اللوجستي متعدد الحدود Multinomail Logistic Regression .

الكشف عن مشكلة التعدد الخطي :[26]

ان الكشف عن مشكلة التعدد الخطي تعني الكشف عن درجة التعدد الخطي وليس الغرض منه هو وجود او عدم وجود التعدد الخطي ولغرض البيان عن التعدد الخطي في انموذج يحتوي على أكثر من متغيرين توضيحيين لا بد من الأخذ بنظر الاعتبار النقاط الآتية:

1) ارتفاع قيمة R^2 وعدم معنوية المعلمات المقدرة للمتغيرات التوضيحية .
2) ارتفاع معاملات الارتباط بين المتغيرات التوضيحية .
3) حساب معامل تضخم التباين (VIF) لكل متغير من المتغيرات التوضيحية

$$VIF_j = \frac{1}{1-R_j^2} \qquad j = 1,2,\ldots,p$$

p: تمثل عدد المتغيرات التوضيحية .

R_j^2 : تمثل معامل التحديد للمتغير التوضيحي X_j .

4) ايجاد الدليل الشرطي: Condition Index (C I) ونسب التباين Proportions Variance و يؤخذان سوية ونستفيد منهما في بيان درجة التعدد الخطي والمتغيرات المرتبطة مع بعضهما البعض فاذا كانت قيمة الدليل الشرطي بحدود (5− 10) وان أثنين او اكثر من نسب التباين اقل من 0.50 فانه يدل على أن الارتباط ضعيف . أما اذا كانت

$10 \leq CI \leq 30$ فهذا يعني ان هناك تعدداً خطياً من المعتدل الى العالي . اما اذا تجاوزت (30) فهذا يدل على ان التعدد الخطي بدرجة اكبر . ولغرض ايجاد الدليل الشرطي لابد اولا من احتساب الجذور المميزة (eigen values) لانها توضح كمية الاختلافات الكلية بين المتغيرات . فعندما تكون الجذور المميزة مساوية للصفر فانه يدل على التعدد الخطي اما اذا كانت قريبة من الصفر فهذا يدل على وجود تعدد خطي عالٍ . اما اذا كانت تساوي (1) فتعد الحالة الامثل في عدم وجود مشكلة التعدد الخطي ويمكن ايجاد العدد الشرطي (K) بالصيغة الاتية :

$$\boldsymbol{K = \frac{Maximum\ eigen\ value}{Minimum\ eigen\ value}}$$

ومن ثم فان الدليل الشرطي CI يعرف بالصيغة الاتية :

$$\boldsymbol{CI = \sqrt{\frac{Maximum\ eigen\ value}{Minimum\ eigen\ value}} = \sqrt{K}}$$

طرائق معالجة مشكلة التعدد الخطي :[19],[23]

فضلاً عن حذف قسم من المتغيرات واضافة بعض البيانات هناك عدة طرائق لحل مشكلة التعدد الخطي منها:

1) تحليل المركبات الرئيسة : Principle Components Analysis (PCA):

يعد تحليل المركبات الرئيسية احد الاساليب المهمة المستعملة في تحليل البيانات متعدد المتغيرات ويستعمل عندما تكون المتغيرات التوضيحية ذات إرتباط عالٍ أي وجود مشكلة التعدد الخطي . كما يستعمل في تحليل البيانات ذات العدد الكبير للمتغيرات التوضيحية نسبة الى عدد المشاهدات والذي يعد أحد التقنيات المستعملة لتلخيص البيانات واختصارها. ويستند الى ايجاد الجذور المميزة والمتجهات المميزة لمصفوفة الارتباطات .

تعتمد طريقة تحليل المركبات الرئيسة على اسلوب تحليل المتغيرات التوضيحية الاصلية الى متغيرات جديدة غير مرتبطة . هذه المتغيرات الجديدة تسمى بالمركبات الرئيسة اذ ان كل مكون رئيسي هو عبارة عن تركيبة خطية في المتغيرات التفسيرية الأصلية ويتم تحويل المتغيرات التفسيرية الى المكونات الرئيسة بالشكل الآتي :

$$Y^* = \beta_0 \underline{I} + XVV'\underline{\beta} + \underline{e}$$

و(V) عبارة عن مصفوفة المتجهات المميزة لمصفوفة الارتباط بين المتغيرات التوضيحية فاذا عوضنا XV بقيمة ثابتة PC مصفوفة ذات بعد $(n \times p)$ أعمدتها عبارة عن معاملات انحدار النموذج المحدد الذي يأخذ الشكل الآتي :

$$Y^* = \beta_0 \underline{I} + PC \ \underline{\propto^*} + \underline{e}$$

كما ويعرف مقدار التحيز للمكونات الرئيسة كالاتي :

$$b(\beta_{pc}^0)E(\beta_{pc}^{\wedge}) - \beta = VV'\beta$$

ونستفيد من تحليل المركبات الرئيسة في عدة جوانب منها :

أ- يلخص المتغيرات في عدد اقل من العوامل الرئيسة التي يمكن ان تفسر الظاهرة وذلك عن طريق الحصول على مجموعة جديدة من المتغيرات (العوامل) وبعدد اقل لتحل جزئيا او كليا محل المجموعة الاصلية من المتغيرات .

ب- التعرف على المتغيرات التي لها دلالة احصائية مهمة والتي تتطلب مزيداً من عمليات التحليل الاخرى .

ت-يحل مشكلة الارتباطات العالية بين المتغيرات التفسيرية والتي تؤدي الى تقديرات خاطئة لمعاملاتها في حال بقاء الارتباط بينها .

2- طريقة انحدار الحرف: [26] Ridge Regression (R.R) :

تعتبر طريقة إنحدار الحرف أحد طرائق معالجة مشكلة التعدد الخطي وتتلخص هذه الطريقة بإضافة كمية صغيرة موجبة تقع قيمتها بين الصفر والواحد $(0 < C < 1)$ الى العناصر القطرية لمصفوفة المعلومات $(X'\mathrm{X})$ للحصول على مقدرات أكثر دقة . و تعمل هذه الطريقة على فك الارتباطات بين المتغيرات التوضيحية . وتستعمل الصيغة الآتية في ايجاد تقديرات قيم المعلمات باستعمال طريقة انحدار الحرف على ان يتم تحويل المتغير المعتمد والمتغيرات التوضيحية الى صيغتها المعيارية .

$$\hat{\beta}_R = (X'\mathrm{X} + CI_P)^{-1}X'Y$$

اذ ان (I) مصفوفة الوحدة Identity matrix . وعندما تكون قيمة $(C = CI = 0)$ فان تقديرات معلمات طريقة انحدار الحرف تساوي تقديرات معلمات طريقة المربعات الصغرى الاعتيادية . وعندما تكون $(C > 0)$ فان مقدرات انحدار الحرف تميل الى الاستقرار عند قيمة

معينة نسبة للتغيرات في البيانات ولكنها تكون متحيزة . كما ان متوسط مربعات الخطأ لمقدرات انحدار الحرف تكون أقل من متوسط مربعات الخطأ لطريقة المربعات الصغرى الاعتيادية أي ان $MSE_{B.RR} < MSE_{B.OLS}$. لهذا نقبل مقدار معين في التحيز مقابل تقليل التباين للمقدرات .

3- استعمال طريقة المربعات الصغرى الجزئية [47]: Partial Least Squares (PLS):

وهي شائعة الاستعمال في الطب ولاسيما في المعالجات السريرية اذ قلة عدد المشاهدات (المرضى) مع كثرة عدد المتغيرات $(X's)$ المتمثلة بأعراض المريض مع عدد المتغيرات المعتمدة $(Y's)$ والمتمثلة بالمستوى الصحي للمريض . وتتميز بامكانية تحليل البيانات لمصفوفة من المتغيرات المعتمدة مع مصفوفة في المتغيرات التوضيحية في آن واحد. وتستعمل في حال عدد المتغيرات التوضيحية يفوق عدد المشاهدات ففي هذه الحالة ستكون مصفوفة المعلومات شاذة singular وهنا ستظهر أيضا مشكلة التعدد الخطي بين المتغيرات التوضيحية. تعتمد طريقة انحدار المربعات الصغرى الجزئية على مصفوفة التباين المشترك $Cov(X,Y)$ إذ أن هذه الطريقة تسمح بتحديد العوامل والتي هي عبارة عن تراكيب خطية للمتغيرات التوضيحية X. وتعرف هذه العوامل بالمتغيرات الصماء والتي بدورها تكون أفضل نموذج للمتغيرات المعتمدة $(Y's)$.

2.6.2 مشكلة الارتباط الذاتي: [5],[35] Autocorrelation Problem :

تظهر هذه الظاهرة نتيجة مخالفة أحد فرضيات نموذج الانحدار الخطي، وتتعلق المخالفة في فرضيات عنصر الخطأ ei التي تنص بعدم وجود ارتباط ذاتي بين قيم المتغير العشوائي و تحدث هذه المشكلة عندما يحصل ارتباط طردي بين عنصري الخطأ *error terms* في مشاهدتين متتاليتين وتأخذ هذه الفرضية في النموذج الخطي الصيغة الآتية :

$$E(e_i e_j) = 0 \qquad i \neq j$$

يمثل عنصر الخطأ المسافة بين القيمة الفعلية للمتغير التابع والقيمة المتوقعة . ومفهوم الارتباط الذاتي هو كون المتغير العشوائي الذي يحدث في مدة يرتبط بالمتغير العشوائي الذي يسبقه او يليه مما يؤدي الى $Cov(e_t e_{t-1}) \neq 0$ وهذا لا يعني وجود ارتباط بين المتغيرات العشوائية كما في مشكلة التعدد الخطي .

والنوع الاول من الارتباط الذاتي وهو الاكثر شيوعاً عندما يميل عنصر الخطأ إلى أن يتأثر بعنصر الخطأ الذي يسبقه مباشرة في المدة الزمنية السابقة . ونحن نسمي هذا الارتباط الذاتي الأول لأن مدة زمنية واحدة فقط تفصل بين اثنين من المشاهدات المتصلة .

الكشف عن مشكلة الارتباط الذاتي :

هناك عدة طرائق للكشف عن وجود الارتباط الذاتي بين قيم الخطأ العشوائي ei من أهمها اختبار Durbin-Watson وصيغة هذا الاختبار هي:

$$D.W = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \quad ; \quad e_t = Y_t - \widehat{Y_t}$$

وكذلك اختبار Durbin – h وصيغته :

$$h - D = \hat{\rho}\sqrt{\frac{n}{1 - n\,var(\hat{B})}} \quad ; \quad \hat{\rho} = 1 - \frac{D.W}{2}$$

ويستعمل للمتغيرات التي تحتوي على التباطؤ الزمني Lagged Variable .

طرائق معالجة مشكلة الارتباط الذاتي :

قبل ان يتم تطبيق احدى طرق المعالجة لا بد من تدقيق مايلي :

1) عندما يكون السبب اهمال احدى المتغيرات التفسيرية او اكثر تتم إضافة ذلك المتغير .
2) عندما يكون السبب الصياغة غير الدقيقة تتم إعادة الصياغة وفق الواقع الصحيح للعلاقة.
3) عندما يكون السبب وجود علاقة بين قيم حد الخطأ يتم تحويل المتغيرات التفسيرية وفق الطرائق المذكورة فيما يأتي لغرض التخلص من الارتباط الذاتي .

أ- طريقة التحويل Transformation Method او طريقة كوكران اوركتCokran –Orkut:

وتعتمد على قيمة (ρ معامل الارتباط الذاتي) في حالة كونها معلومة القيمة .

ب – طريقة التكرار Iteration method :

وفي هذه الطريقة نحسب اولا قيمة معامل الارتباط الذاتي $\hat{\rho}$ كما يأتي :

$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$$

ثم نحسب الفرق الاول للمتغير التابع والتفسيري كما يأتي :

$$Y^* = Y_t - \rho Y_{t-1} \quad ; \quad X^* = X_t - \rho X_{t-1}$$

ولكي لا نفقد قيمة المشاهدة الاولى نقدرها كما يأتي :

$Y_1^* = Y_1\sqrt{1-\rho^2}$; $X_1^* = X_1\sqrt{1-\rho^2}$

ثم نعيد تقدير المعادلة الاولية بطريقة OLS على البيانات المحولة . وهكذا يتم تكرار العمليات آنفة الذكر لحين التخلص من المشكلة .

3.6.2 مشكلة عدم تجانس تباين حد الخطأ [10],[7] : Heteroscedasticity

:

حالة تجانس التباين وهي الحالة التي يكون تباين المتغير العشوائي مساوٍ لقيمة ثابتة. وفي حالات كثيرة قد لا يساوي التباين قيمة ثابتة ، وعليه نحصل على حالة تسمى عدم تجانس التباين (Heteroscedasticity) , إن ظاهرة عدم التجانس تؤثر في تقديرات تباين مقدرات النموذج .

وأن الاختبارات المستخدمة كاختبار t واختبار F تصبح في هذه الحالة غير واقعية ولا يمكن الاعتماد عليها ، و إذا لم يتحقق الافتراض الخاص بتجانس تباين الخطأ فإنه سيترتب على ذلك أن الصيغ الخاصة بتباينات المقدرات $\hat{Y}, \hat{B}$ لا يمكن تطبيقها لانها حسبت تحت فرض ثبات التباين δ_e^2 .

الكشف عن مشكلة عدم التجانس:

ويتم الكشف عن مشكلة عدم تجانس التباين بواسطة عدة اختبارات منها :

1- اختبار معامل ارتباط الرتب لسبيرمان Spearman's Rank Correlation Test :

ويمكن تطبيق هذا الاختبار في حالة العينات الصغيرة والكبيرة على حد سواء . وذلك عن طريق حساب قيم البواقي عن طريق الصيغة الاتية :

$ei = Y_i - \hat{y}i$

ثم ترتيبها (باهمال الاشارة) تصاعدياً او تنازلياً مع قيم Xi واعطاء الرتب وحساب فرق الرتب Di وتطبيق قانون سبيرمان للرتب بالصيغة الآتية :

$$r = 1 - \frac{6\sum Di^2}{n(n^2-1)}$$

وكلما اقتربت قيمة (r) من الواحد الصحيح دل على وجود علاقة قوية بين ei و Xi ومن ثم وجود مشكلة عدم تجانس التباين .

2- اختبار كولد فيلد وكوانت Goldfeld and Quandt test :

وهو من الاختبارات المهمة ويكون استعماله في العينات الكبيرة . ويعتمد على أساس تقسيم العينة الى عينتين جزئيتين $(N = n_1 + n_2)$ ثم تحسب قيمة (F) عن طريق الصيغة الآتية :

$$F_C = \frac{S_{i2}^2}{S_{i1}^2}$$

اذ ان :

$$S_{i1}^2 = \frac{\sum ei^2}{n_1 - 2}$$

$$S_{i2}^2 = \frac{\sum ei^2}{n_2 - 2}$$

فاذا كانت قيمة F_C المحسوبة أصغر من نظيرتها الجدولية عند مستوى معنوية محدد ودرجة حرية $(n_2 - 2)$ للبسط و $(n_1 - 2)$ للمقام , فان الانموذج خالي من مشكلة عدم تجانس التباين لحد الخطأ .

3- اختبار بارتليت Bartlett Test :

يستعمل هذا الاختبار في حالة توفر اكثر من مشاهدة لكل قيمة من قيم المتغير التفسيري

طرائق معالجة مشكلة عدم التجانس :

وتكون معالجة المشكلة بتحويل صيغة النموذج الى صيغة اخرى ينتج عنها قيم متساوية في قطر المصفوفة $\sigma^2 I$ ويتحقق ذلك بتحويل المتغيرات الى نسب عن طريق قسمة طرفي المعادلة على احد المتغيرات او على الجذر التربيعي لاحد المتغيرات ومنها :

1- تحويلة الجذر التربيعي Square Root Transformation :

وتستعمل هذه التحويلة اذا كانت تباينات المشاهدات متناسبة مع متوسطاتها .

$$Y^* = \sqrt{X}$$

2- التحويلة اللوغارتمية Logarithmic Transformation :

تستعمل اذا كانت الانحرافات المعيارية للمشاهدات تتناسب طرديا مع متوسطاتها .

$$Y^* = \log(X)$$

3- تحويل المقلوب Reciprocal Transformation :

تستعمل اذا كان هناك تناسب بين الانحرافات المعيارية وجذور المتوسطات .

$$Y^* = \frac{1}{X}$$

كما توجد اساليب اخرى لحل مشكلة عدم التجانس منها :

1) استعمال الاحصاء اللامعلمي (كروسكال واليس) : اذا كانت البيانات لا تحقق فرض تجانس التباين يعد اختبار كروسكال واليس بديلاً لا معلمياً لتحليل التباين الأحادي .
2) استعمال طريقة البوتستراب : والفكرة الاساسية لهذه الطريقة هي المعاينة بدون ارجاع لاكبر عدد من العينات ذات الحجوم المتساوية وفي كل مرة نحسب التباين حتى يتكون لدينا عدد كبير من التباينات تسمى "توزيع البوتستراب للتباين" والقيمة المتوقعة له تصبح تقديرا لتباين المجتمع , ويشترط في هذا التوزيع الاستقلالية في البيانات .
3) كما تستعمل طريقة المربعات الصغرى الموزونة (Weighted Least Squares) لمعالجة هذه المشكلة ويمكن الحصول على الوزن Wi عن طريق إيجاد معكوس التباين و عن طريق تقليل المجموع المرجح لمربعات الخطأ اذ ان الاوزان تتناسب عكسياً مع تباين الخطأ .

7.2 طرائق تقدير معاملات النماذج : Methods of estimation :

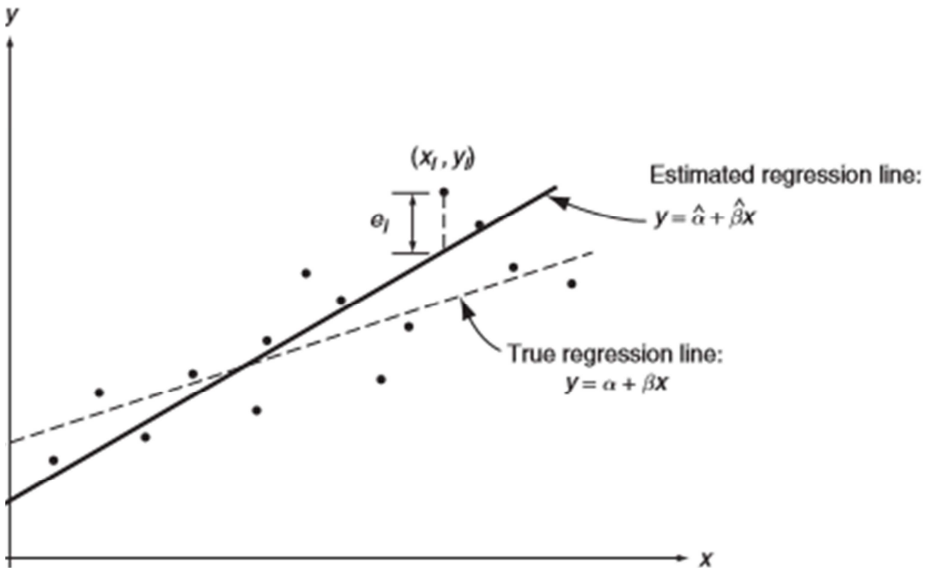
يمكن تقدير العلاقة بين المتغيرات عن طريق تقدير معاملات الانحدار باستعمال الطرائق الرياضية والاحصائية واختبار مدى معنويتها . وهناك عدة طرائق للتقدير الا ان اهمها طريقة المربعات الصغرى (OLS) وطريقة الامكان الاعظم (MLE) التي سيتم اعتمادهما كما سنتطرق الى بعض الطرائق الشائعة .

1.7.2 طريقة المربعات الصغرى الاعتيادية :[46]

Ordinary Least Square method (OLS) :

تعد هذه الطريقة من الطرائق الشائعة وتعتمد على وجود علاقة بين متغيرين او اكثر ولتقدير معلمات الانحدار B بطريقة المربعات الصغرى تشير الى ان المقدرات $\hat{B}$ يتم إختيارها بحيث مجموع مربعات الانحرافات بين قيم المشاهدات Y_i والقيم المقدرة المتوقعة أقل ما يمكن وتكتب بالشكل :

$$ei = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1)$$



شكل(2-1)

مقدرات طريقة المربعات الصغرى الاعتيادية OLS

الشكل السابق يعطي عرضاً بيانياً لهذه الاجزاء , نلاحظ ان البواقي هي المسافة العمودية بين القيم المشاهدة لy ومقدرات المربعات الصغرى $(\hat{\beta}_0 + \hat{\beta}_1 \mathrm{x}_i)$ التي يمثلها خط الانحدار الحقيقي $(\beta_0 + \beta_1 X_1)$.

نجد المقدرات $\widehat{B}_1$و $\hat{B}_0$ عن طريق أخذ المشتقات الجزئية لـ (Q) المعطاة في المعادلة الآتية ومساوات المشتقة الى الصفر.

$$Q = \sum_{i=1}^{n} ei^2 = \sum \left[\mathrm{y}_i - (\hat{\beta}_0 + \hat{\beta}_1 \mathrm{x}_i)\right]^2 \quad \ldots\ldots\ldots\ldots (24\text{-}2)$$

$$\frac{\partial Q}{\partial B_0} = \sum_{i=1}^{n} -2\left[\mathrm{y}i - (\hat{\beta}_0 + \hat{\beta}_1 \mathrm{x}_i)\right]$$

$$\frac{\partial Q}{\partial B_1} = \sum_{i=1}^{n} -2\mathrm{x}_i\left[\mathrm{y}i - (\hat{\beta}_0 + \hat{\beta}_1 \mathrm{x}_i)\right]$$

وعند تبسيط المعادلتين نحصل على:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum \mathrm{x}_i = \sum \mathrm{y}_i$$

$$\hat{\beta}_0 \sum \mathrm{x}_i + \hat{\beta}_1 \sum \mathrm{x}_i^2 = \sum \mathrm{x}_i \mathrm{y}_i$$

باستعمال الحذف والتعويض نحصل على:

$$\hat{\beta}_1 = \frac{n\sum \mathrm{x}_i \mathrm{y}_i - \sum \mathrm{x}_i \sum \mathrm{y}_i}{n\sum \mathrm{x}_i^2 - (\sum \mathrm{x}_i)^2} \quad \ldots\ldots\ldots\ldots (25\text{-}2)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \text{.....(26-2)}$$

للتأكد من أن هذه الحلول تتوافق مع الحد الأدنى من مجموع المربعات للتحقق من ذلك:

$$\frac{\partial^2 Q}{\partial \beta_0^2} = 2n > 0 \ ; D = \begin{bmatrix} \frac{\partial^2 Q}{\partial \beta_0^2} & \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 Q}{\partial \beta_1^2} \end{bmatrix} = 4n \sum (\mathrm{x}_i - \bar{\mathrm{x}})^2 > 0$$

نلاحظ ان D=0 اذا أخذت كل قيم xi القيمة نفسها , ومن ثم هناك على الاقل قيمتين مختلفتين لتحديد β_1 و β_0 .

أما في حالة انموذج الانحدار المتعدد ولعينة حجمها (n) من المشاهدات وباستعمال المصفوفات تكون بالشكل الآتي :

$$\begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \vdots \\ \mathrm{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{x}_{11} & \mathrm{x}_{21} & \cdots & \mathrm{x}_{m1} \\ 1 & \mathrm{x}_{12} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathrm{x}_{1n} & \mathrm{x}_{2n} & \cdots & \mathrm{x}_{mn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\underline{Y} = X\underline{\beta} + \underline{e}$$

.....(27−2)

$\underline{Y}$: تمثل متجهاً لمتغير الاستجابة ذا رتبة (n×1) .

X : تمثل مصفوفة المتغيرات التفسيرية ذات رتبة $[n \times (m+1)]$.

$\underline{\beta}$: تمثل متجهاً لمعلمات الانموذج المراد تقديرها ذا رتبة $[(m+1) \times 1]$.

$\underline{e}$: تمثل متجهاً للاخطاء ذا رتبة (n×1) .

وهناك طرائق متعددة لحساب قيم $\widehat{\boldsymbol{\beta}}$ منها [22] :

1) طريقة المحددات (Gramer Method) (القيم الاصلية) :

2) طريقة المعكوس Inverse Method (قيم اصلية):

3) طريقة كاوس Gauss Method :

4) طريقة كاوس (المصفوفة الموسعة) Expand Matrix Gauss Method

5) الحل بطريقة المعكوس للمصفوفة الثنائية (مصفوفة الانحرافات)

6) طريقة المحددات كريمر (المصفوفة الثنائية) :

7) طريقة كاوس للمصفوفة الثنائية :

8) طريقة كاوس المصفوفة الثنائية الموسعة :

خصائص مقدرات المربعات الصغرى[40],[46]:

يمكن التعبير عن خصائص مقدرات معاملات الانحدار $\hat{\beta}$عن طريق التعبير عن الانموذج بصيغة (المتجه– المصفوفة) بالمعادلة :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

لتكن : $\hat{\beta}$ تعبر عن مقدرات β في طريقة المربعات الصغرى للمشاهدات

$Y = (Y_1, Y_2, \dots, Y_n)'$ اذ ان Y_j (j=1,2,….,n) مستقلة وموزعة بشكل متماثل

$$E(e) = 0 \quad ; cov(e) = \sigma^2 I$$

يمكن تحديد متوسط وتباين المقدر $\hat{\beta}$ عن طريق :

$$\begin{aligned} E(\hat{\beta}) &= (X'X)^{-1}X'E(Y) \\ &= (X'X)^{-1}X'(X\beta + E(e)) \\ &= (X'X)^{-1}(X'X)\beta = \beta \end{aligned}$$

1- المقدرات $\widehat{\boldsymbol{\beta}}$ غير متحيزة للمعلمات $\boldsymbol{\beta}$:

$$cov(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = \sigma^2(X'X)^{-1}$$

2- العناصر القطرية في المصفوفة $\boldsymbol{\sigma^2(X'X)^{-1}}$ تعطي التباين لـ $\widehat{\boldsymbol{\beta}}$:

$$var(\hat{\beta}_0) = [\sigma^2 \sum x_i^2][n\sum(x_i - \bar{x})^2]^{-1}$$

$$var(\hat{\beta}_1) = \sigma^2[\sum(x_i - \bar{x})^2]^{-1}$$

التباينات المعطاة في المعادلتين السابقتين هي الحد الادنى من التباينات المرتبطة بالمقدرات غير المتحيزة ل $\hat{\beta}$.

3- طريقة المربعات الصغرى تقودنا الى مقدرات خطية غير متحيزة بأقل تباين :

Linear unbiased minimum-variance estimator for $\hat{\beta}_0$ and $\hat{\beta}_1$ من بين جميع المقدرات الخطية غير المتحيزة في y . ولبرهان ذلك :

تعد تقديراً خطياً غير متحيز للنموذج: $\beta^* = [(X'X)^{-1}X' + \mathrm{G}]\mathrm{Y}$

اذا كانت β^* هو الحد الادنى للتباين فان G=0. ويؤدي شرط عدم التحيز الى :

GX=0

نعد الآن مصفوفة التباين المشترك

$$cov(\beta^*) = E[(\beta^* - \beta)(\beta^* - \beta)']$$
$$cov(\beta^*) = \sigma^2[(X'X)^{-1} + GG']$$

من اجل تقليل التباين بمكونات β^* يجب تقليل كل عنصر قطري ل GG' . اذ العناصر القطرية iith للمصفوفة GG' تعطى عن طريق :

$$(GG')_{ii} = \sum_{i=1}^{n} g_{ij}^2$$

عندما g_{ij} هو العنصرijth في G يجب ان يكون $g_{ij}=0$ لكل i و j .هذه يعني G=0 .

وهذا حالة خاصة من نظرية كاوس– ماركوف Gauss-Markov Theorem .

أما خصائص عنصر الخطأ (ei) فيمكن الاشارة الى اهمها فيما ياتي[41] :

الخاصية الاولى : $\sum \hat{e}i = 0$

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}\mathrm{x}_i$$
$$\sum \hat{e}_i = \sum y_i - \hat{\beta}\sum x_i = N\bar{y} - \hat{\beta}N\bar{x} = 0$$

اذ ان $\bar{x} = \bar{\mathrm{y}} = 0$ عندما البيانات بصيغة الانحرافات .

الخاصية الثانية : $\sum \hat{e}i\mathrm{x}_i = 0$

$$\sum \hat{e}i\mathrm{x}_i = \sum \hat{e}ix_i + \sum \hat{e}i\bar{\mathrm{x}} = \sum \hat{e}ix_i$$

عن طريق الخاصية الاولى فان:

$$\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0$$

الخاصية الثالثة :

$$\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum \hat{\mathrm{e}}_\mathrm{i}(\hat{\mathrm{Y}}_\mathrm{i} - \bar{\mathrm{Y}}) = 0$$

(عن طريق الخاصية الاولى والثانية)

الخاصية الرابعة : S^2 **هو تقدير غير متحيز ل** σ^2 .

بطريقة الانحرافات (deviations form):

$\hat{e}_{\mathrm{i}} = y_{\mathrm{i}} - \hat{\beta} x_{\mathrm{i}}$

ولكن :

$$y_i = \beta x_i + e_i$$

$$\hat{e}_{\mathrm{i}} = \beta x_i + e_i - \hat{\beta} x_{\mathrm{i}} = (\beta - \hat{\beta}) x_i + e_i$$

وعلى فرض ان الأخطاءei تمتلك متوسطاً عينه يساوي صفراً وعند اعمامه على N من المشاهدات

$$\sum \hat{e}_i^2 = (\hat{\beta} - \beta)^2 \sum x_i^2 + \sum e_i^2 + 2(\hat{\beta} - \beta) \sum x_i e_i$$

وبأخذ التوقع لكلا الجانبين .

$$E(\sum \hat{e}_i^2) = E(\hat{\beta} - \beta)^2 \sum x_i^2 + E(\sum e_i^2) - 2E\left[(\hat{\beta} - \beta) \sum x_i e_i\right]$$

ولكن $var(\hat{\beta}) = E(\hat{\beta} - \beta)^2 = \frac{\sigma^2}{\sum x_i^2}$

$$var(e_i) = E(\sum e_i^2) = (N-1)\sigma^2$$

بالأضافة الى $(\hat{\beta} - \beta) = \sum x_i e_i / \sum x_i^2$ وحقيقة $\sum c_i x_{\mathrm{i}} = 1$

$$-2E\left[(\hat{\beta} - \beta)^2\right] \sum x_i^2 = -2\sigma^2$$

وبجمع النتائج السابقة نجد :

$$E(\sum \hat{e}_i^2) = \sigma^2 + (N-1)\sigma^2 - 2\sigma^2 = (N-2)\sigma^2$$

$$E(s^2) = \frac{E(\sum \hat{e}_i^2)}{N-2}$$

2.7.2 طريقة المربعات الصغرى الموزونة : [44]

Weighted Least Squares Method (WLS) :

وتستعمل في نماذج الانحدار الخطي غير متجانسة الاخطاء والتي يمكن ان تكون اكثر ملاءمة في حال تطبيق المربعات الصغرى الموزونة اذ يتم الحصول على تقديرات المعلمة عن

طريق تقليل المجموع المرجح لمربعات الخطأ اذ ان الاوزان تتناسب عكسياً مع تباين الخطأ. وهذا على النقيض من المربعات الصغرى الاعتيادية والتي يتم الحصول على تقديرات المعالم عن طريق التقليل الى الحد الادنى من الوزن المرجح لمربعات الخطأ.

$$\sum_{i=1}^{n} \frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2$$

ويتم تقديرالمعلمات $(\beta_0, \beta_1, \ldots, \beta_p)$ بطريقة المربعات الصغرى الموزونة WLS من خلال تقليل المقدار الآتي :

$$\sum_{i=1}^{n} \omega i(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$

اذ ان ωi هي اوزان تتناسب عكسياً مع تباينات الخطأ اي $\omega i = 1/\sigma_i^2$

على افتراض ان دالة التباين var(Y/X) هو نفسه لجميع قيم X ، ولنفرض ان لدينا دالة الانحدار $E(Y/X=xi)=\beta' Xi$ ولكن بدل ان نفترض ان الاخطاء ثابتة فاننا نفرض أن :

$$var(Y/X = xi) = var(ei) = \frac{\sigma^2}{\omega i}$$

اذ ان $\omega i = \omega 1, \ldots, \omega n$ هي اعداد موجبة معلومة .

اختلاف التباين لكل حالة يقودنا الى استعمال المربعات الصغرى المرجحة WLS بدلاً من المربعات الصغرى الاعتيادية OLS للحصول على التقدير للمعلمات .

لتكن $W_{n\times n}$ مصفوفة قطرية عناصر القطر الرئيسي ωi والانموذج الذي سوف نستعمله هو:

$$Y = X\beta + e \quad ; \quad var(e) = \sigma^2 W^{-1}$$

$\hat{\beta}$ هو مقدر β يتم الحصول على التقدير بواسطة الترجيح . المقدر $\hat{\beta}$ سنختاره لتقليل البواقي المرجحة من مجموع المربعات .

$$RSS(\beta) = (Y - X\beta)' W (Y - X\beta)$$
$$= \sum \omega i (yi - x')^2$$

استعمال البواقي المرجحة تبين ان بعض الاخطاء لاكثر المتغيرات ذات القيم الكبيرة لـ(ωi) ستكون اقل تباينا ومن ثم ستعطي وزن اكبر قي RSS الموزون .

WLS : مقدر عن طريق $\hat{\beta} = (X'WX)^{-1}X'WY$ وهذه المعادلة نستطيع ايجادها عن طريق OLS ثم ان كل نتائج OLS يمكن تطبيقها على WLS .

نفرض ان $W_{n\times n}^{1/2}$ مصفوفة قطرية عناصر القطر هي $\sqrt{\omega i}$

وحينئذ $W_{n\times n}^{-1/2}$ مصفوفة قطرية بعناصر القطر $\frac{1}{\sqrt{\omega i}}$ كذلك $W^{-1/2}W^{1/2}=I$

فان التباين المشترك $W^{1/2}$e هو :

$$var(W^{1/2}e) = W^{1/2}var(e)W^{1/2}$$
$$= W^{1/2}(\sigma^2 W^{1/2})W^{1/2}$$
$$= W^{1/2}(\sigma^2 W^{-1/2}W^{-1/2})W^{1/2}$$
$$= \sigma^2(W^{\frac{1}{2}}W^{-\frac{1}{2}})(W^{-\frac{1}{2}}W^{\frac{1}{2}}) = \sigma^2 I$$

هذا يعني ان المتجه $W^{1/2}e$ هو متجه عشوائي ذات مصفوفة تباين مشترك تساوي σ^2 عناصرها مصفوفة الوحدة . وبضرب جانبي معادلة الانحدار بـ$(W^{1/2})$:

$$W^{1/2}Y = W^{1/2}X\beta + W^{1/2}e$$

ليكن : $d = W^{1/2}e$; $M = W^{1/2}X$; $Z = W^{1/2}Y$

$$Z = M\beta + d$$

مما سبق $var(d) = \sigma^2 I$ و β في معادلة الانحدار الاخيرة يمكن حلها باستعمال OLS :

$$\hat{\beta} = (M'M)^{-1}M'Z \quad \text{..... (28-2)}$$
$$= \left[(W^{1/2}X)'(W^{1/2}X)\right]^{-1}(W^{1/2}X)'(W^{1/2}Y)$$
$$= (X'W^{1/2}W^{1/2}X)^{-1}(X'W^{1/2}W^{1/2}Y)$$
$$= (X'W\,X)^{-1}(X'W\,Y) \quad \text{..... (29-2)}$$

المربعات الصغرى الموزونة WLS لانحدار Y على X مع الاوزان التي تعطيها العناصر القطرية للمصفوفة W(مصفوفة الاوزان) هي نفسها كما في OLS لانحدار Z على M عندما :

$$M = \begin{bmatrix} \sqrt{\omega_1} & \sqrt{\omega_1}x_{11} & \cdots & \sqrt{\omega_1}x_{1p} \\ \sqrt{\omega_2} & \sqrt{\omega_2}x_{21} & \cdots & \sqrt{\omega_2}x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\omega_n} & \sqrt{\omega_n}x_{n1} & \cdots & \sqrt{\omega_n}x_{np} \end{bmatrix}; Z = \begin{bmatrix} \sqrt{\omega_1}y_1 \\ \sqrt{\omega_2}y_2 \\ \vdots \\ \sqrt{\omega_n}y_n \end{bmatrix}$$

مقارنة التقديرات بين طريقة OLS وطريقة WLS :

وللمقارنة بين كفاءة المقدرات في ظل فرضية عدم تجانس تباين الخطا العشوائي , نقوم بحساب كفاءة التقدير لطريقة المربعات الصغرى الموزونة نسبة الى تقدير طريقة المربعات الصغرى الاعتيادية فمثلا نقوم بحساب كفاءة تقدير الحد الثابت $\hat{\beta}_0$ كما يأتي :

كفاءة $\widehat{\boldsymbol{\beta}}_{\mathbf{0}} = \dfrac{var\,(\widehat{\boldsymbol{\beta}}_{\mathbf{0}})_{WLS}}{var(\widehat{\boldsymbol{\beta}}_{\mathbf{0}})_{OLS}}$

وكذلك يمكن حساب كفاءة $\hat{\beta}_1$ بنفس الطريقة

$$\text{كفاءة } \widehat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{var\,(\widehat{\boldsymbol{\beta}}_{\mathbf{1}})WLS}{var(\widehat{\boldsymbol{\beta}}_{\mathbf{1}})OLS}$$

فاذا كانت نتيجة النسبة مساوية الى (1) الواحد الصحيح هذا يعني تساوي كفاءة الطريقتين اما اذا كانت نتيجة اقل من الواحد دل ذلك على افضلية طريقة WLS بالتقدير على طريقة OLS .

3.7.2 طريقة تقديرالامكان الاعظم:[30] : Maximum Likelihood Estemator

من البديهي ان حساب قيم المعلمات وفق طريقة المربعات الصغرى OLS لا تتضمن افتراض عن شكل توزيع حد الخطأ العشوائي في الانموذج. اما في طريقة الإمكان الأعظم نكون بعض الافتراضات عن هذا التوزيع (وغالباً يتوزع طبيعياً) , ومن ثم يتم تعظيم عينة المشاهدات التي تمثلها البيانات .

إذ أن Yi=Y1,….,Yn تمثل مشاهدات المتغير المعتمد Y لعينة حجمها n وان قيم المتغير المعتمد تكون مستقلة الواحدة عن الاخرى وعلى افتراض ان :

$$Y \sim N(\beta X, \Sigma) \quad ; \quad e \sim N(0, \Sigma)$$

اذ ان :(βX) تمثل الوسط(μ), وان (Σ) تمثل التباين.

فانه يمكن التعبير عن دالة الكثافة الاحتمالية لكل مشاهدة كالاتي :

$$(\mathrm{y}_1, \mathrm{y}_2, \dots\dots, \mathrm{y}_n) = f(\mathrm{y}_i, \mu, \Sigma) \qquad \text{.....(30-2)}$$

$$L(\mathrm{y}_1, \mathrm{y}_2, \dots\dots, \mathrm{y}_n) = \prod_{i=1}^{n} f(\mathrm{y}_i, \mu, \Sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{(\sqrt{2\pi})^{p}|\Sigma|^{1/2}}\, e^{-(\mathrm{y}_i-\mu)'\Sigma^{-1}(\mathrm{y}_i-\mu)/2}$$

$$= \frac{1}{(\sqrt{2\pi})^{np}|\Sigma|^{n/2}}\, e^{-\sum_{i=1}^{n}(\mathrm{y}_i-\mu)'\Sigma^{-1}(\mathrm{y}_i-\mu)/2}$$

وتكون الدالة اعظم ما يمكن عندما $\hat{\mu} = \bar{\mathrm{y}}$ وباضافة وطرح $\bar{\mathrm{y}}$ في الأس نحصل :

$$-\frac{1}{2}\sum_{i=1}^{n}(\mathrm{y}_i - \bar{\mathrm{y}} + \bar{\mathrm{y}} - \mu)'\Sigma^{-1}(\mathrm{y}_i - \bar{\mathrm{y}} + \bar{\mathrm{y}} - \mu)$$

$$L = \frac{1}{(\sqrt{2\pi})^{np}|\Sigma|^{n/2}}\, e^{\frac{-\sum_{i=1}^{n}(\mathrm{y}_i-\bar{\mathrm{y}})'\Sigma^{-1}(\mathrm{y}_i-\bar{\mathrm{y}})}{2} - n\frac{(\mathrm{y}_i-\bar{\mathrm{y}})'\Sigma^{-1}(\mathrm{y}_i-\bar{\mathrm{y}})}{2}}$$

Σ^{-1} هي محدد موجب ونحن لدينا $-n\frac{(\mathrm{y}_i-\bar{\mathrm{y}})'\Sigma^{-1}(\mathrm{y}_i-\bar{\mathrm{y}})}{2} \leq 0$ وكذلك

$$0 \le e^{-n\frac{(\bar{y}_i-\bar{y})'\Sigma^{-1}(y_i-\bar{y})}{2}} \le 1$$

فان الحد الاقصى يكون عندما الأس يساوي(0) .اي عندما $\hat{\mu} = \bar{y}$.

وباخذ المشتقات الجزئية للمعلمات(β, Σ) نحصل على تقدير أقصى احتمال وذلك بعد أخذ (Ln) لطرفي المعادلة ومساوات المشتقة للصفر :

$$L = \frac{1}{(\sqrt{2\pi})^{np}|\Sigma|^{n/2}} e^{-\sum_{i=1}^{n}(y_i-\beta X)'\Sigma^{-1}(y_i-\beta X)/2}$$

$$\ln(L) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2} ln|\Sigma| - \sum_{i=1}^{n}(y_i - \beta X)'\Sigma^{-1}(y_i - \beta X)/2$$

$$\frac{\partial \ln(L)}{\partial \beta_0} = \frac{1}{2|\Sigma|}\sum_{i=1}^{n}(-2X'Y + 2X'X\beta) = 0$$

$$X'Y = X'X\beta$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \ldots\ldots (31\text{-}2)$$

وعن طريق الاشتقاق نحصل على تقدير المعلمات . فعند الاشتقاق ومساوات المشتقة للصفر يكون :

$$\frac{\partial \ln(L)}{\partial|\Sigma|} = \frac{n}{2|\Sigma|} + \frac{(Y'Y-2\beta'X'Y+\beta'X'X\beta)}{2|\Sigma|} = 0$$

$$\Sigma = \frac{(Y-X\beta)'(Y-X\beta)}{n} \; ; \quad e = Y - X\beta$$

8.2 معايير اختيار افضل انموذج انحدار: [2],[1].[29]

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 الانموذج والتي لها تاثير أو يعتقد أنها تؤثر في المتغير التابع (y) الذي يمكن التنبؤ به . كذلك ان عملية تحديد افضل انموذج 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 قليلاً ما قد يؤدي الى تنبؤات غير واقعية ومتحيزة بمعنى ان لا تكون آلية الاختصار تؤثر في المعلومات المستحصلة , بحيث نحصل على المعلومات نفسها كما لو استعملنا جميع المتغيرات . والصعوبة تكمن في اختيار المتغيرات الداخلة والمتغيرات المستبعدة , بمعنى اننا نفاضل بين ما يضيفة المتغير التفسيري في تفسير المتغير التابع وبين ما يضيفه من زيادة في التباين في حال

ضعف تاثيرة . لذلك هناك معايير خاصة للمفاضلة بين المتغيرات يجب معرفتها قبل الدخول في طرائق اختيار افضل انموذج ومنها :

أ- قيمة معامل التحديد R^2 The coefficient of determination :

ويحسب كالآتي : $R^2 = \frac{SSR(X1,\dots,Xp)}{SST} = 1 - \frac{SSE(X1,\dots,Xp)}{SST}$

اذ ان : $SSE = \sum(Yi - \hat{Y})^2$; $SSR = \sum(\hat{Y} - \bar{Y})^2$; $SST = \sum(Yi - \hat{Y})^2$

ونأخذ الانموذج الذي له أعلى R^2

ب- قيمة متوسط مربعات الخطأ Mean square error (MSE) :

وتحسب كالآتي: $MSE = \frac{SSE(X1,\dots,Xp)}{n-p}$

وناخذ الانموذج الذي له اقل MSE .

ج- قيمة اختبار F. F-Test

اختبار المعنوية الكلية للانحدار وتحسب كالآتي :

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

في حال F_C المحسوبة اقل او تساوي القيمة الجدولية ($F_C \leq F_T$) سوف لا نرفض فرضية العدم أي ان المتغيرات ($x_1, x_2, \dots, x_p$) ليس لها تاثير معنوي مع بعضها في انموذج الانحدار.

اما في حالة $F_C > F_T$ سوف نرفض فرضية العدم (H_o) أي انه يوجد على الاقل متغير واحد له أهمية في تفسير المتغير المعتمد Y_i أي أن لها تاثير معنوي في انموذج الانحدار.

ويمكننا ايضا عن طريق إختبار F إختبار جزء من عدد معاملات الانحدار الكلية وسوف يستعمل بهذا إختبار F الجزئي Partial F test والغرض منه لمعرفة في حال إضافة المتغيرات ($X_{k+1}, \dots X_p$) الى الانموذج الذي يحوي ($x_1, x_2, \dots, x_p$) هل له قيمة معنوية في التنبؤ .

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

كما يمكن استعمال هذا الاختبار لبيان تاثير كل متغير في انموذج الانحدار .

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

وغالباً ما يستعمل اختبار t (t-test) لهذا الغرض اذ ان :

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

د – قيمة إحصاءة مالو C_p Mallow 's statistics :

تعتمد قيمة هذا المقياس على متوسط مجموع مر بعات الأخطاء لقيم البيانات الداخلة في الأنموذج الذي يضم كل المتغيرات (P) المؤثرة في الانموذج وكذلك المتغيرات المراد معرفة تاثيرها m.

وتحسب كالآتي: $$C_p = \frac{SSE(X1,\ldots,Xp)}{MSe(X1,\ldots,Xm)} - (n - 2p)$$

اذ ان:

p: عدد المعالم في الانموذج ومن ضمنها B_0.

n: عدد المشاهدات.

m: عدد المتغيرات التفسيرية.

وهناك علاقة بين Cp و MSe و R^2 وكما ياتي [18] :

نحن نحاول إيجاد إنموذج انحدار يمثل المتغيرات التوضيحية المختارة أفضل تمثيل فيكون هذا الانموذج المفضل الذي يحتوي على أقل قيمة لـ Cp وأقل قيمة لـ MSe وأعلى قيمة لـ R^2 وبأقل عدد من المتغيرات.

ونلاحظ بان Cp تتناسب طردياً مع MSe, وكل من Cp و MSe يتناسبان عكسياً مع R^2.

ويمكن ان تأخذ قيمة Cp إشارة سالبة عندما تكون :

$$\frac{SSe(X_1 \ldots X_k)}{MSe(X_1 \ldots X_m)} < (n - 2p)$$

في حين تكون كل من MSe و R^2 إشارتهما موجبة وعند عمل ارتباط بين المقاييس الثلاثة نلاحظ ان الارتباط بين Cp و MSe يساوي واحد وبين MSe و R^2 يساوي سالب واحد وبين Cp و R^2 يساوي واحد .

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

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

اذ ان n عدد المشاهدات وp عدد المتغيرات التفسيرية ويؤخذ الانموذج الذي يعطي أقل قيمة للمعيار.

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

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

ويؤخذ الانموذج 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 الانموذج الملائم الذي يقابل اقل قيمة للمعيار

وبعد ادخال جميع المتغيرات في الانموذج نحاول اختصار عدد المتغيرات ليشمل الانموذج المتغيرات المؤثرة والتي تكون ذات دلالة معنوية للحصول على الانموذج الافضل وهناك طرائق متعددة لاختيار افضل انموذج (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 متغيراو للانموذج (p-1) من المتغيرات ومقارنتها بالجدولية ونقوم بحذف المتغير الذي لديه اقل قيمة ل F_c .وبعد حذف المتغير يصاغ انموذج آخر بالمتغيرات المتبقية ونكرر العملية . اما اذا كانت قيمة F_c اكبر من الجدولية فلا نحذف المتغير لانه معنوي . فضلاً عن احتساب قيمة إحصاءة مالوC_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 اعلى قيمة F_c .

وليس بالضرورة ان تؤدي نتائج الحذف العكسي الى نتائج الحذف الامامي نفسها.

3- طريقة الانحدار التدريجي The stepwise regression procedure

وهي عبارة عن دمج بين طريقتي الاختيار الامامي والاختيار العكسي. اذ يتم في الخطوة الاولى اختيار المتغير الذي يملك أعلى F_c من بين باقي المتغيرات المفسرة واضافته الى انموذج 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 الانموذج اما اذا كانت F الجزئية اقل من الجدولية لا ندخل المتغير ونتوقف عن الاختيار, واختيار متغير آخر يملك اعلى F_c ليدخل مع المتغير الاول .والخطوة الثالثة نجري عملية الاختيار العكسي لاختبار المتغيرات المختارة بالاعتماد على قيمة F_c الجزئية .

الفصل الثالث

الجانب التطبيقي لنماذج الانحدار والمؤشرات الاحصائي

الفصل الثالث
الجانب التطبيقي لنماذج الانحدار

1.3 تمهيد :

يعد معدل وفيات الأطفال الخدج من أهم المؤشرات الصحية التي تعكس مستوى الأحوال الصحية والمستوى المعاشي في أي مجتمع ، ومما لاشك فيه أنه إذا حقق مجتمع ما تقدماً صحياً أو اقتصادياً أو اجتماعياً ينعكس هذا التقدم أولاً على صحة ورفاهية المجتمع ولا سيما الاطفال. كذلك تعد وفيات الأطفال الخدج أكثر المؤشرات التي تعكس الأوضاع الاقتصادية والاجتماعية والصحية للمجتمع ، ولقد اهتم العديد من الباحثين بدراسة العلاقة بين وفيات الأطفال الخدج والمتغيرات المؤثرة في الوفاة للوقوف على مستويات الوفيات واتجاهات التغير في تلك المستويات .

وتمثل وفيات الأطفال مؤشراً حساساً لمستوى التنمية في اي بلد ، وتقدم دليلا على أولوياتها وقيمها. وعلى الصعيد العالمي، يموت أكثر من 000 26 طفل في المتوسط دون سن الخامسة لأسباب يمكن الوقاية منها في اغلب الاحيان. ويعيش جميعهم تقريبا في البلدان النامية. ويموت أكثر من ثلث هؤلاء ألاطفال خلال الشهر ألاول من الولادة عادة لعدم الحصول على الخدمات الصحية ألاساسية . وهناك حاجة إلى معلومات عن أسباب الوفاة لتحديد أولويات التدخلات والتخطيط لتنفيذها، والتأكد من فعالية التدخلات الخاصة بالأمراض، وتقييم الاتجاهات على اساس تلك الامراض. [56]

كما تناول العديد من الباحثين والدارسين وفيات الأطفال الخدج من جوانب عديدة وقد تباينت هذه الدراسات من ناحية أسلوت التحليل الاحصائي المتبع في تحديد العوامل المؤثرة في وفيات الاطفال ولا سيما في الشهر الاول من العمر وكانت من اهم هذه الاسباب هي مدة الحمل ووزن الطفل وامراض الجهاز التنفسي , والتشوهات الخلقية. وفي هذا البحث تم اختيار نماذج انحدار خطية ولاخطية كونها تتلائم مع هذه البيانات .والسبب في اختيار هذين الأسلوبين، هو أن هذه الدراسات يكون المتغير التابع نوعياً او كمياً وأن المتغيرات المستقلة تكون نوعية أو كمية أو مختلطة ما بين النوعين، لذا وجدنا أن أسلوب الانحدار المتعدد والانحدار اللوغارتمي وانموذج الانحدار الأسي وكذلك انموذج الانحدار اللوجستي والدالة التمييزية ينطبق عليها لوصف وبيان تأثير المتغيرات المستقلة في المتغير التابع.

2.3 ماهية البيانات التي اعتمدت في الدراسة :

لقد اعتمد البحث على شكلين من البيانات المسجلة , الشكل الاول بيانات كمية (تجميعية) تم فرزها بمجاميع لعدد المتوفين شهريا وتتضمن [معدل مدة الحمل , معدل عمر الام , معدل وزن الطفل , معدل عمر الطفل , حالة الولادة(مبكرة أم تامة), ونوع الولادة , واسباب الوفاة , والخديج المعقم والمنزلي] وهي ما تتطابق مع انموذج الانحدار المتعدد وانموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي. (لاحظ ملحق رقم 1 الخاص بالبيانات)

والشكل الاخر كانت بيانات وصفية تبين حالة كل طفل خديج عند الولادة وحالة خروجه من المستشفى اما حيا او ميتا مع معلومات الخديج المتمثلة بـ (مدة الحمل , وزن الطفل, عمر الطفل, عمر الام , حالة الولادة(مبكرة أم تامة) , ونوع الولادة , واسباب الوفاة , الجنس , والخديج المعقم والمنزلي) وهي بيانات تتطابق مع انموذج الانحدار اللوجستي ثنائي الاستجابة وتحليل الدالة التمييزية .

تم جمع البيانات عن طريق البيانات الموثقة في مركز دائرة صحة بابل عن كافة مستشفيات محافظة بابل لمدة ثلاث سنوات (2014 , 2015 , 2016) وذلك على شكل برامج (برنامج الاطفال الخدج) واستمارة وفيات الاطفال الخدج بعد اعادة فرزها وتدقيقها من قبل الباحث ومن ثم مقارنتها مع البيانات المستحصلة من المستشفيات ومركز الرعاية الصحية لغرض تدقيقها والتاكد من سلامة وصحة البيانات .

3.3 تحليل البيانات التي اعتمدت في الدراسة:

التحليل الاحصائى للبيانات يعتمد اساسا على مجموعة من الاختبارات الاحصائية ومنها الاختبارات المعلمية Parametric Tests وعادة تكون افضل من الاختبارات اللامعلمية Non Parametric Tests , واحد اهم الشروط للاختبارات المعلمية هي ان البيانات تمتلك التوزيع الطبيعى Normal Distribution واذا لم يتحقق هذا الشرط نستطيع التحويل عن طريق قائمةTransform في برنامج SPSS او اعتماد اساليب التحويل الاخرى. فضلاً عن اختبارات الخطية Linear test الواجب اتخاذها لمعرفة طبيعة العلاقة بين البيانات .

4.3 التطبيق العملي الاول :البيانات الكمية (التجميعية)

تطبيق انموذج الانحدار المتعدد وانموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي على البيانات الكمية والتي شملت مجاميع الوفيات للمدة (2014–2016) اي 36 شهراً

والمفاضلة بين النماذج عن طريق البرنامج الاحصائي SPSS.V25. والجدول الآتي يبين المتغيرات التوضيحية والمتغير المعتمد التي اعتمدت في التحليل .

جدول (3−1)

المتغيرات والرمز المقابل لكل متغير

الرمز	المتغير	الرمز	المتغير
X6	عدد الولادات المبكرة(الخدج)	Y	عدد الوفيات
X7	سبب الوفاة P07	X1	معدل مدةالحمل بـ(الاسابيع)
X8	سبب الوفاة P22	X2	معدل وزن الطفل بـ(كغم)
X9	سبب الوفاة Q00-Q99	X3	معدل عمر الطفل بـ(الايام)
X10	خديج منزلي	X4	معدل عمر الام بـ(السنوات)
		X5	عدد الولادات الطبيعية

ونختبر الفرضية الآتية :

فرضية العدم :− لا يوجد تأثير للمتغيرات التوضيحية في المتغير التابع $H_0: \beta_i = 0$

الفرضية البديلة :− يوجد تأثير للمتغيرات التوضيحية في المتغير التابع $H_1: \beta_i \neq 0$

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

1.4.3 تطبيق انموذج الانحدار الخطي المتعدد :

The Multiple Regression Model

تم تطبيق انموذج الانحدار المتعدد على 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 انموذج الانحدار بطريقة الحذف الخلفي لافضليته على 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) اكايكي وقيمته (3.46) وقيمة المعيار (H.Q) حنان كوين وقيمته (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 (3-2)

معايير اختيار الأنموذج الملائم للانحدار المتعدد بالطرائق الثلاث

Method	R	R^2	R^2Adj	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

وبالامكان الرجوع الى الملحق (2) للاطلاع على جداول التحليل للطرائق الثلاث فضلاً عن طريقة الحذف الامامي (forward) والتي اعطت نتائج مطابقة لطريقة الانحدار التدريجي stepwise .والجداول الاتية تبين نتائج التحليل للبيانات وكما ياتي :

جدول (3-3)

جدول المؤشرات الاحصائية لانموذج الانحدار المتعدد (Model Summary)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
5	.985[e]	0.971	0.965	2.8168035

من **الجدول (3-3)** (Model Summary) نلاحظ قيمة 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 الخطا,وكذلك قيمة Adjuted R-Square وتساوي(0.965) وهي لتصحيح التحيز في المعامل السابق . وهناك ايضا قيمة الانحراف المعياري لانموذج الانحدار Standrd Deviation للاخطاء وقيمته (2.8).

ولمعرفة مدى موافقة الانموذج للبيانات نختبر الفرضية الاتية :

H_0 : الانموذج المقدر يوافق البيانات بشكل جيد

H_1 : الانموذج المقدر لا يوافق البيانات بشكل جيد

جدول (3-4)

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 (3-4)** (ANOVA) تحليل التباين نجد ان قيمة اختبار (F =160.907) وبمعنوية (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 الخطا (MSE) وقيمتها (7.934) . وقد تم عرض قيم المعاملات(β) للمتغيرات في الجدول الاتي :

جدول (3-5)

جدول المعاملات (Coefficients) والمعنوية ومعامل التضخم للانحدار المتعدد

Model			B	Std. Error	Partial correlation coefficient $r_{Y,Xi}$	T	Sig.	Collinearity Statistics VIF
5		(Constant)	39.373	15.604		2.523	0.017	
	X4	عمر الام	-1.422	0.612	0.375	-2.323	0.027	1.164
	X6	طفل خديج	0.230	0.113	0.915	2.039	0.051	6.219
	X7	P07	0.742	0.118	0.8	6.303	0.000	2.806
	X8	P22	0.964	0.105	0.843	9.159	0.000	3.474
	X9	Q00-Q99	0.799	0.140	0.51	5.722	0.000	1.354
	X10	خديج منزلي	0.444	0.075	0.435	5.890	0.000	1.234

ومن **جدول (3-5)** المعاملات (Coefficients) نلاحظ قيم معاملات الانحدار المتعدد ومعنويتها ومنها يمكن ملاحظة عدم معنوية المعاملات التي قيمة المعنوية اكبر من 0.05 ومنها المتغير (حالة ولادة طفل خديج) اما اذا كانت قيمة sig. اقل من 0.05 تعني ان للمتغير تاثيراً معنوياً على المتغير المعتمد .وتظهر تاثير 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 , والعامل (خديج منزلي) في سلوك المتغير المعتمد (الوفيات). نلاحظ من المعاملات انه في حالة انخفاض معدل 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 (المنزلي) فان زيادته تؤدي الى زيادة في عدد الوفيات بمقدار 44.4% . وكذلك الجدول يبين قيم معامل التضخم Variance Factor (VIF) Inflation 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 للمعلمات المقدرة لكل المتغيرات المستقلة في الانموذج واذا كانت قيمة معامل التضخم تفوق (10) هذا يعني ان المتغير يعاني من تضخم في تباين معامله وانه مرتبط مع بقية المتغيرات المستقلة وهذا يكون سببا كافيا لإهمال المتغير من التحليل أو استعمال طريقة اخرى .

وبذلك تكون معادلة الانحدار التقديرية :

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

t = (2.523) (-2.323) (6.303) (9.159) (5.722) (5.89)

2.4.3 تطبيق انموذج الانحدار اللوغارتمي : Log Regression model:

تم تطبيق انموذج الانحدار اللوغارتمي على 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 $(R^2Adj. = 0.942)$ عن باقي الطرائق وكذلك انخفاض قيمة متوسط مربعات الخطأ (MSE=0.00092) وانخفاض قيمة (AIC=5.38) عن مثيلاتها في الطرائق الاخرى وكما مبن في الجدول الاتي :

جدول (3-6)

معايير اختيار الانموذج الملائم للانحدار اللوغارتمي بالطرائق الثلاث

Method	R	R^2	R^2Adj	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

وبالامكان الرجوع الى الملحق (3) للاطلاع على جداول التحليل للطرائق الثلاث فضلاً عن طريقة الحذف الأمامي (forward) والتي اعطت نتائج مطابقة لطريقة الانحدار التدريجي stepwise . والجداول الآتية تبين نتائج التحليل للبيانات وكما ياتي :

جدول (3-7)

جدول المؤشرات الاحصائية لانموذج الانحدار اللوغارتمي (Model Summary)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.979[a]	0.959	0.942	0.03003649

من **الجدول (3-7)** (Model Summary) نلاحظ قيمة 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 (Adjuted R-Square) وتساوي (0.942) وهي لتصحيح التحيز في المعامل السابق , وهناك ايضا قيمة الانحراف المعياري لانموذج الانحدار (Standrd Deviation) للأخطاء وتساوي (0.03) .

جدول (3-8)

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 (3–8)** تحليل التباين(ANOVA) نجد ان قيمة اختبار (F = 57.751) وبمعنوية (sig= 0) وهي أقل من 0.05 ومن ثم نرفض فرضية العدم القائلة بعدم معنوية انموذج الانحدار , أي عدم رفض الفرضية البديلة وتعني ان انموذج الانحدار معنوي . وقد تم عرض المعاملات للمتغيرات في الجدول الآتي :

جدول (3–9)

جدول المعاملات (Coefficients) والمعنوية ومعاملات التضخم

Model			B	Std. Error	Partial correlation coefficient $r_{Y,Xi}$	T	Sig.	Collinearity Statistics VIF
1		(Constant)	0.501	1.350		0.371	0.714	
	X1	مدةالحمل	0.355	0.876	0.884	0.405	0.689	4.593
	X2	وزن الطفل	0.189	0.258	0.853	0.730	0.472	3.670
	X3	عمر الطفل	0.034	0.077	0.6	0.435	0.667	1.557
	X4	عمر الام	-0.462	0.406	0.5	-1.138	0.266	1.317
	X5	ولادة طبيعية	0.156	0.111	0.92	1.401	0.173	6.447
	X6	طفل خديج	0.265	0.093	0.93	2.847	0.009	3.930
	X7	P07	0.127	0.045	0.9	2.814	0.009	5.722
	X8	P22	0.224	0.070	0.915	3.178	0.004	3.161
	X9	Q00-Q99	0.047	0.029	0.65	1.633	0.115	1.733
	X10	خديج منزلي	0.125	0.041	0.7	3.066	0.005	1.965

ومن **جدول (3–9)** (Coefficients) نلاحظ قيم معاملات الانحدار المتعدد ومعنويتها ومنها يمكن ملاحظة عدم معنوية المعاملات التي قيمتها المعنوية أكبر من 0.05 أما اذا كانت قيمة sig. اقل من 0.05 أي ان للمتغير تأثير معنوي على المتغير المعتمد . وتظهر تاثير المتغيرات (طفل خديج , سبب الوفاة P07 والذي يمثل نقص الوزن وقلة مدة الحمل , وسبب الوفاة P22 والذي يمثل عسر التنفس , والعامل (خديج منزلي) في سلوك المتغير التابع (الوفيات).ومن المعاملات نلاحظ في حال زيادة عدد الولاداة الخدج 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 المنزلي) تؤدي الى زيادة في عدد الوفيات بمقدار 12.5% .

وكذلك من الجدول يمكن ملاحظة قيم معامل التضخم (VIF) Variance Factor Inflation 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 للمعلمات المقدرة لكل المتغيرات المستقلة في الانموذج واذا كانت قيمة معامل التضخم تفوق (10) هذا يعني ان المتغير يعاني من تضخم في تباين معامله وانه مرتبط مع بقية المتغيرات المستقلة وهذا يكون سببا كافياً لإهمال المتغير من التحليل أو استعمال طريقة أخرى . وبذلك تكون معادلة الانحدار التقديرية :

$$\hat{Y} = 0.501 + 0.265(X_6) + 0.127(X_7) + 0.224(X_8) + 0.125(X_{10})$$
$$t = (0.371) \quad (2.847) \quad (2.814) \quad (3.178) \quad (3.066)$$

3.4.3 تطبيق انموذج الانحدار الأسي : Exponential Regression model :

تم تطبيق انموذج الانحدار الأسي على البيانات الكمية ايضا وحصلنا على النتائج المبينة في الجداول الاتية :

جدول (3-10)

المؤشرات الاحصائية للانموذج الأسي

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996[a]	0.994	0.991	4.45219

من **الجدول (3-10)** (Model Summary) نلاحظ قيمة معامل التحديد ($R^2 = 0.994$) والذي يمثل نسبة التغيرات التي تحدث في المتغير التابع والتي تشرحها المتغيرات المستقلة وبعبارة اخرى ان المتغيرات المستقلة تفسر ما قيمته (99.4%) من التغيرات التي تحدث في المتغير المعتمد والمتبقي (0.6) يعزى الى عوامل اخرى منها الخطا , وكذلك قيمة Adjuted R-Square وتساوي (0.991) وهي لتصحيح التحيز في المعامل السابق . وهناك ايضا قيمة الخطا المعياري لنموذج الانحدار الذي تم استخلاصه وهو عبارة عن الانحراف المعياري Standrd Deviation للاخطاء وقيمته (4.45) .

جدول (3-11)

تحليل التباين (ANOVA)

Source	Sum of Squares	df	Mean Squares	F	Sig.
Regression	116845.461	11	10622.315	535.897	0.00
Residual	495.539	25	19.822		
Uncorrected Total	117341.000	36			
Corrected Total	7890.306	35			

ومن **جدول (3-11)**(ANOVA) تحليل التباين نجد ان قيمة اختبار(F =535.897) وبمعنوية (sig =0) وهي اقل من 0.05 ومن ثم نرفض فرضية العدم القائلة بعدم معنوية نموذج الانحدار , اي عدم رفض الفرضية البديلة وتعني ان انموذج الانحدار معنوي. وكذلك نلاحظ قيمة متوسط مربعات الخطا (MSE) وقيمتها (19.822) . وهي عالية اذا ما قورنت بنموذج الانحدار المتعدد MSE=7.934 وكذلك الانموذج اللوغارتمي MSE=0.001 . ونلاحظ قيم المعاملات للانموذج عن طريق الجدول الآتي :

جدول (3-12)

معاملات الانموذج الاسي

Parameter	Estimate	Std. Error	Partial correlation coefficient $r_{Y,Xi}$	T
Constant	17.725	16.568		1.070
مدةالحمل X_1	0.025	0.033	0.897	0.773
وزن الطفل X_2	-0.127	0.146	0.844	-0.868
عمر الطفل X_3	0.017	0.021	0.58	0.799
عمر الام X_4	-0.016	0.019	0.459	-0.809

X_5 ولادة طبيعية	0.000	0.003	0.897	-0.094
X_6 طفل خديج	0.004	0.003	0.93	1.077
X_7 P07	0.013	0.004	0.878	3.476
X_8 P22	0.017	0.004	0.91	4.538
X_9 Q00-Q99	0.013	0.005	0.718	2.558
X_{10} خديج منزلي	0.007	0.002	0.672	2.812

من **جدول (3-12)** نلاحظ دخول المتغيرات جميعها في الانموذج وعن طريق قيم (t) نلاحظ معنوية المتغيرات (P07 , P22 , Q00-Q99, خديج منزلي). وبذلك تكون معادلة الانموذج الأسي كما ياتي :

$$\hat{Y} = 17.725\, EXP(0.013X_7 + 0.017X_8 + 0.013X_9 + 0.007X_{10}$$

t = (1.07) (3.476) (4.538) (2.558) (2.812)

5.3 التطبيق العملي الثاني: البيانات الوصفية : Descriptive Data :

تم تطبيق انموذج الانحدار اللوجستي وتحليل الدالة التمييزية على بيانات المتغير المعتمد الوصفي , والمفاضلة بين النماذج عن طريق البرنامج الاحصائي SPSS.V25. نرمز للمتغيرات التوضيحية التي اعتمدت في التحليل كما في الجدول الاتي :

جدول (3-13)

المتغيرات والرمز المقابل لكل متغير

المتغير	الرمز	
الوفيات	Y	1=وفاة , 0= حياة

مدةالحمل بـ(الاسابيع)	X1	
وزن الطفل بـ(كغم)	X2	
عمر الطفل بـ(الايام)	X3	
عمر الام بـ(السنوات)	X4	
نوع الولادة	X5	1= طبيعية , 2= قيصرية
حالة الولادة	X6	1= طفل غير خديج , 2= طفل خديج
سبب الرقود	X7	P07 =1 , P21 =2 , P22=3 , P23 =4 , Q00-Q99 =7 , P59 =6 , P36 =5
التعقيم	X8	1= معقم , 2=غير معقم (خديج منزلي)

1.5.3 تطبيق انموذج الانحدار اللوجستي : Logistic regression :

تم تطبيق انموذج الانحدار اللوجستي على 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 انموذج الانحدار بطريقة الحذف الخلفي لأفضليته على باقي الطرائق ونلاحظ ذلك عن طريق الجداول الآتية :

جدول (3-14)

ملخص البيانات التي دخلت في التحليل اللوجستي

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	12044	100.0
	Missing Cases	0	.0
	Total	12044	100.0
Unselected Cases		0	.0
Total		12044	100.0

يبين **الجدول (3-14)** ملخص البيانات التي دخلت في التحليل اللوجستي , يظهر من الجدول دخول كل المفردات في التحليل اي عدم وجود اي قيمة مفقودة .

جدول (3-15)

معايير اختيار الانموذج الملائم للانحدار اللوجستي بالطرائق الثلاث

Method	Hosmer and Lemeshow Test	-2Log likelihood	Classification	AIC	AIC_C	H.Q	Nagelkerke R Square
Enter	0.052	3867.764	94.7	3885.764	3885.78	3908.09	0.264
forward	0.281	3905.802	94.8	3915.802	3915.806	3946.8	0.256
backward	0.126	3867.805	95	3879.805	3879.81	3908.13	0.264

وبالامكان الرجوع الى الملحق (4) للاطلاع على نتائج التحليل بالطرائق الثلاث .

عن طريق **جدول (3-15)** تبين المؤشرات الاحصائية للمفاضلة بين النماذج أن الانموذج بطريقة backward افضل من بقية النماذج اذ نلاحظ ارتفاع قيمة التصنيف البالغة 95% وكذلك انخفاض قيمة المؤشرين (AIC_c , AIC) البالغة (3879.805 , 3879.81) على الترتيب عن باقي الطرائق. و نتائج التحليل للانموذج مبينة في الجداول الآتية .

جدول (3-16)

اختبار الدلالة الاحصائية للانموذج

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 3[a]	Model	1133.853	12	0.000

من **جدول (3-16)** نلاحظ اختبار كاي تربيع (χ^2) والذي يعتمد على المقارنة بين القيم المشاهدة والقيم المتوقعة للانموذج الذي يحتوي الحد الثابت فقط والانموذج الذي يحتوي كافة المتغيرات .نجد قيمة مربع كاي وتساوي (1133.853) عند درجة حرية (12) ومستوى معنوية (P-Value=0.00 < 0.05) وهذا يعني ان الانموذج ذو دلالة احصائية ويوفق البيانات بشكل جيد اي ان المتغيرات المتضمنة في الانموذج لها تاثير ودلالة ومساهمة في التصنيف .

جدول (3-17)

اختبار هوزمر- ليمشو

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	15.373	8	0.052
2	12.700	8	0.123
3	12.603	8	0.126

من **جدول (3-17)** اختبار هوزمر- ليمشو وهو احد الاختبارات المهمة لقياس جودة توفيق انموذج الانحدار اللوجستي ويعتمد على مدى قرب الاحتمالات المشاهدة من الاحتمالات المتوقعة . ويختبر الفرضية الاتية :

H_0 : الانموذج المقدر يوافق البيانات بشكل جيد

H_1 : الانموذج المقدر لا يوافق البيانات بشكل جيد

وتتم مقارنة قيم $(P-Value)$ لكاي تربيع (χ^2) مع مستوى المعنوية ونلاحظ ان قيمة $(P-Value = 0.126 > 0.05)$ وهي معنوية لجودة التوفيق للانموذج .وهذا يعني لا يوجد دليل كافٍ لرفض فرضية العدم أي ان الانموذج ملائم ويوافق البيانات بشكل جيد .

جدول (3-18)

تصنيف البيانات (Classification Table[a])

Observed			Predicted		
			وفاة او حياة		Percentage Correct
			حياة	وفاة	
Step 1	وفاة او حياة	حياة	11371	30	99.7
		وفاة	576	64	10.0
	Overall Percentage				95.0
Step 2	وفاة او حياة	حياة	11371	30	99.7
		وفاة	576	64	10.0
	Overall Percentage				95.0
Step 3	وفاة او حياة	حياة	11371	30	99.7
		وفاة	574	66	10.3
	Overall Percentage				95.0

يمثل **جدول (3-18)** جدول تصنيف البيانات للقيم المتوقعة من الانموذج النهائي الذي تم تقديره و التحقق من مدى قدرة الانموذج المقدر في تصنيف عدد الحالات التي تنتمي الى المجوعة المعنية بالتصنيف فعلاً من تلك التي لا تنتمي الى المجموعة المعنية بالتصنيف وبشكل إجمالي ان الانموذج قد نجح في تصنيف 95% من الحالات محل الدراسة .

جدول (3-19)

المتغيرات الداخلة في المعادلة (Variables in the Equation)

			B	S.E.	Wald	Df	Sig.	Exp(B)
Step 3[a]	X1	مدةالحمل	-0.170	0.018	86.004	1	0.000	0.844
	X2	وزن الطفل	-0.491	0.086	32.985	1	0.000	0.612
	X6	حالة الولادة			7.666	2	0.022	
	X6.(1)	طفل غير خديج(1)	-1.012	1.294	0.612	1	0.434	0.363
	X6,(2)	طفل خديج(2)	-1.259	1.295	0.946	1	0.331	0.284
	X7	سبب الرقود			363.690	7	0.000	

	X7,(1)	سبب الرقود(1)	0.470	0.208	5.117	1	0.024	1.600
	X7,(2)	سبب الرقود(2)	0.434	0.200	4.712	1	0.030	1.543
	X7,(3)	سبب الرقود(3)	0.046	0.168	0.074	1	0.786	1.047
	X7,(4)	سبب الرقود(4)	-0.166	0.334	0.246	1	0.620	0.847
	X7,(5)	سبب الرقود(5)	0.455	0.209	4.727	1	0.030	1.577
	X7,(6)	سبب الرقود(6)	-2.488	0.300	68.665	1	0.000	0.083
	X7,(7)	سبب الرقود(7)	2.269	0.200	128.194	1	0.000	9.671
	X5,(2)	نوع الولادة(2)	-0.238	0.118	4.082	1	0.043	0.788
		Constant	5.783	1.413	16.758	1	0.000	324.728

اظهرت نتائج التحليل دخول المتغيرات (مدة الحمل ، وزن الطفل ، سبب الرقود , نوع الولادة) واستبعاد العوامل الأخرى لأنها غير معنوية . و**الجدول (3–19)** يمثل المعالم المقدرة للمتغيرات الداخلة فضلاً عن قيمة الثابت والخطأ المعياري . ويبين إحصاءه (wald) والقيمة الاحتمالية المقابلة (sig.) ويستعمل لاختبار الدلالة لكل متغير فان كانت ذات دلالة احصائية فتعني ان قيمة $(\hat{b})$ لذلك المتغير لا تساوي صفراً . وكذلك يبين الجدول معنوية المعالم (sig.) ومنه يمكن ملاحظة المتغيرات المعنوية وغير المعنوية فان كانت $(sig. < 0.05)$ يدل على ان المتغير ذو تاثيرمعنوي. ويظهر ايضا في الجدول قيم $(\exp(B))$ التي تمثل نسبة الارجحية (Adds Ratio) وهي تشير الى مقدار التغير الحاصل في نسبة أرجحية وقوع الحدث (الوفاة) عند حدوث تغير في قيمة المتغير المستقل المرتبط بالمعلمة (B) , وتزداد نسبة الارجحية لوقوع الوفاة كلما زادت عن الواحد .وتحسب كالاتي :

$$Exp(X_1) = e^{0.844} = 2.32$$

وهذا يعني أرجحية ان يكون الخديج عرضة للوفاة أعلى بمقدار (2.4547) ما هو عليه عند مدة الحمل الطبيعية

$$Exp(X_2) = e^{0.612} = 1.844$$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (1.84) ما هو عليه عند الوزن الطبيعي .

$Exp(X_{7,1}) = e^{1.6} = 4.95$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (4.95) ما هو عليه عند غير المصاب (P07) .

$Exp(X_{7,2}) = e^{1.543} = 4.67$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (4.67) ما هو عليه عند غير المصاب (P22)

$Exp(X_{7,5}) = e^{1.577} = 4.84$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (4.84) ما هو عليه عند غير المصاب (P36)

$Exp(X_{7,6}) = e^{0.083} = 1.086$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (1.08) ما هو عليه عند غير المصاب (P59)

$Exp(X_{5,1}) = e^{0.788} = 2.2$

وهذا يعني ارجحية ان يكون الخديج عرضة للوفاة اعلى بمقدار (2.2) ما هو عليه عند الولادة القيصرية .

كما أشارت قيمة نسبة الترجيح Adds Ratio ((Exp(B)) الى ان الزيادة في المتغير (مدة الحمل) تؤدي الى انخفاض في الوفيات بنسبة 16% والزيادة في المتغير (وزن الطفل) تؤدي الى انخفاض في الوفيات بنسبة 39%. في حين ان الارتفاع للمتغير (سبب الرقود P07:(1)) يؤدي الى زيادة في الوفيات بنسبة 60% , والمتغير (سبب الرقود(2): P21) تؤدي الزيادة فيه الى ارتفاع الوفيات بنسبة 54%. والمتغير (سبب الرقود (5):P36) تؤدي الزيادة فيه الى زيادة الوفيات بنسبة 57%. وزيادة المتغير (سبب الرقود (6) :P59) الى انخفاض الوفيات بنسبة 92% . والزيادة في المتغير (سبب الرقود (7): Q00-Q99) يؤدي الى زيادة الوفيات بنسبة 67%. وزيادة المتغير (نوع الولادة (2) :قيصرية) الى انخفاض الوفيات بنسبة 79%.

ومن كل ما سبق يمكن ان تكون معادلة الانحدار اللوجستي بالصيغة الاتية :

$$Log(\left(\frac{\rho}{1-\rho}\right) = 5.783 - 0.17X_1 - 0.491X_2 + 0.47X_{7,1} + 0.434X_{7,2} + 0.455X_{7,5} - 2.488X_{7,6} + 2.269X_{7,7} - 0.238X_{5,1}$$

2.5.3 تطبيق تحليل الدالة التمييزية : Discriminant Function Analysis :

تم تطبيق الدالة التمييزية على ذات البيانات في الانحدار اللوجستي بواسطة برنامج spss.v25 وكانت النتائج كما ياتي :

جدول (3-20)

ملخص البيانات التي دخلت في التحليل

Analysis Case Processing Summary			
Unweighted Cases		N	Percent
Valid		12044	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		12044	100.0

يبين **الجدول (3-20)** ملخص البيانات التي دخلت في التحليل التمييزي , يظهر من الجدول دخول كل المفردات في التحليل أي عدم وجود أي قيمة مفقودة .

جدول (3-21)

بيانات المتغيرات المستقلة واحصاءات المجاميع (Group Statistics)

الوفيات		Mean	Std. Deviation
حياة	مدةالحمل	36.7363	2.02159
	وزن الطفل	2.9116	.56329
	عمر الطفل	6.2794	5.53203
	عمر الام	25.9360	5.60113
	نوع الولادة	1.4337	.49613
	حالة الولادة	1.1992	.39938
	اسباب الرقود	4.5968	1.92173
	معقم ومنزلي	1.5752	.49433
وفاة	مدةالحمل	33.6864	4.41658

	وزن الطفل	2.1756	.87042
	عمر الطفل	4.3994	4.74107
	عمر الام	25.5523	5.59275
	نوع الولادة	1.3401	.47411
	حالة الولادة	1.5850	.49310
	اسباب الرقود	3.6225	2.12744
	معقم ومنزلي	1.3619	.48094
Total	مدةالحمل	36.5740	2.31833
	وزن الطفل	2.8724	.60661
	عمر الطفل	6.1793	5.50882
	عمر الام	25.9156	5.60112
	نوع الولادة	1.4287	.49541
	حالة الولادة	1.2197	.41406
	اسباب الرقود	4.5449	1.94546
	معقم ومنزلي	1.5638	.49593

يبين **الجدول (3–21)** اعداد المفردات في الفئات المختلفة للمتغير التابع (وفاة – حياة).

من الجدول نلاحظ الفرق بين المتوسطات فان كان هناك فرق دل على معنوية المتغير فمثلاً متوسط مدة الحمل والبالغة (36.73) اسبوع في المجموعة الأولى بينما بلغت (33.68) اسبوع في المجموعة الثانية وهذا الفرق يدل على ان المتغير مدة الحمل لها تأثير معنوي بينما المتغير (عمر الام) كان بمتوسط (25.93) في المجموعة الاولى وبمتوسط (25.55) في المجموعة الثانية مما يدل على عدم معنويته وكذلك الحال بالنسبة لبقية المتغيرات..

جدول (3–22)

تحليل التباين والمعنوية للمتغيرات

	Wilks' Lambda	F	df1	df2	Sig.
مدةالحمل	**.913**	**1150.568**	**1**	**12042**	**.000**
وزن الطفل	**.926**	**964.967**	**1**	**12042**	**.000**
عمر الطفل	**.994**	**71.094**	**1**	**12042**	**.000**
عمر الام	**1.000**	**2.849**	**1**	**12042**	**.091**
نوع الولادة	**.998**	**21.684**	**1**	**12042**	**.000**
حالة الولادة	**.956**	**551.134**	**1**	**12042**	**.000**

اسباب الرقود	.987	154.149	1	12042	.000
التعقيم	.991	113.277	1	12042	.000

يبين **الجدول (3–22)** نتائج تحليل التباين للمتغيرات المستقلة وذلك لتحديد ما اذا كان هنالك فروق ذات دلالة احصائية بين متوسطات المتغيرات المستقلة لكل فئة من فئات المتغير التابع, ويتضح من الجدول ان جميع الفروق بين المتوسطات لها دلالة احصائية اي الفروق معنوية لان قيمة الاحتمال (sig) اقل من مستوى المعنوية (0.05) لكل المتغيرات المستقلة ما عدا المتغير (عمر الام) لم يكن معنوياً اذ كانت قيمته المعنوية 0.091 > 0.05 .

جدول (3–23)

الاحصاءات المتعلقة بدالة التمييز

(Summary of Canonical Discriminant Functions)

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.126	100.0	100.0	.335

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.888	1430.198	8	.000

يبين **الجدول (3–23)** الإحصاءات المتعلقة بدالة التمييز, يلاحظ من الجدول ان الدالة التمييزية تقابلها قيمة ذاتية (0.126) بنسبة ارتباط قانوني (0.335) لكل المتغيرات المميزة وقد فسرت الدالة التمييزية 100% من التباين, ولغرض معرفة جودة التمييز للدالة نلاحظ نتائج كل من إحصاءة Wilks Lambda واختبار Chi-square, اذ نلاحظ ان قيمة Wilks Lambda بلغت (0.888) وهي قيمة قريبة من الواحد وبقيمة ل Chi-square بلغت (1430.198) وبمستوى دلالة (0.000) وهذا يشير إلى القدرة الجيدة للدالة على التمييز بين المجموعتين, وبذلك فان الاختلاف في التصنيف بين المجموعتين هو اختلاف جوهري ولا يعود إلى الصدفة. هذا يعني ان الدالة معنوية ولها القدرة على التمييز .

جدول (3–24)

يبين مصفوفة الارتباطات

	X1	X2	X3	X4	X5	X6	X7	X8
X1	1.000	.496	-.043	.053	.086	-.564	.156	.066
X2	.496	1.000	.025	.068	.052	-.474	.179	.133
X3	-.043	.025	1.000	-.044	-.120	-.033	.283	.573
X4	.053	.068	-.044	1.000	.132	-.022	-.026	-.081
X5	.086	.052	-.120	.132	1.000	-.020	-.094	-.186
X6	-.564	-.474	-.033	-.022	-.020	1.000	-.204	-.166
X7	.156	.179	.283	-.026	-.094	-.204	1.000	.478
X8	.066	.133	.573	-.081	-.186	-.166	.478	1.000

يشير **الجدول (3–24)** الى معاملات الارتباط الثنائي بين المتغيرات المستقلة السبعة فمثلا معامل الارتباط بين X3 عمر الطفل وX4 عمر الام بلغ (0.044-) وبين X5 نوع الولادة و X4 عمر الام كان (0.132) وهكذا...

ولغرض صياغة انموذج الدالة التمييزية المعيارية لابد من تقدير المعاملات المميزة المعيارية, وعن طريق المعاملات المميزة المعيارية يمكن معرفة مدى تأثير المتغيرات على الانموذج, إذ كلما كانت قيمة المعامل المطلقة كبيرة دل ذلك على المساهمة العالية للمتغير سواء أكانت سلبا أم إيجابا ويكون ذلك بسحب إشارة العامل. **والجدول (3–27)** يبين قيم المعاملات المعيارية للمتغيرات الداخلة في التحليل.

جدول (3–25)

معاملات الدالة التمييزية المعيارية

Standardized canonical Discriminant Function Coefficients

function 1	مدةالحمل	وزن الطفل	عمر الطفل	عمر الام	نوع الولادة	حالة الولادة	اسباب الرقود	التعقيم
	0.632	**0.458**	**0.201**	**-0.018**	**0.081**	**-0.001**	**0.071**	**0.035**

وبذلك يمكن صياغة دالة التمييز المعيارية بالشكل الآتي:

$$\hat{L} = 0.632X_1 + 0.458X_2 + 0.201X_3 - 0.018X_4 + 0.081X_5 - 0.001X_6 + 0.071X_7 + 0.035X_8$$

بعد إيجاد الدالة التمييزية لابد من التحقق من تصنيف المجاميع فيما إذا كانت فعلا تقع ضمن المجموعة الذي صنفت إليه, أي هل ان التصنيف كان صحيحا أم خاطئاً؛ ولغرض إجراء ذلك لابد من إيجاد النقطة الفاصلة (نقطة الوسط) وتحديدها بين المجموعات والتي تم توضيحها

في الجانب النظري, ولانه كان لدينا مجموعتان لذلك سوف يكون لكل مجموعة نقطة فاصلة وكما هو موضح في الجدول أدناه.

الجدول(3-26)

النقاط الفاصلة

Functions at Group Centroids	
Y	Function
	1
0	.084
1	-1.498-

ونلاحظ من **الجدول (3-26)** انه إذا كانت قيمة الدالة التمييزية سالبة سوف تصنف المشاهدة إلى المجموعة الأولى, أما إذا كانت القيمة موجبة فتصنف المشاهدة إلى المجموعة الثانية.

جدول (3-27)

نتائج دقة تصنيف

		الوفيات	Predicted Group Membership		المجموع
			المجموعة (الاولى حياة)	المجموعة الثانية (وفاة)	
Original	Count	المجموعة الاولى (حياة)	9756	1647	11403
		المجموعة الثانية (وفاة)	285	356	641
	%	المجموعة الاولى (حياة)	85.6	14.4	100.0
		المجموعة الثانية(وفاة)	44.5	55.5	100.0
Cross-validated[b]	Count	المجموعة الاولى(حياة)	9754	1649	11403
		المجموعة الثانية(وفاة)	286	355	641
	%	المجموعة الاولى (حياة)	85.5	14.5	100.0
		المجموعة الثانية (وفاة)	44.6	55.4	100.0

الجدول (3–27) يبين نسبة التنبؤ بالتصنيف الصحيح لبيانات العينة, إذ هناك (9756) حالة من المجموعة الاولى (الحياة) وبنسبة 85.6% قد تم تصنيفها بشكل صحيح وان (1647) حالة وبنسبة 14.4% تم تصنيفها بشكل خاطئ إذ صنفت بانها ضمن المجموعة الاولى (الحياة) وهي تعود الى المجموعة الثانية (الوفاة), وفي المجموعة الثانية يتبين ان (356) وبنسبة 55.5% تم تصنيفها بشكل صحيح وبناءً عليه هناك (285) حالة وبنسبة 44.5% تم تصنيفها بشكل خاطئ إذ صنفت ضمن المجموعة الثانية (الوفاة) وهي تعود الى المجموعة الاولى (الحياة), وكنتيجة عامة فقد دلت النتائج بان الحالات المصنفة تصنيفاً صحيحاً كانت (9756+356=10112) حالة اي ما نسبته 83.95% من حالات العينة البالغة (12044).

الفصل الرابع

الاستنتاجات والتوصيات

الفصل الرابع
الاستنتاجات والتوصيات

1.4 الاستنتاجات :

تم في الفصول السابقة تحليل بيانات وفيات الاطفال الخدج وأهم العوامل المؤثرة فيها وذلك باستعمال نماذج الانحدار الخطية عن طريق انموذج الانحدار المتعدد , والنماذج اللاخطية عن طريق انموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي واستعمال النماذج ثنائية الاستجابة عن طريق انموذج الانحدار اللوجستي وتحليل الدالة التمييزية وهي وسيلة لتحليل مثل هذا النوع من البيانات للحصول على افضل انموذج يفسر تاثير المتغيرات المستقلة في سلوك المتغير التابع (الوفيات) ، وقد توصل البحث الى استنتاجات متعددة أهمها ما يأتي :

أولاً: من التطبيق العملي الاول والذي تضمن تطبيق انموذج الانحدار المتعدد وانموذج الانحدار اللوغارتمي وانموذج الانحدار الأسي تبين ما يأتي :

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 , مدة الحمل , عسر التنفس , التشوهات الخلقية , الخديج المنزلي) .
2) اما في حالة الانحدار اللوغارتمي نلاحظ من قيم (.sig) وقيم (T) معنوية العوامل المؤثرة في حياة الطفل الخديج وهي (طفل خديج (مولود مبكراً) , سبب الوفاة P07, وسبب الوفاة P22, والخديج المنزلي)
3) ومن تحليل البيانات الداخلة في انموذج الانحدار الأسي تبين ان حياة الطفل الخديج مرتبطة بالعوامل (Q00-Q99 , P22 , P07 , والخديج المنزلي) .ويتبين ذلك عن طريق معنوية قيم (t) المحسوبة .
4) اثبتت النتائج ان النماذج الثلاثة المستعملة كانت معنوية اعتماداً على معنوية قيمة F الحسابية في جدول تحليل التباين , وان نماذج الانحدار(المتعدد و اللوغارتمي والأسي) لم تعطِ دلالة لبعض المتغيرات التي يعتقد ان لها تاثيراً 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 والبالغة (3 سنوات) .

5) تشير قيمة معامل التحديد المصحح لانموذج الانحدار الأسي $\bar{R}^2 = \%\ 99.1$ على القدرة التفسيرية العالية للبيانات في الانموذج وقد يعزى ارتفاع قيمة $\bar{R}^2$ الى دخول معظم المتغيرات التفسيرية في الانموذج . في حين بلغت في انموذج الانحدار المتعدد ($\bar{R}^2 = \%\ 96.5$) والذي بدوره أكبر من مقدار تفسير البيانات للانحدار اللوغارتمي والتي بلغت قيمته ($\bar{R}^2 = \%\ 94.2$) . ومن جهة ثانية أعطى الانموذج الأسي أعلى قيمة لمتوسط مربعات الخطأ (MSE=19.822), في حين بلغت في الانموذج المتعدد (MSE=7.934) , وكانت أقل قيمة لمتوسط مربعات الخطأ في الانموذج اللوغارتمي (MSE=0.001) .

ثانياً : من التطبيق العملي الثاني والذي تضمن انموذج الانحدار اللوجستي وتحليل الدالة التمييزية تبين ما يأتي:

1) اثبتت النتائج ان الانموذج اللوجستي وتحليل الدالة التمييزية معنويان اي يمكن استعمالهما في دراسة وتحليل مثل هذه الظواهر الثنائية الاستجابة.وتتضح معنوية الانموذج اللوجستي عن طريق اختبار (هوزمر– ليمشو) . وتحليل الدالة التميزية عن طريق قيمة Wilks وقيمة Chi-square.

2) اعطى الانموذج اللوجستي دلالة معنوية للمتغيرات (مدة الحمل, وزن الطفل ,حالة الولادة, سبب الرقود , نوع الولادة) كمتغيرات ذات تاثير في سلوك المتغير التابع ونلاحظ ذلك عن طريق (sig.) في جدول **(3–21)** . كما أعطى التحليل التمييزي دلالة معنوية للعوامل نفسها فضلاً عن (عمر الطفل وعمر الأم والخديج المعقم والمنزلي) ونلاحظ ذلك عن طريق (sig.) في جدول (3–24) .

3) النسبة الكلية للتنبؤ الصحيح باستعمال الانحدار اللوجستي بلغت 95% , بينما بلغت نسبة التنبؤ الصحيح باستعمال الدالة التمييزية 83.95% , مما يدل على افضلية الانموذج اللوجستي لمثل هذه البيانات .

2.4 التوصيات :

1) نوصي الاهتمام بالعوامل المؤثرة في حياة الطفل المولود حديثا مثل (قلة فترة الحمل, و عسر التنفس, و التشوهات الخلقية , والخديج المنزلي) واعطائها اولويات في التداخلات الصحية والطبية وحسب درجة معنوية تلك العوامل .

2) يفضل الاعتماد على انموذج الانحدار اللوغارتمي لاعطائه اقل قيمة لمتوسط مربعات الخطأ MSE للبيانات التجميعية وكذلك لارتفاع قيم معامل الارتباط الجزئي . كما يفضل زيادة حجم العينة في حال توفر البيانات .

3) اعتماداً على نتائج التحليل الاحصائي للتطبيق العملي الثاني نوصي الباحثين استعمال الانحدار اللوجستي لدراسة الظواهر ثنائية الاستجابة لكفاءته العالية في التصنيف .

4) ضرورة اعداد الجداول الاحصائية الخاصة بالقطاع الطبي وفق الجداول المنشورة عالميا ضمن التقارير الاحصائية العالمية مثل منظمة الصحة العالمية واضطلاع الجهاز المركزي للاحصاء بهذه المهمة عن طريق دوائر الصحة في المحافظات. ونؤكد على تثبيت المعلومات بدقة في شهادة الوفاة وضرورة اعتماد التصنيف الدولي للامراض, وتطويرعملية تسجيل الوفيات الكترونياً لمنع بعض الاخطاء وتحسين نوعية البيانات .

5) ومن تدقيق البيانات تبين ان قسم من المتوفين الخدج لم يسجلوا بحجة عدم اصدار لهم شهادة وفاة وهذا يؤدي الى ضعف البيانات لذا نوصي دوائر الصحة بتسجيل جميع حالات الوفاة للحصول على بيانات أكثر دقة وواقعية .

6) نقترح زيادة عدد الوحدات الخاصة بمعالجة عسر التنفس للسيطرة على عدد حالات المصابة للمواليد الجدد . وزيادة الرعاية الصحية بالأم الحامل واعطائها نوعاً من التحصينات واللقاحات اللازمة اضافة الى التوعية الثقافية لضمان عدم الاصابة بالتشوهات الخلقية .

7) على الجهات المختصة في وزارة الصحة السعي لزيادة عدد وحدات الخدج في المستشفيات العامة والزام المستشفيات الخاصة لفتح وحدات خدج لاستيعاب عدد الاطفال المحتاجين للعناية اللازمة بعد الولادة مباشرة .

8) ضرورة ادخال متغيرات اخرى في برنامج الخدج الاحصائي لدراستها مثل الحالة الصحية للام (وجود امراض مزمنة) ونوع العقاقير المستخدمة من قبل الام اثناء مدة الحمل وقبلها كذلك نوع الرضاعة وعدد الاسقاطات والولادات المعاقة والتدخين وغيرها مما يعتقد ان لها تاثير في وفيات الخدج .

9) ضرورة الاهتمام بالجانب الاحصائي في دوائر الصحة وتطوير الكوادر من خلال عمل دورات تدريبية لتحسين قدرات واداء العاملين ورفدهم بالبرامج والاجهزة الحديثة باعتباره مصدر البيانات .

المصادر

المصادر

The Refrences

*** القرآن الكريم**

أولاً : المصادر العربية : Arab References :

الكتب :

[1] بري, عدنان ماجد عبد الرحمن , **(تحليل الانحدار الخطي)** ,جامعة الملك سعود,2003 .

[2] التميمي, زهرة حسن عباس وآخرون , **(تحليل الانحدار)** , مديرية دار الكتب للطباعة والنشر / جامعة البصرة ,2014 .

[3] راو, بوتلري و ميلر, روجر ليروي . ترجمة كاظم ,اموري هادي و د. هادي , سعيد علي , **(القياس الاقتصادي التطبيقي)** ,جامعة واشنطن , مطبعة دار الحكمة للطباعة والنشر– الموصل ,1990 .

[4] الراوي , خاشع محمود ,(مقدمة في تحليل الانحدار), مديرية دار الكتب للطباعة والنشر– جامعة الموصل .

[5] السيفو , وليد اسماعيل , **(مشاكل الاقتصاد القياسي التحليلي)** , طبعة رقم 1 , الاهلية للنشر والتوزيع , 2006 .

[6] عبد, حميد عبيد , 2017, **(الاقتصاد القياسي)** , دار الكتب ,كربلاء المقدسة .

[7] كاظم, اموري هادي و مسلم,باسم شليبه, **(القياس الاقتصادي المتقدم النظرية والتطبيق)**, بغداد مكتبة دنيا الامل ,2002 .

[8] ـ ,منظمة الصحة العالمية , **(كتاب الاحصاءات الصحية العالمية)**, طبع في فرنسا ,2009

الرسائل والاطاريح :

[9] احمد,ابراهيم محمد ابراهيم سيد, **(مقارنة بين نموذج الانحدار الخطي المتعدد ونموذج الدالة التميزية : دراسة تطبيقية على عينة من الاصحاء ومرضى الفشل الكلوي بمستشفى ابن سينا)** رسالة ماجستير–جامعة السودان للعلوم والتكنولوجيا ,2010 .

[10] احمد ,الطيب عمر , واخرون, **(اثر مشاكل الانحدار على التقدير بالتطبيق على الارباح في المصارف السودانية)** , جامعة جيهان –سليمانية,2014 .

[11] البير, خولة علي, **(تقييم الواقع الصحي للطفولة في العراق بالعمر 4-0 سنة والخدمات المتاحة لهم ومقترحات تطويرها)** , وزارة التخطيط والتعاون الانمائي العراقية- دائرة التنمية البشرية ,2009 .

[12] التلباني ,شادي اسماعيل, **(استخدام نموذجي الانحدار اللوجستي وانحدار كوكس لدراسة العوامل المؤثرة على وفيات الاطفال الرضع في فلسطين),** جامعة الازهر- غزة , 2004 .

[13] حسين, حسن محمد , **(وفيات الاطفال تحت تاثير العوامل الطبيعية والبشرية في محافظة ديالى – دراسة ميدانية)**,مجلة الفتح- العدد 26 .2006 .

[14] زيد, افنان محمد محمود , **(اثر العوامل الديموغرافية والاقتصادية والاجتماعية في وفيات الاطفال الرضع في الاراضي الفلسطينية من واقع مسح الاسر الفيسطينية 2010)** , رسالة ماجستير – جامعة النجاح الوطنية –نابلس ,2014 .

[15] السباح, شروق عبد الرضا, **(بناء أنموذج انحدار لوجستي معدل لحياة الأطفال الخدج في محافظة كربلاء)**,رسالة دكتوراه- جامعة بغداد ,2009 .

[16] سليمان , علي ابشر,واخرون , **(مقارنة بين النموذج اللوجستي الثنائي والدالة التمييزية في التصنيف)**, جامعة السلام كلية الاقتصاد وتنمية المجتمع ,السودان . مجلة العلوم الاقتصادية(2) 17 .2016,Vol .

[17] الشمراني ,محمد موسى,**(مشكلات استخدام تحليل التباين الأحادي والمقارنات البعدية وطرق علاجها)** ماجستير احصاء وبحوث , جامعة ام القرى , مكة المكرمة ,2000 .

[18] العبيدي, ندوة خزعل رشاد , **(دراسة مقارنة لبعض طرائق اختبار مجموعة جزئية في نماذج الانحدار المتعددة مع تطبيق في سرطان الثدي)** رسالة ماجستير- جامعة الموصل , 2005 .

[19] العزاوي , دجلة ابراهيم , عبد القادر, زينة ياور,**(مقارنة الاساليب المستخدمة في تحديد عدد المركبات الرئيسية)** مجلة العلوم الاقتصادية والادارية المجلد 13/ ع45, 2007 .

[20] فتاح, احمد فاضل , **(التنبؤ بمعدلات الوفيات وبناء جداول الحياة لدولة فرنسا)** رسالة ماجستير-الجامعة المستنصرية ,2006 .

[21] ـ ,المكتب المركزي للاحصاء , **(دراسة أسباب وفيات الأطفال دون الخمس سنوات)** , مديرية الرعاية الصحية الأولية في وزارة الصحة السورية بالتعاون مع منظمة اليونسيف , 2008 .

[22] نجم الدين، عدنان كريم, **(محاضرات في القياس الاقتصادي لطلبة الدراسات العليا)** , جامعة كربلاء –كلية الادارة والاقتصاد , 2017 .

[23] النعيمي , اسوان محمد طيب , **(معالجة البيانات غير التامة وتقديرها بطريقة انحدار المركبات الرئيسية)**, قسم نظم المعلومات الادارية / كلية الادارة والاقتصاد / جامعة الموصل 2009, Dec./7-6 .

[24] الهاشمي , سما سعدي علي , **(دراسة إحصائية لوفيات الأطفال الرضع لمحافظة نينوى للفترة 2004 – 1987)** .رسالة ماجستير – جامعة بغداد ,2005 .

[25] ياغوبيان, انكين انترانيك هايك , **(استكشاف وتقدير القيم الشاذة في بعض النماذج اللاخطية)** ,رسالة دكتوراه , الجامعة المستنصرية ,2005 .

[26] يحيى , مزاحم محمد ,عبد الله ,محمود حمدون , **(تشخيص التعدد الخطي واستخدام انحدار الحرف في اختيار متغيرات دالة الاستثمار الزراعي في العراق للفترة 1980–2000)** .جامعة الموصل ,مجلة تكريت للعلوم الإدارية والاقتصادية . المجلد 3 العدد 8 ,2007 .

مواقع الانترنت :

[27] http://aawsat.com/home/article/532821

خوجة , عبد الحفيظ , **(وفيات الاطفال الأطفال حديثو الولادة وعوامل الخطورة التي تهدد حياتهم)**, جدة, العدد 13548, 2016 .

[28] . http://www.almadapress.com/ar/news/11279

منظمة أطفال الحروب **(وور تشايلد)** ,H.A.A،التقرير الفصلي لشهر أيار , المحرر, AHF المدى برس / بغداد ,2013 .

ثانياً : المصادر الاجنبية : Foreign References :

الكتب :

[29] Alan Miller**,(Subset Selection in Regression, A CRC Press Company)** Boca Raton London New York Washington, D.C. Second Edition ,2002 .

[30] Alvin. C. Rencher **,(Methods of multivariate analysis),** Brigham Young University, Copyright by John Wiley & Sons, Inc,Second Edition,2002

[31] B.S Everitt**,(Medical statistics from A to Z).** Institute of Psychiatry, King's College, University of London, cambridge university press , Second Edition.2006

[32] Bolch ,B.W,& Hung,C.J **,(Multivariate statistical methods for Business and Economics prentice)** , Hall, Inc. New Jersey .1974

[33] C. A. Graver and H. E. Boren. Jr **,(multivariate logarithmicand exponential regression models)** , united states air force project rand,1967

[34] Douglas C. Montgomery & George C. Runger**,(Applied statistics and probability for engineers)** , Printed in the United States of America. Third Edition.2003 .

[35] Fan, Xitao, **(Comparing linear discriminant function with logistic regressionfor the two-group)**, Department of Psychology Utah State University Logan,1998 .

[36] Jennifer Peat , Belinda Barton **,(Medical Statistics A Guide to Data Analysis and Critical Appraisal)** , Sydney, Australia, by Blackwell Publishing Ltd.2005.

[37] John.O.Rawlings & others**,(Applied regression analysis A research Tool)** North Carolina State University , USA , Second Edition,1998 .

[38] Michael H. Kutner,& others,**(Applied linear statistical models),** Published by McGraw-Hill!Irwin-New York. Fifth Edition.2005 .

[39] N.R.DRAPER &H.Smith **,(Applied Regression Analysis)** , New York . printed in the United States . Second Edition, 1981

[40] Paul Newbold & others**,(Statistics for Business and Economics)**, Publisher :Pearson,7th Release ,2010.

[41] Robert.S.Pindyck & Daniel.I.Rubinfeld**,(Econometric models and economic forecasts)**, New York ,McGraw-Hill Book company. Second Edition.2000.

[42] S.R.SEARLE**,(Linear models)** , N.Y State College of Agriculture. New York-London. printed in the United States ,1971

[43] Samprit Chatterjee & Alis .Hadi **,(regression analysis by Example),** by John Wiley & Sons, Inc . Fourth Edition, 2006.

[44] Sanford Weisberg**,(Applied linear regression).** University of Minnesota ,Copyright by John Wiley & Sons, Inc. Third Edition,2005 .

[45] Sheldon M. Ross **,(Probability and statistics for engineers and scientists),** University of California, Berkeley, Printed in Canada . Fourth Edition,2009.
[46] T.T. Soong **,(Fundamentals of probability and statistics for engineers)** , State University of New York at Buffalo, Copyright John Wiley & Sons Ltd, England,2004

الرسائل والاطاريح :

[47] Aozgaur Yeniay , Atilla Gaokta ,(**A Comparison of partial least squares regression with other prediction methods)** , Hacettepe Journal of Mathematics and Statistics Volume 31 ,99,111, 2002
[48] ENDING NEWBORN DEATHS **,(Save the Children)** ,2012.
[49] Kenneth Benoit**,(Linear Regression Models with Logarithmic Transformations)** Methodology Institute ; London School of Economics, 2011
[50] L Khikmah, H Wijayanto, and U D Syafitri , **(Modeling Governance KB with CATPCA to overcome multicollinearity in the logistic regression),** Department of statistics,Faculty of mathematics and vatural science. Bogor university ,Indonesia ,2017
[51] Mahmood, Hanady J. & Saad J. Sulaiman**, (Assessment Of Factors Causing Mortality Rate Of Neonate In Al-Batool Teaching Hospital In Mosul City(2010)).** Kufa Journal for Nursing Sciences.2013
http://www.uokufa.edu.iq/journals/index.php/kjns/article/view/2212
[52] Nathan Favero, **(Revisiting Multicollinearity: When Correlated Predictors Exhibit Nonlinear Effects or Contain Measurement Error)** ,Department of Public Administration & Policy, School of Public Affairs, American University ,favero@american.edu .2016
[53] Panagiota Kitsantas**, (Underlying and Multiple Causes of Death in Preterm Infants)** , George Mason University, 2008 .
[54] Salman Shah1, Najam Khalique2, Zulfia Khan2,) **determinants of childhood mortalitym)** Indian J. Prev. Soc. Med. Vol. 42 No.2, 2011
[55] Samuel I.Baker **,(Non-Linear Regression).** 2008.
[56] Shalabh ,**(Linear regression analysis),** Indian Institute of Technology Kanpur.
[57] The Pennsylvania State Universit **,(Exponential Regression Example) ,** Eberly college of science ,2017
https://onlinecourses.science.psu.edu/stat501/node/372

.

الملاحق

الملاحق

ملحق (1) الخاص بالبيانات

وفياة	فترة الحمل	وزن الطفل	عمر الطفل	عمر الام	طبيعية	طفل خديج	P07	P22	Q00-Q99	غير معقم
Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
63	32.54	1.95	4.82	26.78	32	47	13	28	11	11
31	34.58	2.47	5.93	23.93	24	12	4	8	6	17
61	33.13	1.95	4.8	24.95	42	32	13	23	10	21
34	34.7	2.4	4.17	24.61	23	17	6	10	8	16
45	34.17	2.05	3.9	25.5	33	29	6	27	5	7
48	34.27	2.21	4.68	26.5	32	24	6	25	7	17
34	34.67	2.44	4.97	26.5	24	15	2	14	8	18
50	33.1	1.88	5.12	24.88	35	31	14	17	4	14
56	34.6	2.15	6.75	25.2	33	28	11	12	18	25
60	35.25	2.57	5.75	25.35	38	39	13	19	18	14
62	35.08	2.2	5.5	26.3	41	38	5	28	5	29
27	32.4	1.84	3.167	26.4	17	21	6	13	3	9
53	34.15	2.29	4.26	24.6	42	31	9	18	7	25
38	32.5	1.82	3.02	25.7	29	25	4	20	7	11
43	33.5	2.27	3.58	26.6	34	30	9	15	7	19
42	31.64	1.92	3.52	24.3	25	32	14	13	6	6
83	32.91	2.09	3.97	24.02	49	49	15	33	15	29
79	33.96	2.2	5.27	24.67	53	36	23	28	10	29
72	33.87	2.14	3.63	26.5	40	43	10	42	5	14
73	32.32	1.98	4.34	26.3	44	50	31	22	6	39
60	31.9	1.92	3.5	26.1	41	34	25	14	9	14
71	32.6	1.96	4.2	24.8	40	46	23	22	11	22
90	32.92	2.11	4.6	25.3	68	60	21	39	14	22
69	32.67	2.08	5.5	26.14	37	46	10	31	7	28
56	32.84	2.19	4.23	25.55	29	36	7	29	1	20
40	34.3	2.23	4.85	26.175	31	28	5	17	8	15
52	33.73	2.03	3.23	25.94	31	35	12	14	17	14
50	34.02	2.29	3.22	25.48	29	27	10	10	9	19
50	32.64	2.03	4.7	24.7	34	32	16	17	6	15
55	35	2.38	4.6	25.54	33	29	5	29	7	20
60	33.67	2.09	4.38	24.88	37	38	6	26	12	22
51	34	2.3	5.25	25.5	35	24	5	19	10	27
59	33.17	2.1	4.42	26.4	44	34	9	32	6	17
35	34.91	2.4	4.02	24.02	24	16	4	12	8	16
69	33.42	2.08	4.5	25.16	48	36	15	25	10	22
64	33.54	2.1	4.7	26.67	38	40	14	28	7	25

ملحق (2) الخاص بالطرائق الثلاث للانحدار المتعدد

1- طريقة ادخال كل المتغيرات :

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		.987[a]	0.973	0.963	2.8938627	0.973	91.719	10	25	0.000	2.562

a. Predictors: (Constant), وزن الطفل, ولادة طبيعية, ولادة توأم, P07, عمر الطفل, P22, Q00-Q99, غير معقم, عمر الام, فترة الحمل

b. Dependent Variable: الوفيات

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7680.945	10	768.094	91.719	.000[b]
	Residual	209.361	25	8.374		
	Total	7890.306	35			

a. Dependent Variable: الوفيات

b. Predictors: (Constant), غير معقم, عمر الام, فترة الحمل, Q00-Q99, P22, عمر الطفل, P07, وزن الطفل, ولادة طبيعية, خديج

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	20.798	34.155		.609	.548		
	فترة الحمل	.311	1.160	.020	.268	.791	.194	5.166
	وزن الطفل	.430	4.882	.005	.088	.930	.287	3.479
	عمر الطفل	.200	.713	.011	.280	.782	.662	1.511
	عمر الام	-1.205-	.656	-.067-	-1.837-	.078	.789	1.267
	ولادة طبيعية	.155	.115	.099	1.350	.189	.195	5.118
	طفل خديج	.279	.127	.196	2.191	.038	.133	7.543
	P07	.664	.151	.300	4.406	.000	.229	4.358
	P22	.828	.141	.465	5.879	.000	.170	5.892
	Q00-Q99	.658	.177	.174	3.712	.001	.484	2.064
	غير معقم	.374	.094	.175	3.963	.001	.546	1.831

a. Dependent Variable: الوفيات

2- طريقة الانحدار التدريجي :

Model Summary[d]

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	.897[a]	0.805	0.800	6.7185898	0.805	140.798	1	34	0.000	
2	.955[b]	0.912	0.906	4.5917317	0.106	39.792	1	33	0.000	
3	.964[c]	0.930	0.923	4.1683944	0.018	8.043	1	32	0.008	1.756

a. Predictors: (Constant), ولادة توأم

b. Predictors: (Constant), ولادة توأم, ولادة طبيعية

c. Predictors: (Constant), ولادة توأم, ولادة طبيعية, غير معقم

d. Dependent Variable: الوفيات

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6355.564	1	6355.564	140.798	.000[b]
	Residual	1534.741	34	45.139		
	Total	7890.306	35			
2	Regression	7194.534	2	3597.267	170.616	.000[c]
	Residual	695.772	33	21.084		
	Total	7890.306	35			
3	Regression	7334.289	3	2444.763	140.702	.000[d]
	Residual	556.016	32	17.376		
	Total	7890.306	35			

a. Dependent Variable: الوفيات

b. Predictors: (Constant), طفل خديج

c. Predictors: (Constant), ولادة طبيعية , طفل خديج

d. Predictors: (Constant), غير معقم , ولادة طبيعية , طفل خديج

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	12.908	3.731		3.460	.001		
	طفل خديج	1.278	.108	.897	11.866	.000	1.000	1.000
2	(Constant)	2.911	3.002		.970	.339		
	طفل حديج	.739	.113	.519	6.562	.000	.427	2.344
	ولادة طبيعية	.776	.123	.499	6.308	.000	.427	2.344
3	(Constant)	1.095	2.800		.391	.698		
	طفل خديج	.744	.102	.523	7.275	.000	.426	2.345
	ولادة طبيعية	.643	.121	.414	5.312	.000	.363	2.757
	غير معقم	.335	.118	.157	2.836	.008	.721	1.388

a. Dependent Variable: الوفيات

3- طريقة الحذف الخلفي : تم عرضها في الجانب العملي

4- طريقة الحذف الامامي

Model Summary[d]

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 Chang	
1	.897[a]	0.805	0.800	6.719	0.805	140.798	1	34	0.000	
2	.955[b]	0.912	0.906	4.592	0.106	39.792	1	33	0.000	
3	.964[c]	0.930	0.923	4.168	0.018	8.043	1	32	0.008	1.756

a. Predictors: (Constant), ولادة توأم

b. Predictors: (Constant), ولادة توأم, ولادة طبيعيـــة

c. Predictors: (Constant), ولادة توأم, ولادة طبيعيـــة, غير معقم

d. Dependent Variable: الوفيات

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6355.564	1	6355.564	140.798	.000[b]
	Residual	1534.741	34	45.139		
	Total	7890.306	35			
2	Regression	7194.534	2	3597.267	170.616	.000[c]
	Residual	695.772	33	21.084		
	Total	7890.306	35			
3	Regression	7334.289	3	2444.763	140.702	.000[d]
	Residual	556.016	32	17.376		
	Total	7890.306	35			

a. Dependent Variable: الوفيات

b. Predictors: (Constant), طفل خديج

c. Predictors: (Constant), طفل خديج , ولادة طبيعية

d. Predictors: (Constant), طفل خديج , ولادة طبيعية , غير معقم

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	12.908	3.731		3.460	.001		
	طفل خديج	1.278	.108	.897	11.866	.000	1.000	1.000
2	(Constant)	2.911	3.002		.970	.339		
	طفل حديج	.739	.113	.519	6.562	.000	.427	2.344
	ولادة طبيعية	.776	.123	.499	6.308	.000	.427	2.344
3	(Constant)	1.095	2.800		.391	.698		
	طفل خديج	.744	.102	.523	7.275	.000	.426	2.345
	ولادة طبيعية	.643	.121	.414	5.312	.000	.363	2.757
	غير معقم	.335	.118	.157	2.836	.008	.721	1.388

a. Dependent Variable: الوفيات

ملحق (3) الخاص بالطرائق الثلاث للانحدار اللوغارتمي

1- طريقة ادخال جميع المتغيرات : تم عرضها في الجانب العملي

2- طريقة الانحدار التدريجي :

Model Summary[d]

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	.905[a]	0.820	0.815	0.05367252	0.820	154.694	1	34	0.000	
2	.956[b]	0.914	0.909	0.03756632	0.095	36.404	1	33	0.000	
3	.966[c]	0.933	0.927	0.03373310	0.019	8.926	1	32	0.005	1.640

a. Predictors: (Constant), ولادة طبيعيـة

b. Predictors: (Constant), ولادة طبيعيـة, ولادة تـوأم

c. Predictors: (Constant), ولادة طبيعيـة, ولادة تـوأم, غير معقم

d. Dependent Variable: وفيات

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.446	1	.446	154.694	.000[b]
	Residual	.098	34	.003		
	Total	.544	35			
2	Regression	.497	2	.249	176.091	.000[c]
	Residual	.047	33	.001		
	Total	.544	35			
3	Regression	.507	3	.169	148.565	.000[d]
	Residual	.036	32	.001		
	Total	.544	35			

a. Dependent Variable: وفيات

b. Predictors: (Constant), طبيعية ولادة

c. Predictors: (Constant), طفل خديج, طبيعية ولادة

d. Predictors: (Constant), معقم غير, طفل خديج, طبيعية ولادة

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.229	.121		1.895	.067		
	ولادة طبيعية	.972	.078	.905	12.438	.000	1.000	1.000
2	(Constant)	.245	.084		2.899	.007		
	ولادة طبيعية	.591	.084	.551	7.078	.000	.429	2.331
	طفل خديج	.381	.063	.469	6.034	.000	.429	2.331
3	(Constant)	.249	.076		3.277	.003		
	ولادة طبيعية	.450	.089	.420	5.084	.000	.307	3.253
	طفل خديج	.423	.058	.521	7.241	.000	.404	2.475
	غير معقم	.120	.040	.168	2.988	.005	.660	1.515

a. Dependent Variable: وفيات

3- طريقة الحذف الخلفي :

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	.979[a]	0.959	0.942	0.03003649	0.959	57.751	10	25	0.000	
2	.979[b]	0.958	0.944	0.02954987	0.000	0.164	1	25	0.689	
3	.979[c]	0.958	0.945	0.02916702	0.000	0.305	1	26	0.586	
4	.978[d]	0.956	0.944	0.02936130	-0.002	1.374	1	27	0.251	
5	.975[e]	0.951	0.941	0.03029015	-0.005	2.864	1	28	0.102	2.219

a. Predictors: (Constant), وزن الطفل, ولادة طبيعيــة, ولادة تـوأم, P07, عمر الطفل, Q00-Q99, P22, غير معقم, عمر الام, فترة الحمل

b. Predictors: (Constant), وزن الطفل, ولادة طبيعيــة, ولادة تـوأم, P07, عمر الطفل, Q00-Q99, P22, غير معقم, عمر الام

c. Predictors: (Constant), وزن الطفل, ولادة طبيعيــة, ولادة تـوأم, P07, Q00-Q99, P22, غير معقم, عمر الام

d. Predictors: (Constant), وزن الطفل, ولادة طبيعيــة, ولادة تـوأم, P07, Q00-Q99, P22, غير معقم

e. Predictors: (Constant), ولادة طبيعيــة, ولادة تـوأم, P07, Q00-Q99, P22, غير معقم

f. Dependent Variable: وفيات

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.521	10	.052	57.751	.000[b]
	Residual	.023	25	.001		
	Total	.544	35			
2	Regression	.521	9	.058	66.280	.000[c]
	Residual	.023	26	.001		
	Total	.544	35			
3	Regression	.521	8	.065	76.496	.000[d]
	Residual	.023	27	.001		
	Total	.544	35			
4	Regression	.519	7	.074	86.077	.000[e]
	Residual	.024	28	.001		
	Total	.544	35			
5	Regression	.517	6	.086	93.910	.000[f]
	Residual	.027	29	.001		
	Total	.544	35			

a. Dependent Variable: وفيات

b. Predictors: (Constant), طفل خديج, طبيعية ولادة, الطفل وزن, P07, الطفل عمر, Q00-Q99, P22, الحمل فترة, الام عمر, معقم غير

c. Predictors: (Constant), طفل خديج, طبيعية ولادة, الطفل وزن, P07, الطفل عمر, Q00-Q99, P22, الام عمر, معقم غير

d. Predictors: (Constant), طفل خديج, طبيعية ولادة, الطفل وزن, Q00-Q99, P22, P07, الام عمر, معقم غير

f. Predictors: (Constant), طفل خديج, طبيعية ولادة, Q00-Q99, P22, P07, معقم غير

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	VIF
		B	Std. Error	Beta			
1	(Constant)	.501	1.350		.371	.714	
	فترة الحمل	.355	.876	.035	.405	.689	4.593
	وزن الطفل	.189	.258	.057	.730	.472	3.670
	عمر الطفل	.034	.077	.022	.435	.667	1.557
	عمر الام	-.462-	.406	-.053-	-1.138-	.266	1.317
	ولادة طبيعية	.156	.111	.145	1.401	.173	6.447
	طفل خديج	.265	.093	.327	2.847	.009	3.930
	P07	.127	.045	.274	2.814	.009	5.722
	P22	.224	.070	.321	3.178	.004	3.161
	Q00-Q99	.047	.029	.088	1.633	.115	1.733
	غير معقم	.125	.041	.175	3.066	.005	1.965
2	(Constant)	.995	.571		1.743	.093	
	وزن الطفل	.260	.186	.079	1.400	.173	1.961
	عمر الطفل	.041	.074	.027	.552	.586	1.474
	عمر الام	-.448-	.398	-.052-	-1.126-	.270	1.308
	ولادة طبيعية	.161	.109	.150	1.478	.151	6.370
	طفل خديج	.262	.091	.323	2.870	.008	3.872
	P07	.122	.043	.263	2.858	.008	5.281
	P22	.225	.069	.323	3.247	.003	3.153
	Q00-Q99	.050	.027	.093	1.836	.078	1.610
	غير معقم	.124	.040	.173	3.094	.005	1.954
3	(Constant)	1.022	.562		1.819	.080	
	وزن الطفل	.286	.178	.086	1.606	.120	1.841
	عمر الام	-.460-	.392	-.053-	-1.172-	.251	1.304
	ولادة طبيعية	.161	.107	.150	1.499	.145	6.370
	طفل خديج	.255	.089	.314	2.858	.008	3.725
	P07	.123	.042	.267	2.949	.007	5.249
	P22	.230	.068	.331	3.405	.002	3.028
	Q00-Q99	.052	.027	.096	1.928	.064	1.593
	غير معقم	.130	.038	.182	3.435	.002	1.796
4	(Constant)	.375	.107		3.495	.002	
	وزن الطفل	.302	.178	.091	1.692	.102	1.830
	ولادة طبيعية	.184	.106	.171	1.734	.094	6.153
	طفل خديج	.231	.087	.284	2.640	.013	3.307
	P07	.132	.042	.286	3.177	.004	5.093
	P22	.223	.068	.320	3.284	.003	3.971
	Q00-Q99	.052	.027	.098	1.946	.062	1.592
	غير معقم	.127	.038	.178	3.336	.002	1.786
5	(Constant)	.463	.097		4.763	.000	
	ولادة طبيعية	.214	.108	.199	1.982	.057	5.982
	طفل خديج	.209	.089	.258	2.348	.026	3.156
	P07	.111	.041	.241	2.718	.011	4.657
	P22	.212	.070	.304	3.046	.005	3.921
	Q00-Q99	.059	.027	.110	2.135	.041	1.561
	غير معقم	.147	.037	.206	3.972	.000	1.601

4- طريقة الحذف الامامي

Model Summary[d]

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	.905[a]	0.820	0.815	0.05367252	0.820	154.694	1	34	0.000	
2	.956[b]	0.914	0.909	0.03756632	0.095	36.404	1	33	0.000	
3	.966[c]	0.933	0.927	0.03373310	0.019	8.926	1	32	0.005	1.640

a. Predictors: (Constant), ولادة طبيعية

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	0.446	1	0.446	154.694	.000[b]
	Residual	0.098	34	0.003		
	Total	0.544	35			
2	Regression	0.497	2	0.249	176.091	.000[c]
	Residual	0.047	33	0.001		
	Total	0.544	35			
3	Regression	0.507	3	0.169	148.565	.000[d]
	Residual	0.036	32	0.001		
	Total	0.544	35			

a. Dependent Variable: وفيات

b. Predictors: (Constant), ولادة طبيعية

c. Predictors: (Constant), ولادة طبيعية , طفل خديج

d. Predictors: (Constant), ولادة طبيعية , طفل خديج, غير معقم

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	0.229	0.121		1.895	0.067		
	ولادة طبيعية	0.972	0.078	0.905	12.438	0.000	1.000	1.000
2	(Constant)	0.245	0.084		2.899	0.007		
	ولادة طبيعية	0.591	0.084	0.551	7.078	0.000	0.429	2.331
	طفل خديج	0.381	0.063	0.469	6.034	0.000	0.429	2.331
3	(Constant)	0.249	0.076		3.277	0.003		
	ولادة طبيعية	0.450	0.089	0.420	5.084	0.000	0.307	3.253
	طفل خديج	0.423	0.058	0.521	7.241	0.000	0.404	2.475
	غير معقم	0.120	0.040	0.168	2.988	0.005	0.660	1.515

a. Dependent Variable: وفيات

ملحق (4) الخاص بالطرائق الثلاث للانحدار اللوجستي

1- طريقة ادخال جميع المتغيرات :

Classification Table[a,b]

			Predicted		
	Observed		حياة او وفاة		Percentage Correct
			حياة	وفاة	
Step 0	حياة او وفاة	حياة	11401	0	100.0
		وفاة	640	0	.0
	Overall ercentage				94.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.880-	.041	5026.223	1	.000	.056

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients														
			Constant	قرة الحمل	وزن الطفل	توام او مفرد(1)	توام او مفرد(2)	سبب الرقود(1)	سبب الرقود(2)	سبب الرقود(3)	سبب الرقود(4)	سبب الرقود(5)	سبب الرقود(6)	سبب الرقود(7)	عمر الام	معقم او لا(1)	نوع الولادة(1)
Step 1	1	5585.344	1.755	-0.083	-0.001	-0.467	-0.515	0.379	0.084	0.009	-0.056	0.127	-0.152	1.057	0.000	0.020	-0.039
	2	4274.221	4.363	-0.163	-0.003	-0.815	-0.939	0.609	0.222	0.036	-0.127	0.302	-0.445	1.784	-0.001	0.040	-0.142
	3	3977.187	5.873	-0.206	-0.006	-0.994	-1.211	0.675	0.378	0.077	-0.202	0.462	-0.998	2.186	-0.001	0.048	-0.312
	4	3911.928	6.289	-0.216	-0.019	-1.031	-1.298	0.660	0.438	0.091	-0.244	0.509	-1.708	2.300	-0.001	0.042	-0.405
	5	3886.454	6.265	-0.207	-0.134	-1.021	-1.291	0.585	0.425	0.065	-0.243	0.477	-2.269	2.279	-0.001	0.023	-0.377
	6	3868.091	5.828	-0.174	-0.449	-1.003	-1.251	0.481	0.435	0.051	-0.182	0.448	-2.471	2.259	0.000	-0.017	-0.257
	7	3867.764	5.779	-0.170	-0.493	-1.009	-1.255	0.474	0.443	0.053	-0.172	0.450	-2.494	2.270	0.000	-0.021	-0.240
	8	3867.764	5.779	-0.170	-0.493	-1.010	-1.255	0.474	0.443	0.053	-0.172	0.450	-2.495	2.271	0.000	-0.021	-0.240

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1133.893	14	.000
	Block	1133.893	14	.000
	Model	1133.893	14	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3867.764[a]	.090	.264

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	15.373	8	.052

Contingency Table for Hosmer and Lemeshow Test

		حياة = حياة او وفاة		وفاة = حياة او وفاة		Total
		Observed	Expected	Observed	Expected	
Step 1	1	1196	1198.701	5	2.299	1201
	2	1190	1193.545	7	3.455	1197
	3	1201	1199.314	3	4.686	1204
	4	1177	1182.182	25	19.818	1202
	5	1176	1171.684	28	32.316	1204
	6	1177	1163.702	27	40.298	1204
	7	1155	1155.133	49	48.867	1204
	8	1137	1141.908	67	62.092	1204
	9	1112	1104.695	92	99.305	1204
	10	880	890.135	337	326.865	1217

Classification Table[a]

	Observed		Predicted		
			حياة او وفاة		Percentage Correct
			حياة	وفاة	
Step 1	او وفاة حياة	حياة	11371	30	99.7
		وفاة	576	64	10.0
	Overall Percentage				95.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	فترة الحمل	-.170-	.018	85.526	1	.000	.844
	وزن الطفل	-.493-	.086	32.866	1	.000	.611
	حالة الولادة			7.343	2	.025	
	غير خديج(1)	-1.010-	1.295	.608	1	.436	.364
	خديج (2)	-1.255-	1.296	.938	1	.333	.285
	سبب الرقود			363.752	7	.000	
	سبب الرقود(1)	.474	.209	5.156	1	.023	1.606
	سبب الرقود(2)	.443	.205	4.670	1	.031	1.557
	سبب الرقود(3)	.053	.172	.095	1	.757	1.055
	سبب الرقود(4)	-.172-	.336	.262	1	.609	.842
	سبب الرقود(5)	.450	.211	4.542	1	.033	1.569
	سبب الرقود(6)	-2.495-	.302	68.249	1	.000	.083
	سبب الرقود(7)	2.271	.201	128.168	1	.000	9.686
	عمر الام	.000	.008	.002	1	.967	1.000
	معقم او لا(1)	-.021-	.107	.040	1	.842	.979

	نوع الولادة(1)	-.240-	.118	4.121	1	.042	.787
	Constant	5.779	1.426	16.412	1	.000	323.425
a. Variable(s) entered on step 1: الحمل فترة ,الطفل وزن ,مفرد او توام ,سبب الرقود ,عمر الام, الولادة نوع ,لا او معقم.							

2- **طريقة الحذف الامامي :**

Iteration History[a,b,c]			
Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	6015.170	-1.787-
	2	5088.164	-2.522-
	3	5003.206	-2.830-
	4	5001.658	-2.879-
	5	5001.658	-2.880-
	6	5001.658	-2.880-
a. Constant is included in the model.			
b. Initial -2 Log Likelihood: 5001.658			
c. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.			

Classification Table[a,b]					
			Predicted		
			حياة او وفاة		Percentage Correct
	Observed		حياة	وفاة	
Step 0	حياة او وفاة	حياة	11401	0	100.0
		وفاة	640	0	.0
	Overall Percentage				94.7
a. Constant is included in the model.					
b. The cut value is .500					

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.880-	.041	5026.223	1	.000	.056

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	717.579	7	.000
	Block	717.579	7	.000
	Model	717.579	7	.000
Step 2	Step	354.804	1	.000
	Block	1072.384	8	.000

	Model	1072.384	8	.000
Step 3	Step	14.071	1	.000
	Block	1086.455	9	.000
	Model	1086.455	9	.000
Step 4	Step	9.401	2	.009
	Block	1095.856	11	.000
	Model	1095.856	11	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4284.078[a]	.058	.170
2	3929.274[a]	.085	.251
3	3915.203[a]	.086	.254
4	3905.802[a]	.087	.256

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	.000	3	1.000
2	50.938	7	.000
3	41.715	8	.000
4	9.773	8	.281

Classification Table[a]

Observed			Predicted		
			حياة او وفاة		Percentage Correct
			حياة	وفاة	
Step 1	حياة او وفاة	حياة	11401	0	100.0
		وفاة	640	0	.0
	Overall Percentage				94.7
Step 2	حياة او وفاة	حياة	11366	35	99.7
		وفاة	595	45	7.0
	Overall Percentage				94.8
Step 3	حياة او وفاة	حياة	11364	37	99.7
		وفاة	593	47	7.3
	Overall Percentage				94.8
Step 4	حياة او وفاة	حياة	11366	35	99.7
		وفاة	594	46	7.2
	Overall Percentage				94.8

a. The cut value is .500

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	سبب الرقود			512.172	7	.000	
	سبب الرقود(1)	1.898	.186	104.072	1	.000	6.673
	سبب الرقود(2)	.593	.192	9.500	1	.002	1.809
	سبب الرقود(3)	.289	.162	3.193	1	.074	1.336
	سبب الرقود(4)	-.217-	.329	.437	1	.509	.805
	سبب الرقود(5)	.749	.201	13.904	1	.000	2.114
	سبب الرقود(6)	-2.469-	.298	68.745	1	.000	.085
	سبب الرقود(7)	2.351	.192	149.756	1	.000	10.495
	Constant	-3.109-	.148	444.096	1	.000	.045
Step 2[b]	فترة الحمل	-.259-	.013	374.498	1	.000	.771
	سبب الرقود			390.051	7	.000	
	سبب الرقود(1)	.807	.204	15.619	1	.000	2.241
	سبب الرقود(2)	.505	.197	6.544	1	.011	1.657
	سبب الرقود(3)	.146	.166	.768	1	.381	1.157
	سبب الرقود(4)	-.249-	.333	.562	1	.454	.779
	سبب الرقود(5)	.554	.208	7.090	1	.008	1.741
	سبب الرقود(6)	-2.471-	.301	67.599	1	.000	.084
	سبب الرقود(7)	2.355	.198	141.353	1	.000	10.539
	Constant	6.314	.501	158.561	1	.000	552.101
Step 3[c]	فترة الحمل	-.225-	.016	197.951	1	.000	.798
	سبب الرقود			377.117	7	.000	
	سبب الرقود(1)	.697	.205	11.603	1	.001	2.007
	سبب الرقود(2)	.425	.199	4.555	1	.033	1.529
	سبب الرقود(3)	.098	.167	.343	1	.558	1.103
	سبب الرقود(4)	-.253-	.333	.575	1	.448	.777
	سبب الرقود(5)	.528	.208	6.419	1	.011	1.696
	سبب الرقود(6)	-2.463-	.300	67.299	1	.000	.085
	سبب الرقود(7)	2.313	.199	134.589	1	.000	10.100
	نوع الولادة(1)	-.424-	.112	14.342	1	.000	.655
	Constant	5.405	.552	95.932	1	.000	222.531
Step 4[d]	فترة الحمل	-.219-	.016	184.974	1	.000	.803
	حالة الولادة			9.387	2	.009	
	غير خديج(1)	-1.029-	1.291	.636	1	.425	.357
	خديج(2)	-1.303-	1.292	1.018	1	.313	.272
	سبب الرقود			376.883	7	.000	
	سبب الرقود(1)	.676	.205	10.890	1	.001	1.966
	سبب الرقود(2)	.468	.200	5.506	1	.019	1.597
	سبب الرقود(3)	.111	.167	.443	1	.506	1.118
	سبب الرقود(4)	-.261-	.333	.616	1	.433	.770
	سبب الرقود(5)	.505	.209	5.867	1	.015	1.658
	سبب الرقود(6)	-2.489-	.300	68.674	1	.000	.083
	سبب الرقود(7)	2.321	.200	135.089	1	.000	10.190

	نوع الولادة(1)	-.431-	.112	14.838	1	.000	.650
	Constant	6.330	1.405	20.301	1	.000	561.210
a. Variable(s) entered on step 1: الرقود سبب.							
b. Variable(s) entered on step 2: الحمل فترة.							
c. Variable(s) entered on step 3: الولادة نوع.							
d. Variable(s) entered on step 4: مفرد او توام.							

3- طريقة الحذف الخلفي : تم عرضها في الجانب العملي

ABSTRACT

The level of premature mortality is an important indicator of the progress of health services and health education in any society. This topic has been selected due to a few studies that have been examined in this area, especially in the province of Babylon and in Iraq in general.

To achieve this, the researcher relied on the data obtained from the Department of Health of Babylon, which included the monthly totals of the number of preterm infants for the period 2014-2016, and included data for each child in the unit of preterm infants and the condition of exit (alive or dead).

The hypothesis of the study was to determine the effect of a number of factors (pregnancy, child weight, age of the child, age of the mother, Birth status , type of birth, sterilization) on the dependent variable. With a view to developing solutions to reduce this phenomenon and to develop measures to counter or control it in the future through the formulation of medical and health policies that contribute to reducing mortality, as well as the introduction of maternal and newborn health programs, continuity of care including prenatal services, the adoption of qualified staff during childbirth and follow-up during the first month of life.

The purpose of this thesis is to analyze this phenomenon and to show its side effects, in addition to building a suitable statistical model for the nature of the data, through which we can know the effect of these factors on the mortality of premature infants of each variable, and predict the number of deaths in the next stage , Such as multiple regression model, logarithmic regression model, exponential regression model for quantitative data study, logistic regression model, differential function analysis for the study of descriptive response data.

The study found a significant relationship between the different variables (influencing factors) and the dependent variable. The study also showed the suitability of statistical models based on a number of statistical indicators and tests.

The study finding some conclution, and recommendations , the most important of which is the need to pay attention to the level of medical care for preterm infants , Paying attention to factors affecting child mortality and developing appropriate solutions. The thesis also recommends the use of the logarithmic regression model in the case of quantitative data and logistic regression model in the case of dependent binary variable data.

Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Karbala
College of Administration and Economics
Department of Statistics

Selecting the best statistical model for analyzing the mortality of preterm infants (applied study)

A Thesis

Submitted to the Council of Faculty of Administration and Economics at the University of Karbala as a partial Fulfilment of the Requirements for the Degree of Master in statistics science

By

Maithem Abdel Wahab Saleh

Under supervision of

PROF. Dr . Adnan Karim Najm ALdin

1439 ***Karbala*** ***2018***