



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

تقدير نموذج انحدار لوجستي لمتغير معتمد متعدد
المستويات

مع تطبيق عملي

رسالة مقدمة إلى

مجلس كلية الإدارة والاقتصاد / جامعة كربلاء وهي جزء من متطلبات

نيل درجة ماجستير في علوم الإحصاء

من الباحثة

تقى عبد محمد الخزعلي

أشرف

أ.د عبد الحسين حسن الطائي

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَإِذْ تَعْدُوا

نِعْمَةَ اللَّهِ لَا تَحْصُوهَا إِنْ لَمْ يَغْفُرْ

رَحِيمٌ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سورة النحل

الآية ﴿18﴾

الإهداء

- إلى من قدم للبشرية أسمى معانيها ... الرسول الكريم محمد (صلى الله عليه وآله).
- إلى من أذهب الله عنهم الرجس وطهرهم تطهيرا .. الأئمة الأطهار (عليهم السلام).
- إلى من ساندتني في صلاتها ودعاؤها ... إلى من سهرت الليالي تنير دربي
- إلى من تشاركني أفراحي وأحزاني ... إلى نبع العطاء والتضحية الذي لا ينهي .
- إلى أجمل ابتسامة في حياتي إلى أروع امرأة في الوجود ... (أمي الغالية).
- إلى من أحمل اسمه بكل افتخار إلى الشهيد ... (والدي الحبيب).
- إلى من افتقد وجوده في حياتي إلى الشهيد ... (زوجي العزيز).
- إلى سندي وعزوتي ... (أخواتي وإخوتي).
- إلى ريحانة قلبي وقرّة عيني وروح فؤادي ابنتي الغالية ... (ريّان).
- إلى الإنسانية التي تحمل كل معاني الإنسانية والحب ... (خالتي العزيزة).
- وإلى كل من دعم وساهم في انجاز هذا الرسالة اهديه هذا الجهد المتواضع وله مني كل الامتنان .

الباحثة

شكر وتقدير

الحمد لله الذي أنار لنا درب العلم والمعرفة وأعاننا على إداء هذا الواجب ووفقنا إلى إنجاز هذا العمل،
يطيب لي بعد أن اتمت بحشي هذا ومن الواجب والعرفان أن أتقدم بجزيل الشكر
والامتنان إلى الأستاذ الدكتور (عبد الحسين حسن الطائي) لأشرفه على هذه الرسالة ومتابعته
المستمرة لي بالنصح والإرشاد والتوجيه، وكان لإشرافه العلمي الدقيق وتوجيهاته السديدة وما جاءت به
نفسه الكريمة من خلق عالي، الأثر البالغ في إظهار الرسالة وإنجازها على نحو نافع ومفيد، وتبقي
المفردة اللغوية قاصرة عن التعبير عما في داخلي من مشاعر الود والاحترام اتجاهه وفقه الله لكل خير.
كما أتقدم بالشكر والامتنان للأستاذة الأفاضل لجنة المناقشة المحترمين (رئيساً وأعضاء) على تفضلهم
بقبول مناقشة هذه الرسالة وما سيبدونه من ملاحظات وراء علمية قيمة تسهم في اغناء هذه الرسالة.
ولا يفوتني أن أقدم الشكر والتقدير (لأساتيد قسم الإحصاء) لما قدموه لي من العلم والمعرفة بكل
تفان وإخلاص على مدى السنوات السابقة.
شكري وتقديري إلى كادر (شعبة الدراسات العليا) في الكلية.
كما وإن الشكر موصول إلى زملائي (طلبة الدراسات العليا) في قسم الإحصاء لما لمستهم منهم من
روح التعاون والمساندة متمنية لهم التوفيق والسداد.

الباحثة

قائمة المحتويات

رقم الصفحة	العنوان
أ	الآية
ب	الإهداء
ت	شكر وتقدير
ي-ك	المستخلص
د-ز	قائمة المحتويات
ح-ط	قائمة الجداول
ح-ط	قائمة الأشكال
8-1	الفصل الأول : منهجية الرسالة
1	المقدمة
2	مشكلة الرسالة
3	هدف الرسالة
8-3	الاستعراض المرجعي
56-9	الفصل الثاني : الجانب النظري
9	التمهيد
10	الانحدار اللاخطي
10	البيانات المصنفة
14	نماذج الانحدار للمتغيرات المعتمدة الوصفية
15	أنموذج الانحدار اللوجستي
15	مجالات استعمال الانحدار اللوجستي
15	شروط تطبيق الانحدار اللوجستي
15	أنواع الانحدار اللوجستي
16	أنموذج الانحدار اللوجستي ثنائي الاستجابة
18	أنموذج الانحدار اللوجستي المتعدد
18	بعض المشاكل الخاصة بالمتغير المعتمد الوصفي

20	نماذج دوال التحويل الخطي	8-2
21	انموذج اللوغاريتم النسبة المضافة	1-8-2
23	طرائق تقدير أنموذج الانحدار اللوجستي	9-2
24	طريقة تقدير الإمكان الأعظم	1-9-2
25	الشبكات العصبية الاصطناعية	2-9-2
26	لمحة تاريخية عن الشبكات العصبية	1-2-9-2
26	الشبكات العصبية الاصطناعية	2-2-9-2
27	خصائص الشبكات العصبية الاصطناعية	3-2-9-2
27	مفاهيم تقنيات الشبكة العصبية الاصطناعية	4-2-9-2
28	مكونات الشبكات العصبية الاصطناعية	5-2-9-2
29	الصياغة الرياضية للعصبون	6-2-9-2
30	هيكلية الشبكة العصبية الاصطناعية	7-2-9-2
31	الشبكات وحيدة الطبقة	1-7-2-9-2
32	الشبكات متعددة الطبقات	-7-2-9-22
33	انواع الشبكات العصبية الاصطناعية	8-2-9-2
34	الشبكات العصبية عكسية الانتشار	-8-2-9-21
36	تدريب الشبكات العصبية الاصطناعية	9-2-9-2
36	التعلم المراقب	-9-2-9-21
37	التعلم غير المراقب	-9-2-9-22
37	طريقة تقدير الخوارزمية الجينية	3-9-2
37	مفهوم الخوارزمية الجينية	1-3-9-2
40	خطوات عمل الخوارزمية الجينية	2-3-9-2
47-40	مكونات الخوارزمية الجينية	3-3-9-2
47	خطوات إجراء الخوارزمية الجينية لتقدير المعلمات أنموذج الانحدار اللوجستي المتعدد	4-3-9-2

53-48	طريقة المربعات الصغرى الموزونة	4-9-2
53	اختبار جودة التوفيق لأنموذج الانحدار اللوجستي	10-2
53	اختبار Wald	1-10-2
54	اختبار هوزمر_ليمشو	2-10-2
55	معايير المقارنة بين طرائق التقدير المستعملة	11-2
55	معيان متوسط مربعات الخطأ	1-11-2
56	معيان متوسط المطلق الخطأ	2-11-2
89-57	الفصل الثالث : الجانب العملي (التجريبي- التطبيقي)	
57	تمهيد	1-3
57	القسم الاول - الجانب التجريبي	2-3
57	مفهوم المحاكاة	1-2-3
58	توليد المتغيرات العشوائية	2-2-3
58	مراحل بناء ووصف تجربة المحاكاة	3-2-3
73-58	تحليل نتائج تجربة المحاكاة	4-2-3
73	القسم الثاني - الجانب التطبيقي	3-3
73	وصف التجربة	1-3-3
77	تحليل بيانات التجربة	2-3-3
87-80	تقدير معلمات أنموذج الانحدار اللوجستي متعدد الاستجابة	3-3-3
89-88	الفصل الرابع : الاستنتاجات و التوصيات	
88	الاستنتاجات	1-4
89	التوصيات	2-4
95-90	المصادر	
107-96	الملاحق	
A	المستخلص باللغة الانجليزية (Abstract)	

قائمة الجداول

رقم الصفحة	عنوان الجدول	الجدول
11	الاستجابات المصنفة	(1-2)
13	القيم الاحتمالية	(2-2)
59	يوضح القيم الافتراضية	(1-3)
61	(قيم الوزن مع الحد الثابت للشبكات العصبية عند حجوم العينات (125,75,25) حسب قيم الافتراضية (5.0-,3.0,4.0) فضلا عن معايير المفاضلة.	(2-3)
63	(قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدره باستعمال (الأماكن الاعظم ، المربعات الصغرى الموزونة، الخوارزمية الجينية) عند حجوم العينات (125,75,25) حسب قيم الافتراضية (5.0-,3.0,4.0) فضلا عن معايير المفاضلة	(3-3)
66	قيم الوزن مع الحد الثابت لشبكات العصبية عند حجوم العينات (125,75,25) حسب قيم الافتراضية (4.0,3.0-, 2.0)فضال عن معايير المفاضلة.	(4-3)
67	قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدره باستعمال (الأماكن الاعظم، المربعات الصغرى الموزونة، الخوارزمية الجينية) عند حجوم (125,75,25) حسب قيم الافتراضية (2.0-,4.0,3.0) فضلا عن معايير المفاضلة:	(5-3)
70	قيم الوزن مع الحد الثابت لشبكات العصبية عند حجوم العينات (125,75,25) حسب قيم الافتراضية (5.0,6.0-,2.0) فضلا عن معايير المفاضلة.	(6-3)
71	قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدره باستعمال (الأماكن الاعظم ، المربعات الصغرى الموزونة، الخوارزمية الجينية) عند حجوم العينات (125,75,25) حسب قيم الافتراضية (5.0,6.0-,2.0) فضلا فضلا عن معايير المفاضلة	(7-3)

74	الاستجابة الاولى للقيم الحقيقية والمتوقعة للتقيس بيض حشرة عثة التمر	(8-3)
75	الاستجابة الثانية للقيم الحقيقية والمتوقعة للمعدل اوزان يرقات حشرة عثة التمر	(9-3)
75	الاستجابة الثالثة للقيم الحقيقية والمتوقعة للبروتينات الكلية في السوائل الجسمية ليرقات حشرة عثة التمر	(10-3)
76	الاستجابات الكلي للقيم الحقيقية والمتوقعة	(11-3)
79	قيم احتمال الاستجابات الحقيقية لقيم اللوغاريتم النسبة المضافة $\log(Y^*)$	(12-3)
81	نتائج تقدير انموذج الانحدار اللوجستي وفق الاستجابات الحقيقية باستخدام الشبكات العصبية المتمثل بالوزن	(13-3)
83	نتائج تقدير انموذج الانحدار اللوجستي وفق الاستجابات الحقيقية باستخدام الأماكن العظم	(14-3)
84	نتائج تقدير انموذج الانحدار اللوجستي وفق الاستجابات الحقيقية باستخدام المربعات الصغرى الموزونة	(15-3)
85	نتائج تقدير انموذج الانحدار اللوجستي وفق الاستجابات الحقيقية باستخدام الخوارزمية الجينية	(16-3)
86	معايير المفاضلة لطرائق التقدير المستخدمة	(17-3)

قائمة الأشكال

رقم الصفحة	عنوان الشكل	الشكل
17	شكل الدالة اللوجستية	(1-2)
21	العلاقة غير خطية بين متغير الاستجابة والمتغيرات التوضيحية (X)	(2-2)
23	العلاقة الخطية بين متغيرات التوضيحية ومتغير الاستجابة الجديد (Logit)	(3-2)

29	شكل مكونات الشبكات العصبية الاصطناعية	(4-2)
30	شكل الخوارزمية الرياضية للعصبون الاصطناعي	(5-2)
31	شكل شبكة وحيدة الطبقة	(6-2)
32	شكل الشبكة العصبية ذات طبقة واحدة وخلية خليفة واحدة	(7-2)
32	شكل الشبكة العصبية ذات طبقة واحدة وخلايا خليفة متعددة	(8-2)
33	مخطط معمارية الشبكة العصبية الاصطناعية النموذجية	(9-2)
34	شكل نموذج لشبكة عصبية ذات تغذية أمامية.	(10-2)
45	شكل العبور بنقطة واحدة	(14-2)
45	شكل العبور بنقطتين	(15-2)
81	شكل الاوزان والحد الثابت للشبكات العصبية	(1-3)
83	شكل علاقة البيانات الحقيقية والاختفاء العشوائية للإمكان الأعظم مع أنموذج الانحدار اللوجستي متعدد الاستجابة	(2-3)
84	شكل علاقة البيانات الحقيقية والاختفاء العشوائية للمربعات الصغرى الموزونة مع أنموذج الانحدار اللوجستي متعدد الاستجابة	(3-3)
85	شكل البيانات الحقيقية للخوارزمية الجينية مع أنموذج الانحدار اللوجستي متعدد الاستجابة	(4-3)

المستخلص: -

شكل علم الاحصاء بدأ ممتدة تقدم خدماتها عبر الزمن الى العلوم الأخرى ورافداً يساهم في تحليل نتائج ما تتوصل اليه الدراسات في شتى المجالات ، تعتبر التجارب المختبرية عنصراً رئيسياً يستند اليه الباحثون في الجانب الحياتي من حيث التحليل والاستنتاج من أجل الوصول الى أبرز العوامل المؤثرة في تحسين كفاءة التجربة وتحديد العوامل المؤثرة على مجريات التجربة وتطويرها والوصول الى النتائج المطلوبة ، في هذه الرسالة تم استخدام الانحدار اللاخطي والتركيز على انحدار المنحني اللوجستي.

تكمن مشكلة الرسالة في كيفية التعامل مع النموذج انحدار اللوجستي متعدد المستويات في عملية نمذجة البيانات والوصول الى الطرائق التقدير المثلى في تقدير المعلمات

تهدف الرسالة الى تقدير انموذج الانحدار اللوجستي لمتغير معتمد (متعدد المستويات) (Y_{ijk}) , اعتماداً على متغيرات وصفية لمعرفة مدى تأثيرها على المتغير المعتمد (محل الدراسة) , حيث تم استعمال طرائق تقدير الكلاسيكية منها وهي (طريقة الامكان الأعظم [MLE] وطريقة المربعات الصغرى الموزونة [WLS]) وأخرى ذكية وهي (طريقة الشبكات العصبية الاصطناعية [ANN] وطريقة الخوارزمية الجينية [GA]) , ومن ثم المقارنة بين الطرائق لمعرفة الأفضل والأكفاً من بينهما بغية الحصول على المقدرات الكفوة للاعتماد عليها في الوقوف على أبعاد ومتغيرات الظاهرة (محل الدراسة) .من خلال استعمال مقاييس المقارنة منها متوسط مربعات الخطأ (MSE) ومتوسط الخطأ المطلق (MAE).

عملياً ولتحقيق أهداف الرسالة تم اجراء تجربة المحاكاة (Simulation) باستخدام طريقة مونت كارلو (Monte Carlo) لحجوم عينات مختلفة (صغيرة ،متوسطة، كبيرة) بأحجام (125،75،25) ثلاث تجربة التطبيق العملي للوصول الى الطريقة المثلى ومعرفة التغيرات التي تطرأ في عملية التجريب ومن ثم تم تلخيص نتائج تجربة المحاكاة ومن مجريات تجربة المحاكاة تم استنتاج ان طريقة الشبكات العصبية الاصطناعية هي الأفضل من بين الطرائق المستخدمة في عملية التقدير , وكذلك تم التطبيق على بيانات حقيقية تخص تجربة مختبرية حياتية لحشرة (عثة التمر) وتم استعمال جميع الطرق المذكورة انفاً في عملية التقدير في الجانب التطبيقي ايضاً وليس فقط في الجانب التجريبي بغية مطابقة طرائق التقدير حيث اظهرت نتائج الجانب التطبيقي تفوق طريقة الشبكات العصبية الاصطناعية أيضاً من بين

قريناتها وهذا ما يتوافق مع الجانب التجريبي مما يدل على ملائمة طريقة التقدير مع انموذج الانحدار اللوجستي .

وأخيرا من اهم الاستنتاجات التي التوصيل تفوق طريقة الشبكات العصبية الاصطناعية ANN في الجانب التجريبي والتطبيقي على باقي الطرائق المستخدمة، كما أظهرت الرسالة إمكانية تحويل النموذج اللوجستي من النموذج لا خطي الى النموذج الخطي.



الفصل الأول

منهجية الرسالة

الفصل الأول

منهجية الرسالة

Introduction

(1-1) المقدمة :

يعد علم الإحصاء من العلوم المهمة التي تلعب دوراً أساسياً في تطور العلوم الأخرى وتقدمها ومدى مساهمته في كافة المجالات حيث (لا تقدم بدون تخطيط ولا تخطيط بدون إحصاء) وبالإضافة إلى كونه من أهم وسائل الرسالة العلمي المهمة في عملية جمع البيانات وتحليلها والمعلومات اللازمة من خلال استخدام قواعد وقوانين خاصة. وقد شهدت السنوات الأخيرة كثرة استعمال النماذج الإحصائية الذكية في تحليل البيانات الوصفية لاسيما في المجالات الاجتماعية والزراعية والطبية وغيرها.

إن تحليل الانحدار هو أحد تلك النماذج التي تستخدم في وصف العلاقة بين متغيرين أو أكثر حيث تقسم هذه النماذج إلى قسمين الأول النماذج الخطية و الثاني النماذج اللاخطية، وسوف يتم التركيز في دراستنا هذه على النماذج اللاخطية (أنموذج الانحدار اللوجستي المتعدد)، حيث يعد أنموذج الانحدار اللوجستي المتعدد من النماذج الإحصائية التي شهدت اهتماماً كبيراً وملحوظاً في العديد من الدراسات، وتم استخدامه بشكل واسع في التجارب الحياتية. إذ يعد من أهم الأساليب الإحصائية التي تدرس تحليل البيانات المصنفة و لاسيما في حالة متغير الاستجابة الذي يرتبط مع المتغيرات المستقلة من النوع الأسمي أو الرتبي ويتألف من مستويين فأكثر. والهدف الأساس من تحليل الانحدار اللوجستي هو إيجاد أفضل توفيق للنموذج، حيث يوضح العلاقة بين المتغير التابع والمتغير المستقل (واحد أو أكثر). وينقسم أنموذج الانحدار اللوجستي إلى قسمين الأول هو أنموذج الانحدار اللوجستي الثنائي (Binary Logistic Regression Mode) و يستخدم في حالة كون متغير الاستجابة مؤلف من مستويين فقط والثاني هو أنموذج الانحدار اللوجستي المتعدد (Multiple Logistic Regression Mode) يستخدم في

حالة كون متغير الاستجابة مؤلف من ثلاثة مستويات فأكثر، وسوف يتم التركيز في دراستنا على النوع الثاني من أنموذج الانحدار اللوجستي المتعدد.

ولغرض توضيح فكرة الرسالة بشكل عام فقد تم تقسيم الرسالة الى أربعة فصول ، تناول الفصل الأول المقدمة والمشكلة والهدف من الدراسة والاستعراض المرجعي لبعض الدراسات والبحوث السابقة، اما الفصل الثاني (الجانب النظري) فقد تناول مفهوم الانحدار اللوجستي وانواعه وخصائصه وكذلك تضمن بعض التحويلات الخاصة بالأنموذج وطرائق تقدير أنموذج الانحدار لمتغير معتمد متعدد المستويات, حيث تم استخدام طريقة الإمكان الأعظم (MLE) بالاعتماد على تطبيق طريقة نيوتن رافسون التكرارية وكذلك طريقة الشبكات العصبية (ANN) وأيضا تم استخدام طريقة الخوارزمية الجينية (GA) وطريقة المربعات الصغرى الموزونة (WLS) وسبب اختيار هذه الطرائق هو لغرض المقارنة بين الطرق الذكية والطرق الكلاسيكية من خلال معايير المفاضلة لاختيار أفضل طريقة, وهما معيار متوسط مربعات الخطأ (MSE) ومعيار متوسط مطلق الخطأ (MAE)، اما الفصل الثالث فقد تضمن الجانب العملي لمبنيين: الأول يتضمن الجانب التجريبي (باستخدام المحاكاة) حيث تناول مفهوم المحاكاة والطرائق المستخدمة و توليد المتغيرات العشوائية ومراحل بناء تجربة المحاكاة اما مبحث الثاني فقد تناول الجانب التطبيقي حيث تم استخدام احدى التجارب الحياتية الخاصة (بحشرة عثة التمور المخزونة) في حين تضمن الفصل الرابع اهم الاستنتاجات والتوصيات التي توصلت اليها الباحثة.

Problem of the A Thesis

(2-1) مشكلة الرسالة :

تكمن مشكلة الرسالة في كيفية التعامل مع النموذج انحدار اللوجستي متعدد المستويات في عملية نمذجة البيانات والوصول الى الطرائق التقدير المثلى في تقدير المعلمات.

Objective of the A Thesis (2-1) هدف الرسالة :

توظيف الطرائق الذكية (الشبكات العصبية NN ، الخوارزمية GA) لتقدير
انموذج انحدار اللوجستي لمتغير معتمد متعدد المستويات ومقارنتها مع الطرائق
الكلاسيكية (الإمكان الأعظم، المربعات الصغرى الموزونة ومن ثم تحديد افضل
طريقة للتقدير من خلال المعايير الإحصائية (MAE،MSE)

Reference Review (4-1) الاستعراض المرجعي :

لغرض وضع أسس تقدير أنموذج الدراسة فقد اطلعت الباحثة على مجموعة من الكتب
والبحوث والدراسات التي تناولت موضوع انموذج الانحدار اللوجستي والمشاكل التي تواجهه
وطرائق تقدير معلمات هذا الأنموذج ومن هذه الدراسات والبحوث هي :

❖ في عام (1984) اثبت كل من الباحثين (Albert & Anderson) [19] وجود مقدرات
خاصة لأنموذج الانحدار اللوجستي متعدد الاستجابة للبيانات، حيث تعاني هذه البيانات من
مشكلة الفصل لذلك تم تقسيمها الى ثلاث فئات وهي الفصل التام (Complete
separation) والفصل شبة التام (quasi Complete separation) والتداخل (Overlap) وذلك من خلال اجراء التكرار حيث توصل الباحثان الى ان المقدر
الذي يعظم دالة الإمكان الأعظم ليس له صيغة نهائية والسبب يعود الى استخدام بعض
الطرائق التكرارية للحل كطريقة نيوتن رافسون .

❖ في عام (1992) قدم الباحث (بيثون) [10] اختبار فرضيات خاصة بمعلمات انموذج
الانحدار اللوجستي وذلك من خلال دراسة خصائصه وأسلوب بناء حدود الثقة لتقدير
المعلمات المتعلقة بأنموذج الانحدار اللوجستي .

❖ في عام (2000) قام (الطائي)^[4] بدراسة بعض مواصفات أنموذج الانحدار الخطي المتعدد من خلال عرض وتحليل المتغيرات المعتمدة الوصفية، وقد استخدم طريقة خوارزمية التوقع الأعظم (EM) وطريقة نيوتن رافسون (N.R)، فضلا عن طريقة تحويلات المنحى اللوجستي (Logit) لتقدير معالم أنموذج، من خلال تجربة محاكاة ولثلاثة مستويات من احجام العينات ،وتوصل الباحث باستعمال طريقة مونت كارلو في المحاكاة الى ان مقدرات المنحني اللوجستي هي الأفضل في العينات الصغيرة، بينما كانت طريقة خوارزمية التوقع الأعظم هي الأفضل في العينات المتوسطة ، اما في العينات الكبيرة فقد كانت جميع الطرائق متقاربة ومتكافئة.

❖ في عام (2001) قدم الباحثان (Joel & Savin)^[35] بحثا بوصف تقدير لوجيت (logit) ووحدة الاحتمال (probit) لنماذج ثنائية الاستجابة. لنماذج ثنائية الاستجابة باستخدام الطرق البيرية والتقليدية وتطبيقها على البيانات الطولية وتوصي الدراسة الى استخدام الطرائق الحديثة.

❖ وفي عام (2002) قدم كل من (Time)^[61] وآخرين بحثا باستخدام انموذج الانحدار اللوجستي متعدد الحدود لدراسة إثر العديد من العوامل وتأثيرها في المستوى العلمي للأسرة بالاعتماد على معيارين لاختبار الأنموذج الملائم، حيث يمثل المعيار الأول المعلومات البيزية (Bayesian information Criterion)، والمعيار الثاني يمثل احصاءه نسبة الإمكان، حيث توصلوا الى ان تعليم الاب يعد العامل الأول المؤثر ومن ثم عامل الجنس ومن ثم يأتي بعد ذلك عامل الوراثة .

❖ في عام (2005) اقترح كل من (pasia)^[52] وآخرين بتطبيق الخوارزميات الجينية في تقديرات معالم الانموذج اللوجستي مع انموذج (Gauss) غير الخطي من خلال تطبيق المحاكاة. حيث تم المقارنة مع الطرائق الكلاسيكية وكانت الأفضلية للخوارزميات الجينية كبديل جيد عن الأسلوب العددي تحت انموذج الانحدار اللوجستي .

❖ وبعد عامين (2007) قدم الباحث (Mc carthy)^[43] دراسة حول كيفية استخدام انموذج الانحدار اللوجستي في جدول توافق ثنائي الأبعاد حيث تم استخدام عدة طرائق لتقدير معالم انموذج الانحدار اللوجستي الثنائي لمجموعة من البيانات الوصفية بوجود فئات من نقاط بيانات الفصل التام وشبه التام والتداخل فضلا عن وجود مشكلة تعدد التعداد الخطي

وتوصل الباحث الى ان الإمكان الأعظم في حالة التداخل تكون غير صحيحة ودقيقة في حالة وجود مشكلة الفصل التام وغير التام .

❖ وفي عام (2009) قامت الباحثة (Yoder)^[63] باستعمال المنهجية الخاصة بالخوارزمية الجينية لغرض إنشاء أنموذج للانحدار اللوجستي وبدقة كبيرة. ولكن هذا الأنموذج قد يواجه بعض المشاكل، منها طريقة الاختيار لبعض المتغيرات التوضيحية والتي تسمح بالتحيز لغرض الدخول بداخل الأنموذج الخاص بالانحدار اللوجستي، كما ويمكن للطريقة التكرارية التي تم استخدامها أن تؤدي لقبول النماذج التي تكون دون المستوى، بالتالي فإن أتباع مثل هذه التقنية الأفضل في تقنية التحسين الحالية ستعالج وبشكل فعال هذه المشاكل. وقد تبين ذلك عن طريق النتائج التي قد استحصلت عليها الباحثة والتي تتمحور حول البيانات الخاصة بالحوادث الجوية مع (500) حالة و(13) من المتغيرات التوضيحية .

❖ في عام (2010) قدم الباحث (AL-affi)^[18] بحثا حيث تناول فيه استخدام الانحدار اللوجستي متعدد المستويات ذلك من خلال دراسة تطبيقية أجريت هذه الدراسة على بيانات العنف الجسدي , حيث توصل الباحث الى ان انموذج الانحدار اللوجستي يحدد العلاقة بين متغير الاستجابة ومجموعة من المتغيرات التوضيحية و ذلك وفق انموذج رياضي .

❖ في عام (2011) قام (عبد الرزاق)^[15] باستخدام انموذج الانحدار اللوجستي في دراسة وتحليل بعض المتغيرات على بيانات اصابة اللثة حيث توصل الباحث الى مجموعة من الاستنتاجات والتوصيات واهمها كانت هي إن عامل الترسيبات الكلسية يعد السبب الرئيسي في الإصابة .

❖ وفي العام نفسه قدم كل من (Weng & Meng)^[45] دراسة لتقييم مخاطر مكان العمل حيث طبقوا فيها تحليل الانحدار اللوجستي ومنهجية الخوارزمية الجينية واطهرت النتائج بان أداء الخوارزمية الجينية أفضل من انموذج الانحدار اللوجستي الثنائي .

❖ وفي عام (2012) قدمت الباحثة (مراد)^[8] دراسة مقارنة بين الشبكات العصبية والانحدار الكلاسيكي في التنبؤ بمستويات نتائج بحوث طلبة كلية التربية الرياضية، واثبتت الدراسة وفق المعايير الاحصائية بان طريقة الشبكات العصبية الاصطناعية لها افضلية على طريقة الانحدار الكلاسيكي .

❖ وفي العام نفسه قدم (العابدي)^[5] أسلوب تحليل البيانات المصنفة المترابطة والتي تكون فيها الاستجابات عبارة عن أزواج مترابطة فيما بينها وذلك من خلال دراسة تطبيقية أجريت على معدلات الطلبة المتخرجين وعلاقتها مع معدلات قبولهم في المعهد وذلك من خلال الاعتماد على تصنيف هذه المشاهدات تحت صفتي (Negative, Positive) وبالاعتماد على أعلى معدل للطلاب مع المعدل العام للظاهرة، وتوصل الباحث الى إن زيادة نسبة الترجيح (odds Ratio) تكون أعلى لصالح معدلات التخرج .

❖ وفي عام (2013) قام (Johnson)^[36] وآخرون بتقديم دراسة بتشخيص مرض الزهايمر باستعمال الخوارزمية الجينية والانحدار اللوجستي حيث توصلوا الى أن الخوارزمية الجينية تعطي نتائج أفضل وناجحة الى حد كبير .

❖ في عام (2014) قدمت الباحثة (جبارة)^[11] دراسة لتشخيص أمراض العيون وذلك عن طريق مقارنة بين الدالة التمييزية والانحدار اللوجستي حيث تم وصف البيانات على وفق ثلاثة نماذج وهي انموذج الدالة التمييزية الخطية والثاني انموذج احتمالات الاستجابة لدالة التمييز اما الثالث فهو انموذج الدالة الاحتمالية لدالة الانحدار اللوجستي متعدد الاستجابة. حيث تمت المقارنة بين هذه النماذج على أساس احتمال خطأ التصنيف للوصول الى أفضل انموذج لتشخيص أنواع امراض العيون حيث توصلت الباحثة الى أن طريقة تصنيف دالة التمييز الخطي تعطي اقل احتمال للخطأ .

❖ وفي العام نفسه قدم الباحثان (حمدي و سلمان)^[12] دراسة مقارنة بين الشبكات العصبية الاصطناعية وانموذج الانحدار اللوجستي الثنائي للتمييز بين دخل الأسرة وكانت النتائج تشير الى أن الشبكات تعطي نتائج أفضل من انموذج الانحدار اللوجستي الثنائي .

❖ وفي عام (2015) قدم كل من (Akkus & Demir)^[26] دراسة مقارنة بين الخوارزمية نيوتن-رافسون (NR) والخوارزمية الجينية (GA) حول كيفية تنفيذها في تقدير معاملات الأنموذج اللوجستي الثنائي على مرض الثعلبية حيث توصلوا الى أن الخوارزمية الجينية تعطي نتائج أكثر مرونة في الحل من خوارزمية نيوتن رافسون (NR) في تقدير المعلمات المجهولة للوصول الى الحل الأمثل .

❖ في عام (2016) أجرى كل من (زعلان وعبد الرزاق)^[13] دراسة حول تحليل أثر الضغط النفسي على الإصابة بضغط الدم باستعمال الأنموذج اللوجستي. حيث تم تقدير معلمات الأنموذج بطريقة الإمكان الأعظم على مجموعة من المرضى المصابين في قضاء الزبير معتمدين على مقياس هولمز وراهي لقياس ضغط الدم وقد توصلوا الى ان هناك تأثير طردي للضغط النفسي على الإصابة بضغط الدم من خلال اختباري (T,Wald) .

❖ في عام (2017) قام الباحثان (Hawamdh & Alrahammed)^[20] بدراسة لتصنيف مرضى العيون المصابين بالماء الابيض والعوامل المؤثرة على المرض في الأردن على عينة عشوائية مؤلفة من 116 مريض بتطبيق أنموذج الانحدار اللوجستي حيث تم اقتراح طريقة الامكان الأعظم لتقدير معلمات الانموذج وأثبتت النتائج التي ظهرت ان انموذج الانحدار اللوجستي هو فعال وممثلا للبيانات بنسبة (100 %) عن طريق اختبارات (R square Nagelkerke) و (likelihood Ratio) و (Hosmer- Lemeshow) .

❖ وفي العام نفسه قدم الباحث (الحيالي)^[2] دراسة بعنوان "اختيار أفضل طرائق تقدير لمتغير معتمد متعدد الاستجابة" حيث تم استخدام عدة طرائق للتقدير كطريقة الإمكان الأعظم (MLE) وطريقة المربعات الصغرى الموزونة (WLS) وطريقة الجاكنيف (Jackknife) . حيث تم المقارنة بين هذه الطرائق بالاعتماد على بعض مقاييس ومنها مقياس متوسط مربعات الخطأ (MSE) واطهرت النتائج أن الأفضلية كانت للإمكان الأعظم لحصولها على أقل خطأ .

❖ في عام (2019) قامت الباحثة (الرديني)^[3] بدراسة تطبيقية بعنوان " تقدير معلمات انموذج الانحدار اللوجستي الثنائي باستعمال الخوارزمية الجينية" حيث تم توظيف بعض الطرائق الاعتيادية بعد تحسينها من خلال اتباع منهجية الخوارزمية الجينية في التقدير. ولقد توصلت الباحثة بشكل عام الى أن طريقة المربعات الصغرى الموزونة (WLSE) قد جاءت بالمرتبة الأولى من بين الطرائق الاعتيادية وطريقة الخوارزمية الجينية المحسنة (MESE.GA) هي الأفضل من بين طرائق التقدير المحسنة لغرض تقدير معلمات انموذج الانحدار اللوجستي الثنائي .

* تتميز هذه الدراسة عن الدراسات السابقة بأجراء مقارنة بين الطرائق الكلاسيكية المتمثلة بـ(طريقة الإمكان الأعظم MLE وطريقة المربعات الصغرى الموزونة WLS) والطرائق الذكية المتمثلة بـ(طريقة الشبكات العصبية NN وطريقة الخوارزمية الجينية GA) لتقدير أنموذج الانحدار اللوجستي لمتغير معتمد متعدد الاستجابة على حد علم الباحثة، لذلك تم استعمال هذه الطرائق الذكية في هذه الرسالة والتي ميزتها عن البحوث والدراسات السابقة.



الفصل الثاني

الجانِب النظري

الفصل الثاني

الجانب النظري

Preamble

(1-2) التمهيد:

يعرف الانحدار بأنه التحليل الذي يختص بدراسة اعتماد متغير و احد يعرف بالمتغير التابع (Dependent Variable) أو متعدد الاستجابة على متغير واحد أو أكثر يعرف بالمتغير المستقل أو المتغيرات المستقلة (المفسرة) (Independent Variable) وذلك بغرض التقدير أو التنبؤ بمتوسط قيمة المتغير التابع بمعلومية المتغيرات المستقلة (المفسرة). بناء على ذلك فإن أسلوب الانحدار يستخدم للتوصل الى أنموذج رياضي يوضح العلاقة الكمية بين المتغير التابع والمتغيرات المفسرة، ينقسم الانحدار بصورة عامة الى الانحدار الخطي والذي يكون على نوعين الانحدار الخطي البسيط والانحدار الخطي المتعدد، والقسم الثاني هو الانحدار اللاخطي .

في هذا الفصل سيتم التركيز على الانحدار اللاخطي (أنموذج الانحدار اللوجستي) من حيث الصيغة العامة والافتراضات الخاصة به وخصائصه وأنواعه، وسيتم التطرق أيضاً الى نماذج التحويل التي من خلالها يتم تحويل أنموذج اللوجستي الى الانحدار الخطي وأخيراً سيتم دراسة طرائق تقدير معاملات أنموذج الانحدار اللوجستي بالطرائق الكلاسيكية طريقة الإمكان الأعظم Maximum Likelihood Method وطريقة المربعات الصغرى الموزونة Weighted Least Squares Method مقارنة بالطرائق الذكية الشبكات العصبية Network Neural والخوارزمية الجينية Genetic Algorithm.

Nonlinear Regression

(2-2) الانحدار اللاخطي [16]:

يعد الانحدار اللاخطي شكلا من أشكال تحليل الانحدار ،حيث يتم نمذجة البيانات من خلال توظيف مجموعة من المعلمات غير الخطية وتعتمد على واحد وأكثر من المتغيرات المستقلة. اذ يعد أنموذج الانحدار اللاخطي من النماذج ذات الأهمية العالية بالرغم من ندرة الدراسات التي تتعلق به مقارنة بالنماذج الخطية الا أنه ذو تطبيقات واسعة في الدراسات التطبيقية والطبيعية ،حيث توجد أنواع كثيرة من النماذج اللاخطية ومن هذه النماذج نموذج الانحدار اللوجستي، النموذج الاسي ،النموذج التربيعي و التكعيبيالخ التي تهتم بدراسة البيانات المصنفة.

Categorized Data

(3 -2) البيانات المصنفة [42][49][16] :

ان عملية تحليل البيانات المصنفة تعتمد بالشكل الأساسي على عناصر المجاميع الجزئية وعلى طبيعة الاستجابة، اذ تهتم البيانات المصنفة بدراسة المتغيرات التي يمكن أن تصنف فيها الاستجابات بشكل صفات مثل (موافق /غير موافق) (جيد/غير جيد)...الخ وكذلك تهتم بدراسة التخصصات الطبية حول الإصابة بمرض معين (خفيفة /متوسطة /شديدة) الخ ، ولهذه البيانات المصنفة توزيعات إحصائية مختلفة مثلا (Poisson ،Binomial dist.) (Multinomial dist.،dist).

وإن تحليل هذه البيانات يعتمد على جداول التوافق وهي عبارة عن أزواج مرتبة بشكل تكرارات في جداول حيث تحتوي تلك الجداول على متغيرين أو أكثر وتكون هذه المتغيرات كمية مثل الطول والوزن ودرجة الحرارة أو نوعية مثل لون الشعر ولون العين.... الخ. وتكون هذه الجداول على أنواع مختلفة منها جداول ثنائية الأبعاد وجداول متعددة الأبعاد وسوف يتم التركيز في دراستنا على النوع الأخير من الجداول حيث يستعمل هذا النوع من الجداول في حالة متغيرين فأكثر (حيث يضم كل متغير بداخلة عدة مستويات) ولهذه الجداول أهمية بالغة في التخصصات الطبية ،ويمكن وصف التوزيع التكراري للاستجابات المصنفة بالجدول (1-2).

الجدول (1-2) يمثل الاستجابات المصنفة

أصناف الاستجابة المجاميع الجزئية	t_1	t_2 t_j	Total
a_1	r_{111} r_{112} r_{113}	r_{121} r_{122} r_{123}	... r_{1jm} ... r_{1jm} ... r_{1jm}	N_1
a_2	r_{211} r_{212} r_{213}	r_{221} r_{222} r_{223}	... r_{2jm} ... r_{2jm} ... r_{2jm}	N_2
\vdots	\vdots \vdots \vdots	\vdots \vdots \vdots	\vdots
A_i	r_{i11} r_{i12} r_{i13}	r_{i21} r_{i22} r_{i23}	r_{ij1} r_{ij2} r_{ijm}	N_i
Total	N.11 N.12 N.13	N.21 N.22 N.23	... $N_{.jm}$... $N_{.jm}$.. $N_{.jm}$	N...

الجدول من اعداد الباحثة بالاعتماد على مصدر [4]

حيث تمثل الصفوف المجاميع الجزئية والتي يعبر عنها بأنّها عينات سحبت من مجتمع متعدد القيم وباحتمالات مختلفة استناداً الى بعض القيود المتعلقة بتوزيع متعدد الحدود (Multinomial distribution), اما الاعمدة فهي تمثل أصناف الاستجابة خلال الجدول (1-2) والمعبر عنها بـ r_{ijm} حيث أنّ:

i : عدد مستويات العامل الأول (a).

j : عدد مستويات العامل الثاني (t).

m : رقم الاستجابة.

وعليه يمكن تعريف صفات الاستجابة بالآتي:

1...r : الاستجابة (الصفة) الأولى .

2...r : الاستجابة (الصفة) الثانية .

3...r : الاستجابة (الصفة) الثالثة .

ويمكن تعريف المتغيرات التوضيحية أعلاه:

t : متغير المستخلص الأول والذي يشمل عدة مستويات

a : متغير المستخلص الثاني ويمثل نوعيات مختلفة من المستخلصات المسلطة والذي يشمل عدة مستويات

ويمكن عرض الجدول (2-1) باستخدام القيم الاحتمالية (احتمال استجابة العامل (n) لصف
الاستجابة (m) كما موضح بالجدول (2-2)

أصناف الاستجابة العامل	t ₁	t ₂	t _j	Total
a ₁	π ₁₁₁ π ₁₁₂ π ₁₁₃	π ₁₂₁ π ₁₂₂ π ₁₂₃	..π _{1jm} ..π _{1ij} ...π _{1jm}	π ₁
a ₂	π ₂₁₁ π ₂₁₂ π ₂₁₃	π ₂₂₁ π ₂₂₂ π ₂₂₃	...π _{2ij} ..π _{2ij} ...π _{2jm}	π ₂
⋮	⋮ ⋮ ⋮	⋮ ⋮ ⋮	⋮
a _i	π _{i11} π _{i12} π _{i13}	π _{i21} π _{i22} π _{i23}	...π _{ij1} ..π _{ij2} ...π _{ijm}	π _m
Total	π _{.11} π _{.12} π _{.13}	π _{.21} π _{.22} π _{.23}	...π _{.jm} ...π _{.jm}π _{.jm}	π _{...}

الجدول من اعداد الباحثة بالاعتماد على مصدر [4]

π_{ij}: يمثل احتمال استجابة المتغير المعتمد التي تم الحصول عليها تحت تأثير المستوى (i) من العامل المؤثر الأول والمستوى (j) من العامل المؤثر الثاني وتعرف بانها نسبة الجزء الى الكل ويمكن التعبير عنها بالمعادلة (1-2):

$$\pi_{ij} = \frac{r_{ijm}}{N_i} \dots \dots \dots (1-2)$$

حيث r_{ijm} يمثل وحدات التجربة للعامل (i) والتي استجابت للصف (j) وإنَّ

$$N_i = \sum_{j=1}^n r_{ijm}$$

وحيث أنَّ كل مجموعة أو كل عامل هو موضوع خاص لصف الاستجابة عند ذلك فإنَّ π_{ij} سيمثل احتمال الاستجابة، وعلى فرض أنَّ الاستجابة مستقلة من صف آخر لذلك يمكن

التعبير عن دالة التوزيع المشترك ($r_{i1}, r_{i2}, \dots, r_{im}$) بتوزيع متعدد الحدود (Multinomial distribution) كمايلي.

$$P_r(r_{i1}, r_{i2}, \dots, r_{ij}) = \binom{Ni}{r_{i1}, \dots, r_{ij}} \pi_{i1}^{r_{i1}} \dots \pi_{ij}^{r_{ij}}$$

$$\sum_{i=1}^m r_{ij} = Ni \quad , \quad \sum_{i=1}^k \pi_i = 1 \quad \text{حيث أن :}$$

(2-4) نماذج الانحدار للمتغيرات المعتمدة الوصفية [4]:

Regression models of qualitative dependent variables

تعد نماذج الانحدار الوصفية من النماذج الإحصائية الشائعة استعمالاً، حيث تسهم في بناء أنموذج احصائي يستعمل لغرض تقدير العلاقة بين متغير الاستجابة (y) وبين واحد أو أكثر من المتغيرات المستقلة، إذ لا يضع أنموذج الوصفي أي قيود على المتغيرات المستقلة (X) سواء أكانت هذه المتغيرات قابلة للقياس العددي أم الوصفي أم الكمي.

في حالة نماذج الانحدار الكمية يكون المتغير المعتمد (y) متغيراً يأخذ قيم كمية محصورة بين $(-\infty, \infty)$ غير مقيدة بقيود كما في المنحني الطبيعي ولكن عندما يكون المتغير المعتمد وصفي يأخذ قيم محصورة بين $(0, 1)$ عندها ستكون النماذج الوصفية على شكل دالة احتمالية خطية (Linear probability function) كما في المعادلة الآتية :

$$Y_i = B_0 + \sum_{j=1}^n B_{ij} X_{ij} + U_i \quad \dots \dots \dots \quad (2-2)$$

إذ ان:

y_i : يمثل المتغير الاستجابة

X_i : تمثل المتغير المستقل

B_0, B_1, \dots, B_i : تمثل معاملات الانحدار

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

ويعد أنموذج اللوجستي حالة خاصة من نماذج الانحدار الوصفي.

Logistic Regression Model : (5-2) نموذج الانحدار اللوجستي [28]

أنموذج الانحدار اللوجستي هو أنموذج احصائي شائع ينتمي الى نماذج الانحدار اللاخطية الذي يمتاز بالمرونة العالية، ويعرف بانه أحد الأساليب الإحصائية المهمة التي تستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك من خلال ملاءمة البيانات والمعلومات على المنحى اللوجستي. حيث يعمل الانحدار اللوجستي عادة لوصف طبيعة العلاقة بين الاستجابة وبين متغيراً أو مجموعة من المتغيرات المستقلة.

لذلك يعتبر أنموذج الانحدار اللوجستي حالة خاصة من نماذج الانحدار الاعتيادي بسبب طبيعته الاسمية التي يحملها متغير الاستجابة مثل (الحياة، الموت). ويستخدم أنموذج الانحدار اللوجستي بشكل واسع في تحليل البيانات والتطبيقات الإحصائية. ويكون الانحدار اللوجستي على عدّة أنواع منها الانحدار اللوجستي الثنائي (Binary Logistic Regression)، ويستخدم هذا النوع عندما يأخذ متغير الاستجابة قيمتين وهما (0,1) وكذلك الانحدار اللوجستي متعدد الاستجابة (Multinomial Logistic Regression) وهو امتداد لنموذج الانحدار اللوجستي الثنائي عندما يقع المتغير التابع في أكثر من فئتين، أما الانحدار اللوجستي الرتبي أو يدعى الرتبوي (Ordinal Logistic Regression) ويستخدم هذا النوع من الانحدار في حالة متغير الاستجابة فيكون متغيراً رتبياً سوف نقتصر في العرض النظري على الانحدار اللوجستي متعدد الاستجابة.

(1-5-2) مجالات استعمال الانحدار اللوجستي :

- 1- يهدف الانحدار اللوجستي الى التنبؤ وشرح قيم متغيرات كيفية، حيث يعتبر أفضل أسلوب احصائي في الدراسات التي تهدف لبناء نماذج تنبؤية.
- 2- يستخدم الانحدار اللوجستي في المجال الصحي والطبي لذلك فهو يستعمل أيضا في مجالات حياتية كثيرة ويشاع توظيفه بشكل واسع في الطب والعلوم الاجتماعية.

(2-5-2) شروط تطبيق الانحدار اللوجستي [64] :

يرتكز أنموذج الانحدار اللوجستي على مجموعة من الافتراضات التي تعتبر شروطاً أساسية ويجب توفيرها قبل تطبيق وهي:

- 1- يجب ان يكون المتغير التابع متغيراً اسماً ثنائياً بالنسبة لصنف الانحدار اللوجستي ثنائي الاستجابة، أو يكون تصنيفي بالنسبة للانحدار اللوجستي المتعدد، أو قد يكون ترتيباً بالنسبة للانحدار اللوجستي الرتبي.
- 2- يجب ان يكون أنموذج الانحدار اللوجستي في المستوى الفئوي أو النسبي أو التصنيفي أو الترتيبي، محتوياً على أكثر من متغير مستقل.
- 3- يجب ان يسلك متغير الاستجابة سلوكاً احتمالياً واحداً لكل مشاهدة بمعنى لا يمكن ان تكون المشاهدة في فئتين في الوقت نفسه.
- 4- يجب إن تكون هنالك علاقة بين المتغيرات المستقلة النسبية والتحويل اللوغاريتمي للمتغير التابع ويظهر هذا الشرط عند تطبيق أنموذج الانحدار اللوجستي.
- 5- يجب ألا توجد هنالك قيم شاذة في المتغيرات المستقلة ويمكن أن نحدد القيم الشاذة من خلال اختبار (test mahala Nobis) المتوفر على برنامج (spss).

(6-2) أنواع الانحدار اللوجستي:

ينقسم الانحدار اللوجستي الى ثلاثة نوعين من النماذج وهي:

(1-6-2) أنموذج الانحدار اللوجستي ثنائي الاستجابة [33]:

Binary Logistic Regression

يعرف أنموذج الانحدار اللوجستي الثنائي على أنه أحد نماذج الانحدار اللاخطية حيث يبنى على فروض أساسية لذلك فإن متغير الاستجابة الذي نهتم بدراسته يتبع توزيع برنولي (Bernoulli) باحتمال نجاح π_i عندما $(y=1)$ واحتمال فشل $1-\pi_i$ عندما $(y=0)$ بذلك تكون صيغة دالة الكثافة الاحتمالية تكتب كالاتي

$$P(y_i \setminus X_i) = [\pi(X_i)]^{y_i} [1 - \pi(X_i)]^{1 - y_i} \dots \dots \quad (3 - 2)$$

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

$$P(Y) = \left\{ \begin{array}{ll} \pi_i & \text{عند حدوث الاستجابة } Y = 1 \text{ عندما} \\ 1 - \pi_i & \text{عند عدم حدوث الاستجابة } Y = 0 \text{ عندما} \end{array} \right\} \dots \dots (4-2)$$

π_i : يمثل احتمال النجاح (Success)

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

وعندما $y_i = 1$ نحصل على

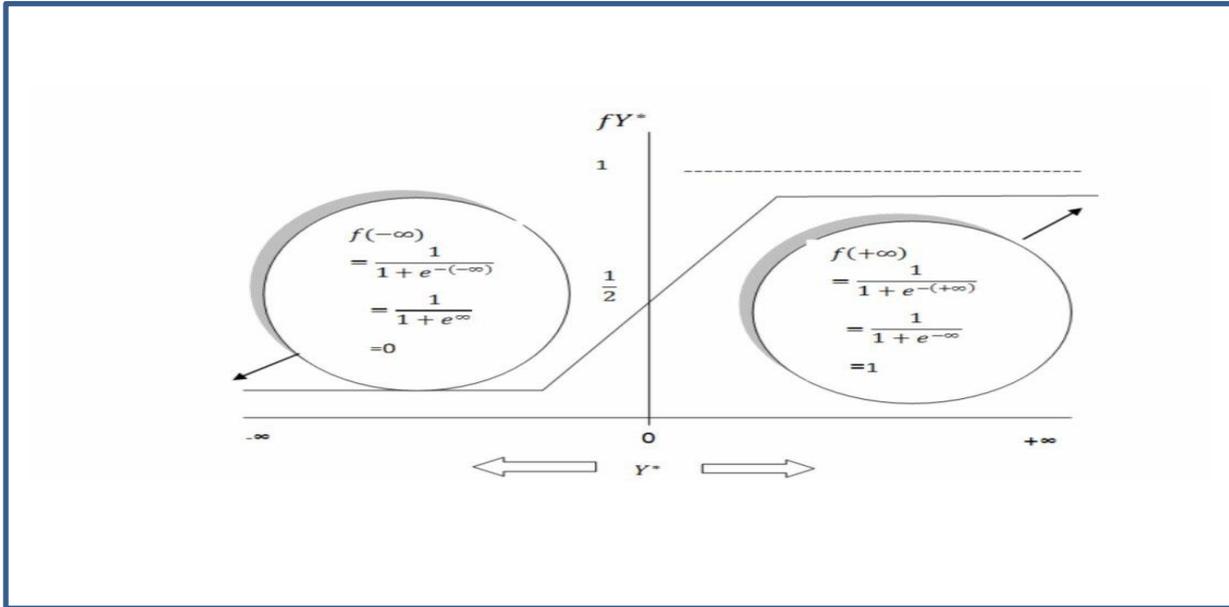
$$\pi_i = \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}} \dots \dots \dots (5 - 2)$$

وعندما $y_i = 0$ نحصل على :

$$1 - \pi_i = 1 - \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

$$1 - \pi_i = \frac{1}{1 + e^{B_0 + B_1 X_1}} \dots \dots \dots (6 - 2)$$

ويمكن عرض الدالة اللوجستية بالشكل الآتي [2]:



الشكل (1-2) شكل الدالة اللوجستية

من الشكل أعلاه، يتضح بأنّ الدالة اللوجستية $f(y^*)$ هي دالة احتمالية تكون محصورة

ضمن القيم $(0,1)$

(2-6-2) نموذج الانحدار اللوجستي المتعدد [27]:

Multinomial Logistic Regression

يعتبر أنموذج الانحدار اللوجستي متعدد الاستجابة من النماذج الاحصائية ذات الأهمية الكبيرة في تحليل البيانات المصنفة، حيث يستخدم بصورة عامة في حالة إذا كان متغير الاستجابة يعود لمتغيرات التي تكون من النوع الاسمي (Nominal) أو الرتبي (Ordinal) والمكونة من تصنيفين أو مستويين ف أكثر.

ويعد أنموذج الانحدار اللوجستي المتعدد الاستجابة امتداداً بسيطاً لأنموذج الانحدار اللوجستي الثنائي الاستجابة. ويعتمد أنموذج الانحدار اللوجستي متعدد الاستجابة بصورة اساسية على التوزيع المتعدد الحدود (Multinomial Distributionom).

$$P_r(r_{i1}, r_{i2}, \dots, r_{ij}) = \binom{N_i}{r_{i1}, \dots, r_{ij}} \pi_{i1}^{r_{i1}} \dots \pi_{ij}^{r_{ij}}$$

(2-7) بعض المشاكل الخاصة بالمتغير المعتمد الوصفي:

Special problems when dependent variable

1- الخطأ العشوائي قد لا يتوزع طبيعياً [30] Nonnormal error term:

في حالة كون المتغير ثنائي الاستجابة أو متعدد الاستجابة يأخذ القيمة (1) عند حدوث الاستجابة وقيمة (0) عند عدم حدوث الاستجابة وسيتم احتساب الخطأ وفق المعادلة

$$U_i = y_i - (B_0 + B_1 X_i) \quad \dots \quad (7-2)$$

فعندما $y=1$ فإن الخطأ سيكون

$$u_i = 1 - B_0 - B_1 X_i$$

وعندما $y=0$ فإن الخطأ سيكون

$$u_i = -B_0 - B_1 X_i$$

وبعبارة أخرى إن الخطأ العشوائي (ui) لن يتوزع طبيعياً بمتوسط صفر وانحراف معياري σ في حالة كون المتغير المعتمد ثنائي الاستجابة أو متعدد الاستجابة وهذه المشكلة لا تعتبر مشكلة رئيسية لأن بعض طرائق التقدير لا تحتاج الى أن يكون توزيع الخطأ (ui) طبيعياً بمتوسط صفر وانحراف معياري σ كطريقة المربعات الصغرى (OLS) والمرجحة (GLS).

2- عدم ثبات تباين الخطأ Non constant error variance [4]:

من المشاكل التي تصادف المتغير الثنائي الاستجابة والمتعدد الاستجابة هي مشكلة عدم ثبات تباين الخطأ بمعنى إنَّ هذه النماذج تبايناتها معتمده على متوسطاتها. فعلى سبيل المثال لو كان متغير الاستجابة y_i يمثل لتوزيع احصائي معين ((بواسون مثلاً))، فإنَّ التباين يساوي المتوسط وإنَّ المتوسط يزداد كلما ازدادت قيم المتغيرات المستقلة x_i وبالتالي فإنَّ التباين سيعتمد أيضاً على المتغيرات المستقلة مما يجعل التباين غير ثابت (غير متجانس) لذلك لابد من اجراء بعض التحويلات الرياضية والإحصائية لجعل التباين متجانس، ومن هذه التحويلات طريقة المربعات الصغرى الموزونة.

3- عدم تقيد دالة الاستجابة بحدود الاحتمال $\{0 \leq E(y_i/X_i) \leq 1\}$ [4]:

طالما إنَّ توقع دالة الاستجابة (المتغير المعتمد الوصفي) يمثل احتمال Probabilitr فإنَّ توقع هذه الدالة يجب إنَّ يقع ضمن الفترة (0,1) أي إنَّ $P_i \leq 1$ ولكن ليس هناك ما يضمن بأنَّ مقدر $E(y_i/X_i)$ سوف يقع ضمن الفترة (0,1)، ولاسيما ان القيمة التقديرية لمتغير الاستجابة (الذي يمثل لتوزيع برنولي) في معادلة انحدار خطية

$$\hat{E}(y_i/X_i) = b_0 + \sum_{j=1}^n b_{ij}X_{ij} = P_i$$

ولجعل قيم مقيدة بالفترة $\hat{E}(y_i/X_i) \in (0,1)$ يمكن اتباع إحدى الطريقتين الآتيتين:

* الطريقة الأولى وهي احتساب القيم المتوقعة للمعاملات بأحد طرائق التقدير الملائمة (كطريقة المربعات الصغرى الاعتيادية (OLS) أو غيرها ذلك وفق أنموذج الانحدار الخطي الاحتمالي:

$$Y_i = B_0 + \sum_{i=1}^n B_{ij}X_{ij} + U_i$$

ثم نعوض القيم التقديرية لمعاملات أنموذج الانحدار بالمعادلة فاذا كانت قيم $\hat{E}(Y_i/X_i)$

$$\hat{E}(y_i/X_i) = 0 \text{ (قيمة سالبة) في هذه الحالة سنفرض}$$

وعندما تكون قيمة $\hat{E}(y_i/X_i)$ أكبر من (الواحد) في هذه الحالة سنفرض

$$\hat{E}(y_i/X_i) = 1$$

*الطريقة الثانية هي إنَّ تستخدم نماذج لدوال التحويل Transformation وهذه الدوال كفيلا بجعل قيم $\hat{E}(y_i/X_i)$ تقع ضمن الفترة (1،0).

(8-2) نماذج دوال التحويل الخطي [38]:

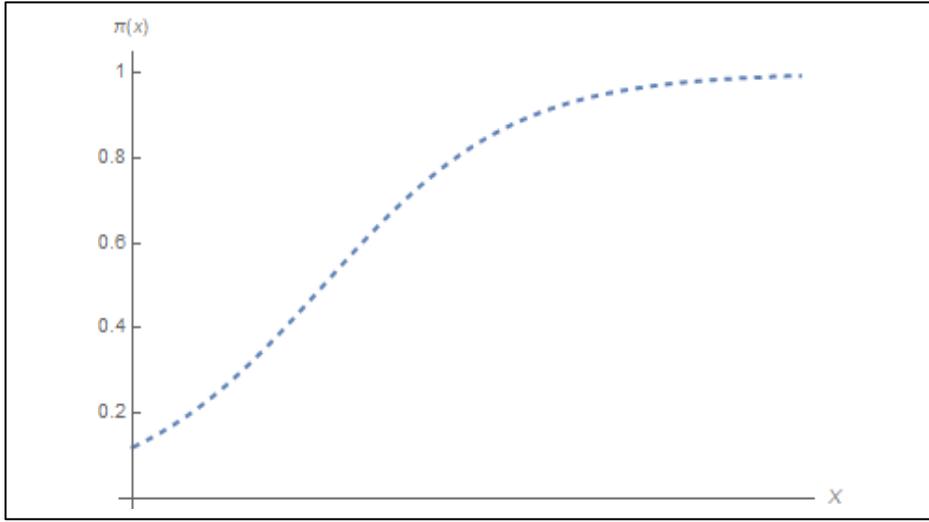
Linear Transformation Function

هنالك عدّة تحويلات يمكن استخدامها لتحويل نموذج من الانحدار اللاخطي الى الانحدار الخطي، إنَّ الغاية الأساسية من اجراء عملية التحويل هو لكي يتغير شكل التوزيع التكراري لبيانات الافتراضات النظرية في نموذج، وهنالك عدّة تحويلات يمكن استخدامها لتحويل أنموذج من الانحدار اللاخطي الى الانحدار الخطي، فقد تم استخدام أنموذج (Logit) لتحويل أنموذج اللوجستي الى معادلة خطية يمكن استعمالها والحصول على قيم المعلمات .

(1-8-2) أنموذج لوغار يتم النسبة المضافة [23]:

Models of odds Ratio(logit)

من المعلوم إنَّ طبيعية العلاقة التي تربط المتغير متعدد الاستجابة مع مجموعة من المتغيرات التوضيحية تكون علاقة غير خطية في الشكل (2-2)



الشكل (2-2) يمثل العلاقة غير الخطية بين المتغير الاستجابة $\pi(x)$ والمتغيرات التوضيحية

(x) من اعداد الباحثة بالاعتماد على مصدر [2]

اذن إنَّ π_i يمثل احتمال النجاح في الانحدار اللوجستي و $(1-\pi_i)$ يمثل احتمال الفشل، فان نسبة النجاح الى نسبة الفشل تنتج من خلالها معادلة غير الخطية، قد نلاحظ وجود صعوبة عند التحويل الى أنموذج الخطي لغرض تقدير أنموذج، لهذا اقترح الباحث الأمريكي (berkson) عام (1944). تحويل صيغة الدالة اللوجستية الى دالة خطية من خلال قسمة احتمال النجاح الى احتمال الفشل و أخذ اللوغاريتم الطبيعي لطرفي معادلة، أي إنَّ

$$\text{Logit} (\pi_i) = \text{Ln} \frac{\pi}{1-\pi}$$

وبقسمة المعادلة (5-2) و(6-2) فان:

$$\text{Logit} (\pi_i) = \text{Ln} \frac{e^{B_0 + \sum_i^n B_i X_i}}{1 + e^{B_0 + \sum_i^n B_i X_i}} \frac{1}{1 + e^{B_0 + \sum_i^n B_i X_i}}$$

$$\text{Logit} (\pi_i) = B_0 + \sum_i^n B_i X_i$$

لنفترض إنَّ $y^* = \text{Ln} \left(\frac{\pi}{1-\pi} \right)$ فإنَّه يمكن اعاده كتابة *y بالشكل الآتي:

$$Y^* = B_0 + \sum_{i=1}^n B_i X_i \quad \dots \quad (8-2)$$

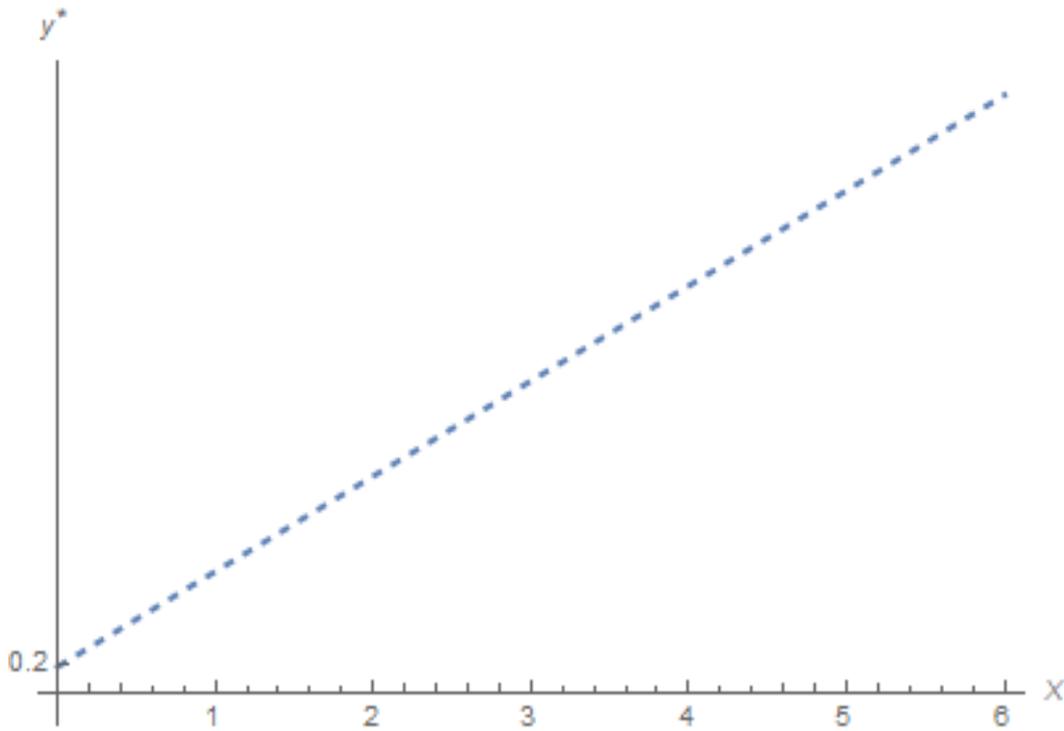
حيث أن:

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

Y^* حيث تمثل متغير اللوجستيك الجديد والتي قيمته تساوي (1) عند حدوث الاستجابة وتساوي صفر عند عدم حدوث الاستجابة وبذلك سوف تكون قيم الانحدار اللوجستي محصورة بين (0,1)

$$0 \leq Y^* \leq 1$$

ستكون طبيعة العلاقة في معادلة (8-2) ما بين متغير الاستجابة ومجموعة المتغيرات التوضيحية علاقة خطية عند تطبيق تحويل (logit)، ويوضح ذلك بالشكل (3-2) الآتي:



الشكل (3-2) العلاقة الخطية بين متغيرات التوضيحية ومتغير الاستجابة الجديد

(Logit) من اعداد الباحثة بالاعتماد على مصدر [2]

(9-2) طرائق تقدير أنموذج الانحدار اللوجستي :

بعد ان تم وصف أنموذج وتحويله من اللاخطي الى الخطي، بعد ذلك تأتي مرحلة تقدير أنموذج حيث يتم تقدير أنموذج ببعض الطرائق التي تكون خاصة بالمتغيرات متعدد المستويات، والتي تعد أحد أهم الجوانب في الاستدلال الإحصائي لأغلب النماذج، ومن طرائق التقدير هذه طريقة تقدير الإمكان الأعظم (Maximum Likelihood Method) من خلال تطبيق خوارزمية نيوتن رافسون (Newton Raphson)، وطريقة المربعات الصغرى الموزونة (Weighted least squares Method)، بالمقارنة مع الشبكات العصبية (Neural Networks)، والخوارزمية الجينية (Geneti Algorithm) كطرائق لتقدير أنموذج الانحدار اللوجستي.

(1-9-2) طريقة تقدير الإمكان الأعظم [17][24][33] :

Maximum Likelihood Method

تعرف طريقة الإمكان الأعظم إنها من الطرائق الإحصائية الأوسع استخدام في تقدير معلمات أنموذج الإحصائية والرياضية، لذا تعتمد في تقديرها على تكرار العمليات الحسابية عدّة مرات أي إنها طريقة تكرارية لكي يتم الوصول الى أفضل تقديرات. لم تكن تستخدم هذه طريقة بشكل واسع الابعاد ظهور البرامج الإحصائية، بما ان أنموذج الانحدار اللوجستي المتعدد يتبع توزيع متعدد الحدود (Multinomial Distribution) والذي يمكن وصف دالة الكثافة الاحتمالية له كالآتي:

$$P_r(r_{i1}, r_{i2}, \dots, r_{ij}) = \binom{ni}{r_{ij}, \dots, r_{ij}} \pi_{i1}^{r_{i1}} \dots \pi_{ij}^{r_{ij}} \dots \quad (9 - 2)$$

إذ أن:

$$\pi_{i1}^{r_{i1}} \dots \pi_{ij}^{r_{ij}}$$

تمثل الاستجابة في المجتمع.

حيث يتم تعظيم أنموذج الانحدار اللوجستي من خلال دالة الإمكان الأعظم (MLE) بدلالة معلمات $L(\beta)$ يتم الحصول على معادلة:

$$L(\beta, \pi_{i1} \dots \pi_{ij}) = \prod_{i=1}^n \prod_{j=1}^J [\pi_{ij}(x)]^{r_{ij}} \dots \quad (10-2)$$

وبأخذ اللوغاريتم لطرفي المعادلة أعلاه سوف يتم الحصول على:

$$\text{Log } L(\beta, \pi_{i1} \dots \pi_{ij}) = \sum_i \sum_j r_{ij} \text{Log } \pi_{ij}(x) \dots \quad (11-2)$$

بالمعادلة (11-2) نحصل على: $\pi_{ij}(x)$ وبالتعويض عن

$$\begin{aligned} \text{Log}(\beta, \pi_{i1} \dots \pi_{ij}) &= \sum_i \sum_j r_{ij} \text{Log} \frac{e^{B_0 + \sum_i^n B_i X_i}}{1 + e^{B_0 + \sum_i^n B_i X_i}} \\ &= \sum_i \sum_j r_{ij} (B_0 + B_i X_i) - \sum_i \sum_j r_{ij} \text{Log}(1 + e^{-B_0} + \sum_i^n B_i X_i) \end{aligned}$$

وباشتقاق دالة الإمكان بالنسبة للمعلمات (β) :

$$\frac{\partial \text{Log} L(\beta)}{\partial (\beta)} = \frac{\partial \sum_i \sum_j r_{ij} (B_0 + B_i X_i)}{\partial (\beta)} - \frac{\partial \sum_i \sum_j r_{ij} \text{Log}(1 + e^{-B_0} + \sum_i^n B_i X_i)}{\partial (\beta)} \dots \dots (12-2)$$

وعند مساواة المشتقة بالصفر ($\frac{\partial \text{Log} L(\beta)}{\partial (\beta)} = 0$) نحصل على معادلات طبيعية، وإنَّ حلول هذه المعادلات تكون عبارة عن قيم المعلمات التقديرية التي تنتج من خلال عملية التعظيم ولكون المعادلات التي تم الحصول عليها غير خطية نلجأ الى استخدام أحد الأساليب التكرارية والتي تنتج تقديرات مناسبة. ومن ضمن هذه الأساليب هي نيوتن رافسون.

2-9-2) الشبكات العصبية الاصطناعية :

Artificial Neural Networks (ANN) [7]

بعد التقدم الهائل في تركيبية الحاسبات الالكترونية وزيادة وتنوع استخدامها في حل المشكلات المعقدة فقد استخدم أسلوب الذكاء الاصطناعي في الحاسبات مما أدى الى تسريع ومحاكاة الحلول.

حيث تعد الشبكات العصبية من اهم مجالات الذكاء الاصطناعي وهي برمجة لمحاكاة عمل المخ البشري، تدور فكرة الشبكات العصبية حول محاكاة العقل البشري باستخدام الحاسب الآلي وقد اكتشف العلماء طريقة عمل المخ البشري من حيث قابليته على تمييز الأشياء وقابليته على التعلم وعلى التذكر وكذلك أيضا قابليته على اتخاذ القرارات , ومن المعلوم إنَّ المخ يتكون من مليارات الخلايا العصبية والتي تكون متشابكة فيما بينها بطريقة معقدة جداً عن طريق الزوائد العصبية.

وبعبارة أخرى إنَّ الشبكات العصبية الاصطناعية (ANN) تكون ذات تشابه كبير مع الدماغ البشري حيث إنَّها أيضا تكسب المعرفة بالتدريب وتستطيع التعلم من الماضي، وفق لنظام ارسال واستقبال النبضات الكهربائية بين الخلايا العصبية، هذه الحقيقة دفعت الكثير من العلماء والباحثين في مجال العلوم الإدراكية والتي تعرف باسم الذكاء الاصطناعي. حيث تم التوصل الى الشبكات العصبية وتطويرها عام 1992 من قبل العالم Zurad.

(1-2-9-2) الشبكات العصبية الاصطناعية [41]:

Artificial Neural Networks ANN

تعرف الشبكات العصبية الاصطناعية بإنَّها الطريقة التي يؤدي بها الدماغ البشري لمهمة ما، وذلك من خلال معالجة ضخمة تكون موزعة على التوازي، ومتكونة من وحدات معالجة بسيطة، هذه الوحدات ما هي إلا عناصر حسابية تسمى عصبونات أو عقد (Nodes)، (Neurons)، والتي تمتلك خاصية عصبية، إذ إنَّها تقوم بتخزين المعرفة العملية وكذلك المعلومات التجريبية لجعلها متاحة للمستخدم وذلك عن طريق ضبط الأوزان. إذ إنَّ (ANN) تتشابه مع الدماغ البشري في إنَّها تكتسب المعرفة بالتدريب وتخزن هذه المعرفة باستخدام قوى وصل داخل العصبونات تسمى الأوزان التشابكية. وهناك أيضاً تشابه عصبي حيوي مما تعطي الفرصة لعلماء البيولوجيا في الاعتماد على (ANN) لفهم تطور الظواهر الحيوية.

(2-2-9-2) خصائص الشبكات العصبية الاصطناعية [6]:

Properties of Artificial Neural Networks:

تتميز الشبكات العصبية الاصطناعية بمجموعة من الخصائص الآتية:

- 1- تستند الشبكات الى برمجيات ذات أسس رياضية قوية.
- 2- قدرة الشبكات العصبية على تخزين المعرفة المكتسبة من البيانات السابقة لنفس الظاهرة قيد الدراسة وتكيفها مع البيانات الجديدة.
- 3- تمتاز الشبكات العصبية بمرونة تتيح للمستخدمين إمكانية استخدامها في كافة المجالات
- 4- الشبكات العصبونية الاصطناعية يتم تنفيذها بشكل متوازي فيما بينها.
- 5- تقبل أي نوع من البيانات سواء كانت نوعية أو كمية.

(3-2-9-2) مفاهيم تقنيات الشبكة العصبية الاصطناعية [46][54]:

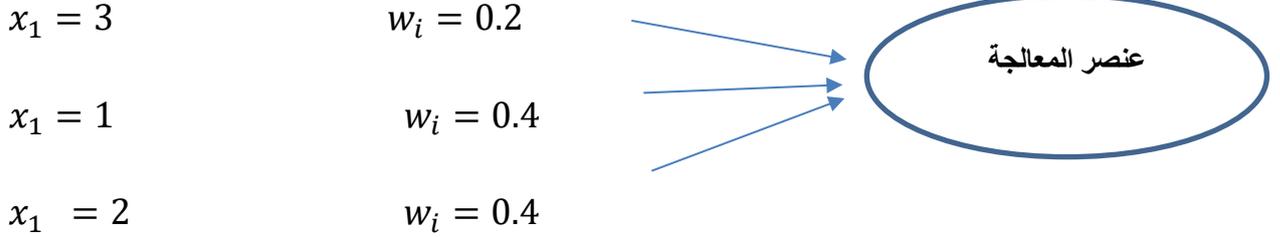
Concepts of Artificial Neural Network Techniques

لغرض دراسة الشبكات العصبية الاصطناعية (ANN) لابد من توضيح بعض المفاهيم الخاصة بيها وكمايلي:

- طبقات (Layers): وهي عبارة عن العدد الكلي لطبقات الشبكة العصبية، وحيث تتكون هذه الطبقات من طبقتين هما طبقات المدخلات Layers Input وطبقات المخرجات Layers Output مع ملاحظة إنَّ عقد Nodes المدخلات تستخدم فقط لتوزيع الأوزان.
- حجم الطبقة (size Layers): يتم تحديد حجم الذاكرة المحددة بالبرنامج الخاص بتشغيل الشبكة العصبية الاصطناعية من خلال حجم الطبقة.
- الدالة العصبية المستخدمة (Function Neural) وهي الدالة اللوجستية (Logistic Function).

$$f(x) = \frac{1}{1+e^{-x_i}} \quad \dots \quad (13 - 2)$$

ويمكن توضيح فكرة الشبكة العصبية الاصطناعية من خلال المثال الآتي:



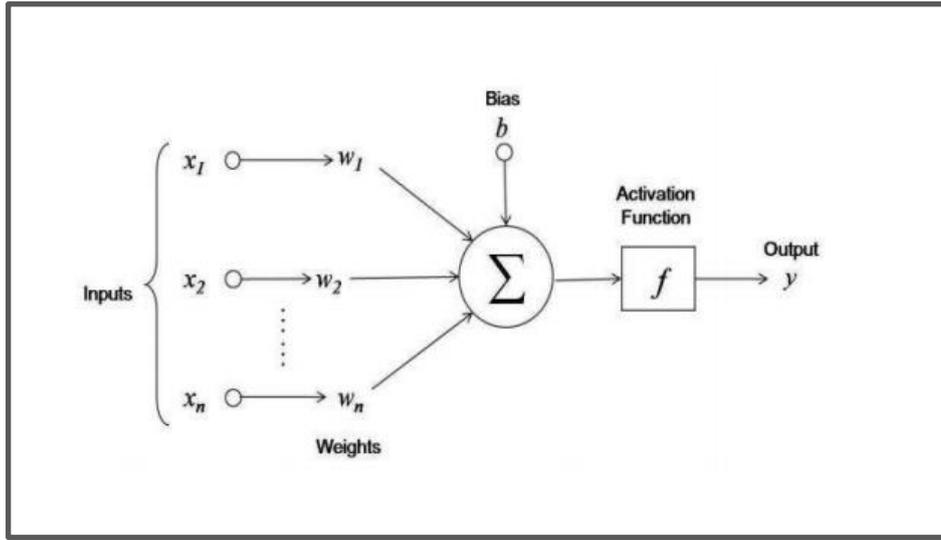
$$X=3 (0.2) +1 (0.4) +2 (0.4) =1.8 \quad \text{Summation Function.}$$

$$f(x)=\frac{1}{1+e^{-1.8}}=0.86 \quad \text{Transformation Function.}$$

(4-2-9-2) مكونات الشبكات العصبية الاصطناعية [40] :

بما أنَّ للإنسان وحدات ادخال يتم من خلالها إيصاله بالعالم الخارجي، والتي تمثل حواسه الخمس. فالشبكات العصبية الاصطناعية كذلك لها مدخلات تتمثل بوحدات المعالجة التي تتم فيها العمليات الحسابية التي تضبط بها الأوزان وسوف نحصل عن طريق هذه العمليات على ردة الفعل لكل مدخل من المدخلات للشبكة. فوحدات الإدخال تكون طبقة تعرف بطبقة المدخلات، ووحدات المعالجة سوف تكون طبقة المعالجة، والتي من خلالها يتم إخراج نتائج الشبكة وتدعى طبقة المخرجات. وبين كل طبقة من هذه الطبقات هناك طبقة من الوصلات البيئية التي تربط كل طبقة بالطبقة التي تليها، ويتم فيها ضبط الأوزان الخاصة بكل وصلة بينية.

وتحتوي الشبكة على طبقة واحدة فقط من وحدات الإدخال، لكنها قد تحتوي على أكثر من طبقة من طبقات المالح. وكما في الشكل الآتي:



الشكل (4-2) مكونات الشبكات العصبية الاصطناعية [40]

يتألف العصبون من:

- إشارات المدخلات: x_1, x_2, \dots, x_n
- الأوزان w_1, w_2, \dots, w_n : حيث يتم من خلال الوزن التعبير عن شدة الترابط ما بين العنصر القبل والعنصر البعد.

• عنصر المعالجة (دالة التنشيط) y : وهي التي تشمل قسمين:

-الجامع: وهو الذي عن طريقه يتم جمع الإشارات في المدخل الموزون

- تابع التفعيل: وهذا التابع يحد من مخرج العصبون ويسمى أيضا بتابع التخمين

•المخرجات: y

(5-2-9-2) الصياغة الرياضية للعصبون :

يتم تمثيل العمليات الرياضية للعصبون بالصيغة الآتية:

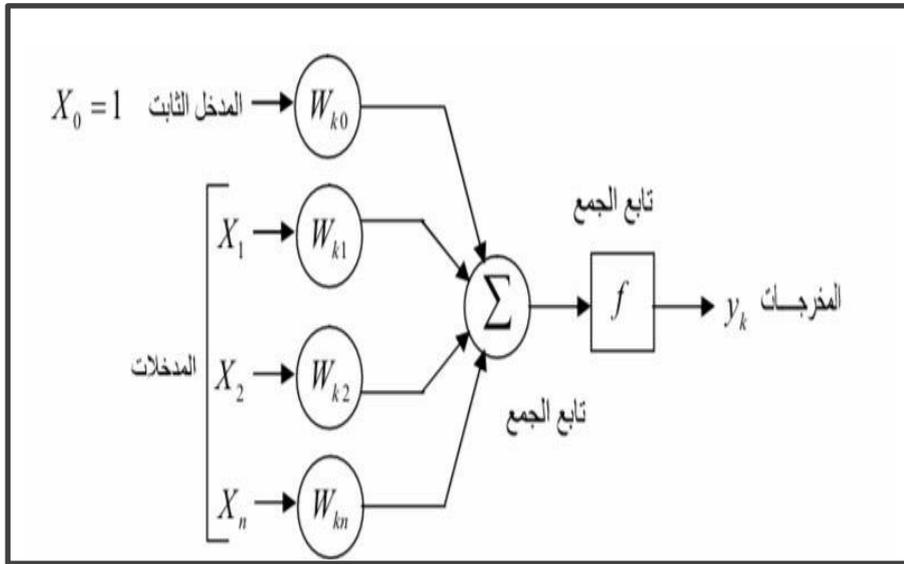
$$y_k = f \left(\sum W_{j k} X_j \right) \quad \dots \dots \dots \quad (14 - 2)$$

حيث إن:

إشارات المدخل: $X_1 X_2 \dots X_j$

الأوزان للعصب: $W_1 W_2 \dots W_j$

تمثل تابع عملية الجمع حيث B_k الإنحياز والذي يمكن عدّه على أنّه أحد الأوزان $W_0 = 1$ ومدخلة $X_0 = 1$ وليكون أنموذج كالآتي:



الشكل (5-2) الخوارزمية الرياضية للعصبون الاصطناعي

(7-2-9-2) هيكلية الشبكة العصبية الاصطناعية [41][42]:

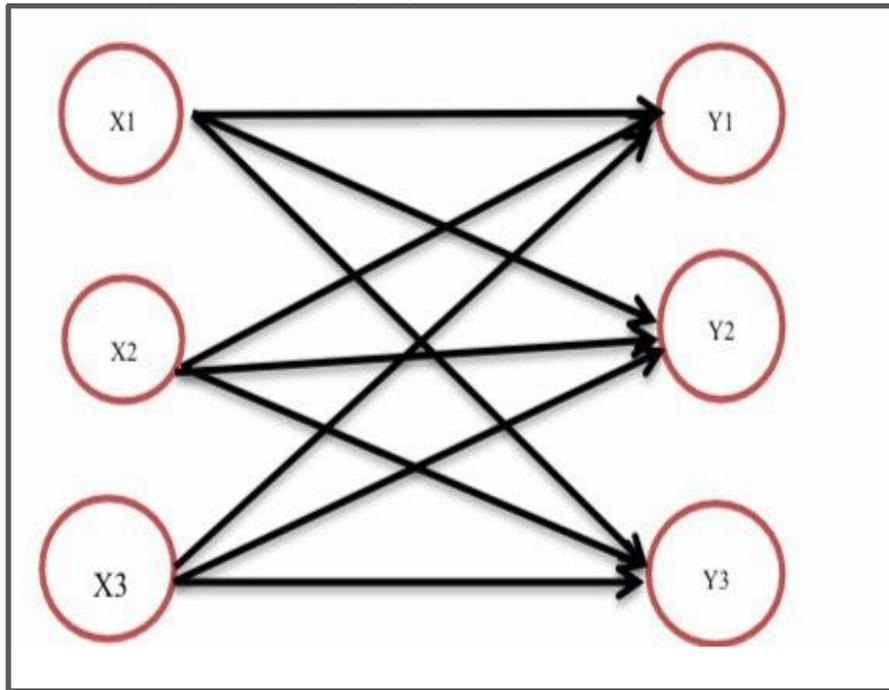
Artificial Neural Network Structure

ترتيب العصبونات في مستويات الطبقات، وشكل الترابطات الذي يكون بين طبقات أو ضمنها يدعى بالبنية المعمارية أو هندسية الشبكة العصبية. ويمكن حساب عدد طبقات في الشبكة أو معرفتها من خلال الترابطات بين العصبونات بدون حساب طبقة الإدخال وتصنف الشبكات العصبية من حيث عدد الطبقات الى نوعين:

(1-7-2-9-) الشبكات وحيدة الطبقة :

Single Layer Networks

إنَّ هذا النوع من الشبكات يتكون من طبقتين فقط، وهما طبقة المدخلات (Input Layer) وطبقة المخرجات (Output Layer)، وتكون انسيابية الإشارات التي تدخل من طبقة المدخلات الى طبقة المخرجات بالاتجاه الأمامي كما في الشكل الآتي:



الشكل(2-6) شبكة وحيدة الطبقة [41]

حيث إنَّ :

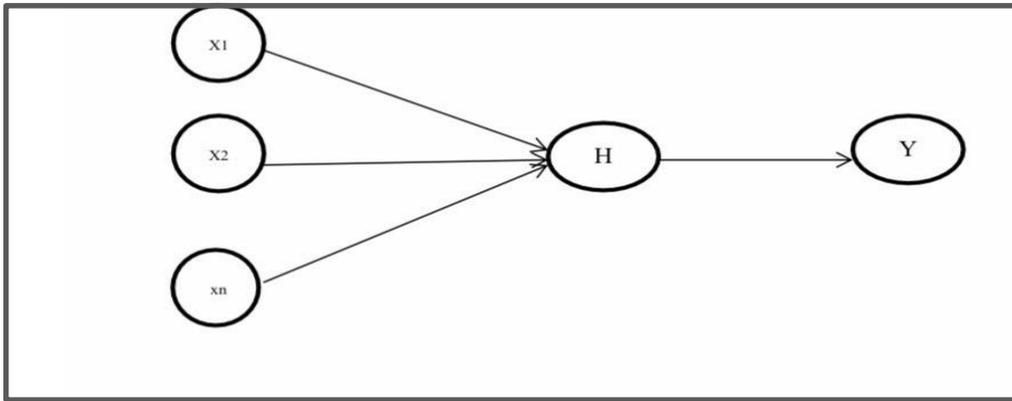
(X_1, X_2, X_3) تمثل طبقة المدخلات.

(y_1, y_2, y_3) تمثل طبقة المخرجات.

(2-7-2-9-2) الشبكات متعددة الطبقات [21] [6] :

Multilayers Networks

يتميز هذا النوع من الشبكات بقدرتها على حل المسائل التي تكون أكثر تعقيدا من تلك المسائل التي يمكن حلها في الشبكات ذات الطبقة المفردة، والسبب يعود الى وجود الطبقات الوسيطة التي تعطي مرونة أكثر في بناء الدوال الناقل ما بين المخرجات والمدخلات، وكما يتميز هذا نوع من الشبكات باحتوائه على طبقة خفية واحدة أو أكثر حيث تكون هذه الطبقة كافية لحل معظم المشكلات كما في الشكلين الآتيين

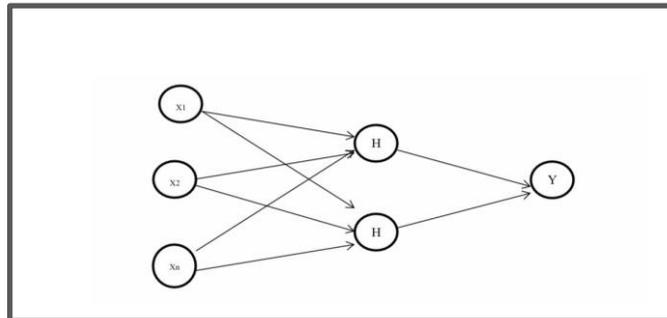


الشكل (7-2) الشبكة العصبية ذات طبقة واحدة وخلية خفية واحدة [21]

(X) تمثل طبقة المدخلات.

(R) تمثل طبقة المخرجات .

(H) تمثل طبقة المخفية

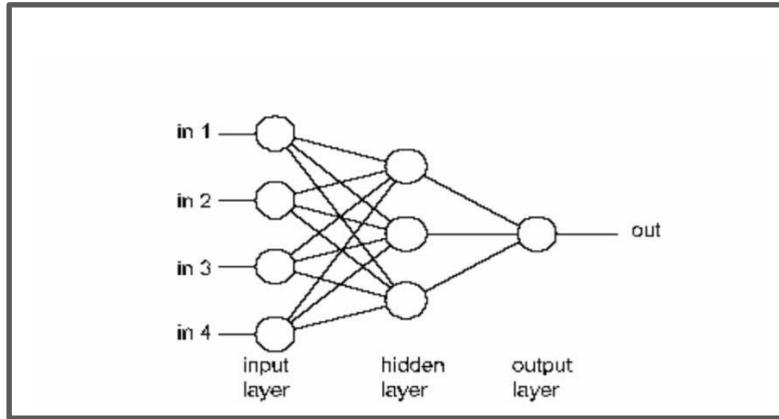


الشكل (8-2) الشبكة العصبية ذات طبقة واحدة وخلايا خفية متعددة [21]

وتنقسم هذه الشبكات الى نوعين أساسيين هما:

الشبكات ذات التغذية الامامية والشبكات ذات التغذية الخلفية (العكسية)، وتوجد في الشبكات العصبية ثلاث طبقات من الأوزان وهي:

- طبقة الوزن: تكون بين مستوى المخفي ومستوى الإدخال (Input to hidden weights).
- طبقات الأوزان: تكون بين مستويات المخفية (Hidden weights).
- طبقة الوزن ما بين مستوى الإخراج ومستوى المخفي (Hidden to out put).



الشكل (9-2) مخطط يوضح معمارية الشبكة العصبية الاصطناعية النموذجية

(8-2-9-2) أنواع الشبكات العصبية الاصطناعية [9] [46] [64]:

Types of Artificial Neural Networks:

ان الشبكات العصبية الاصطناعية يتم تنظيمها بعدة هيئات مختلفة أي بمعنى أن يتم توصيل

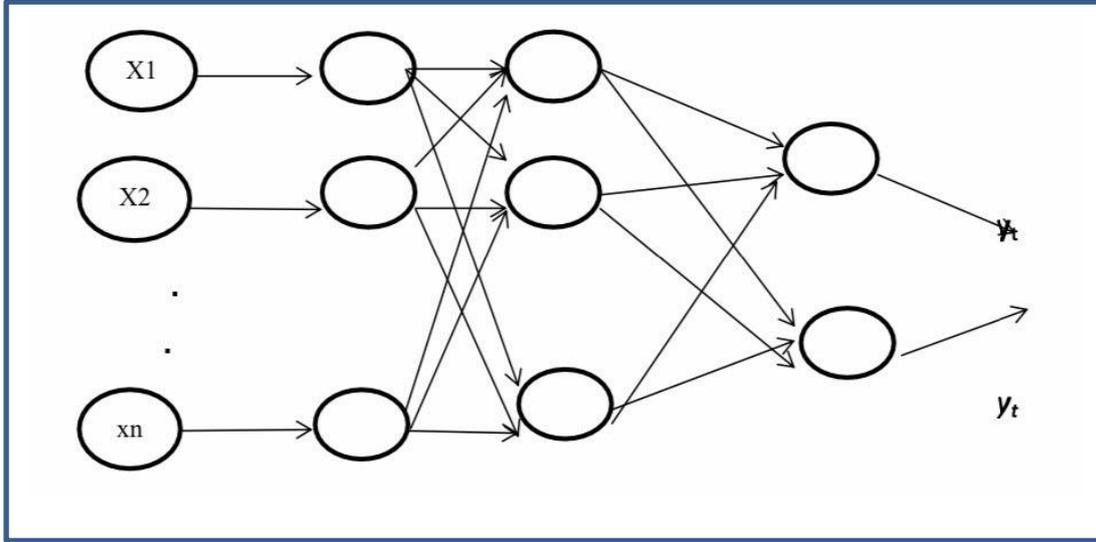
الخلايا العصبية بعدة طرق مختلفة ومنها:

(1-8-2-9-2) الشبكات العصبية امامية الانتشار:

Feed Forward Neural Networks

وهي الشبكات التي يتميز تركيبها بأنه يكون خالي من وجود حلقة مغلقة من الترابطات بين الوحدات المكونة لها لذا تعد هذا نوع من الشبكات من أكثر الشبكات استخداماً. وعادة ما يكون

هذا النوع من الشبكات من طبقتين على الأقل هما طبقة المدخلات وطبقة المخرجات وتتواجد بينهما الطبقة المخفية في كثير الأحيان، وتنتقل العمليات الحسابية من طبقة المدخلات الى طبقة المخرجات باتجاه واحد الى الامام كما في الشكل الاتي:



الشكل (10-2) نموذج لشبكة عصبية ذات تغذية أمامية [49].

2-8-2-9-2) الشبكات العصبية عكسية الانتشار :

Back Propagation Neural Networks

تتميز الشبكات ذات التغذية العكسية أو الخلفية بإتّها من الشبكات التي بالعادة تكون معقدة وقوية جدا وذلك لامتلاكها إشارات تنتقل أو تتحرك بصورة عكسية في كلا الاتجاهين ، حيث يعد هذا نوع من الشبكات الحركية التي تقوم بتغيير استقرارها بالاستمرار لحين الوصول الى نقطة التوازن، وتبقى عند تلك النقطة الى إنّ تم تغيير في الإدخال فيتم حينها الرسالة عن توازن جديد بناء على مدخلات جديد، لذا يستند مفهوم الشبكة العصبية المعتمدة على متوسط مربعات الخطأ ، وذلك من خلال إيجاد القيمة الصغرى لمتوسط مربعات الخطأ الكلي لقيمة المخرجات لتلك الشبكة العصبية ، اذ تبقى في اطار الرسالة عن أفضل توازن والتي تحديد أفضل توافق للنموذج ككل .

• خوارزمية الشبكة العصبية ذات الانتشار العكسي

تعد خوارزمية شبكة الانتشار العكسي للخطأ واحدة من أهم خوارزميات شبكات التدريب حيث تعتمد على اختيار دالة خطأ مناسب والتي تتحدد قيمتها بالنتائج الحقيقية والقيم المطلوب الحصول عليها والتي تعتمد أيضاً على معلمات الشبكة مثل الأوزان والعتبة وإن حساسية العصبون يمكن الحصول عليها كما في المعادلة (15-2)

$$\mu = \sum_{i=1}^N W_i X_i \quad \dots \dots \quad (15 - 2)$$

وكما إن مخرجات الشبكة يمكن صياغتها بالدالة كما في المعادلة (16-2)

$$y = f(\mu - \theta) \quad \dots \dots \quad (16 - 2)$$

حيث إن:

يمثل المدخلات (X_1, X_2, \dots, X_n)

يمثل الأوزان (W_1, W_2, \dots, W_n)

μ : يمثل حساسية العصبون.

θ : تمثل العتبة.

y : تمثل مخرج الشبكة.

$f(\cdot)$: تمثل دالة الحساسية .

يمكن لخوارزمية الانتشار العكسي إن تنفذ الخطأ بطريقتين هما on-line mode و batch

mode

1- في حالة أنموذج موجود على شبكة الإنترنت on-line يمكن إن تحسب دالة الخطأ بعد كل نموذج ادخال و اشارة الخطأ

تبت رجوعها من خلال الشبكة وتعديل الأوزان. ودالة الخطأ المحسوبة هي عبارة عن متوسط مربعات خطأ (MSE) بين القيم الحقيقية والقيم التقديرية، وأن الأوزان الجديدة تستخدم في حساب أنموذج.

2- اما في حالة أنموذج دفعة واسطة batch mode فان اشارات الخطأ يمكن إنَّ تحسب لكل ادخال، والأوزان تعدل وتكيف عندما يقيم ادخال القيم جميعاً، ودالة الخطأ المحسوبة هي عبارة عن مجموعة الخطأ (MSE) كل الاشارات وجميع النماذج قبل تمثيل أنموذج الآتي، الأمر الذي يعني إنَّ هذا أنموذج يقوم بحساب MSE لكل وحدة من وحدات المخرجات أو لجميع النماذج.

(9-2-9-2) تدريب الشبكات العصبية الاصطناعية [39] [62]:

Artificial Neural Network Training

ان الشبكة العصبية الاصطناعية تستند على مفهوم التدريب الذي يعتمد على أساس تقليل متوسط مربع الخطأ. حيث أنَّ الاشارات الخارجة في الشبكات من أي خلية تعتمد على الإشارات الداخلة وهذه الشبكات عادة تكون غير تكرارية أو غير دورية بمعنى آخر أنَّ الارتباطات الموجودة بين العقد تكون غير دورية.

في حين تجري عملية تعلم الشبكات العصبية بواسطة خوارزميات مختصة تسمى خوارزميات التعلم والتي تتولى عملية التدريب اذ تعمل على تعديل أوزان الشبكة للوصول الى أقل حد خطأ ممكن ومن ثم تحسين اداء الشبكة ككل.

ان الهدف الأساسي من خوارزميات التعلم هو تحديد الأوزان المرجحة المثلى، وهناك طريقتان للتدريب والتعلم هي:

(1-9-2-9-2) التعلم المراقب :

Attended Learning

تعتمد الشبكة العصبية في هذا النوع من التعلم على الفرق بين مخرجات الشبكة والمخرجات الحقيقية لتحديد أوزان الشبكة، حيث تتم عملية التعلم المراقب للشبكات العصبية الاصطناعية من خلال ادخال البيانات التدريبية للشبكة بشكل نمطين الأول نمط المدخلات و الثاني نمط المخرجات، وكل منها يرتبط بمتجه المخرجات المستهدفة. ومن تطبيقات هذا النوع من التعلم بناء ذاكرة باستخدام الشبكات العصبونية.

(2-9-2-9-2) التعلم غير المراقب :

Unattended Learning

إنَّ أسلوب التعلم غير المراقب يتم بإدخال البيانات التدريبية للشبكة بشكل نمط واحد فقط نمط المدخلات، فتعتمد هذه الطريقة على أساس قدرة الشبكة في اكتشاف ملامح نمط المخرجات التي يتم تعميمها على باقي المخرجات، أي بدون سابق معرفة بقيم المخرجات الحقيقية وهذا عكس الأسلوب السابق (التعلم المراقب) لذا يستند هذا الأسلوب من التعلم على الارتباطات بين قيم الإدخال .

(3-9-2) طريقة تقدير الخوارزمية الجينية :

Genetic Algorithm

تعد الخوارزمية الجينية أحد أساليب الذكاء الاصطناعي فهي من طرق الرسالة العشوائية التي تعالج مشكلة ما من أجل التوصل الى أفضل النتائج الممكنة حيث تعتمد على آلية الانتقاء الطبيعي ونظام الجينات الطبيعية لذا تمتلك كماً هائلاً من الحلول البديلة، والحل الناتج من تطبيق الخوارزمية يكون في أغلب الأحيان حلاً قريباً الى الجينات الطبيعية لذلك برزت أهميتها في حل المسائل المعقدة.

ان فكرة عمل الخوارزمية الجينية تعتمد على أفكار الهندسة الوراثية، التي تتميز بالإنتاج 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.

(1-3-9-2) مفهوم الخوارزمية الجينية [22] [53]:

Concept of a genetic algorithm

تصنف الخوارزمية الجينية بأنها إحدى طرق الرسالة العشوائية حيث برزت تقنية الخوارزمية الجينية كأداة حقيقية وقوية لإيجاد الحل الأمثل مستخدمة الجينات بالاعتماد على مبدأ داروين في التطور الذي ينص على أنَّ كل الكائنات الحية على مر الزمن تتحدر من

أسلاف مشتركة، ومن خلال تطبيق مبدأ الانتخاب الطبيعي (البقاء للأفضل) يتم الحصول على الأفراد ذوي التقييم الأفضل، والقادرين على التكيف مع البيئة المتجددة وتمتاز الخوارزمية الجينية بسمتين أساسيتين وهما :

- العشوائية: إنَّ القاعدة الأساسية في الخوارزمية الجينية هي العشوائية حيث إنَّ كلا عمليتي إعادة الإنتاج والاختيار تم بطريقة عشوائية.
- فضاء الرسالة : إنَّ الخوارزمية الجينية تعتمد على فضاء الرسالة أو ما يسمى بالمجتمع الذي يحتفظ عند عملية إعادة الإنتاج وتشكيل مجتمع جديد بالحلول التي تنتج أفضل الصفات.
- أهم ما يميز الخوارزمية الجينية :

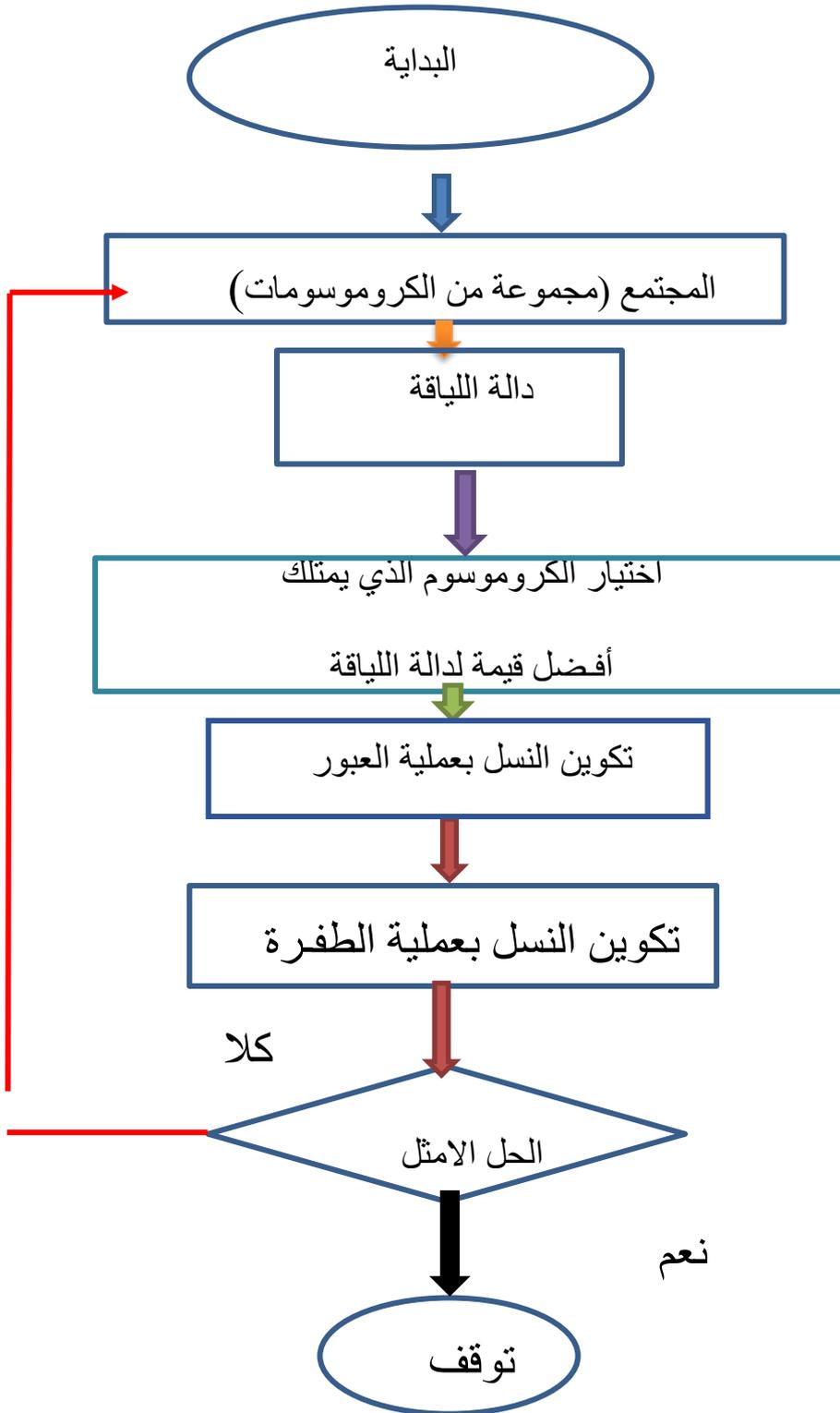
1- إنَّ أهم ما يميز الخوارزمية بأنَّها تبحث في مجتمع يكون عبارة عن مجموعة من المشاهدات وليس مشاهدة واحدة .

2- تتميز الخوارزمية الجينية بأنَّها لا تتوسع في معلومات إضافية بل تستخدم دالة الهدف مباشرة (Objective function) .

3- تمتاز بأنَّها تستخدم القوانين الاحتمالية ولا تستخدم القوانين الكلاسيكية.

4- تمتاز الخوارزمية الجينية بأنَّها لا تتعامل مع المتغيرات نفسها ولكن تتعامل مع التفسير لمجموعة المتغيرات.

ويمكن توضيح خطوات الخوارزمية الجينية من المخطط الانسيابي الآتي:



مخطط (1) يبين خطوات الخوارزمية الجينية من المخطط الانسيابي من اعداد الباحثة

(2-3-9-2) خطوات عمل الخوارزمية الجينية [47] [53]:

Steps Work Genetic Algorithms

توجد عدة خطوات تجرى داخل الخوارزمية الجينية يمكن حصرها بالنقاط الآتية:

- 1- يتم تغيير الكروموسومات الأصلية بالاعتماد على تقييم الأبناء وإن كل كروموسوم يتكون من عدد من القيم ويتم تحديد عدد هذه القيم حسب المسألة وقيمة البداية، وتكون عشوائية ضمن المحددات من المسألة المراد حلها.
- 2- يجب ان يتم اختيار أفضل الأباء لكي تتم عملية انتاج الابناء.
- 3- يتم توليد جيل جديد باستخدام الطفرة والعبور.
- 4- بالاعتماد على الدالة الأصلية يتم تقييم الأبناء الجدد.

(3-3-9-2) مكونات الخوارزمية الجينية:

Components A Genetic Algorithm

تتألف الخوارزمية الجينية من عدة عناصر وهي:

1- المجتمع (Population) [22]:

يتكون المجتمع عادة من الافراد التي تتمثل بالكروموسومات وكل كروموسوم يتكون من عدد من الجينات. ويمكن تحديد حجم المجتمع حسب طبيعة الدراسة المراد حلها حيث اشارت بعض الدراسات الى أنّ حجم المجتمع المثالي حوالي (30-20) فرد وتكون عادة بدايته قيمة عشوائية ضمن حدود الدراسة.

2- دالة اللياقة (التقييم) (Fitness Function (Evaluation) [29]:

وتعرف بأنها دالة الهدف للخوارزمية الجينية ومن أهم المكونات الأساسية للخوارزمية، حيث يتم إعطاء كل كروموسوم قيمة معينة تمثل مدى كفاءته أو مدى اقترابه من الحل من خلالها، لذا تكون دالة الهدف (دالة اللياقة) اما عظمى فهي تعتمد على تعظيم (Maximization) المسألة (أي الحصول على أعلى قيمة لدالة اللياقة) وتستخدم في حساب

الإنتاج والربح، أو صغرى فهي تعتمد على تصغير (Minimization) المسألة (أي الحصول على أدنى قيمة لدالة اللياقة) وتستخدم في حساب الخطأ والجذور التي تحقق المعادلات.

1-التشفير (Encoding)^{[32][37][50]}:

يعد التشفير عاملاً أساسياً مهماً لنجاح الخوارزمية الجينية حيث يقصد به إيجاد التمثيل البرمجي المناسب لمجموعة الحلول العشوائية (الكروموسومات) بهدف الوصول الى الحل الأمثل. اذ هناك عدة طرائق للتشفير وهذا يعتمد على المسألة المراد حلها، ومن هذه الطرائق المستخدمة في تشفير الحلول هي:

أ- التشفير الثنائي (Encoding Binarr):

يعد هذا النوع من التشفير من أقدم أنواع التشفيرات وأكثرها استخداماً مع الخوارزمية الجينية فضلاً عن بساطته النسبية. لذلك يعد من أشهر الطرائق في تمثيل الكروموسومات على شكل سلسلة أرقام متكونة من رقمين تأخذ مرة (1) ومرة (0) ويمكن توضيح هذا النوع من التشفير من خلال الشكل الآتي:

1	0	1	0	1	0	1
---	---	---	---	---	---	-------	---

الشكل (11-2) التشفير الثنائي للكروموسوم^[37]

ب- التشفير الاستبدالي (Permutation Encoding):

يعد هذا النوع من التشفير كأحدى خوارزميات التشفير الالكلاسيكية، حيث كل موقع بالكروموسوم يحمل رقماً بين (0-9) ويمكن توضيح هذا نوع من التشفير، كما في الشكل الآتي:

2	1	5	4	8	3	9	2	1
---	---	---	---	---	---	---	---	---

الشكل (12-2) التشفير الاستبدالي للكروموسوم^[32]

ت- التشفير الحقيقي (Real Value Encoding):

يعد هذا النوع من التشفير من أسهل التشفيرات وأكثرها استخداماً من قبل المبرمجين، حيث يمكن تمثيل الكروموسومات على هيئة سلسلة من القيم حسب طبيعة المسألة قد تكون سلسلة على شكل حروف أو اتجاهات أو سلسلة من الأعداد الحقيقية. ويمكن توضيح هذا النوع من التشفير بالشكل الآتي:

4.33	2.43	5.89	1.45
------	------	------	------

الشكل (2-13) التشفير الحقيقي للكروموسوم^[50]

ث- التشفير الشجري (Tree Encoding):

ينفذ هذا النوع من التشفير في البرامج الجينية والبرامج التطورية، يكون فيها الكروموسوم على هيئة شجرة تحتوي على عدد من العقد (Nodes) وكل عقده تمثل هدف معين كمتغيرات أو تنفيذ أوامر عملية رياضية في لغات البرمجة.

2- الانتقاء (Selection) [25] [37] [59]:

هو عملية اختيار الأبناء من المجتمع لأجل التزاوج وتكوين الجيل الجديد. وان لعملية الانتقاء (الاختيار) لها دور كبير في تطور الخوارزمية الجينية، حيث بعد أن يتم اختيار التشفير المناسب يصبح القرار الواجب اتخاذه هو كيفية اجراء الاختيار وفقاً لنظرية دارون التي تنص على مبدأ البقاء للأفضل أي بمعنى حصول الأفراد ذو الصلاحية الأعلى على فرصة أكبر للتزاوج والعيش من الأفراد ذو صلاحية الواطئة. وهناك عدة طرائق لانتقاء الكروموسومات المناسبة من الجيل القديم لتكوين الجيل الجديد وهي:

أ- الانتقاء النسبي (Proportional Selection):

يتم هذا النوع من الانتقاء (الاختيار) من خلال الاعتماد على نسبة قيمة اللياقة لكل فرد مقسوماً على مجموعة قيم دوال اللياقة لأفراد المجتمع، حيث يتم الاختيار حسب المسألة المراد حلها أي انتقاء زوج من الكروموسومات التي تمتلك أفضل دالة لياقة. وذلك حسب المعادلة الآتية:

$$Pepo(x) = \frac{f(X_i(t))}{\sum_{i=1}^N f(X_i(t))} \dots \dots \dots (17 - 2)$$

حيث إن:

$Pepo(x)$ تمثل قيمة نسبة اللياقة

$i =$ تمثل رقم الكروموسوم

$f(x) =$ تمثل قيمة دالة اللياقة لكروموسوم

$t =$ تمثل رقم الجيل

$N =$ تمثل عدد الكروموسومات

ج- الانتقاء الصفي (Rank Selection):

هذا النوع من الانتقاء يتم من خلال ترتيب الكروموسومات ثم احتساب درجة اللياقة لكل واحد ترتيباً تصاعدياً فكل كروموسوم (x) يكون له رتبة $Rank(x)$ أي موقع بعد الترتيب لذلك يطلق عليه بالانتقاء الرتب، وبعد ذلك يتم حساب نسبة الهدف (Target sampling Rate) لكل كروموسوم حسب المعادلة الآتية:

$$TSR(x) = \text{Min} + (\text{Max} - \text{Min}) \frac{\text{Rank}(x) - 1}{N - 1} \quad \dots \quad (18-2)$$

حيث أن:

$TSR(x)$: تمثل نسبة الهدف لكل كروموسوم.

N : تمثل عدد الكروموسومات بالجيل، إن الكروموسوم الأسوأ يأخذ القيمة 1 والثاني يأخذ 2 من قيمة TSR الى إن نصل إلى الكروموسوم الأفضل.

Min : تمثل الحد الأدنى لدرجة اللياقة لكل كروموسوم.

Max : تمثل الحد الأعلى لدرجة اللياقة لكل كروموسوم.

د- انتقاء النخبة (Elitist Selection):

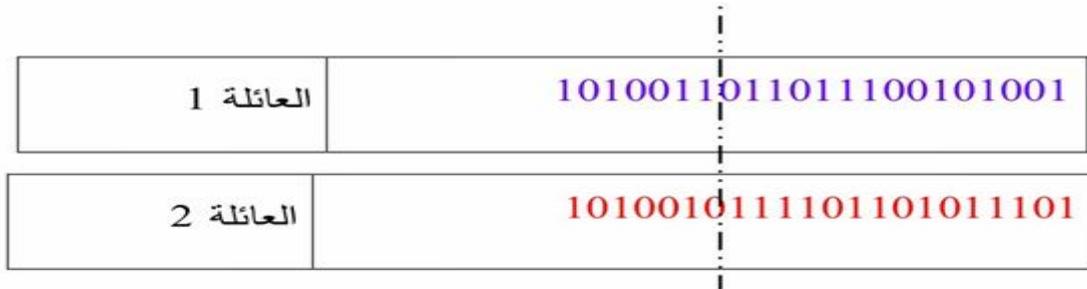
يعد هذا النوع من الانتقاء هو الأسهل على الإطلاق حيث يتم اختيار الفرد الذي يملك أكبر قيمة لدالة اللياقة فيما إذا كانت المشكلة تحقق الحد الأقصى (Maximization) وبالعكس يتم اختيار الفرد الذي يملك أقل قيمة لدالة اللياقة في حالة المشكلة تحقق الحد الأدنى (Minimization)، أي بمعنى يتم انتقاء الحلول الأفضل من المجتمع وإلغاء باقي الحلول منه.

3-العبور (Crossover) [25]:

هي العملية التي بموجبها يتم تبادل بعض أجزاء من الحلول بين أبوين لينتج حلول جديدة، أي بمعنى بعد اختيار الكروموسومات الجيدة من الجيل الأول يتم التزاوج بين كل كروموسومين لتكوين الجيل الجديد (الابناء) اعتمادا على الكروموسومات الأم ويأخذ صفات الجيدة منها، والغاية من عملية العبور هي السماح للطفل لتسلم أفضل الأجزاء المختلفة من آبائه. وهناك عدة طرق لعملية العبور وهي:

أ- العبور بنقطة واحدة (Single Point Crossover):

في هذا النوع من العبور يتم اختيار كروموسومين وتحديد نقطة ما عشوائية على طول السلسلة ويتم القطع، حيث يتم تبادل المادة الوراثية عند هذه النقطة لتكون لدينا طفلين جديدين. ويمكن توضيح ذلك من خلال الشكل الآتي:



الشكل (2-14) العبور بنقطة واحدة [25]

حيث أنّ الإشارة | تمثل موقع نقطة العبور، بعد التبادل بين الآباء عند هذه النقطة يكون لدينا طفلان جديدان.

ب- العبور بنقطتين (Two Point Crossover):

في هذا النوع من العبور يتم اختيار نقطتين عشوائياً ضمن حدود الكروموسوم ويتم القطع عندهما، بعد ذلك يتم تبادل المادة الوراثية بينهما لإنتاج طفلين جديدين كما في الشكل الآتي:

العائلة 1	1010011011011100101001
العائلة 2	101001011110110101011101
الطفل 1	1010010111101100101001
الطفل 2	101001101101110101011101

الشكل (2-15) العبور بنقطتين^[25]

ح- العبور المنتظم (Uniform Crossover):

هو أحد أهم أنواع العبور وأكثرها موثوقية حيث يتم في هذا النوع من العبور تخصيص الصور (للاب)، عند رمي قطعة النقود سوف تأخذ جينات الطفل الأول جينات الأب الذي سيظهر عند رمي قطعة النقود، أما الطفل الثاني سوف يأخذ عكس جينات الأب الآخر .

د- العبور نقطة (n) البسيطة (Simple n-Point Crossover):

يتم هذا النوع من العبور في عدة نقاط عشوائية من سلسلتين من الكروموسومات , بعد ذلك يتم تبادل الكروموسومات الوراثية أي الجينات الواقعة بين نقطة العبور وبين أجزاء الكروموسومات ويتم هذا التبادل بشكل دائري .

4-الطفرة (Mutation) [31] [57] :

هي العملية التي تسهم في خلق الأجيال الجديدة عن طريق الخطأ العشوائي بتغيير أو تبديل بين الجينات المحدودة ضمن الكروموسوم الواحد لإتشاء كروموسومات تعطي حولا

جديدة في الجيل اللاحق. ويتم عادة هذا التغير بالاعتماد على نسبة الطفرة التي تكون قيمتها صغيرة جدا وفي حالة حصول الطفرة فإنَّ حولا جديدة (لم يسبق تكوينها في الأجيال السابقة) تضاف إلى المجتمع الجديد (قد يكونون أفضل وأساء من الأفراد السابقين)، الهدف من هذه العملية هو زيادة أكبر عدد ممكن من الأفراد ضمن المجتمع. وهناك عدة طرائق للطفرة الوراثية:

أ- الطفرة الطبيعية (Normal Mutation):

يتم تطبيق هذه الطفرة على الجينات ذات الاعداد الصحيحة، حيث يتم إضافة قيمة عشوائية من أنموذج الطبيعي على بعض الجينات المختارة من كروموسومات الآباء لإنتاج الأبناء.

ب- الطفرة المنتظمة (Uniform Mutation):

يستخدم هذا النوع من الطفرة مع التشفير الحقيقي أو الصحيح حيث يتم استبدال قيمة الجين المختار من الكروموسوم بقيمة عشوائياً يعرفها المستخدم ضمن الحدود، لذا يختار موقعا عشوائية يكون محصور بين أعلى قيمة وأقل قيمة في الكروموسوم.

ت- الطفرة الرتب (Bit Mutation):

يستعمل هذا النوع من الطفرة مع التشفير الثنائي، حيث يتم استبدال قيمة الجين من الكروموسوم حيث يختار موقعا ضمن حدود الكروموسوم فاذا كانت القيمة هي (0) تستبدل بالقيمة (1) وإذا كانت القيمة (1) تستبدل بالقيمة (0).

5- مقياس التوقف (Stop Criterion) [29]:

يقصد بمقياس التوقف في الخوارزمية الجينية بأنه يحدد اذا كانت الجينية سوف تستمر بالرسالة لإنتاج الأجيال المتعاقبة بهدف تحسين الحل او انها سوف تتوقف اذا تحققت شروط التوقف، الذي 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 أو الحصول على أفضل قيمة لدالة اللياقة في المجتمع الذي تم التوصل إليه عندما تكون أقل من قيمة اللياقة التي حددت في حالة (Minimize) أو تتوقف عندما تكون أحسن قيمة لدالة اللياقة أكبر من قيمة اللياقة التي حددت في حالة (Maximize)

وبعد تكوين كل جيل جديد يتم فحص مقياس التوقف للخوارزمية الجينية لنرى اذا تم الوصول الى الحل الأمثل.

(2-9-3-4) خطوات إجراء الخوارزمية الجينية لتقدير المعلمات لأنموذج الانحدار اللوجستي

المتعدد:

Steps of conducting a genetic algorithm to estimate parameters of a multiple logistic regression model

1- يتم توليد مجتمع ابتدائي بشكل عشوائي يحتوي على N من الكروموسومات وكل كروموسوم يكون طوله (n) إذ إن n تمثل عدد المعلمات الموجودة في أنموذج النهائي لدالة اللوجستية.

2- يتم استخدام تشفير حقيقي، أي بمعنى يتم تشفير الكروموسوم وذلك من خلال تشفير المعلومات الموجودة ضمن كل كروموسوم لكي يتم تمييز الكروموسومات .

3- دالة اللياقة: إن قيمة دالة اللياقة في هذه الخوارزمية تمثل قيمة معيار متوسط مربعات الخطأ.

4- الانتقاء: هي عملية اختيار الآباء في المجتمع لأجل التزاوج وإنتاج جيل جديد، وبذلك يتم اختيار الآباء من خلال انتقاء عجلة.

5- العبور: تتم اختيار العبور بنقطة القطع، أي بمعنى يتم العبور بأجراء تبديل بين قيم متقابلة من كروموسومي الأبوين المنتقيين لتكوين الفرد الجديد.

6- الطفرة: تتم عملية الطفرة باستخدام الطفرة المنتظمة. أي إنه يتم التبديل بين قيم الفرد الواحد لتكوين أفراد تعطي حلاً جديداً إلى الجيل اللاحق لم يسبق تكوينها في الأجيال السابقة وذلك بهدف توسيع الحلول الممكنة.

7- شرط التوقف: إن المقياس الذي تم اعتماده في الخوارزمية الجينية لتحقيق شرط التوقف في حالة عدم الوصول للحل المثالي أو الأمثل، هو $(n \times 100)$ حيث إن n تمثل عدد المعلمات في أنموذج النهائي لدالة اللوجستية .

(2-9-4) طريقة المربعات الصغرى الموزونة [48] [33] [4]:

Weighted Least Squares Method (WLSM)

في هذه الطريقة يتم تحويل معادلة المنحني اللوجستي الى معادلة خطية يمكن منها تقدير معلمات أنموذج المختار. إنَّ معادلة المنحني اللوجستي هي

$$\pi_i = \frac{e^{B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_k X_{ki}}}{1 + e^{B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_k X_{ki}}} \quad \dots \dots \quad (19 - 2)$$

حيث π تمثل احتمال النجاح الحقيقي والتي يمكن التعبير عنها بالدالة التجميعية (C.D.F) ، وبذلك فإنَّ احتمال نجاح العينة المشاهدة وقيمة P_i يمكن التعبير عنها بانها احتمال الحقيقي π_i مضافاً اليها الأخطاء العشوائية (u_i) أي أنَّ

$$P_i = \pi_i + u_i \quad \dots \dots \quad (20 - 2)$$

وعلى فرض أنَّ الأخطاء العشوائية u_i مستقلة (independent) و تتوزع توزيعاً ذي الحدين (binomial distribution) بمتوسط صفر وتباين $\left(\frac{\pi_i(1-\pi_i)}{n_i}\right)$ ، فإنَّ احتمال فشل استجابة العينة المشاهدة $(1 - P_i)$ سيكون

$$1 - P_i = 1 - \pi_i - u_i \quad \dots \dots \quad (21 - 2)$$

وبقسمة المعادلة (20-2) على المعادلة (21-2) نحصل على

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

حيث

$$Q_i = 1 - \pi_i$$

وبأخذ (ln) لطرفي المعادلة

$$\ln \frac{P_i}{1 - P_i} = \ln \frac{\pi_i}{1 - \pi_i} + \ln \left(\frac{1 + u_i/\pi_i}{1 - u_i/Q_i} \right) \quad \dots \dots \quad (22-2)$$

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

ومن معادلة (20-2) فإن

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

وللبحث عن العلاقة بين القيم الحقيقية والمشاهدة للوغاريتم النسبة المضافة (logit) سنفرض

$$f(P_i) = \ln\left(\frac{P_i}{1-P_i}\right) \quad , \quad f(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) \quad \dots\dots \quad (23-2)$$

وباستخدام صيغة تايلر من درجة الـ 1 أولى حول π_i للمعادلة (19-2) فإن

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

وبمقارنة هذه المعادلة بالمعادلة (21-2) ومساواة المشتقة الى صفر نحصل على :

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

وباشتقاق المعادلة (21-2) 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 فإن معادلة (21-2) ستكون

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

على فرض أن Logit للقيم الحقيقية هو y_i^* أي أن $\ln \frac{\pi_i}{1-\pi_i} = y_i^*$ للقيم الحقيقية فإن

$$y^*_i = B_0 + \sum_{i=1}^n B_i X_i \quad \dots \dots \quad (24 - 2)$$

وبذلك تم تحويل المعادلة (19-2) غير الخطية الى معادلة خطية يمكن تقدير معاملات الانحدار منها وإن المتغير الذي تم تحويله y^*_i له متوسط مساو تقريبا الى $\ln\left(\frac{\pi_i}{1-\pi_i}\right)$ بعبارة أخرى مساواه الى $(X_i B)$ وتباين σ^2_i حيث أن:

$$\sigma^2_i = \frac{1}{n\pi_i(1-\pi_i)} \quad \dots \dots \quad (25 - 2)$$

ان انحدار المتغير الذي تم تحويله y^*_i يعتمد على X_i له تباين غير ثابت (غير متجانس) لأنه يعتمد على قيمتي (n, P_i) وطالما إن قيمة P_i تعتمد على X_i وب الأتي فإن y^*_i يعتمد على X_i بذلك سيفقد نموذج أغلب افتراضاته الخاصة ومن هذه الافتراضات سيكون تباين الخطأ العشوائي غير ثابت وب الأتي قد يعاني من مشكلة عدم تجانس التباين وعندها لا يمكن تقدير معاملات نموذج باستعمال طريقة المربعات الصغرى الاعتيادية (OLS) وبذلك تم اللجوء الى طريقة المربعات الصغرى الموزونة (WLS) لتقدير معلما أنموذج اللوجستي المتعدد ليتم الحصول على أفضل تقدير. لذلك لابد من ضرب المعادلة الانحدار (24-2) بالجذر التربيعي للوزن المرافق $\sqrt{W_i}$ لكي نحصل على تجانس للتباين وان :

$$W_i = n_i P_i (1 - P_i)$$

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

وبذلك تصبح المعادلة (24-2) بعد ضربها بالوزن كالآتي:

$$\sqrt{W_i} y^*_i = B_0 + \sqrt{W_i} X_i B_i + u_i \quad \dots \dots \quad (26 - 2)$$

بذلك يمكن تقدير معاملات أنموذج (B_0, B_1, \dots, B_p) في المعادلة (26-2) 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.

$$SSE_i = \sum W_i (Y_i - \hat{Y}_i)^2 \quad \dots \dots \quad (27-2)$$

$$\beta = X'WY X'WX$$

حيث إن:

Y: يمثل متجه التحويل الخطي (اللوجت) لنموذج ذي رتبة (n×1)

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

X'WX: تمثل مصفوفة مربعة حاصلة ضرب مصفوفة الأوزان مع مصفوفة المتغيرات

التوضيحية ذات رتبة [(p+1) (p+1)]

XWY: تمثل متجه مبدلة مصفوفة المتغيرات التوضيحية ومصفوفة الأوزان مع متجه التحويل

الخطي ذي رتبة [(p+1) × n]

وبضرب طرفي المعادلة ب XWX^{-1} وبذلك سوف نصل الى قيمة B^{\wedge} المقدرة والتي تكون:

$$\hat{B} = (X'WX)^{-1}X'WY \quad \dots (29-2)$$

(10-2) اختبار جودة التوفيق لأنموذج الانحدار اللوجستي [14][33][34]:

يتم اختبار جوده التوفيق لا نموذج الانحدار اللوجستي بعدة من اختبارات ومن هذه الاختبارات (اختبار مربع كاي لنسبة الترجيح، اختبار هوزمر_ليمشو، اختبار معامل التحديد R^2 ، اختبار wald، لكي يتم معرفة إلى أي مدى يتناسب أنموذج المقدر مع البيانات. وسوف نختصر بدراستنا على اختبارين فقط هما:

(1-10-2) اختبار (Wald):

يقيم اختبار Wald (الذي سمي على اسم ابراهام والد) القيود المفروضة على معاملات الإحصائية بناء على المسافة الموزونة بين التقدير غير المقيد وقيمه المفترضة في ظل الفرضية الصفرية والتي تنص على (إن تأثير معامل لوجت ما يساوي صفرا)، لذا يعد أحد الأساليب الكلاسيكية (الثلاثة) لاختبار الفرضيات، وميزة هذا الاختبار عن الاختبارات

الأخرى هو أنه لا يتطلب سوى تقدير أنموذج غير المقيد، وإنَّ الفرضية الإحصائية للنموذج كالاتي:

H_0 : تأثير معامل لوجيت يساوي صفر.

H_1 : تأثير معامل لوجيت لا يساوي صفر.

ويمكن حساب احصاء الاختبار حسب الصيغة الآتية

$$wald = \frac{b}{SE_b} \quad \dots \dots \dots (30 - 2)$$

حيث إنَّ :

b : هي قيمة معامل الانحدار اللوجستي للمتغير المستقل .

SE_b : هي قيمة الخطأ المعياري لمعامل الانحدار اللوجستي للمتغير المستقل .

علماً إنَّ الاحصاء Wald تتبع توزيع (χ^2) بدرجة حرية $(c-1)(r-1)$ ، كما إنَّ هذا الاختبار هو اختبار من الطرفين ويجب إنَّ تكون قيمة معنوية المعلمات المناظرة لقبول أو رفض فرضية العدم باستعمال الاحتمالات التي تكون أقل من (0.05) لكي يتم رفض الفرضية وقبول الفرضية البديلة والتي تنص على (إنَّ تأثير المعامل لوجيت لا يساوي صفر)، أي إنَّ المتغير المستقل له تأثير بالتنبؤ بقيمة المتغير التابع.

(2-10-2) اختبار هوزمر ليمشو (Homsre Lemeshow) [18]:

يعتبر اختبار هوزمر ليمشو أحد الاختبارات التي تستخدم في جوده التوفيق لأنموذج الانحدار اللوجستي. ويعتمد هذا الاختبار على مدى قرب الاحتمالات المتوقعة والاحتمالات المشاهدة. حيث يعمل على أساس جميع حالات قيم الاحتمالات المتوقعة، لذا يعد هذا الاختبار مشابه لحد ما لاختبار (χ^2) لحسن المطابقة حيث يستخدم لتقييم حسن المطابقة لأنموذج وهو يسمح بأي عدد من المتغيرات التوضيحية سواء كانت متقطعة أو مستمرة، وإنَّ هذه الاحصاءة تتوزع توزيع (χ^2) بدرجة حرية $(c-1)(r-1)$ ، واذا كانت قيمتها أكبر من مستوى المعنوية فإنَّ هذا يؤكد جودة التوفيق لأنموذج بالكامل، حيث يقوم هذا الاختبار على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل اعمده

، ما صفوف تقسم على أساس القيم المشاهدة لمتغير المعتمد وهما الصفر والو أحد. تكون
الفرضية الإحصائية للنموذج كالاتي:

H_0 : الأنموذج المقدر يوافق البيانات بشكل جيد .

H_1 : الأنموذج المقدر لا يوافق البيانات بشكل جيد .

ويمكن ترميز احصاءة الاختبار هذه بالرمز C ويتم احتسابها وفقاً للصيغة الآتية: -

$$C = \sum_{i=1}^k \frac{(O_i - n_i \bar{P}_i)^2}{n_i \bar{P}_i (1 - \bar{P}_i)} \quad \dots \dots \dots (31 - 2)$$

إذ إن:

$$O_i = \sum_{j=1}^{n'_i} Y_j \quad \text{تمثل القيم المشاهدة للمجموعة}$$

$$\bar{P}_i = \sum_{j=1}^{n'_i} \frac{p_i}{n'_i} \quad \text{تمثل القيم المتوقعة للمجموعة}$$

(11-2) معايير المقارنة بين طرائق التقدير المستعملة :

Comparison Criterion Between Used Estimation Methods

إنَّ المقارنة بين الطرائق المستخدمة تعتمد على عدة معايير حيث يمكن من خلالها تحديد
افضل طريقة لتقدير، ومن ضمن هذه المعايير هي:

(1-11-2) معيار متوسط مربعات الخطأ :

Criterion mean square error (MSE)

يعد معيار متوسط مربعات الخطأ كمقياس مقارنة بين طرق التقدير المستخدمة حيث يشير
الى مدى دقة التقدير، وإنَّ تناقص قيمته تشير الى جودة ودقة المقدرات والهدف من هذه
المعايير هو الحصول على مقدر $\hat{Y}_i = \hat{Y}(X_i)$ في صيغة الدالة اللوجستية ويتم حسابه
كالاتي :

$$MSE = \frac{1}{n - k} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad \dots \dots \dots (32 - 2)$$

حيث أن:

Y_i : تمثل القيمة التقديرية للملاحظات.

n : تمثل حجم العينة .

K : تمثل عدد المعلمات.

(2-11-2) معيار متوسط المطلق الخطأ :

Mean Absolute Error(MAE)

وهو أحد المقاييس الدقيقة الشائعة الاستعمال، لذا يكون عبارة عن معدل الانحرافات عن القيم الحقيقية وصيغته:

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad \dots \dots \dots (33 - 2)$$

حيث إن:

e_t : الفرق بين القيم الحقيقية والقيم المشاهد



الفصل الثالث

الجانِب العملي

(التجريبي والتطبيقي)

الفصل الثالث

الجانب العملي (التجريبي والتطبيقي)

preamble (1-3) التمهيد :-

يتضمن هذا الفصل مبحثين (الجانب التجريبي والجانب التطبيقي). وفي الجانب التجريبي تم استخدام أسلوب المحاكاة على بيانات مولدة باستخدام أسلوب مونت-كارلو (Monte-Carlo) اما الجانب التطبيقي فيتضمن تطبيق عملي على بيانات حقيقية (تجربة حياتية).

(2-3) الجانب التجريبي :-

يتناول هذا الجانب استعراض مفهوم المحاكاة ومراحل وصف تجربة المحاكاة من حيث حجم العينات والقيم الافتراضية وطرائق التقدير المستخدمة لتقدير الانحدار اللوجستي متعدد الاستجابة والمعايير الإحصائية المستخدمة في تجربة المحاكاة، فضلا عن ان المحاكاة تم اعدادها بالاستعانة بالبرنامج الاحصائي (R.14) .

(1-2-3) مفهوم المحاكاة [56][58] :-

هي طريقة علمية عددية تعطي القدرة على التجريب والاختيار من خلال تكرار العملية لمرات عديدة. اي بمعنى ان المحاكاة هي عبارة عن محاولة إيجاد صورة طبق الأصل من أي نموذج ونظام بدون أخذ أو الرجوع الى ذلك النظام والأنموذج ذاته. لذا يتم استخدام أسلوب المحاكاة بواسطة مجموعة من العينات ويتم توليد هذه العينات من مجتمع افتراضي بعد ذلك يتم تطبيق النماذج والأساليب الإحصائية والرياضية عليها لغرض اجراء المقارنة بين الطرائق المعتمدة في تقدير الأنموذج المدروس من خلال المقاييس الاحصائية

وعندها يتم اتخاذ القرار المناسب على ضوء هذه النتائج. حيث كلما كانت نتائج المحاكاة واقعية ودقيقة كلما كانت التجربة أقرب الى الحقيقية.

(2-2-3) توليد الارقام والمتغيرات العشوائية:

يتم توليد المتغيرات العشوائية وفق التوزيع المنتظم المستمر (Continues uniform distribution) ضمن الفترة {1,0} حسب الدالة الآتية

$$f(x) = \frac{1}{b - a} \quad a \leq X \leq b$$

حيث تعد عملية توليد الأرقام العشوائية عملية تكرارية تعطي في كل مرة عدداً عشوائياً مختلفاً عن العدد الاخر ويكون محدد بالقيم (1,0) ومن ثم يتم تحويل العدد العشوائي الى متغير عشوائي يصف الانموذج قيد الدراسة.

اما توليد المتغيرات المستقلة فيتم بالاعتماد على الدالة التجمعية للتوزيع توزيع ثنائي الحدين والتي تكتب بالصيغة الآتية:

$$P(y_i \setminus X_i) = [\pi(X_i)]^{y_i} [1 - \pi(X_i)]^{1 - y_i}$$

والغرض من توليد المتغيرات المستقلة (X) هو الحصول على استجابات متعددة للنموذج، وهذا بدوره سوف يعطي المتغيرات المستقلة استجابات مناضرة للمتغير المعتمد (Y) وهذه الاستجابات تكون ضمن الفترة (1,0).

(3-2-3) مراحل بناء ووصف تجربة المحاكاة:

تتضمن مرحلة بناء تجربة المحاكاة أربع مراحل لغرض تحليل البيانات وهي:

المرحلة الأولى: وهي من أهم المراحل التي تعتمد عليها بقية مراحل المحاكاة حيث يتم اختيار ثلاثة حجوم للعينات تمثل العينات (صغيرة, متوسط, كبيرة) وكانت على التوالي (N=25,75,125) وبتكرار تجربة (R=1000) مرة عند كل حجم , حيث تم اختيار القيم

الافتراضية استناداً للدراسات السابقة القريبة من هذه الدراسة وكذلك من خلال خبرة الطالب بهذا الموضوع كما موضح في الجدول الآتي :

جدول (1-3) يوضح القيم الافتراضية

Model	b_0	b_1	b_2
1	0.5	-0.4	0.3
2	0.2	-0.3	0.4
3	0.6	0.5	-0.2

المرحلة الثانية : وهي المرحلة التي يتم فيها توليد قيم عشوائية للمتغيرات المستقلة تتبع توزيع ثنائي الحدين (Binomial) للنموذج الافتراضي الذي يعطي استجابة مناظرة للمتغير المعتمد (y) ويخضع هذا المتغير الى التوزيع المتعدد الحدود (Multinomial) .

المرحلة الثالثة : وهي المرحلة التي يتم فيها إيجاد تقدير معالم نموذج الانحدار اللوجستي متعدد الاستجابة التي تم التطرق اليها في الجانب النظري وفق الطرق التي تم استخدامها وذكرها في الجانب النظري المتمثلة بـ (الشبكات العصبية, الإمكان الأعظم, الخوارزمية الجينية, المربعات الصغرى الموزونة).

المرحلة الرابعة : وهي المرحلة الأخيرة من مراحل وصف تجربة المحاكاة وفي هذه المرحلة تم المقارنة بين طرائق التقدير بعد إيجاد المقدرات في المرحلة الثالثة باستخدام المقاييس الإحصائية المستخدمة لغرض الحصول على أفضل طريقة تقدير للنموذج قيد الدراسة ومن هذه المقاييس:

1- معيار متوسط مربعات الخطأ (MSE) وحسب الصيغة الآتية:

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

حيث ان :

\hat{Y}_i : تمثل القيمة المقدرة للمشاهدات

n : تمثل حجم العينة, K : تمثل عدد المعلمات.

2- معيار متوسط مطلق الخطأ (MAE) وحسب الصيغة الآتية:

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$$

(4-2-3) تحليل نتائج تجربة المحاكاة:

سيتم تحليل وعرض نتائج تجربة المحاكاة والوصول الى أفضل مقدرات للنموذج اللوجستي متعدد الاستجابة حسب ثلاث حالات من القيم الافتراضية التي تم افتراضها بالاعتماد على المقاييس الإحصائية و يتم الحصول على كافة نتائج تجربة المحاكاة من خلال استخدام برنامج لغة (R) وحسب طرائق التقدير المستخدمة [الشبكات العصبية (ANN), الإمكان الأعظم (MLE), المربعات الصغرى الموزونة (WLS), الخوارزمية الجينية (G)].

الحالة الأولى : من القيم الافتراضية (0.5,-0.4,0.3)

تم اختيار قيم افتراضية لمعالم النموذج بناءً على دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة على ان تكون قيم هذه المعالم $(B_0 = 0.5, B_1 = -0.4, B_2 = 0.3)$ حيث تم استخدامها في كل من الطرائق التقدير وكما يأتي :

1- طريقة الشبكات العصبية (ANN) Neural Networks :

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة بطريقة الشبكات العصبية باستخدام برنامج لغة (R) عند حجوم العينات (125,75,25) حسب القيم الافتراضية التي تم افتراضها على التوالي (0.5,-0.4,0.3) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة ومن خلال تحديد الدالة العصبية المستخدمة وتحديد طبقات المدخلات التي تعتمد على عقد Nodes تستخدم لتوزيع الأوزان وباستخدام المعادلات التي تم ذكرها في الجانب النظري (2-13) (2-14) (2-15) فتم الحصول على النتائج مع معايير المفاضلة كما موضح بالجدول (3-2) ولحجوم عينات (صغيرة ، متوسطة، كبيرة)

جدول (2-3) يوضح قيم الأوزان مع الحد الثابت للشبكات العصبية عند حجوم العينات (25,75,125) حسب القيم الافتراضية (0.5,-0.4,0.3) فضلا عن معايير المفاضلة.

الحجوم			Hidden Layer1		Hidden Layer2		Output Layer
25	Input Layer		W(1,1)	W(1,2)	W(2,1)	W(2,2)	
		Bais	-0.1389	0.0646			
		$input_1$	0.6602	-0.4499			
		$input_2$	-0.7312	1.9678			
	Hidden Layer1	Bias			-1.9881	-0.3039	
		W (1,1)			-0.1975	-0.4447	
		W (1,2)			0.8809	2.6232	
	Hidden Layer2	Bias					-1.0288
		W (2,1)					-0.7124
		W (2,2)					2.1541
MSE							0.7860
MAE							0.7005
75	Input Layer	Bias	-1.2828	-1.9613			
		$input_1$	0.1901	-19.6751			
		$input_2$	1.1588	6.5108			
	Hidden Layer1	Bais			0.6241		
		W (1,1)			1.5566		
		W (1,2)			1.4116		
	Hidden Layer1	Bais					-0.8438
		W (2,1)					1.6040
		W (2,2)					
	MSE						
MAE		0.7559					
125	Input Layer	Bais	-1.4603	0.3233			
		$input_1$	-1.1805	0.2635			
		$input_2$	1.6657	0.4012			
	Hidden Layer1	Bais			0.6344	-0.9870	
		W (1,1)			1.6612	0.1794	
		W (1,2)			-0.8142	-0.6685	
	Hidden Layer2	Bais					-1.0594
		W (2,1)					1.9789
		W (2,2)					1.3368
	MSE						
MAE							0.7749

الجدول من اعداد الباحثة

حيث أن :

Input Layer : تمثل طبقة المدخلات للشبكات العصبية وهي .

Hidden Layer1 : تمثل الطبقة الوسطى المخفية الأولى للشبكات العصبية.

Hidden Layer2 : تمثل الطبقة الوسطى المخفية الثانية للشبكات العصبية.

Output Layer : تمثل طبقة المخرجات للشبكات العصبية.

باستخدام برنامج (R) تم الحصول على نتائج بالجدول (2-3) حيث ان (Bais) يمثل الحد الثابت اما بالنسبة $(W(1,1))$ يمثل الوزن بين طبقة المدخلات وبين طبقة الوسطى المخفية الأولى و $(W(1,2))$ يمثل الوزن بين طبقة المدخلات وبين الطبقة الوسطى المخفية الثانية, وهكذا بالنسبة لباقي نتائج الجدول

2- طرائق التقدير (GA,WLS,MLE):

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستخدام طرائق التقدير المتمثلة بـ(الإمكان الأعظم Maximum Likelihood Method من خلال المعادلة (2-12)، المربعات الصغرى الموزونة Weighted Least Squares Method من خلال المعادلة (2-28)، الخوارزمية الجينية Genetic Algorithm) من خلال مخطط (1) و باستخدام برنامج لغة (R) عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على توالي (0.3,-0.4,0.5) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة وكما موضح بالجدول الاتي:

جدول (3-3) يوضح قيم معاملات الانحدار اللوجستي متعدد الاستجابة المقدرة باستخدام (الإمكان الأعظم، المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعاملات الطرائق	b_0	b_1	b_2	MSE	MAE
		25	MLE	-0.1525	-0.1725	0.4339
	WLS	-0.1861	-0.1541	0.4390	0.9068	0.7559
	GA	-0.1469	-0.1618	0.4183	0.9002	0.7541
75	MLE	0.2789	-0.2613	0.3045	0.9580	0.7804
	WLS	0.2389	-0.2349	0.3055	0.9588	0.7853
	GA	0.2789	-0.2613	0.3045	0.9582	0.7805
125	MLE	0.3389	-0.3727	0.3151	0.9782	0.7874
	WLS	0.3387	-0.3589	0.3015	0.9785	0.7897
	GA	0.2224	-0.3382	0.3583	0.9784	0.7887

الجدول من اعداد الباحثة

نلاحظ من الجداول (2-3) و(3-3) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستخدمة (NN, MLE, WLS, GA) التي تم ذكرها في الجانب النظري.

من خلال النتائج الموضحة في الجداول (2-3) و(3-3) وحسب الحالة الأولى من القيم الافتراضية (0.5, -0.4, 0.3) تبين ان :

1- عند حجم (25) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.7860) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7005) ثم تليها بعد ذلك طريقة الإمكان

الأعظم (MLE) حيث بلغ متوسط مربعات الخطأ (MSE) (0.8998) ونسبة الخطأ المطلق (MAE) (0.7307) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9002), (0.7541) بفرق بسيط جدا بينها وبين المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9068), (0.7559).

2- عند حجم (75) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.9019) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7559) ثم تليها بعد ذلك طريقة الإمكان الأعظم (MLE) حيث بلغ متوسط مربعات الخطأ (MSE) (0.9580) وبلغ نسبة الخطأ المطلق (MAE) (0.7804) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة خطأ مطلق (MAE) على التوالي (0.9582), (0.7805) بفرق بسيط جدا بينها وبين المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الأخطاء مطلق (MAE) على التوالي (0.9588) (0.7853).

3- عند حجم (125) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.9421) وأقل خطأ مطلق (MAE) ومقداره (0.7749) ثم تليها بعد ذلك طريقة الإمكان الأعظم (MLE) حيث بلغ متوسط مربعات الخطأ (MSE) (0.9782) وبلغ نسبة الخطأ المطلق (MAE) (0.7874) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9584), (0.7887) بفرق بسيط جدا بينها وبين المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9585), (0.7897).

الحالة الثانية : من القيم الافتراضية (0.2,-0.3,0.4)

تم اختيار قيم افتراضية لمعامل النموذج بناءً على دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة على ان تكون قيم هذه المعالم

(حيث تم استخدامها في كل من الطرائق $B_0 = 0.2, B_1 = -0.3, B_2 = 0.4$)

التقدير وكما يأتي :

1-طريقة الشبكات لعصبية (NN) Neural Networks :

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة بطريقة الشبكات العصبية بالاستخدام برنامج لغة (R) عند حجوم العينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على التوالي (-,0.2, 0.3,0.4) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة كما موضح بالجدول الاتي:

جدول (3-4) يوضح قيم الأوزان مع الحد الثابت للشبكات العصبية عند حجوم العينات (25,75,125) حسب القيم الافتراضية (0.2,-0.3,0.4) فضلا عن معايير المفاضلة.

الحجوم			Hidden Layer1		Hidden Layer2		Output Layer	
	Input Layer		W(1,1)	W(1,2)	W(2,1)	W(2,2)		
25	Input Layer	Bais	-0.1432	1.4454				
		input ₁	-6.2450	-8.0987				
		input ₂	6.2300	-1.1770				
	Hidden Layer1	Bais			0.2120	-0.4133		
		W (1,1)			1.5110	-1.3967		
		W (1,2)			-0.4194	3.7079		
	Hidden Layer2	Bais					-0.1759	
		W (2,1)					1.1408	
		W (2,2)					-0.2396	
	MSE							0.9070
MAE							0.7577	
75	Input Layer	Bais	-0.2607	1.5554				
		input ₁	-8.3450	-9.8652				
		input ₂	7.1388	-1.1890				
	Hidden Layer1	Bais			0.2120	-0.4133		
		W (1,1)			1.5110	-1.3967		
		W (1,2)			-0.4194	3.7079		
	Hidden Layer2	Bais					-0.1759	
		W (2,1)					1.1408	
		W (2,2)					-0.2396	
	MSE							0.9190
MAE							0.7370	
125	Input Layer	Bais	1.4226	-1.3762				
		input ₁	2.1092	-0.2197				
		input ₂	-2.7715	-4.1259				
	Hidden Layer1	Bais			0.5869			
		W (1,1)			-0.6618			
		W (2,2)			-2.5447			
	Hidden Layer2	Bais					0.3214	
		W (2,1)					1.6456	
	MSE							0.9499
	MAE							0.7766

الجدول من اعداد الباحثة

يمكن تعريف المصطلحات جدول (3-4) كما تم تعريفها في الجدول (3-2).

2-طرائق التقدير (GA, WLS, MLE)

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستخدام طرائق التقدير المتمثلة بـ(الإمكان الأعظم Maximum Likelihood Method، المربعات الصغرى الموزونة Weighted Least Squares Method، الخوارزمية الجينية Genetic Algorithm) باستخدام برنامج لغة (R) عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على التوالي (0.2,-0.3,0.4) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة وكما موضح بالجدول الآتي:

جدول (3-5) يوضح قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدره باستخدام (الإمكان الأعظم، المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعلمات	b_0	b_1	b_2	MSE	MAE
25	MLE	-0.1325	-0.1945	0.3369	0.9999	0.7559
	WLS	-0.1971	-0.1731	0.5590	0.8990	0.7449
	GA	-0.1589	-0.1917	0.3193	0.8003	0.7336
75	MLE	0.3468	-0.3594	0.5055	0.9471	0.7813
	WLS	0.3369	-0.3889	0.5698	0.9577	0.7815
	GA	0.3478	-0.3995	0.5743	0.9373	0.7812
125	MLE	-0.0524	-0.1668	0.4855	0.9801	0.5055
	WLS	-0.0674	-0.1634	0.4916	0.9811	0.7884
	GA	-0.0987	-0.1600	0.5055	0.9804	0.7883

الجدول من اعداد الباحثة

نلاحظ من الجداول (3-4) و(3-5) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستخدمة (NN,MLE,WLS,GA) التي تم ذكرها في الجانب النظري.

من خلال النتائج الموضحة في الجداول (4) و(5) وحسب الحالة الأولى من القيم الافتراضية (0.2,-0.3,0.4) تبين ان :

1- عند حجم (25) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.9070) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7577) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ متوسط مربعات الخطأ (MSE) (0.8003) ونسبة الخطأ المطلق (MAE) (0.7336) ثم تليها بعد ذلك طريقة المربعات الصغرى الموزونة (WLS) حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.8990), (0.7449) ثم تأتي من بعد ذلك طريقة الإمكان الأعظم حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9999), (0.7559).

2- عند حجم (75) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.9091) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7370) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ متوسط مربعات الخطأ (MSE) (0.9373) وبلغ نسبة الخطأ المطلق (MAE) (0.7812) ثم تليها طريقة الإمكان الأعظم (MLE) حيث بلغ كل من متوسط مربعات الخطأ (MSE) (0.9471) و نسبة الخطأ المطلق (MAE) (0.7813) بفرق بسيط جدا بينها وبين المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) لمربعات على التوالي (0.9577), (0.7815).

3- عند حجم (125) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ (MSE) ومقداره (0.9499) وأقل نسبة خطأ مطلق

(MAE) ومقداره (0.7766) ثم تليها بعد ذلك طريقة الإمكان الأعظم (MLE) حيث بلغ متوسط مربعات الخطأ (MSE) (0.9801) وبلغ نسبة الخطأ المطلق (MAE) (0.5055) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9804), (0.7883) بفرق بسيط جدا بينها وبين المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9811), (0.7884).

الحالة الثالثة: من القيم الافتراضية (0.6,0.5,-0.2)

تم اختيار قيم افتراضية لمعالم النموذج بناءً على دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة على ان تكون قيم هذه المعالم $(B_0 = 0.6, B_1 = 0.5, B_2 = -0.2)$ حيث تم استخدامها في كل من الطرائق التقدير كمايلي:

1- طريقة الشبكات العصبية (NN) Neural Networks :

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة بطريقة الشبكات العصبية بالاستخدام برنامج لغة (R) عند حجوم العينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على توالي (0.6,0.5,-0.2) لمعلمات أنموذج (b_0, b_1, b_2) وبتكرار التجربة (1000) مرة كما موضح بالجدول الاتي:

جدول (6-3) يوضح قيم الأوزان مع الحد الثابت للشبكات العصبية عند حجوم العينات (25,75,125) حسب القيم الافتراضية (0.6,0.5,-0.2) فضلا عن معايير المفاضلة.

الحجوم			Hidden Layer1		Hidden Layer2		Output Layer	
			W(1,1)	W(1,2)	W(2,1)	W(2,2)		
25	Inpout Layer	Bais	0.6787	-2.2535				
		$input_1$	-3.2904	2.2354				
		$input_2$	1.3888	0.6919				
	Hidden Layer1	Bais			0.0326			
		W (1,1)			-2.8669			
		W (1,2)			3.0362			
	Hidden Layer2	Bais					-0.0437	
		W(2,1)					1.2547	
			MSE					0.8021
			MAE					0.7103
75	Inpout Layer	Bais	0.8513					
		$input_1$	3.0942					
		$input_2$	-2.0985					
	Hidden Layer1	Bais		0.8513				
		W(1,1)		-1.9214				
	Hidden Layer2	Bais					0.1644	
		W (2,1)					0.9958	
			MSE					0.8853
		MAE					0.7481	
125	Inpout Layer	Bais	0.5184					
		$input_1$	-3.6951					
		$input_2$	1.7521					
	Hidden Layer1	Bais			0.3055	0.8884		
		W (1,1)			-2.5922	3.855		
	Hidden Layer2	Bais					0.2899	
		W (2,1)					-0.4908	
			W (2,2)				0.4375	
			MSE					0.9453
		MAE					0.7758	

الجدول من اعداد الباحثة

يمكن تعريف المصطلحات جدول (3-6) كما تم تعريفها في الجدول (3-2).

2-طرائق التقدير (GA,WLS,MLE):

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستخدام طرائق التقدير المتمثلة بـ(الإمكان الأعظم Maximum Likelihood Method، المربعات الصغرى الموزونة Weighted Least Squares Method، الخوارزمية الجينية Genetic Algorithm) باستخدام برنامج لغة (R) عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على التوالي (0.6,0.5,-0.2) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة وكما موضح بالجدول الآتي:

جدول (3-7) يوضح قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدرة باستخدام (الإمكان الأعظم، المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعلمات الطرائق	b_0	b_1	b_2	MSE	MAE
25	MLE	10.304	101.339	-12.406	0.893	0.7537
	WLS	10.181	101.246	-12.504	0.900	0.755
	GA	-9.973	0.3157	0.2785	0.880	0.7409
75	MLE	0.5123	-0.3412	0.2318	0.942	0.7725
	WLS	0.5075	-0.3900	0.2804	0.943	0.7726
	GA	0.4044	-0.3157	0.2784	0.942	0.7724
125	MLE	0.9402	-0.5460	0.2285	0.987	0.7945
	WLS	0.9366	-0.5466	0.2385	0.987	0.7946
	GA	0.9185	-0.5356	0.2317	0.977	0.7943

الجدول من اعداد الباحثة

نلاحظ من الجداول (6-3) و(7-3) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستخدمة (NN,MLE,WLS,GA) التي تم ذكرها في الجانب النظري.

من خلال النتائج الموضحة في الجداول (6) و(7) وحسب الحالة الأولى من القيم الافتراضية (0.6,0.5,-0.2) تبين ان:

1- عند حجم (25) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ(MSE) ومقداره(0.8021) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7103) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ متوسط مربعات الخطأ(MSE) (0.8804) ونسبة الخطأ المطلق (MAE) (0.7409) ثم تليها بعد ذلك طريقة الإمكان الأعظم حيث بلغ كل من متوسط مربعات الخطأ(MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.8938), (0.7537) ثم تأتي من بعد ذلك طريقة المربعات الصغرى الموزونة(WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ(MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9009), (0.7550).

2- عند حجم (75) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث ظهرت بأقل متوسط مربعات خطأ(MSE) ومقداره(0.8853) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7481) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ متوسط مربعات الخطأ(MSE) (0.9425) وبلغ نسبة الخطأ المطلق (MAE) (0.7724) ثم تليها بعد ذلك طريقة الإمكان الأعظم(MLE) حيث بلغ كل من متوسط مربعات الخطأ(MSE) و نسبة الأخطاء مطلق (MAE) على التوالي (0.9427), (0.7725), ومن ثم بعد ذلك تليها طريقة المربعات الصغرى الموزونة(WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ(MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9433), (0.7726).

3- عند حجم (125) كانت أفضل طريقة لتقدير الأنموذج اللوجستي هي طريقة الشبكات العصبية حيث كانت تمتلك أقل متوسط مربعات خطأ (MSE) ومقداره (0.9453) وأقل نسبة خطأ مطلق (MAE) ومقداره (0.7758) ثم تليها بعد ذلك طريقة الخوارزمية الجينية (GA) حيث بلغ متوسط مربعات الخطأ (MSE) (0.9773) وبلغ نسبة الخطأ المطلق (MAE) (0.7943) ثم تليها بعد ذلك طريقة الإمكان الأعظم حيث بلغ كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9584), (0.7887) ومن بعد ذلك تليها طريقة المربعات الصغرى الموزونة (WLS) حيث كانت نتائج كل من متوسط مربعات الخطأ (MSE) و نسبة الخطأ المطلق (MAE) على التوالي (0.9876), (0.7946) بفروق بسيط جدا بينها وبين المربعات الصغرى الموزونة.

(3-3) الجانب التطبيقي :-

تم الاعتماد على بيانات تجربة حقيقية لتطبيق طرائق التقدير والوصول الى النتائج الدقيقة المستندة الى اسس علمية .

(1-3-3) وصف التجربة :

اعتمدت الباحثة على تجربة حيث أجريت هذه التجربة في مركز مكافحة المتكاملة / وزارة التعليم العالي والرسالة العلمي / بغداد والتي تخص دراسة ثلاث صفات (استجابات متعددة) لحشرة (عثة التمور)، حيث تم الحصول على البيانات الحقيقية من أطروحة دكتوراه في علوم الحياة/ علم الحشرات "تأثيرات الحيوية والكيوحيوية لجسيمات الفضة النانوية ومستخلصات نبات الدمس النباتي في حشرة عثة التمور المخزونة" التي تم مناقشتها في عام (2020)، اذ سلطت على حشرة عثة التمور عدة معالجات بتركيز مختلفة متكونة من مستخلصات احيائية متنوعة مقسم كلا منهما الى عدة مستويات جزئية والمتمثلة بالاتي:

D₁ :يمثل المستخلص الأول كل من الكحولي الايثانولي وبتركيز(0.163)والمائي الحار وبتركيز(0.161)، والمركب الاستري Benzyl Benzoate وبتركيز(0.36)، إضافة الى مركب الألكاني Dodecane وبتركيز(0.35) .

D₂ :يمثل المستخلص الثاني من التراكيز المذابة المحضرة من جسيمات الفضة النانوية بوزن

بملغم [1500,1000,500]

وبعد مزج المستخلصات الاحيائية (D_1) و (D_2) واجراء عدة رشات بفترات زمنية مختلفة على حشرة تم الحصول على اعلى نسبة الاستجابات المتمثلة بـ (تفقيس بيض لحشرة عثة التمور، معدل أوزان ليرقات حشرة عثة التمور، البروتينات الكلية في السوائل الجسمية ليرقات حشرة عثة التمور)، حيث تناولت التجربة دراسة تخص الحشرات. امافي هذه الدراسة فقد استخدمتها الباحثة لتقدير معالم الانموذج اللوجستي المتعدد لغرض الحصول على الاستجابات المتعددة لغرض معرفة مدى تأثير هذه العوامل ومستوياتها الجزئية من خلال توظيف اسلوب تحليل الانحدار اللوجستي المتعدد الاستجابات والوصول الى القرارات المفيدة، والمخطط الاستجابات الحقيقية والقيم المتوقعة لكل الاستجابة موضح بالجدول (8)

الجدول (8-3) يمثل الاستجابة الأولى لقيم الحقيقية والمتوقعة لفقس بيض حشرة عثة التمور

D2 \ D1	500	1000	1500	Total D1j	العدد الكلي لتجربة (N)
0.163	77 (85)	75 (75)	69 (60)	221	100
0.161	60 (78)	75 (68)	66 (55)	201	100
0.36	75 (59)	54 (53)	25 (42)	154	100
0.35	66 (56)	42 (49)	37 (40)	145	100
Total D1j	278	246	197	721	

الجدول من اعداد الباحثة

القيمة (77) تمثل القيم الحقيقية والقيمة (85) تمثل القيم المتوقعة وهكذا بالنسبة لباقي القيم وباقي الجداول

الجدول (9-3) يمثل الاستجابة الثانية للقيم الحقيقية والمتوقعة للمعدل أوزان يرقات حشرة عثة التمور

D2 \ D1	500	1000	1500	Total D1j	العدد الكلي لتجربة (N)
0.163	12.1 (11.7)	12 (11.8)	7.6 (8.3)	31.7	100
0.161	9.5 (10.0)	9.9 (10.1)	7.8 (7.1)	27.2	100
0.36	6.1 (6.5)	6.1 (6.6)	5.5 (4.6)	17.7	100
0.35	10.5 (10.0)	10.5 (10.1)	6.1 (7.1)	27.1	100
Total D1j	38.2	38.5	27	103.7	

الجدول (10-3) يمثل الاستجابة الثالثة للقيم الحقيقية والمتوقعة للبروتينات الكلية في السوائل الجسمية ليرقات حشرة عثة التمور

D2 \ D1	500	1000	1500	Total D1j	العدد الكلي لتجربة (N)
0.163	11 (10.4)	11 (11.0)	11 (11.6)	33	100
0.161	41 (37.9)	40 (39.9)	39 (42.2)	120	100
0.36	35 (30.0)	31 (31.6)	29 (33.4)	95	100
0.35	40 (35.7)	38 (37.5)	35 (39.7)	113	100
Total D1j	114	120	127	361	

الجدول (11-3) يمثل الاستجابات الكلي للقيم الحقيقية والمتوقع

D ₂ \ D ₁	500	1000	1500	500	1000	1500	500	1000	1500	العدد الكلي المستخدم لكل تجربة (n)
0.163	77 (85)	12.1 (11.7)	11 (10.4)	75 (75)	12 (11.8)	11 (11.0)	69 (60)	7.6 (8.3)	11 (11.6)	100
0.161	60 (78)	9.5 (10.0)	41 (37.9)	75 (68)	9.9 (10.1)	40 (39.9)	66 (55)	7.8 (7.1)	39 (42.2)	100
0.36	75 (59)	6.1 (6.6)	35 (30.0)	54 (53)	6.1 (6.5)	31 (31.6)	25 (42)	5.5 (4.6)	29 (33.4)	100
0.35	66 (56)	10.5 (10.0)	40 (35.7)	42 (49)	10.5 (10.1)	38 (37.5)	37 (40)	6.1 (7.1)	35 (39.7)	100

لمعرفة سلوك بيانات التجربة لاستجابات الحقيقية هل تخضع لتوزيع متعدد الحدود (multinomial) ام لا حيث تم معالجة هذا الامر من خلال تطبيق اختبار (chi-square) لحسن المطابقة و عن طريقة اختيار الفرضية الإحصائية :

H_0 : البيانات تتبع توزيع متعدد الحدود.

H_1 : البيانات لا تتبع توزيع متعدد الحدود .

وبعد ان تم تطبيق اختبار (*chi-square*) لحسن المطابقة حسب الصيغة الآتية:

$$X^2 = \sum_{i=1}^4 \sum_{j=1}^9 = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

إذ إن:

$\sum_{i=1}^4$: تمثل مجموع الصفوف.

$\sum_{j=1}^9$: تمثل مجموع الاعمدة.

E_{ij} : تمثل قيم الاستجابة المتوقعة.

O_{ij} : تمثل قيم الاستجابة الحقيقية.

من خلال النتائج التي تم التوصل اليها كانت القيمة المحسوبة لبيانات الحقيقية هي (0.183) اما قيمة مربع كاي الجدولية لمستوى معنوية (0.01) وبدرجة حرية (24) كانت (36.42) لذا فان البيانات لتجربة سوف تخضع لتوزيع متعدد الحدود وعندها يمكن تطبيق أنموذج الانحدار اللوجستي متعدد على بيانات التجربة.

(2-2-3) تحليل بيانات التجربة :

التجربة المستخدمة تحتوي على نوعين من المستخلصات الاحيائية والتي تم التعبير عنهما بـ (D_1, D_2) الحقيقية

التي عاملت معاملة الرشاش أو على شكل جرعات ($dose$) وكما تم الحصول على المتغيرات المستقلة (X_1, X_2) من خلال أخذ لوغاريتم لتركيز المستخلصات المتمثلة بـ $\ln(Tolerance)$ كما موضح أدناه

الجدول (12-3) يمثل قيم احتمال الاستجابات الحقيقية لقيم اللوغاريتم النسبة المضافة $\text{logit}(Y^*)$

LN(D ₁)	LN(D ₂)	Y _{ijk}			π_i			LN($\frac{\pi}{1-\pi}$) _i			i		
		Y... ₁	Y... ₂	Y... ₃	π_i	π_i	π_i	LN($\frac{\pi}{1-\pi}$) ₁	LN($\frac{\pi}{1-\pi}$) ₂	LN($\frac{\pi}{1-\pi}$) ₃	Y* ₁	Y* ₂	Y* ₃
6.215	-1.814	9.9	75.6	11	0.03	0.13	0.12	-3.37	-2.02	-2.13	0.033	0.117	0.106
6.908	-1.826	5.5	25	40	0.14	0.11	0.12	-1.98	-2.25	-2.15	0.121	0.095	0.104
7.313	-1.022	10.5	66.6	29	0.10	0.06	0.04	-2.34	-2.88	-3.33	0.088	0.053	0.035
6.215	-1.050	12	75	40	0.14	0.11	0.10	-1.98	-2.18	-2.29	0.121	0.101	0.092
6.908	-1.814	7.8	66.6	11	0.03	0.13	0.12	-3.37	-2.03	-2.16	0.033	0.116	0.104
7.313	-1.826	6.1	75	39	0.13	0.08	0.10	-2.01	-2.51	-2.29	0.118	0.075	0.092
6.215	-1.022	10.5	42.6	35	0.12	0.06	0.12	-2.13	-2.77	-2.16	0.106	0.059	0.104
6.908	-1.050	7.6	69	38	0.13	0.11	0.06	-2.04	-2.18	-2.77	0.115	0.101	0.059
7.313	-1.814	9.5	60	11	0.03	0.08	0.11	-3.37	-2.54	-2.25	0.033	0.073	0.095
6.215	-1.826	6.1	54.6	41	0.14	0.10	0.09	-1.95	-2.29	-2.40	0.124	0.092	0.083
6.908	-1.022	6.1	37.6	31	0.10	0.06	0.08	-2.26	-2.77	-2.51	0.094	0.059	0.075

العملي

(X_1, X_2) تمثل المتغيرات المستقلة والناجمة من أخذ اللوغاريتم لكل من التركيزين $(\ln D_1, \ln D_2)$.

و تم الحصول على الاحتمال الخاص بالاستجابات الحقيقية عن طريق معادلة (2-1) التي التطرق اليها في الجانب النظري ويمكن التعبير عنها :

$$\pi_i = \frac{r_i}{n_i}$$

تمثيل عدد الاستجابات الحقيقية. r_{ijm} :

n_i : تمثيل حجم العينة.

π_i : تمثيل احتمال الاستجابة الحقيقية.

اما بالنسبة لقيمة (Y^*) تم استخراجها من أخذ لوغاريتم النسبة المضاعفة لمعادلة (2-8) التي تم ذكرها في الجانب وقد تم تحويلها الى قيمة موجبة ذلك من خلال أخذ الأساس (e) لكل قيمة من القيم.

(3-2-3) تقدير معلمات نموذج الانحدار اللوجستي متعدد الاستجابة :

يتم الحصول على تقدير معلمات انموذج الانحدار اللوجستي بالنسبة لاستجابات الحقيقية باستخدام طرائق التقدير التي تم التطرق اليها في الجانب النظري على التوالي [الشبكات العصبية(ANN), الإمكان الأعظم(MLE), المربعات الصغرى الموزونة(WLS), الخوارزمية الجينية(GA)] حيث يتم عرضها كالاتي :

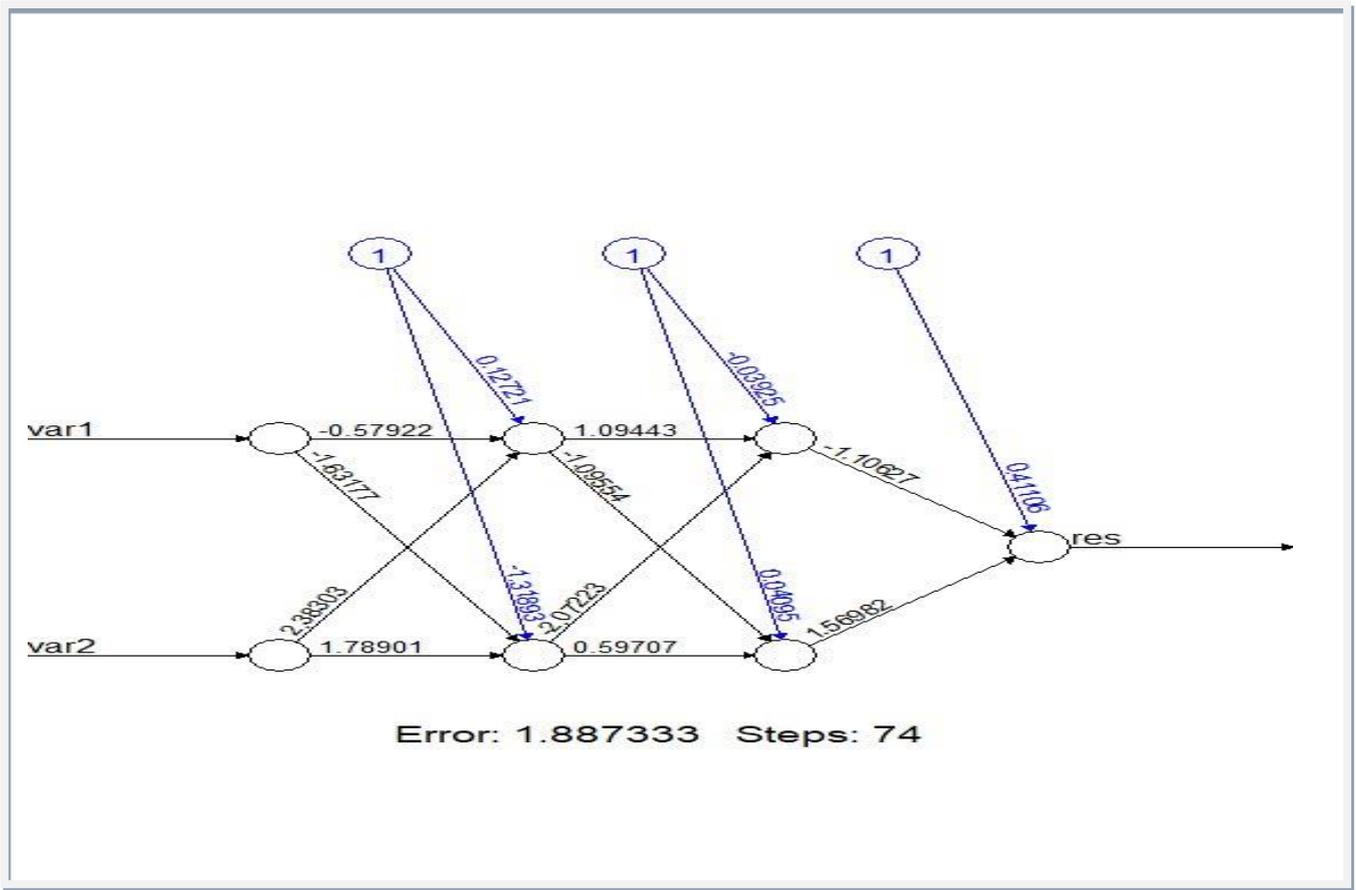
1-3-2-3 الشبكات العصبية:Neural Networks(ANN)

باستخدام برنامج (R) في الجانب التجريبي بحجوم عينات مختلفة (25,75,125) تم التوصل الى أنَّ أفضل الأوزان بطريقة الشبكات العصبية عند حجم عينة(25) وباستخدام البيانات الحقيقية تم التوصل الى الأوزان المثلى لشبكات العصبية كما موضح في الجدول الآتي

الجدول (3-13) يبين نتائج تقدير النموذج الانحدار اللوجستي وفق لاستجابات الحقيقية بالاستخدام الشبكات العصبية المتمثل بالأوزان

	Hidden Layer1		Hidden Layer2		Output Layer
	W(1,1)	W(1,2)	W(2,1)	W(2,2)	
Bais	0.127	-1.318			
Input1	-0.579	-1.631			
Input2	2.383	1.789			
Bais			-0.039	0.040	
W(1,1)			1.094	-1.094	
W(1,2)			-2.072	0.597	
Bais					0.411
W(2,1)					-1.106
W(2,2)					1.569

الجدول من اعداد الباحثة



شكل (1-3) يوضح الأوزان والحد الثابت للشبكات العصبي من اعداد الباحثة باستخدام برنامج

R

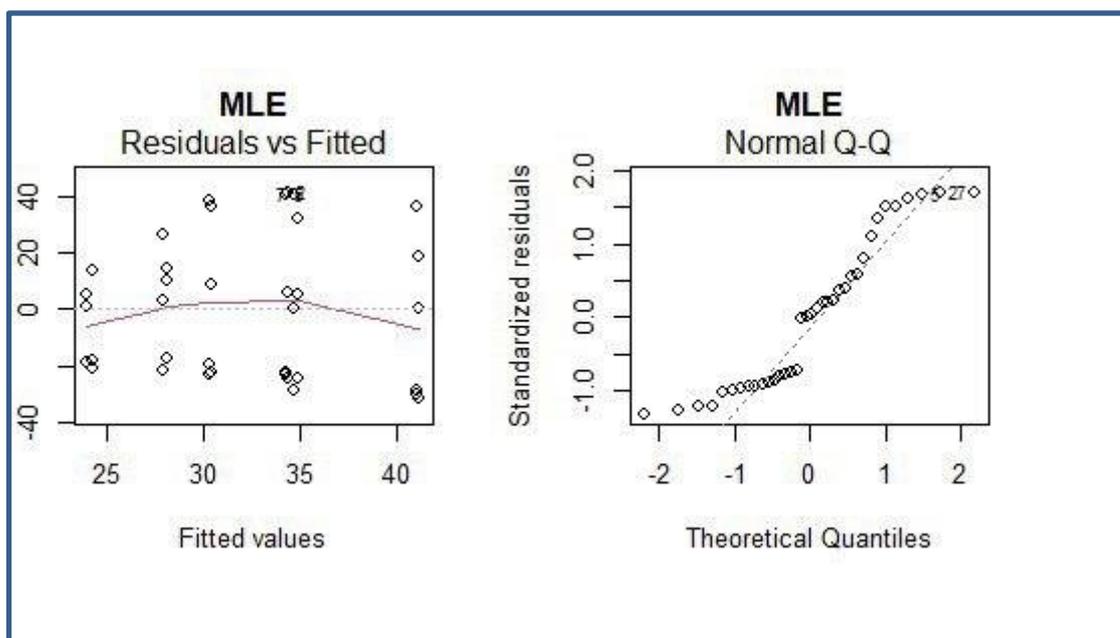
Maximum Likelihood Method

2-3-2-3 الإمكان الأعظم:

باستخدام برنامج لغة (R) في الجانب التجريبي بحجوم عينات مختلفة (25,75,125) تم التوصل الى أنّ أفضل المقدرات بطريقة الإمكان الاعظم عند حجم عينة (25) وباستخدام البيانات الحقيقية تم التوصل الى مقدرات الإمكان الأعظم كما موضح في الجدول الآتي:

الجدول (3-14) يبين نتائج تقدير نموذج الانحدار اللوجستي وفق لاستجابات الحقيقية باستخدام الإمكان الأعظم

MLE	المعالم المقدرة
86.80	b_0
-9.69	b_1
-7.99	b_2



شكل (3-2) يمثل علاقة البيانات الحقيقية والاختفاء العشوائية لإمكان الاعظم مع أنموذج اللوجستي متعدد الاستجابة من اعداد الباحثة باستخدام برنامج R

Weighted Least

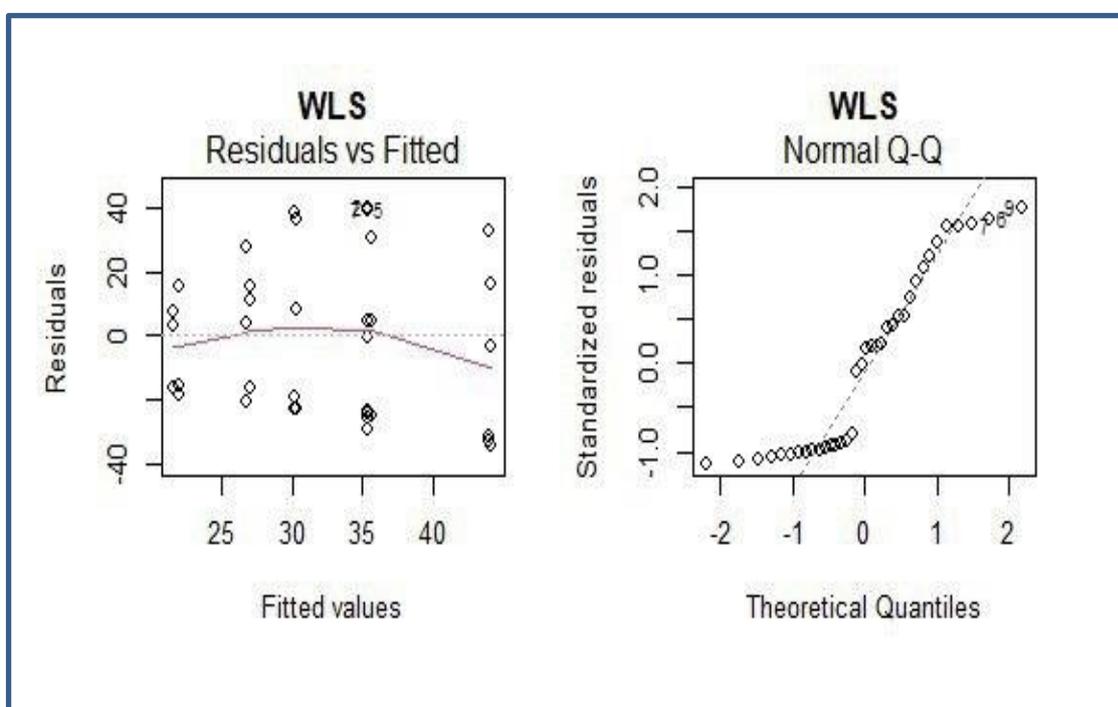
(3-3-2-3) المربعات الصغرى الموزونة :

Squares Method

باستخدام برنامج لغة (R) في الجانب التجريبي بحجوم عينات مختلفة (125,75,25) تم التوصل الى أنَّ أفضل المقدرات بطريقة المربعات الصغرى الموزونة عند حجم عينة (25) وباستخدام البيانات الحقيقية تم التوصل الى مقدرات المربعات الصغرى الموزونة كما موضح في الجدول الآتي:

الجدول (3-15) يبين نتائج تقدير نموذج الانحدار اللوجستي وفقاً لاستجابات الحقيقية بالاستخدام المربعات الصغرى الموزونة

المعالم المقدرة	WLS
b_0	101.41
b_1	-12.40
b_2	-10.76



شكل(3-3) يمثل علاقة البيانات الحقيقية والاختفاء العشوائية لمربعات الصغرى الموزونة مع نموذج اللوجستي متعدد الاستجابة من اعداد الباحثة باستخدام برنامج R

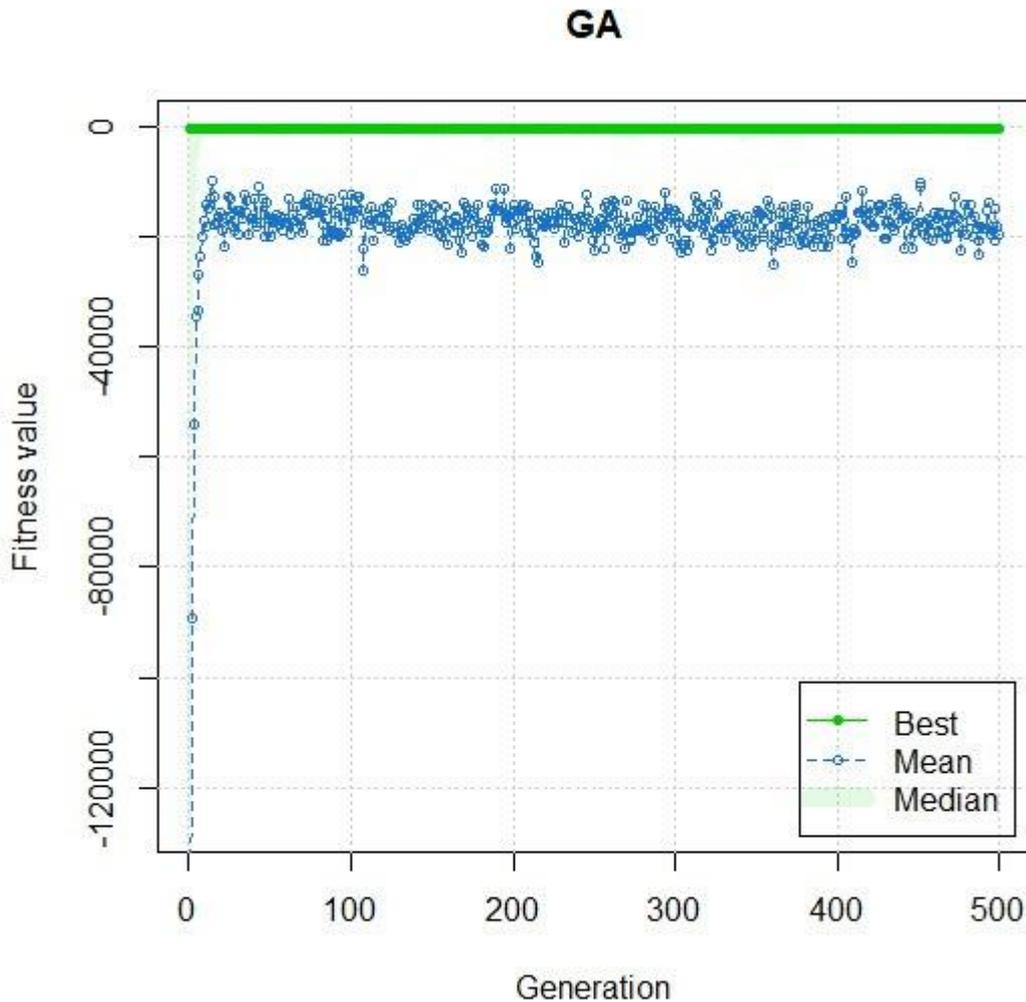
Genetic Algorithm

(4-3-3-3) الخوارزمية الجينية :

باستخدام برنامج (R) في الجانب التجريبي بحجوم عينات مختلفة (125,75,25) تم التوصل الى أنّ أفضل المقدرات بطريقة الخوارزمية الجينية عند حجم عينة(25) وباستخدام البيانات الحقيقية تم التوصل الى مقدرات الخوارزمية الجينية كما موضح في الجدول الاتي:

الجدول (3-16) يبين نتائج تقدير انحدار اللوجستي وفقاً لاستجابات الحقيقية بالاستخدام الخوارزمية الجينية

المعالم المقدره	GA
b_0	81.71
b_1	-9.03
b_2	-8.40



شكل (3-4) يمثل تصغير الخطأ عند تقدير أنموذج الانحدار اللوجستي المتعدد باستخدام الخوارزمية الجينية من اعداد الباحثة باستخدام برنامج R

(5-3-2-3) المقاييس الإحصائية لطرائق التقدير (MLE, NNA, GA, WLS)

الجدول (3-17) معايير المفاضلة لطرائق التقدير المستخدمة

المقاييس الإحصائية	ANN	MLE	WLS	GA
MSE	564.20	580.05	582.75	580.36
MAE	20.77	20.88	21.08	20.92

أظهرت النتائج التي تم توصل إليها في الجداول (3-16) للقيم الخاصة بالاستجابات الحقيقية بالنسبة للوغاريتم النسبة المضافة (logit) بان الأفضلية لطريقة الشبكات العصبية (ANN) حيث حصلت على أقل قيمة بالنسبة للمقاييس الإحصائية المتمثلة بمعيار متوسط مربعات الخطأ (MSE) ومقداره (564.20) ومعيار متوسط مطلق الخطأ (MAE) مقداره (20.77) وهذا يدل الى أفضلية للشبكات العصبية من حيث الكفاءة والأفضلية ثم تليها الخوارزمية الجينية (GA) وتكون نتائج مقارنة بين طريقة المربعات الصغرى الموزونة وطريقة الإمكان الأعظم (MLE) , وهي تنطبق على النتائج التي تم التوصل إليها في الجانب التجريبي عند القيام بتجربة المحاكاة.

4-2-3 اختبار جودة توفيق النموذج

لاختبار جودة توفيق النموذج تم

1-4-2-3 اختبار هوز مر_ ليمشو (Homsre_Lemeshow)

تم اجراء اختبار هوز مر_ ليمشو (Homsre_Lemeshow) لاختبار جودة النموذج باستخدام برنامج (R). وإنَّ الفرضيات التي تخص هذا الاختبار و تنص على أنَّ:

H_0 : الانموذج المقدر يوفق البيانات بشكل جيد

H_1 : الانموذج المقدر لا يوفق البيانات بشكل جيد

من خلال النتائج التي تم التوصل إليها حيث كانت قيمة $P-Value$ هي (0.099) أكبر من قيمة مستوى معنوية (0.05) اذن لا نرفض فرضية الصفرية H_0 أي بمعنى ان أنموذج المقدر يوفق البيانات بشكل جيد.

2-4-2-3 اختبار والـ *wald*

تم اجراء اختبار *Wald* لاختبار جودة ومعنوية النموذج باستخدام برنامج لغة (R) وفق الفرضيات التي تخص هذا الاختبار والتي تنص على ان:

H0 : تأثير لوجيت يساوي صفر.

H0 : تأثير لوجيت لا يساوي صفر.

من خلال النتائج التي تم التوصل إليها حيث كانت قيمة $P-Value$ هي (0.54) أكبر من قيمة مستوى معنوية (0.05) اذن لا نرفض فرضية الصفرية H_0 أي بمعنى إن أنموذج المقدر يوفق البيانات بشكل جيد.



الفصل الرابع

الاستنتاجات والتوصيات

الفصل الرابع

الاستنتاجات والتوصيات

Conclusions

(1-4) الاستنتاجات :-

توصلت الدراسة من خلال ما تم عرضه في الجانب النظري و الجانب التجريبي وكذلك التطبيقي الى جملة من الاستنتاجات والتوصيات الآتية :

- 1- اتضح من خلال الجانب التجريبي عند جميع حجوم العينات (كبيرة - متوسطة - صغيرة) تفوق طريقة تقدير الشبكات العصبية (Neural Networks) على باقي الطرائق المستخدمة في تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة اعتمادا على المقاييس الإحصائية متوسط مربعات الخطأ (MSE) و متوسط الخطأ المطلق (MAE).
- 2- تطابق تسلسل الأفضلية لطرائق تقدير معلمات المستعملة في تقدير الانموذج اللوجستي متعدد الاستجابة عند الجانب التجريبي (باستخدام المحاكاة) والتطبيقي (باستخدام تجربة حياتية).
- 3- اتضح من خلال نتائج الجانب التطبيقي باستعمال كل من اختبار هوزمرليمشو (-Homsre Lemeshow) واختبار والد (Wald) بأن انموذج الانحدار اللوجستي متعدد الاستجابة أنموذج معنوي احصائي.

Recommendations

(2-4) التوصيات :-

استنادا الى ما توصلت اليها الدراسة في الجانبين التجريبي والتطبيقي للدراسة وفي ضوء

الاستنتاجات التي توصلت اليها الباحثة نوصي بالآتي :

- 1- نوصي باستخدام طريقة الشبكات العصبية (Neural Networks) والخوارزمية الجينية (Genetic algorithm) لتقدير الأنموذج اللوجستي متعدد الاستجابة لكونهما يعطيان نتائج تقدير أفضل من باقي الطرائق المستخدمة.
- 2- نوصي باستخدام طرائق تقدير ومعايير مفاضلة أخرى على الأنموذج اللوجستي متعدد الاستجابة.
- 3- نوصي باستخدام نماذج جديدة غير النماذج اللوجستية مثلا نماذج التربيعية ونماذج التمييزية عند تطبيقها على البيانات المصنفة في حالة التجارب البيولوجية.
- 4- نوصي بدراسة أنموذج الانحدار اللوجستي متعدد الاستجابة بوجود مشاكل أخرى غير مشكلة تجانس التباين كمشكلة الارتباط الذاتي أو مشكلة التعدد الخطي وتطبيقها على بيانات زراعية وطبية.
- 5- نوصي بتجريب عينات مختلفة في الجانب التطبيقي لمعرفة دور انموذج الانحدار اللوجستي واختباراته المرافقة في تفسير التغيرات التي تطرأ على العينة ومدى مساهمته في حل مشاكل العينة .
- 6- نوصي بضرورة استخدام المنحني اللوجستي خصوصاً مع البيانات التي لا تتمثل بمعادلات خطية والبيانات الوصفية التي لا تعبر عنها بأرقام اعتيادية .



المصادر

المصادر

• القرآن الكريم

أولاً : المصادر العربية :-

Arabic reference

- 1- ادريس، حسن نوار ، (2009) ، " استخدم النموذج اللوجستي المتعدد لتحديد العوامل المؤثرة في مرض الجلجوم " كلية العلوم / قسم الإحصاء التطبيقي / جامعة السودان للعلوم والتكنولوجيا .
- 2- الحياي، ليث علي،2017،"اختيار أفضل طرائق تقدير المتغير المعتمد متعدد الاستجابة" رسالة ماجستير كلية الإدارة والاقتصاد، جامعة كربلاء.
- 3- الرديني، ساره عادل،2019"استعمال الخوارزمية الجينية في تقدير معلمات انموذج الانحدار اللوجستي الثنائي مع تطبيق عملي" رسالة ماجستير كلية الإدارة والاقتصاد جامعة بغداد.
- 4- الطائي ، عبد الحسين حسن ،2000"تقدير وتحليل معادلة الانحدار المتعدد في حال كون المتغيرات المعتمدة وصفية ومحددة" أطروحة دكتوراه، كلية الإدارة والاقتصاد، جامعة بغداد.
- 5- العابدي، فاضل عبد العباس "أسلوب تحليل البيانات المصنفة المترابطة"، مجلة أبحاث البصرة، المعهد التقني الكوفة العدد 28 الجزء 4.
- 6- العتيبي ، فوزي، (2003)،"استخدام السلاسل الزمنية والشبكات العصبية في التنبؤ بالأرقام القياسية"، رسالة ماجستير في الاحصاء التطبيقي، جامعة قناة السويس، مصر.
- 7- الشرقاوي ، محمد، (1996م)،"الذكاء الصناعي والشبكات العصبية"، مطابع المصري الحديث ، القاهرة ، مصر .

- 8- المراد ، عائدة يونس محمد، (2012م)، "مقارنة بين الانحدار الكلاسيكي والشبكات العصبية الاصطناعية في التنبؤ بمستويات نتائج بحوث طلبة كلية التربية الرياضية"، دراسة منشورة، المجلة العراقية الاحصائية، جامعة الموصل، العدد 21.
- 9- الناصر ، عبدالمجيد حمزة والعبيدي ، مروان عبدالحميد، (2003م)، "استخدام الشبكات العصبية للتنبؤ المستقبلي بقيم السلاسل الزمنية غير المنتظمة في طول الموسمية"، دراسة منشورة ، المجلة العراقية للعلوم الاحصائية، كلية علوم الحاسبات والرياضيات، جامعة الموصل ، المجلد 3.
- 10- بيثون، نغم نافع، (1992) ، " خواص قوه الاختبار وحدود الثقة لمعاملات الانموذج اللوجستي دراسة مقارنة " رسالة ماجستير في الإحصاء كلية الإدارة والاقتصاد جامعة بغداد.
- 11- جبارة، ازهار كاظم، 2014 "تحليل البيانات متعددة الاستجابة لتشخيص أمراض العيون باستخدام الدالة التمييزية والانحدار اللوجستي (دراسة مقارنة)" رسالة ماجستير كلية الإدارة والاقتصاد، الجامعة المستنصرية.
- 12- سليمان ، علي ابشر وحمدي ، احمد محمد، (2014م)"المقارنة بين النموذج اللوجستي الثنائي ونماذج الشبكات العصبية الاصطناعية للتمييز بين دخل الاسرة" دراسة منشورة، مجلة جامعة بخت الرضا العلمية، كلية العلوم، العدد 12.
- 13- عبد الرزاق ، محمد صادق وزعلان، ريسان عبد الامام، (2016)، "استخدم أسلوب الانحدار اللوجستي لتحليل أثر الضغط النفسي على الإصابة بضغط الدم"، البصرة، مجلة الاقتصاد الخليجي، الصفحات 48،66.
- 14- فهمي ، محمد شامل بهاء الدين " (2005) 12، الاحصاء بلا معاناة: المفاهيم مع التطبيقات على برنامج spss" الادارة العامة للباعة والنشر، معهد الادارة العامة، الرياض، المملكة العربية السعودية.
- 15- قاسم ، بهاء عبد الرزاق 2011"تحليل أثر بعض المتغيرات في ا لإصابة بمرض اللثة باستخدام انموذج الانحدار اللوجستي"مجلة العلوم الاقتصادية والإدارية، العدد 27، المجلد 17 .

ثانياً : المصادر الاجنبية :-

Foreign reference

- 16- Agrestic ، A. (2003) " Categorical data analysis " John Wiley & Sons (Vol. 482).
- 17- Akkus، Ö. ، Demir ، E. ، (2016)، " Comparison Som Classical And Meta·Heuristic Optimization Techniques in The

- Estimation Of The Logit Model Parameters "، IJAR، pp.1026،1042.
- 18- ALI، AFIFI ،R .M51 ،(2010)"The Use of Multinomial Logistic Regression Model on Physical Violence Data" degree of Master of Applied Statistics، Al، Azhar University – Gaza.
 - 19- Albert، A.، and Anderson، J. A. ،1984، "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models" ، Biometrika ، Vol. 71، No. 1 (Apr.، pp. 1،10،1984.
 - 20- Alrahamneh، A. & Hawamdeh، O. ،(2017) ، "The Factors Affecting Eye Patients (Cataract) In Jordan by Alzheimer's disease diagnosis and prognosis"، NARI، pp. 455،45.
 - 21- Andreescu، A. (2004). Forecasting corporate earnings: a data mining approach (Doctoral dissertation، M. Sc. Thesis in Accounting The Swedish school of Economics and Business administration).
 - 22- Bala، J.، Huang، J. ، Vafaie، H.، Dejonny، k. and Wechsler، H.، (1995)، " Hybrid learning using Genetic Algorithms and decision trees for pattern classification"، IJCAI Conference، Montreal، PP.19،25.
 - 23- Berkson، J. (1944). Application of the logistic function to bio-assay. Journal of the American statistical association، 39(227)، 357-365.
 - 24- Cramer، J. S. (2003). Logit models from economics and other fields. Cambridge University Press.
 - 25- Davis، R.، Lee، T.، and Rodriguez، G.، (2006)،" Structural Break Estimation for Nonstationary Time Series Models"، Journal of the American Statistical Association، Vol. 101، No. 473، PP. 229–239.
 - 26- Demir، E. ، Akkus ، Ö.، (2015)، " An Introductory Study on How the Genetic Algorithm Works in the26 Parameter Estimation of Binary Logit Model"، IJS:BAR، pp.162،180.
 - 27- El،Habil، A. M. (2012). An application on multinomial logistic regression model. Pakistan journal of statistics and operation research، 271،291.
 - 28- Garson، G. D. (2014). Logistic regression: Binary and multinomial. Asheboro، NC.
 - 29- Goldberg ، D.، (1989)،" Genetic Algorithm in Search Optimization & Machine Learning"، Addison،Wesley Publishing company، Inc.
 - 30- International Edition McGraw،Hill/Irwin، A Business Unit of The McGraw،Hill Companies. Inc.، New York.Gujarati، N. D.، & Porter، D. C. (2009). Basic Econometrics.

- 31- Guo, C. and Yang, X., (2011), " **A Programming of Genetic Algorithm in Matlab7.0**", Modern Applied Science, Vol. 5, No.1.
- 32- Haupt, R. and Haupt, S., (2004), " **Practical genetic Algorithms**", John Wiley and Sons, Inc.
- 33- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.
- 34- Hussain, J. N. ; Low, H. C. & AL, Karkhi, A. F. M , (2007) " **An Overview Of Evaluation Criteria In Logistic Regression Model**", International Conference on Mathematical Sciences, University Teknologi, Malaysia .
- 35- Joel L.Horowitz and N.E.Savin (2001) " **Binary Response models : Logits , Probits and Semiparametrics** " Journal of Economic perspectives .Vol 15 , No 4.
- 36- Johnson, P., Graham, P. & et al., (2013), " **Genetic algorithm with logistic regression for Alzheimer's disease diagnosis and prognosis**", NARI, pp. 455,456.
- 37- 1٤٧.
- 38- Kleinbaum, D. G., Klein, M., & Pryor, E. R. (2002). Logistic regression: a self-learning text.
- 39- Kneale, p.; See, L. and Smith, A. (2000). " **Towards Defining Evaluation Measures for Neural Network Forecasting Models**", School of Geography, University of Leeds, UK.
- 40- Krose, B. and Smagt, V.D.P (1996) . " **An Introduction to neural networks**", Eighth edition. The university of Amsterdam, pp 33,34.
- 41- Kumar, D., Negesh, K. & Sathish, T. (2004), " **River Flow Forecasting Using Recurrent Neural Networks** ", Water Resources Management, vol 18.
- 42- M.S Bartlett .F.R.S & D.R.COX, F.R.S. 1977. " **The Analysis of Contingency Tables** " , Jone Wily & Sons, Inc, New York" .
- 43- McCarthy, W.F, 2007 " **The Existence of Maximum Likelihood Estimates for the Binary Response Logistic Regression Model**", Collection of Biostatistics Research Archive, COBRA Preprint Series, Paper.
- 44- McCullagh, P., & Nelder, J., (1983), " **Generalized Linear Models**", London: Chapman and Hall.
- 45- Meng, Q., Weng, J., (2011) , " **A genetic Algorithm approach To assessing Work Zone casualty Risk** " , ELSEVIER , pp. 1283,1288
- 46- Mitchell, T. M. (1997) , " **Machin learning Boston**", WCB/Mcgraw, hill, new york .

- 47- Mitra, S., Pal, S. and Mitra, p. (2002), "**Data Mining in Soft Computing Framework: A Survey**", IEEE Transactions on neural networks, Vol. 13, No. 1, PP.3,14.
- 48- Muller, Marlene (2004) , "**Generalized Linear Models**", Fraunhofer Institute for Industrial Mathematics (ITWM) , (Germany) , WWW. Marlenmuller . ed / publication / hand book CS. Pdf.
- 49- Neamah, M. W. & Mohammed, L. A. (2020). Used Contingency Tables in Categorical Data Analysis for The Multi-Response. PalArch's Journal of Archaeology of Egypt/Egyptology, 17(6), 15297,15308.
- 50- Pao, Y. (1989), "**Adaptive Pattern Recognition and Neural Networks**", Publishing Company, Inc., New York.
- 51- Park, H. A. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. Journal of Korean Academy of Nursing, 43(2), 154,164.
- 52- Pasia, J., Hermosilla, A. & et al. (2005), "**A useful tool for statistical estimation genetic algoritlm**", JSCS, pp. 237 – 251.
- 53- Prebys, E. (2007), "**The Genetic Algorithm in Computer Science**", MIT Undergraduate Journal of Mathematics, PP.165,170.
- 54- Rajasck, S. & Vijaylak sh mi, G. A. (2003) neural networks, Fuzzy logic, and Genetic Algorithms syanthesis and Application prentice, Hall of India, newdelhi.
- 55- Rakotomalala, R. (2014). Pratique de la Régression Logistique. Régression Logistique Binaire et Polytomique. Version 2.0. Lyon, Université Lumière Lyon,2, multigr.
- 56- Rassam F.(1982)"**operations Research**"Military technical collage.
- 57- Ruan, D. (1997), "**Intelligent Hybrid Systems : Fuzzy Logic, Neural Networks and Genetic Algorithms**", Springer Science Business Media New York.
- 58- Shreider A.(1966) "**The Monte Carlo Method, the method of statistics**"Volume 2, fourth edition, London.
- 59- Sivanandam, S. and Deepa, S. (2008), "**Introduction to Genetic Algorithms**", Springer Berlin Heidelberg New York.
- 60- Stergiou, Ch. & Sigonos, D. (2001), "**Neural Networks**".
- 61- Titma (Tuma ,Rooma) (2002)"**Education as a factor in integeneatjonal mobility in Sovier Society**" ,Integrating Theory and Research.
- 62- Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. International Journal of Forecasting, 17(1), 57,69.

- 63- Yoder, S. E., (2009), " An Investigation on the use and flexibility of Genetic Algorithm for Logistic Regression" , Clemson University, TP, pp. 1,77.

Internet sites

ثالثاً : مواقع الأنترنيت :

- 64- <http://www.emsl.pnl.gov:2080/docs/cie/techbrief/NN.techbrief.html>



الملاحق

الملاحق

شفرات برنامج R (للمحاكاة والتطبيقي العملي)

1-برنامج المحاكاة Simulation Program

```
#install.packages('GA')
#install.packages("neuralnet")
#install.packages("gmm")
library(GA)
library(neuralnet)
library(gmm)
=====#requirment functions =====#
####MODEL ####
L.model <- function(data, b0, b1, b2){
  attach(data, warn.conflicts=F)
  Y_hat <- b0 + b1*var1 + b2*var2
  SSE = l2(res-Y_hat)
  detach(data)
  return(SSE)
####minimizition function ####
l2 = function(e)
#L2 norm
#e is n x 1 of residuals
  sum(e^2)/length(e)
#=====#
=====#Simulation part =====#
n<- 75
```

```
p<- 2
it.no<- 1000
size<-c(rep(3,p))
pr<-c(rep(0.5,p))
x<-NULL
for(j in 1:p){
x<-cbind(x,rbinom(n,size[j],pr[j]))
B<-c(-0.4, 0.3)
b0<- 0.5
mse.GA.i<-NULL
mse.NN.i<-NULL
mse.MLE.i<-NULL
mse.WLS.i<-NULL
mse.GMM.i<-NULL
mae.GA.i<-NULL
mae.NN.i<-NULL
mae.MLE.i<-NULL
mae.WLS.i<-NULL
mae.GMM.i<-NULL
dim(x)
dim(B)
for(it in 1:it.no){
if(it==1) monitor = F else monitor = F
u<-rnorm(n)
y<- b0 + x%*%B + u
data<- data.frame(y,x=x)
xnam <- paste0("var", 1:ncol(x))
Bnam <- paste0("b", 1:ncol(x))
data<-data.frame(y,x)
```

```

names(data)<-c("res",xnam)

#head(data)

#fmla <- as.formula(paste("b0+", paste(xnam, collapse= "+")))
#fmlaB <- paste("b0,", paste(Bnam, collapse= ","))
#fmlaB <- paste("b0+", paste(Bnam, xnam, collapse= "+"))

#=====
#####Genetic Algorithn #####

GA<- ga(type='real-valued', min=rep(-100,p+1),
max=rep(100,p+1), popSize=500, maxiter=500, keepBest=T, pmutation = 0.2,
optimArgs = list(method = "Nelder-Mead"), monitor = monitor,
fitness = function(b) -L.model(data, b[1],b[2], b[3]))

GA.B<- summary(GA)$solution
y.GA<- exp(GA.B[1]+ x%*%GA.B[-1])/(1+exp(GA.B[1]+ x%*%GA.B[-1]))
mse.GA<-sum((y-y.GA)^2)/n
mae.GA<-sum(abs(y-y.GA))/n

#plot(GA)

#####MLE #####

MLE<-lm(formula= res ~., data=data)

MLE.B<- MLE$coefficients
#y.MLE<- MLE.B[1]+ x%*%MLE.B[2:(p+1)]
y.MLE<- exp(MLE.B[1]+ x%*%MLE.B[2:(p+1)])/(1+exp(MLE.B[1]+
x%*%MLE.B[2:(p+1)]))

mse.MLE<-sum((y-y.MLE)^2)/n
mae.MLE<-sum(abs(y-y.MLE))/n

#plot(MLE)

#####WLS #####

wts <- 1/fitted(lm(abs(residuals(MLE)) ~ fitted(MLE)))^2
WLS<-lm(res~., data=data, weights=wts)

WLS.B<- WLS$coefficients
#y.WLS<- WLS.B[1]+ x%*%WLS.B[2:(p+1)]

```

```

y.WLS<- exp(WLS.B[1]+ x%*%WLS.B[2:(p+1)]/(1+exp(WLS.B[1]+
x%*%WLS.B[2:(p+1)]))

mse.WLS<-sum((y-y.WLS)^2)/n

mae.WLS<-sum(abs(y-y.WLS))/n

#plot(WLS)

=====#NN =====#

#index <- sample(1:nrow(data),round(0.75*nrow(data)))

index <- sample(1:nrow(data),round(nrow(data)))

#train <- data[index,]

#test <- data[index,]

train <- data

test <- data

maxs <- apply(data, 2, max)

mins <- apply(data, 2, min)

scaled <- as.data.frame(scale(data, center = mins, scale = maxs - mins))

#train_ <- scaled[index,]

#test_ <- scaled[index,]

train_ <- scaled

test_ <- scaled

node<- ncol(train)-1

ns <- names(train_)

f<- as.formula(paste("res ~", paste(ns[!ns %in% "res"], collapse = " + ")))

error1<-matrix(0,node,node)

error2<-matrix(0,node,node)

for(i in 1:node){

for(j in 1:node){

#set.seed(i+j)

nn <- neuralnet(f,data=train_,hidden=c(i,j),linear.output=T)

pr.nn <- compute(nn,test_[,1:ncol(test_)])

pr.nn_ <- pr.nn$net.result*(max(data$res)-min(data$res))+min(data$res)

```

```

test.r <- (test_$res)*(max(data$res)-min(data$res))+min(data$res)
MSE.nn <- sum((test.r - pr.nn_)^2)/nrow(test_)
MAE.nn <- sum(abs(test.r - pr.nn_))/nrow(test_)
error1[i,j]<-MSE.nn
error2[i,j]<-MAE.nn
mse.NN<-min(error1)
bestnode = which(error1 == min(error1), arr.ind=TRUE)
mae.NN<-error2[bestnode[1],bestnode[2]]
nn <- neuralnet(f,data=train_,hidden=c(bestnode[1], bestnode[2]),linear.output=T)
#plot(nn)
#nn$weights
plot(GMM)
=====#Results =====#
mse.GA.i<-cbind(mse.GA.i, mse.GA)
mse.NN.i<-cbind(mse.NN.i, mse.NN)
mse.MLE.i<-cbind(mse.MLE.i, mse.MLE)
mse.WLS.i<-cbind(mse.WLS.i, mse.WLS)
mae.GA.i<-cbind(mae.GA.i, mae.GA)
mae.NN.i<-cbind(mae.NN.i, mae.NN)
mae.MLE.i<-cbind(mae.MLE.i, mae.MLE)
mae.WLS.i<-cbind(mae.WLS.i, mae.WLS)
YH#=====#
pb <- txtProgressBar(min = 0, max = it.no, style = 3)
  setTxtProgressBar(pb, it)
# {end of iterations
imse<-cbind(
t(mse.GA.i),
t(mse.NN.i),
t(mse.MLE.i),

```

```
t(mse.WLS.i),
t(mse.GMM.i)
colnames(imse)<-c("GA", "NN", "MLE", "WLS", "GMM")
imae<-cbind(
t(mae.GA.i),
t(mae.NN.i),
t(mae.MLE.i),
t(mae.WLS.i),
t(mae.GMM.i)
colnames(imae)<-c("GA", "NN", "MLE", "WLS", "GMM")
mse<-cbind(
apply(mse.GA.i,1,mean),
apply(mse.NN.i,1,mean),
apply(mse.MLE.i,1,mean),
apply(mse.WLS.i,1,mean),
apply(mse.GMM.i,1,mean)
colnames(mse)<-c("GA", "NN", "MLE", "WLS", "GMM")
rownames(mse)<-c("MSE")
mae<-cbind(
apply(mae.GA.i,1,mean),
apply(mae.NN.i,1,mean),
apply(mae.MLE.i,1,mean),
apply(mae.WLS.i,1,mean),
apply(mae.GMM.i,1,mean)
colnames(mae)<-c("GA", "NN", "MLE", "WLS", "GMM")
rownames(mae)<-c("MAE")
dev.new()
plot(GA, main="GA")
savePlot(filename = "c:\\users\\dell\\desktop\\1\\GA",
```

```

type = c("jpeg"),device = dev.cur(),restoreConsole = TRUE)
dev.new()
plot(nn, main="NN")
savePlot(filename = "c:\\users\\dell\\desktop\\1\\nn",
type = c("jpeg"),device = dev.cur(),restoreConsole = TRUE)
results<-rbind(mse, mae)
dev.new()
boxplot(imse, main="MSE boxplot")
savePlot(filename = "c:\\users\\dell\\desktop\\1\\imse",

```

2- برنامج تطبيقي العملي (البيانات الحقيقية) :

```

#install.packages('GA')
#install.packages("neuralnet")
library(GA)
library(neuralnet)
library(gmm)
x1<-
c(6.214608,6.907755,7.31322,6.214608,6.907755,7.31322,6.214608,6.907755,
7.31322,6.214608,6.907755,7.31322,6.214608,6.907755,7.31322,6.214608,6.90
7755,7.31322,6.214608,6.907755,7.31322,6.214608,6.907755,7.31322,6.21460
8,6.907755,7.31322,6.214608,6.907755,7.31322,6.214608,6.907755,7.31322,6.
(214608,6.907755,7.31322
x2<-c(-1.81401,-1.82635,-1.02165,-1.04982,-1.81401,-1.82635,-1.02165,-
1.04982,-,1.02165,-1.82635,-1.81401,-1.04982,-1.02165,-1.82635,-1.81401-
,1.04982
-1.04982-,1.02165-,1.82635-,1.81401-,1.04982-,1.02165-,1.82635-,1.81401-
,1.04982-,1.02165-,1.82635-,1.81401-,1.04982-,1.02165-,1.82635-,1.81401
(1.04982-,1.02165-,1.82635-,1.81401-
y<-c(77,75.6,25,66.6,75,66.6,75,42.6,69,60,54.6,37.6,12.1,9.9,5.5,10.5,12,

```

```

(7.8,6.1,10.5,7.6,9.5,6.1,6.1,11,40,29,40,11,39,35,38,11,41,31,3.5
=====#requirment functions =====#
####MODEL ####
L.model <- function(data, b0, b1, b2){
  attach(data, warn.conflicts=F)
  Y_hat <- b0 + b1*var1 + b2*var2
  SSE = l2(res-Y_hat)
  detach(data)
  return(SSE)
####minimizition function ####
l2 = function(e)
  #L2 norm
  #e is n x 1 of residuals
  sum(e^2)/length(e)
=====#practical part =====#
n<- length(y)
p<- 2
B<-c(-0.4, 0.3)
b0<- 0.5
x<-cbind(x1,x2)
data<- data.frame(y,x=x)
xnam <- paste0("var", 1:ncol(x))
Bnam <- paste0("b", 1:ncol(x))
data<-data.frame(y,x)
names(data)<-c("res",xnam)
=====#Genetic Algorith =====#
GA<- ga(type='real-valued', min=rep(-100,p+1),

```

```
max=rep(100,p+1), popSize=500, maxiter=500, keepBest=T, pmutation =
0.2,
```

```
optimArgs = list(method = "Nelder-Mead"),
```

```
fitness = function(b) -L.model(data, b[1],b[2], b[3]))
```

```
GA.B<- summary(GA)$solution
```

```
y.GA<- exp(GMM.B[1]+ x%*%GMM.B[2:(p+1)]/(1+exp(GMM.B[1]+
x%*%GMM.B[2:(p+1)]))
```

```
mse.GA<-sum((y-y.GA)^2)/n
```

```
mae.GA<-sum(abs(y-y.GA))/n
```

```
#plot(GA)
```

```
=====MLE =====#
```

```
MLE<-lm(formula= res ~., data=data)
```

```
MLE.B<- MLE$coefficients
```

```
y.MLE<- exp(GMM.B[1]+ x%*%GMM.B[2:(p+1)]/(1+exp(GMM.B[1]+
x%*%GMM.B[2:(p+1)]))
```

```
mse.MLE<-sum((y-y.MLE)^2)/n
```

```
mae.MLE<-sum(abs(y-y.MLE))/n
```

```
#plot(MLE)
```

```
=====WLS =====#
```

```
wts <- 1/fitted(lm(abs(residuals(MLE)) ~ fitted(MLE)))^2
```

```
WLS<-lm(res~., data=data, weights=wts)
```

```
WLS.B<- WLS$coefficients
```

```
y.WLS<- WLS.B[1]+ x%*%WLS.B[2:(p+1)]
```

```
mse.WLS<-sum((y-y.WLS)^2)/n
```

```
mae.WLS<-sum(abs(y-y.WLS))/n
```

```
#plot(WLS)
```

```
=====NN =====#
```

```
#index <- sample(1:nrow(data),round(0.75*nrow(data)))
```

```

index <- sample(1:nrow(data),round(nrow(data)))
#train <- data[index,]
#test <- data[index,]
train <- data
test <- data
maxs <- apply(data, 2, max)
mins <- apply(data, 2, min)
scaled <- as.data.frame(scale(data, center = mins, scale = maxs - mins))
#train_ <- scaled[index,]
#test_ <- scaled[index,]
train_ <- scaled
test_ <- scaled
node<- ncol(train)-1
ns <- names(train_)
f <- as.formula(paste("res ~", paste(ns[!ns %in% "res"], collapse = " + "))
error1<-matrix(0,node,node)
error2<-matrix(0,node,node)
for(i in 1:node){
for(j in 1:node){
#set.seed(i+j)
nn <- neuralnet(f,data=train_,hidden=c(i,j),linear.output=T)
pr.nn <- compute(nn,test_[,1:ncol(test_)])
pr.nn_ <- pr.nn$net.result*(max(data$res)-min(data$res))+min(data$res)
test.r <- (test_$res)*(max(data$res)-min(data$res))+min(data$res)
MSE.nn <- sum((test.r - pr.nn_)^2)/nrow(test_)
MAE.nn <- sum(abs(test.r - pr.nn_))/nrow(test_)
error1[i,j]<-MSE.nn

```

```

error2[i,j]<-MAE.nn
mse.NN<-min(error1)
bestnode = which(error1 == min(error1), arr.ind=TRUE)
mae.NN<-error2[bestnode[1],bestnode[2]]
nn <- neuralnet(f,data=train_,hidden=c(bestnode[1],
bestnode[2]),linear.output=T)
#plot(nn)
nn$weights
=====#GMM=====#
fmla <- as.fo
mse.GMM<-sum((y-y.GMM)^2)/n
mae.GMM<-sum(abs(y-y.GMM))/n
#plot(GMM)
=====#Results=====#
mse<-cbind(mse.GA,mse.NN,mse.MLE,mse.WLS,mse.GMM)
colnames(mse)<-c("GA", "NN", "MLE", "WLS", "GMM")
rownames(mse)<-c("MSE")
mae<-cbind(mae.GA,mae.NN,mae.MLE,mae.WLS,mae.GMM)
colnames(mae)<-c("GA", "NN", "MLE", "WLS", "GMM")
rownames(mae)<-c("MAE")
dev.new()
plot(GA, main="GA")
dev.new()
par(mfrow=c(2,2))
plot(MLE, main="MLE")
dev.new()
par(mfrow=c(2,2))

```

```
plot(WLS, main="WLS")
dev.new()
par(mfrow=c(2,2))
plot(GMM, main="GMM")
dev.new()
plot(nn, main="NN")
results<-rbind(mse, mae)
dev.new()
rmula(paste("res~", paste(xnam, collapse= "+")))
hh<- x^2
GMM<-gmm(fmla, data=data, x=hh)
GMM.B<- GMM$coefficients
y.GMM<- GMM.B[1]+ x%*%GMM.B[2:(p+1)]
write.csv(results,"C:/Users/DELL/Desktop/data/results.csv")
summary(GA)
summary(MLE)
summary(WLS)
summary(GMM)
summary(nn)
```

ABSTRACT :-

Statistics form an extended hand that provides its services over time to other sciences and a tributary that contributes to analyzing the results of studies in various fields. The experiment and the identification of the factors affecting the course of the experiment, its development and access to the required results. In this thesis, nonlinear regression was used and a focus on the regression of the logistic curve.

The problem of the thesis lies in how to deal with the multi-level logistic regression model in the data modeling process and to reach the optimal estimation methods in estimating parameters

The thesis aims to estimate the logistic regression model for a dependent variable (multi-levels) (Y_{ijk}), depending on descriptive variables to know the extent of their impact on the dependent variable (under study), where the classical estimation methods were used, including (the method of greatest possibility [MLE] and the method of least squares). weighted [WLS]) and intelligent (the method of artificial neural networks [ANN] and genetic algorithm [GA]), and then compare between the methods to find out the best and most efficient of them in order to obtain the efficient capabilities to rely on in standing on the dimensions and variables of the phenomenon (under study Through the use of comparison scales, including mean squared error (MSE) and mean absolute error (MAE).

In practice, and to achieve the objectives of the thesis, a simulation experiment was conducted using the Monte Carlo method for different sample sizes (small, medium, large) with sizes (25, 75, 125) that fit the experience of practical application to reach the optimal method and to know the changes that occur in The process of experimentation, and

then the results of the simulation experiment were summarized, and from the course of the simulation experiment, it was concluded that the method of artificial neural networks is the best among the methods used in the estimation process. In the estimation process on the applied side as well, and not only in the experimental side, in order to match the estimation methods, where the results of the applied side showed the superiority of the method of artificial neural networks also among the Its peers, and this is consistent with the experimental side, which indicates the suitability of the estimation method with the logistic regression model.

Finally, one of the most important conclusions that the connection is superior to the method of artificial neural networks (ANN) in the experimental and applied aspect to the rest of the methods used, the thesis also showed the possibility of converting the logistic model from the non-linear model to the linear model.

Republic of Iraq
Ministry of Higher Education and
Scientific Research.
University of Kerbala
College of Administration and
Economics.
Statistical Department



(Estimating a logistic Model for a Multi-level Dependent Variable with Practical Application)

A Thesis

submitted to the council of the college of administration and
economics - university of kerbala as a partial fulfillment of
the requirements for the degree of master of the science in
statistics

By

Tuqa. Abid.Mohammed.AL-Khazail

Supervised by

A.Dr.Abdul hussian.H.Habeab.AL.Tai

2021 AD

Kerbala

1443 AH