جامعة كربلاء
كلية الادارة والاقتصاد
قسم الاحصاء

إختيار أفضل طريقة اختبار لمشكلة عدم تجانس التباين في نموذج الانحدار المتعدد (مع تطبيق عملي)

رسالة مقدمة الى
مجلس كلية الادارة والاقتصاد في جامعة كربلاء وهي جزء من متطلبات نيل درجة الماجستير في علوم الاحصاء

تقدم بها

الطالب
رائد اسمر عبد الله

بإشراف
الاستاذ الدكتور
عدنان كريم نجم الدين

١٤٣٩هـ ٢٠١٨م

بِسْمِ اِللّه الرَّحْمنِ الرَّحِيمِ

{لن تنالوا البر حتى تنفقوا مما تحبون
و ما تنفقوا من شيئ فأن اللّه به
عليم }

صدق اللّه العلي العظيم

الاهداء

الى ..... الرحمة الواسعة والانسانية بمعناها الكامل
الرسول محمد واهل بيته (عليهم الصلاة السلام)

الى الشموع التي جعلت الحياة جميلة ومستمرة.
الشهداء رحمهم الله

الى الذين اعطوا من اجسادهم لنا لتكتمل ارواحنا في الحياة.
الجرحى شفاهم الله

الى الذين اناروا للاجيال الطريق من خلال النور والمعرفة .

الى الشموع التي انارت لي طريق العلم .

أساتذتي .....وفاءً وعرفاناً

الى الذي غرس في نفسي العلم والمعرفة .....بذرة.

والدي رعاه الله

الى التي رعت تلك البذرة بحبها وحنانها ..... قطرة بعد قطرة.

والدتي رحمها الله براً وإحساناً

الى التي ازرتني في ايام شدتي ورخائي ..... رحمة ومودة.

زوجتي ورفيقة حياتي

الى الذين وقفوا بجانبي وشدوا من أزري .....وعاشوا معي عناء السفر.

إخوتي وأخواتي الطلبة أعزهم الله

أهدي هذا الجهد المتواضع

رائد

شكر وتقدير

الحمد لله ذي القدرة القاهرة ، والآيات الباهرة، والالاء الظاهرة، حمداً يليق بكماله، وأستعين به، وأشكره جزيل الشكر، واثني عليه العمر كله لنعمه الفضيلة، وصلى الله على خير الأولين والآخرين وأشرف الخلق أجمعين محمد "صلى الله عليه وعلى آله " الى يوم الدين .وبعد بتوفيق من الله تعالى ودوام وبركاته تم انجاز هذا الجهد العلمي. لذا كان من الواجب التوجه بوافر الشكر والامتنان لكل من ساهم في هذا العمل لاسيما الاستاذ الفاضل الدكتور عدنان كريم نجم الدين الذي شرفني بأشرافه على هذه الرسالة وتقويمها وانضاجها بملاحظاته وتوجيهاته القيمة ، لذا ارجوا ان يتقبل تقديري وشكري على الجهود التي بذلها كما اتقدم بالشكر الجزيل لاستاذي الدكتور جاسم ناصر حسين رئيس قسم الاحصاء وذلك لرعايته الأبوية وللدعم المتواصل ولتذليله الصعاب والمعوقات التي واجهت هذه الرسالة فنسأل الله تعالى ان يجزيه خيرا الجزاء وان يمد له العمر بالصحة والعافية إن شاء الله. والشكر موصول الى استاذي القدير الاستاذ الدكتور طاهر ريسان في جامعة القادسية على المجهود الكبير والمتواصل وتفانيه وتواضعه وتحملني بروحه المثابرة واخلاصه لما ابداه من ملاحظات قيمة كان لها الاثر الكبير في اتمام هذه الرسالة.

كما اتقدم بالشكر والتقدير الى كل اساتذتي الافاضل في قسم الاحصاء على الدعم المتواصل لي خلال مدة الدراسة.

وكذلك الشكر موصول الى الاخوة والاصدقاء في القسم الداخلي وخارجه لما قدموه من دعم معنوي في اكمال هذه الرسالة.

المستخلص

يعد هذا البحث محاولة لتسليط الضوء على احدى مشاكل تحليل الانحدار و التي تعد واحدة من الافتراضات الاساسية التي تقوم عليها طريقة المربعات الصغرى الاعتيادية (OLS) ، كذلك تعد واحدة من الشروط الاساسية لتحليل التباين. وهي مشكلة عدم تجانس التباين (Heteroscedasticity) ، وأحيانا يكون من الصعوبة تطبيق هذا الشرط وبالتالي تكون طريقة (OLS) لتقدير معالم نموذج الانحدار الخطي غير مجدية في اعطاء نتائج صحيحة ودقيقة وهذا بسبب وجود هذه المشكلة .

اذ كان هدف الرسالة هو اختيار اوتحديد افضل اختبار للكشف عن وجود هذه المشكلة ضمن مجموعة من الاختبارات المعلمية ، اذ تم كتابة برنامج بلغة (R) لاجل عمل مقارنة يكون فيها المعيار هو نسبة الكشف الصحيح عن وجود أو عدم وجود مشكلة عدم تجانس التباين عن طريقة دراسة محاكاة بالاعتماد على بيانات افتراضية لانموذجين الاول يحتوي على مشكلة عدم تجانس التباين والاخر لايحتوي على المشكلة ، اذا تم اعتماد خمس اختبارات هي (كولد فيلد-كوانت ، بروش- بيجين – جود فري ، وايت، NCV ، هاريسون مكابي) .

كما اعتمدت الرسالة في الجانب التطبيقي على البيانات الحقيقية، واستعمال الاختبار الافضل الذي تم تحديده للكشف عن وجود المشكلة ثم تقدير أنموذج الانحدار استنادا الى وجود المشكلة من عدمها.

.

المحتوى

التسلسل	الموضوع	رقم الصفحة
	الآية	أ
	الإهداء	ب
	الشكر والتقدير	ج
	المستخلص	د
	قائمة المحتويات	هـ - و- ز- ح
	قائمة الاشكال	و
	قائمة الجداول	ز - ي
الفصل الأول	**المقدمة و منهجية البحث و الاستعراض التاريخي**	**6-1**
1-1	المقدمة	**2-1**
2-1	مشكلة البحث	**2**
3-1	هدف البحث	**3**
4-1	اهمية البحث	**3**
5-1	حدود البحث	**3**
6-1	فرضية البحث	**3**
7-1	الاستعراض المرجعي	**6-3**
الفصل الثاني	**الجانب النظري**	**39-8**
1-2	مفهوم الانحدار	**8**
2-2	المفهوم الحديث للانحدار	**9-8**
3-2	الانحدار وعلاقته بالسبب	**9**

4-2	انواع الانحدار	10-9
1-4-2	الانحدار الخطي البسيط	10
2-4-2	الانحدار الخطي المتعدد	10
5-2	النموذج الخطي العام	12-10
6-2	الافتراضات	15-12
7-2	تقدير المعالم	18-15
8-2	المربعات الصغرى المعممة(GLS)	23-19
9-2	خصائص المعلمات المقدرة	24-23
10-2	إختبار الفرضيات	24
11-2	مفهوم معنوية الاختبار	25
12-2	اختبار (t) للمعلمات	26-25
13-2	طبيعة مشكلة عدم التجانس	28-26
14-2	اسباب مخالفة فرض ثبات التباين	29-28
15-2	تفسير مشكلة عدم تجانس التباين	30-29
16-2	الثبات واثار المخالفة	31-30
17-2	طرق الكشف عن مشكلة عدم تجانس التباين	37-32
1-17-2	اختبار كولد فيلد Cold field-Quant test	34-33
2-17-2	اختبار بروش- بيجين – جود فري Breusch-Pagan test	35-34
3-17-2	اختبار وايت Test White	35-34
4-17-2	الأنموذج الثاني اختبار Harrison-McCabe test	35
5-17-2	اختبار NCV Test NCV	37
18-2	معالجة مشكلة عدم التجانس	39-37
الفصل	**الجانب التجريبي**	**64-41**

الثالث		
1-3	مفهوم المحاكاة	**41**
1-1-3	تطبيق اسلوب المحاكاة	**42-41**
2-1-3	الاختبارات المستعملة في تجارب المحاكاة	**42**
3-1-3	تجارب المحاكاة	**43-42**
2-3	الحصول على المعلمات	**45-43**
3-3	تطبيق البيانات الافتراضية	**56-46**
4-3	تحليل النتائج	**57-47**
5-3	تطبيق لبيانات واقعية حول مرض اليرقان	**58**
6-3	تهيئة البيانات	**59-58**
7-3	اختبار طبيعية توزيع البيانات	**61-59**
8-3	تطبيق (الاختبار الافضل)	**61**
9-3	الكشف عن وجود او عدم وجود مشكلة عدم تجانس التباين	**62-61**
1-9-3	اختبار (Gold field-Quant)	**61**
2-9-3	إختبار (NCV)	**62**
10-3	تقديرات (OLS)	**63**
11-3	تحليل النتائج	**64-63**
الفصل الرابع	**الاستنتاجات والتوصيات**	
1- 4	الاستنتاجات	**66**
2- 4	التوصيات	**67**

المصادر	**72-69**
الملاحق	**81-74**

قائمة الأشكال

التسلسل	اسم الشكل	الصفحة
1	يوضح حالات ثبات او عدم ثبات التباين	28
2	يوضح شكل توزيع بيانات المتغير المعتمد (المدرج التكراري) مع المنحنى الطبيعي	60

قائمة الجداول

رقم الجدول	اسم الجدول	الصفحة
1	يبين المعلمات والمتوسطات والتباينات للبيانات الحقيقية	44
2	يبين المعلمات والمتوسطات والتباينات للبيانات الافتراضية العليا	44
3	يبين المعلمات والمتوسطات والتباينات للبيانات الافتراضية الدنيا	45
4	يوضح محاكاة نموذج انحدار بالمعلمات الافتراضية العليا للمعلمة (b_i).	46`
5	يوضح محاكاة نموذج انحدار بالمعلمات الافتراضية الدنيا للمعلمة (b_i).	48
6	يوضح محاكاة نموذج انحدار بالمعلمات الحقيقية للمعلمة (b_i)..	50
7	يوضح محاكاة نموذج انحدار بالمعلمات الحقيقية مع القيم المفترضة الدنيا للمتوسط وللتباين.	52
8	يوضح محاكاة نموذج انحدار بالمعلمات الحقيقية مع القيم المفترضة العليا للمتوسط وللتباين.	54
9	يوضح محاكاة نموذج انحدار بالمعلمات الحقيقية مع القيمة الحقيقية للمتوسط وللتباين.	56
10	يوضح اختبار التوزيع الطبيعي للبيانات الحقيقية للمتغير المعتمد (Y).	60
11	يوضح احصاءة اختبار(Gold field-Quant) على البيانات الحقيقية، مستوى المعنوية (0.05) في حالة احتواء النموذج على مشكلة عدم تجانس التباين.	61
12	يوضح احصاءة اختبار(NCV) على البيانات الحقيقية، مستوى المعنوية (0.05) في حالة عدم احتواء النموذج على مشكلة عدم تجانس التباين.	62
13	يبين تقديرات (OLS) للنموذج المقدر مع الانحرافات المعيارية وكذلك قيم (t) للمعالم المقدرة والقيمة المعنوية لاحصاءة الاختبار.	63

الفصل الأول

المقدمة

والاستعراض المرجعي

الفصل الاول
منهجية البحث والاستعراض المرجعي
Methodology of research and reference review

1-1 : المقدمة : Introduction

يعد تحليل الانحدار من الأساليب المهمة في التحليل الأحصائي ذات التطبيقات الواسعة اذ يقوم على عدة افتراضات اساسية وفي حالة عدم توفر احد هذه الافتراضات فأن ذلك يؤدي الى نتائج غير صحيحة ومن اهم هذه الافتراضات هو ثبات تجانس التباين اذ ان عدم تحقق هذا الشرط يسبب مشكلة تدعى بمشكلة عدم تجانس التباين (عدم ثبات تباين حد الخطأ) والتي قد تأتي من اختلاف العينات والتي تتبع مجتمعات ذات تباينات مختلفة ، وهذا بدوره يؤدي الى قرارات غير دقيقة او مضللة . ففي حالة اختبار الفرضيات اذ يرتفع مستوى المعنوية بشكل تلقائي لذلك يجب ان تكون الاختلافات العشوائية ضمن المعاملات على درجة كبيرة من التجانس وهذا بدوره يجعل الاختلافات العشوائية متساوية بالنسبة للعينات المختلفة ما يجعل معه الحصول على تباين مشترك للخطأ العشوائي لجميع العينات ممكنا، وتعد مشكلة عدم تجانس التباين من مشاكل القياس الاقتصادي التي تواجه الباحثين عند وصفهم وبنائهم لأنموذج إنحدار خطي لمجموعة من الظواهر وعلى هذا الاساس اصبح لزاماً على الباحث ان يختبر ويفحص البيانات والمعلومات لاجل معرفة وتوضيح العلاقة بين العوامل المختلفة وكذلك تحليل النتائج بشكل ادق واشمل . وبذلك تتجلى اهمية لجوء الباحث الى استعمال الطرائق الاحصائية التي تمهد له الوصف الموضوعي الدقيق لاجراء تجربته على اسس خالية من الاجتهادات والعوامل الشخصية التي بدورها قد يكون لها تأثير مباشر على نتائج البحث ، ولاجل الحيلولة دون حصول هذه الامور فقد تم إختيار عدد من الاختبارات تستعمل للكشف عن وجود هذه المشكلة او عدم وجودها . اذ سيتم التركيز في هذه الرسالة على المقارنة بين مجموعة من الاختبارات المعلمية Parametric (Tests). ولتوضيح منهجية هذه الرسالة فقد تم تقسيمها الى اربعة فصول، اذ يشمل الاول المقدمة وهدف ومشكلة البحث وكذلك الاستعراض المرجعي ، اما الثاني فقد تضمن الجانب النظري والذي تم عرض فيه توضيحا لموضوع الانحدار وايضا الاختبارات الخاصة بتجانس التباينات المعلمية في هذا المجال ، اما الثالث فقد تضمن مفهوم المحاكاة وفائدة استعمال هذا الأسلوب واساليب توليد البيانات واجراء تجربة محاكاة (لغة R للبرمجة الاحصائية) وتحليل النتائج الخاصة بكل اختبار والمقارنة بين الاختبارات على اساس

معيار نسبة نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين (Non-homogeneity detection ratio) ، كذلك تضمن الفصل الثالث تطبيقا عمليا لبيانات واقعية حول مرض اليرقان ، واما الفصل الرابع فتم فيه عرض الأستنتاجات التي تم التوصل اليها وذكر التوصيات الازمة .

1-2 مشكلة البحث [15]: Research problem

أولا: المشكلة النظرية :

عندما نعتمد على أنموذج انحدار خطي عام لتمثيل البيانات فأننا نستعمل طريقة المربعات الصغرى الاعتيادية (OLS) لتقدير معالم هذا الأنموذج وتعتمد هذه الطريقة عند تطبيقها على تحقق فرضيات وشروط اساسية معينة حتى تكون هنالك دقة في تقدير معالم أنموذج الأنحدار وواحدة من الفرضيات الاساسية التي نعتمدها عند عملية التقدير بطريقة (OLS) هي فرضية ثبات تجانس التباين (Homoscedasticity) لحدود الاخطاء والتي يعبر عنها رياضيا بالصيغة الاتية :

$$\mathrm{Var}\,(\underline{U}) = \mathrm{E}\,(u_i^2) = \sigma^2 I_n \qquad (1\text{-}1)$$

وفي كثير من التطبيقات العملية لا يمكن تحقيق هذه الفرضية مما يجعل طريقة (OLS) لتقدير معالم أنموذج الانحدار الخطي غير مجدية في إعطاء نتائج صحيحة ودقيقة وهذا يؤدي الى الوقوع في مشكلة عدم التجانس (Heteroscedasticity) التي تفقد المعالم المقدرة خاصية عدم التحيز وتجعلها لاتمتلك صفة أقل تباين ممكن ويعبر عن هذه المشكلة رياضيا بالصيغة الاتية:

$$\mathrm{Var}\,(\underline{U}) = \mathrm{E}\,(u_i^2) = \sigma^2 \Omega \qquad (1\text{-}2)$$

ثانيا : المشكلة العملية :

تعتمد الاصابة بمرض اليرقان لدى الاطفال الخدج على عدة عوامل حسب رأي واضعي هذه العوامل ، وهذه العوامل هي عمر الطفل ووزنه وفترة الرقود وعمر الام واحيانا تكون هذه العوامل غير مؤثرة على تحديد نسبة الاصابة بهذا المرض مما يجعل التعرف على العوامل الحقيقية من الصعوبة في شيئ .

1-3 هدف البحث: Research goal

الهدف من هذا البحث هو اختيار افضل اختباراللكشف عن مشكلة عدم تجانس التباين بالاعتماد على المقارنة بين بعض الطرائق الأحصائية المعلمية في ظل إختلال إحدى فرضيات تحليل الانحدار المتعدد وهي ثبات التباين لحد الخطأ .

1 -4 أهمية البحث: Research importance

العديد من الدراسات تناولت كيفية اختبار ومعالجة مشاكل الانحدار الناتجة عن إختلال احد فروض الانحدار، وأحدى هذه المشاكل المهمة هي مشكلة عدم تجانس التباين وتأتي اهمية هذا البحث في اختيار أنموذجين الأنموذج الاول يحتوي على مشكلة عدم تجانس التباين والاخرلايحتوي على مشكلة عدم تجانس التباين ليتم بعد ذلك اختيار الاختبار الافضل والذي يستعمل للكشف عن وجود او عدم وجود مشكلة عدم تجانس التباين .

1-5 حدود البحث: Research limits

_ الحدود المكانية : محافظة المثنى _ مستشفى الرميثة العام

_الحدود الزمانية : عام 2018 للاشهر(كانون الثاني ، شباط ، آذار)

1 -6: فرضية البحث: Research Hypotheses

_ فرضية العدم: أن تباين حد الخطأ لجميع مشاهدات العينة ثابت بمعنى عدم وجود مشكلة عدم تجانس التباين :

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \cdots = \sigma_i^2$$

_ الفرضية البديلة : على الاقل اثنين من التباينات مختلفة اي وجود مشكلة عدم التجانس :

$$H_1 : \text{At least two differ}$$

1-7 : الاستعراض المرجعي: Reference review

توالت الدراسات والبحوث التي اعتمدت على تحليل الانحدار وعلى الفرضيات الخاصة به و البحث في المشاكل التي تنتج عن إختلال بعض فرضياته لاسيما فرضية ثبات حد التباين (مشكلة

عدم تجانس التباين)، و كذلك اهمية الكشف عن وجود هذه المشكلة عن طريق استعمال الطرائق الاحصائية المعلمية في الكشف هذه المشكلة .

وفي عام (1994) قدم الباحثان (William R. Rice and Steven D. Gainest) بحثا بعنوان (توسيع اختبارات عدم التجانس غير الاتجاهي لتقييم فرضيات بديلة مرتبة ببساطة) اذ اكدا فيه ان علماء الاحياء في كثير من الاحيان بستعملون اختبارات عدم التجانس غير الاتجاهية عند مقارنة ثلاث او اكثر من السكان لان الاختبار الاتجاهي يكون في اغلب الاحيان غير متوفر او غير عملي. الباحثان قاما بوصف اختبارالتجانس المرتب يسمح باجراء الاختبار على فرضيات بديلة مرتبة ببساطة في سياق اي اختبار غير مباشر تقريبا **[25]** .

وفي عام (1998) قدم الباحث (By Chris D. Orme) بحثا بعنوان (على عدم حساسية اختبار النقطة لعدم تجانس التباين للمتغيرات المشتركة المحذوفة في نماذج وقت الفشل متعدد المتغيرات)، وكان من الملاحظات المهمة اتي تم التوصل اليها انه من الممكن ان يكون اختبار النقطة لعدم تجانس التباين حساسا نسبيا بالنسبة للمتغيرات المشتركة المحذوفة. اي ان هنالك ضعفا كبيرا في هذا الاختبار بالمقارنة مع اختبار نسبة الاحتمال الدقيق **[17]** .

عام (2002) قدم الباحثون (Luc D. , Paul J., Patrick L., Catherine L., Rosemary N. , Richard S.) بحثا بعنوان (نماذج الضعف والقوة لاختبارات عدم تجانس التباين في التجارب متعددة الدخول) اذ اعتقد الباحثون انه من الممكن تشكيل أنموذجا لعدم تجانس التباين بين تجارب متعددة المراكز مع الوقت المناسب لنتائج البحث وذلك عن طريق أنموذج المخاطرة النسبي الضعيف ، اذ استنتج الباحثون ان عدد المراكز وعدد المرضى في كل مركز يؤثر في نوعية التقديرات في وضع معين من التجارب السريرية اذ اظهر هذا الاسلوب التقديرات لمعلمة التجانس تقريبا نفسها **[18]**.

وقدم الباحثان (Maoa, C. X. Lindsayb, B. G.) عام (2003) بتقديم بحثا بعنوان (اختبارات وتشخيصات مشكلة عدم تجانس التباين في مشكلة الانواع) اذ افترض الباحثان ان عينة عشوائية من الافراد تم اخذها ضمن عدد سكان غير معلوم ومن فئات منفصلة ، ويقال ان السكان متجانسون اذا كانت جميع الفئات لها النسب نفسها وخلاف ذلك تكون غئير متجانسة . وعلى الرغم من ان هنالك مؤلفات كثيرة وواسعة عن التقدير ،اذ يرتبط عدد الفئات واداء المقدرين بل بافتراض التجانس ، وفي الحقيقة لم يحصل التجانس سوى القليل من التحقيق ، وكذلك تمت مناقشة اختبار (goodness-of-fit test) واختبار درجة التشتت ، وتم استعمال الرسوم البيانية لغرض الكشف عن مشكلة عدم تجانس التباين **[23]** .

وفي عام (2006) تم تقديم بحث بعنوان (تقدير معاملات الانحداربعد اجراء اختبار عدم تجانس التباين) من الباحث (آيدن حسن الكناني)، وكان هدف البحث هوالحصول على مقدر جديد يعرف بمقدر الاختبار الاولي والذي ينتج عن عملية دمج وربط مقدر طريقة المربعات الصغرى الاعتيادية اذ استنتج الباحث بان مقدر الاختبار الاولي والذي نتج عن دمج مقدر المربعات الصغرى الاعتيادية (OLS) بمقدر(Aitken) ذو المرحلتين (2 SAE) يعطي او يقدم نتائج مثل متوسط مربعات الخطأ صغيرة و تكون قليلة نسبيا ، واوصى الباحث بعدة توصيات كان منها بأجراء مقارنة بين المتوسط لمربعات الخطأ العائد لمقدر الاختبار الاولي وبين المقدر لطريقة المربعات الصغرى ومقدر ذي المرحلتين (Aitken) **[10]** .

وقدم الباحث (Jesse Frey) عام (2010) بحثا بعنوان (اختبار تكافؤ الفروق باستخدام نسبة هارتلي) معتمدا على فكرة اختبار (هارتلي) اذ يعتمد هذا الاختبار في طريقة حسابه على النسبة بين اقصى تباين وادنى تباين للعينة ، في هذا البحث استعمل الباحث احصاءة (هارتلي) نفسها لاختبار التكافؤ لعدد (k) من التباينات، وقد استنتج الباحث ان اختبار (هارتلي) لعدم تجانس التباينات يكون غير متحيز عندما تكون احجام العينات غير متساوية **[19]** .

وقدم الباحث (Donald W. Zimmerman) عام (2012) بحثا بعنوان (عدم تجانس التباين واختبار الفرضية المتحيزة) اذ تم فيه بحث تأثير عدم تجانس التباين في معدلات الخطأ من النوع الاول وقوة العينات المستقلة من اختبار (t) من اذ المساواة في العينات من الدرجات الطبيعية، اذ تم اجراء اختبارالمساواة نفسه على تحويل رتبة الدرجات المقابلة بالنسب لكثير من التوزيعات غير الطبيعية، اذ انتجت دوال شاذة في جميع الحلات التي وقع فيها التحيز واظهر اختبار (t) في الصفوف تحيزا اكبر بكثير من اختبار (t) على الدرجات. وهذه النتيجة الشاذه كانت مستقلة **[33]** .

وفي عام (2014) قدم الباحثان (Lan Wang and Xiao-Hua Zhou,) بحثا بعنوان (اختبار التشخيص الامعلمي تماما لتجانس الفروق) اذ اقترح الباحثان اختبارا تشخيصيا جديدا غير معلمي للتحقق من ثبات دالة التباين $\sigma^2 = (x)$ في أنموذج الانحدار $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$ وهذا الاختبار لايملك شكلا معلميا معروفا لدالة المتوسط الشرطية ، وهذه الفكرة مستوحاة من النظرية المقاربة في تحليل التباين عندما يكون عدد مستويات العامل كبيرا ، اذ تبين من نتائج المحاكاة انه باستعمل قيمة حرجة يتم الحصول عليها من الاختبار فأن الاختبار المقترح له اداء مرضي على عينة محددة **[30]** .

وفي عام (2015) قدمت الباحثة علا هادي صادق رسالة بعنوان (الطرائق المعلمية واللامعلمية لاختبارات تجانس التباين مع تطبيق عملي وكان هدف الرسالة امكانية تطبيق اختبارات التجانس في حالة عدم ثبات تجانس الخطأ بالاعتماد على مقارنة الاساليب المعلمية والامعلمية من خلال اسلوب المحاكاة معتمدين على معدلات الخطأ من النوع الاول (Type I error) وقوة الاختبار كمعايير ، اذ تم استنتاج وجود علاقة طردية بين حجم العينة وقوة الاختبار فكلما كبر حجم العينة زادت قوة الاختبار **[14]**.

وفي عام (2016) قدم (م.م احمد رازق عبد رمضان) بحثا بعنوان (تحليل مشكلة عدم تجانس التباين والمقارنة بين اختبار اوبراين واختبار اوبراين المعدل باستعمال المحاكاة) وكانهدف البحث هو المقارنة بين اختبار اوبراين المعدل واختبار اوبراين ، اذ كان المعيار هو قوة الاختبار لمعرفة اي من الاختبارين افضل ، وقد استنتج الباحث مجموعة من الاستنتاجات كان منها وجود علاقة طردية بين حجم العينة وقوة الاختبار **[8]** .

وفي عام (2016) قدم كل من (رمضان اللآمي، احمد رزاق عبد. طالب، حيدر رائد.) بحثا بعنوان (مقارنة اختبار Z المعدل للتباين مع اختبار ليفين المعدل بوجود القيم الشاذة باستعمال المحاكاة) ، اذ كان هدف البحث هو مقارنة اختبار ليفين المعدل مع مقارنة اختبار Z المعدل اذ كان معيار المفاضلة هو قوة الاختبار، وقد تم الاستنتاج عدة استنتاجات منها ان اختبار Z المعدل للتباين كان افضل لكونه يقبل الفرضية الصفرية والتي تؤكد ان التباينات متساوية وهذا سيثبت على عدم تأثر الاختبار بالقيم الشاذة **[11]** .

وفي عام (2016) قامت الباحثة (ايلاف بهاء علوان) بتقديم رسالة بعنوان (مقارنة بين اختبار Gold field Quant الحصين مع اختبارات أخرى للكشف عن مشكلة عدم تجانس التباين بوجود القيم الشاذة)، وكان هدف البحث هو عمل مقارنة بين الاختبارات الحصينة لاجل معرفة اي من الاختبارات المستعملة هي الاكفئ والافضل من بين الاختبارات المستعملة عند وجود مشكلة عدم تجانس التباين ، وتم الاعتماد على قوة الاختبار كمعيار للمفاضلة اذ لوحظ وجود علاقة طردية بين قوة الاختبار وحجم العينة اي كلما زاد حجم العينة ازدادة قوة الاختبار **[7]**.

الفصل الثاني

الجانب النظري

الفصل الثاني
المبحث الاول
مفهوم الانحدار ومشكلة عدم تجانس التباين طبيعتها واسبابها
The concept of regression and heteroscedasticity Nature and causes

1-2 : مفهوم الانحدار [13] [29] : Regression Concept

يقال ان الانحدار هو دراسة الاعتمادية وهذا يعني اعتماد متغير على متغير اخر كتأثير تغيير حجم الصف على نجاح الطلاب ، ويعد الانحدار من الموضوعات الاساسية ويعد جزءا مهما من النظرية الاحصائية ، ويصنف الانحدار ضمن الموضوعات المميزة بسبب استعمالاته الواسعة في مختلف العلوم الادارية منها والاقتصادية وكذلك الطبيعية ، فعلى سبيل المثال لا الحصر ، في المجال الاقتصادي ، يعد الاداة العلمية التحليلية في القياس الاقتصادي ، والاقتصاد الكلي التحليلي ، اذ من الممكن استعماله للتعبير عن العلاقات المستعملة في ربط المتغيرات الاقتصادية فيما بينها ، وذلك على شكل نماذج رياضية تسمى (نماذج الانحدار) ، اذ تم تقديم الانحدار عام (1886) من قبل كالتون (Galton) وذلك في مقالته التي التي تم فيها دراسة استقرارية توزيع الاطوال في المجتمع اذ استعمل عينة تزيد على الف عائلة. اذ اكدت النتائج وجود ميل لجميع الاباء طويلي القامة للحصول على اطفال طويلي القامة، بينما الاباء قصار القامة لهم ميل ان يحصلوا على أبناء قصار القامة ، وان متوسط طول الاطفال الذين ولدوا لاباء من طول معين يكون تحركه (انحداره) باتجاه المتوسط لطول اطفال المجتمع ككل . وقد تم استعمال مصطلح الانحدار من قبل كالتون للاشارة الى اتجاه الاطوال لاطفال المجتمع ككل نحو المتوسط العام ، وقد توالت وتعددت استعمالات هذا النوع من التحليل لتشمل مختلف جوانب الحياة .

2-2 : المفهوم الحديث للانحدار [3] : The modern interpretation of regression

يوجد إرتباط لموضوع تحليل الانحدار بدراسة الاعتمادية لمتغير معين (يعرف بالمتغير المعتمد) ، على متغير اخر او متغيرات اخرى يعبر عنها (المتغيرات التوضيحية) لاجل الحصول

على تقديرات، وكذلك التنبؤ بمتوسط المجتمع للمتغير المعتمد وذلك يتم بدلالة قيم معلومة (تكون ثابتة) للمتغير (او المتغيرات) التوضيحية ويحصل ذلك بتكرار العينة . وبشكل مبسط يعد تحليل

الانحدار طريقة احصائية لتحليل البيانات التي تتضمن متغيرين فأكثر عندما يكون الهدف هو اكتشاف العلاقة . ويعد تحليل الانحدار اكثر الطرائق الاحصائية شيوعا اذ يستعمل بكثرة في العلوم المختلفة لوصفه العلاقة بين المتغيرات بشكل معادلة.

3-2: الانحدار وعلاقته بالسبب [3]: Regression analysis and reason

يكون تعامل الانحدار بشكل اعتمادي بين احد المتغيرات و بين متغيرات اخرى تشخص مشكلة معينة ، مع ذلك فأن هذا لايعني بشكل مطلق استنتاج اتجاه السببية . ولقد تم التاكيد من قبل (كاندال وستيوارت) بأن العلاقات الاحصائية حتى وان كانت قوية فأنها لايمكنها ان تحدد الارتباط السببي لان الفكرة التي تعتمد عليها السببية تأتي من خارج الاحصاء، وكذلك تعتمد على نظريات اخرى . فعلا سبيل المثال دراسة الناتج لمحصول معين ، فانه لايوجد سبب احصائي لاجل افتراض ان المعدل الامطار لايعتمد على ناتج المحصول . وان اختيار الناتج وجعله متغيرا معتمدا اعتمادا على افتراضات هي غير احصائية . فالمنطق يقترح بأنه من غير الممكن ان تكون العلاقة بالشكل العكسي ، لانه من غير الممكن السيطرة على معدل الامطار اعتمادا على التغيرات في الناتج، ويمكن القول إن العلاقات الاحصائية في جميع الامثلة التي توجد في الانحدار من غير الممكن ان تحدد السببية، اذا لابد من التركيز على الجوانب االمسبقة والنظرية.

4-2 :انواع الانحدار [15] : Types of regression

يبنى التقسيم او التفرقة بين انواع الانحدار على اساس عدد المتغيرات المستقلة فالانحدار الذي يشتمل على متغير مستقل واحد ومتغير تابع واحد يطلق عليه الانحدار الخطي البسيط ويعد من ابسط انواع التحليل، ويسمى الانحدار الذي يتضمن اكثر من متغيرين مستقلين بالانحدار المتعدد.

1-4-2 :الانحدار الخطي البسيط [24] : Simple regression

يتالف هذا النوع من الانحدار من متغير معتمد واحد يرمز له (Y) ومن متغير واحد مستقل يرمز له (X) وتكون المعادلة :

$$Y=f(X, b) \qquad (2\text{-}1)$$

اذ ان (b_i) تمثل معلمات مجهولة فهي يمكن ان تكون ثوابت او تكون بصيغة متجه. وهي عبارة عن ثوابت غير معروفة تسمى المعلمات تتحكم في سلوك الأنموذج ويتم تقدير المعلمات اعتمادا على البيانات وان التعقيد الرياضي للنموذج والدرجة الواقعية له تعتمد على ماهو معروف من هذه العملية (عملية بناء الموذج) .

2-4-2 :الانحدار الخطي المتعدد [22] : Multiple Linear Regression

في الانحدار الخطي البسيط تقدر العلاقة بين اثنين من المتغيرات، اما في الانحدار المتعدد تكون المتغيرات مضاعفة، اي ان العلاقة ان وجدت فهي علاقة تتضمن اكثر من متغيرين وكذلك العلاقة ان وجدت فقد تكون غير خطية و يعد أنموذج الانحدار المتعدد الامتداد المنطقي والطبيعي للنموذج الخطي بمتغيرين ، فعندما يتم استعمال (K) من المتغيرات المستقلة ($X_1.X_2.\ldots.X_k$) لاجل تفسير التباين للمتغير المعتمد (Y) في معادلة الانحدار. فأن تعدد المتغيرات التوضيحية تجعل من التعامل مع طرائق الجبر الخطي اي (جبر المصفوفات) هي المستعملة لاجل تقدير واختبار ثم تحليل نماذج الانحدار المتعدد . ليتم بعد ذلك اعمامها ثم تطبيقها في حالة المتغيرين او

ثلاث متغيرات او اي عدد ممكن من المتغيرات ولكن بشرط ان لايتجاوز عدد المتغيرات عدد المشاهدات المستعملة للتقدير.

5-2 :الانموذج الخطي العام [28] : General Linear Model

يعد الانموذج الخطي اساسيا لكل تحليل لكل من البيانات الاحادية المتغيرات والمتعددة المتغيرات ، لنفترض ان المتغير المعتمد (Y) يمثل دالة خطية بدلالة (k) من المتغيرات التوضيحية ($X_1.X_2.\dots.X_k$) وبذلك يتم صياغة أنموذج الانحدار الخطي العام

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + \cdots + b_kX_{Ki} + u_i \qquad (2-2)$$

اذ ان :

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

u_i : تمثل قيمة المتغير العشوائي غير المعلومة

b_0 : يمثل الحد الثابت اومعامل التقاطع

b_j : عبارة عن المعالم للأنموذج التي تقوم بمهمة قياس استجابة المتغير التابع للمتغير المستقل.

المقصود مع بقاء المتغيرات التوضيحية الاخرى ثابتة وتمثل مقدار التغير في (Y) الناتج عن التغير بوحدة واحدة من (X_j) مع بقاء المتغيرات الاخرى ثابتة . ولعينة من المشاهدات (n) فأن العلاقة (2-2) تكون متحققة لكل مشاهدة فيكون لدينا (n) من المعادلات وكالتالي :

$$Y_1 = b_0 + bX_{11} + b_2X_{21} + \cdots + bX_{K1} + u_1 \qquad (2-3)$$

$$Y_2 = b_0 + b_1X_{12} + b_2X_{22} + \cdots + b_kX_{K2} + u_2 \qquad (2-4)$$

$$Y_n = b_0 + b_1X_{1n} + b_2X_{2n} + \cdots + b_kX_{Kn} + u_n \qquad (2-5)$$

وحينما يتم استعمال المصفوفات يتحول نظام المعادلات الى مايأتي :

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & . & . & . & X_{k1} \\ 1 & X_{12} & X_{22} & . & . & . & X_{k2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{1n} & X_{2n} & . & . & . & X_{kn} \end{bmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ . \\ . \\ . \\ b_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix}$$

ويمكن اختصار نظام المصفوفات اعلاه كما يأتي :

$$Y= X\beta+u \qquad (2\text{-}6)$$

اذ ان :

Y: متجه عمودي يمثل المشاهدات للمتغير المعتمد وبترتيب (n*1).

X: يمثل مصفوفة تتضمن المشاهدات للمتغيرات المستقلة ($X_1.X_2.\ldots.X_k$) وعمودها الاول يمثل متغيرا وهميا ليدل على المقطع الصادي. اذ ان اول صف من هذه المصفوفة هو ($1 \quad X_{11} \quad X_{21} \ldots.X_k$). وتسمى هذه المصفوفة مصفوفة المعلومات (Data matrix).

β: يمثل متجه عمودي يتضمن معالم الانموذج الخطي والتي يرغب في تقديرها وكالاتي : n (k+1)

u: يمثل متجه عمودي يتضمن القيم للمتغير العشوائي المجهولة ، وكالاتي : (n*1) .

في المعادلة (2-6) يمثل الطرف الايمن الجزء المتضمن المتغيرات التوضيحية (Xβ) يضاف لها الجزء الاحتمالي العشوائي (u).

6-2 : الأفتراضات [3] [5] Assumptions

الفرضية الاولى : المتوسط لمتجه المتغير العشوائي (u_i) يكون مساويا للصفر بمعنى اخر $[\ E(u_i)=0\ ،i= 1,..., n\ ]$

وبصيغة المصفوفات :

$$E\ (\underline{u}_i) = E\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ \cdot \\ u_n \end{pmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ E(u_3) \\ \cdot \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \cdot \\ 0 \end{pmatrix} \qquad (2-7)$$

الفرضية الثانية : التباين للمتغير العشوائي يكون متجانس ، ونعني بذلك ان تكون قيمة تباين لكل مشاهدة متساوية ويمكن التعبير عنها كما يأتي : $E(u_i u_j) = Var\ (u_i) = \sigma^2$ ، $i=j$ وهذا مايسمى فرضية ثبات تجانس التباين (Homoscedasticity).

الفرضية الثالثة : التباين المشترك يكون مساويا للصفر لأي مشاهدتين ضمن مشاهدات المتغير العشوائي وبعبارة أخرى $E(u_i u_j) =0,\ i\neq j$ وتسمى بفرضية عدم وجود إرتباط ذاتي. ومن الممكن التعبير عنها بصيغة المصفوفات من دمج الفرضيتين الثانية والثالثة وكالتالي :

$$E\ (\underline{u}_i \underline{u}_j) = E\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ \cdot \\ u_n \end{pmatrix} (u_1 \quad u_2 \quad \cdot \quad \cdot \quad \cdot \quad u_n)$$

وبعد ضرب المصفوتين المذكورتين آنفا ينتج :

$$E(\underline{uu'}) = E\begin{pmatrix} u_1^2 & u_1u_2 & u_1u_3 & \cdot & \cdot & \cdot & u_1u_n \\ u_2u_1 & u_2^2 & u_2u_3 & \cdot & \cdot & \cdot & u_2u_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ u_nu_1 & u_nu_2 & u_nu_3 & \cdot & \cdot & \cdot & u_n^2 \end{pmatrix}$$

وبعد ادخال التوقع على عناصر المصفوفة المذكورة آنفا تنتج المصفوفة الاتية :

$$E(\underline{uu'}) = \begin{pmatrix} Eu_1^2 & Eu_1u_2 & Eu_1u_3 & \cdot & \cdot & \cdot & Eu_1u_n \\ Eu_2u_1 & Eu_2^2 & Eu_2u_3 & \cdot & \cdot & \cdot & Eu_2u_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ Eu_nu_1 & Eu_nu_2 & Eu_nu_3 & \cdot & \cdot & \cdot & Eu_n^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

وللتبسيط تكتب بالشكل التالي :

$$E(\underline{u}\,\underline{u}') = \sigma^2 I_n \qquad (2-8)$$

تمثل المصفوفة بالعلاقة (2-8) مصفوفة التباين والتباين المشترك العائدة للمتغير العشوائي (u) بالشكل المصفوفي في هذه الصيغة فأن العناصر للقطر الرئيسي تمثل التباين لمشاهدات المتغير العشوائي . في حال ان عناصر المثلث السفلي تكون مساوية لعناصر المثلث العلوي .

الفرضية الرابعة : المتغيرات $(X_1 . X_2 . \ldots . X_k)$ تمثل متغيرات ثابتة للعينة التي تم اختيارها فهي متغيرات غير عشوائية ومعنى هذا ان المتغيرات التوضيحية هي متغيرات خارجية (Exogenous) ، وبذلك تكون المصفوفة (X) والتي لها ترتيب n*(k+1) مصفوفة غير عشوائية اي انها تحتوي ارقاما ثابتة .

الفرضية الخامسة : عدم وجود علاقة خطية تامة بين تلك المتغيرات التي تقوم بتمثيل المصفوفة (X) ومعنى هذا عدم وجود لمشكلة التعدد الخطي (Multicollinearity). ويمكن كتابة الفرضية مجزئة وكالاتي :

$$r_{X_0 X_j} = 0 \quad \forall \quad j = 1.2 \ldots . k \qquad (a)$$

ومعنى هذا ان العمود الاول مستقل مع اي من الاعمدة للمصفوفة الاخرى ، وبشكل عملي يعني ذلك وجود تغيرات بشكل واضح وملموس في مشاهدات اي متغير من المتغيرات التوضيحية.

$$COV\,(X_i . X_j) = 0$$

$$r_{X_0 X_j} = 0 \quad \forall \quad i \neq j \ , \ \ i , j = 1,2,\ldots, k \qquad (b)$$

وهذا يعني افتراض استقلالية المتغيرات التوضيحية عن بعضها الاخر والعمود الثاني والثالث و..... والعمود (k+1) تكون مستقلة عن بعضها الاخر . ويمكن ان نلخص هذه الفرضية بالصيغة المصفوفية الاتية :

$$P(X) = k+1 < n \qquad (2\text{-}9)$$

الفرضية السادسة : المشاهدات للمتغير العشوائي (u) يكون توزيعها بشكل متماثل اي انها تمتلك توزيع طبيعي (Normal) . ويمكن التعبير عن الفرضيات (الثانية والثالثة والسادسة) مصفوفيا بالصياغة الاتية : $u \sim N(0 , \sigma^2 I_n)$.

7-2 : تقدير المعالم بطريقة [27] : Parameter Estimating by OLS

عندما نستعمل عينة عشوائية من (n) من المشاهدات $Y_1. Y_2. \ldots Y_n$ واخرى من المشاهدات لمتغيرات مستقلة فمن الممكن تقدير كل من المعلمات $(b_0. b_1 \& \sigma^2)$ وكالاتي :

اولا:المقدر $\hat{\beta}$: Estimated $\hat{\beta}$

تستند طريقة المربعات الصغرى الاعتيادية الى معيار مهم وهو جعل المجموع لمربعات الخطأ والتي تعرف بالبواقي اقل مايمكن . وان البواقي تمثل قيم (Y) الحقيقية تطرح منها القيم التقديرية $(\hat{Y})$

وبعبارة اخرى : $e_i = Y_i - \hat{Y}_i \quad \forall \quad i = 1.2. \ldots . n$

وفي حال استعمال المصفوفات تكون :

$e = Y - \hat{Y}$

وبذلك تكون الدالة بصيغة المصفوفات : $\sum_{i=1}^{n} e_i^2 = \underline{e}' \, \underline{e}$

$$\sum_{i=1}^{n} e_i^2 = (e_1 \;\; e_2 \;\; \cdots \;\; e_n) \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + \cdots + e_n^2$$

$$\underline{e}'\,\underline{e} = \left(\underline{Y} - X\underline{\hat{\beta}}\right)' \left(\underline{Y} - X\underline{\hat{\beta}}\right)$$

$$\underline{Y}'\,\underline{Y} - 2\hat{b}'X'\underline{Y} + \underline{\hat{b}}'X'X\,\underline{\hat{b}} \qquad (2-10)$$

اذ ان ($\underline{\hat{b}}'X'\underline{Y}$) تمثل متجه بدرجة (1*1) ثابت او تمثل عدد حقيقي ويكون بذلك :

$$\underline{Y}'X\underline{\hat{b}} = (\hat{b}'X'Y)' = Y'X\hat{b}$$

العلاقة (2-10) هي الدالة التي نسعى ان تكون صغيرة وهي بدلالة متجه المعلمات $\hat{b}$ ·

وبالاشتقاق الدالة بالنسبة للمعلمة ($\hat{b}$) ومساواة الدالة للصفر وكالاتي :

$$\frac{\partial e'e}{\partial \underline{\hat{b}}'} = 0$$

$$\frac{\partial e'e}{\partial \underline{\hat{b}}'} = -2X'Y + 2X'X\hat{\beta} = 0 \qquad (2-11)$$

$$X'X\,\underline{\hat{b}} = X'\underline{Y} \qquad (2-12)$$

المصفوفة $X'X$ تعرف بمصفوفة فيشر للمعلومات وهي بدرجة (K+1) * (K+1) وهي ايضا مصفوفة مربعة (Square) كذلك هي متماثلة (Symmetric) وهي مصفوفة غير شاذة (Nonsingular) وذلك لكونها تمتلك رتبة تامة من جهة الاعمدة وحسب المعادلة (2-9) ولهذا فأن معكوسها موجود $(X'X)^{-1}$. ثم يضرب طرفي العلاقة (2-12) بالمقدار $(X'X)^{-1}$ من جهة اليسار وبتبسيط الطرفين نحصل على القانون للمربعات الصغرى الاعتيادية وبصيغة المصفوفات وكالتالي :

$$\underline{\hat{b}} = (X'X)^{-1}X'Y \qquad (2-13)$$

ثانيا :المقدر σ^2 : **Estimated σ^2**

لاينتج اسلوب المربعات الصغرى دالة من قيم (X) وقيم (Y) من العينة التي يمكننا تقليلها حتى نحصل على قيم (S^2) . ومع ذلك يمكننا استنتاج مقدرات غير متحيزة لل(S^2) استنادا لمقدرات المربعات الصغرى ($\hat{b}$) وكالاتي :

$$\sigma^2 = E[Y_i - E(Y_i)]^2 \qquad (2-14)$$

$$E(\underline{Y_i}) = b_0 + b_1x_{1i} + b_2x_{2i} + \cdots + b_ix_{ki} = x_i'\underline{b} \qquad (2-15)$$

اذ ان X_i' هو صف (i) ضمن المصفوفة (X)، هكذا يصبح (σ^2)

$$\sigma^2 = E[Y_i - x_i'\beta]^2$$

ويتم تقدير $\hat{\sigma}^2$ بواسطة المتوسط المطابق والماخوذ من العينة

$$S^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(\underline{Y_i} - X_i'\underline{\hat{\beta}})^2 \qquad (2-16)$$

اذ ان (n) هو حجم العينة و (k) هو عدد من (X's) ، ($X_i'\hat{b}$) هي افضل تقدير خطي غير متحيز من ($x_i'b$) وباستعمال معادلة (2-10) يمكن كتابة معادلة (2-16) كالاتي :

$$S^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(\underline{Y} - X\underline{\hat{b}})'(Y - X\underline{\hat{b}}) \qquad (2-17)$$

$$= \frac{\underline{Y}'\underline{Y} - \widehat{b'}X'Y}{n-k-1} \qquad (2-18)$$

اذ ان :

$$SSE = \left(Y - X\hat{b}\right)'\left(Y - X\hat{\beta}\right) = Y'Y - \hat{b}'X'Y \cdot$$

وبدرجة حرية (n-k-1)، وحسب العلاقة (2-16)،(2-17) او (2-18) واذ كان $E(Y) = X\beta$ وكذلك $Cov(Y) = I$ فأن :

$$E(S^2) = \sigma^2 \qquad (2-19)$$

8-2 :المربعات الصغرى المعممة [5] Generalized Least-Squares (GLS)

ومن دواعي استعمال هذه الطريقة هو تصحيح كفاءة المربعات الصغرى عند مخالفة فرضية ثبات التباين، وتعرف هذه الطريقة في بعض الاحيان بطريقة المربعات الصغرى الموزونة اذ يتم الحصول على المعلمات للانحدار عن طريق اعطاء وزن لكل مشاهدة وفقا لحجم تباين البواقي ففي حالة اختلال فرض ثبات التباين فأن مصفوفة التباين والتباين المشترك الخاصة بالمتغير العشوائي، يتم تحديدها وفق التالي:

$$Var - Cov(u) = E = \sigma^2\Omega \qquad (2-20)$$

اذ ان (Ω) هي مصفوفة مربعة موجبة وبترتيب (n*n) وعلى هذا الاساس فأن التقدير للانموذج الخطي لايمكن ان يتحقق بأستعمال طريقة المربعات الصغرى (OLS) تعطي مقدرات تكون غير كفوءة ومن ثم يكون من الواجب استعمال طرائق التحويل التي تكون مناسبة ، في الانحدار الخطي العام .

$$Y=X\beta+u$$

وحسب الافتراض

$$E\,(uu')\neq\sigma^2 I_n$$

وانما يكون

$$E\,(uu')=\sigma^2\Omega$$

اذ ان (Ω) هي مصفوفة مربعة موجبة .

وبالاعتماد على نتائج التحويل بواسطة صيغ المصفوفات يصبح من الممكن تحويل المتغير العشوائي ويتم ذلك بالاستناد الى القاعدة :لأي مصفوفة موجبة (Ω) ممكن ايجاد مصفوفة غير شاذة (P)

إذ إن :

$$PP'=\Omega \qquad (2-21)$$

إذ إن :

$$P=\begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_3}} & \dots & 0 \\ \vdots & \vdots & \vdots & \frac{1}{\sqrt{\lambda_4}} & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix}$$

ومن الممكن الحصول على قيمة P^{-1} وكما يأتي:

$$\underline{P}^{-1} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \dots & 0 \\ 0 & 0 & \sqrt{\lambda_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \sqrt{\lambda_4} & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{\lambda_n} \end{bmatrix}$$

وعليه فأن $(PP')^{-1} = \Omega^{-1}$

$$(PP')^{-1} = \Omega^{-1} = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \lambda_4 & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

وبذلك يتم تحويل الانموذج ($Y=X\beta+u$) ويتم ذلك بضرب طرفي المعادلة بالمقدار(P^{-1}) للحصول على :

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \qquad (2-22)$$

ولنفترض ان :

$$P^{-1}u = u^*. \quad P^{-1}X = X^* \quad . \quad P^{-1}Y = Y^*$$

ومن ثم يكون الانموذج المحول :

$$Y^* = X^*\beta + u^* \qquad (2-23)$$

علما ان u^* متغير عشوائي له متوسط مساوي للصفر وتباين يحسب كالاتي :

$$Var(u^*) = Var(P^{-1}u)$$

$$= P^{-1}Var(u)(P^{-1})'$$

$$=\sigma^2 P^{-1}\Omega(P^{-1})' = \sigma^2 P^{-1}(PP')(P^{-1})'$$

$$\text{Var (u)} =\sigma^2 P^{-1}PP'(P')^{-1} = \sigma^2 I_n \qquad (2-24)$$

وبذلك فالانموذج المحول تم تنقيته من عدم تجانس المتغير العشوائي وبذلك اصبحت مصفوفة التباين والتباين المشترك الخاصة بالمتغير العشوائي الذي تم تحويله (u^*) عبارة عن ثابت (σ^2) وهي مضروبة بمصفوفة الوحدة (I_n) ، وعلى هذا الاساس يصبح تطبيق طريقة المربعات الصغرى (OLS) على الانموذج المحول يسهم في اعطاء أفضل تقدير خطي غير متحيز (BLUE) للمعلمة β . وبذلك فأن المعيار لطريقة المربعات الصغرى المعممة هو تقليل (تصغير) مجموع مربعات البواقي للنموذج المحول .

$$\text{Min} \sum u^* u^{*\prime} \qquad (2-25)$$

أو بصيغة اخرى

وباستعمال المصفوفات :

$$\text{Min } (P^{-1}e)\ (P^{-1}e)'$$

بعبارة اخرى

$$\text{Min } P^{-1}e\ e'(P')^{-1}$$

ولاجل الحصول صيغة لحساب المعلمة ($\hat{\beta}_{GLS}$) يتم اجراء بعض الخطوات الرياضية و كمايأتي :

$$X^{*\prime}X^*\beta_{GLS} = X^{*\prime}Y^* \qquad (2-26)$$

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

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

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

$$X'\Omega^{-1}X\hat{\beta}_{GLS} = X'\Omega^{-1}Y \qquad (2-27)$$

وبذلك اصبحت معلمة الانحدار حسب طريقة (GLS) تحسب كالاتي :

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \qquad (2-28)$$

9-2 :خصائص المعلمات المقدرة [32]: Estimated parameters

في هذا الموضوع سيتم مناقشة خصائص المربعات الصغرى الاعتيادية ويتم ذلك عبر نقاط وكالاتي :

1- المعلمات المقدرة هي تقديرات غير متحيزة لقيم معلمات المجتمع الحقيقية. ويمكن توضيح ذلك بصيغة المصفوفات وكما يأتي :

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

$$=(X'X)^{-1}X'(Xb+u)$$

$$=(X'X)^{-1}X'Xb+(X'X)^{-1}X^{-1}u$$

$$=b+(X'X)^{-1}X'u \qquad (2-29)$$

وبأخذ التوقع للطرفين :

$$E(\hat{\beta}) = \beta+(X'X)^{-1}X'E(u)$$

وأن $E(u) = 0$

$E(\hat{\beta}) = \beta$

$$E(\underline{\hat{b}}) = E\begin{pmatrix} \underline{\hat{b}}_0 \\ \underline{\hat{b}}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\hat{b}}_k \end{pmatrix} = \begin{pmatrix} \underline{b}_0 \\ \underline{b}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{b}_k \end{pmatrix}$$

2- يعد المتجه $\underline{\hat{b}}$ تركيب خطي بدلالة المتجه $\underline{Y}$ وان المتغير X يكون ثابت للعينة المختارة وهذا يعني ان $(X'X)^{-1}$ X تكون ثابتة ايضا وهذا يؤكد إن:

$$\underline{\hat{b}} = (X'X)^{-1}X \qquad (2-30)$$

3- المعلمات التي تم تقديرها يكون لها اقل تباين وعلى هذا الاساس تكون المقدرات بطريقة (OLS) افضل المقدرات الخطية غير المنحازة او تعرف احيانا بأفضل تقدير خطي غير متحيز (BLUE).

10-2 : إختبار الفرضيات [12] : Hypotheses testing

يعد اختبار الفرضيات من الموضوعات الاساسية في الاستدلال الاحصائي وبواسطته يستطيع الباحث الخروج باستنتاجات للمشكلة قيد البحث والفرضية التي يقوم الباحث بوضعها عبارة عن تفسيريحتمله الباحث للعوامل المراد دراستها وفحصها، وهذا الاستنتاج تارة يكون صائبا او يكون خاطئا ومن المفترض توفر بعض الشروط في الفرضية من اهمها:

1- إن تكون صياغة الفرضية بسيطة ومفهومة.

2- إن تكون الفرضية قابلة للقياس وتوضح العلاقة بين متغيرات البحث.

3- أن تكون للفرضية قوة تفسيرية.

4- أن يكون للفرضية إجابة محددة وواضحة.

5- أن تكون الفرضية منسجمة مع الحقائق التي تم التوصل اليها عن طريق البحث العلمي سواء بالجانب النظري أم الجانب التطبيقي .

11-2 : مفهوم معنوية الاختبار [12] : Significance Concept Test

ان البحوث التي تعتمد على الطريقة العلمية تقوم على أساس معرفة صحة الفرضيات التي يفترضها الباحث فبعد ان يقوم الباحث بتصميم التجربة وجمع المعلومات الكاملة واستخراج المؤشرات الإحصائية ينتقل إلى عمليتي التحليل واختبار صحة فرضياته ، وان الاختبارات التي تجري لمعرفة صحة الفرضية من عدمها تسمى بالاختبارات المعنوية ،وهي تعمل على معرفة درجة ودقة النتائج التي تم التوصل إليها .

12-2 :اختبار (t) للمعلمات [3] [2] : (t) Test

يستعمل اختبار (t) لاجل تقيم المعنوية لتأثير المتغيرات التوضيحية $(X_1.X_2.\dots..X_k)$ في المتغير التابع (Y) في انموذج الانحدار المتعدد، ويستعمل هذا الاختبار عند مستوى معنوية معين ودرجة حرية (n-k)، وان اختبار (t) في انموذج الانحدار المتعدد يعتمد على نوعين من الفرضيات

فرضية العدم : $H_0: B_1 = B_2 = B_3 = \cdots = B_K = 0$

الفرضية البديلة : $H_1: B_j \neq 0 \qquad j = 1.2 \dots . k$

وبعد الانتهاء من حساب قيمة (t) يجرى مقارنتها مع القيمة الجدولية ليتم اتخاذ القرار برفض او عدم رفض فرضية العدم لاجل تقييم المعلمات للأنموذج المقدر، والصيغة الرياضية هي كما يأتي :

أولا: اختبار معنوية المعلمة (B_i) :

$$t_{\hat{b}_i} = \frac{\hat{b}_i}{S_{\hat{b}_i}} \qquad (2-31)$$

اذ ان :

$$S_{\hat{b}_i} = \sqrt{S^2_{\hat{b}_i}}$$

$$S^2_{\hat{b}_i} = Var(\hat{b}_i) = S^2 e\, a_{11}$$

$$Var(\hat{b}) = S^2 e (X'X)^{-1}$$

$$S^2 e = \frac{e'e}{n-k-1} = \frac{Y'Y - \hat{B}'X'Y}{n-k-1}$$

$$= \frac{\sum y^2 - (\hat{B}_1 \sum x_1 y + \hat{B}_2 \sum x_2 y)}{n-k-1} \qquad (2-32)$$

المبحث الثاني

مشكلة عدم تجانس التباين

Heteroscedasticity

13-2: طبيعة المشكلة [5] [4] [3] [20] : Nature of the problem

من الافتراضات المهمة هو مايحدده احد فروض طريقة المربعات الصغرى (OLS) اي ان يكون تباين حد الخطأ ثابتا ويكون متساويا لجميع مشاهدات العينة بمعنى اخر:

$$\sigma_u^2 = E(u_i)^2 = Var\ (U_i) = E[u_i - E(u_i)]^2 \qquad (2\text{-}33)$$

وهذا الفرض يسمى بفرض تجانس تباين الخطأ (Homoscedasticity) .وتشير كلمة (homo) الى معنى التساوي وكلمة (Scedasticity) الى معنى الانتشار ومعنى هذا الفرض ان احتمالية توزيع حدود الخطأ العشوائي (u_i) تبقى نفسها على جميع مشاهدات قيم (X_i) اي ان تباين حد الخطأ

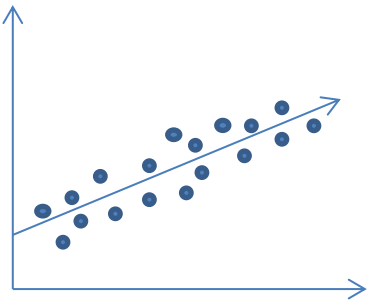
(u_i) هو نفسه لكل قيم المتغير المستقل (التوضيحي) . وعندما لا تتحقق هذه الفرضية في اي حالة معينة فأن هذا يؤدي الى مخالفة فرض التجانس اي ان تباين العينة اصبح يختلف باختلاف مشاهدات العينة وهذا ما يدعى بعدم ثبات تجانس تباين الخطأ (heteroscedasticity) . كذلك اذا كانت فرضية التجانس للتباين غير متحققة ، فأن ذلك يؤدي الى ان تكون مصفوفة التباين – التباين المشترك للأخطاء معرفة كالاتي:

$$\Omega_\varepsilon = E\ (\varepsilon\varepsilon') = \begin{pmatrix} \sigma_{\varepsilon.1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon.2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\varepsilon.n}^2 \end{pmatrix} \neq \sigma_\varepsilon^2 I_n \qquad (2\text{-}34)$$

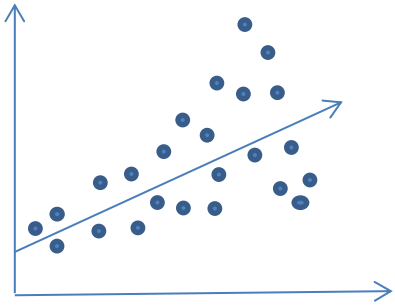
نلاحظ ان تباينات الاخطاء غير ثابتة على القطر الرئيسي ويؤدي هذا الى ان يكون تباين الاخطاء

مرتبط بقيم المتغير المستقل . ويمكن توضيح شكل العلاقة المتوقعة بين المتغير التابع والمتغير التوضيحي في حالة ثبات التباين وفي حالة عدم ثبات التباين كالاتي :

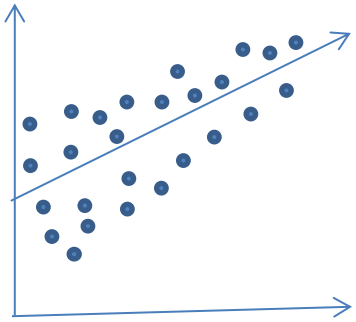
شكل (1) يوضح حالات ثبات وعدم ثبات التباين



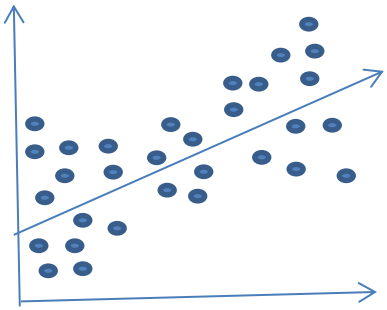
شكل (1-A) يوضح حالة ثبات التباين التباين



شكل (1-B) يوضح حالة عدم ثبات



شكل (1-C) يوضح حالة عدم ثبات التباين مع زيادة في المتغير التوضيحي(x_i)



شكل (1-D) يوضح حالة عدم ثبات التباين مع نقصان في المتغير التوضيحي (x_i)

المصدر : **[26]**

14-2 : اسباب مخالفة فرض ثبات التباين : Causes of the problem

هنالك العديد من العوامل التي تؤدي الى مخالفة فرض تجانس التباين (تغير تباين u_i) منها **[3]** **[5]** :

اولا : التعلم او نماذج خطأ التعلم (Learning Error models) ونقصد به سلوكية وتصرف الافراد التي تقل الاخطاء فيها بمرور الزمن ومن ثم فأن تباين (σ_i^2) يتناقص ايضا خلال المدة

الزمنية مثال ذلك عدد الاخطاء على ماكنة معينة فأنها تقل بزيادة الخبرة من قبل العامل خلال الزمن.

ثانيا : يزداد التباين لحد الخطأ (σ_i^2) وذلك بزيادة مستوى الدخل من لدن الافراد وذلك عائد الى اختلاف وتعدد الناس في اختياراتهم مثال ذلك تباين الانفاق على المواد الغذائية يمكن ان يزيد بزيادة دخل الاسرة .

ثالثا : مع تحسن اساليب جمع البيانات فأن تباين حد الخطأ (σ_i^2) يقل ،لان جمع البيانات الدقيقة والتي تكون قريبة من الواقع تقلل الاخطاء مثال ذلك حساب واردات البلد بالاعتماد على الحاسب الالي تكون ادق من التي يتم حسابها بطريقة يدوية.

رابعا : تحدث المشكلة احيانا بسبب وجود القيم (الشاذة)، فالقيمة الشاذة تعد مشاهدة مختلفة بشكل يمكن ملاحظته من بين بقية المشاهدات (سواء أكانت كبيرة جدا ام كانت صغيرة جدا)، بعبارة ادق المشاهدة الشاذة هي مشاهدة مأخوذة من مجتمع يختلف عن المجتمع الذي تم توليد منه المشاهدات البقية، وان وجود او حذف مثل هذا النوع من البيانات لاسيما عندما يكون حجم العينة صغيرا فمن الممكن ايكون هذا مؤثرا في نتائج تحليل الانحدار .

خامسا : قد يكون سبب الاختلاف هو التوصيف الخاطئ لانموذج الانحدار، فالعديد من اختلاف التباين يكون بسبب حذف متغيرات مهمة من الانموذج، ومن ثم فأن البواقي التي يتم الحصول عليها من أنموذج الانحدار قد توحي ان تباين الاخطاء غير ثابت .

سادسا : من مصادر اختلاف التباين قد يكون التواء توزيع متغير واحد او اكثر من المتغيرات المنحدرة والتي توجد ضمن الانموذج .

سابعا : من الاسباب الاخرى لاختلاف التباين هو التحويل الخاطئ للبيانات او بسبب عدم صياغة الدالة بشكل سليم.

15-2 : تفسير المشكلة [3] : problem Interpretation

عبارة فرض ثبات تجانس التباين (homoscedasticity) تعني ان تباين المتغير العشوائي (u_i) عن وسطه الحسابي الذي يكون مساويا للصفر $E(u_i)=0$ لا يعتمد على قيم المتغير التوضيحي (X_i)، وهذا يشير الى ان تباين (u_i) يبقى نفسه بالرغم من صغر او كبر قيم المتغير التوضيحي لذلك فأن تباين (σ_u^2) ليس دالة بالمتغير (X_i)، اي ان $\sigma_u^2 \neq f(X_i)$ وتظهر هذه الحالة بشكل بياني عندما يكون توزيع او انتشار قيم المتغير العشوائي (u_i) بمسافات ثابتة عن خط معادلة الاتجاه العام. وفي حالة كون تباين المتغير العشوائي (σ_u^2) غير ثابت فأن ذلك يعني ان تباين المتغير العشوائي يعتمد على قيم المتغير التوضيحي (X_i) اي ان $\sigma_u^2 = f(X_i)$ وأن حالة عدم ثبات تجانس التباين (heteroscedasticity) تتوضح عن طريق تزايد او تناقص أنتشار المشاهدات على خط معادلة الاتجاه العام (خط الانحدار) وأن نوع المشاهدات في الشكل الانتشاري يعتمد على شكل وكيفية عدم الثبات ونقصد به شكل العلاقة بين تباين المتغير العشوائي وقيم المتغير المستقل في حالة عدم ثبات تجانس تباين الخطأ العشوائي .

16-2 : الثبات وأثارالمخالفة [5]: Stability and the effects of the violation

ان الفروض الكلاسيكية لحدود الخطأ هي المتوسط المساوي للصفر والتباين ثابت والتغايرصفري اي بمعنى :

$$E(u_i) = 0 \quad ، \quad Var(u_i) = \sigma_u^2 \quad ، \quad Cov(u_i، u_j) = 0$$

وعند مخالفة الفرضيتان (المتوسط المساوي للصفر والتغاير الصفري) فأن ذلك يجعل مقدرات (OLS) غير متحيزة ومتسقة ، اما فرض ثبات التباين فهو فرض مهم لجعل (OLS) مقدرا كفؤا، اذ ان مخالفة هذا الفرض تجعل من غير الممكن الادعاء بأن المقدر هو افضل مقدر خطي غير

متحيز(Blue) بين المقدرات الخطية غير المتحيزة ، اي من الممكن ايجاد مقدر خطي غير متحيز اخر اكثر كفاءة .

ان مخالفة الفروض آنفا يمكن التعبير بما يأتي :

$E(u_i^2)\neq\sigma^2$، $E(u_iu_j)\neq0$ ، $E(uu')\neq\ \sigma_u^2 I_n$ ، $\sigma_1^2\neq\ \sigma_2^2\neq\ldots\neq\sigma_n^2$

لذا فأن تقدير معالم الانحدار بطريقة (OLS) لا يمكن ان تكون افضل تقدير خطي غير متحيز (Blue) . ان الفروض المذكورة آنفا الخاصة بثبات تجانس التباين ان لم تتحقق فسيترتب على ذلك عدة نتائج او اثارمنها :

أولا : من غير الممكن تطبيق صيغة تباين المعلمات لغرض اختبار المعنوية وبناء فترة الثقة ، اذ ان الاختبار هنا لا يمكن تطبيقه لان قيمة (σ_u^2) اختلفت ، اذ ان صيغ التباين للمعلمات كالاتي :

$$Var\ (b_1) =\sigma_u^2\ \frac{1}{\sum x^2_i} \qquad (2-35)$$

$$Var\ (b_0)=\sigma_u^2\left[\frac{1}{n}+\frac{\overline{X^2}}{\sum x^2}\right] \qquad (2-36)$$

و المشكلة نفسها يتعرض لها اختبار (F) واختبار (t) التي تعتمد على هذه المعايير اذ تكون هذه الاختبارات اقل دقة .

ثانيا : لا تمتلك مقدرات (OLS) خاصية اقل تباين في فئة المقدرات غير المتحيزة لذا تكون المقدرات غير كفؤة في العينات الكبيرة والصغيرة .

ثالثا : تكون مقدرات (OLS) غير متحيزة احصائيا معنى ذلك ان القيمة المتوقعة تكون مساوية للقيمة الحقيقية ($E\hat{b}=b$) ومن المعروف ان خاصية عدم التحيز لا تتطلب الثبات في حد الخطأ .

رابعا : يصبح التنبؤ غير كفوء بسبب اعتماده على تقدير المعالم ($\hat{b}s'$) في البيانات والتي لها تباين كبير بسبب وجود مشكلة عدم التجانس.

المبحث الثالث
الاختبارات المعلمية المستعملة للكشف عن مشكلة عدم تجانس التباين [9].
Parametric tests to detect heteroscedasticity

تعد الاختبارات الاحصائية المعلمية واحدة من اهم التطبيقات التي قدمها علم الاحصاء،اذ تعتمد بشكل اساسي على معلمات المجتمع اعتماداً على ما توافر من بيانات من لدن الباحث خاصة بالعينة المأخوذة من هذا المجتمع وتستعمل الأختبارات المعلمية في حالة العينات الكبيرة التي يشترط فيها توافر المعلومات من مجتمعاتها (معلومات الاصل) مثل :

1- ان يكون توزيع البيانات توزيعاً طبيعياً.

2- التباينات تكون متجانسة.

3- العينات مأخوذة بصورة عشوائية.

4- العلاقة تكون خطية.

5- إستقلال العينات.

17-2 طرائق الكشف عن المشكلة [2] [3] : Methods of proble detection

في حالة عدم وجود اي معلومات عن طبيعة مشكلة عدم التجانس ،عند ذلك تكون طريقة الرسم باستعمال البواقي (والتي يمكن عدها تقدير للمتغير العشوائي (u) ولاسيما مع كبر حجم العينة) والتي تدعى تحليل البواقي (Residual Analysis) ، تعد مدخلا اوليا لمعرفة ذلك اذ يتم اجراء الانحدار بافتراض عدم وجود مشكلة عدم تجانس التباين ويتم حساب البواقي ($e_i = \hat{u_i}$) اذ يرسم مربع البواقي (e_i) على المحور العمودي و قيم (X_i) او قيم ($\hat{Y}$) على المحور الافقي . ويحدد وجود او عدم وجود مشكلة عدم التجانس والنمط المناسب لعدم التجانس ، ومن الجدير بالذكر ان استعمال تحليل البواقي يسهم في تحديد نمط عدم التجانس وبذلك تتم الخطوة الاولى في حل المشكلة وذلك عن طريق معرفة النمط لعدم تجانس التباين ليتم بعد ذلك تحويل المشاهدات . وتعد هذه الطريقة تأشيريه بسبب اعتمادها على الحكم الشخصي ولاسيما مع محدودية مشاهدات

العينة ، ومن الضروري ان تقترن هذه الطريقة مع اختبارات احصائية وهذه الاختبارات الاحصائية متعددة فمنها يعتمد على التوزيع الطبيعي والبعض الاخر عام ومن اهم هذه الاختبارات الاحصائية :

1-17-2: اختبار كولد فيلد- كواندت [2] : Gold field-Quant test

يحتسب تباين الخطأ للعينة الجزئية الاولى (S_i^2 1) ، وكذلك للعينة الجزئية الثانية (S_i^2 2) وحسب الصيغ الاتية يعد من الاختبارات المهمة والتي تستعمل للكشف عن مشكلة عدم التجانس لتباين الخطأ ويستعمل في حالة العينات الكبيرة الحجم والخطوات هي كالاتي :

1- ترتب البيانات الخاصة بالمتغير التوضيحي (X_i) ترتيبا تصاعديا.

2- يتم تجزئة العينة الى عينتين جزئيتين تتضمن الأولى المشاهدات الصغيرة والثانية المشاهدات الكبيرة وفي حالة عدم تساوي العينتين يتم حذف القيم الوسطية.

3 - يتم قسمة المشاهدات الباقية بشكل عينتين جزئيتين متساويتين تضم الاولى على قيم (X_i) الاولى (الصغيرة) مع قيم المتغير المعتمد (Y_i) التي تقابلها اما العينة الاخرى فتضم قيم (X_i) الاتية (الاكبر) مع قيم المتغير المعتمد (Y_i) الاتية :

4 - يتم تقدير معلمات العلاقة الخطية ($\hat{b}_2 . \hat{b_1}$) بين المتغير المستقل والمتغير التوضيحي لكل عينة جزئية بشكل منفرد.

5 - يتم حساب S_i^2 1 . S_i^2 2 حيث أن:

$$S_i^2\ 1 = \frac{\sum e_i^2}{T_1 - 2} \qquad (2-37)$$

$$S_i^2\ 2 = \frac{\sum e_i^2}{T_2 - 2} \qquad (2-38)$$

6- تحسب إحصاءة (F^*) اعتمادا على الصيغة الاتية:

$$F = \frac{S_i^2 2}{S_i^2 1} \qquad (2-39)$$

بعد ايجاد (F^*) المحتسبة نقارنها مع (F) الجدولية عند مستوى معين من المعنوية وبدرجة حرية مقدارها ($T_2 - 2$) للبسط و($T_1 - 2$) للمقام فاذا كانت القيمة المحتسبة اصغر من القيمة الجدولية نأخذ بفرضية العدم والتي تؤكد على عدم وجود المشكلة, اما في حالة كون قيمة (F^*) المحتسبة اكبر من قيمة (F) الجدولية عند نفس المستوى من المعنوية ودرجة الحرية فعند ذلك نأخذ الفرضية البديلة :

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$

$H_1:$ At least two differ

2-17-2: اختبار بروش- بيجين – جود فري [2] [3]: BPG test

يعد هذا الاختبارمن الاختبارات المهمة المستعملة للكشف عن وجود مشكلة عدم تجانس تباين الخطأ في الانموذج الخطي المتعدد اذ تم اقتراح هذا الاختبار عام (1979):

$$Y = b_0 + b_1X_1 + bX_2 + \cdots + b_mX_m + u \qquad (2-20)$$

اذ يفترض ان :

$$\sigma_i^2 = f(\alpha_0 + \alpha_1 z + \cdots + \alpha_k z_k) \qquad (2-21)$$

اذ ان (z) تمثل بعض او كل المتغيرات التوضيحية (X_i).

وفي حال ان : $\alpha_1 = \alpha_2 = \cdots = \sigma_k = 0$ فإن ($\sigma_i^2 = \alpha_0$) وهذا معناه ان التباين ثابت (constant) اما خطوات الاختبار فهي كالاتي :

1- يتم اجراء انحدار(Y) على كل (X'S) للحصول على الاخطاء ($e_1. e_2. \ldots . e_n$).

2- يتم تقدير ($\widehat{\sigma}_u^2 = \frac{\sum e_i^2}{n}$) والتي تمثل تقدير للإمكان الاعظم لتباين الخطأ.

3- يتم تعريف: ($P_i = \frac{e_i^2}{\widehat{\sigma}^2}$).

4- يتم تقدير P_i على (Z' S):

$$P_i = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_m z_m + v_i \qquad (2-42)$$

5- الحصول على مجموع المربعات الغير المشروحة (ESS) وذلك مع فرض تحقق التوزيع الطبيعي للمتغير العشوائي (u).

6- يتم تعريف الاحصاءة (BPG) كالاتي: $BPG = \frac{1}{2}ESS$ وان الاحصاءة (BPG) تتوزع حسب توزيع مربع كاي ($BPG \sim X^2_{n-1}$). في حالة $BPG > x^2_c$ فهذا يؤدي الى رفض فرضية العدم (H_0) التي تنص على عدم وجود مشكلة عدم تجانس التباين وهذا بدوره يشير الى وجود مشكلة عدم تجانس التباين، علما ان هذا الاختبار يشترط التوزيع الطبيعي للبيانات.

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$

$H_1:$ At least two differ

2-17-3: اختبار وايت [3] : White test

يعد من الاختبارات المهمة للكشف عن وجود مشكلة عدم تجانس التباين من عدمها ، اذ تم اقتراح هذا الاختبار عام (1980) وهو اختبار عام بينما كان اختبار كولد فيلد – كوانت) يحتاج الى معرفة اي متغير توضيحي هو الذي يسبب المشكلة ، كذلك نجد اختبار بروش- بيجين – جود فري (BPG) يفترض ان يكون التوزيع للمتغير العشوائي توزيعا طبيعيا .

$$Y = b_0 + b_1 X_1 + bX_2 + \cdots + b_m X_m + u \qquad (2-23)$$

ويمكن تلخيص خطوات هذا الاختبار كالاتي :

1- يتم تقدير الانموذج (معادلة 2-23) للحصول على البواقي (e_i).

2- إجراء انحدار (e^2) على $X_1 X_2 . X_2^2 . X_1^2 . X_2 . X_1$ والثابت

$$e_i^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + v_i \qquad (2-24)$$

3- حساب قيمة معامل التحديد (R^2) من معادلة الانحدار، بعد ذلك نجد (nR^2) اذ تمثل (nR^2) احصاءة الاختبار اذ تتوزع حسب توزيع مربع كاي: $nR^2 \sim \chi_5^2$ وبشكل عام فأن $nR^2 \sim \chi^2 \frac{k(k+3)}{2}$ ، اذ ان (k) تمثل عدد المتغيرات التوضيحية في النظام.

4- صناعة القرار: اذا كانت $nR^2 > \chi_c^2$ عند ذلك يتم رفض فرضية العدم اي توجد مشكلة عدم تجانس التباين، اما فرضيات هذا الاختبار هي:

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_5 = 0$
VS.
H_1: At least two differ

4-17-2 : إختبار هارسون مكابي [21]: Harrison-McCabe test

ان الاثار المترتبة من مشكلة عدم تجانس التباين في تقدير المربعات الصغرى الاعتيادية لنماذج الانحدار الخطي تكون معروفة ، ومن المهم ان تكون هنالك مقدرة على كشف مشكلة عدم تجانس التباين على نطاق واسع ، وخلال السنوات الاخيرة تم اجراء عدة اختبارات متقدمة في هذا المجال ،ان معظم الدراسات التجريبية تستند الى الاوقات الزمنية، وقد اشار كل من هاريسون ومكابي عام (1975) وكذلك ايبس (1977) ان صفة عدم التجانس قد تكون في كثير من الاحيان سمة في السلاسل الزمنية وكذلك في نماذج المقاطع العرضية، اذ تستعمل لهذا الاختبار احصاءة معينة والتي هي عبارة عن نسبة معينة لمجاميع بواقي الانحدار، وهذه النسبة هي (مجموع المربعات من مجموعة فرعية اقل الى اجمالي المجموع المتبقي من المربعات وصيغة هذه النسبة:

$$b = \frac{e' A e}{e' e} \qquad (2-45)$$

اذ تقارن قيمة (Sig.)_لاحصاءة اختبار هارسون مع مستوى المعنوية (0.05) فأذا كانت قيمة مستوى المعنوية (Sig.) اقل فهذا دليل على وجود مشكلة عدم تجانس التباين ، اما اذا كانت هذه القيمة اكبر من (0.05) فهذا يدل على عدم وجود مشكلة عدم تجانس التباين .

2-17-5: اختبار (NC V) [31] : NCV Test

يعد اختبار (NCV) من الاختبارات المهمة والمستعملة في الكشف عن مشكلة عدم تجانس التباين اذ تم اقتراح هذا الاختبار مع بعض التعديلات عام 1983 من قبل (Cook & Weisberg)، على اختبار (Gold field-Quant) اذ تتم المقارنة بالاعتماد على قيمة (.Sig) فاذا كانت اقل من (0.05) فهذا يدعونا الى رفض فرضية العدم والتي تفترض ان التباينات ثابتة اي وجود مشكلة عدم تجانس التباين.

2-18 : معالجة المشكلة : The Problem Remedy

ان خرق الفرضية القائلة بثبات التباين لحدود الخطأ العشوائي يؤدي الى ايجاد قيم تكون مختلفة وليست ثابتة لتباينات حدود الخطا العشوائية، وبذلك فأن القطر الرئيسي لمصفوفة التباين-التباين المشترك والخاصة بحدود الخطأ يصبح محتويا على قيم مختلفة وغير ثابتة ، علما ان مشكلة عدم التجانس لاتقوم بالغاء خاصية الاتساق وعدم التحيز للمعلمات المقدرة لكنها (المعلمات المقدرة) تصبح غير كفوءة وكذلك التباين للمعلمات المقدرة يصبح متحيزا وهذا بدوره يكون مؤثرا في قيم (t) لهذه المعلمات التي تم تقديرها وبذلك تكون محلا للشك ، لذا لابد من ايجاد طريقة للمعالجة ، اذ توجد عدة طرائق لعلاج مشكلة عدم تجانس التباين وكما يأتي [2] [3] [5] :

1- عندما يكون سبب المشكلة هو حذف متغيرات مهمة من الانموذج ففي هذه الحالة يتم اضافة متغيرات جديدة للانموذج.

2- اذا كان سبب المشكلة هو إستعمال الصياغة الخطية فالعلاج هو استعمال الصيغة اللوغارتميه المزدوجة بعد ذلك يتم اختبار مدى خلو البواقي الناتجة من عدم التجانس. او تتم المعالجة بتحويل بعض المتغيرات في الانموذج الى صيغ اخرى ، علما ان التحويل اللوغارتيمي من غير الممكن ان يطبق في حال كون احدى قيم المتغيرين (X,Y) سالبة او مساوية للصفر.

3- تتم المعالجة بطريقة المربعات الصغرى المعممة (GLS).

فعندما يتم اختبار مشكلة عدم تجانس التباين وذلك باستعمال واحد من الاختبارات الخاصة بهذه المشكلة فاحد الحلول المناسبة يتمثل بتحويل الانموذج الاصلي بأي من الطرائق حتى يتم الحصول على شكل لانموذج يكون فيه حد التباين ثابتا ليتم بعد ذلك تطبيق طريقة المربعات الصغرى الأعتيادية (OLS) على الانموذج المحول ، هدف اخر من التحويل يتمثل بتعديل البيانات الاصلية للتخلص من المشكلة ، إن التعديل للانموذج يعتمد على شكل الثبات الخاص لتجانس التباين والذي هو الشكل للعلاقة بين تباين الخطأ (σ_u^2) وبين القيم للمتغير او المتغيرات التوضيحية واتي يتم التعبير عنها على شكل دالة اي :$\sigma_u^2 = f(X_i)$ وبشكل عام ان عملية التحويل للانموذج الاصلي تكون عن طريق قسمة العلاقة الاصلية على الجذر التربيعي للمتغير الذي يعتقد انه الذي سبب المشكلة ولتوضيح طرائق المعالجة نورد الحالات الاتية :

الحالة الاولى : بأفتراض ان الانموذج الاصلي هو :

$$Y=b_0 + b_1 x_i + u_i \qquad (2-46)$$

اذ ان (u_i) يواجه مشكلة عدم تجانس التباين بالرغم من انه يكون مستوفيا للفروض التصادفية الاخرى. وان مشكلة عدم التجانس للتباين تكون كالاتي :

$$E(u_i)^2 = \sigma_u^2 = k^2 X^2 \qquad (2-47)$$

اذ ان K هو ثابت ومحدد ويتم تقديره من الانموذج.

تشير العلاقة (2-47) بان حد الخطا (u_i) يزداد بشكل طردي مع X^2 اي :

$$X^2 = \frac{\sigma_u^2}{k^2} \qquad (2-48)$$

وهذا المعنى يوضح ان التحويل المناسب للانموذج الاصلي يكون بقسمة العلاقة الاصلية على $(\sqrt{X^2} = X)$ وبالاتي يكون شكل الانموذج المحول كالاتي :

$$\frac{Y_i}{x_i} = \frac{b_0}{x_i} + b_1 + \frac{u_i}{x_i} \qquad (2-49)$$

والحد الجديد ($\frac{u_i}{x_i}$) يكون ثابت التجانس (Homoscedastic) وهذا يعني ان التباين يكون ثابت .

$$E(\frac{U_i}{X_i})^2 = \frac{1}{X_i^2} E(u_i)^2 = \frac{1}{X^2}\sigma_u^2 \qquad (2-50)$$

ومن الفروض لطريقة (OLS) ان (X'S) يمثل مجموعة من القيم الثابتة في كل العينات ولكن في المعادلة (2-47) تم افتراض ان $\sigma_u^2 = k^2X^2$ ولذا يكون :

$$E(\frac{u_i}{X_i})^2 = \frac{1}{X_i^2} k^2 X_i^2 = K^2 \qquad (2-51)$$

وعلى هذا الاساس فأن حد الخطأ الجديد له تباين ثابت محدود (وهوk^2) والان من الممكن استعمال طريقة (OLS) على الصيغة المحولة .

الحالة الثانية : بأفتراض إن شكل عدم تجانس التباين ياخذ الشكل الاتي :

$$E(u_i)^2 = \sigma_u^2 = k^2 X \qquad (2-52)$$

اذ ان التحويل المناسب للانموذج الاصلي يكون بقسمة العلاقة على $\sqrt{X}$ وكالاتي :

$$\frac{Y_i}{\sqrt{X_i}} = \frac{b_0}{\sqrt{X_i}} + \frac{b_1}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}} \qquad (2-53)$$

ومن ثم يكون لحد الخطأ الجديد تباين ثابت وهو k^2 اي :

$$E\left(\frac{u}{\sqrt{X}}\right)^2 = \frac{1}{X_i} E(u_i)^2 = \frac{1}{X_i} K^2 X_i = K^2 \qquad (2-54)$$

ومن ثم حصلنا على تباين ثابت وتم التخلص من مشكلة عدم التجانس للتباين .

الفصل الثالث

الجانب التجريبي

الفـــصل الثالث
الجـــــانب التجريبي
Experimental side

لاجل تطبيق ماجاء في الجانب النظري تم استعمال اسلوب المحاكاة(Simulation) والذي يمثل تقليد او تمثيل لعمل نظام حقيقي خلال مدة زمنية معينة، او يمكن وصف عملية المحاكاة على إنها عملية إنشاء وتجريب مضافا لها رياضيات محوسبة، ومن مبررات العمل بهذا النظام هو صعوبة الحصول على بيانات توفر معلومات دقيقة عن ظاهرة معينة أو صعوبة القيام بأثبات البرهان الرياضي بشكله النظري لغرض اختيار افضل طريقة تقدير من بين عدد من الطرق المختلفة ، او حتى لغرض التأكد من جانب تطبيقي موجود اصلاً، فضلا عن ذلك فأن المحاكاة تبدي مرونة وحرية عاليتين في اختيار الحجوم للعينات العشوائية المفترض بها تمثيل مجتمع الدراسة افضل تمثيل، وكذلك لها القدرة العالية على التنوع مع تنوع الأخطاء العشوائية، وكذلك من الممكن إظهار حالة التلوث بالبيانات ولها ايضا القدرة على اعادة إجراء التجربة بشكل كامل عدداً كبيراً من المرات، ولكل النماذج المدروسة فضلا عن الفوائد الأخرى المتمثلة باختزال الجهد والوقت والكلفة.

3-1: مفهوم المحاكاة [6] : Simulation Concept

من الممكن اعطاء تعريف للمحاكاة بأنها: تعبير عن الواقع عبر تكوين او بناء الأنموذج الرياضي أو الاحصائي الذي يعبر عن المشكلة لظروف مشابهة للحقيقة لاجل اعطاء صورة واضحة للمشكلة قيد الدراسة وتعد من الوسائل المهمة لحل المشاكل . ومن الممكن تلخيص عدة فوائد لاسلوب المحاكاة نذكر منها مايأتي :

3-1-1: تطبيق اسلوب المحاكاة: Application simulation style

لاجل تطبيق اسلوب المحاكاة يتم اتباع الخطوات الاتية :

1- تم توليد اربعة متغيرات توضيحية هي $(X_1.X_2.X_3.X_4)$، وهي نفس عدد المتغيرات التوضيحية الحقيقية.

2- تم اخذ عينات (n) و بأحجام مختلفة وهي كالاتي (20, 30, 50,100,200).

3- تم تكرار التجربة (1000) مرة.

4- المتوسطات الحسابية تم استخراجها بصور مختلفة.

5- التباينات تم استخراجها ايضا بصور مختلفة.

6- تم وضع المعلمات (b_1) استخرجت بأشكال مختلفة.

علما ان هذه الاشكال المختلفة للنقاط (4,5,6) ستبين حسب كل حالة وبجدول خلال الخطوات القادمة للبحث . علما ان المعيار اللذي سيتم المقارنة (المفاضلة) على اساسه من بين الاختبارات الخمسة هو [نسبة الكشف (Detection ratio) وهي عبارة عن قسمة عدد مرات الكشف الصحيح على عدد المحاولات مضروب * 100].

3-1-2: الاختبارات المستعملة في المحاكاة: Tests used in simulation

تم اختيار خمس اختبارات من بين الاختبارات التي تم ذكرها في الفصل الثاني للكشف عن وجود مشكلة عدم تجانس التباين علما انه سيتم العمل على أنموذجين الاول يحتوي على مشكلة عدم تجانس التباين والانموذج الاخر لايحتوي على مشكلة عدم تجانس التباين اما الاختبارات المستعملة فهي كالاتي :

1- اختبار بروش- بيجين – جود فري. Breusch-Pagan test

2- اختبار NCV. NCV. Test

3- اختبار وايت. White test

4- اختبار جولد فيلد- كواندت. GQ. Test

5- اختبار Harrison-McCabe. Harrison-McCabe test

3-1-3 : تجارب المحاكاة: Simulation experiments

تم استعمال برنامج في تنفيذ تجارب المحاكاة وهذا البرنامج يستند الى لغة(R) وهي لغة متطورة عن لغة (S-Pluss) ومن مميزات هذا البرنامج هو الامكانية العالية في العلوم الإحصائية والهندسية والرياضية اذ يعمل على توظيف الايعازات والادوات بصيغة مباشرة لاجل الحصول على برمجة متقدمة . تعد لغة R من بين اللغات التي تألق نجمها حديثا وبشكل سريع في مجال البرمجيات العلمية في مجالي الإحصاء والمعلوماتية اذ اصبحت لغة معتمدة على نطاق كبير في كثير من مراكز البحث العلمية والجامعات، وصار استعمال هذه اللغة والاشارة إليها في المجلات العلمية المحكّمة والبحوث المنشورة يزداد بشكل طردي بوتيرة متسارعة وبشكل طردي ، مضافا الى هذا تعد لغة البرمجة(R) لغة حرة ذات مصدر مفتوح و توزيعها يخضع لترخيص

GPL االمشهور. كل هذا ساعد في تزايد ما هو متوفر وموجود على شبكة (الإنترنت) من مراجع لها ، فهناك ماهو موجود على شكل الكتب الإلكترونية والدروس التعليمية بل حتى المناهج الأكاديمية وكذلك الدورات التدريبية مضافا لذلك البرمجيات الجاهزة ,والتي تكون مكتوبة بلغة (R)لاجل تنفيذ المهمة هذه أو تلك، حتى اصبحت هذه اللغة تحظى ببعض الامتياز عند مقارنتها مع العديد من العمالقة في قطاع البرامج الإحصائية والرياضية العلمية مثل (SASو SPSS) ولاسيما في مجال توفر الطرق الجديدة والخوارزميات الحديثة ، اذ نلحظ هذا التوجه في معظم الجامعات التي تتمثل بطلاب الدراسات العليا يشجعهم على ذلك سهولة البناء لاضافة هذه اللغة، ويعد استعمال هذا الأسلوب رغم ما قد يتخلله من نقاط ضعف فيما يخص موثوقية وغزارة وجودة هذه الإضافات الجديدة، والتي تتعلق بمهارة وخبرة مطوريها وناشريها، وعلى الرغم من ذلك تبقى في القطاع الأكاديمي و العلمي الافضل من البدائل التجارية التي يعاب عليها ارتفاع ثمنها من جهة، وبطئ إضافة التحديثات التي تعكس تطور القطاعات العلمية المختلفة من جهة اخرى، اذ أنها في العادة تتبع دورة تجارية يتم التحكم بها من لدن الشركات المنتجة(SAS و SPSS) ولاسيما في مجال توفر الطرائق الجديدة والخوارزميات الحديثة .

2-3 : الحصول على المعلمات: Get parameters

اولا : المعلمات(b_i) :

تم تقدير المعلمات $(b_1; b_2; b_3; b_4)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للمعلمات الحقيقية .

ثانيا: المتوسطات(u_i) :

تم تقدير المتوسطات $(u_1; u_2; u_3; u_4)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للمتوسطات الحقيقية .

ثالثا: التباينات (σ_i^2) :

تم تقدير التباينات $(\sigma_1^2; \sigma_2^2; \sigma_3^2; \sigma_4^2)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للتباينات الحقيقية . وفيما يلي تبيان للمعالم الأتية :

جدول (1) يبين المعلمات والمتوسطات والتباينات للبيانات الحقيقية

ت	المعلمة	المتغيرات التوضيحية			
1	المعلمات (b_i)	X_4	X_3	X_2	X_1
		-0.268	0.023	-0.144	0.641
2	المتوسطات $\bar{X}_i$	2.18	26.63	6.14	2.745
3	التباينات σ_i^2	1.462	19.104	13.758	0.195

المصدر : الجدول من عمل الباحث واعتمادا على البرنامج (SPSS)

جدول (2) يبين المعلمات والمتوسطات والتباينات للبيانات الافتراضية العليا

ت	المعلمة	المتغيرات التوضيحية			
1	المعلمات (b_i)	X_4	X_3	X_2	X_1
		-0.123	0.034	-0.112	0.732
2	المتوسطات$(\bar{X}_i)$	3.46	27.32	7.29	3.24
3	التباينات $(\sigma^2{}_i)$	2.35	20.621	14.561	1

المصدر : الجدول من عمل الباحث

جدول (3) يبين المعلمات والمتوسطات والتباينات للبيانات الافتراضية الدنيا

ت	المعلمة	المتغيرات التوضيحية			
1	المعلمات	X_4	X_3	X_2	X_1
		-0.381	0.011	-0.221	0.512
2	المتوسطات $(\overline{X}_i)$	1.94	25.41	5.43	1.92
3	التباينات σ_i^2	0.94	18.123	11.234	0.132

المصدر : الجدول من عمل الباحث

3-3 : تطبيق البيانات الافتراضية:

تجربة رقم (4): تم محاكاة أنموذج انحدار بالمعلمات الافتراضية العليا
$B_i = 0.732, -0.112, 0.034, -0.123$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان $u_1 = u_2 = u_3 = u_4 = 0$ ، $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

الجدول (4)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج لا يحتوي على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	2	3	13	3	4	12	3	3	4	6	6	5
R.NCV	20	21	34	25	28	39	2	3	5	4	4	6
R.white	38	44	36	27	41	51	43	38	39	41	39	41
R.GQ	87	99	100	100	100	100	4	6	4	4	6	6
R.HM	82	98	100	100	100	100	5	4	4	4	6	5

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

4-3: تحليل نتائج محاكاة أنموذج الانحدار:

بالاعتماد على نتائج (جدول 4) يتبين ان نسبة الكشف لاختبار (كولد فيل-كوانت) بلغت 87% عند حجم العينة 20 و99% عند حجم العينة اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% عند احجام العينات الاخرى ، بينما كانت نسب كشف اختبار (بروش- بيجين – جود فري) 2% عند حجم العينة 20 و3% عند حجم العينة 30 و13% عند حجم العينة 50 و 3% عند حجم العينة 100 و4% عند حجم العينة 150 و12% عند حجم العينة 200.

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين كانت النتائج مايأتي:

بلغت نسبة الكشف لاختبار (NCV) 2% عند حجم العينة 20 و3% عند حجم العينة 30 و 5% عند حجم العينة 50 و4% عند حجم العينة 100 و4% عند حجم العينة 150 و6% عند حجم العينة 200 اما اختبار(white) فسجل 43% عند الحجم 20 و38% عند الحجم 30 و39% عند حجم العينة 50 و41% عند حجم العينة 100 و39% عند حجم العينة 150 و 41% عند حجم العينة 200.

تجربة (5): تم محاكاة أنموذج انحدار بالمعلمات الافتراضية الدنيا
$B_i = 0.512, -0.221, 0.011 - 0.381$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان $u_1 = u_2 = u_3 = u_4 = 0$ ، $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

الجدول (5)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج لا يحتوي على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	10	3	5	2	10	1	2	4	5	4	4	5
R.NCV	23	23	24	23	39	26	3	3	3	4	4	5
R.white	62	40	33	41	40	13	43	43	40	38	42	40
R.GQ	82	99	100	100	100	100	4	5	5	4	6	6
R.HM	68	97	100	100	100	100	4	4	5	4	6	6

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

بالاعتماد على نتائج (الجدول 5) يتبين أن نسبة الكشف لاختبار (كولد فيلد كوانت) بلغت 82% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 10% عند حجم العينة 20 و3% عند حجم العينة 30 و5% عند حجم العينة 50 و2% عند حجم العين 100 و 10% عند حجم العينة حجم العينة 150 و1% عند حجم العينة 200.

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين كانت النتائج كما يأتي:

اذ بلغت نسبة الكشف لاختبار (NCV) 3% عند حجم العينة 20 و3% عند حجم العينة 30 و 3% عند حجم العينة 50 و4% عند حجم العينة 100 و4% عند حجم العينة 150 و5% عند حجم العينة 200 اما اختبار(white) فسجل 43% عند الحجم 20 و43% عند الحجم 30 و40% عند حجم العينة 50 و38% عند حجم العينة 100 و42% عند حجم العينة 150 و 40% عند حجم العينة 200.

تجربة رقم (6): تم محاكاة أنموذج انحدار بالمعلمات الحقيقية الاتية
$B_i = 0.641, -0.144, 0.023. -0.268$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان $u_1 = u_2 = u_3 = u_4 = 0$ ،
$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

الجدول (6)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في حالة احتواء الانموذج على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في حالة عدم احتواء الانموذج على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	2	3	10	8	7	4	3	3	4	5	4	6
R.NCV	21	20	32	32	34	29	4	4	4	4	4	5
R.white	70	32	54	11	45	39	56	54	61	62	60	62
R.GQ	82	99	100	100	100	100	5	5	4	5	6	5
R.HM	72	98	100	100	100	100	7	5	4	5	5	5

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

يتبين عن طريق نتائج (الجدول 6) ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 82% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 2% عند حجم العينة 20 و3% عند حجم العينة 30 و10% عند حجم العينة 50 و8% عند حجم العينة 100 و 7% عند حجم العينة حجم العينة 150 و4% عند حجم العينة 200.

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين جاءت النتائج كما يأتي :

كانت نسبة الكشف متكافئة لاختباري (NCV وBP) بشكل تقريبي اما اختبار (white) فسجل 56% عند حجم العينة 20 و54% عند الحجم 30 و61% عند الحجم 50 و 62% عند حجم العينة 100 و60% عند الحجم 150 و 62 % عند الحجم 200.

تجربة (7): تم محاكاة أنموذج انحدار بالمعلمات الحقيقية مع القيم الدنيا للمتوسط والتباين $B_i = 0.641, -0.144, 0.023. -0.268$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان:

$$,u_1. u_2. u_3 . u_4 = 1.92 . 5.43. 25.41 . 1.94$$
$$\sigma_1^2. \sigma_2^2. \sigma_3^2. \sigma_4^2 = 0.132 . 11.234 . 18.123 . 0.9$$

الجدول (7)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج لا يحتوي على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	3	3	4	10	14	2	4	3	4	3	6	4
R.NCV	16	23	28	34	42	34	3	3	4	4	4	4
R.white	29	46	32	46	46	46	40	37	36	43	38	40
R.GQ	86	99	100	100	100	100	5	6	7	6	3	5
R.HM	75	97	100	100	100	100	4	4	6	6	3	5

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

بالاعتماد على نتائج (الجدول رقم 7) يتبين ان نسبة الكشف لاختبار (كولد فيلد كوانت) بلغت 86% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 3% عند حجم العينة 20 و3% عند حجم العينة 30 و4% عند حجم العينة 50 و10% عند حجم العين 100 و 14% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200.

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين كانت النتائج مايأتي :

اذ بلغت نسبة الكشف لاختبار(NCV) 3% عند حجم العينة 20 و3% عن الحجم 30 و 4% عند الحجم 50 و4% عند عند الحجم 100 و4% عند الحجم 150 و4% عند الحجم 200 على الاختبارات الاخرى اما اختبار (white) فسجل 40% عند حجم العينة 20 و37% عند الحجم 30 و36% عند الحجم50 و43% عند الحجم 100 و38% عند الحجم 150 و40% عند الحجم 200.

تجربة رقم (8): تم محاكاة أنموذج انحدار بالمعلمات الحقيقية والقيم الافتراضية العليا للمتوسط والتباين

$B_i = 0.641, \ -0.144, \ 0.023. -0.268$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان:

$$u_1 . u_2 . u_3 . u_4 = 3.24 . 7.29 . 27.32 \ . \ 3.46$$

$$\sigma_1^2 . \sigma_2^2 . \sigma_3^2 . \sigma_4^2 = 1 \ . \ 14.561 \ . \ 20.621 \ . \ 2.35$$

الجدول (8)

نسبة الكشف عن عدم وجود مشكلة عدم تجانس التباين في نموذج لا يحتوي على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	5	2	10	5	5	2	3	4	5	4	5	6
R.NCV	18	28	33	30	31	25	2	5	5	4	4	4
R.white	49	26	50	33	46	31	42	40	38	41	39	41
R.GQ	78	100	100	100	100	100	5	5	4	7	5	6
R.HM	60	99	100	100	100	100	5	5	4	6	5	6

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

بالاعتماد على نتائج (الجدول 8) يتضح ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 78% عند حجم العينة 20 و100% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 5% عند حجم العينة 20 و2% عند حجم العينة 30 و10% عند حجم العينة 50 و5% عند حجم العين 100 و 5% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200 .

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين كانت النتائج مايأتي :

اذ بلغت نسبة الكشف لاختبار (NCV) 2% عند حجم العينة 20 و5% عند الحجم 30 و5% عند الحجم 50 و4% عندالاحجام 100و150و200 اما اختبار (white) فسجل 42% عند حجم العينة 20 و40% عند الحجم 30 و38% عند الحجم 50 و41% عند الحجم 100 و 39% عند الحجم 150 و41% عند الحجم 200 .

تجربة رقم (9): تم محاكاة أنموذج انحدار بالمعلمات الحقيقية والقيم الحقيقية للمتوسط والتباين $B_i = 0.641, \ -0.144, \ 0.023. -0.268$ والجدول الاتي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين اذ ان:

$$u_1 . u_2 . u_3 . u_4 = 2.745 . 6.14 . 26.63 . 2.18$$
$$\sigma_1^2 . \sigma_2^2 . \sigma_3^2 . \sigma_4^2 = 0.195 . 13.758 . 19.104 . 1.462$$

جدول (9)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين							نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج لا يحتوي على عدم تجانس التباين					
Size of sample / Tests	20	30	50	100	150	200	20	30	50	100	150	200
R.BP	2	3	1	2	4	2	2	3	5	5	6	5
R.NCV	19	22	20	21	30	26	3	4	4	4	4	4
R.white	54	47	23	20	60	19	61	43	43	44	39	39
R.GQ	85	99	100	100	100	100	5	4	5	5	5	5
R.HM	70	98	100	100	100	100	6	4	5	6	5	5

المصدر : الجدول من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

بالاعتماد على مخرجات (الجدول 9) تبين ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 85% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون اذ تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 2% عند حجم العينة 20 و3% عند حجم العينة 30 و1% عند حجم العينة 50 و2% عند حجم العينة 100 و 4% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200.

اما في حالة عدم احتواء الانموذج على مشكلة عدم تجانس التباين كانت النتائج مايأتي :

اذ بلغت نسبة الكشف لاختبار (NCV) 3% عند حجم العينة 20 و 4% عند احجام العينات 30 و50 و100 و 150 و 200 اما اختبار(white) فسجل 61% عند الحجم 20 و43% عند الحجم 30 و43% عند الحجم 50 و44% عند الحجم 100 و39% عند الحجم 150 و39% عن الحجم 200.

5-3: تطبيق لبيانات واقعية حول مرض اليرقان:

يعرف المرض اليرقان محليا (ابو صفار) و عالميا (Jaundice) وهو عبارة عن تلوّن في الجلد والعيون بصبغة صفراء نتيجةَ لتراكم مادة البيليروبين في الجسم وبسبب عدم قدرة الكبد على التخلّص من هذه المادة إلى خارج الجِسم وتكثر حالات الاصابة بهذا المرض عند الاطفال حديثي الولادة ؛ ففي الحالة الطبيعية مادّة البيليروبين تعد مادة ترافق كريات الدم الحمراء، وفي حال هرم كرية الدم الحمراء فأن الكبد يتخلص منها ومن مادة البيليروبين الموجودة فيها، ليتمّ إنتاج كريات دم حمراء فتية وجديدة، ولكن بسبب مشكلة معينة في الكبد فإنه لا يستطيع أن يتخلّص من هذه الكريات الهرمة، فتتراكم مادة البيليروبين في الجسم. هناك نوع من اليرقان يعرف بالكاذب، يكون سببه الإكثار من مادة البيتاكاروتين؛ اذ يتلون الجلد باللون الأصفر البرتقالي ولا يتغير بياض لون العين .

تم سحب عينة عشوائية بسيطة من المرضى المصابين بمرض اليرقان حجمها (100) يتضمن هذا المبحث توصيف للبيانات المتمثلة بنسبة الاصابة بمرض اليرقان والعوامل التي تؤثر في هذا المرض سواء ازدياد المرض او نقصانه، ومن ثم اختبار البيانات عن طريق معرفة التوزيع لها وكذلك تشخيص مشكلة عدم تجانس التباين من خلال تطبيق البيانات على الاختبار المعلمي الافضل الخاص بالكشف عن مشكلة عدم التجانس للتباينات .

6-3: تهيئة البيانات: Data Preparing

تم الحصول على البيانات الحقيقية من مستشفى الرميثة العام وهي بيانات طبية اخذت كعينة من شعبة الاحصاء ومختبر التحليلات المرضية للمستشفى عن طريق عينة عشوائية تشمل الاشهر(كانون الثاني– شباط- اذار من عام 2018) لنسبة اليرقان في الدم والعوامل المؤثرة في زيادة او نقصان هذه النسبة، وتم فرز البيانات الاصلية عن طريق استعمال برنامج (Excel) كل قسم على حده

اذ تم اخذ عينة مكونة من (100) مشاهدة ، وتم افتراض المتغيرات وهي :

1-المتغير المعتمد(Y) ويمثل نسبة الاصابة بمرض.
2- المتغير التوضيحي (X_1) ويمثل وزن الطفل (الخديج) بالكغم.
3- المتغير التوضيحي (X_2) ويمثل عمر الطفل (الخديج) بالايام.
4- المتغير التوضيحي (X_3) ويمثل عمر الام (بالسنين).
5- المتغير التوضيحي (X_4) ويمثل مدة رقود الطفل (الخديج) في المستشفى بالايام.

7-3: اختبار طبيعية توزيع البيانات [12] :

حسب احد شروط الانحدار الخطي المتعدد هو ان مشاهدات المتغير يجب ان تتوزع طبيعيا، لذا كان من الضروري معرفة مدى كون البيانات تتوزع طبيعيا ام لا ، ولان المشاهدات المستعملة هي اكثر من (50) مشاهده لذلك من الممكن استعمال اختبار مربع كاي لحسن المطابقة (χ^2goodnees of fit) او اختبار كولموكروف ـ سيمرنوف (Smirnov test – The Kolmogorov)، وهنا سيتم استعمال اختبار كولموكروف ـ سيمرنوف. وبافتراض امتلاكنا عينة عشوائية بحجم (100) مشاهدة ، ستكون الفرضية :

ولاجل اختبار طبيعة التوزيع بالنسبة للمتغير المعتمد (Y) عند حجم العينة (n=١٠٠) تم استعمال البرنامج الاحصائي الجاهز (SPSS) وهذا البرنامج منفذ على الحاسب الالكتروني ، عن طريق مشاهدة الجدول (10) يتضح لنا بان البيانات تتوزع توزيعا طبيعيا ، وتتم معرفة ذلك عن طريق قيمة مستوى المعنوية (.Sig) فاذا كانت هذه القيمة اكبر من (0.05) فان هذا يدل ان البيانات تتوزع توزيعاً طبيعياً ، لكن في حالة كون قيمة (.Sig) اقل من (0.05) فهذا يدل ان البيانات لا تتبع التوزيع الطبيعي.

جدول (10) يوضح اختبار التوزيع الطبيعي للبيانات الحقيقية للمتغير المعتمد (Y)

	Kolmogorov-Smirnov test		
	Statistic	Df	Sig.
y	0.064	100	0.200

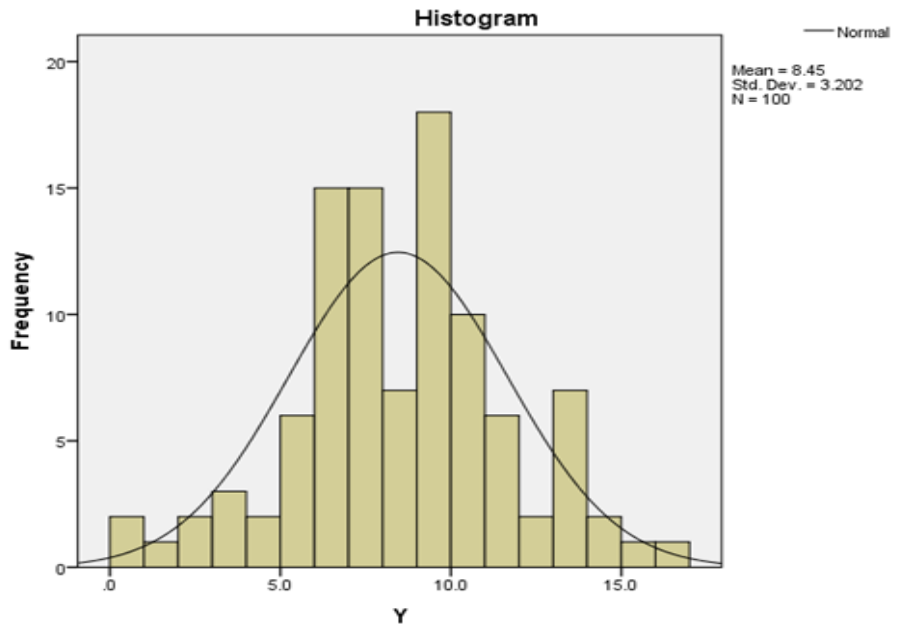
المصدر : الجدول من عمل الباحث واعتمادا على نتائج برنامج (Spas)

فرضية العدم (H_0) : العينة مسحوبة من مجتمع بياناته تتوزع حسب التوزيع الطبيعي.

الفرضية البديلة (H_1) : العينة مسحوبة من مجتمع بياناته لاتتبع التوزيع الطبيعي .

بالاعتماد على الجدول (10) يتبين عن طريق قيمة (Sig.) والبالغة (0.2) ان بيانات المتغير المعتمد تتوزع حسب التوزيع الطبيعي

الشكل (2) يوضح توزيع بيانات المتغير المعتمد (Y)



الشكل من عمل الباحث بالاعتماد على نتائج البرنامج الإحصائي SPSS

كذلك لتوضيح اكبر لطبيعية توزيع البيانات عن طريق الرسم ، من الممكن ملاحظة الشكل رقم (3) اذ يوضح المدرج التكراري التوزيع الطبيعي للبيانات ولا سيما عن طريق منحنى التوزيع الطبيعي.

8-3: تطبيق (الاختبار الافضل): Apply the best test

سيتم تطبيق الطرائق المعلمية اي الاختبار الافضل الذي اختير بالاعتماد على اسلوب المحاكاة في حالة وجود مشكلة عدم تجانس التباين (علما انه تم اعتماد الأنموذج الذي يحتوي مشكلة عدم تجانس التباين) على البيانات لاختبار وجود او عدم وجود مشكلة عدم تجانس التباين بواسطة الاختبار الافضل وهو اختبار (كولد فيلد-كوانت) و أختبار(NCV) في حالة حجم عينة (n=100) ، ليتم بعد ذلك تقدير المعلمات بطريقة المربعات الصغرى المعممة (GLS) في حال وجود مشكلة عدم تجانس التباين اما في حالة عدم وجود المشكلة فيتم اللجوء الى طريقة المربعات الصغرى الاعتيادية (OLS). اذ تم الاعتماد على أسلوب (لغة البرمجة R)

9-3 : الكشف عن وجود او عدم وجود مشكلة عدم تجانس التباين:

تم حساب إحصاءه الاختبار لاختباري (كولدفيلد-كوانت و NCV) اذ تم تطبيقهما على البيانات الحقيقية بعد ان كانا هما الافضل في اسلوب المحاكاة .

1-9-3: اختبار (Gold field-Quant):

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$
$H_1: \text{At least two differ}$

جدول (11) يوضح إحصاءة اختبار(Gold field-Quant) على البيانات الحقيقية، مستوى المعنوية (0.05)

Test	Statistic	D f	D f	Sig.
GQ	0.59971	45	45	0.9551

المصدر : الجدول من عمل الباحث واعتمادا على نتائج برنامج (لغة R)

التعليق :
عن طريق النظر في قيم (Sig.) والخاصة لاختبار كولد فيلد- كوانت في الجدول (11) نجد ان قيمة (Sig.) والتي تساوي (0.9551) هي اكبر من مستوى المعنوية (0.05) .

2-9-3: اختبار (NCV): Test the NCV

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$
$H_1:$ At least two differ

جدول رقم (12) يوضح إحصاءة اختبار(NCV)
على البيانات الحقيقية، مستوى المعنوية (0.05)

Test	Statistic	Df	Sig.
NCV	0.8036538	1	0.3700032

المصدر : الجدول من عمل الباحث واعتمادا على نتائج برنامج (لغة R)

التعليق :
عن طريق النظر في قيم (Sig.) والخاصة لاختبار(NCV) في الجدول (12) نجد ان قيمة (Sig.) والتي تساوي (0.37) هي اكبر من مستوى المعنوية (0.05).

من خلال النظر في قيمة مستوى المعنوية (Sig) لاختبار كولد فيلد-كوانت والبالغة (**0.9551**) وكذلك مستوى المعنوية (Sig) والبالغة (0.37) نجد إنها أكبر من (0.05) وهذا يعني ان البيانات لا تعاني من مشكلة عدم تجانس التباين. وبالتالي فأن الطريقة المناسبة لتقدير معالم الأنموذج هي طريقة المربعات الصغرى الاعتيادية (OLS) وكما يلي :

10-3: تقديرات المربعات الصغرى الاعتيادية (OLS):

جدول (13) يبين تقديرات (OLS) للنموذج المقدر مع الانحرافات المعيارية وكذلك قيم (t) للمعالم المقدرة والقيمة المعنوية لاحصاءة الاختبار.

Variables	Estimate	Std. Error	t value	Sig.
constant	7.58409	2.88726	2.627	* 0.0101
X_1	0.64148	0.71677	0.895	0.3731
X_2	-0.14370	0.09041	-1.589	0.1153
X_3	0.02268	0.07236	0.313	0.7546
X_4	-0.26774	0.27402	-0.977	0.3310

المصدر : الجدول من عمل الباحث بالاعتماد على نتائج برنامج (لغة R)

11-3 تحليل النتائج:

عن طريق النتائج في الجدول (13) يتبين مايأتي :

اولا: بلغت المعلمة للمتغير (X_1) والذي يمثل وزن الطفل (0.64148) بينما بلغ الانحراف المعياري (0.71677) اما قيمة اختبار (t) للمعلمة (b_1) فكان (0.895) بينما بلغ مستوى المعنوية (Sig.) (0.3731) وهو اكبر من (0.05) وهذا يعني ان المتغير التوضيحي (X_1) الذي يمثل وزن الطفل ليس له اثر في المتغير المعتمد (Y) والذي يمثل نسبة الاصابة بمرض اليرقان.

ثانيا : بلغت المعلمة للمتغير (X_2) والذي يمثل عمر الطفل (-0.14370) بينما بلغ الانحراف المعياري (0.09041) اما قيمة اختبار (t) للمعلمة (b_1) فكان (-1.589) بينما بلغ مستوى المعنوية (Sig.) (0.1153) وهذا يعني ان المتغير التوضيحي (X_2) والذي يمثل عمر الطفل ليس له أثر على المتغير المعتمد (Y) والذي يمثل نسبة الاصابة بمرض اليرقان.

ثالثا : بلغت المعلمة للمتغير (X_3) والذي يمثل عمر الأم (0.02268) بينما بلغ الانحراف المعياري (0.07236) اما قيمة اختبار (t) للمعلمة (b_1) فكان (0.313) بينما بلغت قيمة

Sig. (0.7546)وهذا يعني ان المتغير التوضيحي (X_3) ليس له تأثير على المتغير المعتمد (Y) والذي يمثل نسبة الاصابة بمرض اليرقان.

رابعا : بلغت المعلمة للمتغير (X_4) والذي يمثل مدة رقود الطفل (0.26774-) بينما بلغ الانحراف المعياري (0.27402) اما قيمة اختبار (t) للمعلمة (b_1) فكان (0.977-) بينما بلغت قيمة Sig. (0.3310) وهذا يعني ان المتغير التوضيحي (X_4) ليس له تأثيرعلى المتغير المعتمد (Y).

الفصل الرابع

الاستنتاجات

و

التوصيات

الفصل الرابع

الاستنتاجات والتوصيات

4-1: الاستنتاجات: Conclusions

1- عن طريق النتائج المعروضة لاختباري كولد فيلد- كوانت و(NCV) نلحظ ان قيمة (.Sig) هي اكبر من مستوى المعنوية ومعنى هذا ان البيانات لاتعاني من مشكلة عدم تجانس التباين اي الاخذ بالفرضية الصفرية للبحث .

2- بالاعتماد على نتائج تقديرات المعلمات نلاحظ عدم وجود أثر للمتغيرات المستقلة ($X_1.X_2.X_3.X_4$) على المتغير المعتمد (Y)، وهذا يعني ان العوامل (وزن الطفل وعمر الطفل وعمر الام وفترة الرقود) ليس لها تأثير معنوي على نسبة الاصابة بمرض اليرقان.

3- تفوق اختبار (كولد فيلد –كوانت) على بقية الاختبارات في الأنموذج الذي يحتوي على مشكلة عدم تجانس التباين وذلك لامتلاكه اكبر نسبة كشف صحيح.

4- تفوق اختبار (NCV) على بقية الاختبارات في الأنموذج الذي لايحتوي على مشكلة عدم تجانس التباين اذ كانت له اعلى نسبة كشف صحيح.

5- تسجيل اقل نسبة كشف لاختبار (بروش- بيجين – جود فري) لمشكلة عدم تجانس التباين في حالة الأنموذج الذي يحتوي على مشكلة عدم تجانس التباين.

6- هنالك تقارب بين اختبار (كولد فيلد- كوانت) واختبار (بروش- بيجين – جود فري) في حالة احتواء الأنموذج على مشكلة عدم تجانس التباين.

7- في حالة الأنموذج الذي يحتوي على مشكلة عدم تجانس التباين نجد ان إختبار (كولد فيلد- كوانت) يتفوق على جميع الاختبارات الا انه يتساوى مع اختبار (Harrison-McCabe) عند احجام العينات (50,100,150,200).

8- في حالة الأنموذج الذي لايحتوي على مشكلة عدم تجانس التباين نجد ان الاختبارات الاربعة عدا اختبار (وايت) كانت متفوقة بشكل متقارب في الكشف عن عدم وجود مشكلة عدم تجانس التباين.

9- في حالة الأنموذج الذي لايحتوي على مشكلة عدم تجانس التباين نجد ان اختبار (وايت) يسجل ادنى نسبة كشف عن عدم وجود مشكلة عدم تجانس التباين وهذه دلالة على ان هذا الاختبار يحمل كفاءة اقل من الاختبارات الاربعة الاخرى.

2-4: التوصيات: Recommendations

اعتمادا على الاستنتاجات التي تمت الاشارة اليها ، هنالك عدد من التوصيات التي نرى انها ضرورة الاخذ بها من قبل الباحثين وكما يأتي :-

1- اعتماد اختبار (NCV) في الكشف الصحيح كونه اعطى افضل نسبة كشف صحيح عن عدم وجود مشكلة عدم تجانس التباين.

2- يمكن اعتماد اختباري (Harrison-McCabe اوGold field-Quant) للعينات ذات الحجم اكبر من (50) في الكشف عن المشكلة كونهما يكونان متكافئين.

3- استعمال اسلوب المقارنة بين اختبارات معلميه أخرى باستعمال أسلوب البرمجة (R).

4- اعتماد اختبار (كولد فيلد-كوانت) في الكشف الصحيح كونه اعطى افضل نسبة كشف صحيح عن وجود مشكلة عدم تجانس التباين.

5- تم في هذه الرسالة استعمال معادلة خطية من الدرجة الثانية في إسلوب للمحاكاة والتي تستعمل للكشف الصحيح عن وجود مشكلة عدم تجانس التباين، لذا نوصي بأجراء دراسة يتم فيها استعمال معادلة لحساب الجذر التربيعي، او معادلات مثلثية.

6- نوصي شعبة الاحصاء الصحي والحياتي في وزارة الصحة العراقية بتصميم استمارة احصائية تظم الاسباب الرئيسة والتي تؤثر بشكل مباشر واساسي في نسبة اليرقان (ابو صفار) وذلك لان العوامل المثبتة حاليا (المتغيرات التوضيحية) التي تمت دراستها اظهرت عدم معنويتها (عدم تثيرها) في نسبة الاصابة بمرض اليرقان (المتغير المعتمد)، كذلك اختيار او تدريب عاملين يدركون اهمية الرقم والتوثيق الاحصائي.

المصادر

اولا: المصادر العربية:

القران الكريم

1- امين، اسامة ربيع .(2007) ," .التحليل الاحصائي باستخدام برنامج SPSS "، الطبعة الثانية، مصر، القاهرة، مكتبة الأنجلو المصرية.

2- بخيت، حسين علي. فتح الله، سحر. (2012)،" الاقتصاد القياسي "، مطبعة دار الكتب ، بغداد.

3- التميمي، زهرة حسن عباس. واخرون.(2014)، "تحليل الانحدار"، الطبعة الاولى، مديرية دار الكتب للطباعة والنشر، جامعة البصرة، العراق، البصرة.

4- شيخي ، محمد ، 2011 ، " طرق الاقتصاد القياسي " ، الطبعة الاولى، دار الحامد ،عمان ، الاردن .

5- عبد، حميد عبيد. 2016، "الاقتصاد القياسي"، الطبعة الاولى، مطبعة دار الكتب، العراق، كربلاء.

6- عبدالرحمان، عدنان ماجد. (2004)،" نظام المحاكاة Arena بلامثلة "، ط١، المملكة العربية السعودية، جامعة الملك سعود.

7- علوان، إيلاف بهاء. (2016) ، " مقارنة بين اختبار Goldfield-Quant الحصين مع اختبارات أخرى للكشف عن مشكلة عدم تجانس التباين بوجود القيم الشاذة " ، بحث منشور ، مجلة الكوت للعلوم الاقتصادية والادارية، كلية الادارة والاقتصاد ، جامعة واسط.

8- عبد رمضان، احمد (2016)" تحليل مشكلة عدم تجانس التباين والمقارنة بين اختبار اوبراين واختبار اوبراين المعدل باستعمال المحاكاة "، بحث، منشور، مجلة الكوت للعلوم الاقتصادية والادارية، الى كلية الادارة والاقتصاد، جامعة واسط.

9- القريشي، احسان كاظم شريف. (2007)، " الطرائق المعلمية والطرائق اللامعلميه في الأختبارات الأحصائية"، الطبعة الأولى ،كلية الأدارة والاقتصاد /الجامعة المستنصرية.

10- الكناني، ايدن حسن. (2006)، "تقدير معاملات الانحداربعد اجراء اختبار عدم تجانس التباين"، بحث منشور، مجلة الادارة والاقصاد، الجامعة المستنصرية.

11- اللآمي، احمد رزاق عبد رمضان. طالب، حيدر رائد.(2016)، "مقارنة اختبار Z المعدل للتباين مع اختبار ليفين المعدل بوجود القيم الشاذة باستعمال المحاكاة"، كلية الادارة والاقتصاد، جامعة واسط.

12- المشهداني ، كمال علوان. البياتي، محمود مهدي .(2014) ،"التصاميم التجريبية وتحليل التجارب بأستعمال برنامج SPSS" ، كلية الأدارة والأقتصاد ، جامعة بغداد. د. ط.

13- النعيمي، محمد عبد العال.طعمة، حسن ياسين. (2015)، " الاحصاء التطبيقي" ، ،ط٢،الاردن، عمان، دار وائل للنشر.

14- الوائلي، علا هادي صادق. (2015)، "الطرائق المعلمية واللامعلمية لاختبارات تجانس التباين مع تطبيق عملي" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

ثانيا : المصادر الاجنبية:

15. Anderson, David R., Sweeney, Dennis J., Williams, Thomas A. (2008),"STATISTICS FORBUSINESS ANDECONOMICS" , Eleventh Edition, USA, South-Western Cengage Learning.

16. Chung, Christopher A. (2004) "Simulation modeling handbook: a practical approach "Printed in the United States of America Library of Congress Card.

17. D. ORME, CHRIS. (1998), "On the insensitivity of the score test for heterogeneity to omitted covariates in multivariate failure time models ", Printed in Great Britain, Biometrika, 85, 2, pp. 457-161.

18. Duchateau, Luc. Janssen, Paul. Lindsey, Patrick. Legrand, Catherine. Nguti, Rosemary. Ylvester, Richard. (2002), "The shared frailty model and the power for heterogeneity tests in multicenter trials", Computational Statistics & Data Analysis pp. 603 – 620.

19. FREY, Jesse. (2010)," Testing for equivalence of variances using Hartley’s ratio", The Canadian Journal of Statistics, Vol. 38, No. 4, pp. 647–664.

20. Gujarati, Damodar N. (1995)," Basic econometrics", fourth edition, McGraw-Hill, Inc.

21. Harrison, M. J., McCabe, B. P. M. (1979)," A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals ", USA, Journal of the American Statistical Association.

22. Leeckley, Robers M. (2010)," Applied Statistics in Business and Economics ", CRC Press Taylor & Francis Group, Boca Raton London New York.

23. Maoa Chang Xuan Lindsayb, Bruce G. (2003)," Tests and diagnostics for heterogeneity. In the species problem ", Computational Statistics & Data Analysis pp. 389 – 398.

24. Rawlings, John O., Pantula, Sastry G., Dickey, David A.(1998)," Applied Regression Analysis " , Second Edition, USA, Springer-Verlag New York, Inc.

25. Rice, William R., Gaines, Steven D. (1994) "Extending non directional heterogeneity tests to evaluate simply ordered alternative hypotheses", Proc. Natl. Acad. Sci. USA, Vol. 91, pp. 225-226, January Statistics.

26. Salvatore, Dominick. And Reagie, Derrick (2002) "Statistics and Econometrics", Second Edition, London.

27. Schaalje, G. Bruce., Rencher, Alvin C. (2008)," Linear models in statistics", Second Edition, Canada, John Wiley & Sons, Inc., Hoboken, New Jersey.

28. Timm, Neil H. (2002)," Applied multivariate analysis ", USA, Springer-Verlag New York, Inc.

29. Weisberg, Sanford. (2005)," Applied Linear Regression" Third Edition, USA, A John Wiley & Sons, Inc., Publication.

30. Wang, lan. Hua Zhou, Xiao. (2014)," A Fully Nonparametric Diagnostic Test for Homogeneity of Variances", Source: The Canadian Journal of Statistics / La Revue Canadienne de Statistique, Vol. 33, No. 4(Dec.), pp. 545-558

31. Weisberg, S. (2014) Applied Linear Regression, Fourth Edition, Wiley, Printed in the United States of America.

32. Yan, Xin. Gang Su, Xiao (2009)," Regression Analysis Theory and Computing", London, World Scientific Publishing Co. Pte.Ltd.

33. Zimmerman, Donald W. (2013)," Heterogeneity of variance and biased hypothesis tests", Journal of Applied Statistics, Vol. 40, No. 1, January 169–193.

الملاحق

الملاحق

ملحق(1)

بيانات نسبة الاصابة بمرض اليرقان (Y) والعوامل التي تؤثر على المرض (X_i) لعام ٢٠١٨ للاشهر (كانون الثاني - شباط - اذار)

Y	X_1	X_2	X_3	X_4
0. 1	2.5	20	30	8
9.8	2.5	5	35	3
2	3	1	25	2
5	3	10	25	3
6.6	2.5	15	30	2
13.3	2.5	6	36	2
0.3	2.5	20	30	4
6	3	3	25	3
11.1	4	7	23	3
5.5	2.5	6	30	2
14.1	1.5	7	25	1
13	3	12	25	4
7.9	3	6	35	1
13.4	3	4	35	3
8.1	2.5	10	25	1
9.8	2.5	10	25	2
4.1	2.5	10	25	6
1.2	3	2	29	1
3	2	1	24	1
3.7	3	7	30	2
7	2	5	21	2
9.7	3	4	30	2
6.6	2.5	7	27	2
9.2	3	9	30	3
11.9	2.5	10	35	3

13.6	2.5	3	20	4
6.1	3	3	30	5
5.1	3	3	25	3
10	3.5	7	25	3
10.8	3	6	30	2
8.2	3.5	6	30	2
9.9	2.5	6	40	1
7.6	2.5	6	31	1
9.1	3	1	25	1
9.7	2.5	8	30	2
6.8	3	17	25	3
10.1	3	1	25	1
15.3	3	4	20	1
11	3	4	25	2
9.5	4	4	24	1
9.7	2.5	5	25	2
8.1	2	11	22	3
8.3	3	3	21	3
12.5	3	5	25	3
7.1	2.6	7	23	2
11.6	3	8	25	1
7.5	3	4	26	2
11.8	3	6	35	1
6.3	2.5	4	30	1
10.2	3	1	25	1
9.9	2.5	4	30	3
6.5	2.5	3	28	2
9.9	3	5	30	1
5.6	3	5	20	2
6.8	2.5	4	30	4
9.9	2.5	7	25	2
6.6	2.5	3	25	4
5.4	2.5	5	22	4
7.8	3	6	29	2

7	2.5	10	30	3
7.1	2.5	7	30	2
16.5	2.5	7	30	4
7.8	2.5	4	30	2
5.9	2.5	5	30	2
6.8	3	7	30	1
10	2.5	8	25	2
6	3	6	30	2
13.2	3	3	21	2
9.3	2.5	5	30	3
12.9	2.5	5	35	3
13.7	2.5	3	25	2
6.4	1.9	3	23	5
11.1	3	9	25	1
8.7	3	4	30	2
10.3	3	3	29	2
10	3	9	25	2
7.4	3	3	22	1
7.4	3.5	5	22	2
10.4	2.5	3	30	1
7.6	3	6	25	3
8.9	3	6	30	3
7.6	3	11	22	2
10.6	2.5	7	23	2
14	1.9	5	22	2
9.1	2.5	7	18	2
7.3	2.8	4	18	2
3.1	2.5	17	30	1
13.8	4	6	20	2
9.2	3.5	5	25	1
9.2	3.5	6	25	1
8	1.9	12	20	1
9.6	3	4	25	1
10	2.5	5	21	1

6.8	2.5	6	25	1
6	2	4	25	1
4.4	1.9	7	23	1
6.7	2.5	6	30	2
9.3	2.5	7	30	2
7.2	2.5	1	21	1
5.3	3	1	22	1

المصدر : وزارة الصحة العراقية / دائرة صحة المثنى / مستشفى الرميثة العام

Y : المتغير المعتمد واللذي يمثل نسبة الاصابة للطفل بمرض اليرقان.

X_1 : المتغير التوضيحي والذي يمثل وزن الطفل المصاب (بالكغم).

X_2 : المتغير التوضيحي واللذي يمثل عمر الطفل المصاب (بالايام).

X_3 : المتغير التوضيحي واللذي يمثل عمر الام للطفل المصاب.

X_4 : المتغير التوضيحي واللذي يمثل فترة رقود الطفل المصاب (بالايام).

ملحق رقم (2)

البرنامج (لغة R) المستعمل في اسلوب المحاكاة

```
#Install. Packages ("tseries")

#Install. Packages ("olsrr")

#Install. Packages ("car")

#Install. Packages ("lmtest")

library (tseries)

library (olsrr)

library (car)

library (lmtest)

iteration<-1000

n<-20

alpha<-0.05

mu1=0; mu2=0; mu3=0; mu4=0

sg1=1; sg2=1; sg3=1; sg4=1

x1<-rnorm (n,mu1,sg1)

x2<-rnorm (n,mu2,sg2)

x3<-rnorm (n,mu3,sg3)

x4<-rnorm (n,mu4,sg4)

x<- cbind (x1, x2, x3, x4)

Beta<-c (0.732,-0.112, 0.034,-0.123)

nn<-rep(1:n,2)

sigma2 = nn^23

R.BP=0
```

```
R.NCV=0

R.white=0

R.GQ=0

R.HM=0

i.BP.p.value<-NULL

i.NCV.p.value<-NULL

i.white.p.value<-NULL

i.GQ.p.value<-NULL

i.HM.p.value<-NULL

h = function(x) {15+.4*x^2}

a<-1: n

for (it in 1:iteration) {

err1 <- c (rnorm (n/2,  sd=1), rnorm (n/2, sd=1))

err2 <- rnorm (n,a,h(a) )

y<-x%*%Beta+err1

data<- data.frame (y,x)

model <- lm(y ~., data = data)

par (mfrow=c(2,2))

plot (model)

# Breusch-Pagan test

BP.test<-bptest (model)

BP.p.value<-BP.test$p.value

# Non-constant Variance Score Test (NCV test)

NCV.test<-ncvTest (model)

NCV.p.value<-NCV.test$p
```

```
# White test:

# ===========

model2 <- lm ((model$residuals)
(^2~x1+x2+x3+x1^2+x2^2+x3^2+x1*x2+x1*x3+x2*x3)

summary (model2)

# To get the R-squared:

summary (model2)$r.squared

 # A quick way to get n (the number of observations used in model1) is:

length (model$residuals)

 # The White test statistic (n times R-squared) is

White.test.statistic <- length (model$residuals)*summary (model2) $r.squared

White.test.statistic # displays the value

 # The p-value of obtaining a chi-squared value of as much as this test statistic or
more:

white.p.value<-pchisq (White.test.statistic, 2, lower.tail = FALSE)

# (Carefully check what the degrees of freedom are! In this example, df= 2).

## generate a regressor

#  x <- rep(c (-1, 1), 50)

## generate heteroskedastic and homoskedastic disturbances

# err1 <- rnorm(100,  sd=rep(c(1,2), 50))

## generate a linear relationship

## perform Goldfeld-Quandt test

GQ.test<-gqtest (model)

GQ.p.value<-GQ.test$p.value

## perform Harrison-McCabe test

HM.test<-hmctest (model)
```

```
HM.p.value<-HM.test$p.value

if (BP.p.value<alpha){R.BP=R.BP+1

if (NCV.p.value<alpha){R.NCV=R.NCV+1

if (white.p.value<alpha){R.white=R.white+1

if (GQ.p.value<alpha){R.GQ=R.GQ+1

if (HM.p.value<alpha){R.HM=R.HM+1

i.BP.p.value<-cbind (i.BP.p.value,BP.p.value

i.NCV.p.value<-cbind (i.NCV.p.value,NCV.p.value

i.white.p.value<-cbind (i.white.p.value,white.p.value

i.GQ.p.value<-cbind (i.GQ.p.value,GQ.p.value

i.HM.p.value<-cbind (i.HM.p.value,HM.p.value

{  # end of iteration

rbind (i.BP.p.value,i.NCV.p.value,i.white.p.value,i.GQ.p.value,i.HM.p.value)

 # There is a problem

round (cbind (R.BP,R.NCV,R.white,R.GQ,R.HM)/iteration*100 ,0)
```

المصدر : برنامج لغة البرمجة (R)

Abstract

This research is considered on attempt to highlight one of the problems of regression, which is one of the basic assumptions underlying ordinary least squares method (OLS). It is considered one of the fundamental conditions of variance analysis, and in many practical applications this hypothesis cannot achieve thus (OLS) to estimate the model of linear regression is use less (not benefited) to give true and accurate, results and this leads to face the problem of (heteroscedasticity).

The aim of this thesis is to choose the best parametric test to detect this problem from a group of parametric tests. We wrote an program in (R) language for a comparison in the best of the ratio of the detection true for the presence or absence of a problem through simulation within chapter III has a virtual reality two of the models, the first has a problem of heteroscedasticity the other not contain the problem depending on five tests (Gold field Quant, Prush bejen, White, NCV, Harrison-McCabe).

Chapter four contains the Applied side of the study and depending on real data, in this chapter we choose the best test to detect the problem or not, then we estimate the regression model.

Conclusions and recommendations are found in the end of the study.

University of Karbala
College of administration and Economic
Department of statistic

Choosing the best test method for the problem of Heteroscedasticity in a multiple regression model with practical application

A thesis
Submitted to the council of the college of administration and Economic\ University of Karbala, as partial fulfillment of the requirements for the degree of Master of Science in statistic

By

Raed Asmar Abdullah

Super vised by

Prof. Dr.

Adnan Karim Najim-Aldin

2018 **1439**