جامعة كربلاء
كلية الإدارة والاقتصاد
قسم الإحصاء

تقدير معلمات انحدار *log-logistic* باستخدام الخوارزمية الجينية مع تطبيق عملي

رسالة مقدمة إلى

مجلس كلية الإدارة والاقتصاد في جامعة كربلاء وهي جزء من متطلبات نيل درجة ماجستير في علوم الإحصاء

تقدم بها الطالب
حسين خليل عبيد مخيلف

بإشراف
أ.م.د مشتاق كريم عبد الرحيم

1444هـ كربـــلاء المقدســـة 2022م

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

﴿ وَلَتَعْلَمُنَّ نَبَأَهُ بَعْدَ حِينٍ ﴾

صدق الله العظيم

سورة ص الآية ﴿88﴾

إقرار المشـرف

أشـهد أن إعـداد هـذه الرسـالة الموسـومة بــ(**تقـدير معلمــات انحـــــدار Log-Logistic بأستخدام الخوارزميـة الجينيـة مـع تطبيق عملـي**) والتـي تقدم بها الطالب **" حسين خليل عبيد"** قد جرى بإشرافي في قسم الاحصـاء - كلية الادارة والاقتصاد - جامعة كربلاء، وهي جزء من متطلبات نيل درجة ماجستير علوم في الاحصاء.

أ.م.د. مشتاق كريم عبد الرحيم

التاريخ: / / 2022

توصية رئيس قسم الاحصاء

بناءً على توصية الاستاذ المشرف، أرشح الرسالة للمناقشة.

أ.د. شروق عبد الرضا السباح

رئيس قســم الاحصاء

التاريخ: / / 2022

إقرار الخبير اللغوي

أشـهد أن الرسـالة الموسـومة بـ(**تقـدير معلمـات انحـدار Log-Logistic بأستخدام الخوارزمية الجينية مع تطبيق عملـي**) من الطالب (**حسين خليل عبيد**) / قسم الإحصـاء قد جرى مراجعتها من الناحيـة اللغويـة حتـى غدت خاليـة من الاخطاء اللغوية والاسلوبية ولأجله وقعت.

الخبير اللغوي

م. صلاح مهدي جابر

جامعة كربلاء – كلية الادارة والاقتصاد

إقرار لجنة المناقشة

نشهد نحن أعضاء لجنة المناقشة بأننا قد اطلعنا على الرسالة الموسومة (**تقدير معلمات انحدار Log-Logistic بأستخدام الخوارزمية الجينية مع تطبيق عملي**) والمقدمة من قبل الطالب "**حسين خليل عبيد** " وناقشنا الطالب في محتوياتها وفيما له علاقة بها، ووجدنا بأنها جديرة بنيل درجة ماجستير علوم في الإحصاء بتقدير ( ).

أ.د. عواد كاظم شعلان	**أ.م.د صدى فايض محمد**
رئيساً	**عضواً**
/ / 2022	**/ / 2022**
م.د نمارق قاسم حسين	**أ.م.د مشتاق كريم عبد الرحيم**
عضواً	**عضواً ومشرفاً**
/ / 2022	**/ / 2022**

إقرار رئيس لجنة الدراسات العليا

بناء على اقرار المشرف العلمي والخبير اللغوي على رسالة الماجستير للطالب "**حسين خليل عبيد**" الموسومة بـ (**تقدير معلمات انحدار Log-Logistic بأستخدام الخوارزمية الجينية مع تطبيق عملي**) ارشح هذه الرسالة للمناقشة.

أ.د محمد حسين كاظم الجبوري

رئيس لجنة الدراسات العليا

معاون العميد للشؤون العلمية والدراسات العليا

مصادقة مجلس الكلية

صادق مجلس كلية الادارة والاقتصاد/ جامعة كربلاء على قرار لجنة المناقشة.

أ.د. محمد حسين كاظم الجبوري

عميد كلية الادارة والاقتصاد- جامعة كربلاء

١٩ / ١ / ٢٠٢٢

الإهداء

إلى معلم الإنسانية ومنقذ البشرية من الضلال والظلام إلى الهدى والنور

رسول الحق والهدى والعلم سيدنا محمد (ص)

إلى مَنْ أُفضِّلها على نفسي، مَنْ ضحَّت من أجلي ولم تدَّخر جُهدًا في سبيل إسعادي على الدَّوام

(أُمِّي الحبيبة).

عندما نسير في دروب الحياة يبقى من يُسيطر على أذهاننا في كل مسلك نسلكه، صاحب الوجه الطيب، والأفعال الحسنة من لم يبخل عليَّ طوال حياته الذي تعب وبذل مجهود لأجل كل لحظة فرح وسعادة عشتها الذي زرع القيم والمبادئ التي اوصلتني لما انا عليه إلان

(أبي حبيبي رحمك الله يا قرة عيني).

هنالك الكثير ممن يتمنون لك كل خير ويدعون لك في ظهر الغيب، هم أغلى ما نملك في هذه الحياة ولولا وجودهم ماوصلت الى هذا المكان

(إخوتي وأخواتي)

إلى مَنْ اعزها الله في قلبي احتراماً وفي الوجود تميزاً...من تطيب الحياة بقربها وتزهو الأيام بها (شمسي وقمري عزيزةُ روحي)

تواجهنا في الحياة عقبات كثيرة، نتعثّر ببعضها ونتجاوز الآخر بهمة وتفاؤل وحبّ للحياة

(رفقاء الدرب إصدقائي)

الى كل من التقيت بهم في مسيرتي العلمية ..... اسأتذتي الافاضل.. زملائي الكــرام

إلى الأكرم منا جميعا.. من قدموا أرواحهم وبذلوا أنفسهم من أجل الوطن

شهداء العراق

أهدي لهم هذا الجهد المتواضع

حسين

شكر وتقدير

الحمد الله والشكر له كما ينبغي لجلال وجهه وعظيم سلطانه، عدد خلقه ورضا نفسه وزنة عرشه ومداد كلماته على أن منَّ عليَّ بإنجاز هذه الرسالة، والصلاة والسلام على أفضل الخلق نبينا محمد وعلى آله وصحبه وسلم تسليما كثيرا. أتقدم بوافر الشكر والتقدير والاحترام إلى قدوتي وأعظم الرجال في حياتي الى من كان صديقي وحبيبي وسندي والدي العزيز رحمه الله وشكري الى أمي الحبيبة الذي تدعو لي دائما وتتمنى لي كل خير في هذه الحياة.

كما أتقدم بالشكر الجزيل وخالص الامتنان والاعتزاز لمشرفي الفاضل الأستاذ المساعد الدكتور (مشتاق كريم عبد الرحيم) لقبوله الإشراف على رسالتي ولما قدمه من دعم وتوجيهات علمية وآراء سديدة ومساندتي وإرشادي بالنصح والتصحيح إذ كان له الأثر الكبير في إخراج هذه الرسالة بالصورة التي هي عليها، فضلا عن نُبل أخلاقه الرفيعة التي يشهد بها القاصي والداني فحفظه الله ورعاه ذخرا للعلم والمعرفة.

كما يسرني ويسعدني أن أتقدم بوافر الشكر والامتنان للأساتذة الفضلاء رئيس لجنة المناقشة وأعضائها المحترمين بقبول مناقشة هذه الرسالة.

وأتقدم بالشكر الجزيل إلى المقوّمَين العلميّيَن والمقوّم اللغوي لتفضلهم بمراجعة هذه الرسالة وتدقيقها، فجزاهم الله خيرا ووفقهم لكل خير.

وشكري موجه إلى إدارة كلية الإدارة والاقتصاد/ جامعة كربلاء والى قسم الإحصاء، والى جميع أساتذتي الفضلاء الذين لم يدّخروا جهدا ولا معلومة كي يرتقوا بنا علميا وأفاضوا علينا مما كرّمهم الله به من علم.

والشكر الجزيل لرفقاء الدرب زملائي الأعزاء وبالاخص زميلتي صفا نجاح، وأخيرا يسرني أن أتوجه بالشكر لكل من نصحني وأرشدني ووجهني وأسهم معي ومدّ يد العون في إعداد هذا البحث ولم أذكره.

أسأل الله العليّ القدير أن يوفق الجميع لما يحبه ويرضاه

والله وليّ التوفيق

الباحث

قائمـــة المحتويات

العنـــــــــوان		رقم الصفحة
الآية الكريمة		أ
الاهداء		ب
شكر وتقدير		ج
قائمة المحتويات		د ــ ز
قائمة الجداول		ح
قائمة الأشكال		ط
قائمة الرموز والمصطلحات		ي
المستخلص		ك
الفقرة	**الفصل الأول المقدمة والاستعراض المرجعي**	**رقم الصفحة**
1-1	المقدمة	2
2-1	مشكلة الدراسة	2
3-1	هدف الدراسة	3
4-1	الاستعراض المرجعي	6-3
الفقرة	**الفصل الثاني: الجانب النظري**	**رقم الصفحة**
1-2	تمهيد	9
2-2	إنموذج الانحدار الخطي	9
3-2	إنموذج الانحدار اللاخطي	11
4-2	الثنائي لوجستي	11

الفقرة	الفصل الثاني: الجانب النظري	رقم الصفحة
5-2	إنموذج الانحدار (Log-Logistic) الثنائي	14
1-5-2	شروط (Log-Logistic) الثنائي	14
2-5-2	أفتراضات (*log-logistic*)الثنائي	15
3-5-2	مجالات التي يستعمل بها (log-logistic) الثنائي	15
6-2	إنموذج الانحدار المتعدد	16
7-2	الدالة الوجستية (Logistic function)	17
8-2	إنموذج اللوجت (*logit model*)	19
9-2	*طرائق تقدير معلمات إنموذج (Log-Logistic) الثنائي*	20
1-9-2	طريقة تقدير الامكان الاعظم	20
1-1-9-2	نيوتن رافسون (*Newton Raphson*)	25
2-1-9-2	خطوات خوارزمية نيوتن رافسون	26
3-1-9-2	نسبة الارجحية	26
2-9-2	طريقة المربعات الصغرى الموزونة	27
3-9-2	طريقة تصغير مربع كاي	32
10-2	الاختبارات المتعلقة بأنموذج الانحدار (Log-Logistic)الثنائي	34
1-10-2	اختبار هوزمر- ليمشو لجودة المطابقة	35
2-10-2	اختبار ولد	36
3-10-2	معيار المقارنة بين طرائق التقدير المستعملة	37
11-2	مفهوم الخوارزمية الجينية	38
1-11-2	اهم ما يميز الخوارزمية الجينية	42

2-11-2	المشاكل التي تعالجها الخوارزمية الجينية	42
3-11-2	عمل الخوارزمية الجينية	42
12-2	منهجيـــــــــة الخوارزمية الجينية	43
1-12-2	محتويات الخوارزمية الجينية	44
2-12-2	مراحل الخوارزمية الجينية	46
3-12-2	المعايير التي تستعمل في الخوارزمية الجينية	49
13-2	تطبيق مراحل الخوارزمية الجينية في انموذج الانحدار	53
الفقرة	**الفصل الثالث: الجانب التجريبي والتطبيقي**	رقم الصفحة
1-3	تمهيد	63
2-3	مفهوم المحاكاة	63
3-3	مراحل بناء تجربة المحاكاة	64
1-3-3	تحديد القيم الافتراضية	64
2-3-3	توليد البيانات	65
3-3-3	التقديرات	65
4-3-3	المقارنة بين طرائق التقدير الاعتيادية والحديثة	66
4-3	تحليل نتائج المحاكاة	66
5-3	تمهيد	70
6-3	أمراض القلب	70
7-3	البيانات الحقيقة	71
8-3	تحليل النتائج التطبيقية	75
الفقرة	**الفصل الرابع: الاستنتاجات والتوصيات**	رقم الصفحة

1-4	الاستنتاجات	84
2-4	التوصيات	85
المصادر		90-88
الملاحق		95-91
Abstract		A

قائمـــــة الجـــــــداول

رقم الجدول	عنـــوان الجـــــدول	رقم الصفحة
(1-3)	القيم الافتراضية للمعلمات في انموذج الانحدار (Log-Logistic) الثنائي	65
(2-3)	متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار (Log-Logistic) الثنائي في الانموذج(1) للمعلمات باستعمال الطرائق الاعتيادية والجينية وكافة احجام العينات	67
(3-3)	متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار(Log-Logistic) الثنائي في الانموذج (2) للمعلمات باستعمال الطرائق الاعتيادية والجينية وكافة احجام العينات	68
(4-3)	متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار (Log-Logistic) الثنائي للانموذجين باستعمال افضل الطرائق الاعتيادية وافضل الجينية وكافة احجام العينات	68
(5-3)	البيانات الحقيقة لمتغير الاستجابة للمرضى المصابين بأمراض القلب وكذلك المتغيرات التوضيحية لسنة 2015 لعينة حجمها (90) (اعداد الباحث).	72
(6-3)	جدول (3 - 6) قيمة اختبار 2 للمتغيرات التوضيحية	75
(7-3)	يمثل المعلمات المقدرة وكذلك الخطأ المعياري لكافة المتغيرات التوضيحية بطريقة مربع كاي المحسنة (MCSE.GA)(أعداد الباحث)	76
(8-3)	قيمة المعلمات المقدرة وألخطا المعياري لجميع المتغيرات التوضيحية بطريقة المربعات الصغرى الموزونة (WLSE) (أعداد الباحث).	78
(9-3)	تقدير معلمات أنموذج الانحدار (Log-Logistic)الثنائي بطريقة المربعات الصغرى الموزونة الأعتيادية وطريقة أصغر مربع كاي الجينية (أعداد الباحث)	79
(10-3)	تصنيف البيانات للعينة عن طريق استعمال الأنموذج المقدر بطريقة تصغير مربـــــــــــع كاي الجينية (MCSE.GA) (أعداد الباحـــث)	80
(11-3)	تصنيف البيانات للعينة عن طريق استعمال الأنموذج المقدر بطريقة المربعـــــــــات الصغرى الموزونة (WLSE). (اعداد الباحـــث)	81

قائمـــة الأشكال

رقم الشكل	عنـوان الشــكـل	رقم الصفحة
(1-2)	الانحدار الخطي بين المتغير التابع (y) والمتغير المستقل (x)	10
(2-2)	العلاقة غير الخطية بين متغير الاستجابة $\pi(X)$ والمتغير التوضيحي (X)	11
(3-2)	الدالة اللوجستية	18
(4-2)	العلاقة الخطية بين i و Z_i	20
(5-2)	المخطط العام للخوارزمية الجينية من تصميم الباحث	41
(6-2)	التشفير الاستبدالي للكروموسوم	45
(7-2)	التشفير الثنائي للكروموسوم	45
(8-2)	عجلة الروليت	50
(9-2)	التهجين المنظم	52
(10-2)	يوضح التهجين البسيط	52
(11-2)	مخطط انسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات MLE (المخطط من عمل الباحث)	57
(12-2)	المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات WLSE (المخطط من عمل الباحث)	58
(13-2)	المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات MCSE (المخطط الانسيابي من عمل الباحث) .	59

قائمة الرموز والمصطلحات

الرمز	المصطلح باللغة العربية	المصطلح باللغة الانكليزية
MLE	طريقة الامكان الاعظم	Maximum Likelihood Method
WLSM	طريقة المربعات الصغرى الموزونة	weighted least squares method
MCSE	طريقة تصغير مربع كاي	Minimum Chi - Square Method
G.A	الخوارزمية الجينية	Genetic Algorithm
$\pi(Xi)$	احتمال النجاح	probability of success
$1\pi(X_i)$-	احتمال الفشل	probability of failure
$logi\ (\pi_i)$	أنموذج اللوجت	logit model
$\beta_\circ, \beta_1$	معلمتا انموذج الانحدار اللوجستي	The two parameters of the logistic regression model
α	قيمة ثابتة	Fixed value
$V^{(m)}$	مصفوفة مربعة	square matrix
$P^{(m)}$	القيم الاحتمالية	probability values
OLS	طريقة المربعات الصغرى الاعتيادية	Ordinary Least Squares Method
NR	نيوتن رافسون	Newton Raphson
$R\left(\underline{\beta}\right)$	احصاءة مربع كاي لبيرسون	Pearson chi square statistics
R^2	معامل التحديد	The coefficient of determination
OR	نسبة الارجحية	Odds Ratio
SE6	الخطا المعياري للوجستي المستقل	Standard error of the independent logistician
$S.E\ (\hat{\beta}_j)$	الخطأ المعياري للوجستي التوضيحي	Standard error of explanatory logistic
MSE	متوسط مربعات الخطأ	Mean square error
W	اختبار والد	Wald Test
A.E	الخوارزميات التطويرية	Algorithm Evaluation
TSR(x)	نسبة الهدف لكل كروموسوم	Target ratio for each chromosome
$f(i)$	دالة التقيم	Evaluation Function
$o.f$	دالة الهدف	objective function
$C_{(i)}$	احتمالية الفرد الكروموسوم	Chromosome individual probability
MCMC	طريقة مونت كارلو	Markov Chain Monte Carlo

المستخلص

يعد انموذج (*log-logistic*) أحد أهم نماذج الأنحدار غير الخطية الذي يستعمل في نمذجة وتقدير الكثير من التطبيقات الاحصائية ، بعدها تم تقدير معلمات هذا أنموذج بطرائق التقدير الأحصائية المعروفة و عند تقدير معلماته ظهرت لنا مشكلة وخاصة عندما يكون عدد المعلمات (P+1).وإذا تم استعمال الطرائق العددية لتقدير معلماته احيانا هذه الطرائق لا تعطي حل أفضل وذلك لأنها تعتمد بشكل مباشر على المقدرات الابتدائية للأنموذج ،وعليه سوف نقوم بأستعمال وتوظيف الطرائق الاعتيادية للتقدير بعد تحسينها من خلال أتباع أحد الخوارزميات الحديثة وهي (الخوارزمية الجينية)، بعدها نقوم بالمقارنة بين جميع طرائق التقدير . من أجل اختيار أفضل الطرق من حيث التقدير عن طريق عدد من النماذج وأحجام عينات مختلفة في المحاكاة. وبألاعتماد على المعيار الاحصائي متوسط مربعات الخطأ. وتمت المقارنة بين طرائق التقدير الأعتيادية التي تضمنت (طريقة الامكان الأعظم، وطريقة تصغير مربع كاي، وطريقة المربعات الصغرى الموزونة) وفي المقابيل طرائق التقدير المحسنة عن طريق (*Genetic Algorithm*) والتي تضمنت: -

1- طريقة تقدير الأمكان الأعظم المحسنة.
2- طريقة تقدير تصغير مربع كاي المحسنة.
3- طريقة تقدير المربعات الصغرى الموزونة المحسنة.

وبشكل عام تم التوصل الى ان طريقة المربعات الصغرى الموزونة هي الطريقة الفضلى من بين جميع الطرائق الاعتيادية في تقدير معلمات الأنموذج .أما طريقة تصغير مربع كاي كانت الأفضل من بين جميع الطرائق التقدير المحسنة لتقدير أنموذج (*log-logistic*) الثنائي والسبب في ذلك يعود ان الطريقتين تملك أقل (*MSE*) في برنامج المحاكاة لجميع المقدرات مقارنة مع باقي الطرائق .أما في الجانب التطبيقي تم أستعمال بيانات حقيقية لعينة عددها (90) مصاب بأمراض القلب وتم نمذجة البيانات وتقدير معلمات هذا الأنموذج وأختيار الطريقة الفضلى التي تم التوصل اليها في الجانب التجريبي من خلال حدوث حالات الوفاة الحقيقة للمصابين مع حالات الوفاة المقدرة ، إذ تبين مدى ملائمة أنموذج (*log-logistic*) الثنائي في نمذجة هذه البيانات وكذلك تبين ان طريقة المربعات الصغرى الموزونة و طريقة تصغير مربع كاي لهما الدقة في تقدير معلمات أنموذج(*log-logistic*) الثنائي .

الفصل الأول

المقدمة والاستعراض المرجعي

(1-1) المقدمة: *Introduction*

عند تحليل مجموعة من المتغيرات يهتم الكثير من الباحثين بطبيعة العلاقة بين هذه المتغيرات ويعد تحليل الانحدار واحد من اهم الموضوعات التي تدرس شكل العلاقة بين متغيرين أحدهما يسمى المتغير التابع والاخر يسمى مجموعة من المتغيرات التوضيحية (المفسرة) عن طريق معادلة رياضية تربط بين هذه المتغيرات وتوضح العلاقة فيما بينهم، وهذه المعادلة قد تكون خطية فتسمى معادلة الانحدار الخطي او غير خطية فتسمى معادلة الانحدار غير خطي وهو ما يحصل في اغلب الظواهر الطبيعية،

يعد انموذج الانحدار (*log-logistic*) الثنائي من النماذج اللاخطية التي تصف العلاقة بين متغير تابع ثنائي القيمة يأخذ اما الصفر او الواحد الصحيح أذ يعبر الصفر عن عدم حدوث حدث معين والواحد الصحيح يعبر عن حدوث ذلك الحدث اما باقي المتغيرات المستقلة تأخذ قيماً كمية وصفية،

تم تقسيم ال الرسالة اربعة فصول احتوى الفصل الاول على المقدمة والمشكلة وهدف البحث والاستعراض المرجعي اما الفصل الثاني فتضمن الجانب النظري لعرض انموذج الانحدار (*log-logistic*) الثنائي من حيث صيغه وبعض المفاهيم الاحصائية والاختبارات وكذلك مفهوم الخوارزمية الجينية والطرائق الكلاسيكية والاعتيادية. اما الفصل الثالث يتضمن الجانب التجريبي والتطبيقي إذ تضمن الجانب التجريبي تطبيق المحاكاة اما التطبيقي فتضمن الطريقة الفضلى من حيث التقدير للأنموذج المقدر.

اما الفصل الرابع تضمن اهم الاستنتاجات والتوصيات التي توصلنا اليها من خلال تجربة المحاكاة لهذا البحث.

(2-1) مشكلة الرسالة: *Problem Thesis*

عند أستعمال طرائق التقدير (*ML*) و (*WLS*) و (*MCS*) لتقدير معلمات انموذج الانحدار (*log-logistic*) تعاني مقدرات معلمات انموذج الانحدار اللوغاريتمي اللوجستي من التحيز وضعف الدقة عند استعمال طرائق التقدير المعروفة لكون المعادلات الانموذج هي معادلات غير خطية ولحل هذه المشكلة نلجأ الى تحويل هذه المعادلات الى خطية. تم اخذ مجموعة من العوامل المؤثرة على امراض القلب ومنها (التدخين، السكر، العمر،...) ولمعرفة ايهما اكثر تاثير على مرض القلب ويسبب حالة الوفاة للمريض .

(3-1) هدف الرسالة: *Objective Thesis*

ترمي الرسالة إلى:

1- تحسين مقدرات معلمات انموذج الانحدار اللوغاريتمي اللوجستي بما يجعها أقل تحيزاً (اكثرقُرباً من المعلمات الحقيقة).
2- بناء انموذج الانحدار اللوغاريتمي اللوجستي وتقدير معلماته لبيانات ترتبط ببيانات معلومة لمرضى القلب.
3- تقدير معلمات الانموذج (*log-logistic*) بأستعمال طرائق التقدير (طريقة الامكان الاعظم) و(طريقة المربعات الصغرى الموزونة) و(طريقة تصغيرمربع كاي) وبأستعمال طريقة (*Genetic Algorithm*) وطريقة (*Newton Raphson*) لتحسين النتائج والمقارنة بين هذه الطرائق باستعمال معيار (*Wald Test*) واختبار هوزمر-ليمشولاختبار جودة المطابقة.

(4-1) الاستعراض المرجعي *Review of Literature*

دخلت الخوارزمية الجينية في مجالات متعددة واستعملت لتحسين الحل، وفي هذه الرسالة تم استعمال الخوارزمية الجينية لتقدير معلمات الانحدار (*log-logistic*) الثنائي وفيما يأتي بعض الدراسات السابقة التي تخص هذا الموضوع:

- **في عام (2002) قدم (*Czepiel*)** نظرة عامة عن انموذج الانحدار اللوجستي للمتغيرات التابعة التي تحتوي على مستويين أو أكثر من المستويات الفئوية المنفصلة وتم اشتقاق معادلات الغير خطية بطريقة دالة الامكان الاعظم باتباع خوارزمية نيوتن رافسون لأنظمة المعادلات غير الخطية وناقش تنفيذ اسلوب الخوارزمية وتوصل الى انها تعطي نتائج افضل في الاستعمال من تقدير المربعات الصغرى لأنها غير قادرة على انتاج الحد الأدنى من تقديرات التباين غير المتحيزة للمعلمات الفعلية.[7]

- **في عام (2005) اقترح (pasiaواخرون)** وآخرون مقارنة النتائج المحاكاة بتطبيق الخوارزميات الجينية في تقديرات معلمات الانموذج اللوجستي مع انموذج (*Gauss*) غير الخطي بالطرائق العددية الكلاسيكية وأثبتوا افضلية الخوارزميات الجينية كبديل جيد عن الأسلوب العددي الأمثل عند تقديرمعلمات انموذج الانحدار اللوجستي عن طريق هذه الدراسة.[34]

- **وفي العام نفسه** اثبتـت النتـــــــــــــــائج التــي حصــــــــــــــل عليهـا (*Groenwald and Mokgatlhe*) من استعمـــــال محاكـاة معاينــــــــــة جـبس (*Gibbs Sampler*) عند تشكيل التوزيعـــات اللاحقـــة لأنموذج الانحدار اللوجستي التـي يسـتند عملهـا الى مبـدأ التكـرارات ان جميع التوزيعـات الشـرطية تتـوزع (*Uniform*) وان بالإمكان تطبيق معاينة جبس على المتغيرات الاســمية والترتيبية [14].

- **في عام 2007 قدم الباحث (*Mccarthy*)** عدة طرائق لتقدير معلمات انموذج الانحدار اللوجستي الثنائي لمجموعة من البيانات الوصفية بوجود فئات من نقاط بيانات طريقة الفصـل التام وشـبه التام والتداخل فضــلا عن وجود مشـكلة التعدد الخطي وتوصـل الباحث الى ان الإمكان الأعظم في حالة التداخل تكون غير صـحيحة وغير دقيقة في حالة وجود مشكلة الفصل التام وغير التام.[29]

- **وفي عام 2008 قام (*Vandewater*) وآخرون** باستعمال انموذج الانحدار اللوجستي (LR) بتطبيقه على الخوارزمية الجينية (*GA*) لاختيار المؤشـرات الحيوية المقاسـة من عينات الدم للتنبؤ بتطور مرض الزهايـــــــــــمر ومن النتـــــــــائج تم استكشاف فائدة (*GA*) في اختيار الانموذج الجيد من حيث تحســين المعلمات للتنبؤ بدقة عالية في تشخيص المرض.[42]

- **وفي العام نفســه قام الباحث (*Gayou*) وآخرون** باختبار الخوارزمية الجينية (*GA*) على مجموعة من البيانات التي حققت في حدوث إصـــابة الرئة بالســرطان بتحديد المتغير كأفضــل متنبئ للإشــعاع لعلاج الالتهاب الرئوي المســتحدث في مرضــى

سرطان الرئة ومن خلال النتائج تم التأكد انه أفضل انموذج لتحليل جميع التوليفات الممكنة لعوامل الانحدار اللوجستي (*LR*) في تحليل وتوقع نتائج معالجة العلاج الإشعاعي.[10]

❖ **وفي عام 2009 قامت (*Yoder*)** بأنشاء أنموذج للانحدار اللوجستي (*LR*) باستعمال المنهجية الخاصة بالخوارزمية الجينية (*GA*) ولكن وقد واجه هذا الانموذج بعض المشاكل في الاختبار منها طريقة الاختيار لبعض المتغيرات التوضيحية والتي تسمح بالتحيز لغرض الدخول بداخل هذا الأنموذج وباستعمال تقنية التحسين الحالية ستعالج وبشكل فعال هذه المشاكل وتبين ذلك عن طريق النتائج التي حصلت عليها من البيانات الخاصة بالحوادث الجوية مع (500) حالة و(13) من المتغيرات التوضيحية.[44]

❖ **وفي عام 2010 قدم الباحثان (*Meng and Weng*)** المنهجية الخوارزمية الجينية (*GA*) لتحليل الانحدار اللوجستي(*LR*) لتقييم مخاطر مكان العمل وأظهرت النتائج أن منهجية الخوارزمية الجينية (*GA*) تكون أفضلاً لانموذج للانحدار اللوجستي من الاداء الثنائي.[32]

❖ **في عام 2014 قام (*Johnson*) وآخرون** باستعمال الانحدار اللوجستي (*LR*) والخوارزمية الجينية (*GA*) في تشخيص مرض الزهايمرو اثبتوا أن الخوارزمية الجينية تعطي نتائج ناجحة إلى حد كبير.[20]

❖ **في عام 2015 قام الباحثان (*Demir and Akkus*)** بمقارنة الخوارزمية الجينية (*GA*) مع خوارزمية نيوتن رافسون (NR) في كيفية تنفيذهما في تقدير معلمة انموذج اللوجت الثنائي على بيانات مرض الثعلبة وتوصلا الى أن خوارزمية (GA) تكون اكثر مرونة في تحسينها للحل وتفوق خوارزمية (NR) في تقدير المعلمة المجهولة بالوصول الى الحل الامثل.[8]

- **في العام 2016 قام الباحث (*Mahmoud*) وآخرون** باشـــتقاق تقديرات معلمات انموذج مقترح وهو إنموذج الانحدار اللوجستي لوجيستي بيتا المستند إلى التوزيع اللوجيستي التجريبي بيتا والذي يستعمل على نطاق أوسع في التطبيقات، وطُبق على الانموذج المقترح على مجموعة بيانات حقيقية. وتمت فحوصات الانموذج بناءً على بقايا مارتينجال وإحصاءات AIC وBIC لاقتراح النماذج المناسبة.[26]

- **في عام 2021 قام الباحث (*Al-Bayati and Muhammad*) وآخرون** بأستعمال طريقة الإمكان الأعظم وطريقة انحدار الحرف في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة باعتماد أسلوب الجاكنايف واجراء المقارنة بين المقدرات وفق معيار المعلومات (AIC). وتم تطبيق اسلوب الجاكنايف والمقدرات الاحصائية المذكورة لدراسة العلاقة بين متغير الاستجابة (اصابة وعدم الاصابة بسرطان الثدي) لعينة حجمها (100) عينة لعام (2020) وبين المتغيرات التوضيحية (نسبة الهيموغلوبين الموجودة في الكريات الحمراء في الدم، كريات الدم الحمراء، كريات الدم البيضاء، الصفيحات الدموية ، نسبة الهيموغلوبين الموجودة في الدم ، نسبة اللمفاويات في الدم ، نسبة كريات الدم الوحيدة ، نسبة كريات الحامضية ، نسبة كريات القاعدية) واتضح عن طريق المقارنة ان طريقة انحدار الحرف في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة هي الفضلى في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي.[4]

تتميز هذه الدراسة عن الدراسات السابقة باجراء مقارنة إذ تم أستعمال الخوارزمية الجينية وخوارزمية نيوتن رافسون لتحسين الطرائق الكلاسيكية الاعتيادية المتمثلة بـ (طريقة الإمكان الأعظم وطريقة المربعات الصغرى الموزونة وطريقة تصغير مربع كاي) لتقدير معلمات انحدار(Log-$Logistic$) الثنائي واختيار الفضلى من بينهم. وتم أستعمال بيانات حقيقية ونمذجة هذ البيانات وتقدير معلمات هذا الأنموذج وأختيار الطريقة الفضلى التي تم التوصل اليها في الجانب التجريبي إذ تبين مدى ملائمة أنموذج (log-$logistic$) الثنائي في نمذجة هذه البيانات وكذلك تبين ان طريقة المربعات الصغرى الموزونة وطريقة تصغير مربع كاي المحسنة لهما الدقة في تقدير معلمات أنموذج(log-$logistic$) الثنائي.

الفصل الثاني

الجانب النظري

1-2 تمهيد: *preamble*

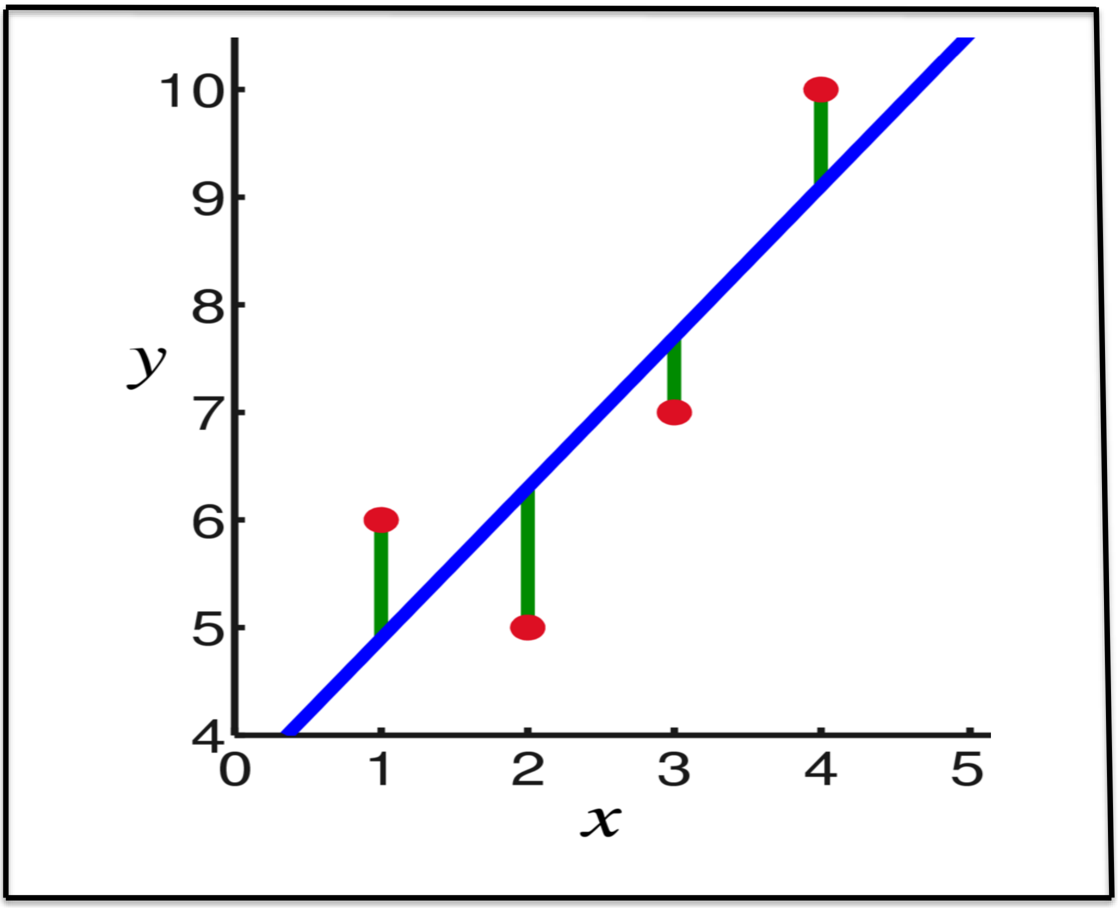
يعرف الانحدار بانه طريقة احصائية تستعمل التقدير والتنبؤ و التمويل والاستثمار والتخصصات الاخرى وكذلك يقيس مقدار التأثير بين متغير تابع واحد (يشار إليه عادة بواسطة Y) (*Dependent Variable*) وسلسلة من متغيرات الاخرى تسمى المتغيرات المفسرة (المستقلة) (*Independent Variable*) يرمز له ب(X). وبذلك يتم التقدير او التنبؤ بمتوسط قيمة المتغير التابع بمعلومية المتغيرات المستقلة . بناء على ذلك فان أسلوب الانحدار يستعمل للتوصل الى أنموذج رياضي يوضح العلاقة الكمية بين المتغير التابع والمتغيرات المستقلة ، مثال على ذلك أرتفاع ضغط دم شخص متمثل ب(Y) و عمر الشخص نفسه ب(X) هذا الارتباط و التابعية بين (X) و(Y) هي ما ندعوه بالانحدار .ينقسم الانحدار بصورة عامة الى الانحدار الخطي والذي يكون على نوعين الانحدار الخطي بسيط والانحدار الخطي متعدد، والقسم الثاني هو الانحدار اللاخطي.في هذا الفصل سوف يتم التطرق الى الانحدار اللاخطي متمثلاً بانموذج الانحدار(*log-logistic*) من حيث الافتراضات الخاصة به والصيغة العامة وكذلك خصائصه وأنواعه . ودراسة الطرائق الكلاسيكية الاعتيادية لتقدير معلماته المتمثلة بطريقة المربعات الصغرى الموزونة (*weighted least squares method*) وطريقة الإمكان الأعظم (*Maximum Likelihood Method*) وكذلك طريقة تصغير مربع كاي (*Minimum Chi - Square Method*). وسوف يتم تحسين هذه الطرائق بأستعمال الخوارزمية الجينية (*Genetic Algorithm*) أذ ان الخوارزمية الجينية سوف تعمل هنا (*Optimization*) لتقدير معلمات إنموذج الانحدار (*log-logistic*) بأستعمال الطرائق المذكوره آنفاً.

2-2 إنموذج الانحدار الخطي (*linear regression model*)

هو أنموج إحصائي يستعمل في تفسير متغير Y بمتغير أخر X او مجموعة من المتغيرات $X_1, X_2,\ldots, X_p$ () وفق دالة خطية.

يسمى المتغير Y بالتابع والمتغيرات X بالمتغيرات المستقلة او المفسرة ، بمعنى أنها تفسر إحصائيا تغير المتغير التابع .وتقدير الانموذج يعني بالأساس تقدير القيم المقدرة للمعاملات (β) والتي تعد معالم الانموذج ، يعتمد نجاح الانموذج وقابلية تطبيقه

واستغلاله عمليا لا سيما في توقع قيم مستقبلية (أو جديدة) ل Y بمعرفة قيم X .[12] توجد اربع طرائق لتقدير أنموذج الانحدار الخطي (طريقة التقدير حسب القيمة العليا لدالة الامكان و طريقة المربعات الصغرى و طريقة العزوم و الطرائق البايزية) وتُعَّد طريقة المربعات الصغرى الاكثر شيوعاً وتعتبر افضل تقدير خطي غير متحيز (*Best Linear Unbiased Estimator*) ويشار إليه اختصارا ب BLUE .[26]

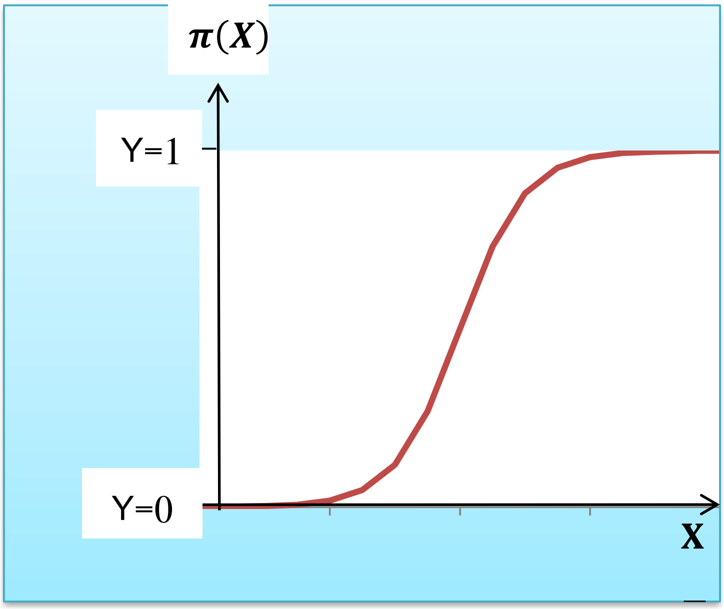


الشكل (2 – 1) يبين الانحدار الخطي بين المتغير التابع (y) والمتغير المستقل (x)

في نموذج الانحدار الخطي يفترض أن النقط الحمراء، هي نتيجة لانحرافات عشوائية، موضحة بالفوارق الخضراء عن الخط الأزرق، الذي ينمذج وفق دالة خطية المتغير التابع (y) بدلالة المتغير المستقل (x).

3-2 إنموذج الانحدار اللاخطي (*Nonlinear regression model*)

هو احد اشكال تحليل الانحدار و يتم نمذجة البيانات عن طريق توظيف مجموعة من المعلمات غير الخطية وتعتمد على واحد او اكثر من المتغيرات المستقلة . ويعد انموذج اللاخطي من النماذج واسعة الاستعمال والمهمة في العديد من المجالات ويكون ذات تطبيقات واسعة في الدراسات التطبيقية والطبيعية على الرغم من قلة الدراسات فيه مقارنة بالانموذج الخطي ، وهناك عدة انواع من النماذج اللاخطية ومن هذه النماذج هو انموذج الانحدار اللوجستي و اللوجستي اللوغاريتمي وانموذج التكعيبي والاسي وانموذج التربيعي ...الخ التي تهتم بدراسة وتحليل هذه البيانات .



الشكل (2 – 2) العلاقة غير الخطية بين متغير الاستجابة$\pi(X)$ والمتغير التوضيحي(X)

4-2 الثنائي لوجستي (*Binary logistic*)

هذا الانموذج من نماذج الانحدار اللاخطي إذ أن متغير الاستجابة (Y) يتبع توزيع ثنائي (توزيع برنولي) (*Bernoulli distribution*) فتكون القيم التي يأخذها ثنائية هي (0) و (1) .

إذ ان متغير الاســـتجابة الفئوي (Y) يأخذ حالتين إذ تتمثل الحالة الاولى عند وقوع حدث معين وعندما تكون (Y=1) والحالة الثانية هي عدم وقوع هذا الحدث عندما (Y= 0) ويكون احتمال وقوع الحدث (النجاح) هو { $\pi(X_i)$ } بالاعتماد على قيم المتغيرات التوضـــيحية للمشـــاهدات ويكون احتمال عدم وقوع الحدث (الفشـــل) هو$[1-\pi(X_i)]$ وبذلك تكون دالة الكثافة الاحتمالية بالصيغة الاتية [12]:

$$P(Y_i \backslash X_i) = [\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i} \qquad \ldots(1-2)$$

$$Y_i = 0,1$$

$$P(Y=1 \backslash X_i) = \pi(X_i) \qquad \ldots(2-2)$$

$$P(Y=0 \backslash X_i)=1-\pi(X_i) \qquad \ldots(3-2)$$

وان:

$\pi(X_i)$: يمثل احتمال النجاح (success)

$1-\pi(X_i)$: يمثل احتمال الفشل (failure)

وان توقع متغير الاستجابة يمثل احتمال النجاح

$$\mathrm{E}(Y_i)=\pi(X_i)$$

وايضا تباين متغير الاستجابة يمثل احتمال حاصل ضرب النجاح مع احتمال الفشل

$$\mathrm{V}(Y_i)=\pi(X_i)(1-\pi(X_i))$$

يمكن التعبير عن انموذج الانحدار اللوجستي في حالة احتوائه على متغير توضيحي واحد (X) بالصيغة الرياضية الاتية:

$$\pi(X_1)=\frac{1}{1+e^{-(\beta_\circ+\beta_1 X_1)}} = \frac{e^{(\beta_\circ+\beta_1 X_1)}}{1+e^{(\beta_\circ+\beta_1 X_1)}} \qquad \ldots(4-2)$$

$$1-\pi(X_1)=\frac{1}{1+e^{(\beta_\circ+\beta_1 X_1)}} \qquad \ldots(5-2)$$

اذ ان :

$\pi(X_1)$: هو التوقع الشرطي $E(Y/X)$ (conditional mean) لمتغير الاستجابة (Y) عند قيمة معينة لـ (X) (احتمال حدوث الاستجابة).

$1-\pi(X_1)$: هو احتمال عدم حدوث الاستجابة.

$\beta_\circ, \beta_1$: هما معلمتا انموذج الانحدار اللوجستي.

اما اذا كان انموذج الانحدارلوك اللوجستي يحتوي على اكثر من متغير توضيحي واحد، فيعبر عن توقعه الشرطي لاحتمال متغير الاستجابة (وقوع الحدث) حسب الصيغة الرياضية الاتية:[26]

$$\pi(X_i)=\frac{1}{1+e^{-(\beta_\circ+\beta_1 X_{i1+\cdots+}\beta_p X_{ip})}} = \frac{e^{(\beta_\circ+\beta_1 X_{i1+\cdots+}\beta_p X_{ip})}}{1+e^{(\beta_\circ+\beta_1 X_{i1+\cdots+}\beta_p X_{ip})}} \qquad \ldots(6-2)$$

وان احتمال عدم وقوع الحدث هو:

$$1-\pi(X_i)=\frac{1}{1+e^{(\beta_\circ+\beta_1 X_{i1+\cdots+}\beta_p X_{ip})}} \qquad \ldots(7-2)$$

وان:

$X_1, X_2, \dots, X_p$: المتغيرات التوضيحية التي تكون المصفوفة $\mathrm{X} = \mathrm{X_{ij}}$ بعد $(\mathrm{n} * \mathrm{p})$

$i = 1,2, \dots, n$; n:عدد المشاهدات.

$j = 0,1, \dots, p$; p:عدد المتغيرات التوضيحية والمعلمات المجهولة.

$\beta_\circ, \beta_1, \dots, \beta_p$: متجه للمعلمات المراد تقديرها.

ويتم تحويل هذا الانموذج الى شكل خطي يتمثل بعلاقة خطية عن طريق المتجه الصفي $(\acute{X_i})$ من المتغيرات التوضيحية مع لوجت الاحتمال $[logit\ \pi(X_i)]$ وحسب الصيغة الرياضية الاتية:

$$Z_i = logit\ \pi(X_i) = ln\left[\frac{\pi(X_i)}{1-\pi(X_i)}\right] = \beta_\circ + \beta_1 X_{i1} + \dots + \beta_p X_{ip}$$

$$= [1 \quad X_{i1} \quad \dots \quad X_{ip}] \begin{bmatrix} \beta_\circ \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \underline{X'_i \beta} \qquad \dots(8-2)$$

$$\therefore \quad Z_i = \underline{X'_i \beta} + \varepsilon_i \qquad \dots(9-2)$$

إذ ان متغير الاستجابة في الانحدار الخطي يفترض: $Y_i = E(Y/X) + \varepsilon_i$

$$Y_i = \pi(X_i) + \varepsilon_i \quad \rightarrow \quad Y_i = \frac{e^{(\beta_\circ + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}}{1+e^{(\beta_\circ + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}} + \varepsilon_i$$

وان التوزيع الشرطي يكون طبيعياً بمتوسط$E(Y/X)$ وبتباين ثابت وان $\varepsilon_i = Y_i - E(Y/X)$ وعند تعويض قيمة $Y = 1$ ينتج $_i = 1 - \pi(X_i)$ وعندما $Y = 0$ ينتج $\varepsilon_i = -\pi(X_i)$ وبذلك ان حد الخطأ (ε_i) في هذه الحالة يتبع توزيع ثنائي الحدين يتوزع بمتوسط مقداره$\mathrm{n\pi(X_i)}$ وتباين مقداره $[\mathrm{n\pi(X_i)(1-\pi(X_i))}]$ وهذا يدل ان التباين لحد الخطأ يكون غير متجانس . يتم تعويض المعادلة (8−2) في المعادلتين (6−2) و (7−2) فيكون انموذج الانحدار اللوجستي(انموذج الدالة اللوجستية) كالاتي:

$$(X_i) = \frac{e^{\underline{X'_i \beta}}}{1+e^{\underline{X'_i \beta}}} \qquad \dots(10-2)$$

$$1 - \pi(X_i) = \frac{1}{1+e^{\underline{X'_i \beta}}} \qquad \dots(11-2)$$

2–5 إنموذج الانحدار (*Log-Logistic*) الثنائي

هو انموذج احصائي يستعمل للبيانات الفئوية (*Categorical data*) وينتمي الى نماذج الانحدار اللاخطية ويعد أحد الاساليب الاحصائية المهمة التي تستعمل للتنبؤ باحتمالية وقوع حدث معين وذلك عن طريق ترتيب البيانات والمعلومات على المنحى اللوجستي . يعمل انحدار (*log-logistic*) لوصف العلاقة بين متغير الاستجابة والمتغيرات التوضيحية [1].

يُعد أنموذج الانحدار (*Log-Logistic*) حالة خاصة من نماذج الانحدار الاعتيادية بسبب طبيعته الاسمية التي يحملها متغير الاستجابة مثل (نجاح ، فشل) (موفق ، غير موفق) ويستعمل الانحدار لوك اللوجستي بشكل كبير في كثير من التطبيقات الاحصائية وتحليل البيانات في معظم الدراسات الطبية والهندسية والاقتصادية والغاية منه هو تقدير الانموذج الذي يمثل العلاقة غير الخطية بين المتغيرات لكي يتم تطبيقها في التنبؤ الاحصائي كالتنبؤ بوقوع حدث معين او عدم وقوعه، ويكون على عدة انواع .الانحدار اللوجستي الثنائي (*Binary Logistic Regression*) ويستعمل هذا النوع عندما يأخذ متغير الاستجابة قيمتين هما (1 ، 0) . و الانحدار اللوجستي المتعدد الحدود (*Multinomial Logistic Regression*) يتم استعماله عندما يكون متغير الاستجابة متعدد القيم (اكثر من قيمتين) وكذلك الانحدار اللوجستي الترتيبي (*Ordinal Logistic Regression*) وفيه تكون متغيرات الاستجابة رتبوية . في دراستنا سوف نركز ونقتصر في العرض النظري والتطبيقي على الانحدار (*Binary Log-Logistic Regression*) الثنائي .[5].

2-5-1 شروط (*Log-Logistic*) الثنائي:

دائما ما يرتكز انموذج الانحدار لوك اللوجستي الثنائي على عدد معين من النقاط التي تعد شروطاً أساسية يجب توفيرها قبل أستعمال الانموذج والتطبيق عليه وكما يأتي [44]:

1- يجب إنَّ تكون هنالك علاقة بين المتغيرات المستقلة النسبية والتحويل اللوغاريتمي للمتغير التابع ويظهر هذا الشرط عند تطبيق أنموذج الانحدار لوك اللوجستي.

2- يجب ان يكون المتغير التابع متغيراً اسمياً ثنائياً بالنسبة لصنف الانحدار اللوجستي الثنائي الاستجابة .

3- يتم تحديد القيم الشاذه عن طريق اختبار (*test mahala Nobis*) (اختبار المسافة) التي توجد في برنامج (*SPSS*) ، إذ يجب ان لا تكون هناك قيم شاذة (*outliers*) في المتغيرات المستقلة .

4- يسلك متغير الاستجابة سلوكاً احتمالياً واحداً لكل مشاهدة واحدة اي بما معنى يجب ان لا تكون هناك فئيتين للمشاهدة نفسها في الوقت نفسه .

5- يحتوي انموذج الانحدار اللوجستي االثنائي على اكثر من متغير مستقل في جميع المستويات .

6- في الانحدار (*Log-Logistic*) الثنائي يوجد متغير ثنائي تابع واحد ، يتم ترميزه بواسطة متغير مؤشر واحد إذ يتم تسمية القيمتين (0 ، 1) بينما يمكن ان تكون كل المتغيرات المستقلة متغيراً ثنائياً (فئتان ، مشفرة بواسطة متغير مؤشر) أو متغيراً مستمراً (اي قيمة حقيقة) .

2-5-2 أفتراضات (*log-logistic*) الثنائي:

1- يجب ان لا تكون المتغيرات التوضيحية مرتبطة فيما بينها ارتباطاً قوياً.
2- الخطا لا يتوزع توزيعا طبيعيا و التباين يكون غير متجانس.
3- يكون حجم العينة كبيراً لذلك ينبغي ان لا يقل عن خمس اضعاف عدد المتغيرات .
4- يتعامل انموذج الانحدار (*log-logistic*) مع جميع العلاقات الخطية وغير الخطية التي تكون بين المتغيرات التوضيحية و متغير الاستجابة ، اي لا يشترط وجود علاقة خطية بينهما .
5- يعامل البيانات الاسمية والترتيبية على أنها متغيرات مستقلة (توضيحية)
6- يفترض ان $P\ (Y=1)$ هو احتمالية وقوع الحدث لذلك يتم اعطاء رمز لمتغير الاستجابة وفقأ لذلك[37] .

2-5-3 مجالات التي يستعمل بها (*log-logistic*) الثنائي .

1- يستعمل في العديد من المجالات إذ يستعمل في المجال الصحي ويشاع استعماله وتوظيفه بشكل أوسع في الطب والعلوم الاجتماعية والاقتصادية كذلك .
2- يهدف الانحدارلوك اللوجستي الى التنبؤ وشرح قيم متغيرات كيفية، حيث يعتبر أفضل أسلوب احصائي في الدراسات التي تهدف لبناء نماذج تنبؤيةَ .
3- تستعمل المتغيرات الثنائية اللوجستية على نطاق واسع في الاحصائيات لنمذجة احتمالية حدوث فئة معينة او حدث معين ، مثلا احتمال فوز فريق ما ، او ان يكون المريض بصحة جيدة .
4- انموذج الانحدار لوك اللوجستي نفسه ببساطة يصوغ احتمالية المخرجات من حيث المدخلات ولا يقوم بتصنيف إحصائي (إنه ليس مصنفًا) ، على الرغم من أنه يمكن استعماله لإنشاء مصنف ، على سبيل المثال عن طريق اختيار قيمة القطع وتصنيف المدخلات مع الاحتمالية أكبر من القطع كفئة واحدة ، أسفل القطع مثل الأخرى ؛ هذه طريقة شائعة لعمل مصنف ثنائي.[34]

2-6 إنموذج الانحدار المتعدد (*Multiple regression model*)

يعد أنموذج الانحدار المتعدد من الأساليب الإحصائية الحديثة التي تضمن اختيار ودقة الإستدلال من اجل تحسين نتائج البحث عن طريق الاستعمال الامثل وتحقيق افضل النتائج . و الانحدار الخطي المتعدد هو عبارة عن أيجاد معادلة رياضية تعبر عن العلاقة بين اكثر من متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ بقيم مستقبلية ، وهو عبارة ايضا عن انحدار للمتغير التابع (Y) على العديد من المتغيرات المستقلة (X_1،X_2،...،X_n) . لذا فهو يستعمل في التنبؤ بتغيرات المتغير التابع التي يؤثر فيه عدة متغيرات مستقلة .[12]

ويعد انموذج الانحدار المتعدد الاستجابة من النماذج الاحصائية ذات الاهمية الكبيرة في تحليل البيانات وتصنيفها ،إذ يستعمل في حالة كان متغير الاستجابة يصنف او يعود للمتغيرات التي تكون من نوع الرتبي (*Ordinal*) او من نوع الاسمي (*Nominal*) والتي تتكون من تصنفين او مستويين فأكثر .

يُستعمل الانحدار الخطي المتعدد لشرح العلاقة بين متغير تابع مستمر ومتغيرين مستقلين او اكثر . يمكن ان تكون المتغيرات المستقلة مستمرة او متقطعة فهي بالاساس تعتمد فكرته على العلاقات الدالية والذي يستعمل بنمط يعرف بالتشتت او الانتشار مثلأ يمكننا التنبؤ بمرض سرطان الثدي بالاعتماد على دراسة حالات اخرى للمرضى مثل عمر الشخص والمواصفات الجسمانية والامراض التي يعاني منها الوالدين وغيرها .

والمعادلة الخطية في الانحدار المتعدد هي

$$Y=\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n+e$$

إذ ان :

Y : المتغير التابع

α : *قيمة ثابتة* Constant

β_1 : *معامل الانحدار على المتغير المستقل الاول*

β_2: *معامل الانحدار على المتغير المستقل الثاني*

β_n: *معامل الانحدار على المتغير المستقل i*

X_1 : المتغير المستقل الاول

X_2 : المتغير المستقل الثاني

X_n: المتغير المستقل i

ويُعد انموذج الانحدار المتعدد الاستجابة امتدادا بسيطا لانموذج الانحدار الثنائي الاستجابة إذ يعتمد هذا الانموذج بشكل اساسي على التوزيع المتعدد الحدود (*polynomial distribution*) كما يأتي :

$$P_r(r، r_{i2}، \; . \; . \; . \; . \; . \; ، r_{ij}) = \binom{Ni}{r_{ij} \text{،......،} r_{iJ}} \pi_{i1}^{r_{i1}} \; . \; . \; . \pi_{iJ}^{r_{iJ}}$$

إذ أنَّ : $\sum_{i=1}^{k} \pi_i = 1 \quad ، \quad \sum_{i=1}^{m} r_{ij} = Ni$

2-7 الدالة الوجستية (*Logistic function*)

دائما ما يعرف انموذج الانحدار اللوجســـتي الثنائي على انه أحد نماذج الانحدار اللاخطية إذ يبنى على فروض اساسية يكون فيها متغير الاستجابة الذي نهتم بدراسته يتبع توزيع برنولي (*Bernoulli*) ، وعليه يكون أحتمال النجاح π_i عندما (y=1) واحتمال فشـــل $1-\pi_i$ عندما (y=0) بذلك تكون صـــيغة دالة الكثافة الاحتمالية تكتب كالآتي:[25]

$$P(Y_i/X_i) = [\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i} \quad ... \qquad (12\text{-}2)$$

وعند تعويض قيم $y_i = 0$،1، سيكون P(Y) فان

$$P(Y) = \left\langle \begin{array}{ll} \pi_i & \text{عند حدوث الاستجابة} \; Y = 1 \; \text{عندما} \\ 1-\pi_i & \text{عند عدم حدوث الاستجابة} \; Y = 0 \; \text{عندما} \end{array} \right\rangle \quad ..(13\text{-}2)$$

π_i: يمثل احتمال النجاح (Success)

$1-\pi_i$: يمثل احتمال الفشل (Failure)

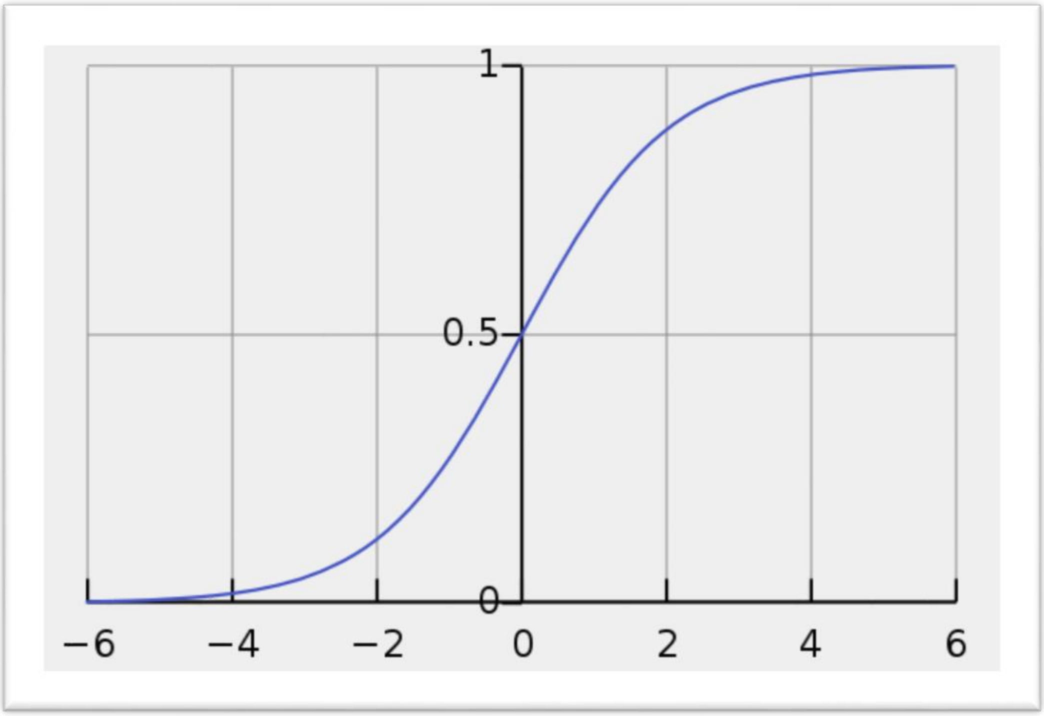
وعندما $y_i = 1$ نحصل على

$$\pi i = \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}} \quad . \; . \; . \; . \; . \; . \; . \qquad (14-2)$$

وعندما $y_i = 0$ نحصل على :

$$1-\pi i \; = 1 - \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

$$1-\pi i \; = \frac{1}{1 + e^{B_0 + B_1 X_1}} \qquad \ldots\ldots\ldots (15-2)$$



الشكل (2-3) يبين منحنى الدالة اللوجستية

2- 8 إنموذج اللوجت (*logit model*)

يوجد العديد من الدوال والتحويلات التي يمكن استعمالها لتحويل انموذج من الانحدار اللاخطي الى الانحدار الخطي ، الغرض منه والغاية الاساسية من اجراء هكذا تحويلات هي نقوم بتغير شكل التوزيع التكراري لكي يلائم الافتراضات النظرية في انموذج الانحدارلوك اللوجستي ، وهناك العديد من التحويلات التي يمكن استعمالها ، سوف نقوم باستعمال أنموذج (*logit*) لتحويل انموذج اللوجستي الى معادلة خطيه يمكن استعمالها والحصول على قيم المعلمات ومتغيراته . ويعرف بانه في الانحدار اللوجستي هو حالة خاصة لوظيفة الارتباط في نموذج خطي معمم (انها وظيفة الارتباط المتعارف عليها لتوزيع برنولي)[12] . اذن هو عبارة عن التحويل الخطي لدالة

الانحدار اللوجستي (دالة logit) إذ يستعمل لإزالة الانحناءات الموجودة بالأنموذج اللوجستي لتاثيرها السلبي على مقدرات المعلمات عن طريق اخذ اللوغارتم لمعاملات المفاضلة لتحويل العلاقة بين احتمالية حدوث الاستجابة والمتغيرات التوضيحية الى علاقة خطية بتطبيق أنموذج الانحدار الخطي ضمن المجال { ∞,∞- } إذ تأخذ الدالة logit القيم بين اللانهاية السالبة والموجبة. وحسب الصيغة الرياضية الاتية [9]:

$$logit\ (\pi_i) = ln \frac{\pi_i}{1-\pi_i} \qquad \ldots (16-2)$$

$$= \ln(odds)$$

$$= ln(e^{\beta_\circ + \beta_1 X_1 + \cdots + \beta_p X_p})$$

$$= ln(e^{\underline{X_i'\beta}})$$

$$= \underline{X_i'\beta}$$

$$= Z_i \qquad \ldots (17-2)$$

وعليه ان التحويل الخطي لدالة الانحدار اللوجستي(z_i) يتوزع توزيعا طبيعيا تقاربيا (*Asymptotically Normal distribution*) اي ان:

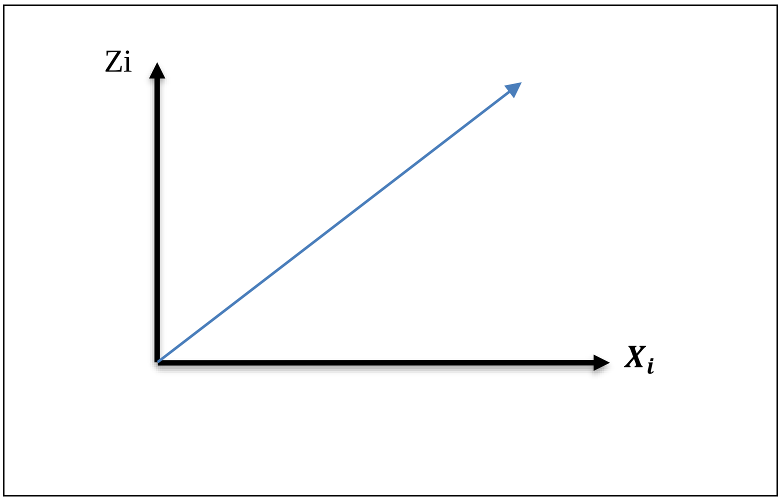
$$Z_i \sim Asym.N.\ \{X_i'\beta\ ,\ \ [\pi_i\ (1-\pi_i)]^{-1}\ \} \qquad \ldots (18-2)$$

2- 9 *طرائق تقدير معلمات إنموذج (Log-Logistic) الثنائي :*

يمكن ان تعرف عملية التقدير للمعلمات على انها عملية استخراج مقدر للمعالم المجهولة للمجتمع عن طريق معلومات العينة التي تكون متوفرة لدينا ، في بعض الاحيان تكون معظم الظواهر ليس بالإمكان دراستها بشكل شامل ، لذلك توجد هناك امكانية لدراسة سلوك الظاهرة وفق توزيع احتمالي معين ، إذ يحتوي هذا التوزيع على معلمات مجهولة بحاجة الى تقديرها بإحدى طرائق التقدير المعروفة لكي نتمكن من التعرف على خصائص الظاهرة المدروسة عن طريق التوزيع الاحتمالي . ومن هذه الطرائق طريقة الامكان الاعظم (*Maximum Likelihood Method (MLM)*)

وطريقة المربعات الصغرى الموزونة (*Weighted Least Squares (WLS)*) وطريقة تصغير مربع كاي (*Minimum Chi - Square Method (MCSM)*) اما طريقة المربعات الصغرى الاعتيادية (OLS) *Ordinary Least Squares Method* لا يمكن استعمالها في تقدير معلمات الانموذج اللوجستي لعدم تحقق احد شروطها وهو عدم تجانس تباين الخطأ وتوجد طريقة ابتدائية ايضاً للتقدير وهي طريقة الرسم المعروفة

في تحليل الانحدار (graphical method) و يتم رسم خط الانحدار بين قيم (Z_i) ولوغاريتم المتغير التوضيحي (X_i) فينتج رسم مقارب للخط المستقيم (انموذج اللوجت) نحسب منه قيم (β) التقـــــديرية كما في الشكل $(2-4)$ الآَتي[38] :



الشكل (2-4) العلاقة الخطية بين X_i و Z_i

2-9-1 طريقة تقدير الامكان الاعظم

تعد طريقة الامكان الاعظم من الطرائق الكلاسيكية واسعة الاستعمال في تقدير معلمات الانموذج الاحصائية و الرياضية ، كونها تتميز بعدة خصائص منها[40] :

1- الاتساق (*consistency*)
2- أقل تباين (*Minimum Variance*)
3- الكفاءة (*Efficiency*)
4- الثبات (*Invariance*)

وعندما يزداد حجم العينة تكون اكثر دقة ، لذا تعتمد في تقديرها على تكرار العمليات الحسابية عدة مرات اي انها طريقة تكرارية كي يتم الوصول فيها الى افضل الطرق لذلك تقوم هذه الطريقة على مبدأ ايجاد مقدرات للمعلمات عن طريق جعل دالة الامكان الاعظم في نهايتها العظمى ، وعليه فان أنموذج الانحدار اللوجستي الثنائي يكون فيه متغير الاستجابة (Y_i) الثنائي والذي يعد أنموذج اللوجستك احدها يتوزع كما ذكرنا

حسب توزيع برنولي اي ان له مستويان هما الصفر والواحد حسب الصيغة $(1-2)$ وكما يأتي :

$$P(Y_i/X_i) = [\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i}$$

وللحصول على تقديرات المعلمة بواسطة الامكان الاعظم (MLE) يتم ضرب الحدود في الصيغة $(1-2)$ لعينة حجمها (n) تعتمد على (X)من المتغيرات التوضيحية (التفسيرية) ومجموعة متغير الاستجابة (Y) حسب الصيغة الرياضية الآتية[25]:

$$P(Y/X) = \prod_{i=1}^{n}[\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i} \qquad \dots (19-2)$$

وان $\pi(X_i)$ و $P(Y/X)$ تعتمد على المعلمات وهدفنا هو تقدير المعلمات غير المعلومة فبإمكاننا تحديد دالة الإمكان $l(\beta)$ للكشف عن التبعية.

$$l(\beta) = P(Y/X)$$

وان $l(\beta)$: يمثل الامكان (الاحتمال) لمشاهدات العينة Y.

والغرض من (MLE) هو ايجاد التقدير الأفضل لمتجه $(\hat{\beta})$ الذي يزيد من احتمال المشاهدات (Y) وحسب ما يأتي:

$$l(\hat{\beta}) = \max_{\beta} l(\hat{\beta})$$

وبالتالي ان دالة الامكان الاعظم في الانموذج اللوجستي الثنائي هي:

$$l(\beta) = \prod_{i=1}^{n}[\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i}$$

وبأخذ اللوغارتم (Log) على الطرفين للحصول على (MLE) لتسهيل عملية الحل

$$ln[l(\beta)] = \sum_{i=1}^{n} Y_i ln[\pi(X_i)] + (1-Y_i)ln[1-\pi(X_i)] \qquad \dots (20-2)$$

وبالتعويض عن$\pi(X_i)$ و$1-\pi(X_i)$ بما يساويهم بحسب الصيغتين $(10-2)$ و$(11-2)$ وكالاتي:

$$ln[l(\beta)] = \sum_{i=1}^{n}\left[Y_i ln\left(\frac{e^{\acute{X}_i\underline{\beta}}}{1+e^{\acute{X}_i\underline{\beta}}}\right) + (1-Y_i)ln\left(1-\frac{e^{\acute{X}_i\underline{\beta}}}{1+e^{\acute{X}_i\underline{\beta}}}\right)\right]$$

$$= \sum_{i=1}^{n}\left[Y_i ln\left(e^{\acute{X}_i\underline{\beta}}\right) - Y_i ln\left(1+e^{\acute{X}_i\underline{\beta}}\right) + ln\left(\frac{1}{1+e^{\acute{X}_i\underline{\beta}}}\right) - Y_i ln\left(\frac{1}{1+e^{\acute{X}_i\underline{\beta}}}\right)\right]$$

$$= \sum_{i=1}^{n}\left[Y_i\left(\underline{\acute{X}_i}\underline{\beta}\right) - ln\left(1+e^{\acute{X}_i\underline{\beta}}\right)\right] \qquad \ldots(21-2)$$

ومن اجل الحصول على تقديرات (β) وهي $(\hat{\beta})$ لتعظيم لوغارتم دالة الامكان $ln[l(\beta)] = L\left(\underline{\beta}\right)$ تؤخذ المشتقات من الدرجة الاولى ومساواة الدالة الناتجة بالصفر لكل j من الصيغ و j من المعلمات اي نستخرج المشتقة الجزئية الاولى لكل β وبحسب ما يأتي:

$$\acute{L}\left(\underline{\beta}\right) = \acute{ln}[l(\beta)] = \frac{\partial L\left(\underline{\beta}\right)}{\partial \underline{\beta}} =$$

$$\frac{\partial \acute{ln}[l(\beta)]}{\partial \underline{\beta}}$$

$$= \sum_{i=1}^{n}\left[Y_i \frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right) - \frac{\partial}{\partial \beta_j} ln\left(1+e^{\acute{X}_i\underline{\beta}}\right)\right]$$

$$when \qquad \frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right) = \frac{\partial}{\partial \beta_j}\left(\beta_\circ + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}\right)$$

$$= \acute{X}_{ij}$$

and $\qquad X_{i0} = 1$

$$\frac{\partial}{\partial \beta_j} ln\left(1+e^{\acute{X}_i\underline{\beta}}\right) = \frac{1}{1+e^{\acute{X}_i\underline{\beta}}} \quad \frac{\partial}{\partial \beta_j}\left(1+e^{\acute{X}_i\underline{\beta}}\right)$$

$$= \frac{1}{1+e^{\acute{X}_i\underline{\beta}}} \quad e^{\acute{X}_i\underline{\beta}} \frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right)$$

$$= \pi(X_i)\acute{X}_{ij}$$

$$\therefore \acute{L}\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[Y_i\acute{X}_{ij} - \pi(X_i)\acute{X}_{ij}\right] = 0 \qquad \ldots(22-2)$$

$$\sum_{i=1}^{n}\left[\left(Y_i - \pi(X_i)\right)\acute{X}_{ij}\right] = 0$$

ونعوض قيمة $\pi(X_i)$ فينتج:

$$\acute{L}\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[\left(Y_i - \frac{e^{\acute{X}_i\underline{\beta}}}{1+e^{\acute{X}_i\underline{\beta}}}\right)\acute{X}_{ij}\right] = 0$$

نستنتج بانه تكونت لدينا $(p+1)$ من المعادلات غير الخطية ونقوم بحلها بإحدى الطرائق التكرارية التقليدية كطريقة نيوتن رافسون (NR) لإيجاد تقديرات دالة الامكان الاعظم في المشتقة من الدرجة الاولى ومساواتها للصفر

ولذلك فان خوارزمية نيوتن رافسون التكرارية لإيجاد قيم (β) التقديرية لدالة الامكان الاعظم في الانموذج لوك اللوجستي ستكون في $(m+1)$ من التكرارات وكالآتي: [8][9]

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\acute{X}\, \mathrm{V}^{(\mathrm{m})}\, X)^{-1}\, \acute{X}\left(Y - P^{(m)}\right) \qquad \dots(23-2)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ P^{(m)} = \begin{bmatrix} \pi_1{}^{(m)} \\ \pi_2{}^{(m)} \\ \vdots \\ \pi_n{}^{(m)} \end{bmatrix},\ X = \begin{bmatrix} \underline{\acute{X}_1} \\ \underline{\acute{X}_2} \\ \vdots \\ \underline{\acute{X}_n} \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\mathrm{V}^{(\mathrm{m})} = \begin{bmatrix} \pi_1{}^m(1-\pi_1{}^m) & 0 & \cdots & 0 \\ 0 & \pi_2{}^m(1-\pi_2{}^m) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \pi_n{}^m(1-\pi_n{}^m) \end{bmatrix}$$

إذ ان:

Y: يمثل متجه متغير الاستجابة ذو بعد $(n*1)$

$P^{(m)}$: يمثل القيم الاحتمالية لحدوث متغير الاستجابة ذو بعد $(n*1)$ للتكرار m

X: تمثل مصفوفة المتغيرات التوضيحية ذو بعد $(n*P+1)$

$\mathrm{V}^{(\mathrm{m})}$: مصفوفة مربعة للتباينات عناصر قطرها الرئيسي $[\pi_i{}^m(1-\pi_i{}^m)]$ مكتسبة من التكرار (m)

وكل حل يحدد نقطة حرجة اما عظمى او صغرى وتكون النقطة الحرجة عظمى عندما تكون مصفوفة المشتقات الجزئية من الدرجة الثانية سالبة اي ان كل عنصر من عناصر القطر اصغر من الصفر ومن خصائصها انها تشكل مصفوفة التباين والتباين المشترك

لتقديرات المعلمة بإيجاد المشتقة الجزئية الثانية لكل $(p+1)$ من المعادلات في المعادلة (2 – 20) وكما يأتي : [3]

$$\acute{L}\left(\underline{\beta}\right) = \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \acute{\beta}_j} = \frac{\partial}{\partial \acute{\beta}_j}\left[\sum_{i=1}^{n}\left(Y_i \acute{X}_{ij} - \pi(X_i)\acute{X}_{ij}\ \right)\right]$$

$$= \frac{\partial}{\partial \acute{\beta}_j}\left[\sum_{i=1}^{n} -\pi(X_i)\acute{X}_{ij}\right]$$

$$= -\sum_{i=1}^{n} \acute{X}_{ij}\ \frac{\partial}{\partial \acute{\beta}_j}\left(\frac{e^{\acute{X}_i\underline{\beta}}}{1+e^{\acute{X}_i\underline{\beta}}}\right)$$

$$= \frac{e^{\acute{X}_i\underline{\beta}}\ \frac{\partial}{\partial \acute{\beta}_j}\left(\underline{\acute{X}_i}\underline{\beta}\right)}{\left(1+e^{\acute{X}_i\underline{\beta}}\right)^2}$$

$$= \ \pi(X_i)\ [1-\pi(X_i)]X_{ij}$$

وبالتعويض نحصل على المشتقة الجزئية من الدرجة الثانية وكالآتي:

$$\acute{L}\left(\underline{\beta}\right) = -\sum_{i=1}^{n} \acute{X}_{ij}\left[\pi(X_i)\left(1-\pi(X_i)\right)\right]X_{ij} \qquad \dots(24-2)$$

ونستمر بتطبيق الصيغة (2 – 23) حتى يتحقق معيار التوقف وهو تجانس المعالم المقدرة اي ان التقديرات تبدأ بالأستقرار بمعنى يبدأ التغيير في المقدرات يتلاشى واحياناً يمكن ان تتوقف دون الوصول الى النقاط العليا ومن فرضيات هذه الخوارزمية هو تحديد نقاط البداية بالعملية التكرارية.

2–9–1–1 نيوتن رافسون (*Newton Raphson*)

اقترح نيوتن رافسون استعمال طريقة المربعات الصغرى الاعتيادية لحل المعادلات غير الخطية لإيجاد التقديرات الابتدائية للمعلمة المجهولة. *[4]*

ويكون هدف هذه الخوارزمية هو إيجاد جذر للمعادلة التي لا تحل خطياً ولا جبرياً ويتم فيها تحديد تقديرات المعلمة التقريبية اعتماداً على نقطة البداية فبعد استخراج المشتقة الاولى للوغارتم دالة الامكان الاعظم $\acute{L}\left(\underline{\beta}\right)$ ومساواتها للصفر بالنسبة الى $\left(\underline{\beta}\right)$ وجدنا ان هنالك $(p+1)$ مِن المعادلات غير الخطية ويكون الحل في مثل هذه المعادلات عبارة عن متجه يتضمن عناصر (β_j) .

وبعد التحقق من المصفوفة ذات المشتقات الثنائية الجزئية هي سالبة وان الحل يكون ذا نهاية عظمى عليا بدلاً من أن يكون ذا نهاية عظمى صغرى فيمكن ان نستنتج ان المتجه يحتوي على تقديرات المعلمة التي تمتلك احتمال عالياً ويكون التقدير باستعمال العمليات التكرارية لهذه الخوارزمية (NR) وهي الاكثر شيوعاً في حل انظمة مثل هكذا معادلات حسب الصيغة الآتية:[38][10]

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[\acute{\acute{L}}(\hat{\beta})^{(m)}\right]^{-1} \left[\acute{L}(\hat{\beta})^{(m)}\right] \qquad \dots(25-2)$$

وان:

$\underline{\hat{\beta}}^{(m)}$: هي تقديرات الامكان الاعظم للتكرار (m)

$\underline{\hat{\beta}}^{(m+1)}$: التقديرات الجديدة بالتكرار اللاحق (m + 1)

$\acute{L}(\hat{\beta})^{(m)}$: مصفوفة المشتقة الاولى للوغارتم دالة الامكان الاعظم عندما $\underline{\beta} = \underline{\hat{\beta}}^{(m)}$

$\acute{\acute{L}}(\hat{\beta})^{(m)}$: مصفوفة المشتقة الثانية بعد تعويض $\underline{\hat{\beta}}$ بالمقدر $\underline{\hat{\beta}}^{(m)}$

2-9- 1- 2 خطوات خوارزمية نيوتن رافسون (*Steps Algorithm Newton Raphso (NR)*)

يمكن اختصار مراحل حساب تقديرات الامكان الاعظم بواسطة خوارزمية (NR) التكرارية كما يأتي:[38] [17]

اكتب المعادلة هنا.

1- نقوم بحساب التقديرات الأولية $\underline{\hat{\beta}}^{(\circ)}$ بواسطة احدى الطرائق البسيطة مثل طريقة المربعات الصغرى الاعتيادية (OLS).
2- بعدها نعوض هذه التقديرات في الصيغة رقم (2 − 5) لينتج عنها التقديرات الجديدة.
3- بعد ذلك نحسب اكبر فرق مطلق بين عناصر التقديرات الجديدة والتقديرات في الخطوة السابقة فإذا كان المقدار صغيراً ويمكن إهماله يتم التوقف.
4- ثم تتم اعادة حساب تقديرات جديدة بواسطة الخطوة (2) وحساب اكبر فرق في الخطوة (3) وفحصه الى ان يتم التوصل إلى الحل المطلوب.

$$\acute{L}(\beta)^{(m)} = \frac{\partial\, L(\underline{\beta})}{\partial \beta_j} = \acute{X}\left(Y - P^{(m)}\right)$$

$$\acute{L}(\hat{\beta})^{(m)} = \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \acute{\beta}_j} = -\acute{X}\ V^{(m)}\ X$$

$$P(\beta / X) = f(X / \beta)\, P(\beta) \qquad \dots(26-2)$$

2-9-1-3 نسبة الارجحية (OR) (Odds Ratio)

هي مقياس لحجم التأثير، يصف قوة وعلاقة الارتباط أو عدم الاستقلال بين قيمتين ثنائيتين. وتستعمل كقيمة إحصائية وصفية، وتؤدي دوراً مهماً في تحليل الانحدار لوك اللوجستي. وتفيد نسبة الأرجحية في تقييم اخطاء حصيلة معينة (مثلاً، مرض ما) عند وجود عامل معين (مثلاً، التعرض لمرض). هي نسبة أرجحية وقوع حدث ما في المجموعة التجريبية إلى أرجحية وقوع الحدث نفسه في مجموعة الحقيقة . وبذلك فإن نسبة أرجحية تساوي 1 تعني عدم وجود فارق بين مجموعتي المقارنة. أما إذا كانت نسبة الأرجحية أقل من 1 فتدل على فعالية التداخل التجريبي في إنقاص خطر الإصابة بالمرض المدروس. هذا، وتقترب قيمة نسبة الأرجحية كثيراً من قيمة الخطر النسبي عندما يكون معدل الحدث صغيراً.*[47]*

وبذلك تكون طريقة للتعبير عن احتمال حدوث شيء ما (وقوع حدث معين) منسوب الى احتمال عدم حدوث ذلك الشيء (عدم وقوع الحدث) ويكون ضمن المجال $\{1, \infty\}$ويعرف بانه

$$odds = \frac{\pi_i}{1-\pi_i} = \frac{\text{احتمال الحدوث}}{\text{احتمال عدم الحدوث}} \qquad \dots(27-2)$$

$$= \frac{\dfrac{e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}}{1-\dfrac{e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}}$$

$$= \frac{\dfrac{e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}}{\dfrac{1}{1+e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}}}$$

$$= e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p}$$

$$= e^{\frac{X_i'\beta}{}} \qquad \ldots(28-2)$$

2-9-2 طريقة المربعات الصـــغرى الموزونة (*Weighted Least Squares Method*) (*WLSM*)

يُعد انموذج الانحدار (*log-logistic*) الثنائي حالة خاصة من النماذج الخطية العامة التي تكون امتداد للانموذج الخطي البسيط.

يمكن كتابة انموذج الانحدار الخطي العام كالآتي[44] :

$$y_i = b_\circ + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi} + \varepsilon_i \qquad \ldots(29-2)$$

$$y_i = b_\circ + \sum_{j=1}^{p} b_j x_{ij} + \varepsilon_i \qquad , i = 1,2,\ldots,n\, ;\ j = 1,2,\ldots,p$$

وبذلك يتم تقدير معلمات أنموذج الانحدار (*log-logistic*) الثنائي بأستعمال طريقة المربعات الصغرى الموزونة (*WLS*) بهدف الحصول على افضل تقدير .

لذلك سيتم في هذه الطريقة تحويل معادلة المنحني اللوجستي الى معادلة خطية يمكن منها تقدير

معلمات أنموذج المختار. إنَّ معادلة المنحني اللوجستي هي

$$\pi_i = \frac{e^{B_0+B_1X_{1i}+B_2X_{2i}+\cdots\ldots+B_kX_{ki}}}{1+e^{B_0+B_1X_{1i}+B_2X_{2i}+\cdots\ldots+B_kX_{ki}}} \qquad \ldots\ldots \qquad (30-2)$$

إذ π_iتمثل احتمال النجاح الحقيقي والتي يمكن التعبير عنها بالدالة التجمعية (C.D.F) ،وبذلك فإنَّ احتمال نجاح العينة المشـــاهدة وقيمة P_i يمكن التعبير عنها بانها احتمال الحقيقي π_i مضافاً اليها الأخطاء العشوائية (u_i) أي أنَّ

$$P_i = \pi_i + u_i \qquad \ldots\ldots \qquad (31-2)$$

وعلى فرض أنَّ الأخطاء العشـوائية $_i$ مسـتقلة (*independent*) و تتوزع توزيعا ذي الحدين (*binomial distribution*) بمتوســط صــفر وتباين $\left(\frac{\pi_i(1-\pi_i)}{n_i}\right)$، فإنَّ احتمال فشل استجابة العينة المشاهدة $(1-P_i)$ سيكون [39]

$$1-P_i = 1-\pi_i - u_i \dots\dots \qquad (32-2)$$

وبقسمة المعادلة(31-2)على المعادلة (32-2) نحصل على

$$\frac{P_i}{1-P_i} = \frac{\pi_i}{1-\pi_i}\left(\frac{1+u_i/\pi_i}{1-u_i/Q_i}\right)$$

إذ

$$Q_i = 1-\pi_i$$

وبأخذ (ln) لطرفي المعادلة

$$\ln\frac{P_i}{1-P_i} = ln\frac{\pi_i}{1-\pi_i} + ln\left(\frac{1+u_i/\pi_i}{1-u_i/Q_i}\right) \quad \dots\dots \qquad (33\text{-}2)$$

إذ $\ln\left(\frac{P_i}{1-P_i}\right)$ يطلق عليها $(\text{logit}(P_i))$ للقيم المشـــاهدة بينما $\left(ln\frac{\pi_i}{1-\pi_i}\right)$ يطلق عليها $(\text{logit}(\pi_i))$ للقيم الحقيقية .

ومن معادلة (31-2) فأن

$$E(P_i) = \pi_i \qquad ، \ \text{Var}(P_i) = \frac{\pi_i(1-\pi_i)}{n}$$

وللبحث عن العلاقة بين القيم الحقيقية والمشـــاهدة للوغاريتم النســبة المضـــافة (logit)سنفرض

$$f(P_i) = \ln\left(\frac{P_i}{1-P_i}\right) \quad ، \ f(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) \quad \dots\dots. \qquad (34\text{-}2)$$

وباستعمال صيغة تايلر من درجة الأولى حول π_i للمعادلة (30-2)فإنَّ

$$f(P_i) = f(\pi_i + u_i)، \quad f(P_i) = f(\pi_i) + u_i\frac{\partial f\pi_i}{\partial P_i}$$

وبمقارنة هذه المعادلة بالمعادلة (33-2) ومساواه المشتقة الى صفر نحصل على :

$$u_i \frac{\partial f\pi_i}{\partial P_i} = Log\left(\frac{1 + u_i/\pi_i}{1 - u_i/Q_i}\right)$$

وباشتقاق المعادلة (2-33) نحصل على

$$\frac{\partial f\pi_i}{\partial P_i} = \frac{1}{\left(\frac{\pi_i}{1-\pi_i}\right)} \cdot \frac{(1-\pi_i)(1)-\pi_i(-1)}{(1-\pi_i)^{\ 2}}$$

$$\therefore \frac{\partial f\pi_i}{\partial P_i} = \frac{1}{\left(\frac{\pi_i}{1-\pi_i}\right)(1-\pi_i)^{\ 2}} = \left(\frac{1}{\pi_i(1-\pi_i)}\right)$$

وبذلك فإنَّ معادلة (2-32) ستكون

$$\mathrm{Ln}\frac{P}{1-P} = \ln\frac{\pi_i}{1-\pi_i} + \frac{u_i}{\pi_i(1-\pi_i)}$$

على فرض أَنَّ Logit للقيم الحقيقية هو ${y^*}_i$ أي أَنَّ ${y^*}_i = ln\frac{\pi_i}{1-\pi_i}$ للقيم الحقيقية فإنَّ

$$ {y^*}_i = B_0 + \sum_{i=1}^{P} B_i X_i \qquad \ldots \qquad (35-2)$$

وبذلك تم تحويل المعادلة (2-30)غير الخطية الى معادلة خطية يمكن تقدير معاملات الانحدار منها وإنَّ المتغير الذي تم تحويله ${y^*}_i$ له متوسط مساوي تقريبا الى $\ln\left(\frac{\pi_i}{1-\pi_i}\right)$ بعبارة أخرى مساواه الى $(X_i B)$ وتباين ${\sigma^2}_i$ إذ أَنَّ:

$$ {\sigma^2}_i = \frac{1}{n\pi_i(1-\pi_i)} \qquad (36-2)$$

ان انحدار المتغير الذي تم تحويله ${y^*}_i$ يعتمد على X_i له تباين غير ثابت (غير متجانس) لإنَّه يعتمد على قيمتي $(n، P_i)$ وما دامت قيمة P_i تعتمد على X_i ومن ثَم فإنَّ ${y^*}_i$ يعتمد على X_i بذلك سيفقد أنموذج أغلب افتراضاته الخاصة ومن هذه الافتراضات سيكون تباين الخطأ العشوائي غير ثابت لذا قد يعاني من مشكلة عدم تجانس التباين وعندها لا يمكن تقدير معلمات أنموذج باستعمال طريقة المربعات

الصغرى الاعتيادية (OLS) وبذلك تم اللجوء الى طريقة المربعات الصغرى الموزونة (WLS) لتقدير معلما ت أنموذج لوك اللوجستي ليتم الحصول على أفضل تقدير. لذلك لابد من ضرب المعادلة الانحدار (2-35) بالجذر التربيعي للوزن المرافق $\sqrt{W_i}$ لكي نحصل على تجانس للتباين وان :

$$\mathrm{W_i} = n_i P_i (1 - P_i)$$

$$\sqrt{\mathrm{W_i}} = \sqrt{n_i P_i (1 - P_i)}$$

وبذلك تصبح المعادلة (2-35) بعد ضربها بالوزن كالاتي:

$$\sqrt{W_i} {y^*}_i = B_0 + \sqrt{W_i} X_i B_i + u_i \qquad \ldots.. \quad (37 - 2)$$

بذلك يمكن تقدير معلمات أنموذج $(\mathrm{B_0}, \mathrm{B_1}, \ldots \ldots, \mathrm{B_p})$ في المعادلة (2-37) لإيجاد افضل المعالم التي تبحث عن القيم التي تجعل الفرق بين الاستجابة المشاهدة والاستجابة المقدرة أقل ما يمكن أي بمعنى تصغير مجموع مربعات الخطأ الموزون.

$$SSE_{\,i} = \sum W_i (Y_i - \hat{Y}_i)^{2} \ldots. \qquad (38 - 2)$$

$$SSE = \sum W_i \left(Y_i - \hat{B}_0 - \hat{B}_1 X_{1i} - \cdots \hat{B}_p X_{pi} \right)^{2} \qquad (39 - 2)$$

وبحل التفاضلات الجزئية بالنسبة $(\mathrm{B_0}, \mathrm{B_1}, \ldots \ldots, \mathrm{B_p})$ ومساواة الناتج الى الصفر نحصل على ماياتي:

$$\frac{\partial SSE}{\partial B_0} = B_0 \sum W_i + B_1 \sum W_i X_{1i} + \cdots + B_p \sum W_i X_{pi} = \sum W_i Y_i$$

$$\frac{\partial SSE}{\partial B_1} = B_0 \sum W_i X_{1i} + B_1 \sum W_i {X^2}_{1i} + \cdots + B_p \sum W_i X_{1i} = \sum W_i \; X_{1i} Y_i$$

$$\frac{\partial SSE}{\partial B_p} = B_0 \sum W_i X_{pi} + B_1 \sum W_i X_{pi} X_{1i} + \cdots + B_p \sum W_i = \sum W_i \ X_{ip} Y_n$$

عند اتباع أسلوب المصفوفات تكتب المعادلات بصورة الشكل الاتي:

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} ، \ W = \begin{bmatrix} W_1 & 0 & \ldots\ldots & 0 \\ 0 & W_2 \ldots\ldots & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 \ldots\ldots & & W_n \end{bmatrix} ،X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1i} \\ 1 & X_{21} & \cdots & X_{2i} \\ \vdots & & \vdots & \vdots \\ 1 & X_{n2} & \cdots & X^2{}_{ni} \end{bmatrix} ،B = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_p \end{bmatrix}$$

$$X'WX = \begin{bmatrix} \sum W_i & \sum W_i \ X_{i1} & \sum W_i \, X_{i2} & \cdots & \sum W_i \, X_{ip} \\ \sum W_i \ X_{i1} & \sum W_i \ X^2{}_{i1} & \sum W_i \ X_{i1} X_{i2} & & \sum W_i \ X_{i1} X_{np} \\ \vdots & & \ddots & & \vdots \\ \sum W_i \ X_{ip} & \sum W_i \ X_{ip} X_{i1} & \sum W_i \ X_{ip} X_{i2} & \cdots & \sum W_i \ X^2{}_{ip} \end{bmatrix} ، \quad X'WY = \begin{bmatrix} \sum W_i \, Y_i \\ \sum W_i X_{i1} \, Y_i \\ \vdots \\ \sum W_i X_{ip} \, Y_n \end{bmatrix}$$

$$\beta = \acute{X} W Y \ \acute{X} W X$$

إذ إنَّ:

Y: يمثل متجه التحويل الخطي (اللوجت) لانموذج ذي بعد (n×1)

W: يمثل مصفوفة مربعة عناصر قطرها الرئيسي يمثل الأوزان ذات الرتبة (n×n)

X′WX: تمثل مصفوفة مربعة حصيلة ضرب مصفوفة الأوزان مع مصفوفة المتغيرات التوضيحية ذات بعد [(p+1) (p+1)]

XWY : تمثل متجه مبدلة مصفوفة المتغيرات التوضيحية ومصفوفة الأوزان مع متجه التحويل الخطي ذي بعد [(p+1) × n]

وبضرب طرفي المعادلة بXWX^{-1} وبذلك سوف نصل الى قيمة $\hat{B}$المقدرة والتي تكون:

$$\widehat{\boldsymbol{B}} = (\mathbf{X'WX})^{-1}\mathbf{X'WY} \qquad \text{..... (40-2)}$$

2-9-3 طريقة تصغير مربع كاي (Minimum Chi - Square Method (MCSM))

أقترح كارل بيرسون طريقة بسيطة ومتعددة الاستعمالات وفعالة لاختبار التوافق بين تنبؤات الانموذج والبيانات التجريبية وهي من الطرائق الشائعة الاستعمال في التقدير وتعتمد على تصغير احصاءة مربع كاي لبيرسون المعروفة حسب الصيغة الرياضية (41-2) *[25]*:

$$x^2 = R(\beta) = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \text{... } (41-2)$$

و ان:

O_i: تمثل القيمة المشاهدة عند المستوى i

E_i: تمثل القيمة المتوقعة عند المستوى i

$R\left(\underline{\beta}\right)$: تمثل احصاءة مربع كاي لبيرسون

وفي حالة الانموذج اللوجستي الثنائي الاستجابة فإن:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{(Y_i - \pi_i)^2}{\pi_i} + \frac{[(1-Y_i)-(1-\pi_i)]^2}{1-\pi_i} \qquad \text{... } (42-2)$$

ويتم اختصارها الى:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{(Y_i - \pi_i)^2}{\pi_i(1-\pi_i)}$$

وبعد تعويض قيمة $(\pi_i, 1-\pi_i)$ في الصيغة آنفا نستنتج:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \left[\left(Y_i - \frac{e^{\underline{\acute{X}_i \beta}}}{1+e^{\underline{\acute{X}_i \beta}}} \right)^2 \frac{\left(1+e^{\underline{\acute{X}_i \beta}}\right)^2}{e^{\underline{\acute{X}_i \beta}}} \right] \qquad \text{... } (43-2)$$

وبعد التبسيط تصبح الصيغة كالآتي:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \left[{Y_i}^2\, e^{-\underline{\acute{X}_i}\, \underline{\beta}} + (1-Y_i)^2\, e^{\underline{\acute{X}_i \beta}} - 2Y_i(1-Y_i) \right] \qquad \text{... } (44-2)$$

لإيجاد $\underline{\beta}$ التي تعطي اقل $R\left(\underline{\beta}\right)$ يتطلب ايجاد المشتقة الاولى ومساواتها بالصفر كما يأتي[27]:

$$\frac{\partial R\left(\underline{\beta}\right)}{\partial \underline{\beta}} = \sum_{i=1}^{n} \acute{X}_{ij}\left[(1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right)-(Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right)\right] \qquad \dots(45-2)$$

وتكون علاقة غير خطية وعند حلها يتطلب تطبيق احدى الطرائق التكرارية كما هو الحال مع طريقة الامكان الاعظم وهنا سنستعمل طريقة نيوتن _ رافسون لإيجاد تقديرات β اذ يتم حساب المشتقة الثانية فضلا عن المشتقة الاولى وكما يأتي:

$$\frac{\partial^2 R\left(\underline{\beta}\right)}{\partial \beta_j \acute{\beta}_j} = \sum_{i=1}^{n} X_{ij}\acute{X}_{ij}\left[(1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right)+(Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right)\right] \qquad \dots(46-2)$$

ويتم تطبيق طريقة نيوتن رافسون حسب الصيغة الرياضية السابقة الذكر $(25-2)$ وكالآتي:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[\acute{\acute{L}}(\hat{\beta})^{(m)}\right]^{-1}\left[\acute{L}(\hat{\beta})^{(m)}\right]$$

$$\acute{L}(\hat{\beta})^{(m)} = \frac{\partial R\left(\underline{\beta}\right)}{\partial \beta_j} = \sum_{i=1}^{n} \acute{X}_{ij}\left[(1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right)-(Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right)\right]$$

$$\acute{\acute{L}}\left(\underline{\hat{\beta}}\right)^{(m)} = \frac{\partial^2 R\left(\underline{\beta}\right)}{\partial \beta_j \partial \acute{\beta}_j} = \acute{X}\ \mathrm{V}^{(\mathrm{m})}\ X$$

وان:

X: تمثل مصفوفة المتغيرات التوضيحية تم ذكرها.

$\mathrm{V}^{(\mathrm{m})}$: مصفوفة مربعة عناصر قطرها الرئيسي $\left[(Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right)+(1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right)\right]$ مكتسبة من التكرار (m).

2−10 الاختبارات المتعلقة بأنموذج الانحدار (*Log-Logistic*)الثنائي

(*Tests related to the binary logistic regression model*)

هناك العديد من الاختبارات التي يتم فيها تقيم جودة التوفيق ومعيار التقيم لأنموذج الانحدار اللوجستي وكذلك حسن المطابقة للانموذج ومن هذه الاختبارات اختبار والد (*wald*) واختبار معامل التحديد (R^2) واختبار نسبة الامكان الاعظم واختبار نسبة

مربع كاي كذلك واختبار هوزمر_ليمشو (*Hosmer- Lemeshow*) وغيرها من المعايير الاخرى . وذلك لكي يتم معرفة الى أي مدى يتناسب انموذج المقدر مع البيانات ، وعليه سوف نختصر على بعض الاختبارات الاتية:[40]

1-10-2 اختبار هوزمر- ليمشو لجودة المطابقة (*Hosmer- Lemeshow Quality of Conformance Test*)

يقوم هذا الاختبار بتجميع حالات العينة بناء على قيم الاحتمالات المتوقعة إذ يُعد اختبار هوزمر- ليمشو أًحد الاختبارات التي تسعمل فـــي جوده التوفـــيق لأنموذج الانحدار (*Log-Logistic*) . ويعتمد هذا الاختبار على مدى قرب الاحتمالات المتوقعة والاحتمالات المشاهدة. إذ يعمل على أساس تجميع حالات قيـم الاحتمالات المتوقعة ، لذا يعد هذا الاخـتبـار مشـابـه لحد ما لاختبار (χ^2) لحسـن المـطـابقـة إذ يستعمل لتقـيـم حســن المطــابــقــة للأنموذج وهــو يســمــح بــأي عــدد مــن المتغيــرات التوضيحيـة سـواء أكانت متقطعة أًو مستمرة، وإنَّ هـذه الاحصـاءة تتوزع توزيـعاً (χ^2) بدرجة حرية $(c\text{-}1)(r\text{-}1)$، واذا كانت احتمالها أكبــر مــن مستوى المعنـويـة فــإنَّ هـذا يـؤكـد جودة التوفـيق للأنموذج بالكامل، إذ يقوم هذا الاختبار على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل اعمده ،ما صفـــوف تقسم على أساس القيم المشاهدة لمتغير المعتمد وهما الصفـــر والواًحد. تكون الفـــرضية الإحصائية للانموذج كالاتي:*[5]*

$\mathbf{H_0}$: الأنموذج المقدر يوافق البيانات بشكل جيد .

$\mathbf{H_1}$: الأنموذج المقدر لا يوافق البيانات بشكل جيد .

ويمكن ترميز احصاءة الاختبار هذه بالرمز Y ويتم احتسابها وفقا للصيغة الأتيـة: -

$$\mathbf{Y} = \sum_{i=1}^{k} \frac{\left(\mathrm{O_i} - n_i\overline{\mathrm{P_i}}\right)^2}{n_i\overline{\mathrm{P_i}}\left(1-\overline{\mathrm{P_i}}\right)} \quad \ldots\ldots\ (47-2)$$

إذ إنَّ:

$\mathbf{O_i} = \sum_{j=1}^{n_i} \mathbf{Y_j}$: تمثل القيم المشاهدة للمجموعة

$\overline{\mathbf{P_i}} = \sum_{j=1}^{n_i} \frac{\mathbf{p_i}}{n_i}$: تمثل القيم المتوقعة للمجموعة

2-10-2 اختبار ولد (*Wald Test*):

يقيم اختبار *Wald* (الذي سـمي على اسـم ابراهام والد) القيود المفـــــروضـة على معلمات الإحصـــائية بناء على المســافــــة الموزونة بين التقدير غير المقيد وقيمته المفترضة فـي ظل الفرضية الصفرية والتي تنص علـــى (إنَّ تأثير معامـل لوجت مـــــا يساوي صفــرا)، ويكون ذات اهمية ومعنوية للمتغيرات والثوابت التوضيحية وتأثيرها في متغير الاسـتجابة في النماذج اللوجسـتية ، لـــــــذا يعد أحد الأسـاليب الكلاسيكية (الثلاثــة)لاختبـــار الفــرضيات، وميزة هـذا الاختبار عن الاختبارات الأخرى هو أنه لا يتطلب سـوى تقدير أنموذج غيـر المقيد، وإنَّ الفرضية الإحصائية للانموذج كالاتي:*[47]*

H_0:تأثير معامل لوجيت يساوي صفر.

H_1:تأثير معامل لوجيت لا يساوي صفر.

ويمكن حساب احصاءه الاختبار حسب الصيغة الأتية:

$$t = \frac{\mathrm{b}}{\mathrm{SE_b}} \quad = t^2 = \left[\frac{\hat{\beta}_j}{S.E\ (\hat{\beta}_j)}\right]^2 \qquad \ldots\ldots\ldots \quad (48-2)$$

إذ إنَّ :

b : هي قيمة معامل الانحدار اللوجستي للمتغير المستقل .

$\mathrm{SE_b}$: هي قيمة الخطأ المعياري لمعامل الانحدار اللوجستي للمتغير المستقل .

علما إنَّ الاحصاءه Wald تتبع توزيع (χ^2) بدرجة حرية $(c\text{-}1)(r\text{-}1)$، كما أنَّ هذا الاختبار هو اختبار من الطرفين ويجب إنَّ تكون قيمة معنوية المعلمات المناظرة لقبول أَو رفض فرضية العدم باستعمال الاحتمالات التي تكون أقل من (0.05) لكي يتم رفض الفرضية وقبول الفرضية البديلة والتي تنص على (إنَّ تأثير المعامل لوجت لا يساوي صفراً)، أي إنَّ المتغير المستقل له تأثير بالتنبؤ بقيمة المتغير التابع. ويرمز

لاحصاءة اختبار ولد بالحرف (W) وتكون عبارة عن مربع اختبار (T) وتحسب حسب الصيغة الاتية: [25]

$$W = t^2 = \left[\frac{\hat{\beta}_j}{S.E\ (\hat{\beta}_j)}\right]^2 \qquad \dots (49-2)$$

إذ ان:

$\hat{\beta}_j$: تقدير معلمة الانحدار اللوجستي للمتغير التوضيحي.

$S.E\ (\hat{\beta}_j)$: تقدير الخطأ المعياري لمعلمة الانحدار اللوجستي للمتغير التوضيحي.

اذا كانت قيمة الاختبار الاحتمالية$(P-value)$ اقل من (0.05) ترفض فرضية العدم وهذا يعني ان معاملات المتغير التوضيحي معنوية ذات دلالة احصائية.

وفي اختبار(wald) تستعمل احصاءه (z) وهي عبارة عن الجذر التربيعي لاختبار ولد وتتوزع توزيعاً طبيعياً قياسياً حسب الصيغة الاتية:

$$Z = \frac{\hat{\beta}_j}{S.E\ (\hat{\beta}_j)} \quad , \ j = 1,2, \dots . p \qquad , \qquad Z \sim N(0,1) \qquad \dots (50-2)$$

وتقارن قيمة الاحصاءه (z) مع القيم الجدولية $Z_{\left(1-\frac{\alpha}{2}\right)}$ ، $Z_{\left(\frac{\alpha}{2}\right)}$

فتقبل فرضية العدم في حالة وقوع قيمة (z) بينهما وبمستوى معنوية (α).

2-10-3 معيار المقارنة بين طرائق التقدير المستعملة :- (*Comparison Criterion Between Used Estimation Methods*)

في تحليل الانحدار يعد التخطيط طريقة طبيعية اكثر لعرض الاتجاه العام للبيانات بأكملها يمكن حساب متوسط المسافة من كل نقطة الى إنموذج الانحدار المتوقع واظهاره على انه متوسط الخطأ التربيعي ، أحد الامثلة على الانحدار الخطي باستعمال (MSE) طريقة المربعات الصغرى التي تبين وتقيم مدى ملأئمة أنموذج الانحدار الخطي لنمذجة مجموعة البيانات ثنائية المتغير ، فيتم الاعتماد على مقياس متوسط مربعات الخطأ $(MSE) Mean\ Square\ Error$ لكل محاولات (تكرارات) الظاهرة قيد الدراسة مقسوما على عدد تلك المحاولات كمقياس مقارنة بين طرائق التقدير المستعملة يشير الى مدى دقة التقدير وان تناقص قيمته تشير الى جودة ودقة المقدرات والهدف الذي يراد

تحقيقه هنا هو الحصول على المقدر $[(\hat{Y}_i) = \hat{\pi}(X_i)]$ في الصيغة $(2-10)$ ويتم احتسابه كالآتي:[40]

$$MSE = \frac{1}{n-p}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \quad \dots(2-51)$$

اذ ان:

$\hat{Y}_i$: تمثل القيمة التقديرية للمشاهدات.

n: تمثل حجم العينة.

p: تمثل عدد المعلمات.

2−11 مفهوم الخوارزمية الجينية (*Genetic algorithm concept*)

الخوارزمية (*Alogarithm*): وتعني اصـطلاحاً ؛ مجموعة خطوات وقواعد رياضـية ، يتم برمجة الحواسـيب عليها لتتعامل مع المسـائل المختلفة ، وايجاد الحلول المناسـبة لها . أمّا الخوارزمية الجينية (*Genetic algorithm*) في علم الجينات فهي عملية تعد الى حد ما عملية ذكاء اصـطناعي يتم فهي ادخال الشـيفرة الجينية (الكروموسـومات) كاملة الى نظام الحاسـوب إذ يقوم على معالجتها بناء على الخطوات المدخله إليه ، والتعليمات لتتنبأ بحدوث طفرات مثلاً ، ودراسـة كيف تتم عملية المضـاعفه الجينية والتعبير الجيني وقد يطلق على هذه العملية ايضا البرمجة الجينية . لذلك هي أسلوب من أساليب الذكاء الاصطناعي (*Artificial Intelligence*) ، في ايجاد افضـل حل للمشـكلة قيد البحث بطريقة فعالة وسـريعة جاءت فكرتها من البروفيسـور في علم الحاسـوب جون هولاند (John Holland) من جامعة ميشـيغان (Michigan University) كان هدفه تطوير مفهوم اجرائية التطور الطبيعية وتصـميم انظمة صـناعية تتمتع بمميزات مشـابهة للأنظمة الطبيعية وان رغبته المسـتمرة لتحسين أداء النظم الحسابية جعل من الخوارزميات الجينية اكثر فعالية في حل مسائل الامثلية (Optimization Issues).[42]

فهي من طرائق الرسالة العشوائي التي تعالج مشكلة ما من أجل التوصل الى أفضل النتائج الممكنة إذ تعتمد على آلية الانتقاء الطبيعي ونظام الجينات الطبيعية لذا تمتلك كماً هائلا من الحلول البديلة، والحل الناتج من تطبيق الخوارزمية يكون في أغلب الأحيان حلاً قريبا الى الجينات الطبيعية لذلك برزت أهميتها في حل المسائل المعقدة .

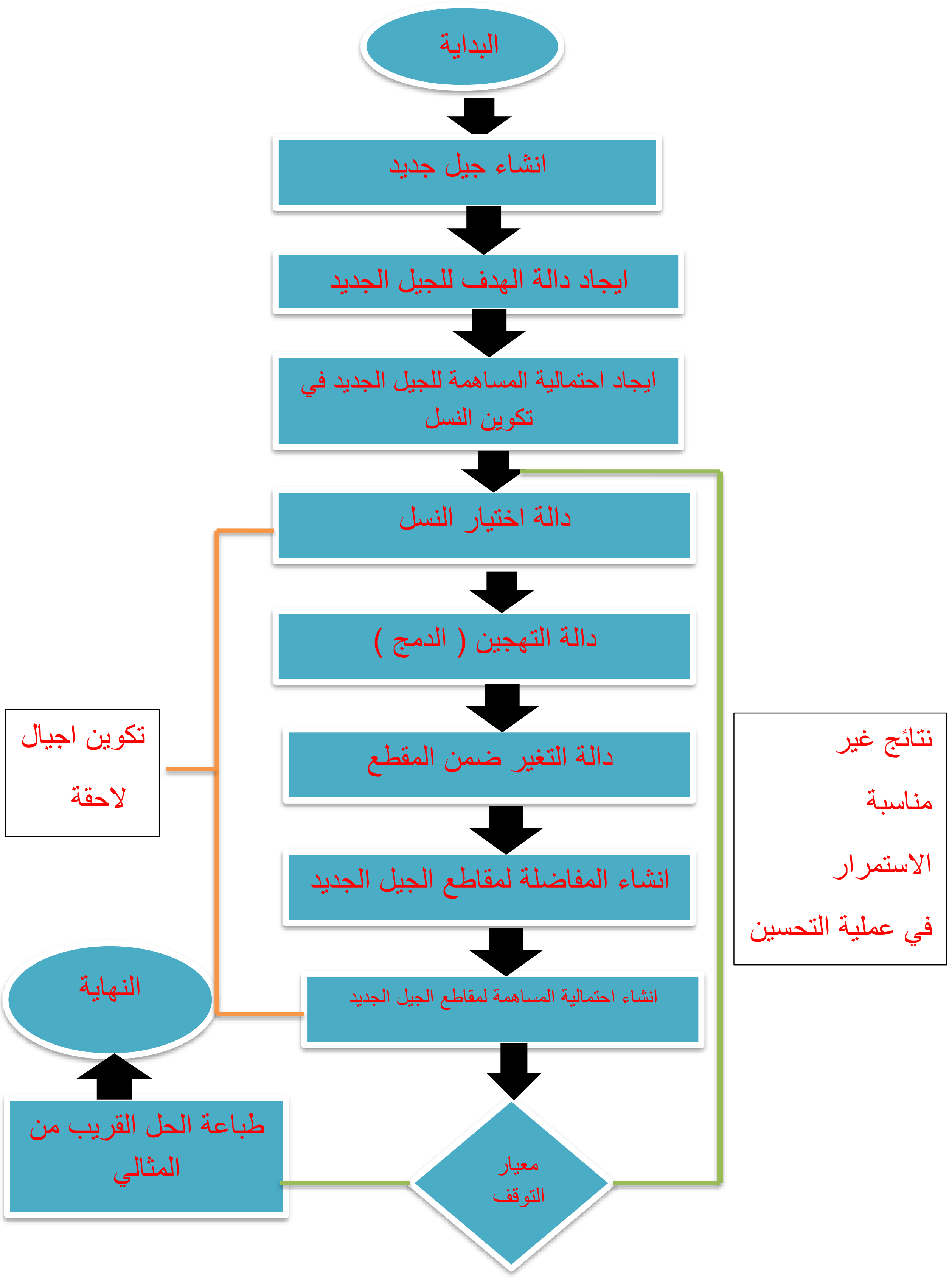
ان فكرة عمل الخوارزمية الجينية تعتمد على أفكار الهندسة الوراثية، التي تتميز بالإنّتاج المقصود للمجموعات المورثة لكي يتم تكوين افراد جديدة ذات صفات جيدة على هذا الأساس تقوم الخوارزمية الجينية باختيار الحلول الأفضل من بين عدد كبير من الحلول وتقوم بأجراء بعض التبديلات والتدخلات بين هذه الحلول بهدف تكوين حل أفضل فهي طريقة آلية. كذلك هي طريقة من طرائق الاستمثال والبحث ويمكن تصنيف هذه الطريقة كإحدى طرائق الخوارزميات التطويرية (*Algorithm Evaluation*) والتي تعتمد على تقليد عمل الطبيعة من منظور دارون (*Drown*) ان الكائنات الحية تتناسب مع الظروف المحيطة بها وتحاول ان تكيف من غير تغير الظروف عبر الاجيال وفي حالة عدم تمكنها من تكيف وتطوير فأنها تنقرض فالكائنات ذات المواصفات القوية هي التي تسود بينما تضمحل وتموت الكائنات ذات الصفات الضعيفة كما تعد الطفرة الوراثية التي تحدث بنسب قليلة جدا من العوامل التي تساعد على تطوير الصفات الوراثية المنقولة عبر الجينات . وتتميز الخوارزمية الجينية بسمتين أساسيتين وهما:[4]

1- العشوائية :إنَّ القاعدة الأساسية في الخوارزمية الجينية هي العشوائية إذ أنَّ عمليتي اعادة الإنّتاج والاختيار تم بطريقة عشوائية.

2- فضاء الرسالة : إنَّ الخوارزمية الجينية تعتمد على فضاء الرسالة أو ما يسمى بالمجتمع الذي يحتفظ عند عملية إعادة الإنّتاج وتشكيل مجتمع جديد بالحلول التي تنتج أفضل الصفات.

وتستعمل الخوارزمية الجينية الوراثية تقنية بحث لايجاد حلول مضبوطة او تقريبية تحقق الامثلية ، تصنف الخوارزميات الوراثية على انها من طرائق البحث الشامل الاستدلالي (*Global search heuristics*). لذلك تكمن فلسفة الخوارزمية الجينية بشكل عام على توليد عدد كبير من الحلول المتوفرة لمشكلة معينة ومن ثم يتم تقييم كل حل من هذه الحلول وصولاً الى الحل الافضل وتكون فرصته اكبر لتوليد حل اخر في تحسين نقل فرصة الحل السيء وبتكرار هذه العملية تتطور نوعية الحلول الموجودة وتصل او

تقترب من الحل الامثل وتطبيقها بالشكل الصحيح تكون فعالة في حل المشكلات المعقدة التي تعجز الطرائق الاخرى عن حلها , ونظراً للافتراضات المقيدة الخاصة بخوارزمية (NR) سوف ندرس كفاءة خوارزمية اخرى كالخوارزمية الجينية (GA) لتقدير المعلمة ويمكن ان ينظر لها على انها بديل جيد والخوارزمية الجينية تفعل ذلك بطريقة سريعة جداً وقريبة من الحل الدقيق:*[10]*



الشكل (2-5) المخطط العام للخوارزمية الجينية من تصميم الباحث

2-11-1 اهم ما يميز الخوارزمية الجينية :

1- تقوم فـي مجتمع يكون عبارة عن مجموعة من المشاهدات وليس يختص بواحدة .

2-لها ميزة في استعمال القوانين الاحتمالية ولا تستعمل القوانين الكلاسيكية.

3- تمتاز الخوارزمية الجينية بإنَّها لا تتعامل مع المتغيرات نفـــــسـها ولكن تتعامل مع التشفير لمجموعة المتغيرات.

4-تتميز الخوارزمية الجينية بإنَّها لا تتوسع فـــي معلومات إضافـــية بل تستعمل دالة الهدف مباشرة (*Objective function*) .

2-11-2 المشاكل التي تعالجها الخوارزمية الجينية

(*Proplems Solving With The Genetic Algorithm*)

1- بعض المشاكل التي تحتاج الى تحليل بيانات في قواعد البيانات الالكترونية الكبيرة كالبيانات المالية وعلم الفلك وغيرها.

2- مشاكل فشل طرائق البحث التقليدية لحل المسألة .

3- المشكلة التي تكون فيها المتغيرات مرتبطة بعلاقة غير خطية او علاقة غير مفهومة بصورة جيدة.

4- المشاكل التي يكون فيها الحل التقريبي مقنع كمعالجة الصورة .[18]

2-11-3 عمل الخوارزمية الجينية (*Work Genetic Algorithms*)

تعتبر من التقنيات الهامة في البحث عن الخيار الامثل من مجموعة حلول متوفرة لتصميم معين، توجد في التطبيقات المعلوماتية الاحيائية (*bioinformatics*) .وتوجد خطوات تعمل بها وهي :

1- يتم تقييم الابناء الجدد بالاعتماد على الدالة الأصلية.

2- تتغير الكروموســومات الاصــلية بالاعتماد على تقييم الابناء وان كل كروموســوم يتكون من عدد من القيم ويتم تحديد عدد هذه القيم حسب المسألة وقيمة البداية.

3- يتم اختيار أفضل الأباء لكي تتم عملية انتاج الابناء.

4- يتم توليد جيل جديد باستخدام الطفرة والعبور.

12-2 منهجيـــــة الخوارزمية الجينية (*Genetic Algorithm Methodology*)

بأستعمال منهجية الخوارزميات الجينية بهدف الحصول على الحلول المثلى للمسائل الرياضية كإحدى الطرائق التكرارية المتبعة حديثاً من اجل اتخاذ القرارات الصحيحة ، ان اهم ما يميز هذه الطريقة هو إيجاد علاقة بين المشكلة قيد البحث وبين الخوارزمية وتنشأ هذه العلاقة عن طريق الترميز Encoding ودالة التقيم Evaluation function ويكون ترميز الحلول بأستعمال سلسلة من الارقام الثنائية Binary Numbers(0،1) وهو من افضل الحلول او استعمال رموز اخرى كأرقام حقيقية وتسمى هذه الارقام كروموسومات . اما دالة التقييم فتأخذ كل كروموسوم على جانب وتقييم أدائه في حل المشكلة بإعطائه قيمة معينة وكلما كانت هذه القيمة كبيرة كان الكروموسوم ذات كفاءة عالية Fitness Function وتسمى دالة المفاضلة. تطبق عملية التهجين والطفرة للحصول في النهاية على مجموعة الكروموسومات التي تمثل الجيل الاخير وباختيار الافضل وصولا للحل الأمثل وتنتهي عنده. [41]

1-12-2 محتويات الخوارزمية الجينية (*Contents of the genetic algorithm*)

تم تطوير وتطبيق عدد كبير من تحورات الخوارزميات الجينية على نطاق واسع لحل مختلف المشاكل التحسينية، وعلى الارجح كان هولاند اول من استخدم عمليات التقاطع والدمج والتحور والانتقاء حيث هذه العمليات تعتبر العمود الفقري للخوارزميات الجينية بكونها استراتيجية لحل المشاكل . وتتالف الخوارزمية الجينية من العناصر الاتية :

1- التشفير (*Encryption*)

يكون التشفـــير عاملاً أساسياً مهماً لنجاح الخوارزمية الجينية إذ يقصد به إيجاد التمثيل البرمجي المناسب لمجموعة الحلول العشوائية (الكروموسومات) وذلك بهدف الوصول الى الحل الأمثل. اذ هناك عدة طرائق للتشفـــير وهذا يعتمد على المسألة المراد حلها، ومن هذه الطرائق المستعملة فـي تشفير الحلول هي:[18]

أ- التشفـير الشجري (*tree cipher*)

يستعمل هذا النوع من التشفـــير فـــي البرامج الجينية والبرامج التطورية، يكون فيها الكروموسوم على هيأه شجرة تحتوي على عدد من العقد (*Nodes*) وكل عقده تمثل هدفاً معيناً كمتغيرات أًو تنفيذ أوامر عملية رياضية فـي لغات البرمجة .[32]

ب- التشفير الاستبدالي (*substitution cipher*)

هو طريقة في التشفير تستبدل فيها أجزاء من النص العادي بالنص المشفر وفقاً لنظام محدد ، و يعد هذا النوع من التشفير كإحدى خوارزميات التشفير الكلاسيكية ، إذ كل موقع بالكروموسوم يحمل رقماً بين (0-9) ويمكن توضيح هذا نوع من التشفير، كما في الشكل الآتي:[32]

2	1	5	4	8	3	9	2	1

الشكل (2-6) التشفير الاستبدالي للكروموسوم

ت- **التشفير الثنائي (Binary encryption)**

هي شفرة تكتب المعلومات في متسلسلة رقمية تتكون من رمزين هما (0) (1) او (صحيح) (غير صحيح) و يعد هذا النوع من التشفير من أقدم أنواع التشفيرات واكثرها استعمالاً مع الخوارزمية الجينية فضلا عن بساطته النسبية. لذلك يعد من أشهر الطرائق في تمثل الكروموسومات على شكل سلسلة أرقام متكونة من رقمين تأخذ مرة (1) ومرة (0) ويمكن توضيح هذا النوع من التشفير عن طريق الشكل الآتي:[40]

1	0	1	0	1	0	....	1

الشكل (2-7) التشفير الثنائي للكروموسوم

2- دالة اللياقة (*fitness function*)

تُعد دالة الهدف و من أهم المكونات الأساسية للخوارزمية الجينية ، إذ يتم إعطاء كل كروموسوم قيمة معينة تمثل مدى كفاءته أَو مدى اقترابه من الحل من خلالها، لذا تكون دالة الهدف (دالة اللياقة) اما عظمى فهي تعتمد على تعظيم المسألة (أي الحصول على أعلى قيمة لدالة اللياقة) (*Maximization*) وتستعمل في حساب الإنَّتاج والربح، أَو صغرى فهي تعتمد على تصغير (*Minimization*) المسالة (أي الحصول على أدنى قيمة لدالة اللياقة) وتستخدم في حساب الخطأ والجذور التي تحقق المعادلات.[13]

3- المجتمع (*Population*)

يتكون المجتمع من عدد من الأفراد $(Individuals)$ التي تمثل الكروموسومات $(Chromosomes)$ وان كل كروموسوم يتكون من عدد من القيم التي تمثل الجينات $(Gens)$ وهذه الكروموسومات تتطور باستمرار من اجل ايجاد حلول جديدة مكونة اجيال $(Generations)$ ويتم تحديد حجم المجتمع حسب طبيعة المسألة المطروقة التي يراد حلها، اشارت بعض الدراسات الى ان حجم المجتمع المثالي ما يقارب $(30 - 20)$ وتكون بدايته قيمة عشوائية ضمن حدود المسألة.[4][41]

4- الترميز (*coding*) يقصد به التمثيل البرمجي المناسب لمجموعة الحلول العشوائية (الكروموسومات) للمسألة المراد حلها بهدف الوصول الى الحل الأمثل وتمثل الخطوة الاولى في الخوارزمية الجينية ولاقت نجاحاً ملحوظاً ومن الطرائق المستعملة في ترميز الحلول:[28]

5- الطفرة (*mutation*)

هي العملية التي تسهم في خلق الأجيال الجديدة عن طريق الخطأ العشوائي بتغيير أو تبديل بين الجينات المحدودة ضمن الكروموسوم الواحد لإنَّشاء كروموسومات تعطي حلولا جديدة في الجيل اللاحق. وهناك عدة طرائق للطفرة الوراثية منها الطفرة المنتظمة والطبيعية وطفرة الرتب .[10]

6- مقياس التوقف (*Stop Criterion*)

هو الذي يحدد اذا كانت الخوارزمية الجينية سوف تستمر بالرسالة لإنَّتاج الأجيال المتعاقبة بهدف تحسين الحل او انها سوف تتوقف اذا تحققت شروط التوقف ، لذلك يعتمد على عدة مقاييس وتختلف هذه مقاييس باختلاف المشكلة المراد معالجتها و منها يتم تنفيذ الخوارزمية لعدد من الأجيال المطلوبة يحددها الباحثون، وكذلك تحديد وقت مستغرق لتنفيذ الخوارزمية أَو الحصول على أَفضل قيمة لدالة اللياقة في المجتمع الذي تم التوصل إليه عندما تكون أقل من قيمة اللياقة التي حددت في حالة (*Minimize*) أَو تتوقف عندما تكون أحسن قيمة لدالة اللياقة أكبر من قيمة اللياقة التي

حددت فـي حالة (*Maximize*). وبعد تكوين كل جيل جديد يتم فـحص مقياس التوقف للخوارزمية الجينية لنرى اذا تم الوصول الى الحل الأمثل.[13]

2-12- 2 مراحل الخوارزمية الجينية (*Stages of the genetic algorithm*)

تتغير الخوارزمية الجينية حسب اختلاف فروع التقنيات التطورية الا انها تشترك بالمراحل الاتية:
1- البداية (*Star*)[15] تتمثل في تعداد سكاني عشوائي من الكروموسومات (فضاء البحث) وبعبارة اخرى هي مجموعة حلول المسألة.

1- البداية (*Star*)[15] تتمثل في تعداد سكاني عشوائي من الكروموسومات (فضاء البحث) وبعبارة اخرى هي مجموعة حلول المسألة

2- التهيئة (*Initialization*)[22] و تتضمن عملية انشاء الجيل الجديد توليد كروموسومات عشوائية بقدر حجم المجتمع وحسب طبيعة المشكلة المدروسة.

3- دالة التقييم (المفاضلة) (*Fitness function (comparison*) *[38]* وعن طريقها يتم تقييم الكروموسوم إذ يعطى كل كروموسوم قيمة معينة تمثل مدى كفاءته اي مدى اقترابه من الحل وتكون دالة الهدف اما عظمى (Maximize) أو صغرى (Minimize) وعليها تعتمد دالة التقييم. ويتم فيها توليد جيل جديد من الافراد(الكروموسومات) التي تم اختيارها عن طريق عملية الانتقاء (*Selection*) وفق مبدأ البقاء للأصلح ومن ثم اجراء عليها عملية التهجين (*Crossover*) وعملية الطفرة (*Mutation*) لانتاج الابناء(الجيل اللاحق)

4- مجتمع جديـــد (*New Population*) [19] يتضمن توليد جيل جديد بتكرار المراحل الاتية الى أن يكتمل الجيل ويتمثل الجيل بما يأتي:

5- **الانتقاء (*Selection*)** هو عملية اختيار الآباء من المجتمع لأجل التزاوج وتكوين الجيل الجديد. وان لعملية الانتقاء (الاختيار) لها دور كبير فـي تطور الخوارزمية الجينية. وهناك عدة طرائق لانتقاء الكروموسومات المناسبة من الجيل القديم لتكوين الجيل الجديد وهي:[37]

الانتقاء النسبي (*Proportional Selection*):

هذا النوع من الانتقاء (الاختيار) يتم بالاعتماد على نسبة قيمة اللياقة لكل فـرد مقسوما على مجموعة قيم دوال اللياقة لأفـــراد المجتمع، إذ يتم الاختيار حسب المسالة المراد حلها أي انتقاء زوج من الكروموسومات التي تمتلك أفـــضل دالة لياقة. وذلك بحسب المعادلة الآتية:

$$Pepo(x) = \frac{f(X_i(t))}{\sum_{i=1}^{N} f(X_i(t))} \quad .\ .\ .\ .\ .\ .\ .\ (52-2)$$

إذ إنَّ:

Pepo(x) تمثل قيمة نسبة اللياقة

i= تمثل رقم الكروموسوم

f(x)= تمثل قيمة دالة اللياقة لكروموسوم

t= تمثل رقم الجيل

N= تمثل عدد المعلمات

الانتقاء الصفي (*Rank Selection*):

هذا النوع من الانتقاء يتم عن طريق ترتيب الكروموسومات ثم احتساب درجة اللياقة لكل واحد ترتيباً تصاعديا فـــكل كروموسوم (x) يكون له رتبة *Rank*(x) أي موقع بعد الترتيب لذلك يطلق عليه بالانتقاء الرتب، وبعد ذلك يتم حســـاب نســبة الهدف (Targe *sampling Rate*) لكل كروموسوم حسب المعادلة الآتية:

$$TSR(x) = Min + (Max - Min)\frac{\mathbf{Rank(x) - 1}}{\mathbf{N - 1}} \quad \ldots \quad (53\text{-}2)$$

إذ أن:

TSR(x):تمثل نسبة الهدف لكل كروموسوم.

N : تمثل عدد الكروموسومات بالجيل، إنَّ الكروموسوم الأسو يأخذ القيمة 1 والثاني يأخذ 0 من قيمة TSR الى إنَّ نصل إلى الكرموسوم الأفضل.

Min:تمثل الحد الأدنى لدرجة اللياقة لكل كروموسوم.

Max: تمثل الحد الأعلى لدرجة اللياقة لكل كروموسوم.

1. التهجين (*Crossover*)

بعد اختيار الكروموسومات الجيدة من الجيل الاول يتم التزاوج بين كل كروموسومين لتكوين الجيل الجديد (الأبناء) اعتماداً على الكروموسومات (الأم) يأخذ الصفات الجيدة منها ومن طرائقها التهجين بنقطة وبنقطتين وغيرها.[4]

2. الطفرة (*Mutation*)

يتم انشاء الجيل الجديد (الأبناء) وباحتمال وجود الطفرة فتتم عملية الطفرة بأحداث تغيرات عشوائية في صبغتها الكروموسومية وهذا يؤدي بنا الى الحفاظ على الصفات الجيدة بين الجينات في الكروموسوم الواحد والوصول الى الحل بشكل اسرع وفيها يحدث التبادل بين الكروموسومات وعندما لا تكون هناك طفرة فتجري عملية استنساخ الكروموسومات (الآباء) مباشرةً دون حدوث عملية التهجين.[10]

4. الاختبار (*Test*)

نقوم باختبار الحل بتوفر شرط التوقف من عدمه فعند توفره تتوقف الخوارزمية الجينية وتعيد الحل الجديد من تكون اخر جيل.[33]

5.معيار التوقف (*Stopping Criterion*)

يستمر تكوين الاجيال بشكل متعاقب من أجل تحسين أمثلية الحل حتى يتحقق شرط التوقف الذي يعتمد على مقياس توقف الخوارزمية الجينية(الحل الأفضل) ويختلف هذا المقياس بحسب المشكلة المراد معالجتها.[13]

6.الانهاء (*Termination*)

تنتهي الخوارزمية الجينية عند وجود احد العوامل الاتية :

* ايجاد الحل الأمثل.
* توفر قيمة معينة مثل حساب كلفة الانتاج .
* الوصول الى عدد الأجيال المطلوبة.
* الوقوع في الصغرى المحلية (*Local Minimum*) ولا يمكن الخروج منها.[30]

3-12-2 المعايير التي تســتعمل في الخوارزمية الجينية (*Criteria that are used in the genetic algorithm*)

يتم وضع عدة معايير في الخوارزمية الجينية لكي يتم أتخاذ القرار الصحيح فيها وكذلك من اجل الوصول الى الحل الأمثل وهذه المعايير تتحكم في الطفرة و الاختيار و التهجين وكالآتي:-

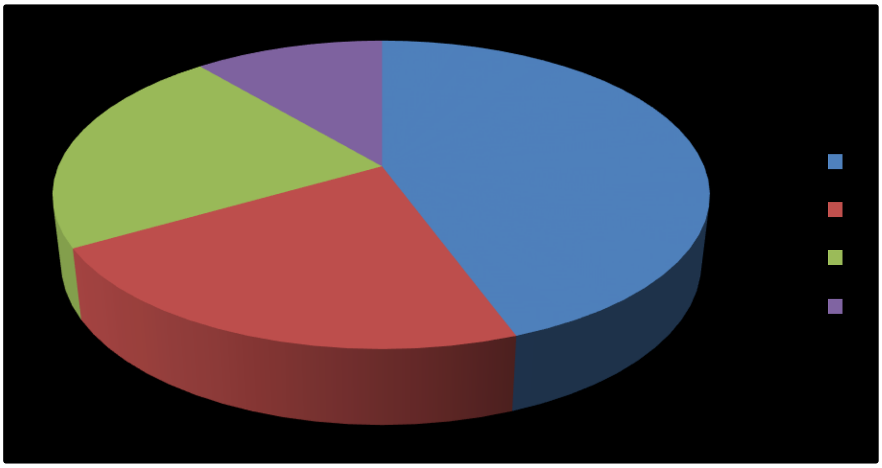
أولا- معيار الاختيار (*Selection Standard*)

حسب نظرية دارون في التطور ان الكروموسومات الفضلى هي التي تؤخذ ومنها تنشأ الكروموسومات (الأبناء) وهناك عدة طرائق لتحديد كيفية اختيار الكروموسوم الافضل ومنها:[40]

1- اختيار عجلة الروليت (*Roulette Wheel Selection*)

تكون عملية اختيار (الاباء) في هذه الطريقة يعتمد على قيم كفاءتهم فكلما زادت قيمة كفاءة الكروموسوم زاد احتمال الاختيار حسب الشكل (2 – 11) الآتي:[17]

(الالوان) تمثل كروموسومات



الشكل (2 – 8) عجلة الروليت

إذ يتم رمي حجر النرد في العجلة وحسب القطاع الذي سيقع فيه سيتم اختيار الكروموسوم ومن ثَمّ فان الكروموسوم ذو القطاع الأكبر (الأكثر كفاءة) سيكون معرضاً للإختيار اكثر من مرة ويتم تلخيص ذلك:-

أ-(المجموع) يتم حساب مجموع قيم دالة التقييم(مفاضلة) كل الكروموسومات الموجودة في المجتمع (تنفذ مرة واحدة فقط من أجل كل تجمع).

ب- (الاختيار) توليد رقم عشوائي (R) ضمن المجال[0,1] .

ج-(الحلقة) نمر بحلقة على الكروموسومات في المجتمع ونقوم بحساب مجموع قيم المفاضلات اثناء مرورنا وعندما نصل لقيمة المجموع أكبر من R فنتوقف ونعيد الكروموسوم الذي وصلنا عنده وأدى الى كسر الحلقة و تواجه هذه الطريقة مشكلة وهي عندما تختلف قيم المفاضلة بشكل كبير وعلى سبيل المثال اذا كانت افضل قيمة مفاضلة (كفاءة) الكروموسوم كبيرة تغطي فيها 90% من عجلة الروليت عندها سيكون لبقية الكروموسومات فرص قليلة جداً للأختيار.[18]

2- اختيار المتبارين (*Selection of contestants*)

هنا يتم اختيار مجموعة من الحلول (المتبارين) عشوائياً من المجتمع وصولاً للحل الافضل كالكروموسومات (الأب) فتكرر هذه العملية بعدد الكروموسومات المختارة (الآباء) اللذين ينتجون اجيال عشوائية (الابناء) ويكون عدد المتبارين في المجتمع الجيني بين(2 – 5) متبارين.[26]

ثانياً- معايير التهجين (*Crossover Criteria*)

1- التهجين المنظم (*Uniform Crossover*)

هذا النوع من التهجين يحتاج الى قيمة احتمالية (probability value) ويرمز لها (P_v) يحددها الباحث قبل تنفيذ الخوارزمية ضمن المجال [0.25,1]. فاذا كان لدينا L جين (خانة) في كل من الكروموسومين الابوين تقوم الخوارزمية بمقارنة كل جين في الكروموسوم الاول مع القيمة (P_v) فاذا كانت لدينا قيمة ترتيبها $I: I \leq L$ واكبر او تساوي (P_v) نقوم بتبديل الجين I من الكروموسوم الاول مع نظيره من الكروموسوم الثاني والاّ يبقى جين الكروموسوم الاول في مكانه في الابن الاول وكذلك الامر بالنسبة لجين الكروموسوم الثاني كما في الشكل(2 – 9) الآتي: [36]

الكروموسوم الاب1	**0.32**	**0.22**	**0.34**	**0.55**	**0.36**	**0.60**	**0.54**	**0.15**
الكروموسوم الاب2	0.12	0.41	0.76	0.32	0.29	0.51	0.64	0.8
ارقام عشوائية	0.32	0.22	0.34	0.32	0.29	0.51	0.54	0.15
Pv=0.3	> Pv	< Pv	> Pv	> Pv	< Pv	> Pv	> Pv	< Pv
الابن 1	0.12	0.22	0.76	0.32	0.36	0.51	0.64	0.15
الابن 2	0.32	0.41	0.34	0.55	0.29	0.60	0.54	0.8

الشكل (2 – 9) يبين التهجين المنظم

2-التهجين البسيط (نقطة واحدة) (*Simple Crossover*)

بعدما يتم اختيار مجموعة الكروموسومات يتم تطبيق هذه الطريقة من التهجين وفق مرحلتين:-

أ- يتم اختيار كروموسومات عشوائية من العينة المختارة.

ب- كل زوج من الكروموسومات يختبر التهجين بحيث يتم توليد رقم عشوائي K ϵ[1,L] إذ L يمثل طول الكروموسوم (عدد الجينات لكل كروموسوم) وثم يتم انشاء كروموسومين جديدين عن طريق تبادل الجينات الواقعة في المجال(L,K+1)كما في الشكل (2_10) الاتي:[36]

K=3,L=8

الكروموسوم الاب1	0.32	0.22	0.34	0.55	0.36	0.60	0.54	0.15
الكروموسوم الاب2	0.12	0.41	0.76	0.32	0.29	0.51	0.64	0.8

يتم التبديل بين كل الجينات من الموقع 4 فما فوق

الذرية الابن 1	0.32	0.22	0.34		0.32	0.29	0.51	0.64	0.8
الذرية الابن2	0.12	0.41	0.76		0.55	0.36	0.60	0.54	0.15

الشكل (2 – 10) يوضح التهجين البسيط

3- التهجين السطحي (*Flat Crossover*)

نفترض لدينا الابوين الآتيين:[33]

الكروموسوم الاب 1 = $(X_{1,1},\dots,X_{1,n})$

الكروموسوم الاب2 = $(X_{2,1},\dots,X_{2,n})$

مجموعة العناصر العشوائية: $r=(r_1,\dots,r_n)$

فبالتالي يكون الابن1= $(x_1{}^1,\dots,x_n{}^1)$

إذ كل قيمة x_i تحسب بالطريقة الاتية:

$$x_i{}^1=r_i x_{1,i}+(1-r_i)x_{2,i} \qquad ; \quad i=1,\dots,n \qquad \dots(54-2)$$

وبنفس العملية بالنسبة الى الابن2.

ثالثا- معايير الطفرة (*Mutation Standard*)

حدوث الطفرة تفيدنا في تأمين التنويع خلال عملية البحث وتمنعها لأن تكون متشابكة مع النقاط المثلى المحلية عن طريق مبادلة جينات مختارة عشوائياً مع قيمة جديدة حصلنا عليها عشوائياً وبذلك نحتاج الى قيمة احتمالية للطفرة (Probability Mutation) ويرمز لها P_m لتحديد الكروموسومات التي لديها طفرات وتكون ضمن المجال [0.01,0.001] وكذلك نحسب الطول الكلي للجينات الموجودة سابقاً في المجتمع عن طريق الدالة (55-2) :[12]

$$S = W + N \qquad (55-2)$$

إذ ان :

S: مجمــــــــــــــــوع اطوال الجينات

W: عدد الجينات في الكروموسوم

N : مجمـــــــــــــــوع السكـان

يتم تحديد احتمال الطفرة عن طريق مدى قابلية الجينات على التحور إذ يتم توليد رقم عشوائي h إذ $h\in[1,\text{مجموع الجين}]$ وان $h<P_m$ وبذلك تكون الطفرة لأجزاء الكروموسوم وبعد عملية الطفرة نستنتج ان الخوارزمية الجينية قضت واحدة من تكراراتها

بما يسمى الجيل. وتتكرر هذه العملية للأجيال المحددة سابقاً الى أن نحصل على الكروموسوم الذي يمثل دالة الهدف المثلى ومن طرائقها ايضاً:عكس حالة الجين (*Reverse Gen Staus*)هذه الحالة تحدث عندما تأخذ الجينات قيماً ثنائية (0) أو(1) فتتم معاكسة القيمة المأخوذة في الجينات بالنسبة للكروموسوم الواحد.تغير الرتب مع التزحيف (*Rank changes with creep*) في هذه الخطوة يتم الاستبدال بمواقع الجينات مع الازاحة العشوائية.[18]

2-13 تطبيق مراحل الخوارزمية الجينية في انموذج الانحدار (*Log-Logistic*)
(*Application of genetic algorithm stages in logistic regression model*)

سوف يتم تطبيق مراحل الخوارزمية الجينية في معادلة دالة الهدف لكل طريقة لإيجاد تقديرات معلمات انموذج الانحدار اللوجستي الثنائي وفقاً لما يأتي:[37][19]

1- البداية: تكوين الكروموسوم عن طريق قيم β_p التي تشكل جينات الكروموسوم و ان ($P = 0,1,...,p$) ضمن الاعداد الحقيقة.
2- التهيئة: انشاء الجيل الابتدائي عن طريق ايجاد قيمة اولية للجينات مع القيم العشوائية لمجموعة القيود الاخرى.
3- في دالة الهدف يتم تقييم الكروموسوم من إذ الكفاءة وصولاً إلى الحل الأفضل بتحديد قيمة β_p.
4- اجراء عملية الاختيار للكروموسوم الذي يمتلك قيمة دالة هدف صغيرة باختيار الاحتمال الكبير لها وايجاد دالة التقييم له عن طريق المعادلة الآتية :

$$fitness\ function = \frac{1}{\mathbf{1} + \text{objective}\ function} \qquad \dots (56-2)$$

$\boldsymbol{f_i}$: تمثل دالة التقييم.

$\boldsymbol{o.f_i}$: تمثل دالة الهدف.

ومن صيغة دالة التقييم نستطيع ايجاد احتمالية هذه الدالة (افضل القيم) بحسب الصيغة الرياضية الآتية[22]:

$$C_{(i)} = \frac{f_{(i)}}{\sum_{i=1}^{n} f_{(i)}} \qquad \dots (57-2)$$

$C_{(i)}$: تمثل احتمالية الفرد (الكروموسوم) i

$f_{(i)}$:تمثل دالة التقييم للفرد i

n: يمثل حجم المشاهدات

وباستعمال احد معايير الاختيار (roulette wheel)عجلة الروليت بتوليد رقم عشوائي $R_{(c)}$ محصور في المجال [1,0] فاذا كان $R_{(c)} < C_{(1)}$ سوف يتم اختيار الكروموسوم الاول (كالأم) او يتم الاختيار بإذ يكون الاحتمال محصور وفق $C_{(p-1)} < C_{(p)} < R_{(c)}$ او يكون الرقم العشوائي محصور وفق $C_{(p-1)} < R_{(c)} < C_{(p)}$ وفي كل مرة يتم تحديد كروموسوم واحد للمجتمع الجديد بالاعتماد على دالة التقييم.

5- بعد اتمام عملية الاختيار تأتي بعدها عملية التهجين للكروموسومات الجيدة في صفاتها عن طريق التزاوج بين كل كروموسومين وبتطبيق احد معاييره وهو التهجين المنظم بالاعتماد على احتمالية التهجين P_c وتقارن هذه القيمة مع قيمة الجينات للكروموسومين (الاباء) لتكوين الجيل الجديد(الابناء) ويحدث التبادل عندما تكون قيمة الجين اكبر او تساوي القيمة الاحتمالية.[32]

6-آخر خطوة ممكن ان تمر بها الكروموسومات هي عملية الطفرة وايضاً تعتمد على مقدار احتمالي (P_m) للمعلمات باستبدال جينات منتقاة عشوائياً مع قيمة جديدة ايضاً حصلنا عليها بشكل عشوائي[19] بتطبيق الصيغة $(2-55)$ التي ذكرناها سابقاً:

ويمكن تلخيص ما ذكره في أعلاه في النقاط الاتية :

1- نحصل على مقياس المقارنة (MSE) لكل طريقة ثم بعد ذلك نقوم بتحديد أفضل طريقة حسب مقياس المقارنة اعلاه .

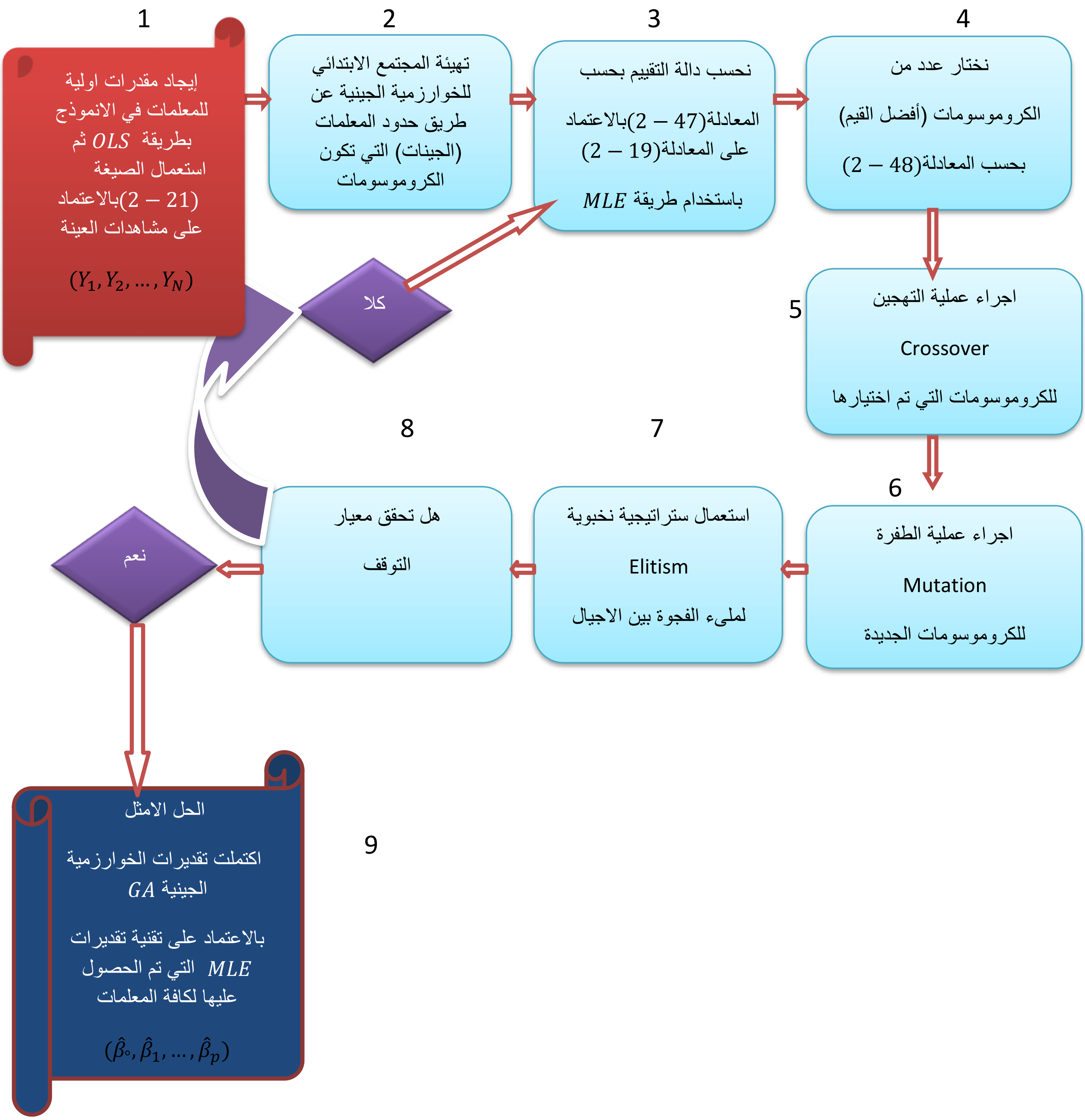
2- توليد معالم مركبة حسب مقياس المقارنة اعلاه .

3- بناء الحل الأساسي عن طريق الطرائق المعرفة وعدد المعالم إذ ان عدد الطرائق المستخدمة هي ثلاثة طرائق(طريقة الامكان الاعظم، طريقة المربعات الصغرى الموزونة ، طريقة تصغير مربع كاي) وان عدد المعالم عشرة.

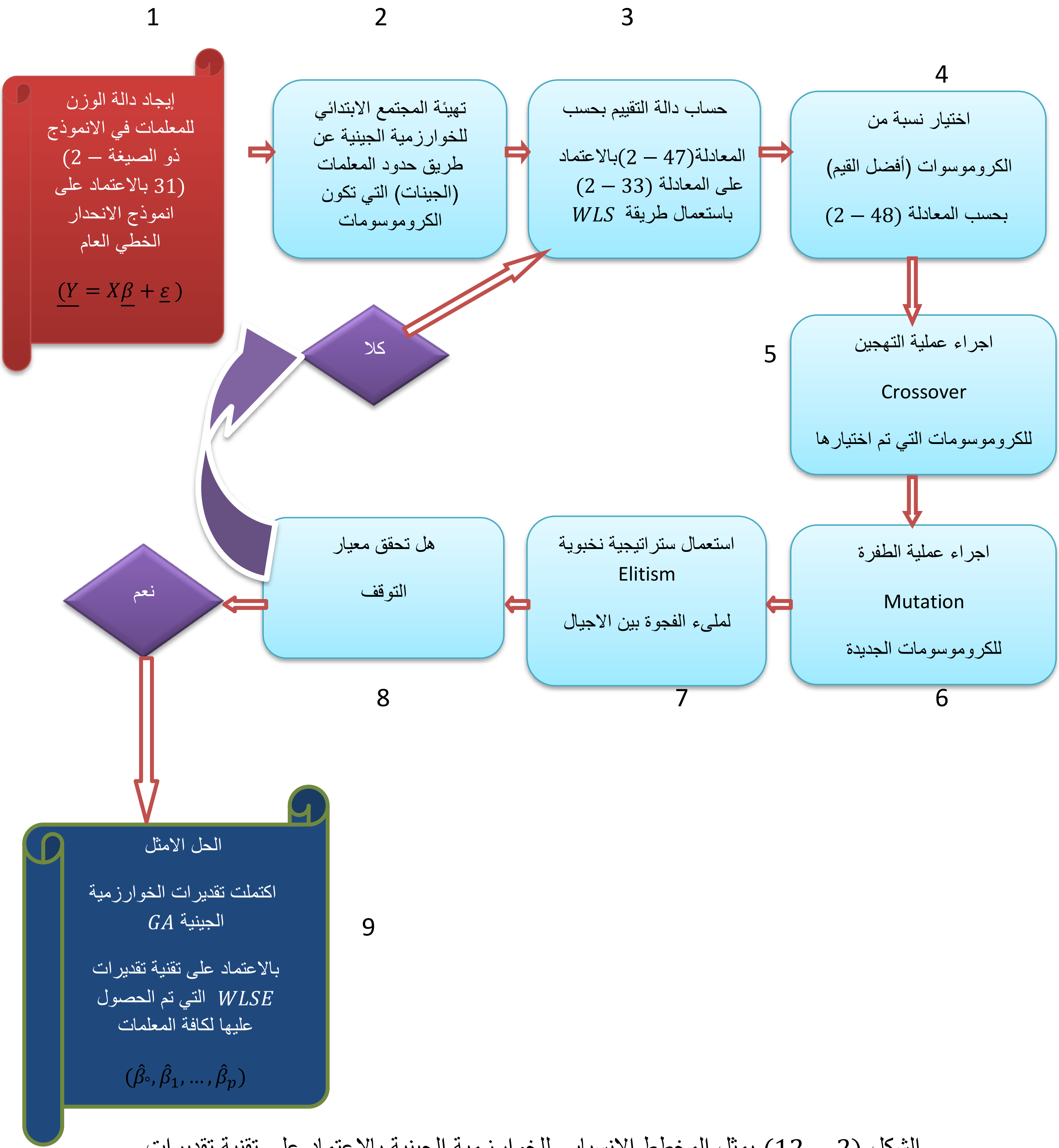
4- بعدها يتم توليد عدد من النماذج لكل خلط تم استخدامه بما في ذلك النماذج الاساسية بواقع محدد مقداره نصف باتجاه الموجب والسالب لـ(m) من المرات وبالتالي سيتولد عدد كبير من النماذج المولدة، يتم احتساب (MSE) لكل نموذج ومقارنته مع المعالم(القيمة المثلى) وعند انتهاء هذا العمل سيتولد لدينا اصغر (MSE) والمعالم التي تم احتساب هذا المقياس من خلالها (افضل المعالم).

5- نقوم بعمل خلط او تزاوج بين معالم الطرائق المستخدمة واجراء احتساب مقاييس المقارنة لهذه النماذج الجديدة والمقارنة مع (MSE) الأفضل الذي يتم تحديده وفق النقطة(1)،فاذا تم الحصول على (MSE) اقل للنماذج الجديدة يتم استبدال قيمة(MSE) المثلى للقيمة الجديدة وكذلك استبدال قيمة المعالم الجديدة.

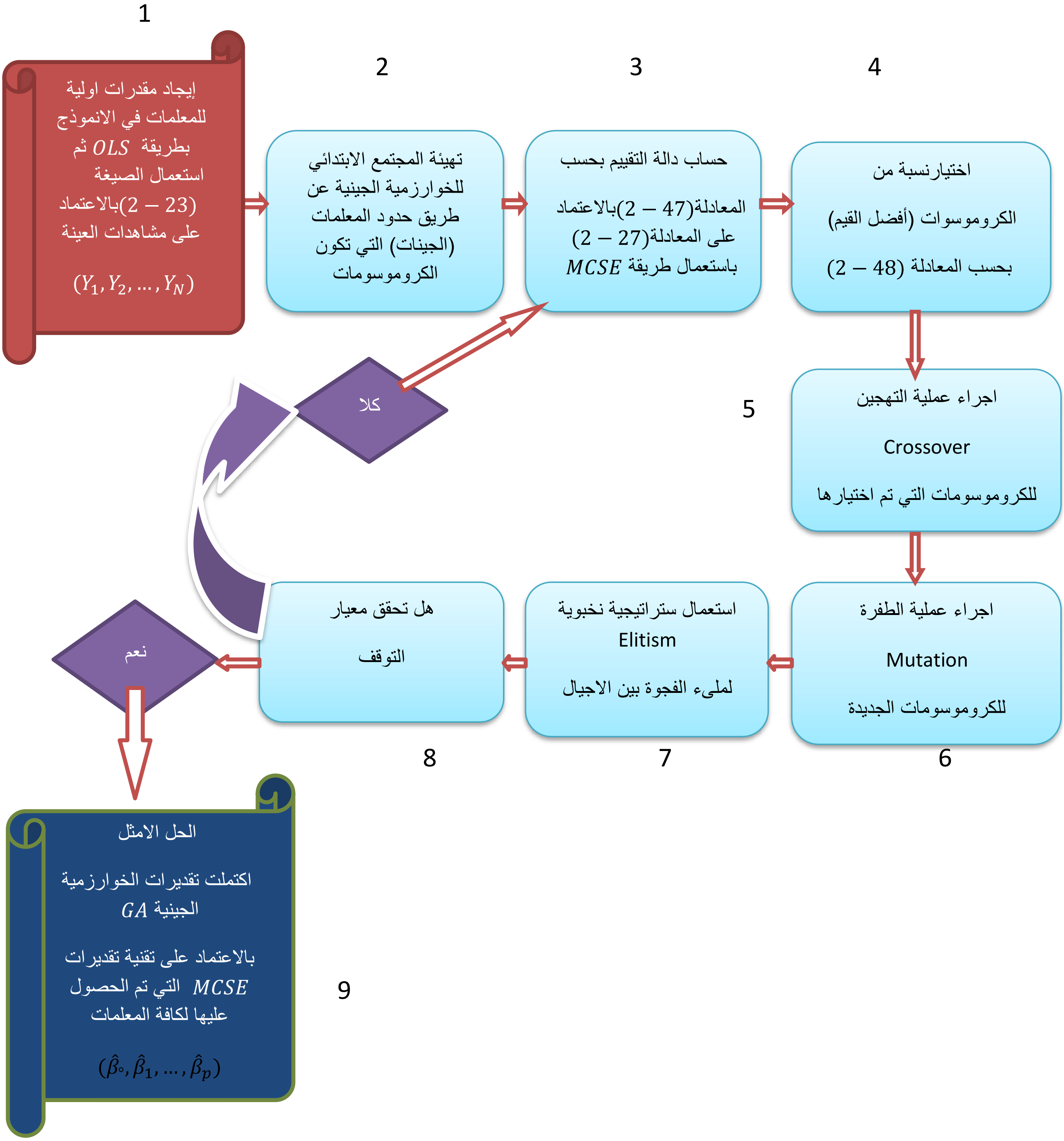
اذ ان المخططات التالية $(2-14)$،$(2-15)$،$(2-16)$، تمثل تطبيق مراحل الخوارزمية الجينية في انموذج الانحدار (*Log-Logistic*) الثنائي وفق الطرائق الثلاثة على التوالي التي عرضناها في الرسالة.



الشكل $(2-11)$ مخطط انسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات MLE (المخطط من عمل الباحث)



الشكل (2 – 12) يمثل المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات $WLSE$ (المخطط من عمل الباحث).



الشكل (13 – 2) المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات $MCSE$ (المخطط الانسيابي من عمل الباحث)

الفصل الثالث

الجانب التجريبي والتطبيقي

1-3 تمهيد *preamble*

يتضـــمن هذا الفصــــل جانبين هما (الجانب التجريبي والجانب التطبيقي). إذ يبحث الجانب التجريبي ما تم التوصـــل اليه في الجانب النظري من طرائق تقدير معلمات إنموذج الانحدار (*Log-Logistic*) الثنائي بما يلائم هدف البحث ولتحقيق ذلك كان لا بدا من دراســة عدد من الحالات التي تختلف من حيث احجام العينات المدروسـة وقيم المعلمات وعدد المتغيرات التوضـــيحية . وعليه يمكن اســـتعمال اســـلوب المحاكاة لأجراء مقارنة ما بين طرائق التقدير المدروســـة في البحث ، عن طريق المؤشــــر الاحصــائي متوسـط مربعات الخطأ (*MSE*) للتوصــل الى الطريقة الفضــلى وهذا المحور الذي انصــب عليه اهتمامنا في هذا المبحث ، أذ تم صـــياغة إنموذج محاكاة يحاكي العديد من الحالات المفترضة التي من الممكن وجودها فعلا في الواقع العملي من حيث (عدد المتغيرات التوضيحية ، عدد المشاهدات الكلية ، وقيم المعلمات) .

2-3 مفهوم المحاكاة (*Simulation Concept*)

هناك بعض العمليات التي تكون معقدة الفهم ولا سيما في بعض المشكلات او النظريات الاحصائية والطبية والهندسية وكذلك الاقتصادية التي يكون تحليلها تحليلاُ منطقيا وذلك عن طريق استعمال البراهين الرياضية في غاية الصعوبة مما يؤدي الى ترجمة هذه النظريات الى مجتمعات حقيقة ، أذن المحاكاة هي طريقة أو اسلوب تعليمي يستعمل عادة لتمثيل الواقع الحقيقي الذي يصعب الوصول اليه بواقع افتراضي تصوري يشبه لحد ما ، اي يعطي نسخة فرضية تكون طبق الاصل من الواقع الاصلي او العالم الواقعي لنظام معين او أنموذج محدد من دون التطرق لهذا الانموذج او النظام بشكل مباشر . وقد ظهرت المحاكاة قديما جدا كما في لعبة الشطرنج وزاد الاهتمام بها في ستينيات القرن الماضي كطريقة مناسبة وفعالة للتعليم ولا سيما بعد ظهور اجهزة الحاسوب . وتوجد عدة طرائق مختلفة للمحاكاة منها الطريقة التناظرية (*Analog method*) والطريقة المختلطة (*mixed method*) وطريقة مونت كارلو ((*MCMC*) (*Markov Chain Monte Carlo*)) وهي الطريقة الاكثر استعمالا في التحليل الاحصائي والتي تقوم على فكرة توليد البيانات العشوائية من المجتمع النظري المماثل

للمجتمع الحقيقي ، فبعد تحديد المجتمع المدروس اذ تولد ارقاما عشوائيا لتكوين عينة تمثل هذا المجتمع . أن اسلوب المحاكاة الذي امتاز بالمرونة ساعد الباحثين في توفير الوقت والجهد والمال وذلك عن طريق الاستعانة بالحاسبات الالكترونية لتكوين البيانات (المشاهـــدات) المطلوبة دون اللجوء الى العمل الميداني للحصول عليها دون الاخلال بالنتائج المطلوبة ودقتها.

تمت المقارنة بين طرائق التقدير الاعتيادية (تقديرات الامكان الاعظم (MLM)، وتقديرات المربعات الصغرى الموزونة (WLS) ، وتقديرات تصغير مربع كاي ($MCSM$) . وتحديد الطريقة الفضلى من بين هذه الطرائق ، أما المقارنة الاخرى سوف تكون بين طرائق التقدير المحسنة بالخوارزمية الجينية ($MLE.\ GA$) و ($WLS.GA$) و ($MCSM.GA$) وتحديد الطريقة الفضلى بينهما ، أن تكوين تجربة المحاكاة التي سوف يتم الحصول عليها تعتمد على عدد من المراحل ، والتي سوف يتم توضيحها في إنموذج الانحدار (*Log-Logistic*) الثنائي الذي تم صياغته في معادلة (10 -2) .

3-3 مراحل بناء تجربة المحاكاة (*Stages of Building Simulation Experiment*)

يتم بناء تجربة المحاكاة على شكل مراحل متسلسلة بهدف الحصول على أفضل تقدير من بين التقديرات المستعملة وتتضمن أربع مراحل مهمة وهي كالاتي :

3-3-1 تحديد القيم الافتراضية :

تعد هذه المرحلة من اهم المراحل التي تعتمد عليها بقية مراحل المحاكاة إذ يتم اختيار القيم افتراضـــيا تجريبا عبر اجراء تجارب واختبار القيم التي اســـتقرت عندها التقديرات واعطت افضل النتائج ، ويتم عن طريق تحديد القيم الافتراضية للمعلمات وكذلك يتم تحديد ثلاثة احجام مختلفة للعينات وبذلك ســـوف تكون عدد المشـــاهدات الكلية (30, 60, 90) مشــاهدة على الترتيب ، وأن هذه القيم تم تحديدها في الانموذج الاول من تقديرات المعلمات للبيانات الحقيقة قيد البحث لأنموذج الخطي العام إذ ان انموذج الانحدار (*Log-Logistic*) الثنائي حالة خاصة منه ، ثم أخذنا (*Standardized*) للمعلمات في الانموذج الثاني على وفق البرنامج الجاهز (*SPSS*) وكما مبين في الجدول (3 – 1)

Para / Mod.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
1	**0.816**	**0.95**	**-0.073**	**-0.002**	**0.013**	**0.004**	**-1.031**	**-0.085**	**-0.0415**	**0.001**	**-0.002**
2	**-0.046**	**0.091**	**-0.088**	**-1.031**	**-0.803**	**-0.002**	**0.150**	**0.092**	**-0.004**	**0.060**	**-0.013**

جدول (1-3) القيم الافتراضية للمعلمات في انموذج الانحدار (Log-Logistic) الثنائي

2-3-3 توليد البيانات

سيتم توليد عشر متغيرات توضيحية كما في الجانب التطبيقي وذلك عن طريق استعمال اسلوب مونت كارلو وذلك عن طريق التوزيع المنتظم (*Uniform Distribution*) أما في ما يتعلق بتوزيع الخطأ فأنه يتبع توزيع برنولي (*Bernoulli Distribution*) وكذلك في هذه المرحلة يتم توليد قيم المتغيرات التوضيحية وقيم متغير الاستجابة ، سوف نستعمل طريقة الرفض والقبول لاحتساب متغيرات ثنائية الاستجابة وتحديد القيم الاحتمالية وفق الاتي :

$$Y = \begin{Bmatrix} 1 \; if \; \pi(X_i) \geq 0.5 \\ 0 \; if \; \pi(X_i) < 0.5 \end{Bmatrix}$$

3-3-3 التقديرات

في هذه المرحلة يتم تقدير معلمات انموذج الانحدار (*Log-Logistic*) الثنائي المعطى في المعادلة (2 ـــــــ 10) وفق الطرائق الاعتيادية ، وأيضا وفق توظيف الخوارزمية الجينية مع طرائق التقدير الاعتيادية التي تم ذكرها في الجانب النظري من الفصل الثاني للبحث وهذه الطرائق هي :

أ- الطرائق الاعتيادية

1- طريقة تقديرات الامكان الاعظم (MLE).

2- طريقة تقديرات المربعات الصغرى الموزونة ($WLSE$).

3- طريقة تقديرات اصغر مربع كاي ($MCSE$).

ب- طرائق التقدير المحسنة عن طريق (GA)

1- طريقة الخوارزميـــة الجينيـــة بالاعتماد على تقنية تقديـــرات الامكـــان الاعظم . $(\mathcal{MLE.GA})$

2- طريقة الخوارزميــة الجينيـــــة بالاعتماد على تقنيــة تقديــــرات المربعــات الصغرى الموزونة $(\mathcal{WLSE.GA})$.

3- طريقة الخوارزميـــة الجينيـة بالاعتماد على تقنيـــة تقديـــرات اصغر مربع كاي . $(\mathcal{MCSE.GA})$

3-3-4 المقارنة بين طرائق التقدير الاعتيادية والحديثة

وهي المرحلة الأخيرة من مراحل وصـــف تجربة المحاكاة وفي هذه المرحلة تم المقارنة بين طرائق التقدير بعد أن تم إيجاد تقديرات المعلمات في المرحلة الثالثة باستعمال المقاييس الاحصائية المستعملة لغرض الحصول على أفضل طريقة تقدير للانموذج قيد الدراســـة ، لذلك ســوف يتم المقارنة بين طرائق تقدير معلمات أنموذج الانحدار (*Log-Logistic*) الثنائي والتي تكون على اساس (الاعتيادية والمحسنة ونبين من هي الفضلى) لجميع الطرائق المستعملة وذلك عن طريق استعمال احد المعايير الاحصـــائية المهمة وهو متوســـط مربعات الخطأ للأنموذج المدروس كما مبين في الجانب النظري من الفصل الثاني حسب المعادلة (2 ـ 51) الاتية:

$$MSE = \frac{1}{R}\sum_{i=1}^{R} MSE_i = \frac{1}{R}\sum_{i=1}^{R}[\frac{1}{N-P}\sum_{i=1}^{N}(Y_i - \widehat{Y_i})^2]$$

إذ **(R)** تمثل عدد مرات تكرار التجربة ، وقد تم تكرار تجربة المحاكاة (R=1000) وذلك بهدف الحصول على افضل نتائج الجانب التجريبي .

4-3 تحليل نتائج المحاكاة (*Analyze the Simulation Results*)

سـوف يتم عرض نتائج عملية المحاكاة ومن ثم تحليلها للوصـول الى أفضـل الطرائق لتقـدير معلمـات أنموذج الانحـدار (*Log-Logistic*) الثنـائي بـالاعتمـاد على المقيـاس الاحصائي متوسط مربعات الخطأ $(\mathcal{MSE})$ لمقدرات الأنموذج .

ســـوف نقوم بعرض نتـائج المحـاكـاة التي تم الحصـــول عليهـا عن طريق برنـامج ($MATLAB$) والتي سـيتم عرضــها في جدول الملاحق، وفيما يأتي النتائج التي سـيتم تحليلها حسب تسلسل الجداول الاتية:

جدول (3-2) متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار (*Log-Logistic*) الثنائي في الانموذج (1) للمعلمات باستعمال الطرائق الاعتيادية والجينية ولكافة احجام العينات)

Sample sizes	Methods	Classic	Genetic	Best
30	MLE	0.231	0.0057	Genetic
	WLSE	0.0191	0.0074	Genetic
	MCSE	0.2159	0.0009	Genetic
	Best	WIs	Mcs.GA	Best
60	MLE	0.1952	0.0024	Genetic
	WLSE	0.0176	0.0036	Genetic
	MCSE	0.1554	0.0005	Genetic
	Best	WIs	Mcs.GA	Best
90	MLE	0.0101	0.0023	Genetic
	WLSE	0.0235	0.0023	Genetic
	MCSE	0.0385	0.0003	Genetic
	Best	MIs	Mcs.GA	

نلحظ في الجدول (3- 2) افضلية وتفوق طريقة الامكان الاعظم المحسنة (MLE.GA) على طريقة الامكان الاعظم الاعتيادية (MLE) وطريقة المربعات الصغرى الموزونة المحسنة (WLSE.GA) على طريقة المربعات الصغرى الاعتيادية ، وكذلك طريقة تصغير مربع كاي المحسنة (MCS.GA) على طريقة مربع كاي الاعتيادية (MCSE) وذلك عند احجام العينة كافة (90 ،60، 30). عن طريق تطبيق جميع الطرائق نشاهد تميز طريقة مربع كاي المحسنة وقد احتلت المرتبة الاولى من بين الطرائق وكذلك لها الافضلية في تقدير المعلمات من حيث امتلاكها اقل متوسط لمربعات الخطـــــأ ، للمقدرات وذلك مقارنة مع بقية طرائق التقدير المحسنة الاخرى ، أما طريقة (MLE) أحتلت المرتبة الاولى مقارنة بطرائق التقدير الاعتيادية عند حجم العينة (90) وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول الذي تم افتراضه .

جدول (3-3) متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار(*Log-Logistic*) الثنائي في الانموذج (2) للمعلمات باستعمال الطرائق الاعتيادية والجينية و احجام العينات كافة

Sample sizes	Methods	Classic	Genetic	Best
30	MLE	0.1351	0.0056	Genetic
	WLSE	0.0367	0.0073	Genetic
	MCSE	0.0905	0.0009	Genetic
	Best	WIs	Msc.GA	Best
60	MLE	0.1839	0.0023	Genetic
	WLSE	0.0156	0.0034	Genetic
	MCSE	0.1395	0.0005	Genetic
	Best	WIs	Msc.GA	Best
90	MLE	0.1029	0.0022	Genetic
	WLSE	0.0058	0.0023	Genetic
	MCSE	0.0739	0.0003	Genetic
	Best	WIs	Msc.GA	

اما جدول (3-3) نلحظ عند احجام العينة كافة (90، 60 ، 30) وكذلك للقيم الافتراضية كافة للمعلمات والانموذج الثاني الذي تم افتراضه هناك افضلية واضحة لطرائق التقديـر المحسنة (*Genetic Algorithm*) على طرائق التقدير الاعتيادية ، ونلحظ هناك تميز طريقـة تصغير مربع كاي الجينية (MCSE.GA) على الطرائق الاخرى إذ كان لها الافضلية الاولى في تقدير المعلمات من حيث امتلاكها اقل (MSE) للمقدرات كافة مقارنة بطرائق التقديـر المحسنة الاخرى . وجائت طريقة المربعات الصغرى الموزونه (*WLSE*) من إذ التقديـر مقارنة بطرائق التقدير الاعتيادية الاخرى وكما موضح في الجدول آنفاً .

جدول (3-4) متوسط مربعات الخطأ (*MSE*) لمقدرات انموذج الانحدار (*Log-Logistic*) الثنائي للانموذجين باستعمال افضل الطرائق الاعتيادية وافضل الجينية وكافة احجام العينات

Methods	First Model		Second Model		
Sample size	WLSE	MCSE.GA	WLSE	MCSE.GA	Best
30	0.0191	0.0009	0.0367	0.0009	Genetic
60	0.0176	0.0005	0.0156	0.0005	Genetic
90	0.0235	0.0003	0.0058	0.0003	Genetic

في الجدول (3-4) ولكلا الانموذجين الأول والثاني الذين تم افتراضهم بأستعمال افضـل الطرائق الاعتيادية (طريقة المربعات الصغرى الموزونة) وافضل الطرائق الجينية المحسنة (طريقة تصغير مربع كاي الجينية) وعند احجام العينة كافة (90 ،60 ، 30) نلحظ ان طريقة

التقدير المحسنة بالخوارزمية الجينية $(MCSE.GA)$ افضل من طريقة التقديـــــــر الاعتيادية $(WLSE)$.

- عند التحسين بين طرائق التقدير الاعتيادية الثلاثة عن طريق الخوارزمية الجينية ، التي بينا عن طريقها من هي افضل الطرائق حسب الافضلية وذلك لامتلاكها اقـــــل (MSE) لمقدرات الانمــــوذج (log-logistic) الثنائي عن طريق الجداول آنفاً (3-2) و (3-3) (3-4) سوف نلحظ ما يأتي :

1- أفضلية طريقة $(WLSE)$ على طرائق التقدير الاعتيادية كافة ، وذلك في تقدير المعلمات لأنموذج الانحدار $(log\text{-}logistic)$ الثنائي ان هذه الطريقة تصدرت باقي الطرائق من حيث أمتلاكها اكثر عدد مرات اقل متوسط مربع خطا مع باقي الطرائق لمقدرات الانموذج ومن ثّم بعدها طريقة $(MCSE)$ وذلك لاغلب حجوم العينات والقيم الافتراضية للمعلمات والنماذج التي تم افتراضها في ما عداً حجم العينة $(n = 90)$ بالنسبة للانموذج الاول إذ طريقة الامكان الاعظم (MLE) الافضلية في الجدول وجاءت بعدها طريقة المربعات الصغرى الموزونة $(WLSE)$ ثانيا .

2- ان طريقة الامكان الاعظم المحسنة $(MLE.GA)$ بأستعمال الخوارزمية الجينية افضل من طريقة الامكان الاعظم الاعتيادية (MLE) بأستعمال خوارزمية نيوتن رافسون وان طريقة تصغير مربع كاي المحسنة $(MCS.GA)$ بأستعمال $(Genetic\ Algorithm)$ افضل من مربع كاي الاعتيادية باستعمال خوارزمية (Newton Raphson) وكذلك طريقة المربعات الصغرى الموزونة المحسنة بأستعمال $(Genetic\ Algorithm)$ افضل من طريقة المربعات الصغرى الموزونة الاعتيادية بالنسبة للقيم الافتراضية للمعلمات وفي كلا الانموذجين وعند احجام العينة كافة (90 , 60 , 30) .

3- أقل متوسط مربعات خطأ (MSE) في الجدول لطريقة مربع كاي المحسنة $(MCS.GA)$ إذ جاءت في المرتبة الاولى لجميع طرائق التقدير المحسنة من إذ عدد مرات لمقدرات الانموذج انحدار $(log\text{-}logistic)$ الثنائي .

4- عند احجام العينة (90, 60 , 30 = n) بالنسبة للقيم الافتراضية للمعلمات والانموذج الاول والثاني في حالة الطرائق الاعتيادية نلحظ تفوق طريقة $(WLSE)$ وهي الافضل من بين الطرائق على كل من (MLE) و$(MCSE)$ ، وعند توظيف واستعمال الخوارزمية الجينية في التحسين في هذه الطرائق ولنفس النماذج عند أستعمال الطرائق الاعتيادية سوف نلحظ تفوق (MCSE.GA) ولها الافضلية من بين الطرائق على كل من $(MLE.GA)$ و$(WLSE.GA)$ وعليه فان مقدرات الانموذج $(log\text{-}logistic)$ الثنائي تحسن اداء الطرائق عند استعمال $(Genetic\ Algorithm)$ وذلك عن طريق ايجاد افضل مقدر للمعلمة المجهولة في انموذج الانحدار $(log\text{-}logistic)$ الثنائي .

ثانياً: الجانب التطبيقي

3-5 تمهيد

يشمل هذا الجزء من الرسالة التطبيق العملي على البيانات الاصلية (الحقيقية) التي حصل عليها الباحث من مركز شهيد المحراب لقسطرة وجراحة القلب/ دائرة صحة بابل ،ونظراً لأهمية حياة الانسان في الجانب الصحي تم التركيز على امراض القلب وتأثيرها في حياة الناس لكونها من الامراض الشائعة الانتشار في المجتمع وكذلك لكثره الوفيات في هذا المرض ،إذ يبحث هذا الجانب في وصف البيانات المتمثلة بإصابة أمراض القلب عن طريق العديد من العوامل منها عوامل تسمى (متغيرات التوضيحية) التي تسبب هذه الامراض ويكون لها تأثير حالة حدوث الوفاة او عدم حدوثه والعامل الاخر هو (متغير الاستجابة) باستعمال أنموذج (*Log-Logistic*) الثنائي الذي اعتمد على (*Genetic Algorithm*) في تحسينها للحل بتقديرها لمعلمات الانموذج .

سيتم التطبيق على البيانات مرضى الراقدين في مركز شهيد المحراب لقسطر القلب لسنة 2015 في مستشفى مرجان التعليمي وبلغ عدد العينة (90) راقد وذلك لمعرفة العوامل التي لها اكثر تأثيراً في وفيات المرضى واخذت المعلومات من طبلات وسجلات المرضى الراقدين فضلا عن برنامج الراقدين المصمم من وزارة الصحة العراقية .

وتتم معالجة البيانات التي تعاني من وجود عشرة عوامل تسبب حالة الوفاة للأشخاص الذين يعانون من امراض القلب وذلك عن طريق تطبيق افضل الطرائق الاعتيادية وافضل الطرائق المحسنة التي تم الحصول عليها من نتائج المحاكاة ومن ثم اجراء الاختبارات الخاصة بالأنموذج الذي يحدد اكثر عامل مؤثر لحدوث الوفاة .

3-6 أمراض القلب (*Heart disease*)

ظلت أمراض القلب السبب الرئيس للوفاة على الصعيد العالمي على مدى السنوات العشرين الماضية. غير أن عدد الأشخاص الذين تفتك بهم اليوم يفوق أي وقت مضى. فقد ارتفع عدد الوفيات الناجمة عن أمراض القلب بأكثر من مليوني حالة منذ عام 2000، ليصل إلى ما يقرب من 9 ملايين حالة وفاة في عام 2019. وتمثل أمراض القلب حالياً 16 في المائة من مجموع الوفيات الناجمة عن جميع الأسباب . إذ دعت

جمعية الصحة العالمية السادسة والثلاثون الى أعداد برنامج طويل الأجل على تعزيز البحوث الخاصـــة بالمرض واســـباب حدوث الوفاة ، فهناك العديد من العوامل التي تسبب حدوث المرض وتأثيرها في حياة الأنسـان مثل تعاطي الكحول وارتفاع ضـغط الدم والدهون كذلك الزيادة المفرطة بالوزن والتدخين وغيرها من الأسباب التي تؤدي الى الوفاة في بعض الأحيان .

7-3 البيانات الحقيقية (*Real data*)

تم جمع البيانات الحقيقية من الســجل الخاص للمرضــى المســجلين في مركز شــهيد المحراب في محافظة بابل حيث يعتبر احد المراكز التخصــصــية التابعة الى مشــفى مرجان التعليمي التي تم تأسـيسـها سـنة 1957م وهي مشـفى تقع في مدينة الحلة في محافظة بابل بالعراق وهو مؤســســه طبية تعليمية عمومية تضــم العديد من المراكز التخصــصــية والتي تقدم خدماتها المجانية لمواطني المحافظة والمناطق المجاورة ، ويشـمل هذا المشـفى مراكز طبية تخصـصـية تعني بعلاج أمراض الجهاز الهضـمي والكبد والعلاج الطبيعي والســكري والأمراض الســرطانية والانعاش وجراحة القلب وغسـل الكلى ،فضـلا عن وجود طوارئ خاصـة وكذلك أجنحة أسـتشـارية متعددة في الباطنية والنفسـية والجلدية وكبار السـن وغير ذلك ،إذ تم اخذ العينة العشـوائية بحجم 90 مريضـــاً وتحديد مدة بقائهم على قيد الحياة لحين الوفاة بالأشـــهر و ادراجها في الجدول ادناه :

جدول (3-5) البيانات الحقيقة لمتغير الاستجابة للمرضى المصابين بأمراض القلب وكذلك المتغيرات التوضيحية لسنة 2015 لعينة حجمها (90) من سجلات المرضى .

ت	Y_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	1	3	2	2	55	1	1	2	1	3	64
2	1	2	2	2	67	4	1	2	1	3	70
3	1	3	1	2	38	5	1	2	1	2	65
4	1	3	1	2	78	2	2	2	2	3	60
5	1	3	1	2	42	7	3	2	1	3	60
6	1	1	1	2	45	3	3	2	1	2	67
7	1	1	2	2	67	5	1	2	1	2	60
8	1	3	2	2	40	2	1	2	1	1	60
9	1	3	2	2	65	8	1	2	1	1	67
10	1	2	2	2	60	5	1	2	1	2	70
11	1	3	2	2	65	1	1	2	1	2	70
12	1	3	2	2	65	5	1	2	1	1	90
13	1	3	1	2	46	5	1	2	2	2	70
14	1	2	1	2	25	2	1	2	1	3	60
15	1	2	2	2	64	7	4	2	1	3	90
16	1	2	1	2	85	7	4	2	2	2	95
17	1	3	1	2	70	5	1	2	2	3	75
18	1	2	1	2	70	6	1	2	2	2	70
19	1	2	1	2	50	4	1	2	3	4	70
20	1	2	2	2	64	4	1	2	3	4	77
21	1	2	2	2	71	6	1	2	2	3	70
22	1	2	2	2	80	4	1	2	3	4	85
23	1	3	1	2	65	2	1	2	1	3	90
24	1	3	1	2	25	4	1	2	1	4	70
25	1	2	2	2	60	2	1	2	1	2	85
26	1	2	1	2	47	4	1	2	1	2	80
27	1	2	2	2	70	2	1	2	1	3	75
28	1	2	2	1	14	5	1	1	1	3	60
29	1	2	1	2	65	2	1	2	1	3	80
30	1	2	1	2	70	3	2	2	1	2	70
31	1	2	2	2	39	2	1	2	1	2	80
32	0	3	1	2	45	8	3	1	1	1	62
33	0	3	1	2	50	7	1	1	1	1	73
34	0	3	1	2	65	7	2	1	1	1	70
35	0	3	1	2	63	8	3	1	2	1	70
36	0	3	1	2	55	8	3	1	1	1	64
37	0	3	1	2	55	8	3	1	1	1	70
38	0	2	1	2	52	8	3	1	1	1	80
39	0	3	1	2	54	2	1	2	1	1	75
40	0	3	1	2	49	3	3	1	1	2	60
41	0	3	1	2	46	8	3	1	1	1	75
42	0	3	1	2	49	8	1	1	1	1	70
43	0	3	2	2	60	8	1	1	1	2	70

ت	Y_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
44	0	3	2	2	35	4	1	1	1	1	60
45	0	1	1	2	65	7	1	1	1	1	69
46	0	1	2	2	65	8	3	1	1	1	83
47	0	2	1	2	70	5	1	1	1	1	60
48	0	3	2	2	65	8	1	1	1	1	60
49	0	3	1	2	65	4	1	2	1	2	62
50	0	1	1	2	63	10	1	1	1	2	73
51	0	3	1	2	61	4	1	2	1	1	75
52	0	3	1	2	44	8	3	1	1	2	65
53	1	2	1	2	60	5	1	2	1	3	85
54	1	2	1	2	66	2	1	2	2	3	90
55	1	2	1	2	52	2	1	2	2	4	70
56	1	2	1	2	65	2	1	2	2	4	70
57	1	2	1	1	32	2	1	2	2	3	68
58	1	2	2	2	62	7	1	2	2	3	68
59	1	2	1	2	49	5	1	2	2	2	67
60	1	2	1	2	66	4	1	2	2	2	77
61	1	3	1	1	17	9	1	2	1	2	60
62	1	2	1	2	47	2	1	2	1	2	76
63	0	3	1	2	44	8	3	1	1	2	65
64	0	3	1	2	60	8	3	1	1	2	80
65	0	4	1	1	1	3	3	2	1	1	12
66	0	3	2	2	60	8	3	1	1	2	70
67	0	3	2	3	72	8	2	1	2	1	77
68	0	3	2	2	68	8	1	1	2	1	63
69	0	3	2	2	65	8	3	1	1	3	65
70	0	3	2	2	56	8	1	1	1	1	70
71	0	3	2	2	60	8	3	1	1	1	60
72	0	3	1	2	40	8	3	1	1	1	74
73	0	3	1	2	63	7	1	1	1	2	72
74	0	3	1	2	61	7	3	1	1	3	72
75	0	3	1	2	60	7	3	1	1	2	70
76	0	3	1	2	63	7	1	1	1	2	64
77	0	3	1	2	53	7	3	1	1	2	65
78	0	3	1	2	43	2	1	2	1	1	65
79	0	3	1	2	70	3	3	1	1	2	70
80	1	3	1	2	70	5	1	2	1	2	65
81	1	2	2	2	50	5	1	2	1	3	60
82	1	3	1	2	82	5	1	2	1	2	75
83	1	2	1	1	24	5	1	2	1	2	80
84	1	3	2	2	50	2	1	2	2	3	90
85	1	3	2	3	80	2	1	2	2	3	85
86	1	3	1	2	64	2	1	2	1	2	75
87	1	2	1	2	63	5	1	2	1	2	80
88	1	3	1	2	61	5	1	2	2	1	83
89	1	3	2	2	45	6	1	2	1	3	60
90	1	3	1	2	70	5	1	2	2	3	65

إذ ان :

متغير الاستجابة (Response Variable):

Y_i: المصابين بأمراض القلب ، $i=0,1$ ،$Y=0$:عدم حدوث الوفاة، $Y=1$: حدوث الوفاة.

المتغيرات التوضيحية(Variables Explanatory):

X_{ij}: العوامل التي تضمنتها الرسالة في دراسة امراض القلب، $j=\ldots1,10$، $i=1..,90$

X_1(الردهة): متغير وصفي يمثل نوع الردهة التي يرقد بها المريض.

X_2(الجنس): متغير وصفي يمثل جنس المريض الراقد.

X_3(الحالة الزوجية): متغير وصفي يمثل الحالة الزوجية للمريض الراقد.

X_4 (العمر): متغير كمي يمثل عمر المريض يقاس بوحدة (سنة).

X_5 (سبب الرقود): متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد.

X_6 (درجة العملية): متغير وصفي يمثل التداخل الجراحي للمريض الراقد.

X_7 (التدخين): متغير وصفي يمثل هل ان المريض من المدخنين او لا،

X_8 (الضـــغط العالي في الدم): متغير كمي يتم تحويله الى وصــفي يمثل نسـبة ضــغط المريض يقاس بوحدة(ملم زئبق) إذ تتراوح النسب الطبيعية بين (120-130ملم/زئبق).

X_9 (نسبة السكر في الدم): متغير كمي يتم تحويله الى وصـفي يقاس بوحدة(mmol/g) إذ تتراوح النســـب الطبيعية بين (85-126g/m) قبل الاكل لمدة ثمان ســـاعات وبين (165-200g/m) بعد الاكل بساعتين.

X_{10} (وزن المريض): متغير كمي يقاس بوحدة (كغم)

وتم اختبار البيانات الحقيقية (المتغيرات التوضيحية) عن طريق أختبارات حسن المطابقة (Goodness of fit tests) :-

اختبار مربع كاي (Chi square test) [Chi]

إن مبدأ هذا الاختبار هو تحديد الفرق بين التكرارات المشاهدة والتكرارات المتوقعة فاذا كان الفرق صغيراً جدا يتحقق على ضوءه مطابقة البيانات ، وان الصيغة الرياضية لهذا الاختبار يمكن توضيحها بالمعادلة الاتية :

إذ أن $\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$

O_i تمثل التكرار المشاهد للفئة i

E_i تمثل التكرار المتوقع للفئة i

m تمثل عدد الفئات

حيث يتم اختبار الفرضية الآتية:

$H_0: \chi^2 = 0$
$H_1: \chi^2 \neq 0$

جدول (3 - 6) قيمة اختبار χ^2 للمتغيرات التوضيحية

المتغيرات	Siq.	P-Value
X1	0.000	0.05
X2	0.003	0.05
X3	0.000	0.05
X4	0.000	0.05
X5	0.000	0.05
X6	0.000	0.05
X7	0.020	0.05
X8	0.000	0.05
X9	0.000	0.05
X10	0.000	0.05

3-8 تحليل النتائج التطبيقية (*Analysis of Applied Results*)

يتضمن هذا القسم عرض النتائج التطبيقية ومن ثم تحليلها للوصول الى مدى ملائمة ودقة البيانات الحقيقية مع انموذج الانحدار الثنائي الذي تم تقديره عن (*log-logistic*) طريق اجراء الاختبارات الخاصة بمقدرات الانموذج ، وفيما يأتي سوف يتم عرض

كافة النتائج التطبيقية التي تم الحصول عليها باستعمال برنامج ($\mathcal{MATLAB}$) والتي نبينها في جداول الملاحق وفيما يأتي الجداول التي تم تحليل البيانات واظهار النتائج بحسب التسلسل وكالاتي :

جدول (3-7) يمثل المعلمات المقدرة وكذلك الخطأ المعياري للمتغيرات التوضيحية كافة بطريقة مربع كاي المحسنة ($\mathcal{MCSE.GA}$)(إعداد الباحث)

المعلمات ($\hat{\beta}_i$)	المعلمات المقدرة	الخطأ المعياري SE($\hat{\ }_i$)	نسبة (Z) $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$	المعنوية
$\hat{\beta}_0$	0.061	0.295	0.206	Non-sig
$\hat{\beta}_1$	0.0323	0.459	0.071	Non-sig
$\hat{\beta}_2$	2.040	0.650	3.137	Sig
$\hat{\beta}_3$	0.331	0.634	0.522	Non-sig
$\hat{\beta}_4$	0.342	0.475	0.719	Non-sig
$\hat{\beta}_5$	0.041	0.142	0.286	Non-sig
$\hat{\beta}_6$	-3.945	22.402	-0.176	Non-sig
$\hat{\beta}_7$	-0.289	0.589	-0.490	Non-sig
$\hat{\beta}_8$	0.332	0.662	0.502	Non-sig
$\hat{\beta}_9$	-0.294	0.877	-0.336	Non-sig
$\hat{\beta}_{10}$	-0.023	0.117	-0.198	Non-sig

كما مبين آنفاً في الجدول (3- 6) تم تقدير المعلمات وقيم الخطأ المعياري لكل معلمة تم تقدير ها بطريقة تصـــغير مربع كاي الجينية ($\mathcal{MCSE.GA}$) ، إذ تم عمل نفس الخطوات والاختبارات التي تم أجرائها بطريقة المربعات الصـــغرى الموزونة الاعتيادية ($\mathcal{WLSE}$) . لذلك يتم تقدير المعلمات لأهمية وأختبار التأثير الكلي في المتغيرات التوضيحية ، و عن طريق النتائج التي حصلنا عليها نلحظ أن نسبة (Z) التي تتبع توزيع الطبيعي القياسي وبمستوى دلالة ($\alpha = 0.05$) وعند مقارنتها مع

قيمة (Z) الجدولية $1.96 = (0.05)\, Z_{\frac{1}{2}}$ نحصــل على العوامل التي يكون لها تأثير وكذلك التي تكون غير مؤثرة في الأنموذج المقدر ،

ومن تجربة المحاكاة في الجدول نشــاهد أن المعلمات المقدرة($\widehat{\beta}_0$ و $\widehat{\beta}_1$ ، $\widehat{\beta}_3$، $\widehat{\beta}_4$، $\widehat{\beta}_5$، $\widehat{\beta}_6$، $\widehat{\beta}_7$، $\widehat{\beta}_8$، $\widehat{\beta}_9$، $\widehat{\beta}_{10}$) تمت أزالتها لأنها لم يكن لها تأثير على الانموذج المقدر عكس ($\widehat{\beta}_2$)
معامل الانحدار الذي أختص بعامل التدخين هو العامل الوحيد الذي كان له تأثير في أنموذج (*log-logistic*) الثنائي وذات تأثير معنوي كبير ، لذلك كان لابداً لنا من أعادة تقدير المعلمات وقيم الانحراف المعياري وكذلك نســبة (Z) .إذ أن المتغيرات التي ليس لها تأثير في حالة حدوث الوفاة للمصابين بأمراض القلب هي (الحالة الاحتماعية ،الجنس ، الردهة ، العمر، درجة العملية ، سبب الرقود ، السكر ، الوزن ، الضغط) كل هذه المتغيرات لم يكن لها تأثير في حالة حدوث الوفاة .

نلحظ كذلك ان معامل أنحدار متغير التدخين هي قيمة موجبة و هذا يدل على ان كلما اقترب متغير التدخين من (0) التي تعبر عن حالة (غير المدخن) يتجه نحو متغير الأســتجابة وذلك لأخذ القيمة (0) وهي (عدم حدوث الوفاة) إذ من خلال هذه القيمة يتبين لنا انه كلما زاد متغير التدخين وأقترب من القيمة (1) يؤدي ذلك الى زيادة أحتمالية حدوث الوفاة وبمعدل (2.040) ، وبذلك يمكننا الحصول على القيم التقديرية لأنموذج الانحدار (*log-logistic*) الثنائي المبين في الجدول (2) في الملاحق التي تم أستخراج القيم فيها وحسب الصيغة الاتية :

$$\widehat{\pi}(x) = \frac{e^{(\widehat{\beta}_2 x_{i2})}}{1 + e^{(\widehat{\beta}_2 x_{i2})}}$$

$$\widehat{\pi}(x) = \frac{e^{(2.040 x_{i2})}}{1 + e^{(2.040 x_{i2})}}$$

وكذلك حصـلنا على قيمة معامل التحديد ($R^2 = 0.41$) من حيث الأعتماد على أعادة تحليل البيانات للمتغير التوضــيحي المعنوي وكذلك تحليل التقديرات لهم ، وهذا يثبت لنا ان متغير التدخين لوحده فقط يفسر نسبة (41%) من كل الأختلافات في الاستجابات لمشاهدات العينة لحالات (عدم حدوث الوفاة) وهذه نسـبة كبيرة تبين أهمية (عامل التدخين) الذي يُعد المتغير في تحديد مستقبل حياة الانسان المصاب بأمراض القلب .

جدول (3-8) قيمة المعلمات المقدرة وألخطا المعياري لجميع المتغيرات التوضيحية بطريقة المربعات الصغرى الموزونة ($WLSE$) (إعداد الباحث).

المعلمة $(\hat{\beta}_i)$	المعلمات المقدرة	الخطأ المعياري SE($\hat{\ }_i$)	نسبة (Z) $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$	المعنوية
$\hat{\beta}_0$	0.013	0.008	1.596	Non-sig
$\hat{\beta}_1$	-0.007	0.019	-0.383	Non-sig
$\hat{\beta}_2$	0.816	0.038	21.285	Sig.
$\hat{\beta}_3$	0.026	0.036	0.702	Non-sig
$\hat{\beta}_4$	0.078	0.021	3.819	Sig.
$\hat{\beta}_5$	0.004	0.002	2.367	Sig.
$\hat{\beta}_6$	-1.011	2.031	-0.498	Non-sig
$\hat{\beta}_7$	-0.072	0.031	-2.284	Sig.
$\hat{\beta}_8$	0.096	0.039	2.426	Sig.
$\hat{\beta}_9$	-0.096	0.069	-1.365	Non-sig
$\hat{\beta}_{10}$	-0.001	0.001	-0.440	Non-sig

إذ نشاهد في الجدول آنفاً (3-7) إن تقدير المعلمات يكون ضرورياً وذلك لأختيار أهمية التأثير الكلي في المتغير التوضيحي عن طريق فرضية العدم وكذلك الفرضية البديلة إذ:

$H_0: \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_{10} = 0$

$H_1: \hat{\beta}_1 \neq \hat{\beta}_2 \neq \cdots \neq \hat{\beta}_{10} \neq 0$

لذلك في المعادلة (2-50) تم توليد أختبار ($Wald$) عن طريق استعمال نسبة (Z) وعرض النتائج التي حصلنا عليها من الجدول آنفاً (3-7) إذ تمت مقارنتها مع قيمة (Z) الجدولية إذ $Z_{\frac{1}{2}}(0.05)=1.96$ و $Z_{\frac{1}{2}}(1-0.05) = -1.96$ (1.96 $\leq Z \leq$ -1.96) والعمود السادس نلحظ أن القيمة اقل من 0.05 الذي يكون متغيراً قيد الاختبار معنويا إذ يمثل $\hat{\beta}_5$ معنوية المتغيرات التوضيحية على حالة المصاب بالمرض .لذلك سوف يتم أزالة المتغيرات التي ليس لها تأثير بشكل كبير في الانموذج المقدر وهي (درجة العملية ، الضغط ، العمر ، سبب الرقود ، الحالة الاجتماعية) وعليه فان المقدرات التي سيتم أزالتها من الجدول هي ($\hat{\beta}_0$ $\hat{\beta}_1$$\hat{\beta}_3$ $\hat{\beta}_6$ $\hat{\beta}_9$$\hat{\beta}_{10}$) أما المتغيرات الاخرى التي لها تأثير في حالة

الوفاة بالنسـبة للمصـابين الذين يعانون من امراض القلب هي (الجنس ، السـكر ، الردهة ،الوزن ،كذلك التدخين) المتمثلة بالمقدرات الاتية ($\hat{\beta}_2$ $\hat{\beta}_4$ $\hat{\beta}_5$ $\hat{\beta}_7$ $\hat{\beta}_8$) لذلك كان لا بداً من أعادة تقدير المعلمات وكذلك قيم الانحراف المعياري ونسبة Z للأنموذج المقدر .

نلاحظ من الجدول ان اشارات المقدرات مثل معامل تأثير متغير الردهة يكون

$$\hat{\pi}(x) = \frac{e^{(\hat{\beta}_2 x_{i2}+\hat{\beta}_4 x_{i4}+\hat{\beta}_5 x_{i5}+\hat{\beta}_7 x_{i7}+\hat{\beta}_8 x_{i8})}}{1+e^{(\hat{\beta}_2 x_{i2}+\hat{\beta}_4 x_{i4}+\hat{\beta}_5 x_{i5}+\hat{\beta}_7 x_{i7}+\hat{\beta}_8 x_{i8})}}$$

جدول (3-9) تقدير معلمات أنموذج الانحدار (*Log-Logistic*)الثنائي بطريقة المربعات الصـغرى الموزونة الأعتيادية وطريقة أصغر مربع كاي الجينية (إعداد الباحث)

Estimated Parameters	Binary Log-Logistic Regression Analysis	
	WLSE	MCSE.GA
$\hat{\beta}_0$	0.013	0.061
$\hat{\beta}_1$	-0.007	0.032
$\hat{\beta}_2$	0.816	2.040
$\hat{\beta}_3$	0.026	0.331
$\hat{\beta}_4$	0.078	0.342
$\hat{\beta}_5$	0.004	0.041
$\hat{\beta}_6$	-1.011	-3.945
$\hat{\beta}_7$	-0.071	-0.289
$\hat{\beta}_8$	0.096	0.332
$\hat{\beta}_9$	-0.095	-0.294
$\hat{\beta}_{10}$	-0.001	-0.023

بما أن طريقة $(\mathcal{WLSE})$ وكذلك طريقة ($\mathcal{MCSE.GA}$) هما أفضـــل الطرائق من إذ تقدير الأنموذج المقدر تم تطبيق مقدرات أنموذج الانحدار $(\mathcal{Log\text{-}Logistic})$ الثنائي عن طريق تحديد مـــتـــجـــه كـــل مـــعـــلـــمـــة ($\hat{\beta}_0 \quad \hat{\beta}_1\ \hat{\beta}_2 . . . \hat{\beta}_{10}$) في الجدول آنفاً (3-8) وتم تعويض هذه المعلمات مع المتغيرات التوضيحية وتم الحصول على مقدرات الأنموذج $(\mathcal{Log\text{-}Logistic})$ الثنائي وعن طريق طريقة $(\mathcal{WLSE})$ يقل مجموع مربعات الخطأ $(\mathcal{MSE})$ الى اصـغر ما يمكن . اما بالنسـبة لطريقة تصغير مربع كاي الجينية $(\mathcal{MCSE.GA})$ كذلك تم اتخاذ الخطوات نفسها لأستخراج مقدرات الانموذج .كما موضح في الجدول (3) وجدول (5) من الملاحق .

جدول (3-10) تصـنيف البيانات للعينة عن طريق اسـتعمال الأنموذج المقدر بطريقة تصـغير مربع كـــاي الجينية ($\mathcal{MCSE.GA}$) (إعداد الباحث)

حالة المريض		التنبؤ		
		المجموع	حدوث الوفاة 1 $\hat{\pi} \geq 0.5$	عدم حدوث الوفاة0 $\hat{\pi} < 0.5$
المشاهدة	عدم حدوث الوفاة0=Y	38	6	32
	حدوث الوفاة1=Y	52	51	1
	المجموع	90	57	33
دقة النموذج		91.39		
حساسية النموذج		95.64		
نسبة التصنيف الصحيح		97.93		

تم أيجاد قيم انموذج الانحدار $(\mathcal{Log\text{-}Logistic})$ الثنائي التي تم توضـــيحها في جدول (2) في الملاحق عن طريق تصــنيف المتغيرات المعنوية للأنموذج إذ تم تصــنيف (41) مصـــاباً من أصل 48 ممن ليس لديهم حالة وفاة تصنيفا صحيحاً . وعليه بلغت نسبة التصنيف الصحيحة لحالة عدم حدوث الوفاة (89%) للذين تم اصـــابتهم بامراض القلب وكذلك تم تصــنيف (63) من اصـل 64 من الذين حدثت لهم الوفاة وبلغت النسـبة لهذه الحالة 98% ويرجع السـبب في هذا الى طبيعة الدراسـة (فيما يتعلق بحالة حدوث الوفاة) ، إذ يمكننا التنبؤ بشكل صـحيح بناء على المتغيرات التوضيحية التي يعاني منها المصاب بامراض القلب وحدثت معهم حالة الوفاة . إذ كانت نسـبة التصـنيف مرتفعة وبلغت 95% و هذا يدل على جودة الانموذج المقدر إذ تعد

هذه النســبة مقبولة جداً وهذا يدل على ان نسـبة الخطأ تســاوي 5% من العدد الكلي الذين تم تصنيفهم بشكل خاطئ .

كذلك نلحظ أن أنموذج الانحدار $(Log\text{-}Logistic)$ الثنائي يكون أفضـل عند أسـتعمال طريقة تصــغير مربع كاي المحســنة $(MCSE.GA)$. عن طريقة المربعات الصــغرى الموزونة $(WLSE)$ ويكون اكثر ملائمة لنمذجة حالات المرضـى المصـابين وذلك من إذ بقائهم على قيد الحياة او سـبب الوفاة . ويدل على دقة $(MCSE.GA)$ إذ تتطابق القيمة المقدرة مع القيمة الحقيقية بشكل معقول في تقدير هذه المعلمات إذ يمكن تقدير الأنموذج بدرجة عالية من الدقة . عن طريق البيانات الحقيقية لمرضـى القلب في الجدول (3-5) وكذلك البيانات المقدرة في الجداول (2)(3)(4)(5) . وكـذلك عن طريق قيمة معامل التحديد التي تم ايجادها في الطريقتين بأستعمال $(WLSE)$ $(MCSE.GA)$.

جدول (3-11) تصـنيف البيانات للعينة عن طريق اسـتعمال الأنموذج المقدر بطريقة المربعات الصـغرى الموزونة $(WLSE)$. (إعداد الباحث)

حالة المريض		التنبؤ		
		المجموع	حدوث الوفاة 1 $\hat{\pi} \geq 0.5$	عدم حدوث الوفاة0 $\hat{\pi} < 0.5$
المشاهدة	عدم حدوث الوفاةY=0	38	2	36
	حدوث الوفاةY=1	52	49	3
	المجموع	90	51	39
دقة النموذج		94.51		
حساسية النموذج		97.87		
نسبة التصنيف الصحيح		96.47		

من الجدول آنفاً نلحظ ان انموذج الانحدار $(Log\text{-}Logistic)$ الثنائي قام بتصـنيف وترتيب البيانات (المتغيرات المعنوية) التي لها تأثير في حالة وفاة المريض فقط وذلك عن طريق ايجاد القيم الاحتمالية المقدرة للانموذج المقدر كما سـوف يتم توضـيحه عن طريق جدول (4) في الملاحق .إذ تم تصـنيف 62 مصـاباً من اصـل 64 ممن حدثت معهم حالة الوفاة وبلغت نسـبة التصـنيف الصـحيحة لحالة حدوث الوفاة 97% من المصـابين بامراض القلب وكذلك تم تصـنيف 43 مصـاب من اصـل 46 ممن ليس

لديهم حالة وفاة تصنيفا صحيحا ، وبلغت نسبتهم ما يقارب فا 94% لحالة عدم حدوث الوفاة للمصـــابين بامراض القلب والسـبب في ذلك هو حســاسـية طبيعة الدراســة (ما يتعلق بحالة حدوث الوفاة للمريض) . وعليه يمكننا التنبؤ بطريقة صـحيحة بناء على المتغيرات التوضـيحية الموجودة فيها ، وكانت نسـبة التصـنيف الصـحيحة بصـورة عامة 96% وهذه نســبة مقبولة جدا إذ تكون بمعدل خطا (4%) للأشــخاص الذين تم تصنيفهم بشكل خاطأ عن طريق أنموذج الانحدار (*Log-Logistic*) الثنائي .

الفصل الرابـــــع

الاستنتاجـــــــات

و

التوصيـــــات

في هذا الفصل سوف يتم عرض الاستنتاجات والتوصيات التي توصل اليها الباحث على ضوء ما تم دراسته وبحثه في الجانبين التجريبي وكذلك التطبيقي.

1-4 الاستنتاجـــــات (*Conclusions*)

بحسب ما تم تحليله من البيانات والحصول على النتائج في الجانب التجريبي والتطبيقي وما تم استخراجه من تجارب المحاكاة فقد تم التوصل الى الاستنتاجات والتوصيات الأَتية :

1- الجانب التجريبي والقيم الافتراضية للمعلمات كافة في الانموذج وقيم المتغيرات التوضيحية ومتغير الاستجابة وأنموذج الانحدار (*Log-Logistic*) التي تم افتراضه عن طريق أمتلاكه أقل متوسط مربعات خطا (*MSE*) من طرائق التقدير لمقدرات الانموذج (*Log-Logistic*) يمكن اجمالها بالنقاط الاتية :

أ- ان طريقة تصغير مربع كاي الجينية (*MCSE.GA*) هي افضل الطرائق التحسين لتقدير المعلمات أنموذج الانحدار (*Log-Logistic*) الثنائي للأنموذج ولجميع حجوم العينات ، وذلك عن طريق التقديرات التي تم أجرائها .

ب- عن طريق استعمال كافة طرائق التقدير الاعتيادية وكذلك المحسنة نلحظ تناقص قيمة المعيار الاحصائي (متوسط مربعات الخطأ (*MSE*) وذلك بأزدياد حجم العينة عند ايجاد مقدرات أنموذج (*Log-Logistic*) الثنائي وهذا ما يطابق النظرية الاحصائية إذ كلما يزداد حجم العينة يقل الوقوع في الخطأ .

ت- أن طرائق التقدير المحسنة أفضل من طرائق التقدير الاعتيادية كما موضح في تجربة المحاكاة وذلك للحجوم العينات كافة التي تم افتراضها وتقدير المعلمات لأنموذج الانحدار (*Log-Logistic*) الثنائي .

ث- أن طريقة المربعات الصغرى الموزونة هي افضل طرائق التقدير الاعتيادية وهذا يعني ان البيانات لا خطية .

ج- من بين كل طرائق التقدير الاعتيادية التي تم التقدير بها أثبتت طريقة (*WLSE*) كفاءتها على باقي الطرائق الاعتيادية كونها الطريقة الوحيدة التي لا تحتاج الى قيم بدائية عند تقدير معلمات الانموذج بها إذ هذا ما يميزها عن البقية وكان لها الافضلية.

2- اما الجانب التطبيقي تم التوصل الى الأتي :

أ- أظهرت نتائج التحليل أن اكثر العناصر تأثير على متغير الأستجابة (أمراض القلب) هو عامل (التدخين) .

ب- أن متغير التدخين يعد متغيراً مهما من المتغيرات التي سببت حالة الوفاة وكذلك أول عامل لتحديد الوفاة وهذا ما تم اثباته وأظهاره في نتائج حسب طريقة (*WLSE*) وكذلك طريقة (*MCSE.GA*) إذ اشارت النتائج التي تم التوصل اليها

في الدراسة أنها مطابقة مع الدراسات المعمول بها سابقا في المتغيرلت التوضيحية التي لها أكثر تأثير في حالة المريض ، وكذلك هذا ما تم تأكيده من الاطباء المتخصصين بهذا المجال .

ت- أن طريقة مربع كاي الجينية (*MCSE*) أفضل من طريقة (*WLSE*) الاعتيادية وعن طريق اختبار جدول التصنيف نستنتج ان الطريقتين لهما القدرة على التصنيف على الرغم من ان طريقة المربعات الصغرى الموزونة (*WLSE*) اقل كفاءة من طريقة (*MCSE.GA*) ، مع ذلك يمكننا أستعمال اي منهما لتصنيف الحالات المرضية الجديدة للمصابين بأمراض القلب والذي يؤدي ذلك الى (حدوث وفاة ، عدم حدوث وفاة) ويكون اعتماد هذه الحالات على قيم المتغيرات التوضيحية للمريض .

ث- أن التدخين هو العامل الاكثر تأثير على أمرض القلب وفق النتائج التي حصلنا عليها بطريقة تصغير مربع كاي الجينينة (*MCSE.GA*) في تحليل البيانات على انموذج الانحدار (*Log-Logistic*) الثنائي على متغير الاستجابة (حالة المصابين بأمراض القلب عن طريق النتائج التي ظهرت لنا في تجربة المحاكاة .

2-4 التوصيات والدراسات المستقبلية (*Recommendations and future studies*)

على ضوء الاستنتاجات التي تم التوصل إليها يوصي الباحـــث بما يأتي :

1- يوصي الباحث أجراء دراسات مستقبلية تتناول الاعتماد على طرائق التقدير الاعتيادية (*MCSE،WLSE،MLE*) وكــــذلك الطرائــق المحسنة بالخوارزميـــة الجينية (**MLE. GA**) و (**WLS.GA**) و (**MCSM.GA**) التي تم تقدير المعلمات الانموذج عن طريقهما والمقارنة بينهم في حالة وجود قيم شاذة (*outliers*) في البيانات التي سوف نجري عليها تجربة المحاكاة .

2- هناك عدة خوارزميات اخرى مثل خوارزمية الشبكات العصبية (*neural network Algorithm*) وخوارزمية محاكاة التلدين (*Annealing simulation algorithm*) وكذلك خوارزمية مستعمرة النمل (*Ant colony algorithm*) ، إذ يمكن أستعمال هذه الخوارزميات المتطورة مع الطرائق الاعتيادية الكلاسيكية التي تم أستعمالها في البحث للمقارنة فيما بينهم عن طريق (*MSE*) للتوصل الى أفضل الطرائق في التقدير معلمات النماذج .

3- أوصي بأستعمال طريقة مربع كاي المحسنة (*MCSE.GA*) في حالة كانت البيانات تتوزع توزيع برنولي (*Bernoulli*) وذلك لتقدير معلمات أنموذج الانحدار (*Log-Logistic*) الثنائي .

4- في أنموذج الانحدار (*Log-Logistic*) الثنائي المتعدد وكذلك الترتيبي يمكننا أستعمال الطرائق التي تم التقدير معلمات بها وعليه يمكن أستعماله في نمذجة البيانات في العلوم الاجتماعية والاقتصادية وغيرها من المجالات الاخرى .
5- هناك العديد من المتغيرات الاخرى التي لم تؤخذ بعين الاعتبار وذلك لعدم توفر المعلومات الكافية عنها مثل وظيفة الشخص وتأثير الكحول وكذلك الكولسترول وتناول الاطعمة غير الصحية وغيرها من العوامل التي تؤثر في حياة الانسان وبالاخص الاشخاص المصابين بأمراض القلب .
6- ينبغي توظيف (*Genetic Algorithm*) مع طرائق التقدير الحصينة بدلا من طرائق التقدير الاعتيادية الكلاسيكية .
7- يمكن أستعمال الخوارزمية الجينية مع الطرائق التي تم التقدير بيها في مجالات اخرى غير مجال طب مثلا على بيانات تخص الصناعة او المجال التكنولوجي او حالة مؤثرة في المجتمع لها تأثير واضح في حياة الأنسان وغيره من المجالات الاخرى .

المصادر

1. Abdulqader, Q. M., (2015),"Comparison Of Discriminant Analysis and Logistic Regression Analysis " ,Zakho Tecnical Institute Duhok, Polytechnic University Duhok, Duhok , Iraq, *INAS*, pp.34-46.
2. Agresti, A., 2002. An Introduction to Categorical Data Analysis, Department of Statistics, University of Florida, Gainesville, Florida. ISBN : 978-0-471-22618-5
3. Akkus ,Ö. , Demir , E. , (2016), " Comparison Som Classical And Meta-Heuristic Optimazation Techniques in The Estimation Of The Logit Model Parameters", *JAR*, pp.1026-1042.
4. Al-Bayati, M. M., Muhammad, M. J. (2021). Using the Jacknife method to estimate the logistic regression model for breast cancer. *Journal of Economics and Administrative Sciences*, *27*(126), 571-582.
5. Alrahamneh, A. and Hawamdeh, O. ,(2017) , "The Factors Affecting Eye Patients (Cataract) In Jordan by Using the Logistic Regression Model", *MAS*, pp. 38-42.
6. Cramer, J. S., (2003),"Logit Models From Economics and other fields", Cambridge University Press cape Town, New York, *ISBN*, pp.33-45.
7. Czepiel, S. A., (2002) ,"Maximum Likelihood Estimation Of Logistic Regression Models " ,*MLMLRM*, Retrieved from , *https://czep.net/stat/mlelr.pdf*, pp.1-23.
8. Demir , E. , Akkus , Ö., (2015), " An Introductory Study on How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model", *IJS:BAR*, pp.162-180.
9. Ferrer, J. ,Wang, L., (1999)," Comparing the Classification Accuracy among ,Parametric Nonparametric, Discriminant Analysis and Logistic Regression Methods", *ERIC* , Quebec, Canada , pp.1-23.
10. Gayou, O., (2008)," A genetic algorithm for variable selection in logistic regression analysis of radiotherapy treatment outcomes", American Association of Physicists in Medicine, pp. 5426 – 5433.
11. Gelfand, A., and Smith, A., (1990)," Sampling Based Approaches to Calculating Marginal Densities", *JASA*, pp. 398-409.
12. Geman, S., and Geman, D. ,(1984)," Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*", IEEE*, pp. 721-741.
13. Goldberg, D. E., and Deb, K., (1991)," A Comparative Analysis of Selection Schemes Used in Genetic Algorithms ,Foundation of Genetic Algorithms", San Francisco: CA: Morgan Kaufmann, pp. 69-93.
14. Groenwald , P. and Mokgatlhe, L., (2005) , " Bayesian Computation for Logistic Regression", University of Bostswana , *ELSEVIER,* pp. 857-868.
15. Gülder ,K. , Özkm, B. , (2014) ," Categorical Principal Component Logistic Regression:A Case Study for Housing lo Approval" , *ELSEVIER*, pp. 730 – 736.

16. **Hadji, S. & et al. ,(2015)," Theoretical and experimental analysis of genetic algorithms based mppt for PV systems", *ELSEVIER*, pp. 772-787.**
17. **Henry , A. ,(2013), "Bayesian Logistic Regression Modelling Via Markov Chain Monte Carlo Algorithm ", University of Cape Coast , *JSTS* , pp. 193 – 197.**
18. **Hussain, J. N. and Nassir, A. J., (2015)," Cluster Analysis as a Strategy Of Grouping to Construct Goodness – Of – Fit Tests when the Continuous Covariates Present in the Logistic Regression Models", *BJMCS*, pp. 1-16.**
19. **Husssain, J. N, (2007), "An Overview Of Evaluation Criteria In Logistic Regression Model", *ICMS*, pp. 1-7.**
20. **Johnson, P., Graham, P, (2014)," Genetic algorithm with logistic regression for alzheimer's disease diagnosis and prognosis" , *NARI*, pp. 455-456.**
21. **Karim , A. J., (2013),"Statistical Study to Classify β-Thalassaemia diseases in Erbil City at Thalassaemia Center by Using ROC Curve Analysis", *IJSR*, pp.2955-2961.**
22. **Lee, K.H., Kim, K.W.,(2015),"Performance comparisons of particle swarm optimization and genetic algorithm for inverse surface radiation problem", *ELSEVIER*, PP. 330-337.**
23. **Liu, H., Ong, C., (2008), "Variable selection in clustering for marketing segmentation using genetic algorithms", *ELSEVIER*, pp. 502-510.**
24. **Magalhaes- M., J. ,(2013), "A comparative study of crossover operators for genetic algorithm to solve the job shoscheduling problem", *WSEAS*, pp. 164-173.**
25. **Mahdavi, I., Paydar, M., ,(2008)," Genetic alogorithm approach for solving acell formation problem in cellubar manufacturing", *ELSEVIER*, pp. 1-7.**
26. **Mahmoud Riad Mahmouda , A. A. EL- Sheikhb , Naglaa A. Moradc , Moshera A. M. Ahmad,(2016)" Log-Beta Log-Logistic Regression Model ",International Journal of Sciences: Basic and Applied Research (IJSBAR)**
27. **Matilainen K. & ,(2013)," Employing A Monti Carlo Algorithm in Newton-Type Methods For Maximum Likelihood Estimation of Genetic Parameters", *PLOSONE* , pp.1-7.**
28. **McCarthy,F. William(2007)," The Existence of Maximum Likelihood Estimates for the Binary Response Logistic Regression Model ", Collection of Biostatistics Research Archive,26.**
29. **Mccullagh, P., & Nelder, J., (1983)," Generalized Linear Models", London: Chapman and Hall.**
30. **Mehejabeen Mahbub and Most. Fatima-Tuz-Zahura(2017) "Factors Affecting Postnatal Care in Bangladesh: Clustered Data Analysis ",Department of Statistics, Dhaka University, Dhaka- 1000, Bangladesh, 66(1): 59-65**
31. **Menard ,S.,(2002),"Applied Logistic Regression Analysis",2nd Edition Thousand Oaks , CA : SAGE Publications , *Series Quantitative Applications in the Social Sciences*.**

32. Meng ,Q., Weng, J., (2010) , " Agenetic Algorithm approach To assessing Work Zonecasualty Risk " , *ELSEVIER* , pp. 1283-1288.
33. Misra, A. ,(2013)," Portfolio optimization of commercial Banks- An Application of Genetic Algorithm", *EJBM*, pp. 120-129.
34. Mitchell ,M., (1999) ," An Introduction to Gentic Algorithms ", London , England fifth printing: Abradford Book the MIT Press Cambridge Massachusetts.
35. Pasia, J., Hermosilla, A., (2005)," A useful tool for statistical estimation genetic algorilhm", *JSCS*,pp. 237 – 251.
36. Raghupathikumar, D., and Raja, K., (2012)," genetic algorithm based scheduling of an input queued switch",*IJCA*, pp.37-42.
37. Sefian , S., Benbouziane , M. ,(2012) ,"Portfolio Selection Using Genetic Algorithm" , *MPRA* , pp. 1-12
38. Soderstrom, R. , Leitner، W. (1997)." The Effects of Base Rate, Selection Ratio, Sample Size and Reliability of Predectors on Predective Efficiency Indic. Associated with Logistic Regression Models", *ERIC*, PP.1-25.
39. Spiegelhalter, D., Best, N., (2002)," Bayesian Maesures of Model complexity and Fit", *JRSS*, pp. 583-639.
40. Sukono, Sholahuddin, A. (2014)," Credit Scoring for Coopera of Financial Services Using Logistic Regression Estimated by Genetic Algorithm", *AMS*, pp. 45-57.
41. Tektas, D., & Gủnay, S. ,(2008)," A Bayesian Approach to Parameter Estimation in Binary Logit and Probit Models", *HJMS*, pp. 167-176.
42. Vandewater, L., ,(2008) ,"An adaptive genetic algorithm for selection of blood-based biomarkers for prediction of Alzheimer's disease progression" , *BMC* , pp.1-10.
43. Wuensch, K. , (2014) , " Binary Logistic Regression with SPSS" , Retrieved from *WWW . Care . ecu . edu / psyc/ Wuensch / MV /Logistic SPSS*, pp.1-29.
44. Yoder, S. E., (2009)," An Investigation on the use and flcxibility of Genetic Algorithm for Logistic Regression", Clemson University , TP, pp. 1-77.
45. Zager, S., & Karim, R., (1991),"Generalized Linear Models with Random Effects : A Gibbs Sampling Approach", *JASA* , pp. 79-86.

الملاحق

جدول رقم (1) (*تعريف متغير الأستجابة وكذلك المتغيرات التوضيحية*)

رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة
Y	متغير الاستجابة	0	احتمال عدم حدوث الوفاة
		1	احتمال حدوث الوفاة
X_1	متغير وصفي يمثل نوع الردهة التي يرقد بها المريض	1	ردهة القلب المفتوح
		2	ردهة انعاش القلب
		3	ردهة انعاش و الاوعية
		4	ردهة التمريض الخاص
X_2	متغير وصفي يمثل جنس المريض الراقد	1	ذكر
		2	انثى
X_3	متغير وصفي يمثل الحالة الزوجية	1	اعزب
		2	متزوج
		3	ارمل او ارمله /مطلق او مطلقة
X_4	متغير كمي يمثل عمر المريض بالسنوات		
X_5	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	1	ذبحة صدرية
		2	احتشاء واعتلال عضلة القلب
		3	اضطرابات في توصيل الدم
		4	توقف او تسارع او رجفان القلب
		5	عجز القلب
		6	جلطة دماغية
		7	انسداد شرايين
		8	اضطراب في جهاز الدوران
		9	وذمة رئوية
		10	تشوهات خلقية ولادية
X_6	متغير وصفي يمثل التداخل الجراحي للمريض الراقد	1	عدم اجراء عملية
		2	اجراء عملية وسطى
		3	اجراء عملية كبرى
		4	اجراء عملية فوق الكبرى
X_7	متغير وصفي يمثل التدخين	1	مدخن
		2	غير مدخن
X_8	متغير وصفي يمثل ضغط المريض	1	(100-149)
		2	(150-199)
		3	(200-249)
X_9	متغير وصفي يمثل سكر المريض	1	(75-179)
		2	(180-284)
		3	(285-389)
		4	(390-494)
X_{10}	متغير كمي يمثل وزن المريض		

جدول 2

القيم المقدرة للانموذج (*log-logistic*) الثنائي للمتغيرات التي لها تأثير(المعنوية) فقط بأستعمال طريقة تصغير مربع كاي المحسنة (*MCSE.GA*) في تقدير المعلمات للبيانات الحقيقية

Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$
1	1	0.508089	31	1	0.508089	61	1	0.508089
2	1	0.508089	32	0	0.524250	62	0	0.524250
3	1	0.508089	33	0	0.508089	63	0	0.524250
4	1	0.516174	34	0	0.516174	64	0	0.524250
5	1	0.524250	35	0	0.524250	65	0	0.524250
6	1	0.524250	36	0	0.524250	66	0	0.516174
7	1	0.508089	37	0	0.524250	67	0	0.508089
8	1	0.508089	38	0	0.524250	68	0	0.524250
9	1	0.508089	39	0	0.508089	69	0	0.508089
10	1	0.508089	40	0	0.524250	70	0	0.524250
11	1	0.508089	41	0	0.524250	71	0	0.524250
12	1	0.508089	42	0	0.508089	72	0	0.508089
13	1	0.508089	43	0	0.508089	73	0	0.508089
14	1	0.508089	44	0	0.508089	74	0	0.524250
15	1	0.532314	45	0	0.508089	75	0	0.508089
16	1	0.532314	46	0	0.524250	76	0	0.524250
17	1	0.508089	47	0	0.508089	77	0	0.508089
18	1	0.508089	48	0	0.508089	78	0	0.524250
19	1	0.508089	49	0	0.508089	79	1	0.508089
20	1	0.508089	50	0	0.508089	80	1	0.508089
21	1	0.508089	51	0	0.508089	81	1	0.508089
22	1	0.508089	52	0	0.524250	82	1	0.508089
23	1	0.508089	53	1	0.508089	83	1	0.508089
24	1	0.508089	54	1	0.508089	84	1	0.508089
25	1	0.508089	55	1	0.508089	85	1	0.508089
26	1	0.508089	56	1	0.508089	86	1	0.508089
27	1	0.508089	57	1	0.508089	87	1	0.508089
28	1	0.508089	58	1	0.508089	88	1	0.508089
29	1	0.508089	59	1	0.508089	89	1	0.508089
30	1	0.516174	60	1	0.508089	90	1	0.508089

جدول 3

القيم المقدرة للمتغيرات التوضيحية كافة (معنوية وغير معنوية) بأستعمال طريقة تصغيرمربع كاي الجينية المحسنة ($\mathcal{MCSE.GA}$) في تقدير المعلمات للأنموذج المقدر

Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$
1	1	0.893838	31	1	0.959318	61	1	0.950400
2	1	0.928881	32	0	0.429198	62	1	0.922764
3	1	0.894126	33	0	0.480412	63	0	0.609115
4	1	0.821805	34	0	0.374163	64	0	0.607103
5	1	0.914173	35	0	0.489084	65	0	0.591035
6	1	0.929219	36	0	0.393001	66	0	0.509747
7	1	0.897563	37	0	0.452604	67	0	0.455864
8	1	0.844409	38	0	0.640106	68	0	0.596469
9	1	0.853749	39	0	0.838495	69	0	0.513535
10	1	0.920593	40	0	0.396136	70	0	0.405882
11	1	0.858423	41	0	0.555149	71	0	0.579168
12	1	0.925457	42	0	0.470966	72	0	0.480689
13	1	0.923000	43	0	0.575304	73	0	0.577056
14	1	0.935749	44	0	0.472112	74	0	0.493812
15	1	0.976651	45	0	0.497720	75	0	0.400528
16	1	0.957286	46	0	0.734792	76	0	0.483310
17	1	0.922364	47	0	0.288580	77	0	0.816634
18	1	0.907161	48	0	0.363260	78	0	0.377719
19	1	0.974284	49	0	0.790302	79	1	0.801286
20	1	0.980707	50	0	0.673589	80	1	0.931868
21	1	0.949322	51	0	0.833008	81	1	0.821217
22	1	0.979864	52	0	0.550329	82	1	0.974713
23	1	0.936299	53	1	0.955693	83	1	0.975818
24	1	0.963043	54	1	0.963915	84	1	0.924594
25	1	0.946801	55	1	0.958328	85	1	0.852938
26	1	0.940768	56	1	0.944540	86	1	0.921034
27	1	0.929715	57	1	0.969779	87	1	0.910946
28	1	0.845732	58	1	0.957610	88	1	0.924346
29	1	0.928869	59	1	0.929659	89	1	0.88769
30	1	0.857894	60	1	0.926583	90	1	0.950305

جدول 4

القيم المقدرة للمتغيرات التوضيحية التي لها تأثير (المعنوية) فقط بأستعمال طريقة المربعات الصغرى الموزونة (*WLSE*) في تقدير معلمات الأنموذج (*log-logistic*) الثنائي

Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$	Obs.	Y_i	$\hat{\pi}(X_i)$
1	1	0.893164	31	1	0.899053	61	1	0.873409
2	1	0.902161	32	0	0.739906	62	1	0.888250
3	1	0.875797	33	0	0.749020	63	0	0.768908
4	1	0.881813	34	0	0.746555	64	0	0.828104
5	1	0.881813	35	0	0.746555	65	0	0.778153
6	1	0.891450	36	0	0.741579	66	0	0.769794
7	1	0.897660	37	0	0.746555	67	0	0.758811
8	1	0.875393	38	0	0.767765	68	0	0.787753
9	1	0.878681	39	0	0.871971	69	0	0.764347
10	1	0.895031	40	0	0.753068	70	0	0.756412
11	1	0.888086	41	0	0.750654	71	0	0.749837
12	1	0.888961	42	0	0.746555	72	0	0.762658
13	1	0.878147	43	0	0.778153	73	0	0.776535
14	1	0.889099	44	0	0.756412	74	0	0.761078
15	1	0.909585	45	0	0.772008	75	0	0.756294
16	1	0.896204	46	0	0.798508	76	0	0.757095
17	1	0.888454	47	0	0.751873	77	0	0.867029
18	1	0.885629	48	0	0.756412	78	0	0.761078
19	1	0.900555	49	0	0.874369	79	1	0.875797
20	1	0.911350	50	0	0.788412	80	1	0.898248
21	1	0.902161	51	0	0.871971	81	1	0.880458
22	1	0.914125	52	0	0.757095	82	1	0.889967
23	1	0.894766	53	1	899389.0	83	1	0.903499
24	1	0.893932	54	1	0.901343	84	1	0.901583
25	1	0.901012	55	1	0.900555	85	1	0.880458
26	1	0.889967	56	1	0.900555	86	1	0.889967
27	1	0.904066	57	1	0.892489	87	1	0.875811
28	1	0.796006	58	1	0.901389	88	1	0.891490
29	1	.8974020	59	1	0.884299	89	1	0.884064
30	1	0.885629	60	1	0.888681	90	1	0.893809

جدول 5

Obs.	Yi	$\hat{\pi}(X_i)$	Obs.	Yi	$\hat{\pi}(X_i)$	Obs.	Yi	$\hat{\pi}(X_i)$
1	1	0.715544	31	1	0.732444	61	1	0.720531
2	1	0.741097	32	0	6 0.4808	62	1	0.708669
3	1	0.692503	33	0	0.492662	63	0	0.517972
4	1	0.695734	34	0	0.485525	64	0	0.623979
5	1	0.706142	35	0	0.493522	65	0	0.531136
6	1	0.715174	36	0	0.481680	66	0	0.502027
7	1	0.733860	37	0	0.488207	67	0	0.512955
8	1	0.683363	38	0	0.517460	68	0	0.544500
9	1	0.703006	39	0	0.674942	69	0	0.515803
10	1	0.729051	40	0	0.481974	70	0	0.500719
11	1	0.703667	41	0	0.494877	71	0	0.494607
12	1	0.715869	42	0	0.492679	72	0	0.509367
13	1	0.701592	43	0	0.534775	73	0	0.529176
14	1	0.712903	44	0	0.495195	74	0	0.503944
15	1	0.760653	45	0	0.522173	75	0	0.500657
16	1	0.734504	46	0	0.560708	76	0	0.499455
17	1	0.719472	47	0	0.487449	77	0	0.666650
18	1	0.716312	48	0	0.503692	78	0	0.469991
19	1	0.749164	49	0	0.683835	79	1	0.688771
20	1	0.770890	50	0	0.555579	80	1	0.736914
21	1	0.750318	51	0	0.679610	81	1	0.696648
22	1	0.775470	52	0	0.503830	82	1	0.741136
23	1	0.720411	53	1	0.738280	83	1	0.745862
24	1	0.728035	54	1	0.739514	84	1	0.717967
25	1	0.734461	55	1	0.739285	85	1	0.690710
26	1	0.717386	56	1	0.737915	86	1	0.718168
27	1	0.740128	57	1	0.742956	87	1	0.695307
28	1	0.581381	58	1	0.751961	88	1	0.725807
29	1	0.726076	59	1	0.713418	89	1	0.710598
30	1	0.701752	60	1	0.717832	90	1	0.727570

القيم المقدرة للمتغيرات التوضيحية كافة (المعنوية والغير معنوية) وذلك عن طريق أستعمال طريقة المربعات الصغرى الموزونة (*WLSE*) في تقدير معلمات أنموذج (*log-logistic*) الثنائي

Abstract

In this thesis, one of the most important non-linear regression models was studied, which is the binary log-logistic model, which is used in modeling and estimating any statistical applications, and then the parameters of this model were estimated by statistical estimation methods, but when estimating its parameters, a special problem (P+1). And if numerical methods appeared to us. When the number of parameters are used to estimate its parameters, sometimes these methods do not give a better solution, because they depend directly on the primitive capabilities of the model, and accordingly we will use and employ the usual methods of estimation after improving them through the followers of one of the modern algorithms genetic algorithm) in order to suit this type A non-linear regression model for estimating its parameters. Then we compare all estimation methods. In order to choose the best methods in terms of estimation by a number of models and different sample sizes in the simulation. And based on the statistical standard (mean square error (MSE)). The comparison was made between the usual estimation methods, which included (the method of greatest possibility. And the method of chi-square minimization, the method of weighted least squares) and in the contrasts the improved estimation methods by (Genetic Algorithm), which included the method of estimating the improved maximum possible (MLE.GA) and the method of estimating the improved chi-square minimization (MCSE.GA) as well as the method of stimating the weighted least squares Enhanced (WLSE.GA). In general, it was concluded that the (WLSE) method is the best among all the usual methods for estimating the model arameters. The (MCSE.GA) method was the best among all the improved estimation methods for estimating the binary (log-logistic) model, and the reason for that is due to These two methods had the lowest (MSE) in the simulation program for all abilities compared to the rest of the methods. As for the practical side, real data was used for a sample of (90) patients with heart disease. The data was modeled and the parameters of this model were estimated and the best method was chosen, which was reached in The experimental side through the occurrence of real deaths of the injured with the estimated deaths, where it was shown the appropriateness of the binary (log-logistic) model in modeling these data and extracting the main cause of death is (smoking), as well as it was found that the (WLSE) method (MCSE.GA) They are accurate in estimating the parameters of the binary (log-logistic) model.

University of Karbala
College of Administration and Economics
Department of statistics

Estimation of log-logistic Regression Parameters Using Genetic Algorithm with Practical Application

A Thesis Submitted

Council of the College of Administration and Economics at the University of Karbala, which is part of the requirements for obtaining a master's degree in Statistics

Written by
Hussein Khalil Obaid Mikhlif

Supervised by

Prof. Dr. Mushtaq Kareem Abdel Rahem

2022 A.D   Karbala   1444 A.H