



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة كربلاء
كلية الادارة و الاقتصاد
قسم الاحصاء

اختيار افضل طرائق تقدير المتغير المعتمد متعدد الاستجابة

رسالة مقدمة الى مجلس كلية الادارة والاقتصاد في جامعة كربلاء

و هي جزء من متطلبات نيل درجة ماجستير في علوم الاحصاء

تقدم بها

ليث علي محمد الحياي

بإشرافه

أ . د . محمد الحسين حسن حبيب الطائي

2017 هـ

كربلاء المقدسة

1438 هـ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

((قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا
مَا عَلَّمْتَنَا إِنَّكَ أَنْتَ الْعَلِيمُ الْحَكِيمُ))

صَدَقَ اللَّهُ الْعَلِيُّ الْعَظِيمُ

سورة البقرة الآية (32)

الاهداء

إلى من سهر الليالي وقدم الغالي والنفيس من أجل تقدمي ونجاحي

إلى بسملة الحياة وسر الوجود

إلى أبي العزيز

إلى الصبر والتضحية , إلى القلب الدافئ الرقيق , وعنفوان الحب والحياة

إلى من بوجودها أكتسب قوة ومحبة لا حدود لها

إلى . . . أمي الغالية

إلى أخوتي الذين زرعوا في قلبي حب العلم وساندوني حتى أوصل

مسيرتي العلمية , الى سندي في الحياة



المستخلص

اختيار افضل طرائق تقدير المتغير المعتمد متعدد الاستجابة

تعد المتغيرات المعتمدة الوصفية من المتغيرات المهمة التي ليس لها وحدات قياس كمية وتخضع لوصف الظاهرة باستخدام دراسة البيانات والمعلومات والتي تمكن متخذ القرار ان يتعرف على طبيعة الاستجابة في حالة كون المتغير المعتمد ثنائي الاستجابة او متعدد الاستجابة , لذا قد تكون دالة الاستجابة غير خطية, ولذلك ظهرت الحاجة الى البحث عن اساليب اخرى لتحويل الى دوال استجابة خطية, ويتم ذلك عن طريق اسلوب التحويل اللوغاريتم النسبة المضافة (logit) المعتمد على التوزيع اللوجستي, وأيضاً استخدام تحويلة وحدة الاحتمال (probit) المعتمدة على التوزيع الطبيعي, ومن خلال استخدام هذه الاساليب يمكن تحويلها الى استجابات خطية. تضمنت الدراسة استعمال انموذج الانحدار الوصفي في حالة كون المتغير المعتمد متعدد الاستجابة وتم ذلك من خلال دراسة بعض النماذج المتعلقة بهذه المتغيرات الوصفية .

وهدف الرسالة هو دراسة وتحليل المتغير الوصفي وتأثيره في المتغير المعتمد متعدد الاستجابة, وذلك عن طريق تقدير معالم النموذج للمتغير المعتمد متعدد الاستجابة باستعمال طريقة المربعات الصغرى الموزونة (WLS) وطريقة الامكان الاعظم (MLE) ويتم ذلك من خلال الاعتماد على تطبيق طريقة نيوتن رافسن (Newton Raphson) وايضاً تم استعمال طريقة الجاكنائف (jackknife), وكما تم استخدام جانبان احدهما هو الجانب التجريبي والاخر هو الجانب التطبيقي, ففي الجانب التجريبي تم استعمال طريقة مونت كارلو (Monte Carlo) في تجربة محاكاة ولثلاث مستويات من العينات (صغيرة, متوسطة, كبيرة) وبأحجام مختلفة وذلك عن طريق توليد ارقام عشوائية لمعالم انموذج الانحدار, فضلاً عن الجانب التجريبي تم تطبيق الجانب العملي على تجربة حياتية لبعض المواد المبيدة للحشرات وبتراكيز مختلفة , ومن ثم اختيار افضل هذه الطرائق عن طريق الاعتماد على مقياس متوسط مربعات الخطأ (MSE) . وتوصلت الدراسة باستعمال طريقة المونت كارلو (Monte Carlo) في المحاكاة الى ان مقدرات معالم الانموذج باستعمال طريقة الامكان الاعظم (MLE) هي الافضل والاكفاً في التقدير في حالة حجوم العينات الصغيرة (n=20) عند الاستجابات الاولى والثانية , بينما كانت طريقة الجاكنائف (JACK) هي الافضل في تقدير معالم انموذج الانحدار عند حجوم العينات المتوسطة (n=50) لجميع الاستجابات إذ حصلت على اقل (MSE), اما في حالة العينات الكبيرة (n=100, 200) فقد كانت طريقة الامكان الاعظم (MLE) وطريقة الجاكنائف (JAK) هي الافضل إذ بينت نتائج الدراسة الى وجود تقارب واضح بالقيم عند حجوم العينات الكبيرة للاستجابات الثلاث. بينما في الجانب التطبيقي فقد تم تطبيق الطرائق المستعملة في الجانب النظري على بيانات حقيقية تم تطبيقها على تجربة حياتية خاصة لبعض المواد المبيدة للحشرات وبتراكيز مختلفة وتم الحصول على استجابات متوقعة , ومن خلال المقارنة تبينت النتائج الى افضلية وكفاءة طريقة الامكان الاعظم (MLE) إذ حصلت على اقل (MSE) وفقاً للنظرية الاحصائية الفائلة عند زيادة حجم العينة يقل متوسط مربعات الخطأ (MSE) .

الصفحة	قائمة المحتويات	
1 – 6	الفصل الاول	
1	المقدمة	1-1
2	مشكلة الدراسة	2 -1
2	اهمية الدراسة	3 -1
2	منهجية الدراسة	4 -1
2	هدف الدراسة	5 -1
2	الافتراض المرجعي	6 -1
7 – 33	الفصل الثاني: نماذج المتغيرات المعتمدة متعددة الافتراضات تجابة	
7	البيانات المصنفة	1-2
7	جداول التوافق	2-2
10	توزيع متعدد الحدود	3-2
11	النماذج الخطية الاعتيادية والعامه	4-2
21	نماذج الانحدار الوصفية	5-2
22	انموذج الانحدار اللوجستي	5 -2- 1
23	انموذج الانحدار اللوجستي الثنائي	5 -2-1- 1
25	انموذج الانحدار اللوجستي متعدد الافتراضات تجابة	5 -2- 1- 2
26	بعض المشاكل الخاصة بالمتغير المعتمد الوصفي	5- 2 -2
29	نماذج دوال التحويل	6-2
30	أنموذج لوغاريتم النفاذ المضافة	6-2 -1
31	أنموذج وحدة الاحتمال	6-2-2
34 -46	طرائق تقدير افتراضات عالم المتغير المعتمد تعدد الاستجابة	7-2
34	طريقة المربعات الصغرى الموزونة	7-2-1

38	طريقة الامكان الاعظم	7-2-2
41	طريقة نيوتن رافن	7-2 -2-1
45	طريقة الجا كناف	7-2-3
46 – 62	الفصل الثالث المبحث الاول: الجانب التجريبي	
46	المقدمة	1 -3
46	مفهوم المحاكاة	1 -3 -1
47	توليد الاعداد العشوائية	1 -3-2
47	توليد المتغيرات المقل	1- 3 -2 -1
48	توزيع المتغير المعتمد (الا تجابة)	1 -3 -2 -2
48	مراحل بناء ووصف تجارب المحاكاة	1 -3 – 3
50	تحليل نتائج تجارب المحاكاة	1 -3- 4
51	احتساب قيم المقاييس الاحصائية المخدمة للا تجابة الاولى والثانية والثالثة	1-3-5
62	المبحث الثاني: الجانب التطبيقي	2 -3
62	بيانات التجربة	2-3 -1
63	تحليل بيانات التجربة	2 -3-2
65	نتائج التجربة	2 -3 -3
67	نتائج التجربة للا تجابات المتوقعة لقيم وحدة الاحتمال (Probit)	2 -3 -3-1
70	نتائج التجربة لقيم الا تجابة المتوقعة الخاصة للوغاريتم النذبة المضافة (logit)	2-3 -3 -2
71 -73	الفصل الرابع : الا نتاجات والتوصيات	
74	المصادر	
80-83	الملحق	

فهرس الجداول والاشكال

رقم الجدول	عنوان الجداول	الصفحة
1-2	تكرار العينة المصنفة	8
2-2	يمثل القيم الاحتمالية	9
3-2	بعض الصيغ التي يمكن تحويلها الى الصيغة الخطية	13
1-3	قيم المعالم الافتراضية	49
2-3	القيم التقديرية (\hat{B}) للاستجابة الاولى لجميع الطرائق وحجوم العينات المستخدمة	53
3-3	قيم متوسط مربعات الخطأ (MSE) للاستجابة الاولى للمعالم المقدره والانموذج لجميع الطرائق وحجوم العينات المستخدمة	54
4-3	قيم التحيز (Biase) المطلقة للاستجابة الاولى للمعالم المقدره وحجوم العينات المستخدمة	55
5-3	القيم التقديرية (\hat{B}) للاستجابة الثانية لجميع الطرائق وحجوم العينات المستخدمة	56
6-3	قيم متوسط مربعات الخطأ (MSE) للاستجابة الثانية للمعالم المقدره والانموذج و لجميع الطرائق وحجوم العينات المستخدمة	57
7-3	قيم التحيز (Biase) المطلقة للاستجابة الثانية للمعالم المقدره وحجوم العينات المستخدمة	58
8-3	القيم التقديرية (\hat{B}) للاستجابة الثالثة لجميع الطرائق وحجوم العينات المستخدمة	59
9-3	قيم متوسط مربعات الخطأ (MSE) للاستجابة الثالثة للمعالم المقدره والانموذج لجميع الطرائق وحجوم العينات المستخدمة	60
10-3	قيم التحيز (Biase) المطلقة للاستجابة الثالثة للمعالم المقدره وحجوم العينات المستخدمة	61
11-3	الاستجابة الحقيقية والمتوقعة المرتبطة بكل تركيز	62
12-3	يمثل قيم المتغيرات التوضيحية X_1, X_2 والاحتمالات المقابلة مضافاً اليها قيم Probit الحقيقية	64
13-3	يمثل قيم المتغيرات التوضيحية X_1, X_2 والاحتمالات المقابلة لها للاستجابات المتوقعة للطرائق المستخدمة	66
14-3	يمثل القيم التقديرية لمعالم الانحدار الخاصة للاستجابات المتوقعة لـ probit للطرائق المستخدمة	67
15-3	يمثل احتمال الاستجابة الحقيقية وقيم logit الحقيقية	68
16-3	يمثل احتمال الاستجابة المتوقعة وقيم logit المتوقعة	69
17-3	القيم التقديرية لمعالم الانحدار الخاصة للاستجابات المتوقعة لـ logit للطرائق المستخدمة	70
رقم الشكل	عنوان الشكل	الصفحة
1-2	شكل الدالة اللوجستية	24
2-2	العلاقة الغير الخطية بين احتمال الاستجابة (p_i) والمتغيرات المستقلة (X_i)	30
3-2	العلاقة الخطية بين المتغيرات التوضيحية (X_i) والمتغير الاستجابة (Y_i)	31

الفصل الاول

منهجية الدراسة

المقدمة	1 -1
مشكلة الدراسة	2-1
اهمية الدراسة	3-1
منهجية الدراسة	4-1
هدف الدراسة	5-1
الاستعراض المرجعي	6-1

1-1 المقدمة : introduction

يشهد العالم تطوراً ملحوظاً ومتسارعاً في مجالات الحياة وذلك عن طريق استعمال الاساليب الرياضية والاحصائية والتي تهدف الى حل المشاكل والمعوقات التي تواجه متخذ القرار في معظم مجالات المعرفة, ومن هذا المنطلق يبرز دور علم الاحصاء في استخراج النتائج وتحليلها لمعظم البحوث والدراسات عن طريق استعمال المقاييس والمؤشرات التي يحتاجها المخطط والمجرب, ويتم ذلك عادةً بناء نماذج لتحليل الانحدار لغرض تحليل اغلب الظواهر التي تتم عن طريق دراسة العلاقة بين المتغير المعتمد وبين متغير واحد او مجموعة من المتغيرات التوضيحية .

إذ تعد النماذج الوصفية من احد النماذج التي تمثل سلوك أغلب الظواهر وذلك عن طريق دراسة مجموعة من المتغيرات والتي يمكن أن يطلق على متغيرات هذه النماذج المتغيرات الوصفية والتي يكون عندها متغير الاستجابة أما ثنائي أو متعدد, لذا تكون دالة الاستجابة غير خطية او قد تكون لوجستية, ولكن من خلال استخدام أساليب التحويل الخطي ومنها اسلوب تحويل اللوغاريتم النسبة المضافة (logit) المعتمد على التوزيع اللوجستي , وايضا اسلوب تحويل وحدة الاحتمال (probit) المعتمد على التوزيع الطبيعي, عندها يمكن تحويلها الى استجابات خطية,^[39] ولكن قد تعاني هذه الدالة التي يتم تحويلها من مشكلة عدم ثبات تباين الخطأ عندها سيكون تقدير معالمها طريقة (OLS) مضللة وغير دقيقه, لذلك تم استخدام طريقة المربعات الصغرى الموزونة (WLS) لمعالجة هذه المشكلة^[10]. وقد تركزت الرسالة على كيفية دراسة تحليل البيانات للمتغير المعتمد الوصفي متعدد الاستجابة, وذلك عن طريق بناء نماذج استجابة ثنائية ونماذج استجابات متعددة ومن الناحيتين التجريبية والتطبيقية ويتم ذلك عادة استعمال طرائق تقدير خاصة للمتغير المعتمد متعدد الاستجابة ومقارنة هذه الطرائق مع المقاييس الاحصائية المستعملة ومنها مقياس متوسط مربعات الخطأ (MSE) لبيان افضلية هذه الطريقة.

وتسليط الضوء على فحوى الرسالة فقد تضمنت, أربعة فصول تضمن الاول مشكلة واهمية وكذلك منهجية والهدف والاستعراض المرجعي لبعض الدراسات والبحوث السابقة, أما الثاني فقد تناول مفهوم البيانات المصنفة وعلاقتها جداول التوافق والتوزيعات المستعملة, وكذلك تضمن الانموذج الخطي الاعتيادي و العام, بالإضافة الى النموذج الوصفي و النموذج اللوجستي , وكذلك تضمن بعض الطرائق التقدير الخاصة معالم أنموذج الانحدار للمتغير المعتمد متعدد الاستجابة, إذ تم استعمال طريقة المربعات الصغرى الموزونة (WLS) وكذلك طريقة الامكان الاعظم (MLE) الاعتماد على تطبيق طريقة نيوتن رافسن , وايضا تم استعمال طريقة الجاكناييف (jack), أما في الثالث فقد تضمن مبحثين الاول يتكون من الجانب التجريبي والمتضمن مفهوم المحاكاة للطرائق المستخدمة ومحاولة توليد المتغيرات العشوائية وكذلك مراحل بناء تجرئة

المحاكاة واستعراض النتائج , أما في المبحث الثاني فقد تناول الجانب التطبيقي والمتضمن استعمال تحليل إحدى التجارب الحياتية الخاصة لبعض المواد المبيدة للحشرات , أما في الرابع فقد تم التطرق الى الاستنتاجات والتوصيات , فضلا عن المصادر والملاحق.

problem of the Study

2-1: مشكلة الدراسة

في بعض التحليلات الاحصائية قد لا تكون المعلومات والبيانات رقمية وإنما قد تكون وصفية لذلك فإن الطرائق الكمية قد لا تفي الغرض آنذاك, كما ان الانموذج الذي يحتوي على بيانات وصفية سيواجه بعض المشاكل والمعوقات والتي تجعل نتائج التحليل مضللة وليس لها ما يؤيدها من الناحية النظرية, ومنها أن الخطأ العشوائي قد لا يتوزع توزيع طبيعي, وأيضا تباين الخطأ العشوائي قد يكون غير ثابت, وكذلك عدم حصر قيمة الاحتمال ضمن المدى المسموح $h(0,1)$, لذلك لا بد من بحث عن اجراء او اسلوب معين يختص بمعالجة مثل هكذا بيانات ومعلومات, ولذلك تم اختيار هذا الموضوع وكما يأتي.

importance of The Study

3-1: اهمية الدراسة

وجد عند تحليل ودراسة البيانات والمعلومات الاحصائية يمكن لمتخذ القرار ان يتعرف على عملية الاستجابة لاسيما اذا كان المتغير المعتمد ثنائي او متعدد الاستجابة , وعن طريق معطيات الحل يمكن لمتخذ القرار ان يتخذ قراراً صائباً اعتماداً على الفرضيات الاحصائية .

methodology of The Study

4-1 : منهجية الدراسة

تم العمل على المنهج الاستدلالي (الاستقرائي) وفيه يبدأ الباحث بملاحظة المشكلة ثم يتم وضع الفروض لها وبعد ذلك يتم اختبارها وقد تم استعمال الاسلوب الاحصائي على وفق هذا المنهج ,

Objective of The Study

5-1: هدف الدراسة

تهدف الدراسة الى ما يأتي:

- 1- دراسة وتحليل المتغير الوصفي وتأثيره في المتغير المعتمد متعدد الاستجابة .
- 2- تقدير معالم نموذج المتغير المعتمد متعدد الاستجابة باستعمال طرائق التقدير منها طريقة المربعات الصغرى الموزونة (WLS) وطريقة الامكان الاعظم (M.L.E) وذلك عن طريق الاعتماد على تطبيق خوارزمية نيوتن رافسن (Newton Raphson) وأيضا تم استعمال طريقة الجا كناف (jackknife) .

3- اختيار افضل طرائق التقدير المذكورة أنفأ وذلك اعتماداً على مقياس متوسط مربعات الخطأ (MSE) باستعمال الجانب التجريبي (محاكاة) والجانب التطبيقي .

-: Review of Literature

6-1: الاستعراض المرجعي

تم مناقشة موضوع البيانات المصنفة صورة عامة منذ القرن الماضي من لدن كثير من الكتاب والباحثين و بدأت تزداد دراسة المتغيرات الوصفية في حالة كون المتغير المعتمد متعدد الاستجابة شكل كبير في السنوات الاخيرة .

ففي عام (1973): قام الباحث (Amemiyal) [32] باستعمال طريقة المربعات الصغرى الموزونة (WLS) لتقدير معالم أنموذج الانحدار الوصفي عندما يكون تباين المتغير المعتمد الوصفي مرتبطاً مع المتغيرات التوضيحية .

قدم الباحثان (MCculagn , p. Nelder , j. A) [52] في عام (1983) : كتاباً تناولوا فيه استعمال اختبار مربع كاي (Chi –Square) واختبار الانحراف (Deviance of Test), وذلك لبيان كيفية توفيق او (حسن مطابقة) النموذج اللوجستي, وتوصل الباحثان الى ان كلا الاختبارين يقتران من توزيع مربع كاي .

وفي عام (1984) م اثبت الباحثان (Albert and Anderson) [33] إمكانية وجود مقدرات خاصة لأنموذج الانحدار اللوجستي متعدد الاستجابة للبيانات التي تعاني من مشكلة الفصل اذ تم تقسيمها الى ثلاث فئات هي الفصل التام (complete separation) والفصل شبه التام (quasicomplete separation) , والتداخل (overlap) وعن طريق إجراءات تكرارية , توصل الباحثان الى ان المقدر الذي يعظم دالة الإمكان ليس له صيغه نهائية ولا سبب في ذلك يعود الى استعمال الطرائق التكرارية للحل بطريقة نيوتن- رافسون.

حدثت (بيثون) [4] في عام (1992) : اختبار فرضيات المتعلقة بمعلمات انموذج الانحدار اللوجستي وذلك عن طريق دراسة خصائص هذا الاختبار وأسلوب بناء حدود الثقة لتقدير معلمات الخاصة بأنموذج الانحدار اللوجستي .

و في عام (2000) حدث (الطائي) [16]: عرض وتحليل المتغيرات المعتمدة الوصفية والمحددة وذلك دراسة بعض مواصفات انموذج الانحدار الخطي المتعدد , وقد استعمل طريقة خوارزمية التوقع الاعظم (EM), وطريقة خوارزمية نيوتن رافسن (N.R), فضلاً عن طريقة تحويلات المنحني اللوجستي (Logit) لتقدير معالم الانموذج, وذلك من خلال تجربة محاكاة ولثلاث مستويات من أحجام العينة, وتوصل الباحث باستعمال طريقة مونت كارلو في المحاكاة الى أن مقدرات تحويلات المنحني اللوجستي هي الافضل في العينات الصغيرة بينما كانت طريقة خوارزمية التوقع الاعظم هي الافضل في العينات المتوسطة , اما في

حالة العينات الكبيرة فقد كانت جميع الطرائق المستعملة متقاربة ومتكافئة, وكذلك توصل الباحث الى افضلية طريقة تحويلات المنحني اللوجستي في تحليل البيانات الحقيقية في الجانب التطبيقي.

في عام (2002) :حث (Tima) وآخرون^[58] : امكانية استعمال أنموذج الانحدار اللوجستي متعدد الحدود لدراسة اثر مجموعة من العوامل في تطور عينة اسرية, الاعتماد على معيارين لاختبار للأنموذج الملائم الاول يمثل معيار معلومات البيزية (Bayesian information Criterion) والآخر يمثل إحصاءه نسبة الامكان, استعمال البيانات الخاصة بالطبقة المهنية, وتوصلوا الى مجموعة من الحقائق منها ان تعليم الاب يعد العامل الاول المؤثر ثم عامل الجنس وأخيرا يأتي عامل الوراثة كأقل تلك العوامل تأثيرا.

وفي العام نفسه, حث (خير)^[8] : موضوعاً تناول فيه النماذج اللوغاريمية الخطية في جداول توافق ثنائية الإبعاد (2*2) والمستعملة في تطبيق البيانات المبوبة والمتمثلة بتأثير العمر في الإصابة بمرض السكر والربو إذ تم تصنيف هذه البيانات تحت فرضية معينة لغرض تحليلها ومن ثم الوصول الى النموذج الذي يمثل تلك البيانات, وتوصل الباحث الى ان نموذج الاستقلالية يمثل البيانات بصورة جيدة وان أكثر الاصابات مرض السكر والربو كانت في المدة العمرية (15-44) وتترايد الإصابة كلما تقدمت اتجاه الفئات العمرية الكبيرة .

قام (العزاوي)^[18] في عام (2005) :دراسة المقارنة بين بعض طرائق تقدير أنموذج الانحدار اللوجستي الاعتيادي والطرائق الحصينة عن طريق الاعتماد على مقياس متوسط منجات الخطأ (MSE) اذا استعمل أسلوب المحاكاة للتجارب الحياتية ذات الاستجابة الثنائية ومن اهم الاستنتاجات التي تم توصل اليها الباحث هي كفاءة الطرق الحصينة في تقدير معلمات انموذج الجرعة ذات الاستجابة الثنائية .

وفي عام (2007) : قام الباحث (McCarthy)^[51] :دراسة انموذج الانحدار اللوجستي ثنائي الاستجابة في جداول توافق ثنائية الإبعاد (2*2) وقد استعمل طرائق تقدير متعددة لتقدير معلمات انموذج اللوجستي لمجموعة من البيانات الوصفية ووجود فئات من نقاط بيانات الفصل التام والفصل شبه التام والتداخل, فضلا عن مشكلة التعدد الخطي , وتوصل الباحث الى ان طريقة الامكان الاعظم المستعملة في التقدير في حالة التداخل تكون غير دقيقة وصحيحة في حالة وجود مشكلة الفصل التام وغير التام .

في العام ذاته : بحث (إيليا)^[3] امكانية تطبيق طريقة (jackknife) والتي تستعمل لتقدير التحيز والتباين للإحصاءات المختلفة, كذلك تهدف الى تقليل التحيز والتباين في تقدير معلمات الانموذج , إذ تم تطبيق هذه الطريقة على بيانات حقيقية خاصة دالة الاستهلاك الاسري ومن ثم تطبيق هذه الطريقة على انموذج انحدار غير خطي لا يجاد تقدير المعلمات والتباينات .

قامت (الثعلبي) [5] عام (2008) : موضوعاً تناولت فيه دراسة أهم العوامل المؤثرة في حدوث التشوهات الولادية والتي اجريت في مستشفى البصرة للنسائية والتوليد , وتوصلت الباحثة الى كفاءة أنموذج الانحدار اللوجستي الخاصة في دراسة العوامل المؤثرة في حدوث الولادات المشوهة و بينت نتائج البحث ايضاً أن عامل الموقع الجغرافي لسكن الإاوين له تأثير كبير ويكون أحد اسباب حدوث التشوه الولادي.

وفي عام (2009) : قام الباحث (حسين) [9] دراسة اسلوب المحاكاة للمقارنة بين مقدرات الإمكان الأعظم الموزونة الحصينة WLME مع طرائق أخرى (مقدرات M_ , مقدرات LP_) والمستعملة في تقدير معالم أنموذج الانحدار اللوجستي ثنائي الاستجابة في حالة كون بيانات الظاهرة المدروسة ملوثة (بيانات شاذة), افتراض ثلاث مستويات لتلوث البيانات (0%, 10%, 30%) وتوصل الباحث الى نجاح طريقة WLME على سائر الطرائق المدروسة اعتماد المقاييس الإحصائية للتحيز (Bias) ومتوسط مربعات الخطأ (MSE) لمعاملات الانموذج.

نشر (AL- AFIFI) [30] عام (2010) : بحثاً تناول فيه استعمال انموذج الانحدار اللوجستي متعدد الاستجابة وذلك عن طريق دراسة تطبيقية والتي اجريت على بيانات العنف الجسدي , وتوصل الباحث الى ان انموذج الانحدار اللوجستي تمكن من تحديد العلاقة بين متغير الاستجابة ومجموعة من المتغيرات التوضيحية وفق نمودج رياضي وكذلك تمكن انموذج الانحدار اللوجستي ان يتنبأ تصنيف كل مفردة من المفردات في مستويات متغير الاستجابة .

وفي عام (2011): بحث (عبد الرزاق) [19] دراسة وتحليل أثر بعض المتغيرات في الإصابة بمرض اللثة استخدام انموذج الانحدار اللوجستي, وتوصل الباحث الى مجموعة من أهم الاستنتاجات كان من أبرزها هو ان عامل الترسبات الكلسية يعد السبب الرئيسي في الإصابة بمرض اللثة .

نشر الباحث (Krieng) [48] عام (2012) : بحثاً تناول فيه امكانية اجراء المقارنة بين الانحدار اللوجستي والتحليل التمييزي في تصنيف مجموعة من المصابين بأمراض سرطان الثدي وتوصل الباحث الى مجموعة من الاستنتاجات كان من أبرزها هي أن نتائج الانحدار اللوجستي يعطي تصنيفاً أكفاً من التحليل التمييزي.

وفي عام ذاته بحث (العابدي) [17] : أسلوب تحليل البيانات المصنفة المترتبة والتي تكون فيها الاستجابات عبارة عن أزواج مترتبة فيما بينها وذلك من خلال دراسة تطبيقية أجريت على معدلات الطلبة المتخرجين وعلاقتها بمعدلات قبولهم في المعهد وذلك من خلال الاعتماد على تصنيف هذه المشاهدات تحت صفتي (positive , Negative) والاعتماد على أعلى معدل للطلاب مع المعدل العام للظاهرة, وتوصل الباحث الى ان زيادة نسبة الترجيح (odds Ratio) تكون أعلى لصالح معدلات التخرج.

في عام (2013): درس الباحثان (Ogoke & others) [55] تعديل مقدرات انحدار الحرف اللوجستي المعمم بواسطة دالة الاستجابة الآسية لمصفوفة الأوزان ما يقلل من التباين, إذ استعمل المقدر المعدل (Jackknife) لتخفيض التباين , وتوصل الدراسة الى أن مقدر (Jackknife) متفوق على كل من مقدر الحرف اللوجستي ومقدر الحرف اللوجستي المعدل من خلال التباين والحد من التحيز.

وفي عام (2014): قامت الباحثة (جبارة) [6] بتحليل البيانات متعددة الاستجابة لتشخيص أمراض العيون عن طريق إجراء المقارنة ما بين الانحدار اللوجستي والدالة التمييزية, إذ تم التركيز على مسألة الاختيار ما بين أنموذج الانحدار اللوجستي متعدد الاستجابة ودالة التمييز الخطي كدوال تصنيف خطية, إذ يتم تصنيف البيانات على وفق ثلاثة نماذج , الصيغة الأولى أنموذج دالة التمييز الخطية, والثاني أنموذج احتمالات الاستجابة لدالة التمييز الخطي أما الثالث فهو أنموذج الدالة الاحتمالية لدالة الانحدار اللوجستي المتعدد الاستجابة وتم المقارنة بينهم على أساس احتمال خطأ التصنيف للوصول الى أفضل نموذج خطي لتشخيص انواع امراض العيون وتوصلت الباحثة الى أن طريقة التصنيف وفق الصيغة الاحتمالية لدالة التمييز الخطي تعطي أقل احتمال لخطأ التصنيف.

في عام (2015): بحث (سعيد) [14] امكانية دراسة العوامل المساعدة على تشخيص حالات الإصابة بمرض سرطان المثانة استعمال أنموذج الانحدار اللوجستي, إذ تم الاعتماد على عينة قوامها (128) مريضاً وعد جمع وتحليل البيانات , توصل الباحث الى مجموعة من الحقائق والاستنتاجات كان من أهمها: هو أن عامل (البيلة الدموية) كان له الاثر الأكبر في إمكانية تشخيص حالات الإصابة بالمرض , ثم عامل العمر , أما العوامل الأخرى فلم يكن لها تأثير معنوي في إمكانية تشخيص الإصابة بالمرض.

قامت الباحثة (سهاد) [2] في عام (2016): بأجراء مقارنة تضمنت إمكانية استعمال اشجار الانحدار التصنيفية مع الانحدار اللوجستي في تقدير أنموذج تجميعي, وقد تم استعمال طريقة الخوارزمية الجينية وخوارزميتي (Id3, c4.5) ولعدم فقدان التعميم تم استعمال التحليل اللوجستي كأداة تحليل مهمة, وتوصلت الباحثة الى أهم الاستنتاجات عن طريق عملية المقارنة ما بين الطرائق المستعملة أسلوب المحاكاة الى أفضلية طريقة خوارزمية التشجيرية (CART) وتفوق طريقة خوارزمية الجينية عند تحليل البيانات الحقيقية في الجانب التطبيقي.

الفصل الثاني

الجانب النظري

نماذج المتغيرات المعتمدة متعددة الاستجابة

المحتويات	
البيانات المصنفة	1-2
جداول التوافق	2-2
توزيع متعدد الحدود	3-2
النماذج الخطية الاعتيادية والعامية	4-2
نماذج الانحدار الوصفية	5-2
انموذج الانحدار اللوجستي	5-2-1
انموذج الانحدار اللوجستي الثنائي	5-2-1-1
انموذج الانحدار اللوجستي متعدد الاستجابة	5-2-1-2
بعض المشاكل الخاصة بالمتغير المعتمد الوصفي	5-2-2
نماذج دوال التحويل الخطي	6-2
انموذج لوغاريتم النسبة المضافة	6-2-1
انموذج وحدة الاحتمال	6-2-2
تقدير معلمات أنموذج الانحدار للمتغيرات المعتمدة متعددة الاستجابة	7-2
طريقة المربعات الصغرى الموزونة (WLS)	7-2-1
طريقة الامكان الاعظم (MLE)	7-2-2
طريقة نيوتن رافسن (Newton Raphson)	7-2-2-1
طريقة الجا كنيف (Jackknife of Method)	7-2-3

الفصل الثاني

نماذج المتغيرات المعتمدة متعددة الاستجابة

(1-2) البيانات المصنفة (Categorized Data) [17], [23], [26]

فاتح التطور الذي حصل في تحليل البيانات المصنفة (Categorized Data) عام (1960) نلد العالم الاحصائي كارل بيرسون (Karl Pearson) [26]، إذ تعتمد البيانات المصنفة بشكل عام على بدأ العد بدلاً من القياس، وهي تهتم بدراسة المتغيرات الوصفية التي يمكن تصنيف فيها الاستجابة بشكل صفات (ناجح/فاشل) (جيد/رديء)، إذ شاهدت الاستجابة التي تصنفها المتغيرات الوصفية للوحدة التجريبية المستعملة تسجل بشكل عدد (أي عدد المشاهدات التي تتحقق في كل صفة)، ولذلك تسمى هذه البيانات في بعض المراجع العلمية بالبيانات المعدودة (Count Data).

إذ تستعمل البيانات المصنفة (النوعية) في مجالات علمية متعددة كالعلوم الصحية والعلوم الاجتماعية وكذلك العلوم النفسية، و تحليلها يعتمد على جداول توافق ثنائي الاتجاه أي تحتوي على صفوف وأعمدة، ومن ثم حساب بعض الوحدات التجريبية التي تقيس العلاقة بين المتغيرات، وتكون الاصناف على نوعين :-

- الاصناف الاسمية (Nominal categories) ونقصد بها فردة الوحدة التجريبية يمكن تصنيف ضمن صنف واحد من بين عدة اصناف كصفة الحالة الاجتماعية (تزوج، اعزب، ارل، ل، طلق) و أيضا كصفة حالة دم الانسل (A او B او O ...)
- الاصناف المرتبة (ordered categories) هي الاصناف التي يمكن ترتيبها وفقاً لمعيار معين ثلاث تصنيف المستوى التعليمي (ابتدائية، متوسطة، جاعية، اعلى من الجاعية) وايضا تصنيف درجة الاصابة بمرض من التي قد تكون (لا شيء، معتدل، حاد) ويمكن اعتبار الاصناف المرتبة حالة خاصة من الاستجابات المتعددة.

(2-2) جداول التوافق (contingency tables) [7], [50]

هي عبارة عن اصناف او تراتيب خاصة بالبيانات الوصفية والتي تتمثل على شكل تكرارات ووضوعة في جداول توافق تضم تغييرين او اكثر وكل تغيير من تغيرات جداول التوافق يجرأ الى صنفين (ستوين) وفقاً لعدد المتغيرات الوصفية الخاصة بها [8].

وتكون جداول التوافق عادة على انواع مختلفة منها جداول التوافق ثنائية الابعاد ($r \times c$) (Two -Dimensional contingency tables)، وهذا النوع من الجداول يقوم على اساس

□ بدأ تقسيم او تجزئة بيانات العينة الى صفوف وأعمدة (Row & colum) وبيانات هذا النوع □ الجداول تكون على شكل تكرارات وتستعمل في الكثير □ التطبيقات □ها في العلوم الطبية والنفسية والحياتية والاجتماعية وغيرها □ التطبيقات الاخرى, على سبيل المثال يتم سحب عينة تتعلق بحالات الاصابة بمرض □ا عندها يتم تصنيف بيانات العينة □ن ثم تقسيمها الى قسمين هما جنس المريض التي تمثل الصفوف (r) والقسم الاخر يتعلق بنوع الاصابة بالمرض الذي يمثل الاعمدة (c), وهناك نوع اخر □ن جداول التوافق يسمى جداول توافق □تعددة الابعاد (multi-*Dimensional contingency tables*), يستعمل هذا النوع □ن الجداول في حالة وجود اكثر □ن □تغيرين ولكن بمستويات □ختلفة (اي □ كل □ستوى يضم بداخله □ستويات اخرى) ولهذه الجداول اهمية بالغة في التطبيقات الطبية والبيولوجية .

ويمكن تمثيل بيانات العينة المصنفة على شكل توزيع تكراري □وضع في جداول توافق ثنائية الابعاد (r*c) وتوضيحها كما في جدول (2-1) والذي يتكو □ن صفوف (i) والتي تمثل العال □ المستعمل في التجربة , وأعمدة (j) (j= 1, 2, . . . s) والتي تمثل اصناف الاستجابة (categories Response).

جدول (2-1)

يمثل تكرارات العينة المصنفة

$i \backslash j$	1	2	s	Total
1	k_{11}	k_{12}	k_{1s}	$k_{1.}$
2	k_{21}	k_{22}	k_{2s}	$k_{2.}$
⋮	⋮	⋮		⋮	⋮
r	k_{r1}	k_{r2}	k_{rs}	$k_{r.}$
total	$k_{.1}$	$k_{.2}$	$k_{.s}$	$k_{..} = n$

الجدول □ن اعداد الباحث بالاعتماد على المصدرين [16], [7]

إذ □

المشاهدة (i, j) تمثل المشاهدات التي تنتمي للمجموعة (i) وقد استجابت للصف (j)

كما و □ :

k_{ij} : تمثل تكرار المشاهدة في الصف (i) والعمود (j)

$k_{r.}$: تمثل المجموع الكلي للملاحظات في الصف (i) اي □ :

$$k_{r.} = k_{r1} + k_{r2} + \dots + k_{rj} = \sum_{j=1}^s k_{ij}$$

$k_{.s}$: تمثل المجموع الكلي للملاحظات في العمود (j) اي □ :

$$k_{.s} = k_{1s} + k_{2s} + \dots + k_{is} = \sum_{i=1}^r k_{ij}$$

$$\sum_{i=1}^r \sum_{j=1}^s k_{ij} = n$$

إذ □ :

n : تمثل □ مجموع □ شهادات الاستجابة الكلي

كما □ احتمال استجابة الصنف (i) للمعالجة (j) والذي يرمز له (P_{ij}) والذي يتضمن حاصل قسمة الاعداد التي استجابت للتجربة على المجموع الكلي لوحدة التجربة اي □ :

$$p_{ij} = \frac{k_{ij}}{n} \quad \dots \quad (2-1)$$

والجدول (2-1) يمكن تحويله الى قيم احتمالية بدلاً من الاعداد اعتماداً على المعادلة (2-1) وكما

□ وضح في الجدول (2-2), ولزيادة المعلومات □ راجعة المصدر [37, 7]

جدول (2-2) يمثل القيم الاحتمالية للعينة المصنفة

$i \backslash j$	1	2	s	Total
1	p_{11}	p_{12}	p_{1s}	1
2	p_{21}	p_{22}	p_{2s}	1
⋮	⋮	⋮		⋮	
r	p_{r1}	p_{r2}	p_{rs}	1

واعتماداً على هذه القيم الاحتمالية (p_{ij}) للبيانات المصنفة يمكن الحصول على عدة توزيعات إحصائية لها علاقة بهذه البيانات, و n هذه التوزيعات هو توزيع متعدد الحدود:

(2-3) توزيع متعدد الحدود (*multinomial Distribution*) [7], [46]

يعد توزيع متعدد الحدود n التوزيعات الاحتمالية المتقطعة ذات المتغيرات المتعددة المستعملة في البيانات الوصفية ويمثل هذا التوزيع الحالة العامة لتوزيع ثنائي الحدين (*Binomial distribution*) أي عند وجود اصناف متعددة n الحوادث او الحالات بدلاً من الحالتين, بعبارة اخرى n تجربة متعددة الحدود (*multinomial*) تكون ماثلة لتجربة ثنائي الحدين (*binomial*) الا انها تشترط ان يكون لها اكثر من محاولتين [15].

يعد توزيع متعدد الحدود n التوزيعات الاحتمالية كثيرة الاستعمال في مجالات علمية مختلفة كجالات الهندسة والفيزياء وغيرها, كما و n توزيع متعدد الحدود له مجالات اخرى في التطبيقات وهي عند ما تكون بيانات العينة المستعملة تأخذ صفوفاً واعمدة فالتغير الاستجابة هو الاخر يأخذ مستوى صفوف واعمدة فمثلاً, لو كل n تغير الاستجابة (y_{ij}) فهذا يعني ان الاستجابة العائدة للمجموعة (i) والواقعة ضمن اصناف الاستجابة (j) على n تكون اصناف الاستجابة مستقلة n صنف الى اخر [57].

كما n دالة الكثافة الاحتمالية المشتركة ل $(k_{i1}, k_{i2}, \dots, k_{im})$ الخاصة بتوزيع متعدد الحدود يعبر عنها بالمعادلة الاتية على فرض n (k) تغير عشوائي و (n) يمثل حجم العينة

$$P(k_{i1}, k_{i2}, \dots, k_{ij}) = \frac{n!}{k_{i1}! k_{i2}! \dots k_{ij}!} p_{i1}^{k_{i1}} p_{i2}^{k_{i2}} \dots p_{ij}^{k_{ij}}$$

$$= \frac{n!}{\prod_{j=1}^s k_{ij}} \prod_{j=1}^s p_{ij}^{k_{ij}} \dots \quad (2-2)$$

إذ n :

$$P_{ij} \geq 0, \quad \sum_{i=1}^r \sum_{j=1}^s P_{ij} = 1, \quad \sum_{i=1}^r \sum_{j=1}^s k_{ij} = n$$

كما \square الوسط الحسابي $E(y_{ij})$ والتباين $V(y_{ij})$ لتوزيع \square تعدد الحدود هو

$$E(y_{ij}) = n p_{ij} \quad . \quad . \quad . \quad (2 - 3)$$

$$\text{var}(y_{ij}) = n p_{ij} (1 - p_{ij}) \quad . \quad . \quad . \quad (2 - 4)$$

$$\text{cov}(y_{ij}) = -p_{ij} (1 - p_{ij}) \quad . \quad . \quad . \quad (2 - 5)$$

له \square صفوفه ارتباط \square تماثله \square فردة شاذة $|V|=|R|=0$ والسبب في ذلك يعود الى شروط المفروضة على هذا التوزيع وهو $(\sum x = n, \sum p_i = 1)$, لذلك \square الناحية العملية يتم حذف احد المتغيرات الاقل اهمية \square بين \square تغيرات المتجه X للتخلص \square من الشذوذ .

(4-2): النماذج الخطية الاعتيادية والعامه [16], [12], [20]

Ordinal and Generalized Linear models

تعرف النماذج الخطية الاعتيادية بشكل عام على انها علاقة رياضية خطية تربط بين \square تغييرين أحدهما المتغير المعتمد يسمى أحياناً بمتغير الاستجابة والاخر قد يكون \square ن واحد او اكثر \square ن المتغيرات المستقلة (التوضيحية), ويمكن التعبير عنها بالصيغة الاتية

$$Y_i = \beta_0 + \beta X_i + U_i \quad . \quad . \quad . \quad . \quad . \quad . \quad (2 - 6)$$

إذ \square

Y_i : تمثل \square تغير المعتمد للمشاهدة (i)

X_i : يمثل المتغيرات التوضيحية للمشاهدة (i)

β_0, β_i : ثوابت (المقطع الصادي والميل) لنموذج الانحدار على الترتيب

U_i : يمثل الخطأ العشوائي للمشاهدة (i)

كما \square الخطأ العشوائي (U_i) يبني على عدة افتراضات \square نها :

1- اعتدالية توزيع الاخطاء (normality of Errors) اي \square تكون الاخطاء \square وزعة بشكل طبيعي عند كل \square مستوى \square ن \square مستويات المتغير المستقل.

2- تجانس تباين الخطأ العشوائي (U_i)

$$E(U_i^2) = \sigma^2$$

$$E(U_i U_j) = 0, \quad i \neq j$$

\square اذا \square تباين الخطأ غير ثابت فسوف تظهر \square مشكلة عدم تجانس التباين الخطأ

(Heteroscedastity Problem)

3- الاخطاء العشوائية تتوزع توزيعاً طبيعياً بمتوسط (0) وتباين ثابت \square قدره (σ^2) اي \square

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

4- الخطأ العشوائي \square مستقل عن قيمة المتغير التوضيحي (X_i)

$$COV(U_i, X_i) = 0$$

5- عدم وجود تعدد خطي بين المتغيرات ولا ارتباط ذاتي بين الاخطاء العشوائية (البواقي).

\square النموذج الذي يمثل \square رحلة توصيف البيانات يجب \square يكون \square خطياً في المتغيرات, \square اذا \square كان النموذج غير خطي فيمكن استعمال اساليب التحويل لهذه القيم عندها تستعمل البيانات المحولة لإيجاد \square عالم النموذج والسبب في ذلك يعود الى \square المتغيرات الاصلية عادة \square تحدث فيها انتهاك لفرض واحد او اكثر \square ن فرضيات النموذج القياسية المهمة والتي يجب \square تتوفر في هذا النموذج, لأ \square دقة التقديرات لمعالم النموذج في الواقع التطبيقي \square توقف \square عتمد على \square دى صحة هذه الفرضيات, \square ن اكثر الفرضيات التي يحدث فيها انتهاك هي تلك الفرضية التي تتعلق بخطية النموذج وكذلك عدم ثبات (تجانس) تباين الخطأ, و تنشأ \square هات \square الفرضيت \square بسبب الطبيعة الثنائية للمتغير المعتمد (Y), و \square عدم تحققها سيؤدي الى الحصول على سلسلة \square ن النتائج والتقديرات لا تحقق صفات او خصائص المقدر الجيد والكفؤ, \square يجعل نتائج الاختبارات \square ضللة و غير دقيقة, \square ن هذه الفرضيات :

1- خطية النموذج : تعد الفرضية الخطية \square ن اهم الفرضيات التي يجب \square تتحقق في نموذج الانحدار, و عند \square ما يعاني النموذج \square ن فقد \square هذا الافتراض عندها تكون تقديرات \square عالمه غير دقيقة وغير صالحة للتنبؤ والسبب في ذلك يعود الى وجود بعض الفجوات التي تظهر نتيجة لفقد \square هذا الافتراض بوصفها \square ن الفرضيات المهمة التي يجب \square تتحقق في النموذج, و هناك العديد \square ن النماذج غير الخطية كالنموذج اللوجستي, اضافة الى الصيغ الأسية واللوغاريتمية, والتي بواسطة تحويلات \square ناسبة, كالتحويل النسبة المضافة (logit) المعتمدة

على التوزيع اللوجستي , وكذلك تحويلة وحدة الاحتمال (probit) المعتمدة على التوزيع الطبيعي يمكن عالجتها وتحويلها الى نماذج خطية, والجدول (3-2) يبين بعض المنحنيات القابلة لتحويل الخطي في المعالم والمتغيرات .

جدول (3-2): يبين بعض الصيغ غير الخطية والتي يمكن تحويلها الى الصيغ الخطية:

الصيغة	التحويل	الصيغة الخطية
$y = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$	$y' = Z = \frac{x - \mu}{\sigma}$	$y' = A + BX$
:	:	:
$y = \frac{e^{A+BX}}{1 + e^{A+BX}}$	$y' = \ln\left(\frac{y}{1-y}\right)$	$y' = A + BX$

للمزيد من التحويلات يمكن مراجعة المصدرين [16], [12]

2- عدم ثبات (تجانس) التباين (heteroscedasticity): وهو احد اكثر الافتراضات التي تفتقد في انموذج الانحدار الخطي, وهذا المصطلح يتكوّن من كلمتين الاولى (hetero) بمعنى غير متساو بينما الاخر (scedasticity) بمعنى الانتشار, كما أنّ شكل عدم ثبات التباين يمكن ملاحظتها في معظم الدراسات والابحاث ولاسيما تلك التي تعتمد على بيانات احصائية تأخذ شكل بيانات المقطع العرضي (cross section Data) والسبب في ذلك يعود الى كل شهادة من شهادات المقطع العرضي ولاسيما شهادات المتغيرات التوضيحية (X_i) تحتوي على مجموعة من المشاهدات لنفس السنة, قد تكون هذه المشاهدات ذات احوام مختلفة وبن ثم فأ هذه المشاهدات قد تؤثر في المتغير التابع. [11], [20]

وايضا يمكن ملاحظة شكل عدم ثبات التباين بكثرة في بعض النماذج التي يكون فيها التباين مرتبط مع توسط, بعبارة اخرى المتغير المعتمد الذي يتم تحليله يخضع لتوزيع احتمالي له تباين يرتبط او يعتمد مع المتوسط فاذا كل المتوسط يرتبط مع قيمة المتغير المستقل (X_i) عندها فأ تباين (Y_i) سوف يتغير مع المتغير المستقل (X_i), وفي هذه الحالة فأ تباين الخطأ العشوائي (U_i) قد تجردت نه حالة الثبات وأصبح تباين الخطأ غير ثابت, كما و توزيع المتغير المعتمد (Y_i) سيخضع الى توزيع احتمالي اخر غير التوزيع الاحتمالي المتعارف عليه الا وهو التوزيع الطبيعي, وهذا بدوره سوف يؤدي الى حصول تغير في توزيع الخطأ العشوائي استناداً الى حكم العلاقة التي تربط المتغير المعتمد مع الخطأ العشوائي (U_i) لا أي تغير في توزيع المتغير المعتمد سيؤدي الى حصول تغير في الخطأ العشوائي (U_i), وبذلك يكون تباين الخطأ العشوائي غير ثابت, عندها لا بد من أخذ حالة عدم الثبات بنظر الاعتبار عند اجراء عمليات التقدير, فاذا لم تؤخذ في الحسبان فأ هذه التقديرات التي يتم الحصول عليها لا تمتلك اقل

تباين قدر مكن, لذلك لابد من استعمال بعض الاجراءات التي تجعل التباين ثابتاً و من ثم يتم تقدير عالم الانموذج [12].

في ظل فرضية عدم تجانس تباين الخطأ (Heteroscedasticity assumption) سوف لن تمتلك تقديرات المربعات الصغرى الاعتيادية (OLS) لمعالم النموذج الخطي الاعتيادي صفة اقل تباين قدر مكن في حالة العينات الكبيرة والصغيرة, بعبارة اخرى سوف لا تتمتع بخاصية افضل تقدير خطي غير تحيز (B. L. U. E), لذلك لابد من اللجوء الى طريقة او أسلوب يهدف الى تصحيح كفاءة قدرات (OLS) في ضل عدم الثبات التباين, لذا ف أسلوب الاث لهذا هو استعمال طريقة المربعات الصغرى الموزونة (weighted least square) إذ يتم وضع أوز (w_i) مع تباين الخطأ العشوائي لتخلص من عدم ثبات التباين وكما يأتي [43]:

$$E(u_i^2) = \sigma_i^2 = \frac{\sigma^2}{w_i} = \sigma^2 W_i \quad \dots \dots \dots \quad (2 - 7)$$

إذ :

$$W_i = \frac{1}{w_i}$$

يمكن نتخلص من مشكلة عدم تجانس التباين وذلك عن طريق التعال مع المتغيرات المحولة, وذلك بضرب طرفي المعادلة (2 - 6) ب (√w_i) وكالاتي :

$$\sqrt{w_i} Y_i = B_0 \sqrt{w_i} + B_1 \sqrt{w_i} X_i + \sqrt{w_i} U_i \quad \dots \dots \dots \quad (2 - 8)$$

كما توقع ربع الخطأ العشوائي الموزون في النموذج الخطي يأخذ الشكل الاتي :

$$E(\sqrt{w_i} U_i)^2 = w_i E(U_i)^2 \\ = w_i \sigma_i^2 = w_i \frac{\sigma^2}{w_i} = \sigma^2$$

كذلك يمكن دراسة انحدار المتغير المعتمد تعدد الاستجابة باستعمال طريقة المصفوفات (الطريقة العامة) او طريقة النموذج الخطي العام (Generalized linear model) الذي يمثل بالصيغة الاتية :

$$\underline{Y} = \underline{X}\underline{B} + \underline{U} \quad \dots \dots \dots \quad (2 - 9)$$

إذ □

Y : □ تجه المتغير المعتمد ذات رتبة (n × 1)

X : □ صفوف □ شاهدات المتغيرات التوضيحية ذات رتبة (n × k)

B : □ تجه ثوابت وهي □ علامات الانحدار بضمنها الحد الثابت ذو رتبة (n × 1)

U : □ تجه الاخطاء العشوائية ذو رتبة (n × 1)

وبتطبيق النموذج الخطي العام في المعادلة (9 - 2) وباستعمال المصفوفات لغرض أستبعاد ظاهرة عدم ثبات التباين وكالاتي :

$$\begin{bmatrix} \sqrt{W_1} Y_1 \\ \vdots \\ \sqrt{W_n} Y_n \end{bmatrix} = \begin{bmatrix} \sqrt{W_1} X_{11} & \sqrt{W_1} X_{12} & \cdots & \sqrt{W_1} X_{1K} \\ \sqrt{W_2} X_{21} & \sqrt{W_2} X_{22} & \cdots & \sqrt{W_2} X_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \sqrt{W_n} X_{n1} & \sqrt{W_n} X_{n2} & \cdots & \sqrt{W_n} X_{nK} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_K \end{bmatrix} + \begin{bmatrix} \sqrt{W_1} U_1 \\ \sqrt{W_2} U_2 \\ \vdots \\ \sqrt{W_n} U_n \end{bmatrix}$$

و على فرض □ :

$$\rho^{-1} = \begin{bmatrix} \sqrt{W_1} & 0 & \cdots & 0 \\ 0 & \sqrt{W_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{W_n} \end{bmatrix}$$

لذلك فأ □ المعادلة (9-2) بعد ضربها بـ (ρ⁻¹) ستكون □ بالشكل الاتي :

$$\rho^{-1} \underline{Y} = \rho^{-1} \underline{X} \underline{B} + \rho^{-1} \underline{U} \quad (2 - 10)$$

و □ (ρ) هي (ρ⁻¹)⁻¹ لذلك فأ □ (ρ) ستكون □ صفوف قطرية بالشكل الاتي إذ □ قيم القطر الرئيس هي عبارة عن □ قلوب الجذر التربيعي للأوز □ :

$$\rho = \begin{bmatrix} \frac{1}{\sqrt{W_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{W_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{\sqrt{W_n}} \end{bmatrix}$$

و عند ضرب المصفوفة (ρ) بالمصفوفة المبدلة سنحصل على :

$$\rho\rho' = \begin{bmatrix} \frac{1}{W_1} & 0 & \dots & 0 \\ 0 & \frac{1}{W_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{W_n} \end{bmatrix}$$

□ من ثم تطبيق طريقة المربعات الصغرى العامة (Generalized least squares) لتقدير □ عالم الانموذج الخطي العام بعد استبعاد ظاهرة عدم التجانس للحصول على افضل تقدير خطي غير □ تحيز.

□ إذ يعد الانموذج الخطي العام (Generalized linear models) حالة خاصة (□ تداد) للنموذج الخطي الاعتيادي والتي تم التعبير عنه على وفق المعادلة (2-9) , ويحتوي النموذج الخطي العام على المركبات الاتية [42].

*□ رتبة الجزء العشوائي (Random part component) عند تحليل هذه المركبة تفترض ثبات (تجانس) وتباين الخطأ □ وهذا الخطأ □ ستقل عن المتغيرات التوضيحية الاخرى, وانه يتوزع توزيعاً طبيعياً بمتوسط (0) وتباين (σ^2), ولذلك ف□ تغير الاستجابة (Y) هو الاخر سيتوزع توزيعاً طبيعياً بمتوسط (μ) وتباين (σ^2), وترتبط □ رتبة الجزء العشوائي بدالة الاستجابة الاحتمالية (Y) و تحتوي هذه المركبة على توزيع احتمالي يعود الى □ مجموعة □ ن التوزيعات الاحتمالية الاسية وتسمى بتوزيعات العائلة الاسية (exponential family) [52]

$$Y \sim N(\mu, \sigma^2)$$

*ركبة الجزء المنتظم (systematic part component) تحتوي هذه المركبة على صنفوفة
شاهدات المتغيرات التوضيحية (X) وتجه المعالم (B) التي تحتويها عادلة نموذج الخطي
العام و XB يمثل الجزء المنتظم .

*دالة الربط (link function) هي الدالة التي تربط بين ركبة الجزء المنتظم مع ركبة
الجزء العشوائي .

و $f_y(y, \theta_i)$ دالة الاستجابة الاحتمالية التي تحتويها ركبة الجزء العشوائي في النموذج
الخطي العام , يمكن كتابتها بصيغة العائلة الاسية كما يأتي [52], [59] :

$$f_y(y, \theta_i) = \exp \left\{ \frac{y\theta_1 - b(\theta_1)}{a(\theta_2)} + c(y, \theta_2) \right\} (2-11)$$

إذ \square :

$a(\cdot)$, $b(\cdot)$, $c(\cdot)$ تمثل دوال \square حدودة , (θ_i) تمثل عالم النموذج

وبإدخال اللوغاريتم لدالة الإك \square الاعظم (Maximum likely hood) الخاص
للمعادلة (2-11) المذكورة انفاً ف \square دالة الاستجابة الاحتمالية ستكون \square :

$$\ln f = \frac{y\theta_1 - b(\theta_1)}{a(\theta_2)} + c(y, \theta_2)$$

ويمكن الحصول على المتوسط والتباين للمتغير الاستجابة (y) عن طريق العمليات الاتية [54] :

اشتقاق المعادلة المذكورة انفاً بالنسبة (θ_1, θ_2) و \square ن ثم \square ساواتها للصفر يمكن الحصول على
المتوسط والتباين وكما يأتي :

$$\frac{\partial \ln L}{\partial \theta_1} = 0$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{y - b'(\theta_1)}{a(\theta_2)} = 0 \quad (2 - 12)$$

و \square انها نحصل على قيمة (y) إذ \square :

$$y = b'(\theta_1) = E(y)$$

وبذلك فإن \square توسط المتغير (y) هو :

$$E(y) = b'(\theta_1) = \mu \quad \quad (2 - 13)$$

ولإيجاد التباين لابد من إيجاد المشتقة الثانية للمعادلة (2-12) بالنسبة لـ θ_1 نحصل على :

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = - \frac{b''(\theta_1)}{a(\theta_2)}$$

$$\begin{aligned} E\left(\frac{\partial^2 \ln L}{\partial \theta_1^2}\right) + E\left(\frac{\partial \ln l}{\partial \theta_1}\right)^2 &= 0 \\ &= \left(-\frac{b''(\theta_1)}{a(\theta_2)}\right) + \frac{E(y - b'(\theta_1))^2}{a(\theta_2)} = 0 \end{aligned}$$

إذ \square :

$E(y - b'(\theta_1))^2$ يعبر عن تباين المتغير المعتمد (y)

$$= -\frac{b''(\theta_1)}{a(\theta_2)} + \frac{\text{var}(y)}{a(\theta_2)^2} = 0$$

وعند ضرب طرفي المعادلة المذكورة انفاً بـ $a(\theta_2)^2$ نحصل على :

$$- a(\theta_2) b''(\theta_1) + v(y) = 0$$

$$\text{var}(y) = b''(\theta_1) a(\theta_2) \quad \quad (2 - 14)$$

$$0 < b''(\theta_1) < \infty$$

□□ عادلة (14 – 2) يتبين أن تباين المتغير المعتمد (y) ناتج عن حاصل ضرب دالتين الدالة الأولى تعتمد على معلمة (θ_1) وتسمى بالمعلمة القانونية (canonical parameter), وتعتمد المعلمة الثانية على (θ_2) وتسمى بالمعلمة التشتت (*Dispersion parameter*).
بينما عندئذ تسلك دالة الاستجابة الاحتمالية توزيع بواسون (Poisson distribution), فأ الصيغة الاحتمالية توصف بالصورة الآتية [54].

$$f(y_i, \theta_i, \mu) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}$$

$$y_i = 0, 1, 2, \dots, n$$

وبأخذ اللوغاريتم لطرفي المعادلة نحصل على

$$\ln f(y_i, \theta_i) = y_i \ln \mu - \mu - \ln y_i!$$

□□ عندئذ مقارنة المعادلة المذكورة انفاً □□ مع المعادلة (11 – 2) نجد □

$$\ln \mu = \theta_1$$

$$e^{\theta_1} = \mu$$

$$b'(\theta_1) = e^{\theta_1} = \mu$$

$$b''(\theta_1) = e^{\theta_1} = \mu$$

$$a(\theta_2) = \theta_2 \quad \text{at } \theta_2 = 1$$

$$c(y_i, \mu) = -\ln(y_i!).$$

و بذلك يمكن الحصول □□ توسط وتباين المتغير المعتمد (y) الذي يتبع توزيع بواسون □□ بالاعتماد على الصيغتين (13 – 2) و (14 – 2)

$$E(Y_i) = b'(\theta_1) = e^{\theta_1} = \mu$$

$$\text{var}(Y_i) = a(\theta_2) \cdot b''(\theta_1) = e^{\theta_1} = \mu$$

نلاحظ □□ □□ توسط التوزيع يكون □□ ساوياً للتباين.

وكذلك عندئذ تسلك دالة الاستجابة الاحتمالية توزيعاً ثنائياً الحدين binomial distribution

□□ الصيغة الاسية للنموذج توصف وفق المعادلة الآتية [42], [16]

$$f(y, \theta_i) = C_y^n p^y (1 - p)^{n-y}$$

كما و دالة الاكالا اعظم لتوزيع ثنائي الحدين ستكو بالشكل الاتي:

$$f(y, \theta_i) = \prod_{i=1}^n C_y^n p^y (1 - p)^{n-y}$$

وبادخال اللوغاريتم لطرفي المعادلة المذكورة انفاً سوف نحصل على :

$$\ln f(y, \theta_i) = y \ln p + (n - y) \ln(1 - p) + \ln C_y^n$$

وبعد ادخال اللوغاريتم الاكالا ستكو المعادلة بالصورة الاتية :

$$\ln f(y, \theta_i) = y \ln \left(\frac{p}{1-p} \right) + n \ln(1 - p) + \ln C_y^n$$

وعند قارنتها مع صيغة المعادلة (11 - 2) نحصل على :

$$\ln \frac{p}{1-p} = \theta_1$$

وبرفع المعادلة المذكورة انفاً للأساس (e) نحصل على

$$\frac{p}{1-p} = e^{\theta_1}$$

ونها فاً :

$$p = \frac{e^{\theta_1}}{1 + e^{\theta_1}}$$

ولذلك فاً احتمال (1 - p) ستكو قيمته بالشكل الاتي :

$$1 - p = \frac{1}{1 + e^{\theta_1}}$$

وبالتعويض عن (1 - p) بما يساويها ونها فاً :

$$b(\theta_1) = n \ln(1 + e^{\theta_1})$$

وبأخذ المشتقة الاولى لـ b لطرفي المعادلة اعلاه بالنسبة الى (θ₁) فاً :

$$b'(\theta_1) = n \frac{e^{\theta_1}}{1 + e^{\theta_1}}$$

$$a(\theta_2) = \theta_2 \text{ and } \theta_2 = 1$$

$$c(y, \theta_2) = \ln C_y^n$$

كما □ المشتقة الثانية لـ b للحصول على التباين :

$$b''(\theta_1) = n \cdot \frac{e^{\theta_1}}{(1 + e^{\theta_1})^2}$$

وبمقارنة المعادلات (2 - 13), (2 - 14) □ النتائج التي تم الحصول عليها ستكون □ قيمة التوقع والتباين للمتغير المعتمد الذي يخضع لتوزيع ثنائي الحدين كما يأتي

$$E(Y) = b'(\theta_1) = n \cdot \frac{e^{\theta_1}}{1 + e^{\theta_1}} = np$$

$$\text{var}(Y) = b''(\theta_1)a(\theta_2)$$

$$= n \cdot \frac{e^{\theta_1}}{(1 + e^{\theta_1})^2} \cdot (1) = n \cdot \frac{e^{\theta_1}}{(1 + e^{\theta_1})} \cdot \frac{1}{(1 + e^{\theta_1})} = np(1 - p)$$

ويمكن وصف دوال الاستجابة الاحتمالية بالطريقة نفسها, اي بصورة الانموذج الخطي □ هما مختلف نوع التوزيع □ معادلات الانحدار للبيانات الوصفية فقد تكون □ قيده ببعض القيود.

(2 - 5): نماذج الانحدار الوصفية [16], [31], [40]

Qualitative Regression models

تعد نماذج الانحدار الوصفية □ النماذج الاحصائية المهمة و الشائعة الاستعمال في كثير □ مجالات المعرفة, والتي تسهم في بناء انموذج احصائي يستعمل لغرض تقدير طبيعة العلاقة بين □ تغير الاستجابة (y) وبين □ تغير واحد او □ مجموعة □ المتغيرات التوضيحية, إذ لا يضع انموذج الانحدار الوصفي اي قيود على قيم المتغيرات التوضيحية (X) سواء أكانت هذه المتغيرات قابلة للقياس الكمي ام الوصفي ولا تؤثر في تفسير □ علماته [31].

في نماذج الانحدار الاعتيادية يكون المتغير المعتمد y متغيراً مستمراً قد يأخذ قيم حرة غير قيّدة بقيود $(-\infty, \infty)$ عند اجراء عمليات التحليل الاحصائي لها كما في المنحني الطبيعي, ولكن عندئذ يكون المتغير المعتمد ثنائي الاستجابة يمثل تغيراً تقطعاً يأخذ القيم $(0, 1)$ عندها ستكون نماذج الانحدار الوصفية على شكل دالة احتمالية خطية (linear probability function) تأخذ القيم $(0, 1)$ وكما في المعادلة الاتية:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_{ij} X_{ij} + U_{ij} \quad \dots \dots \dots \quad (2 - 15)$$

إذ U_{ij} :

y_i هو تغير الاستجابة ويأخذ قيم حدوده وهي $(0, 1)$

X_i : يمثل المتغيرات التوضيحية

$\beta_0, \beta_1, \dots, \beta_i$: معالم الانحدار

U_{ij} : يمثل الخطأ العشوائي

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

بعبارة اخرى U_{ij} تغير الاستجابة قد يأخذ قيمة (1) عند حدوث استجابة ويأخذ القيمة (0) عند عدم حدوث استجابة, لكن عند وضع معادلة الانحدار لابد ان تحقق الفروض الخاصة بالانحدار, واحياناً بعضها قد لا يتحقق, كتبناين الخطأ العشوائي قد لا يتوزع طبيعياً, وكذلك عدم كانية حصر حدود الاحتمال بين $(0, 1)$, لذا فأن اغلب الباحثين يتجاهلون الطبيعة الثنائية للمتغير المعتمد y والتركيز على استخدام الانحدار الاعتيادي لغرض الحصول على تقديرات دقيقة ونطقية.

معظم النماذج الاحصائية للبيانات الوصفية تعتمد على طبيعة السلوك الذي يحكم الاستجابة وكذلك تعتمد على اهداف التحليل الوصفي.

وتعد نماذج الاستجابة الثنائية (dichotomous) والمتعددة (polychotomous) حالة خاصة من نماذج الانحدار الوصفي.

(1 - 2 - 5): نموذج الانحدار اللوجستي [47], [6]

Logistic Regression Model

من المعلوم أن نماذج الانحدار تكون على نوعين اساسيين α نماذج انحدار خطي او نماذج انحدار غير خطي, ويعد النموذج اللوجستي واحد من نماذج الانحدار غير الخطي والذي يمتاز بالمرونة العالية, ولكنه بحد ذاته يكون مائلاً للانحدار الخطي الاعتيادي من ناحية توضيح درجة

العلاقة بين تغير الاستجابة وبين تغير واحد أو مجموعة من المتغيرات التوضيحية , لكن جوهر الاختلاف يكمن في طبيعة المتغير المعتمد (y) الخاص بالنموذج الانحدار اللوجستي وهو يجب أن يكون ثنائي أو تعدد الاستجابة, وهذا الاختلاف بحد ذاته ينعكس على الافتراضات الخاصة بالانحدار الاعتيادي .

والدالة اللوجستية التي تقدر هي دالة غير خطية لذا يتم استعمال المتغيرات المحولة والذي يتم اختيارها لغرض تحويلها الى دالة الاستجابة الخطية ويتم ذلك عادة عن طريق اسلوب تحويل لوغار يتم النسبة المضافة (logit) المعتمد على التوزيع اللوجستي .

يستخدم انموذج الانحدار اللوجستي عادة لوصف طبيعة العلاقة بين تغير الاستجابة وبين تغير أو مجموعة من المتغيرات التوضيحية لغرض التنبؤ بوجود خاصية معينة [13] , ويعد انموذج الانحدار اللوجستي حالة خاصة من نماذج الانحدار الاعتيادي وذلك بسبب الطبيعة الاسمية التي يحملها المتغير المعتمد (مثل حياة, موت), و يستعمل انموذج الانحدار اللوجستي بشكل واسع في تحليل البيانات والتطبيقات الاحصائية في مجالات علمية مختلفة كالتحليل الصحية والاجتماعية, والسلوكية, والتعليمية وغيرها من المجالات العلمية الاخرى [24].

وتوجد انواع متعددة لأنموذج الانحدار اللوجستي منها أنموذج الانحدار اللوجستي الثنائي (binary Logistic Regression), وانموذج الانحدار اللوجستي المتعدد (Multi- Logistic Regression), والذي يعد الاكثر شيوعاً [19].

(5-2-1-1): انموذج الانحدار اللوجستي الثنائي [44], [47]

Binary Logistic Regression Model

يعد الانحدار اللوجستي الثنائي أسلوباً رياضياً يستعمل لغرض تشخيص وتوفيق طبيعة العلاقة بين تغير الاستجابة وبين مجموعة من المتغيرات التوضيحية والتي عادة ما تكون علاقة غير خطية, يعتمد انموذج الانحدار اللوجستي الثنائي على فرض اساسي هو أن المتغير المعتمد (y) تغير ثنائي الاستجابة أي أن استجابة المتغير المعتمد صنفه ضمن مجموعتين, اذا كانت هذه المجموعة هي المطلوبة فتعد (نجاحاً) واذا كانت غير المطلوبة فتعد (فشلاً), ولقد زادت اهمية استعمال التحليل اللوجستي يوماً بعد اخر لكونه يهتم في تحليل البيانات ثنائية القيم , والتي يكون فيها المتغير المعتمد عند تحقق الاستجابة يأخذ (y=1) باحتمال (p) وعند عدم تحقق الاستجابة فأخذ (y=0) وباحتمال (1-p), وبذلك فأن تغير الاستجابة يتبع توزيع برنولي بالصيغة الاتية

$$E(y_i) = 0(1 - p) + 1(p) \quad i = 0,1$$

$$E(y_i) = p \quad \dots \dots \dots (2 - 16)$$

كما و □ الدالة اللوجستية يمكن كتابتها على وفق المعادلة الآتية: [27]

$$P_i = \frac{e^{(\beta_0 + \sum_{i=1}^n \beta_i X_i)}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \quad \dots \dots (2 - 17)$$

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i X_i)}} \quad \dots \dots (2 - 18)$$

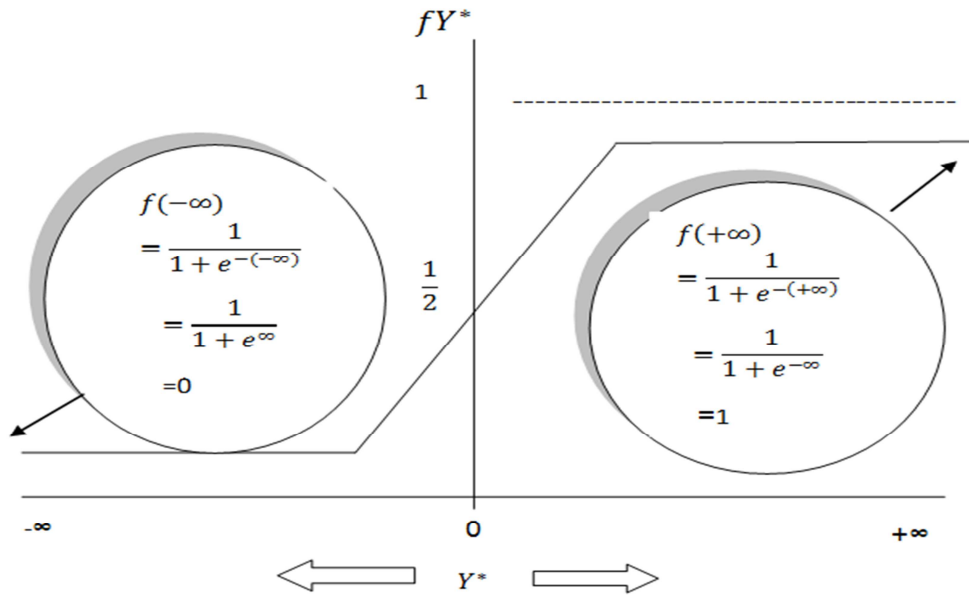
$$p(y = 0) = 1 - \frac{e^{(\beta_0 + \sum_{i=1}^n \beta_i X_i)}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \\ \frac{1}{1 + e^{(\beta_0 + \sum_{i=1}^n \beta_i X_i)}} \quad (2 - 19)$$

حيث □ :

□ عالم □ جهولة □ راد تقديرها $\beta_0, \beta_1, \dots, \beta_n$

□ يمثل دالة الاستجابة $f(y^*)$

ويمكن عرض الدالة اللوجستية بالشكل الآتي [47]:



الشكل (2-1) يبين شكل الدالة اللوجستية

وعن طريق الشكل المذكورة انفاً يتضح بأن الدالة اللوجستية $f(Y^*)$ هي دالة احتمالية واقعة ضمن القيم $(0,1)$.

وبذلك فإن النسبة $\left(\frac{p_i}{1-p_i}\right)$ عبارة عن مقدار واجب واقع ضمن $(0, \infty)$ وبأخذ لوغاريتم هذا المقدار سنحصل على (logit) والتي تهدف الى جعل انحدار اللوجستي نموذجاً خطياً اي $(-\infty \leq p \leq \infty)$, وكما في المعادلة الآتية :

$$\text{logit} = \text{Ln} \left(\frac{p_i}{1-p_i} \right) \quad \dots \quad (2-20)$$

[6], [25] : أنموذج الانحدار اللوجستي متعدد الاستجابة (5-2-1-2)

Multiple Response Regression model

يعد انموذج الانحدار اللوجستي متعدد الاستجابة واحداً من النماذج الاحصائية المهمة في تحليل البيانات المصنفة , فاذا كان انموذج الانحدار اللوجستي الثنائي يستعمل عندنا يكون تغيير الاستجابة له قيمتين فقط $(0,1)$ فإن انموذج الانحدار اللوجستي المتعدد يعد تداً له اذ انه يستعمل عندنا يكون لمتغير الاستجابة اكثر من قيمتين فمثلاً في التجارب الطبية فإن تأثير دواء معين غالباً يتم قياسه بشكل رتب فاذا كانت طبيعة الاستجابة صنفه بصيغة لا يوجد تأثير للدواء (None) او يوجد تأثير نخفض (Little) او تأثير توسط (Moderate) او تأثير عال (High) لذا فإن تغيير الاستجابة في هذه الحالة قد يأخذ اكثر من قيمتين, وعليه فإن انموذج الانحدار اللوجستي المتعدد يعد محاولة لتطوير وتوسيع الانموذج الثنائي .

يعد نموذج الانحدار اللوجستي متعدد الاستجابة حالة خاصة من توزيع تعدد الحدود (multinomial Distribution) كما في معادلة (2-2)

وعندنا يسلك المتغير المعتمد (y) سلوكاً ثنائي الاستجابة فإن انموذج الانحدار الخطي هو نموذج احتمالي يقع بين $(0,1)$ [43] :

$$y_i = \beta_0 + \beta_i X_i + u_i$$

وعندنا يأخذ المتغير المعتمد (y) القيمة $(0,1)$ فإنه يماثل توزيع برنولي ولذلك فإن توقع المتغير المعتمد (y) كما في معادلة (2-16) يساوي p اي $E(y)=p$ لذلك فإن الاحتمال الشرطي $E(Y|X)$ (conditional probability) هو ايضاً يقع بين $(0,1)$ اي $0 \leq E(Y|X) \leq 1$, عند ذلك فإن معادلة الانحدار ستكون بالصيغة الآتية :

$$E(Y_i|X_i) = \beta_0 + \beta_i X_i \quad \dots \quad (2-21)$$

وإدماً □ توقع دالة المتغير المعتمد الوصفي (دالة الاستجابة) يمثل احتمالاً , كما في المعادلة-2
(16) فأ□ توقع دالة الاحتمال الشرطي في المعادلة المذكورة انفاً يجب □ يقع بين (0, 1)

وبمقارنة المعادلة (2 - 16) مع المعادلة (2 - 21) نحصل على التكافؤ الآتي :

$$E(Y_i|X_i) = \beta_0 + \beta_i X_i = p_i \quad \dots \dots \dots (2 - 22)$$

أي □ المعادلة المذكورة انفاً وتساوي قيمة الاحتمال ولذلك فأ□ هذه المعادلة هي ايضاً تكو□
ثنائية القيم □ وحصورة بين (0, 1), لذلك فأ□ المتغير المعتمد (الاستجابة) لا يخلو □ ن □ أشكال
ولاسيما عند استخدام نماذج الانحدار الوصفية.

(2 - 2 - 5): بعض المشاكل المتعلقة بالمتغير المعتمد الوصفي [16], [32]

(Related problems when dependent Variable descriptive)

في الواقع التطبيقي قد نجد بعض المشاكل المتعلقة بالمتغير المعتمد الوصفي □ ن هذه المشاكل:

1 - الخطأ العشوائي قد لا يتوزع طبيعياً (Non normal Random Error) [43]

□ ن شروط الانحدار الخطي □ الخطأ العشوائي يتوزع توزيعاً طبيعياً بمتوسط (0) وتباين (σ^2)
على الرغم □ ن □ بعض طرائق التقدير لا تشترط توزيع الخطأ العشوائي توزيعاً طبيعياً
كطريقة المربعات الصغرى الاعتيادية (OLS), ولكن لو افترضنا □ الاخطاء العشوائية تتوزع
توزيعاً طبيعياً وذلك لغرض الاستدلال الاحصائي والذي يمثل (التقديرات واختبار الفرضيات),
و□ تغير الاستجابة في البيانات الوصفية قد يأخذ قيمة (0,1) وعندها فأ□ الخطأ (U_i) سيتخذ
(0,1) ويتبع توزيع برنولي, وعند تعويض □ تغير الاستجابة ب (1) أو (0) سنحصل على :

$$U_i = Y_i - \beta_0 - \beta_i X_i \quad \dots \dots \dots (2 - 23)$$

□ الاحتمالات المرافقة لقيم المتغير المعتمد Y هي :

Y_i	U_i	Probability
when 1	$1 - \beta_0 - \beta_i X_i$	P
when 0	$0 - \beta_0 - \beta_i X_i$	1-p

(2 - 24)

وفقاً لذلك لا يمكن اعتبار ب□ توزيع الخطأ العشوائي (U_i) يتبع توزيع طبيعي, وإنما سيتبع
احتمالي اخر الا وهو توزيع برنولي .

□ عدم تحقق فرضية التوزيع الطبيعي ليس بالموضوع الحرج جداً وذلك لا نه كما هو
 □ تعارف عليه اذا كانت □ قدرات المربعات الصغرى الاعتيادية (OLS) هي □ قدرات نقطية ذو
 خاصية (غير □ تحيزة) [43] اذا كا □ هدف الباحث هو الحصول على تقديرات نقطية , وفي هذه
 الحالة تكون □ افتراضات (Normality) غير ضرورية , فضلاً عن ذلك, اذا زاد حجم العينة
 بصورة غير □ حدة , فأ □ النظرية الاحصائية وبالتحديد (نظرية الغاية المركزية) تشير الى
 اقتراب □ قدرات المربعات الصغرى □ ن خاصية التوزيع الطبيعي.

لذا يكون □ الاستدلال الاحصائي بالنسبة للعينات الكبيرة هو للنموذج الاحتمالي الخطي (L P M)
 الذي يتبع توزيعاً □ قدرأ للمربعات الصغرى (OLS) في ظل افتراض نظرية التوزيع الطبيعي .

2 – عدم ثبات تجانس تباين الخطأ (Non constant error variance) [44]

□ ن بعض □ أشكال نموذج الاستجابة الثنائية هي خاصية عدم ثبات تجانس تباين الخطأ رغم □
 $E(u_i) = 0$, $[cov(u_i, x_i) = 0 \text{ if } i \neq j]$ و هنا لا يوجد ارتباط □ تسلسل (serial
 Correlation) وقد افترض الانموذج الاحتمالي الخطي (L P M) تجانس تباين الخطأ وهذا
 ليس بالأمر الغريب, وذلك لا □ النظرية الاحصائية توضح □ المتوسط والتباين لتوزيع برنولي
 هما (P, P(1 - P)) على الترتيب ,

إذ □ :

P : يمثل احتمال النجاح

وهذا يعني □ التباين هو على شكل دالة يعتمد على □ توسط التوزيع بعبارة اخرى □ تباين
 التوزيع يكون □ مرتبط دالياً □ مع المتوسط , وعليه يكون □ تباين الخطأ العشوائي هو تبايناً غير ثابت,
 وهذه المشكلة تؤدي الى اختلال التقديرات التي نحصل عليها □ ن النموذج وابتعادها عن
 خصائص الجودة (BLUE) وعن طريق تطبيق تعريف صيغة التباين المذكورة انفاً سيكون □
 تباين الخطأ العشوائي هو كالاتي:

$$\text{var}(u_i) = E(u_i^2) - \{E(u_i)\}^2$$

على فرض □ $E(u_i) = 0$

$$\text{var}(u_i) = E(u_i^2) = (1 - \beta_0 - \beta_i X_i)^2 p_i + (-\beta_0 - \beta_i X_i)^2 (1 - p_i)$$

وبالتعويض عن p_i في المعادلة (22 - 2) نحصل على :

$$\text{var}(u_i) = E(Y_i|X_i)(1 - E(Y_i|X_i)) = p_i(1 - p_i) \quad \dots \quad (2 - 25)$$

لذا فـ $E(Y_i|X_i)$ تبين الخطأ العشوائي في النموذج الاحتمالي الخطي غير-تجانس, ويمكن استبعاد ظاهرة عدم التجانس عن طريق العمل مع المتغيرات المحولة اي بتطبيق المعادلة (10 – 2)

3 - عدم تقييد دالة الاستجابة بحدود الاحتمال

Non fulfillment of $0 \leq E(Y_i|X_i) \leq 1$

بما $E(Y_i|X_i)$ توقع $E(Y_i|X_i)$ تغيير الاستجابة يمثل احتمال الاستجابة في انموذج الاحتمال الشرطي $(L p m)$ وقيمته $(0, 1)$ بصورة بين $(0, 1)$, وبالتالي فـ $E(Y_i|X_i)$ دالة الاستجابة الاحتمالية يجب $E(Y_i|X_i)$ تقع ضمن $(0, 1)$ اي $E(Y_i|X_i)$

$$0 \leq E(Y_i|X_i) \leq 1$$

رغم $E(Y_i|X_i)$ ليس هناك $E(Y_i|X_i)$ يضمن (guarantee) بـ $E(Y_i|X_i)$ قدر (Y_i) سيكو بالضرورة $E(Y_i|X_i)$ قيد ضمن بالقيمة $(0, 1)$, فاذا لم يتم استخدام اي اجراء او تحويل على قيم الطرف الايمن $E(Y_i|X_i)$ فـ $E(Y_i|X_i)$ من الممكن في هذه الحالة $E(Y_i|X_i)$ تتعدى قيم الاحتمال (p_i) المتنبأ بها عن قيمة المدى الخاص للاحتمال وتجاوز القيمة التقديرية للمتغير المعتمد (\hat{Y}_i) عن قيمة الواحد, وبذلك تصبح هذه القيمة $E(Y_i|X_i)$ ضللة وغير-نطقية وليس لها $E(Y_i|X_i)$ يؤيدها $E(Y_i|X_i)$ الناحية النظرية, وبالتالي فـ $E(Y_i|X_i)$ القرارات التي تتخذ على اساس هذا النموذج لا تمثل المجتمع المدروس ولا تتفق مع الواقع.

لذا فـ $E(Y_i|X_i)$ هناك طريقتين لجعل $E(Y_i|X_i)$ قدر الاحتمال الشرطي $E(Y_i|X_i)$ قيد بالقيمة $(0, 1)$ وبن هذه الطرق هي:

الطريقة الاولى: يمكن اعتماد طريقة المربعات الصغرى الاعتيادية او غيرها عند تقدير $E(Y_i|X_i)$ علمات الانموذج الخطي الاحتمالي $E(Y_i|X_i)$ هذه الطريقة لا تأخذ التقييد في الحسب, ويتم ذلك وباستخدام (9 – 2) فـ:

$$\underline{Y} = \underline{X}\underline{B} + \underline{U}$$

$$\underline{U} = \underline{Y} - \underline{X}\underline{B}$$

وبظرب طرفي المعادلة بمبدلة قيمة $E(Y_i|X_i)$ وجه الخطأ العشوائي (U') وبتعويض عنها بما يساويها نحصل على

$$U'U = (Y - XB)'(Y - XB)$$

$$U'U = Y'Y - Y'XB - B'X'Y + B'X'XB$$

يلاحظ $E(Y_i|X_i)$ الصيغة المذكورة بـ الحد الثاني والثالث ذات قيمة $E(Y_i|X_i)$ حدة ويمكن جمعها

$$U'U = Y'Y - 2B'X'Y + B'X'XB$$

و عند اخذ المشتقة الخطأ العشوائي بالنسبة لـ B وسواتها للصفر نحصل على

$$\frac{\partial U'U}{\partial B'} = -2X'Y + 2X'Xb = 0$$

ونها فأ

$$X'Xb = X'Y$$

$$\hat{b} = (X'X)^{-1}X'Y \quad \dots \dots \dots (2 - 26)$$

ثم نعوض القيم التقديرية لمعالم نموذج الانحدار الخطي (2 - 26) بالمعادلة (2 - 22).

في بعض الاحيا قد تكون $E(Y_i|X_i)$ اصغر من الصفر (قيمة سالبة) عندها تكون القيمة التقديرية للمتغير المعتمد $(Y_i = 0)$

$$E(Y_i|X_i) = \hat{Y}_i = 0$$

وإذا كانت اكبر من واحد سوف نفترض بـ

$$E(Y_i|X_i) = \hat{Y}_i = 1$$

الطريقة الاخرى : هي ابتكار تقنية تقدير تضمن (guarantee) بـ توقع الاحتمال الشرطي يقع بين (0, 1), انها أساليب التحويل وهما لوغاريتم النسبة المضافة (logit) وانموذج وحدة الاحتمال (probit).

(2 - 6) : نماذج دوال التحويل الخطي

models of transformation linear function

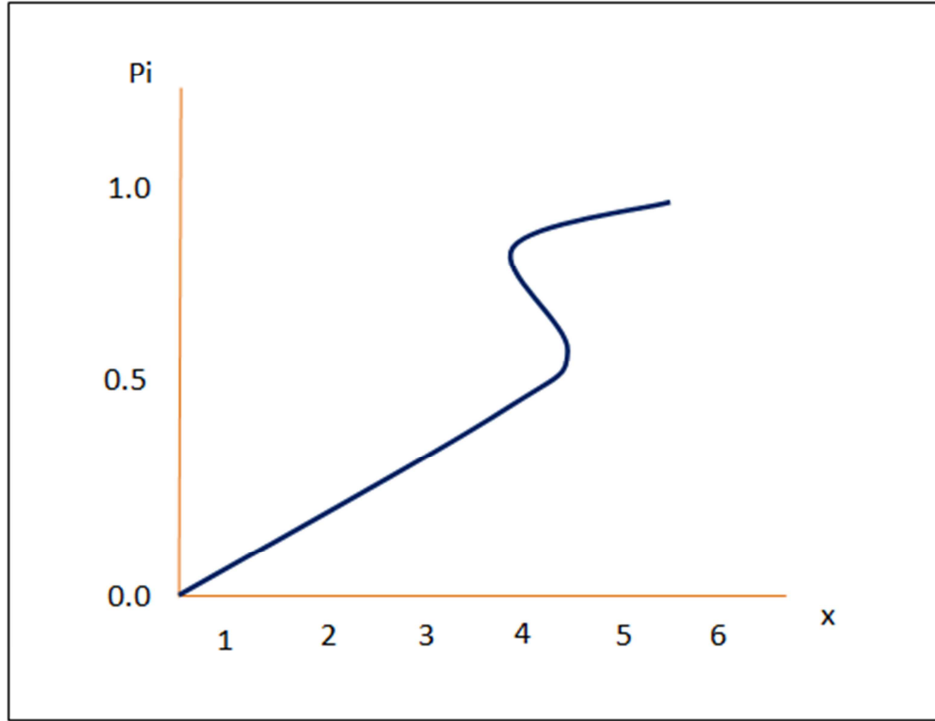
الغاية الرئيسية من اجراء عملية التحويل الخطي هو تغيير شكل التوزيع التكراري لكي يلائم الافتراضات النظرية المتضمنة في الانموذج, فقد تم استعمال التوزيع اللوجستي والتوزيع الطبيعي كنموذج تحويل (logit) ونموذج تحويل (probit), إذ يتم تطبيق البيانات على الحالتين تمثل الحالة الاولى اللوجستيك واستعمل فيها انموذج (logit) والحالة الثانية تمثل التوزيع الطبيعي وقد استعمل فيها انموذج (probit) ون هذه النماذج هي:

(1 - 2 - 6): أنموذج لوغاريتم النسبة المضافة [35], [47]

Models of odds ratio (logit)

بما أن طبيعة العلاقة التي تربط تغير ثنائي الاستجابة مع مجموعة من المتغيرات التوضيحية هي علاقة غير خطية, ويمكن شاهدة هذه العلاقة عن طريق دالة الاستجابة والمعرفة بالمعادلة

(2-17), ويمكن توضيح هذه العلاقة كما في الشكل (2-5) الآتي :



الشكل (2-2) يمثل العلاقة غير الخطية بين احتمال الاستجابة (p_i) والمتغيرات المستقلة (X_i)

فلو كان p هو احتمال النجاح في الانحدار اللوجستي وكان $(1-p)$ هو احتمال الفشل، فأما نسبة النجاح على الفشل ستنتج علاقة غير خطية، وقد يواجه الباحث صعوبة عند التحويل إلى نموذج اللوجستي خطي لغرض تقدير معالم النموذج، ولهذا اقترح الباحث الأيركي (*berkson*) عام (1944) آلية تحويل صيغة دالة اللوجستيك إلى دالة خطية وذلك عن طريق أخذ اللوغاريتم الطبيعي لطرفي المعادلة، واطلق على هذه التحويلة بتحويلة (*logit*)، والمعرفة في المعادلة (2-20) ويمكن إعادة كتابتها لغرض التطبيق التحويل:

$$\text{logit}(p_i) = \ln \frac{p_i}{1 - p_i}$$

وبتطبيق المعادلة (2-17) و (2-19) فأما :

$$\text{logit}(p_i) = \ln \left[\frac{e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \right] \left[\frac{1}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \right]$$

$$\text{Logit}(p_i) = \beta_0 + \sum_{i=1}^n \beta_i X_i$$

وعلى فرض \square $Y^* = \text{logit}(p_i) = \ln \frac{(p_i)}{(1-p_i)}$ \square لذا فـ Y^* يمكن اعادة كتابتها بالشكل التالي

$$Y^* = \beta_0 + \sum_{i=1}^n \beta_i X_i \quad i = 1, 2, \dots, n \quad \dots \quad (2 - 27)$$

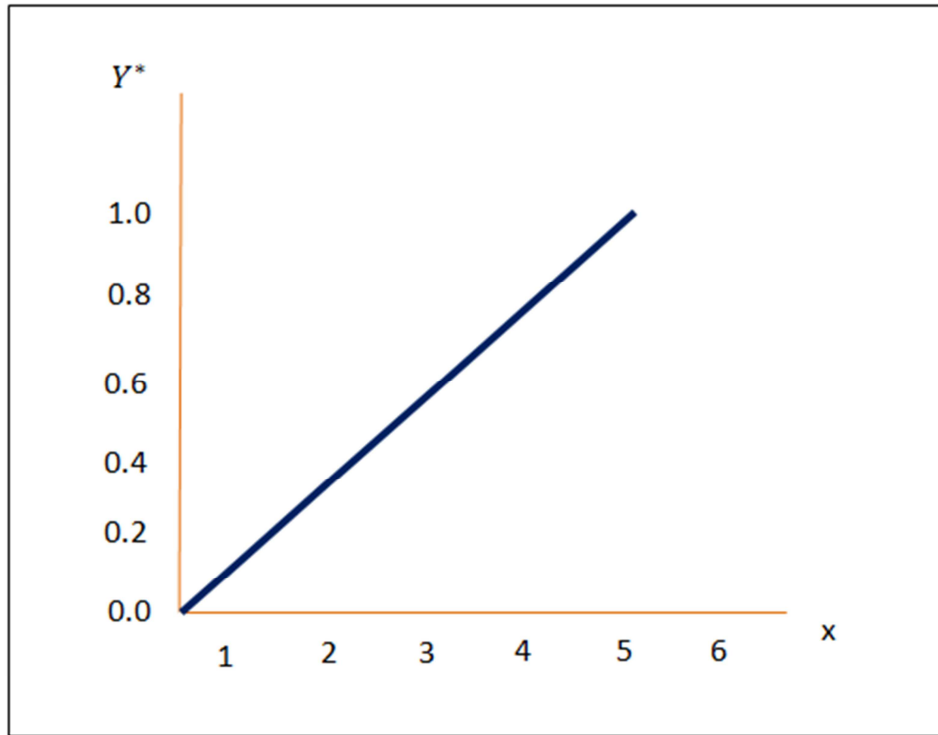
إذ \square :

Y^* : تمثل \square تغير اللوجستيك الجديد (المحول) (logit)

β_0, β_i : تمثل \square معالم النموذج

X_i : تمثل المتغيرات التوضيحية

وعند تطبيق تحويله (logit) ستكو \square طبيعة العلاقة بين \square تغير الاستجابة و \square جموعة المتغيرات التوضيحية علاقة خطية يمكن توضيحها كما في الشكل (2-6)



الشكل (2 - 3): يوضح العلاقة بين (Y^*, X) عند تحويل دالة اللوجستيك الى دالة خطية

(2 - 2 - 6): انموذج وحدة الاحتمال [36],[41] [7]

(unit of the probability model (probit)

يستعمل انموذج وحدة الاحتمال (probit) بكثرة في الابحاث البيولوجية والطبية [41] ويعتمد أسلوب وحدة الاحتمال على فكرة استعمال التوزيع الطبيعي لغرض توصيف البيانات بخط انحدار مناسب وذلك عن طريق تحويل هذه البيانات الى النسب الى الانحرافات الطبيعية, عند ذلك نستعمل انموذج وحدة الاحتمال عند تطبيق التوزيع الطبيعي إذ يمكن تحويلها الى شكل خطي عن طريق استعمال الدرجة المعيارية والتي تتكون من $z = \frac{x-\mu}{\sigma}$ $z \sim N(0,1)$ عندها نفرض $\beta_1 = \left(\frac{1}{\sigma}\right)$ و $\beta_0 = \left(\frac{-\mu}{\sigma}\right)$ عندها فأ \square الدالة الاحتمالية الخطية (linear probability function) تأخذ الصيغة الآتية :

$$Y_i = \beta_0 + \beta_i X_i + U_i. \quad (2 - 28)$$

وبما \square قيمة اي احتمال يجب \square يقع ضمن المدة (0,1), لذلك فأ \square احتمال (p_i) هو دالة ل Y_i ويأخذ الصورة الآتية

$$Y_i = \begin{cases} 1 & \text{if } u_i \leq y_i^* = \beta_0 + \beta_i X_i \\ 0 & \text{if } u_i > y_i^* = -\beta_0 + \beta_i X_i \end{cases}$$

$$u \sim N(0,1)$$

$$p(y_i = 1) = p(u_i \leq y_i^*) = p_i \quad (2 - 29)$$

$$p(y_i = 0) = p(u_i > y_i^*) = 1 - p(u_i \leq y_i^*) = 1 - p_i$$

\square ن المعادلة (2-29) فأ \square $(p \leq y_i^*)$ يساوي الدالة التجميعية , $F(Y)$ والتي تساوي :

$$P(u \leq y_i^*) = \Phi(y_i) = \int_{-\infty}^{y_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2 - 30)$$

و \square ن \square معادلة (2-29) يتم التعويض عن قيمة (Y_i) بما يساويها في \square معادلة (2-30)

$$\Phi(\beta_0 + \beta_i X_i) = \int_{-\infty}^{\beta_0 + \beta_i X_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

اذ \square الطرف الايمن \square من المعادلة (2-30) تسمى بالمنحرف المكافئ الطبيعي (*normal*)
 وقد تكون \square وجبة ولأنها احتمال فلا يمكن \square تكون \square سالبة , وتخلص \square من هذه القيم السالبة اضافة
 الباحث الايريكي (Finney) [41] قيمة (5) الى طرفي \square معادلة المنحرف المكافئ الطبيعي
 (N E D) وقد حصل على \square ما يسمى بوحدة الاحتمال (probit) إذ وضع جداول خاصة بوحدة
 الاحتمال يمكن \square راجعتها في (الملحق الخاص \square من الرسالة) .

لذا ف \square وحدة الاحتمال (probit) يمكن توضيحها بأنها عبارة عن المنحرف المكافئ الطبيعي
 \square ضافاً اليها القيمة (5)

$$\Phi(Y_i) + 5 = N E D + 5$$

$$Probit (\tilde{y}) = N E D + 5 \quad . . . \quad (2 - 31)$$

أذ \square :

\tilde{y} : تمثل وحدة الاحتمال

وهناك جداول خاصة باحتساب قيم وحدة الاحتمال (probit) والاحتمالات المقابلة لها في
 الجدول (انظر الملحق) اذ قام الباحث (Finney) بتحويل النسب الاحتمال الى انحرافات طبيعية
 وضعها في تلك الجدول \square انها نحصل على قيم وحدة الاحتمال (probit) .

(7-2) طرائق تقدير معالم المتغير المعتمد متعدد الاستجابة

بعد ان تم بناء و وصف النماذج, تأتي مرحلة تقدير معالمه ويتم ذلك من خلال استعمال بعض طرائق التقدير الخاصة بالمتغيرات متعددة الاستجابة والتي تعد احد أهم الجوانب في الاستدلال الاحصائي لأغلب النماذج, ومن هذه الطرائق هي طريقة المربعات الصغرى الموزونة (weighted Least squares), وطريقة الامكان الاعظم (Maximum Likelihood Method) وذلك من خلال تطبيق طريقة خوارزمية نيوتن رافسن (Newton- Raphson Method), وايضا تم استعمال طريقة الجا كنايف (Jackknife of Method).

(7-2-1) -: طريقة المربعات الصغرى الموزونة (المرجحة) (18) (12)

weighted Least squares method (WLS)

عندما يكون المتغير المعتمد في أنموذج الانحدار الخطي ثنائي الاستجابة, عندها سيفقد النموذج اغلب الافتراضات الخاصة به, ومن هذه الافتراضات هو ان الخطأ العشوائي قد لا يتبع توزيعاً طبيعياً إنما يتبع توزيعاً متقطعاً وليس مستمراً, وبمجرد اذا افترضنا بأن الخطأ العشوائي يتبع توزيع طبيعى عندها يكون تباين الخطأ العشوائي غير ثابت .

عندما يكون المتغير المعتمد ثنائي الاستجابة (0,1) عندها لا يمكن تقدير علامات انموذج الانحدار الخطي باستعمال طريقة المربعات الصغرى الاعتيادية (OLS) عندها سيكون تباين الخطأ العشوائي في هذه الحالة غير ثابت وبالتالي قد يعاني من مشكلة عدم الثبات, لذا تم اللجوء الى استعمال طريقة المربعات الصغرى الموزونة (WLS) والتي تستعمل لغرض تصحيح كفاءة المقدرات (OLS) في ظل عدم ثبات التباين الخطأ العشوائي و تعطي افضل تقدير خطي غير تحيز (BLUE) .

وفي هذه الطريقة تكون فيها العلاقة بين المتغير المعتمد الخطي (Y^*) (والذي يمثل لوغاريتم النسبة المضافة $\frac{P}{1-P}$ (logit) اي $Y^* = \ln \frac{P}{1-P}$) وبين المتغير واحد او مجموعة من المتغيرات التوضيحية (X_i) والتي يمكن تمثيلها بشكل المعادلة الخطية العكسية لغرض الوصول الى المصفوفات والتي عن طريقها يتم استبعاد مشكلة عدم ثبات التباين وكما في المعادلة الاتية :

$$\underline{Y^*} = \underline{X}\underline{B} + \underline{U} \quad \dots \dots \dots \quad (2 - 32)$$

إذ أن

Y^* : يمثل موجه عمودي ذو رتبة (n^*1)

X : مصفوفة مشاهدات المتغيرات التوضيحية ذات الرتبة (n^*k)

B : متجه المعلمات من الرتبة (n^*1)

U : متجه يمثل الخطأ العشوائي ذو رتبة (n^*1)

وبعد أخذ التوقع والتباين لنموذج المعادلة (2-32) عندها يكون المتغير (Y^*) يحمل النتائج التي يمكن تمثيلها بالصورة التالية: [12]

$$E(Y^*) = XB \quad \dots \quad (2 - 33)$$

$$Var(Y^*) = \frac{1}{\sigma^2} = \frac{1}{np_i(1 - p_i)} \quad \dots \quad (2 - 34)$$

$$Y^* \sim N(XB, \frac{1}{np_i(1 - p_i)})$$

وعند اجراء عملية الاشتقاق الصيغة التقديرية لمعلمات الانموذج المبين في المعادلة (2-32) يتم اولاً ضرب طرفي المعادلة (2-32) بمعكوس المصفوفة القطرية المتماثلة (ρ_i) لغرض استبعاد مشكلة عدم التجانس وبذلك سوف نحصل على الانموذج الآتي:

$$\rho^{-1}Y^* = \rho^{-1}XB + \rho^{-1}U \quad \dots \quad (2 - 35)$$

أذ ان

ρ : تمثل مصفوفة قطرية متماثلة يمكن تمثيلها بالصورة الآتية :

$$\rho = \begin{bmatrix} 1 & \dots & 0 \\ \sqrt{n_1 p_1 q_1} & & \vdots \\ \vdots & \frac{1}{\sqrt{n_2 p_2 q_2}} & \vdots \\ 0 & \dots & \frac{1}{\sqrt{n_r p_r q_r}} \end{bmatrix} \quad \dots \quad (2 - 36)$$

وبعد ذلك يتم تقدير معالم الانموذج باستعمال طريقة المربعات الصغرى (OLS) وكما يأتي:

$$Y^* = XB + U$$

$$U = Y^* - XB$$

$$U'U = (Y^* - XB)'(Y^* - XB)$$

وبضرب طرفي المعادلة بـ (ρ^{-1}) نحصل على:

$$(\rho^{-1}U)'(\rho^{-1}U) = (\rho^{-1}Y^* - \rho^{-1}XB)'(\rho^{-1}Y^* - \rho^{-1}XB)$$

$$(\rho^{-1}U)'(\rho^{-1}U) = (Y^{*'}\rho'^{-1} - B'X'\rho'^{-1})(\rho^{-1}Y^* - \rho^{-1}XB)$$

وبفك الاقواس نحصل على :

$$(U'\rho'^{-1}\rho^{-1}U) = Y^{*'}\rho'^{-1}\rho^{-1}Y^* - Y^{*'}\rho'^{-1}\rho^{-1}XB - B'X'\rho'^{-1}Y^* + B'X'\rho'^{-1}\rho^{-1}XB$$

على فرض أن

$$w = \rho'^{-1}\rho$$

لذلك فإن:

$$(U'W^{-1}U) = Y^{*'}W^{-1}Y^* - Y^{*'}W^{-1}XB - B'X'W^{-1}Y^* + B'X'W^{-1}XB$$

إذ أن :

W^{-1} : يمثل معكوس مصفوفة الاوزان وناتجة عن حاصل ضرب المصفوفة القطرية مع مبدلتها

وبأخذ التفاضل الجزئي الاول بالنسبة الى موجه المعلمة B نحصل على :

$$\frac{\partial U'W^{-1}U}{\partial B} = -2X'W^{-1}Y^* + 2X'W^{-1}XB = 0$$

ومنها فإن :

$$X'W^{-1}Y^* = X'W^{-1}XB \quad \dots \dots \dots \quad (2 - 37)$$

$$\hat{B} = (X'W^{-1}X)^{-1}X'W^{-1}Y^* \quad \dots \dots \dots \quad (2 - 38)$$

ويمكن التعبير عن معكوس مصفوفة الاوزان بما يأتي

$$w^{-1} = \begin{bmatrix} n_1 p_1 q_1 & 0 & \dots & 0 \\ 0 & n_2 p_2 q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_r p_r q_r \end{bmatrix}$$

كما وان Y^* : يمثل موجه عمودي والذي يمكن تمثيلة بالصورة الاتية:

$$Y^* = \begin{bmatrix} \ln \frac{p_1}{1-p_1} \\ \vdots \\ \vdots \\ \ln \frac{p_r}{1-p_r} \end{bmatrix}$$

كما وان المعادلة المذكورة أنفاً لا يمكن تطبيقها عندما تكون $(p=0)$ او $(p=1)$ وذلك لان المقام سوف يساوي (صفرأ) في حالة $p=1$ ومن ثم لا يمكن تطبيق (WLS), لكن الباحثين الأمريكيين (Finney) و (Berkson) قاما بمعالجة هذا الموضوع عن طريق القاعدة الاتية. [34],[41]

$$\left. \begin{array}{l} \text{when } : k_i = 0 \quad , \quad p_i = \frac{1}{2n_i} \\ \text{when } : k_i = n_i \quad , \quad p_i = 1 - \frac{1}{2n_i} \end{array} \right\} \dots \dots (2 - 39)$$

إذ ان:

k_i : يمثل العدد الذي استجاب للتجربة i
 n_i : يمثل العدد الكلي المستعمل في التجربة i

لذا فإ

$$Y_i^* = \ln \left[\frac{p_i = \frac{1}{2n_i}}{1 - \frac{1}{2n_i}} \right] \dots \dots (2-40)$$

$$\text{var}(Y_i)^* = \frac{(n_1 + 1)(n_2 + 2)}{n_i^3 \left(p_i + \frac{1}{n_i}\right) \left(q_i + \frac{1}{n_i}\right)} \dots \dots \dots (2-41)$$

إذ تم الحصول على معادلة (2-41) من المصدر [18] لذا فإن :

$$\text{Var}(Y_i^*) = \frac{1}{w_i^*}$$

أذن :

w_i^* : يمثل الوزن المعدل الجديد

وعلى هذا الأساس يتم حساب الصيغ التقديرية للمعالم على وفق المعادلة الآتية :

$$\hat{B} = (X'W^{*-1}X)^{-1}X'W^{*-1}Y^* \dots \dots (2-42)$$

(7-2-2): طريقة الامكان الاعظم (Maximum Likelihood Method) [16], [41]

تعد طريقة الاكبر الاحتمال من الطرائق الاحصائية الاوسع استعمالا في تقدير نماذج النماذج الرياضية والاحصائية، لما لها من خصائص وميزات المقدر الجيد منها انها تقديرات تسقة واكثر كفاءة من الطرائق الاخرى وايضا لها اقل تباين قدر ممكن [41]، تكمن هذه الطريقة في جعل دالة الاكبر الاحتمال للمتغيرات العشوائية في نهايتها العظمى.

لذا تم استعمال هذه الطريقة لغرض تقدير معالم نموذج الانحدار عند ما يكون المتغير المعتمد الوصفي تعدد الاستجابة، وهذا يعتمد على توزيع تغير الاستجابة، كما و (X_i) تمثل المتغيرات التوضيحية، وأيضا الاخطاء العشوائية التي تفترض انها تتوزع طبيعياً بمتوسط (0) وتباين ثابت قدره (σ^2) ، كما و $\Phi(.)$ التي تمثل دالة التجميعية القياسية (standard Cumulative Distribution) والذي يمكن التعبير عنه بالصيغة الآتية:

$$\Phi(Z) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \dots \dots (2-43)$$

لو رمزنا الى الدرجة المعيارية بـ (Z) والتي يعبر عنها بـ $Z = \left(\frac{X_{ij} - \mu_j}{\sigma} \right)$, $j=1, 2, \dots, \mu$, $(i=1, 2, \dots, n)$, وبالتعويض عن قيمة Z في المعادلة (2-43) سوف نحصل على :

$$P_{ij} = Prob(y \leq Z) = \Phi \left(\frac{X_{ij} - \mu_j}{\sigma} \right) \dots \dots \dots (2 - 44)$$

وعلى فرض بأن :

$$B_j = \frac{1}{\sigma} \quad , \quad A_j = \frac{-\mu_j}{\sigma}$$

لذلك فان المعادلة (2-44) يمكن كتابتها بالشكل الاتي:

$$p_{ij} = prob (y \leq Z) = \Phi(A_j + B_j X_{ij}) \dots \dots \dots (2 - 45)$$

وعند ضرب المعادلة (2- 45) بـ Φ^{-1} نحصل على وحدة الاحتمال بشكل انحدار خطي وكما يأتي :

$$\Phi^{-1} p_{ij} = \Phi^{-1} p(y \leq Z) = A_j + B_j X_{ij} \dots \dots \dots (2 - 46)$$

إذ ان :

A_j, B_j : يمثلان معاملات نموذج الانحدار وحدة الاحتمال (probit)

P_{ij} : يمثل احتمال الاستجابة المصنفة

Φ^{-1} : يمثل معكوس الدالة التجميعية للتوزيع الطبيعي القياسي

وعندما يكون المتغير المعتمد (Y_i) متعدد الاستجابة فإن الدالة التوزيع المشتركة لهذا المتغير سوف تتبع توزيع متعدد الحدود (multinomial Distribution) والذي يعبر عنه بالصيغة الاتية :

$$prob(K_{i1}, K_{i2}, \dots, K_{ij}) = \frac{n_{ij}!}{k_{i1}! k_{i2}! \dots k_{ij}!} p_{i1}^{k_{i1}} \cdot p_{i2}^{k_{i2}} \dots p_{ik}^{k_{ij}} \dots \dots \dots (2 - 47)$$

كما وان

$$k_{i1} + k_{i2} + \dots + k_{ij} = n_{ij} \quad i = 1, 2, \dots, n \quad , j = 1, 2, \dots, \mu$$

$$p_{i1} + p_{i2} + \dots + p_{ij} = 1$$

وبذلك يمكن وصف دالة الامكان الاعظم (like hood) للمعادلة (2-47) بالصيغة الاتية :

$$L = \prod_{i=1}^n \frac{n_i!}{\prod_{j=1}^m k_{ij}!} \cdot \prod_{j=1}^m p_{ij}^{k_{ij}} \quad \dots \quad (2-48)$$

وبإدخال لوغاريتم دالة الامكان للمعادلة المذكورة انفاً والمعرفة بالرقم (2-48) سيكون شكل المعادلة بالصيغة الاتية :

$$\ln L = \sum_{j=1}^m C + \sum_{j=1}^m \sum_{i=1}^n (k_{ij} \ln p_j(X_i) - k_{ij+1} \ln p_{j+1}(X_i)) \quad \dots \quad (2-49)$$

إذ ان :

C : تمثل المقدار الثابت

$$0 \leq p_{ij} \leq 1$$

وعند اجراء عمليات الاشتقاق للمعادلة المعرفة بالرقم (2-49) بالنسبة الى كل من B_j, A_j سوف نحصل على : [28]

$$\left(\frac{\partial \ln L}{\partial A_j} \right) = \sum_{j=1}^m \sum_{i=1}^n \left(\frac{k_{ij}}{p_j(X_i)} - \frac{k_{ij+1}}{p_{j+1}(X_i)} \right) \frac{\partial p_j(X_i)}{\partial A_j} \quad \dots \quad (2-50)$$

$$\left(\frac{\partial \ln L}{\partial B_j} \right) = \sum_{j=1}^m \sum_{i=1}^n \left(\frac{k_{ij}}{p_j(X_i)} - \frac{k_{ij+1}}{p_{j+1}(X_i)} \right) \frac{\partial p_j(X_i)}{\partial B_j} \quad \dots \quad (2-51)$$

ونظراً لأن المعادلتين (2-50) و(2-51) غير خطيتين ولا يمكن حلها بالطرائق الاعتيادية, لذلك لا بد من استعمال طرائق عددية في حل مثل هكذا نماذج, وذلك بهدف ايجاد المقدرات التي تعظم لوغاريتم دالة الامكان الاعظم, ومن ابرز هذه الطرائق هي طريقة نيوتن رافسن.

(7-2-2-1) : طريقة نيوتن رافسن (Newton- Raphson) [49]

وهي طريقة عددية تستعمل في تقدير معالم دالة الامكان, تعتمد على ايجاد المعكوس (inverse) في مصفوفة المعلومات والمعروفة بمصفوفة (Hussein matrix), والتي يمكن الحصول عليها من المشتقة الثانية لدالة لوغاريتم الامكان الاعظم لـ (A_j, B_j) , وايضا تعتمد على متجه المعلمة الابتدائية (initial), ويمكن التعبير عن هذه الطريقة بالصيغة الاتية :

$$\underline{\theta}_{n+1} = \underline{\theta}_n - G^{-1} \underline{g}(n) \quad . \quad . \quad . \quad (2 - 52)$$

إذ أن :

n: تمثل مجموع التكرارات

θ_{n+1} : متجه يمثل المعالم المراد تقديرها

θ_n : متجه يمثل قيم المعالم الاولية (initial)

$g(n)$: متجه يمثل المشتقة الاولى للوغاريتم دالة الامكان الاعظم

$$g(n) = \begin{bmatrix} \frac{\partial \ln l}{\partial A_j} \\ \frac{\partial \ln l}{\partial B_j} \end{bmatrix}$$

G^{-1} : يمثل معكوس مصفوفة المعلومات ليفشر والتي تحتوي على سالب توقع المشتقة الثانية للوغاريتم دالة الامكان وكما هو مبين بالصيغة الاتية :

$$-E \left(\frac{\partial^2 \ln l}{\partial \theta_j^2} \right) = \sum_{j=1}^m \sum_{i=1}^n \left(\frac{n_i}{p_j(X_i) p_{j+1}(X_i)} \left[\frac{\partial p_j(X_i)}{\partial \theta_j} \right]^2 \right) \quad . \quad . \quad . \quad (2 - 53)$$

$$-E \left(\frac{\partial^2 \ln l}{\partial \theta_j \partial \theta_{j+1}} \right) = \sum_{j=1}^m \sum_{i=1}^n \frac{n_i}{p_{j+1}(X_i)} \frac{\partial p_j(X_i)}{\partial \theta_j} \cdot \frac{\partial p_{j+1}(X_i)}{\partial \theta_{j+1}}$$

ومن المعادلتين (2-50) و(2-51) نفرض بأن

$$\theta_j = (A_j, B_j)$$

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

ولذلك يمكن التعبير عن مصفوفة المعلومات (G) والتي تعرف بمصفوفة (Hussein) والتي

تحتوي على المشتقة الثانية للوغاريتم دالة الامكان الاعظم الموجودة في المعادلة (2-53)

ويمكن وصفها بالشكل الآتي :

$$G = \left[\frac{\partial^2 \ln l}{\partial \theta_j^2} \right] = \begin{bmatrix} \frac{\partial^2 \ln l}{\partial A_j \partial A_j} & \dots & \frac{\partial^2 \ln l}{\partial A_j \partial B_j} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln l}{\partial A_j \partial B_j} & \dots & \frac{\partial^2 \ln l}{\partial B_j \partial B_j} \end{bmatrix}$$

$$(m - 1)(m - 1)$$

$$j = 1, 2, \dots, m$$

$$g = \left[\frac{\partial \ln l}{\partial \theta_j} \right] = \begin{bmatrix} \frac{\partial \ln l}{\partial A_1} \\ \frac{\partial \ln l}{\partial A_2} \\ \vdots \\ \frac{\partial \ln l}{\partial A_{m-1}} \\ \frac{\partial \ln l}{\partial B_1} \\ \frac{\partial \ln l}{\partial B_2} \\ \vdots \\ \frac{\partial \ln l}{\partial B_{m-1}} \end{bmatrix} \dots \dots \dots (2 - 54)$$

وعندها تكون صيغة التقديرات بالشكل الآتي :

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \left(\frac{\partial^2 \ln l}{\partial^2 \hat{\theta}_j \partial^2 \hat{\theta}'_j} \right)^{-1} \left(\frac{\partial^2 \ln l}{\partial \hat{\theta}_j} \right) \dots \dots \dots \quad (2 - 55)$$

أما طريقة حساب نيوتن - رافس لإيجاد مقدرات الأماكن الأعظم في حالة نموذج الانحدار اللوجستي فيمكن التعبير عنها بالصيغة الآتية [33]:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + (X'VX)^{-1}X'(Y - \hat{Y}_n) \dots \dots \dots \quad (2 - 56)$$

حيث □

$\hat{\theta}_{n+1}$: يمثل □ وجه عمودي لقيم التقديرات في الدورة (n+1)

$\hat{\theta}_n$: يمثل □ وجه عمودي لقيم التقديرات في الدورة (n)

X : يمثل □ صفوف □ شهادات المتغيرات التوضيحية

V : يمثل □ صفوف قطرية للأوزان □ □ □ □ □ الرتبة (n × n)

و المصفوفة القطرية لـ V يعبر عنها بالمصفوفة الآتية :

$$V = \begin{bmatrix} n_1 \hat{p}_1 \hat{q}_1 & 0 & 0 & 0 & 0 \\ 0 & n_2 \hat{p}_2 \hat{q}_2 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & n_r \hat{p}_r \hat{q}_r \end{bmatrix}$$

ويتم التوقف عن إيجاد التقديرات لمعالم (θ) عندما يكون الفرق بين الدورة السابقة ($\hat{\theta}_{n+1}$) والدورة اللاحقة ($\hat{\theta}_n$) صغيراً جداً ويقترب من الصفر .

(7-2-3) : طريقة تقدير الجاكنيف (Jackknife Estimation Method) [15],[3]

□ المفهوم الاساس لطريقة الجاكنيف يكمن في استبعاد جزء □ ن البيانات عند القيام بأجراء كل عملية تقدير بحيث □ الجزء المستبعد □ ما يهمل عند التقدير او يؤخذ بنظر الاعتبار عند القيام بعملية التقدير عن طريق احتساب هذا الجزء .

إذ استعملت هذه الطريقة الاول □رة لد □ الباحث (Quenouille) وك □ ذلك عام (1949) عند □ا اقترح طريقة لتقليص التحيز والتباين للعديد □ن الاحصاءات عن طريق تقدير ارتباط □ تسلسل يعتمد على تقسيم العينة الى جزأين □ن ثم حساب التقدير لكل جزء كما ويتم القيام بأجراء تقدير آخر اعتماداً على تقديري الجزئين الذين تم حسابهما.

يتم استخراج مقدر (jackknife) كما يأتي:

$$\hat{\theta}_{jackknife} = n\hat{\theta} - (n - 1)\hat{\theta}_* \quad (2 - 57)$$

إذ ان :

$\hat{\theta}$: يمثل مقدر المعلمة حسب الطريقة المعتمدة

n : يمثل حجم العينة

$\hat{\theta}_*$: يمثل مقدر المعلمة الامكان الاعظم

اذ يتم تقدير المعلمة حسب طريقة (jackknife) بالاعتماد على طريقة الامكان الاعظم لتوزيع متعدد الحدود, وذلك من خلال الخطوات الاتية :

$$\hat{\theta}_* = \frac{\sum_{i=1}^n \hat{\theta}_i}{n} \quad (2 - 58)$$

i=1,2, ...n

اما مقدر $\hat{\theta}_i$ يتم حسابه على وفق الخطوات الاتية :

1- ايجاد $\hat{\theta}_1$ وذلك عن طريق حذف المتغير الاول (t_1) من مجموعة المتغيرات (t_1, t_2, \dots, t_n)

ومن ثم يتم ايجاد $\hat{\theta}_1$ بالاعتماد على طريقة الامكان الاعظم بدون المتغير الاول (t_1)

2- ايجاد $\hat{\theta}_2$ يتم حسابه عن طريق ارجاع المتغير الاول (t_1) الى مجموعة المتغيرات

(t_1, t_2, \dots, t_n) وحذف المتغير الثاني (t_2) .

3- نستمر هكذا بإيجاد $(\hat{\theta}_i)$ الى ان نجد $\hat{\theta}_n$.

4- يتم ايجاد $\hat{\theta}_*$ باستعمال المعادلة (2- 58)

5- يتم تطبيق صيغة (jackknife) في المعادلة (2-57) للحصول على المقدر

إذ أن مقدر الامكان الاعظم لمعاملات توزيع متعدد الحدود هو

$$\hat{\theta}_{mle} = \frac{\sum x_i}{n} = \bar{X} \quad . \quad . \quad . \quad (2 - 59)$$

إذ ان :

\bar{X} : يمثل المتوسط الحسابي لمقدر معاملات الامكان الاعظم

الفصل الثالث

المبحث الاول

الجانب التجريبي

(محاكاة)

المقدمة	1-3
مفهوم المحاكاة	1-3-1
توليد الاعداد العشوائية	1-3-2
مراحل بناء تجارب المحاكاة	1-3-3
تحليل نتائج تجارب المحاكاة	1-3-4
احتساب المقاييس الاحصائية المستخدمة في التجربة	1-3-5

1-3 المقدمة Introduction :

يعد المنهج التجريبي (Empirical Approach) من المناهج العلمية والتي لها اساس وجذور في التاريخ الانساني القديم, إذ كانت أغلب الاختراعات والاكتشافات تتم عن طريق اسلوب التجريب وشدة الملاحظة [2].

ونظراً للسرعة الفائقة التي توفرها البرامج الإلكترونية بمختلف انواعها من حزم جاهزة, مما دفع اغلب الباحثين بمختلف تخصصاتهم الى اعتماد اسلوب المحاكاة (Simulation) لغرض تطبيق الطرائق الخاصة بالنموذج المدروس و المتمثل بنموذج الانحدار للمتغير المعتمد متعدد الاستجابة , وذلك من خلال حساب المعلمات المقدرة للطرائق المستعملة والتي تم تناولها في الجانب النظري لمحاكاة اكبر عدد ممكن من الحالات التي تصادفنا في الواقع عند تطبيق البيانات الحقيقية ومن ثم اجراء المقارنة بين هذه الطرائق عن طريق الاعتماد على المؤشر الإحصائي متوسط مربعات الخطأ (MSE) (mean square Error) .

ويشمل هذا الفصل عرض بعض المفاهيم العامة للمحاكاة وكذلك مراحل وصف التجربة للمحاكاة من حيث حجوم العينات والقيم المفترضة والطرائق التقدير المستعملة والمؤشرات الاحصائية المستعملة في تجربة المحاكاة .

1-3-1 مفهوم المحاكاة :

تعرف المحاكاة بأنها طريقة تحليلية عددية علمية تحاول استعمال مناهج وأساليب رياضية منهجية والغرض من ذلك ايجاد صورة طبق الاصل من أي نموذج من دون الرجوع إلى أخذ ذلك النموذج؛ ويتم ذلك عن طريق البرامج الإلكترونية وتشبيه ذلك النموذج في ظروف عدم التأكد, كما وان اسلوب المحاكاة يستخدم عادة لوصف سلوك نظام حركي معين عن طريق تطبيق تجارب تكون مماثلة وملائمة و مقارنة للنموذج الحقيقي والواقعي الموجود اصلاً [22].

ومع انتشار الواسع للبرامج والحاسبات الكترونية عن طريق انتاج انظمة وحزم جاهزة خاصة ببرنامج المحاكاة ومنها نظام المحاكاة العام (General Purpose Simulation System) ويشار له (GPSS) والذي يطبق على الاغلب مع المتغيرات المنفصلة (Discreet variable)

ولنظام المحاكاة طرائق متعددة منها طريقة مونت كارلو (Monte Carlo method) والتي تعد الطريقة الرائدة من طرائق المحاكاة المستعملة في الوقت الحاضر, ومونت كارلو هو مصطلح سري تعود جذوره الى العالمين الشهيرين وهما عالم الفيزياء جون نيومان (Neman) وعالم الرياضيات ألم (A. Elam) وذلك عن طريق بحوثهم عن دراستهم حول (سلوك النيوترونات في المعادن) وذلك خلال فترة الحرب العالمية الثانية .

إذ يتم تطبيق اسلوب المحاكاة بواسطة العينة , وعندها تؤخذ هذه العينة من مجتمع افتراضي عوضاً من ان يتم أخذها من مجتمع حقيقي, وعند ذلك يتم تطبيق النماذج والاساليب الرياضية والاحصائية المعتمدة لغرض اجراء المقارنة بغية الحصول على النتائج المرجوة عندها يتم اتخاذ

القرار المناسب على ضوء هذه النتائج, وكلما كانت نتائج تجربة المحاكاة دقيقة و واقعية كلما كانت التجربة اقرب للحقيقة واقرب تمثيل للواقع التطبيقي .

على الرغم من النتائج الايجابية والفوائد العظيمة التي يحققها اسلوب المحاكاة باستعمال طريقة مونت كارلو من توفير الوقت والجهد على الباحث, الا انها قد يترتب على الباحث ان يراعي المقولة المأثورة (Garbage in Garbage out) والتي تعني ان نتائج الحاسوب الالكتروني غير جازمة بصحتها او جعلها مشابهة للواقع الحقيقي وذلك لان الاصل هو توفير الشروط النظرية والمنطقية لتصميم برنامج المحاكاة للوصول الى النتائج [2] .

2-3-1 توليد الاعداد العشوائية :

تُعد عملية توليد الاعداد العشوائية عملية تكرارية موضوعة في معادلة خاصة ليعطي في كل مرة عدداً عشوائياً مختلفاً عن العدد الاخر و محدداً بالقيم (0,1), ويتم ذلك من خلال توليد ارقام عشوائية تتبع التوزيع المنتظم المستمر (Continues uniform distribution) ضمن الفترة (0,1)

كما ان الغاية من توليد الاعداد العشوائية هو لغرض الحصول على متغيرات عشوائية تمثل الانموذج قيد الدراسة من خلال التوزيع الاحصائي المستخدم في الدراسة, وتنشطر عملية توليد المتغيرات العشوائية الى شطرين من المتغيرات وهي :

1-3-2-1 توليد المتغيرات المستقلة :

يتم توليد ارقام عشوائية تتبع توزيع المنتظم (uniform distribution) ضمن الفترة (0,1) , ويتم ذلك من خلال الصيغة الاتية :

$$f(x) = \left[\frac{1}{b-a} \right] \quad a \leq x \leq b \quad \dots \dots \dots \quad (3-1)$$

يتم بناء ثلاث نماذج من الارقام العشوائية الخاصة بالمتغير المستقل كل نموذج يخضع لتوزيع ثنائي الحدين (Binomial Distribution) ويمكن كتابة الدالة التجميعية (c d f) لصيغة توزيع ثنائي الحدين للتعبير عن النماذج الثلاثة بالمعادلة الاتية :

$$F(x) = \sum_{i=0}^y C_X^n P^x (1-P)^{n-x} \quad \dots \dots \dots \quad (3-2)$$

$$x=0, 1, 2, \dots, n$$

إذ ان :

n : يمثل عدد المحاولات المستعملة في التجربة

p : يمثل احتمال النجاح

والغرض من هذا البناء هو لحصول المتغير المستقل (x) على استجابة في كل محاولة من المحاولات (Binomial) ولكل نموذج, وهذا بدوره سوف يعطي المتغير المستقل (x) استجابة للمتغير المعتمد (y) وكل استجابة محددة ضمن الفترة (0,1) , عندها سيتم الحصول على العديد من الاستجابات للمتغير المعتمد (y) .

1-3-2-2 توزيع المتغير المعتمد (الاستجابة) :

ان معادلة المتغير المعتمد متعدد الاستجابة تتبع توزيع متعدد الحدود ويمكن وصفه بالمعادلة (2-2) التي تم ذكرها مسبقاً :

$$prob(K_{i1}, K_{i2}, \dots, K_{ij}) = \frac{n_{ij}!}{k_{i1}! k_{i2}! \dots k_{ij}!} p_{i1}^{k_{i1}} \cdot p_{i2}^{k_{i2}} \dots p_{ij}^{k_{ij}}$$

يتم توليد متغير عشوائي (y) يخضع على وفق توزيع متعدد الحدود (multinomial Distribution), ويتم ذلك من خلال بناء ثلاث نماذج افتراضية وتمثلة بالنموذج التالي :

$$\hat{Y}_i = B_0 + B_i X_i \quad . . . \quad (3-3)$$

$$i = 1, 2, 3$$

ومن خلال كل نموذج يتم الحصول على استجابة للمتغير (y) فعندما يتم تطبيق ثلاث نماذج فسوف يتم الحصول على ثلاث استجابات للمتغير (y), وكل استجابة تكون محددة بالفترة (0,1) وذلك لان هذه الفترة يمكن ان تتضمن بداخلها اعداد ما لانهاية من الاستجابات .

1- 3-3 : مراحل بناء ووصف تجارب المحاكاة :

من خلال استخدام اسلوب المحاكاة لا بد من تحديد أهم العوامل الخاصة لمرحل وصف وبناء تجربة المحاكاة وذلك لغرض تحليل البيانات ومنها :

المرحلة الاولى : وهي من اهم المراحل التي يمكن الاعتماد عليها في المراحل اللاحقة من التجربة وتتضمن هذه المرحلة وكما يأتي :

1- تحديد حجوم العينات حيث تم تحديد اربعة حجوم للعينات (صغيرة , متوسطة , وكبيرة) وهي (200 , 100 , 50 , 20) على التوالي وبتكرار التجربة في كل حجم عينة بمقدار (R= 1000).

2 - تحديد القيم الافتراضية فقد تم اختيار قيم للمعالم الافتراضية ولثلاث استجابات لـ (y) والتي تخضع لتوزيع متعدد الحدود كما تم ذكرها في المرحلة السابقة, ويمكن توضيح القيم المفترضة في الجدول (3- 1) :

جدول (3- 1)

يوضح قيم المعالم للنماذج المفترضة

Model	B_0	B_1	B_2
1	0.5	-0.4	0.3
2	0.4	0.5	0.4
3	0.3	0.6	-0.2

المرحلة الثانية : تم توليد متغيرات عشوائية للمتغيرات المستقلة تتبع توزيعاً ثنائي الحدين (binomial) ولثلاث نماذج افتراضية لتعطي في كل نموذج استجابة للمتغير المعتمد (y) والذي يخضع لتوزيع متعدد الحدود وبتكرار التجربة (R=1000).

المرحلة الثالثة : يتم في هذه المرحلة حساب تقديرات المعالم للبيانات المولدة لطرائق التقدير المستعملة للمتغير المعتمد متعدد الاستجابة ومن هذه الطرائق التي تم استعمالها في التقدير هي طريقة المربعات الصغرى الموزونة (W LS) وطريقة الامكان الاعظم (M LE) و كذلك تم استعمال طريقة الجاكناييف (jack) .

المرحلة الرابعة : وهي المرحلة النهائية من مراحل وصف تجربة المحاكاة, تضمن هذه المرحلة اهم المقاييس والمؤشرات الاحصائية المستعملة لغرض الحصول على افضل الطرائق المستخدمة في الانموذج الاحتمالي للمتغير المعتمد متعدد الاستجابة ومن هذه المؤشرات:

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

$$MSE = \frac{\sum_{i=1}^r (\hat{B}_i - B_i)^2}{R} \quad \dots \dots \dots \quad (3 - 4)$$

إذ أن :

\hat{B}_i : يشير الى المعالم المقدرة عن طريق اسلوب المحاكاة

B_i : يشير الى القيم الحقيقية للمعالم

R : يشير الى عدد تكرار التجربة

كما تم حساب (MSE) للأنموذج على وفق الصيغة الاتية لبيان افضلية الطريقة المستعملة في النماذج والتي تعد الطريقة المفضلة لتمثيل البيانات .

$$MSE = \frac{\sum_{i=1}^r (Y_i - \hat{Y}_i)^2}{n - K} \quad (3 - 5)$$

أذ ان :

K : يمثل عدد معالم النموذج.

2 - تم حساب معامل التحديد (R^2) للأنموذج وحسب الصيغة الاتية :

$$R^2 = 1 - \frac{\sum_{i=1}^r (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^r (Y_i - \bar{Y}_i)^2} \quad (3 - 6)$$

3 - التحيز (Bias) المطلق للمعالم المقدرة (B_0, B_1, B_2) وحسب الصيغة الاتية :

$$Bais = |\hat{B}_i - B| \quad (3 - 7)$$

وكما هو موضح من الصيغة المذكورة انفاً بأن التحيز يمثل الفرق المطلق ما بين القيمة التقديرية (\hat{B}) وبين القيمة الحقيقية (B) .

1-3-4 : تحليل نتائج تجارب المحاكاة :

في هذه المرحلة سيتم عرض نتائج المحاكاة وتحليلها بالنسبة للقيم التقديرية ومتوسطات مربعات الخطأ بالطرائق المستعملة كافة, إذ سيتم في هذا المبحث استعراض النتائج لثلاث استجابات وهي (الاستجابة الاولى, الاستجابة الثانية, الاستجابة الثالثة), إذ تم الحصول على النتائج عن طريق التطبيق بالبرنامج الاحصائي الذي تمت كتابته بلغة (Mat lab 2011) والذي يتميز بقدرته الهائلة والفائقة في الجوانب الرياضية والاحصائية مقارنة مع بقية تطبيقات البرامج الاخرى.

سيتم دراسة العلاقة بين نتائج مقياس متوسط مربعات الخطأ (MSE) والتحيز مع حجوم العينات, واتضح عن طريق ما تقدم بأن اغلب الاساليب المدروسة تعطي نتائج جيدة عند حجوم العينات الكبيرة .

1-3-5 احتساب قيم المقاييس الاحصائية المستعملة للاستجابة الاولى والثانية والثالثة :

اولاً: تم تقدير معالم الانحدار المتعدد ولجميع الطرائق المستعملة (WLS , MLE , jack) للاستجابة الاولى والثانية والثالثة , وكما هو مبين في الجداول (2-3) , (3-5) , (3-8) نلاحظ الاتي :

1 - يتضح لنا بأن نتائج متوسط مربعات الخطأ (MSE) والتحيز لتقدير المعالم والأنموذج يتناقص مع كبر حجم العينة ولأغلب الطرائق المدروسة , وعندها يصبح (MSE) اقل ما يمكن وهذا بدوره يجعل التباين والتحيز اقل ما يمكن وكما هو معروف في النظرية الاحصائية بأن :

$$MSE = V(Y_i) + B^2 \quad . . . \quad (3 - 8)$$

2- نلاحظ بان قيم المطلقة للتحيز للقيم التقديرية للاستجابة الاولى او الاستجابة الثانية او للاستجابة الثالثة, حيث نلاحظ بأن قيمة العينة الصغيرة (n=20) تبدأ بتحيز عالٍ ولكن عندما يكبر حجم العينة (n=200) تبدأ قيمة التحيز بالانخفاض, مثلاً في الاستجابة الاولى كانت كان اعلى قيمة تحيز في العينة صغيرة (n=20) هي واقعة في طريقة (WLS) وكان مقدارها هو (b₂= 0.8383) ولكن عند زيادة حجم العينة في حالة (n=200) اصبحت قيمة التحيز اقل ما يمكن مقدارها (b₂=0.1607) وهذا يتفق مع النظرية الاحصائية .

3 - ان لحجم العينة تأثيراً واضحاً في معامل التحديد (R²) نلاحظ في العينات الصغيرة يكون معامل التحديد قليل و لكن مع زيادة حجم العينة واقتراب معامل التحديد من الواحد يصبح معامل التحديد كبير مما يدل ذلك على افضلية الطريقة المستعملة .

ثانياً - متوسط مربعات الخطأ (MSE):

استناداً الى النتائج أتضح أن لحجم العينة تأثيراً واضحاً في قيم متوسط مربعات الخطأ (MSE) إذ ان هذه القيم تتناقص كلما ازداد حجم العينة, وعند المقارنة بين حجوم العينات للاستجابة الاولى وتحديدأ في الجدول(3-3) لوحظ ما يأتي:

أ - في العينات الصغيرة (n=20) نلاحظ بأن طريقة الامكان الاعظم (MLE) هي الافضل إذ حصلت على اقل (MSE) مقداره (0.6860) عندها حصل على معامل التحديد مقداره (0.71).

ب - في العينات المتوسطة (n=50) كانت الطريقة الافضل هي طريقة الجاكانايف (Jack) وذلك لامتلاكها اقل (MSE) مقداره (0.5750) في حين تم الحصول على معامل التحديد (R²) مقداره (0.8530) .

ج - في العينات الكبيرة ($n=100, 200$) نلاحظ أن طريقة الامكان الاعظم (MLE) هي الافضل في حجوم العينات الكبيرة إذ حصلت على اقل (MSE) في العينتين مقداره ($0.5082, 0.4841$) على الترتيب وتم الحصول على اعلى معامل التحديد للعينتين الكبيرتين مقداره ($0.92, 0.96$) على التوالي وهذا بدوره ينسجم مع النظرية الاحصائية القائلة كلما تزداد حجم العينة كلما يقل (MSE) ويكبر قيمة معامل التحديد (R^2).

كما ويتم تحليل قيم متوسط مربعات الخطأ (MSE) للاستجابة الثانية والثالثة كما هو مبين في الجدول (6-3) و (9-3) وعلى وفق الشرح المذكورة أنفاً .

أ - في العينات الصغيرة ($n=20$) كانت طريقة الامكان الاعظم (MLE) هي الافضل إذ حصلت على اقل (MSE) مقداره (0.7970) للاستجابة الثانية عندها تم الحصول على معامل التحديد في مقداره (0.79), بينما في الاستجابة الثالثة فقد كانت طريقة الجاكنيف هي الافضل في حجم ($n=20$) حيث حصلت على اقل (MSE) مقداره (0.7932) إذ تم الحصول على معامل التحديد مقداره (0.79).

ب- في العينات المتوسطة ($n=50$) كانت طريقة الجاكنيف هي الافضل للاستجابة الثانية إذ حصلت على اقل (MSE) وكان معامل التحديد مقداره (0.65), بينما في الاستجابة الثالثة كانت ايضا طريقة الجاكنيف هي الافضل إذ حصلت على اقل (MSE) مقداره (0.6601) في حين كان معامل التحديد له مقداره (0.78).

ج - في العينة الكبيرة ($n=100$) كانت طريقة الامكان الاعظم للاستجابة الثانية هي الافضل حيث حصلت على اقل (MSE) مقداره (0.5942) في حين كان معامل التحديد له مقداره (0.49), بينما في الاستجابة الثالثة ولحجم العينة ($n=100$) كانت طريقة الجاكنيف هي الافضل إذ حصلت على اقل (MSE) مقداره (0.5862) في حين كانت له قيمة معامل التحديد له عالية مقدارها (0.92) وقريبة من الواحد الصحيح.

د - تبين عند حجم العينة ($n=200$) للاستجابتين الثانية والثالثة بأن طريقة الجاكنيف هي الافضل في حجوم العينات الكبيرة إذ حصلت على اقل (MSE) للاستجابتين, في حين كان معامل التحديد لهما كبيراً جداً ويقترب من الواحد الصحيح وحسب الفرضية الاحصائية المعروفة القائلة بأن قيم متوسطات مربعات الخطأ تصبح اقل ما يمكن عندما يكبر حجم العينة مع ارتفاع في قيمة معامل التحديد (R^2).

جدول (2-3) :

يبين القيم التقديرية (\hat{B}) للاستجابة الاولى لجميع الطرائق المستعملة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار ($R= 1000$)

حجم العينة (n)	B	Bwls	Bmle	Bjak
20	0.5000	1.1422	0.7820	0.7333
	-0.4000	0.0390	-0.7331	-0.6753
	0.3000	-0.4708	0.1190	0.1555
50	0.5000	0.5625	0.4979	0.5375
	-0.4000	-0.0189	-0.2987	-0.4659
	0.3000	0.8290	0.0646	0.2857
100	0.5000	1.1774	0.4710	0.4981
	-0.4000	0.1772	-0.4593	-0.3928
	0.3000	-0.5383	0.3421	0.3042
200	0.5000	0.8955	0.5250	0.4998
	-0.4000	-0.0256	-0.4893	-0.4007
	0.3000	0.1393	0.3056	0.3029

جدول (3-3) :

يبين قيم متوسطات مربعات الخطأ (MSE) للاستجابة الاولى للمعلمات المقدرة والانموذج ولجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R= 1000).

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b_0	0.4589	0.1958	0.0654
	b_1	0.3332	0.3115	0.0907
	b_2	0.7028	0.1847	0.0327
	MSE	0.7080	0.6880	0.6940
	R^2	0.5636	0.7111	0.7352
50	b_0	0.4124	0.0211	0.0065
	b_1	0.1928	0.0545	0.0149
	b_2	0.5941	0.0908	0.0076
	MSE	0.5845	0.5855	0.5750
	R^2	0.6381	0.8530	0.8540
100	b_0	0.1564	0.0057	0.0019
	b_1	0.1452	0.0106	0.0036
	b_2	0.2798	0.0120	0.0041
	MSE	0.5113	0.5082	0.5160
	R^2	0.6713	0.9222	0.9217
200	b_0	0.0039	0.0018	0.0009
	b_1	0.1402	0.0099	0.0016
	b_2	0.0258	0.0020	0.0013
	MSE	0.4865	0.4841	0.4948
	R^2	0.7915	0.9610	0.9600

جدول (4-3):

يبين قيم التحيز (Bias) المطلقة للاستجابة الاولى للمعالم المقدرة لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R= 1000)

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b_0	0.6774	0.2820	0.2333
	b_1	0.5772	0.3331	0.2753
	b_2	0.8383	0.2354	0.1445
50	b_0	0.6422	0.0290	0.0375
	b_1	0.4390	0.1013	0.0659
	b_2	0.7708	0.1810	0.0143
100	b_0	0.3955	0.0250	0.0019
	b_1	0.3811	0.0893	0.0072
	b_2	0.5290	0.0421	0.0042
200	b_0	0.0625	0.0021	0.0002
	b_1	0.3744	0.0593	0.0007
	b_2	0.1607	0.0056	0.0029

جدول (5-3) :

يبين قيم المقدرات التقديرية (\hat{B}) للاستجابة الثانية لجميع الطرائق المستعملة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار ($R=1000$).

حجم العينة	B	Bwls	Bmle	Bjak
20	0.4000	0.5059	0.5966	0.4093
	0.5000	1.0556	1.5733	0.8542
	0.4000	0.4737	-1.2391	0.0270
50	0.4000	0.9537	0.2964	0.3996
	0.5000	-0.1489	0.7081	0.4865
	0.4000	0.9463	0.4715	0.3903
100	0.4000	1.1114	0.4414	0.3995
	0.5000	0.6757	0.5167	0.5026
	0.4000	-0.3494	0.2964	0.4046
200	0.4000	0.8633	0.4042	0.4044
	0.5000	0.8807	0.4668	0.4965
	0.4000	-0.1905	0.4055	0.3967

جدول (6-3):

يبين قيم متوسطات مربعات الخطأ (MSE) للاستجابة الثانية للمعالم المقدرة والانموذج لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R=1000) .

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b_0	0.3066	0.1867	0.0096
	b_1	0.4210	0.3542	0.1574
	b_2	0.2984	0.2143	0.1966
	MSE	0.8030	0.7970	0.8060
	R^2	0.7512	0.7993	0.7866
50	b_0	0.2146	0.0278	0.0044
	b_1	0.3450	0.0791	0.0077
	b_2	0.2487	0.0409	0.0074
	MSE	0.6595	0.6600	0.6510
	R^2	0.8414	0.9147	0.9150
100	b_0	0.0112	0.0072	0.0025
	b_1	0.3087	0.0089	0.0036
	b_2	0.0054	0.0200	0.0041
	MSE	0.5947	0.5942	0.5962
	R^2	0.8931	0.9493	0.9492
200	b_0	0.0061	0.0010	0.0009
	b_1	0.0309	0.0032	0.0018
	b_2	0.0016	0.0022	0.0019
	MSE	0.5661	0.5578	0.5562
	R^2	0.9242	0.9749	0.9750

جدول (3-7):

يبين قيم التحيز (Bias) المطلقة للاستجابة الثانية للمعالم المقدرة لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R= 1000) .

حجوم العينة	معلمت المقدرة	Bwls	Bmle	Bjak
20	b0	0.5537	0.1966	0.0093
	b1	0.6489	1.0733	0.3542
	b2	0.5463	1.6391	0.3730
50	b0	0.4633	0.1036	0.0044
	b1	0.3807	0.2081	0.0035
	b2	0.5905	0.1036	0.0033
100	b0	0.1059	0.0414	0.0005
	b1	0.5556	0.0167	0.0135
	b2	0.0737	0.0715	0.0097
200	b0	0.0114	0.0042	0.0004
	b1	0.1757	0.0032	0.0026
	b2	0.0494	0.0055	0.0046

جدول (3 - 8):

يبين قيم المقدرات التقديرية (\hat{B}) للاستجابة الثالثة لجميع الطرائق المستخدمة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار ($R= 1000$).

حجم العينة (n)		B	Bwls	Bmle	Bjak
20	b0	0.3000	0.3801	-0.1974	-0.0532
	b1	0.6000	1.0705	1.3582	1.0709
	b2	-0.2000	0.1939	0.3365	0.1138
50	b0	0.3000	1.2264	0.1286	0.2623
	b1	0.6000	0.4919	0.8379	0.6219
	b ₂	-0.2000	-1.0895	-0.2657	-0.1596
100	b ₀	0.3000	1.1100	0.2500	0.3084
	b ₁	0.6000	-0.0459	0.6435	0.5977
	b ₂	-0.2000	-0.3271	-0.2915	-0.2141
200	b0	0.3000	1.1644	0.2887	0.2983
	b1	0.6000	0.3273	0.5863	0.6037
	b2	-0.2000	-0.7203	-0.1990	-0.1988

جدول (3- 9) :

يبين قيم متوسطات مربعات الخطأ (MSE) المطلقة للاستجابة الثالثة للمعالم المقدرة والانموذج لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R=1000).

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b_0	0.8583	0.4156	0.1398
	b_1	0.4171	0.7671	0.2385
	b_2	0.7913	0.4584	0.1186
	MSE	0.8004	0.7991	0.7932
	R^2	0.7868	0.7905	0.7920
50	b_0	0.6561	0.0613	0.0096
	b_1	0.2213	0.0904	0.0078
	b_2	0.1552	0.0529	0.0146
	MSE	0.6699	0.6657	0.6601
	R^2	0.7967	0.8700	0.8713
100	b_0	0.6472	0.0099	0.0028
	b_1	0.0744	0.0118	0.0039
	b_2	0.0707	0.0182	0.0047
	MSE	0.5916	0.5945	0.5862
	R^2	0.7077	0.9266	0.9277
200	b_0	0.0064	0.0011	0.0007
	b_1	0.0117	0.0020	0.0015
	b_2	0.0161	0.0017	0.0014
	MSE	0.5609	0.5581	0.5557
	R^2	0.8339	0.9628	0.9630

جدول (3- 10) :

يبين قيم التحيز (Bias) المطلقة للاستجابة الثالثة للمعالم المقدرة لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R= 1000) .

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b_0	0.9264	0.4974	0.3532
	b_1	0.4705	0.7582	0.4709
	b_2	0.8895	0.5365	0.3138
50	b_0	0.8100	0.1714	0.0377
	b_1	0.6459	0.2379	0.0219
	b_2	0.3939	0.0657	0.0404
100	b_0	0.8044	0.0500	0.0084
	b_1	0.2727	0.0435	0.0023
	b_2	0.2203	0.0515	0.0141
200	b_0	0.0801	0.0113	0.0017
	b_1	0.1081	0.0137	0.0017
	b_2	0.1271	0.0010	0.0012

3-2 الجانب التطبيقي

ولغرض الوصول الى حقيقة التقدير ودقته تم اعتماد احد التجارب الحياتية كجانب تطبيقي

3-1-2 بيانات التجربة

تم سحب عينة عشوائية حجمها (150) ولكل تركيز, إذ تم الحصول على البيانات التي تمثل الاستجابة الحقيقية بمساعدة الباحثة (عذراء عقيل الكريطي)^[21] والتي نفذت في مختبر الدراسات العليا قسم وقاية النبات - كلية الزراعة - جامعة كربلاء خلال الفترة الزمنية (2014-2015), التجربة المستخدمة كانت تأثير نوعين من المبيدات الاحيائية لمكافحة بعض المواد المبيدة لحشرة غدوق النخيل (*Oryctes elegans*). إذ تعرضت هذه الحشرة الى تراكيز مختلفة من المبيدات الاحيائية احدهما المستحضر الفطر (*Beauveria bassiana*) نرمل له (D_1) تم الحصول عليه من مركز الزراعة العضوية / وزارة الزراعة وتم الاعتماد على تراكيز الفعالة والتي تم التوصل اليها مختبرياً وهي غم / 100 ملم (0.3, 0.4, 0.5, 0.6) على الترتيب, والمستحضر الثاني هو (*Biocont -T*) نرمل له (D_2) وهو من انتاج شركة الرؤيا السعودية والمادة الفعالة له هي الفطر *Trichoderma harzianum* تم الحصول عليه من مديرية زراعة كربلاء, وتم الاعتماد على تراكيز مختلفة غم / 100 ملم (0.05, 0.1, 0.2, 0.3) على الترتيب.

وبعد مزج هاتين المادتين مستحضر الفطر (*B. bassiana*) والمبيد الاحيائي (*Biocont T*) واجراء عدد من الرشوات بفترات زمنية مختلفة على الحشرة تم الحصول على اعلى نسبة قتل والتي تمثل (الاستجابة الحقيقية) في اليوم الثامن من بدء اجراء عملية الرش ومن ثم حساب توقع الحشرات المقتولة والتي تمثل (الاستجابة المتوقعة), كما وان (K) يشير الى عدد الحشرات المقتولة من مجموعة (n) و الجدول (3-11) يبين عدد الحشرات المقتولة والمتوقعة المقترنة بكل تركيز

جدول (3-11) يمثل الاستجابة الحقيقية والمتوقعة والمرتبطة بكل تركيز

D1 \ D2	0.05	0.1	0.2	0.3	Total	العدد الكلي المستخدم لكل تجربة (n)
0.3	6 (12.3)	6 (12.3)	30 (26.1)	42 (33.3)	84	150
0.4	14 (16.3)	14 (16.3)	36 (34.6)	48 (44.4)	112	150
0.5	23 (22.1)	23 (22.1)	47 (47.3)	59 (60.31)	152	150
0.6	39 (31.2)	39 (31.2)	62 (66.6)	74 (84.9)	214	150
Total	82	82	175	223	562	

لبيان معرفة سلوك بيانات التجربة هل تخضع لتوزيع ثنائي الحدين أم لا, تم معالجة هذا الامر عن طريق تطبيق اختبار (Chi -Square) لحسن المطابقة عن طريق اختبار الفرضية الاحصائية الاتية :

$H_0 =$ البيانات تتبع توزيع ثنائي الحدين

$H_1 =$ البيانات لا تتبع توزيع ثنائي الحدين

وبعد تطبيق اختبار مربع كاي لبيرسون كانت القيمة المحسوبة له هي (16.03) أما قيمة مربع كاي الجدولية ولمستوى معنوية (0.05 , 0.01) وبدرجة حرية (9) هي (21.666, 16.919) على الترتيب , لذا فإن البيانات الحقيقية ستخضع لتوزيع ثنائي الحدين, وعندها يمكن تطبيق نموذج الانحدار الوصفي على بيانات التجربة .

2- 3 - 2 : تحليل بيانات التجربة

التجربة المستعملة تحتوي على نوعين من المبيدات الاحيائية والتي تم التعبير عنهما بـ (D_1, D_2) والتي أعطيت على شكل رشات او على شكل جرعات (dose), كما وتم الحصول على متغيرات توضيحية (X_1 , X_2) عن طريق أخذ لوغاريتم طاقة التحمل لتراكيز الجرعات (Tolerance) كما في الصورة الاتية :

$$X_1 = \ln D_1$$

$$X_2 = \ln D_2$$

والجدول (12-3) والمتضمن قيم X_1 X_2 وقيم الاستجابة الحقيقية وقيم الاحتمالات المقابلة للمتغيرات التوضيحية ويمكن عرضه بالاتي :

جدول (3- 12) يبين المتغيرات التوضيحية ($\ln D_1$, $\ln D_2$) والاحتمالات المقابلة لهما فضلاً عن قيم probit الحقيقية (y)

$\ln D_1$	$\ln D_2$	k_i	n_i	P_i	القيمة التخمينية (probit)
-1.204	-2.996	6	150	0.04	3.25
-0.916	-2.966	14	150	0.09	3.66
-0.693	-2.996	23	150	0.15	3.96
-0.511	-2.996	39	150	0.26	4.36
-1.204	-2.303	6	150	0.04	3.25
-0.916	-2.303	14	150	0.09	3.66
-0.693	-2.303	23	150	0.15	3.96
-0.511	-2.303	39	150	0.26	4.36
-1.204	-1.609	30	150	0.2	3.92
-0.916	-1.609	36	150	0.24	4.29
-0.693	-1.609	47	150	0.31	4.5
-0.511	-1.609	62	150	0.41	4.77
-1.204	-1.204	42	150	0.28	4.44
-0.916	-1.204	48	150	0.32	4.53
-0.693	-1.204	59	150	0.39	4.72
-0.511	-1.204	74	150	0.52	5

تم الحصول على قيمة الاحتمال الخاصة للاستجابات الحقيقية عن طريق معادلة (2-1) والتي تم ذكرها في الجانب النظري والتي يعبر عنها :

$$\left[p_i = \frac{K_i}{n_i} \right]$$

إذ ان :

X_1, X_2 : تمثل المتغيرات التوضيحية والناجمة من اخذ اللوغاريتم لكل من التركيزين ($\ln D_1, \ln D_2$)

K_i : يمثل الحشرات المقتولة

n_i : يمثل العدد الكلي للحشرات المستعملة في التجربة

كما وان قيم (*probit*) تم الحصول عليها عن طريق جداول خاصة وضعها العالم (*Finney*)^[41] (والموجود في ملحق الرسالة), وقد تم اضافة رقم (5) الى قيم الاحتمال لتخلص من القيم السالبة وبذلك يمكن الحصول على قيم *probit* بتطبيق معادلة (2-31) الموجودة في الجانب النظري .

2-3-3 نتائج التجربة :

يتم الحصول على تقديرات معالم النموذج للاستجابة الحقيقية الخاصة بقيم وحدة الاحتمال (*probit*) تم الحصول على نتائج تقدير الخاصة من تطبيق بيانات حقيقية باستعمال طريقة المربعات الصغرى (OLS) ومن خلال البرنامج الاحصائي (mat lab 2011) حصلنا على:

$$\hat{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.2712 \\ 0.5968 \\ -0.2079 \end{pmatrix}$$

وبذلك فان معادلة الانحدار ستكون :

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

$$\hat{y} = 0.2712 + (-0.327) + 0.623 = 0.567$$

2-3-3-1 نتائج التجربة للاستجابات المتوقعة الخاصة بقيم وحدة الاحتمال (*probit*) المتوقعة بالنسبة لطرائق التقدير المستعملة :

ولاستعمال القيم المتوقعة في جدول (11-3) لغرض الحصول على القيم التخمينية (*Probit*) إذ يتم تحديد قيمة وحدة الاحتمال (*probit*) (\hat{y}) من الجداول *Finney* (تم وضعها في الملحق) بافتراض ان (K^*) يمثل الحشرات المتوقعة المقتولة وباحتمال استجابة متوقعة (π_i) , وتم

إضافة (5) الى قيمة الاحتمال لغرض التخلص من قيمة السالب , وضعت النتائج التي تم الحصول عليها في جدول (13-3).

جدول (13-3) : يوضح نتائج احتمال الاستجابة المتوقعة وقيم وحدة الاحتمال (probit) (\tilde{y}) المتوقعة

$\ln D_1$	$\ln D_2$	K^*_i	n_i	π_i	القيمة التخمينية probit \downarrow (\tilde{y})
-1.204	-2.996	12.3	150	0.08	3.59
-0.916	-2.996	16.3	150	0.11	3.77
-0.693	-2.996	22.1	150	0.15	3.96
-0.511	-2.996	31.2	150	0.20	4.16
-1.204	-2.303	12.3	150	0.08	3.59
-0.916	-2.303	16.3	150	0.11	3.77
-0.693	-2.303	22.1	150	0.15	3.96
-0.511	-2.303	31.2	150	0.20	4.16
-1.204	-1.609	26.1	150	0.17	4.05
-0.916	-1.609	34.6	150	0.23	4.26
-0.693	-1.609	47.3	150	0.31	4.5
-0.511	-1.609	66.6	150	0.44	4.85
-1.204	-1.204	33.3	150	0.22	4.23
-0.916	-1.204	44.4	150	0.21	4.19
-0.693	-1.204	60.31	150	0.40	4.75
-0.511	-1.204	84.9	150	0.57	5.18

إذ ان :

X_1, X_2 : تمثل المتغيرات التوضيحية والنااتجة من اخذ اللوغاريتم لكل من التركيزين
($\ln D_1, \ln D_2$)

(k_i^*): تمثل عدد الحشرات المقتولة المتوقعة

n_i : العدد الكلي للحشرات المستخدمة في التجربة

π_i : احتمال الاستجابة المتوقعة

وعند تطبيق الطرائق المستعملة ($WLS, MLE, jack$) لتقدير المعلمات وفقاً للمعادلات
الآتية (2-57, 2-56, 2-42) على الترتيب والتي تم تناولها في الجانب النظري , وباستعمال

طريقة المربعات الصغرى تم الحصول على النتائج الآتية وعرضها في الجدول (14-3)

**جدول (14-3) : يمثل القيم التقديرية لمعالم الانحدار للاستجابة المتوقعة لوحدة
الاحتمال (probit) وللطرائق المستخدمة كافة**

المعالم المقدره	WLS	MLE	Jack
b_0	1.1675	0.2527	0.2993
b_1	0.1612	0.6136	0.6000
b_2	- 0.2453	- 0.1322	- 0.1971
R^2	0.8571	0.9610	0.9600
MSE	0.6822	0.4841	0.4984

تبين النتائج المذكورة انفاً بأن طريقة الامكان الاعظم (MLE) هي الافضل والامثل لقيم
الاستجابات المتوقعة لوحدة الاحتمال (Probit) إذ حصلت على اقل متوسط مربعات الخطأ
(MSE) والبالغ مقداره (0.4841) مع ارتفاع واضح في قيمة معامل التحديد (R^2) والبالغ
مقداره (0.9610) وهذا يتماشى مع الفرضية الاحصائية , يليها طريقة الجاكناف إذ حصلت
على متوسط مربعات الخطأ مقداره (0.4984) ومعامل التحديد مع ارتفاع ملحوظ في قيمة
معامل التحديد.

**2-3-3-2 نتائج التجربة لقيم الاستجابات الحقيقية الخاصة بلوغاريتم النسبة
المضافة (Logit) الحقيقية (y^*)**

يتم استخراج قيمة احتمال الاستجابة الحقيقية للحشرات المقتولة ومن ثم يتم احتساب قيمة
لوغاريتم النسبة المضافة على وفق المعادلة (2-20) التي تم تناولها في الجانب النظري, ولكن
عندما تكون قيمة الاحتمال ($p=0$) او ($p=1$) عندها لا يمكن تطبيق طريقة المربعات الصغرى

الموزونة (WLS) ولمعالجة هذه المشكلة يمكن الاستعانة بقاعدة (Berkson) والتي تم التطرق لها في الجانب النظري على وفق المعادلة (2-40), ويمكن عرض النتائج الخاصة لقيم (logit) الحقيقية بالجدول (15-3).

جدول (15-3) يوضح احتمال الاستجابة الحقيقية وقيم لوغاريتم النسبة المضافة الحقيقية (y^*)

$\ln D_1$	$\ln D_2$	K_i	n_i	p_i	$\ln \frac{p}{1-p}$	y^*
-1.204	-2.996	6	150	0.04	-3.19	0.041
-0.916	-2.996	14	150	0.09	-2.31	0.099
-0.693	-2.996	23	150	0.15	-1.71	0.177
-0.511	-2.996	39	150	0.26	-1.05	0.341
-1.204	-2.303	6	150	0.04	-3.19	0.041
-0.916	-2.303	14	150	0.09	-2.31	0.099
-0.693	-2.303	23	150	0.15	-1.71	0.177
-0.511	-2.303	39	150	0.26	-1.05	0.341
-1.204	-1.609	30	150	0.2	-1.39	0.249
-0.916	-1.609	36	150	0.24	-1.15	0.317
-0.693	-1.609	47	150	0.31	-0.80	0.449
-0.511	-1.609	62	150	0.41	-0.36	0.810
-1.204	-1.204	42	150	0.28	-0.94	0.390
-0.916	-1.204	48	150	0.32	-0.76	0.473
-0.693	-1.204	59	150	0.39	-0.45	0.638
-0.511	-1.204	74	150	0.52	0.08	0.923

إذ يبين العمود السادس من الجدول المذكورة انفاً قيم (logit) والنتيجة من تطبيق المعادلة (2-20) كما ذكرناها سابقاً , وقد تم تحويلها الى قيم موجبة عن طريق اخذ الاساس (e) لكل قيمة من قيم (logit)

يتم الحصول على تقديرات معالم أنموذج الانحدار باستعمال طريقة المربعات الصغرى الاعتيادية (OLS) بعد التخلص من مشكلة عدم تجانس تباين الخطأ وذلك بتطبيق المعادلة

(2-10) التي تم التطرق اليها في الجانب النظري عن طريق تطبيق هذه الطريقة على بيانات حقيقية, ومن خلال استخدام برنامج (Mat lab 2011) تم الحصول على النتائج الآتية :

$$\hat{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.2311 \\ 0.6033 \\ -0.1719 \end{pmatrix}$$

وكذلك يتم حساب الاستجابات المتوقعة للحشرات المقتولة (k^*) والتي تم استخراجها من جدول (11-3) ومن ثم يتم حساب قيمة الاحتمال المتوقعة (π_i) للوغاريتم النسبة المضافة (logit) المتوقعة , ويمكن عرض النتائج في الجدول (16-3) :

الجدول (16-3): يمثل قيم احتمال الاستجابة المتوقعة لقيم لوغاريتم النسبة المضافة (y^*) logit المتوقعة

$\ln D_1$	$\ln D_2$	k_i^*	n_i	π_i	$\ln \frac{\pi}{1-\pi}$	y_i^*
-1.204	-2.996	12.3	150	0.08	-2.44	0.09
-0.916	-2.996	16.3	150	0.11	-2.09	0.124
-0.693	-2.996	22.1	150	0.15	-1.71	0.177
-0.511	-2.996	31.2	150	0.20	-1.39	0.25
-1.204	-2.303	12.3	150	0.08	-2.44	0.09
-0.916	-2.303	16.3	150	0.11	-2.09	0.124
-0.693	-2.303	22.1	150	0.15	-1.71	0.177
-0.511	-2.303	31.2	150	0.20	-1.39	0.25
-1.204	-1.609	26.1	150	0.17	-1.59	0.20
-0.916	-1.609	34.6	150	0.23	-1.207	0.21
-0.693	-1.609	47.3	150	0.31	-0.80	0.45
-0.511	-1.609	66.6	150	0.44	-0.24	0.786
-1.204	-1.204	33.3	150	0.22	-1.273	0.781

-0.916	-1.204	44.4	150	0.21	-1.324	0.264
-0.693	-1.204	60.31	150	0.40	-0.40	0.670
-0.511	-1.204	84.9	150	0.57	-0.291	0.23

وعند تطبيق الطرائق المستعملة (WLS, MLE, jack) على وفق المعادلات الآتية والتي تم ذكرها مسبقاً (2-57), (2-56), (2-42) المستعملة , عندها يتم الحصول على تقديرات لمعالم نموذج الانحدار الخاص بالاستجابات المتوقعة بالنسبة للوغاريتم النسبة المضافة (logit) باستعمال طريقة المربعات الصغرى (OLS) , وكذلك تم احتساب المقاييس الاحصائية (R^2 , MSE) من نفس الصيغ التي تم احتسابها من الجانب التجريبي اي بتطبيق الصيغتين (3-6, 3-5) من ثم تم الحصول على النتائج التقديرية , ويمكن عرض هذه نتائج في الجدول (17-3)

جدول (17-3) : يوضح معامل التحديد (R^2), MSE, اضافة الى القيم التقديرية لمعالم نموذج الانحدار لقيم الاستجابة المتوقعة بالنسبة لقيم اللوغاريتم النسبة المضافة Logit (y^*) المتوقعة

المعالم المقدرة	WLS	MLE	Jack
b_0	0.2154	0.221	0.2011
b_1	0.2722	0.5929	0.6133
b_2	-0.2911	-0.1866	-0.1923
R^2	0.7933	0.9844	0.9634
MSE	0.6066	0.3213	0.3315

أظهرت النتائج في الجدول المذكورة انفاً للقيم الخاصة للاستجابة المتوقعة بالنسبة للوغاريتم النسبة المضافة logit الى أفضلية طريقة الامكان الاعظم (MLE) حيث حصلت على اقل متوسط مربعات الخطأ (MSE) مقداره (0.3213) وكذلك حصلت هذه الطريقة على أعلى معامل تحديد (R^2) مقداره (0.9844) وهو يمثل ما فسرتة المتغيرات التوضيحية للمتغير المعتمد , وهذا بدوره يدل على افضلية وكفاءة طريقة الامكان الاعظم وحسب النظرية الاحصائية القائلة يكبر حجم العينة كلما يقل (MSE) ويزداد معامل التحديد ويقترّب من الواحد الصحيح , ومن ثم يليها طريقة الجاكنيف (Jack) من حيث الكفاءة والافضلية إذ حصلت على MSE مقداره (0.3315) وهو متقارب مع طريقة (MLE) وايضاً قد حصل على معامل التحديد مقداره (0.9634) , وهي نفسها النتائج التي تم التوصل اليها في الجانب التجريبي عند القيام بتجربة المحاكاة .

الفصل الرابع

الاستنتاجات

و

التوصيات

عن طريق ما تم عرضه في الجانب النظري والتجريبي وكذلك التطبيقي فقد توصلت الدراسة إلى الاستنتاجات والتوصيات الآتية :

أولاً : الاستنتاجات Conclusions

1 - توصلت الدراسة إلى أن طريقة الامكان الاعظم (MLE) وطريقة الجاكنيف (jak) تعطيان أفضل وأكفاً تقديراً للمعالم أنموذج الانحدار اللوجستي للاستجابات الثلاث بالاعتماد على مقياس متوسط مربعات الخطأ (MSE)

2 - نستنتج بأن لحجم العينة تأثيراً واضحاً وصريحاً في قيمة معامل التحديد (R^2) إذ تكون قيمته صغيرة عند حجوم العينات الصغيرة لكن مع زيادة حجم العينة ترتفع قيمة معامل التحديد وتقترب من الواحد الصحيح وهذا بدوره يدل على أفضلية الطريقة المستعملة, على العكس تماماً من قيمة متوسط مربعات الخطأ (MSE) التي تبدأ بالانخفاض كلما زاد حجم العينة يقل (MSE) .

3 - وجوه تقارب واضح في أن (MSE) يتناسب عكسياً مع حجم العينة لقيم التقدير لأغلب طرائق استجابات المتغير المعتمد متعدد الاستجابة.

4 - عند إجراء المقارنة بين الطرائق المستعملة لنموذج الانحدار المتعدد للاستجابات الثلاث بالاعتماد على مقياس (MSE) ولجميع حجوم العينات المستعملة تبين لنا ما يأتي :

أ- في ضوء ما جاء في الجانب النظري والتجريبي والتطبيقي للرسالة , وكذلك على ضوء النتائج التي تم في حالة العينات الصغيرة ($n=20$)

كانت طريقة المكان الاعظم للاستجابة الاولى والثانية هي الافضل اعقبتها طريقة الجاكنيف في الاستجابة الثالثة إذ حصلت على أقل (MSE) .

ب- في حالة العينات المتوسطة ($n=50$)

تعد طريقة الجاكناييف (jak) هي الافضل والاكفأ ولجميع الاستجابا □ إذ حصلت على أقل متوسط مربعا □ الخطأ (MSE) .

ج- في حالة العينات الكبيرة ($n=100, 200$)

عند حجم العينة ($n=100$) كانت طريقة الامكان الاعظم هي الافضل في الاستجابة الاولى والثانية إذ حصلت على أقل متوسط مربعا □ الخطأ (MSE) بينما في الاستجابة الثالثة كانت طريقة الجاكناييف هي الافضل

بينما عند حجم العينة ($n=200$) نلاحظ في الاستجابة الاولى تفوق طريقة الامكان الاعظم (MLE) ثم طريقة الجاكناييف عند الاستجابة الثانية والثالثة إذ حصلت على أقل (MSE).

5 - عند تطبيق الطرائق المستعملة على بيانا □ حقيقية تبين تفوق طريقة الامكان الاعظم (MLE) إذ حصلت على أقل (MSE) تليها طريقة الجاكناييف وهذا يتماشى مع نتائج تجربة المحاكاة .

ثانياً : التوصيات Recommendation

توصل اليها, توصي الدراسة بما يأتي:

- 1 – ضرورة استعمال نماذج \square الية أخرى غير النماذج اللوجستية عند تطبيق بياننا \square متعدد الاستجابة كدوال التربيعية , والدوال التمييزية وغيرها من الدوال الاخرى .
- 2 – ضرورة استعمال طريقتي الامكان الاعظم (MLE) والجاكنايف (jak) في توزيعا \square احتمالية أخرى ومن ثم تطبيق هذه الطرائق على جوانب تطبيقية أخرى كالجوانب الطبية أو الاجتماعية .
- 3 – ضرورة القيام بدراسة أنموذج الانحدار اللوجستي متعدد \square الاستجابة بوجوه \square مشاكل أخرى كمشكلة التعدد \square الخطي أو الارتباط الذاتي وتطبيقها على بياننا \square طبية او حياتية
- 4 – ضرورة استعمال المتغير المعتمد متعدد \square الاستجابة في تطبيقا \square شائعة صحية واجتماعية وحياتية وذلك لما لها من \square ور في تحليل الخاصة بذلك التطبيق.
- 5 – ضرورة استعمال طريقة (jackknife) وتشجيع الباحثين بالتطرق لها في بحوثهم النظرية والتطبيقية بغية الحصول على تقدير المعلما \square باقل تحيز ممكن من تقديرا \square الطرائق الاخرى .

قائمة

المصادر

REFERENCES

المصادر العربية :

- 1- القرآن الكريم
- 2- احمد , سهاد احمد " استعمال اشجار الانحدار التصنيفية والانحدار اللوجستي في تقدير أنموذج تجميعي والمقارنة بينهما مع تطبيق عملي " أطروحة دكتوراه , كلية الادارة والاقتصاد , قسم الاحصاء , جامعة بغداد, 2016
- 3- ايليا , يوربتيوئيل, " استخدام اسلوب الجاكنائف (Jackknife) لتقدير معلمات انموذج انحدار غير خطي مع تطبيق عملي " رسالة ماجستير – كلية الادارة و الاقتصاد , جامعة بغداد, 2007.
- 4- بيثون, نغم نافع "خوائص قوة الاختبار وحدود الثقة لمعاملات الانموذج اللوجستي الخطيدراسة مقارنة " , رسالة ماجستير في الإحصاء, كلية الإدارة والاقتصاد, جامعة بغداد. 1992 .
- 5- الثعلبي , ساهرة حسين زين " تحليل البيانات الثنائية لدراسة العوامل المؤثرة في حدوث التشوهات الولادية في مستشفى البصرة للنسائية والاطفال "رسالة ماجستير , كلية الادارة والاقتصاد , جامعه البصرة. 2008
- 6- جبارة, أزهار كاظم, " تحليل البيانات متعددة الاستجابة لتشخيص أمراض العيون باستخدام الدالة التمييزية والانحدار اللوجستي (دراسة مقارنة) " , رسالة ماجستير, كلية الإدارة والاقتصاد , الجامعة المستنصرية , 2014.
- 7- الجنابي , ضوية سلمان حسن, والشمري, نذير عباس, وعبود , سهيل نجم " تحليل البيانات المصنفة وتطبيقاتها " 2012 .
- 8- الجنابي , مشتاق عبد الغني شخير " بعض تطبيقات النماذج اللوغاريتمية الخطية في جداول التوافق متعدد الاتجاهات " رسالة ماجستير في الرياضيات – كلية التربية – جامعة المستنصرية , 2002 .
- 9- حسين, شرين علي " مقدرات الإمكان الاعظم الموزونة الحصينة ومقارنتها مع طرائق أخرى لأنموذج اللوجستيك مع تطبيق عملي " رساله ماجستير في الإحصاء كلية الإدارة والاقتصاد , جامعة بغداد , 2009.
- 10- الراوي , خاشع محمود , " المدخل الى تحليل الانحدار "مديرية دار الكتب للطباعة والنشر , جامعة الموصل, 1987.

- 11- رزاق , حسن هبوب " مقارنة بعض طرائق تقدير انموذج الانحدار شبه المعلمي بوجود مشكلة عدم تجانس تباين الخطأ مع تطبيق عملي " رسالة ماجستير – كلية الادارة والاقتصاد – جامعة بغداد , 2013.
- 12- سامبريت, برايس " تحليل الانحدار بالأمثلة " ترجمة محمد مناجد عيفان مراجعة دكتور اموري هادي كاظم , مطبعة التعليم العالي, 1990 .
- 13- السباح , شروق عبد الرضا" بناء أنموذج انحدار لوجستي معدل لحياة الاطفال الخدج في محافظة كربلاء " اطروحة دكتوراه – كلية الادارة والاقتصاد , جامعة بغداد , 2009.
- 14- سعيد, رشا عادل" استخدام نموذج الانحدار اللوجستي في دراسة العوامل المساعدة على تشخيص حالات الاصابة بسرطان المثانة". مجلة العلوم الادارية والاقتصادية – المجلد 21- العدد 83- 2015.
- 15- الشيلخي, هند مهند فوزي " تحديد أفضل مناطق ثقة لمعلمات توزيع متعدد الحدود مع تطبيق عملي " , رسالة ماجستير , جامعة بغداد, 2014.
- 16- الطائي, عبد الحسين حسن" تقدير وتحليل معادلة الانحدار المتعدد في حالة كون المتغيرات المعتمدة وصفية ومحددة " اطروحة دكتوراه , كلية الادارة والاقتصاد , جامعة بغداد, 2000
- 17- العابدي, فاضل عبد العباس" اسلوب تحليل البيانات المصنفة المترابطة". مجلة ابحاث البصرة – المعهد التقني الكوفة- العدد 38- الجزء 4, 2012
- 18- العزاوي, احمد ذياب احمد " المقارنة بين طرائق تقدير انموذج انحدار اللوجستيك والطرائق الحصينة لتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة" , رسالة ماجستير , كلية الادارة والاقتصاد , جامعة بغداد , 2005.
- 19- قاسم , بهاء عبد الرزاق " تحليل أثر بعض المتغيرات في الاصابة بمرض اللثة باستخدام نموذج الانحدار اللوجستي " مجلة العلوم الاقتصادية والادارية , العدد 27 , المجلد 17 , 2011.
- 20- كاظم , اموري هادي " مقدمة في القياس الاقتصادي". دار الكتب والوثائق زهران - عمان 2012,
- 21- الكريطي, عذراء عقيل عبد الصاحب " دراسات حقلية ومختبرية عن حشرة حفارة عذوق النخيل في بعض مناطق محافظة كربلاء" رسالة ماجستير – كلية الزراعة – قسم وقاية نبات- جامعة كربلاء, 2015 .

- 22- محمد , ايناس عبد الحافظ " تقدير نماذج مختلطة للبيانات المصنفة مع التطبيق العملي " اطروحة دكتوراه – كلية لإدارة والاقتصاد – جامعة بغداد- 2016.
- 23- المناصير, وضاح صبري "طرائق تقدير انموذج راش للبيانات المصنفة متعددة القياسات مع التطبيق عملي". اطروحة دكتوراه– كلية الادارة والاقتصاد- جامعة بغداد,(2016).
- 24- الوكيل, خولة حسين" استخدام تقنية الانحدار اللوجستي متعدد الحدود في دراسة اسباب الهجرة للشباب بعمر (18-30) سنة " مجلة اتحاد الإحصائيين العرب , المجلد الثالث , العدد الثاني , 2015.

المصادر الاجنبية :

- 25- Abdulla .m .Ellabell, "An Application on multinomial logistic Regression model". HEAD OF the department of Applied statistics facuitr of Economics and Administrative sciences, AL-AZHAR university, Gaza-palest in .Vol.VIII.No (.2).pp. 271-291,2012.
- 26- Agresti. A " **categorical Data Analysis**" second Edition ,university of Florida,2002 .
- 27- Agresti. A" **in trodution to categorical Data Analysis**", Second Edition ,chapter four .pp 99, Johhwiley,2007.
- 28- Aitcheson's, silvey .s .D" **the Generalization of probit Analysis to the case of multiple responseses** " .Biometrika,Vol.44,pp 131-140,,1957 .
- 29- Ajitic.T" **Randomized Response Techniques for multiple sensitive at ributes**" .JASA.Vol.76.pp916-923 ,1981.
- 30- AL- AFIFI ,R .M,"**The Use of Multinomial Logistic Regression Model on Physical Violence Data**" degree of Master of Applied Statistics, Al- Azhar University – Gaza,(2010).
- 31- Aldrich, J. H., & Nelson, F. D. "**Linear probability**", logit , and probit models. Beverly Hills, CA: Sage,1984.

- 32- AMemiya T "Regression analysis when variance of the dependent variance is proportional to the square of its expectation" JASA .VOL .68 .NO. 344 .P 928 - 934 ,1973
- 33- Albert, A., and Anderson, J. A. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models" , Biometrika , Vol. 71, No. 1 (Apr., 1984), pp. 1-10,1984.
- 34- Berkson . j " Approximation of the chi-square by probit and by logistic "JASA,Vol 41 , p 70-74,1946
- 35- Berkson . J,"Application of the logistic function to Bio-Assay". JASA,Vol,39.No,227,pp357-365, 1944.
- 36- Berkson.J,"why prefer Logistic to probit". Biometric.Vol.6.No.4.pp327-339,1951.
- 37- Dale. O .w. and James E, "Analysis of contingency table having ordered response categories" , JASA,Vol.67, No.337,1972.
- 38- David. A. "freedmanStatistical model theory and part ice , Cambridge, New york,.,2009.
- 39- Draper .N. R. and smith. H, "Applied Regression Analysis" 2nd edition. New york. john wiley and sons,.,1981.
- 40- Ethel. s. Gilbert. "On discrimination using Qualitative variable".JASA.Vol.63.No.324.pp1399-1412,1968.
- 41- Finney , D .J" Probit Analysis". Cambridge university. pre ss ,uk.1971 .
- 42- Gill , J" Generalized Linear models : A unified Approach . university of Florida .
- 43- Gujarati. Porter. D," Basic Econometrics", Fifth Edition,Mc craw-hill. Gapter.15. pp541-545,2009.
- 44- Gujarati, "Basic Econometrics " , fourth Edition, Mcg raw-Hill. Gapter.15. p 587-589,.,2004.

- 45- John o .Rawlings , and others. " **Applied Regression analysis** " A Research Tool , Second Edition , North Carolina state university , pp430-428 ,1998 .
- 46- Johnson ,N. and Kotz , S, "**Distribution in Statistics Discrete Distribution** " Willey Series , USA1976 .
- 47- Klein Baum. D. and Klein. M"**logistic Regression**", As elf - Learning text , third Edition , Department of Epidemiology, Emory university,2005.
- 48- Krieng , kitbumrungrat, "**Comparison logistic regression and Discriminant analysis in classification groups for breast cancer**"Faculty of Information Technology, Rang sit University, Thailand,2012.
- 49- Long , J .S. "Regression models for categorical and limited dependent Variable . Thousand oaks , CA .sage , bgn,1997 .
- 50- M.S Bartlett .F.R.S &D.R.COX,F.R.S, "**The Analysis of Contingency Tables**" , Jone Wily & Sons, Inc, New York,, 1977 .
- 51- McCarthy, W.F "**The Existence of Maximum Likelihood Estimates for the Binary Response Logistic Regression Model**", Collection of Biostatistics Research Archive, COBRA Preprint Series, Paper,2007,
- 52- Mcgullage .p . and Nelder. J. A. "**Generalized linear model**". Department of statistics , university, of chic age, second Edition,1983
- 53- Menard, Scott, "**Applied logistic Regression Analysis**", 2nd. Edition. Sage university . paper series on quantitative Application in the social sciencies. No 07-106. Bererly. halls.ca,.2002.
- 54- Nelder. J. A. , and wederburn "**Generalized Linear model theory** " , Appendix B (1972).

- 55- Ogoke, U.P., and Nduka, E.C., and Nja, M.E, "**A New Logistic Ridge Regression Estimator Using Exponentiated Response Function**", Journal of Statistical and Econometric Methods, vol.2, no.4, , PP. 161-171,2013.
- 56- Simith ,D.C. and Pontius ,S.J. "**Jackknife Estimator of Species Richness with S. plus**", Journal of Statistical software , Journal ,Vol 15,2006 .
- 57- Steffen .I . lauritzen, "**Lectures on contain geney tales**" Electronic edition, university of Aalborg ,,2000.
- 58- Titma ,Tuma ,Rooma), "**Education as afactor in integenea-tional mobility in Sovier Society**" ,Integrating Theory and Research,,2002.
- 59- Turner .H, "**In trodution to Generalizedlinear models** ", depart ment of statistics .universecity of Warwick, uk,2008.

الملاحق

ملحق (1)

جدول (A) يوضح تحويل النسب المئوية إلى وحدة الاحتمال الطبيعي (Normit)

y_i										
%	0	1	2	3	4	5	6	7	8	9
0	-	-2.33	-2.05	-1.88	-1.75	-1.64	-1.55	-1.48	-1.41	-1.34
10	-1.28	-1.23	-1.18	-1.13	-1.08	-1.04	-0.99	-0.95	-0.92	-0.88
20	-0.84	-0.81	-0.77	-0.74	-0.71	-0.67	-0.64	-0.61	-0.56	-0.55
30	-0.52	-0.50	-0.47	-0.44	-0.41	-0.39	-0.36	-0.33	-0.31	-0.28
40	-0.25	-0.23	-0.20	-0.18	-0.15	-0.13	-0.10	-0.08	-0.05	-0.03
50	0	0.03	0.05	0.08	0.10	0.13	0.15	0.18	0.20	0.23
60	0.25	0.28	0.31	0.33	0.36	0.39	0.41	0.44	0.47	0.50
70	0.52	0.55	0.58	0.61	0.64	0.67	0.71	0.74	0.77	0.81
80	0.84	0.88	0.92	0.95	0.99	1.04	1.08	1.13	1.18	1.23
90	1.28	1.34	1.41	1.48	1.55	1.64	1.75	1.88	2.05	2.33
99	2.33	2.37	2.41	2.46	2.51	2.58	2.65	2.75	2.88	3.09

ملحق (2)

جدول B يمثل قيم الاوزان W_i

W_i										
Y_i	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-4	0.001	0.001	0.001	0.002	0.002	0.003	0.005	0.006	0.008	0.011
-3	0.015	0.019	0.025	0.031	0.040	0.050	0.062	0.076	0.092	0.11
-2	0.131	0.154	0.180	0.208	0.238	0.269	0.302	0.336	0.370	0.405
-1	0.439	0.471	0.503	0.532	0.558	0.581	0.601	0.616	0.627	0.634
0	0.637	0.634	0.627	0.616	0.601	0.581	0.558	0.532	0.503	0.471
1	0.459	0.405	0.370	0.336	0.302	0.269	0.238	0.208	0.180	0.154
2	0.131	0.110	0.092	0.076	0.062	0.050	0.040	0.031	0.025	0.019
3	0.015	0.011	0.008	0.006	0.005	0.003	0.002	0.002	0.001	0.001

لمزيد مراجعة المصدر : كتاب Finney^[41]

ملحق (3): برنامج

- برنامج الاتي يمثل برنامج خاص با جانب تجريبي وتطبيقي واذي كتب بلغة (mat lab) من إصدار (2011) ويمكن أدراجه باشكل التالي :

```
clc
clear all
Bo1 = [0.5;-0.4;0.3];
Bo2 = [0.4;0.5;0.4];
Bo3 = [0.3;0.6;-0.2];
N=[20 50 100 200];
forrn=4:4
n=N(rn);Bo=Bo1;
x1= ([rand(n,1) rand(n,1)]);
x=( [ones(n,1) x1]);
pi = exp(x*Bo)./(1+exp(x*Bo));pi=pi';
ni=[1:n];
y=binornd(ni,pi);
y=y';
k1 = length(Bo);
k=k1-1;
for q=1:1000
%%%%%%%%%%WLS
vy=(( (ni(1)+1)*(ni(2)+2))./(ni.^3).*(pi+(1./ni))).*(1-
pi)+(1./ni))).^(-1);
w1=diag(sqrt(vy));ys=log((pi+(1./ni))./(1-
pi)+(1./ni));ys=ys';w=inv(w1);
Bwls(:,q)=inv(x'*w*x)*(x'*w*ys);
pwls=exp(x*Bwls(:,q))./(1+exp(x*Bwls(:,q)));
ywls(:,q)=binornd(ni,pwls');ywls(:,q)=ywls(:,q)';
Bo=Bo1;
mwls(:,q)=(((Bo-(Bwls(:,q))).^2));
bwls(:,q)=(((Bo-(Bwls(:,q)))));
mywls(:,q)=(sum((y-(ywls(:,q))).^2));
%%%%%%%%%% mle
v=diag(ni.*pi.*(1-pi));
Bmle(:,q)=Bo+(inv(x'*v*x)*x'*(y-ywls(:,q)));
while abs(Bmle(:,q)-Bo)>0.001
pmle=exp(x*Bmle(:,q))./(1+exp(x*Bmle(:,q)));
ymle=binornd(ni,pmle');ymle=ymle';
Bo=Bmle(:,q);
Bmle(:,q)=Bo+(inv(x'*v*x)*x'*(y-ymle));
end
Bmle(:,q);
pmle(:,q)=exp(x*Bmle(:,q))./(1+exp(x*Bmle(:,q)));
%%%%%%%%%%jackknife
```

```

ymlej=binornd(ni,pmle(:,q)');
Bo=Bo1;
for i=1:n
    y1=y;
    y1(i,:)=[];
    xj=x;
    xj(i,:)=[];
    pil=pi;
    pil(:,i)=[];
    yj=ymlej';
    yj(i,:)=[];
    nil=ni;
    nil(:,i)=[];
    vj=diag(nil.*pil.*(1-pil));
    Bj(:,i)=Bo+(inv(xj'*vj*xj)*xj'*(y1-yj));
    while abs(Bj(:,i)-Bo)>0.5
        pil=exp(xj*Bj(:,i))./(1+exp(xj*Bj(:,i)));
        yj=binornd(nil,pil');yj=yj';
        Bo=Bj(:,i);
        Bj(:,i)=Bo+(inv(xj'*vj*xj)*xj'*((y1-yj)));
    end
end
Bo=Bo1;
ymle(:,q)=binornd(ni,pmle(:,q)');
mmle(:,q)=(((Bo-(Bmle(:,q))).^2));
bmle(:,q)=(((Bo-(Bmle(:,q)))));
mymle(:,q)=(sum((y-(ymle(:,q))).^2));
Bj(:,q)=mean(Bj,2);
Bjak(:,q)=n*Bmle(:,q)-((n-1)*Bj(:,q));
Bjak(:,q)=Bj(:,q);
pjak(:,q)=exp(x*Bjak(:,q))./(1+exp(x*Bjak(:,q)));
yjak(:,q)=binornd(ni,pjak(:,q)');
Bo=Bo1;
mjak(:,q)=(((Bo-(Bj(:,q))).^2));
bjak(:,q)=(((Bo-(Bj(:,q)))));
myjak(:,q)=(sum(((y-(yjak(:,q))).^2)));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end
n
mse_Bwls=mean(mwls,2);mse_Bmle=mean(mmle,2);mse_Bjak=mean
(mjak,2);
Bias_Bwls=mean(bwls,2);Bias_Bmle=mean(bmle,2);Bias_Bjak=m
ean(bjak,2);
mse_ywls=mean(mywls,2);mse_ymle=mean(mymle,2);mse_yjak=me
an(myjak,2);
F1={'n','Bo','mean(Bwls,2)','mean(Bmle,2)','mean(Bjak,2)'}
;...
n,Bo,mean(Bwls,2),mean(Bmle,2),mean(Bjak,2)}
F2=[Bo,mean(Bwls,2),mean(Bmle,2),mean(Bjak,2)];
mse_B={'mse_Bwls','mse_Bmle','mse_Bjak';mse_Bwls,
mse_Bmle,mse_Bjak}

```

```
Rwls=1-((mean(mywls))/(sum((y-mean(y)).^2)));Rmle=1-  
((mean(mymle))/(sum((y-mean(y)).^2)));  
Rjak=1-((mean(myjak))/(sum((y-mean(y)).^2)));  
Rs=[RwlsRmleRjak]  
Bias={'Bias_Bwls', 'Bias_Bmle',  
'Bias_Bjak';Bias_BwlsBias_BmleBias_Bjak}
```

ABSTRACT

Choose the best estimation Methods for the multi – response dependent Variable .

The descriptive dependent variables are important one that do not have quantitative measurement units, they subject to the description of the phenomenon through studying the data and information, and enables the decision–maker to identifying the nature of response when the variable is binary–response or multi–response. The response function may be nonlinear; therefore, there is a need to find other methods to converting it into linear one, this is by means of logarithmic conversion or other transformation that can be converted into linear responses.

We used in our thesis the descriptive regression model, when the dependent variable is multi–response, by studying some models related to these descriptive variables.

The aim of our thesis is to studying and analyzing the descriptive variable, and its effect on the multi–response dependent variable, by estimation the parameter of model of dependent variable, using weighted least squares method, and the method of maximum likelihood, using "Newton Raphson" and "Jackknife estimation" methods.

We used two sides in our study, experimental and empirical, on the experimental we used in our study, experimental and empirical, on the experimental side; we used "The Monte Carlo" method in simulations experiment for three levels of samples (small, middle and large) in different sizes by generate random numbers for the parameter of regression model. In addition, the practical aspect was applied to the life experience of some insecticidal in different concentration, for choosing, then, the best method in reliance on the measurement of Mean Squares Error (MSE), We found, by using The Monte Carlo method of estimation, that the (MLE) method is the best and efficient in the small samples, whereas the (JAK) method is the best in the middle samples, and the two methods of (MLE and (JAK) are equivalent in the large sample size. From the empirical side, on the other hand,

we found the result of estimation is that the (MLE) method is the best and efficient because it gets less (MSE).

Republic of Iraq
Ministry of Higher Education and Scientific
Research
University of Karbala
Faculty of Management and Economics
Department of Statistics



Choose the best estimation Methods for the multi- response dependent Variable

A thesis

*Submitted to College of Administration and Economic-Karbala
University in partial fulfillment of the Requirements for the Degree
of master of Science in statistics*

By

Laith. Ali .M. Al-Hayalli

Under supervision

A. Dr . Abdul hussian. H . HABEAB . AL-Tai

H 1438

Holy Karbala

M 2017