مقارنة اختبار Z المعدل - للتباين مع اختبار ليفين المعدل بوجود القيم الشاذة باستعمال المحاكاة

أ.د عباس لفته كنيهر العقابي alafta@uowasit.edu.iq

م.م احمد رزاق عبد رمضان اللآمي

م.م حيدر رائد طالب

المستخلص : تعتبر مشكلة عدم تجانس التباين احد اهم الافتراضات الأساسية لتحليل الانحدار الخطي وغيرها من الأساليب الأحصائيه لذلك تعد من المواضيع المهمة في علم الإحصاء لما لها من تطبيقات مختلفة في العلوم الإنسانية وغيرها من المجالات الأخرى ، واغلب الباحثين يقع في خطأ كبير وهو تطبيق البيانات الحقيقية بدون اختبارها وبالتالي الحصول على نتائج مظللة ، ومن هذا المنظور اصبح لزاماً على الباحث اختبار وفحص البيانات والمعلومات لمعرفة طبيعة البيانات ولتوضيح العلاقة القائمة بين العوامل المختلفة وتحليل النتائج بصورة ادق واشمل . وبذلك تظهر اهمية البحث في استخدام الأساليب الاحصائية التي تهيئ له الوصف الموضوعي الدقيق لتجربته على اسس بعيده عن الاجتهادات والعوامل الشخصية التي قد تؤثر تأثيراً مباشراً على نتائج البحث، وأخيرا تم عرض أهم اختبارات تجانس التباين المعدلة في حالة وجود القيم الشاذة لتشخيص مشكلة عدم تجانس التباين وتمت المقارنة بين اختبار Z المعدل – للتباين واختبار ليفين المعدل باستعمال المحاكاة والاعتماد على معيار قوة الاختبار.

Abstract: The problem of non-homogeneity variance one of the basic assumptions of the analysis of variance and analysis of linear regression and other statistical methods for it is one of the important topics in statistics because of their different applications in the humanities and other areas, and most of the researchers is a big mistake, a real application data without testing and thus get a shaded results , from this perspective has become imperative for the researcher test and check data and information to determine the nature of the data and to explain the relationship between different factors and analysis of results in a more accurate and comprehensive. Thus, the importance of research in the use of statistical methods that create a substantive exact description of his experience on the foundations appear far from the interpretations and personal factors that have a direct impact may affect the search results, and finally has been the most important homogeneity modified variation tests in the case of abnormal values for the diagnosis of the problem of lack of homogeneity of variance and then after that, the application of tests on the fact that data and the comparison between them.

Keywords: Power of the test, simulation, Modified Z-variance test, Modified Levene test.

1-المقدمة:

إن موضوع عدم تجانس التباين هو موضوع مهم وأوضحته العديد من البحوث والدراسات المختلفة إذ يعد من المواضع المهمة في علم الإحصاء لما له من تطبيقات مختلفة في العلوم الإنسانية والطبيعية ومن أهم الفرضيات الأساسية التي يتم الاعتماد عليها في أنموذج الانحدار الخطي المتعدد هي فرضية تجانس تباين الخطأ التي تمثل بالصيغة الآتية:

$$E(\underline{U}\,\underline{U}') = \sigma_u^2 I_n \qquad \dots\dots (1)$$

إما في حالة عدم تجانس التباين فان الفرضية تمثل بالصيغة الآتية:

$$E(\underline{U}\,\underline{U}') = \begin{bmatrix} \sigma_1^2 & 0 \dots & \dots \\ 0 & \sigma_2^2 & \dots \\ \vdots & \vdots & \ddots \sigma_n^2 \end{bmatrix} \qquad \dots\dots (2)$$

وإن وجود مشكلة عدم تجانس التباين في البيانات تؤثر على كل من التقدير وإجراءات الاختبار لتحليل الانحدار ، لذلك من المهم إن يتم الكشف عن هذه المشكلة لأنها تؤدي بالنتيجة الى قرارات خاطئة ونتائج مظلله ،وإن وجود بعض المشاهدات الشاذة أو غير العادية التي غالبا ما نسميها بالقيم الشاذة هي ميزة شائعة جدا في تحليل البيانات وأن وجود هذهِ القيم الشاذة في البيانات تؤدي إلى ظهور مشكلة عدم تجانس التباين وبالتالي يواجه الباحث صعوبة في تحليل البيانات ، لذلك يجب معرفة كيفية التعامل مع هكذا نوع من البيانات لأنها تؤثر في عملية التقدير. وان وجود القيم الشاذة في النموذج ممكن ان تحول الانموذج المتجانس الى انموذج غير متجانس ، لذا سوف تقوم في هذه الدراسة بالتحقيق في العديد من النقاط وهي كيفية عمل اختبارات عدم التجانس (Heteroscedasticity) في حالة احتواء البيانات على قيم شاذة إذ تبين الدراسات السابقة إن الاختبارات الاعتيادية (الكلاسيكية) تعاني من سوء التحليل عندما تحتوي هذه البيانات على القيم الشاذة ، ومن هنا بدأ الاهتمام بتعديل بعض الاختبارات بما يتناسب مع هكذا حالات وللدور المهم الذي تقدمة هذا الاختبارات في التعامل مع البيانات في حاله احتوائها على القيم الشاذة .

قام الباحث (Levene 1960)، باختبار فرضية تساوي التباينات للمتغيرات المستقلة من خلال طريقة حصينة تتمثل بتحويل الفرضية من حالة القياس(scale) بالنسبة للمتغيرات الظاهرة المدروسة لتمثل حالة الموقع(location) بالنسبة للمتغيرات أخرى تمثل القيمة المطلقة للانحرافات (انحراف القيم عن الوسط الحسابي) ومن ثم تطبيق اختبار (t-test) للعينة في حالة وجود متغيرين مستقلين واختبار (F-test) لتحليل التباين باتجاه واحد (one way of analysis of variance). وكذلك قدم الباحثان (Overall & Woodward 1974) اختبار (z – التباين)، استنادا إلى صيغه اختبار (Fisher & Yates) عام 1963 م، من خلال تحويل إحصاءه مربع كاي (Chi - square) مع درجات حرية (degree freedom) إلى المستوى العادي حيث يكون أداءها جيد جدا مع بيانات التوزيع الطبيعي ولكن تنتج العديد من الأخطاء ومنها خطأ النوع الأول عندما تكون العينات من توزيع (Skewed or leptokurtic)

وتوصل الباحثان (Olsson & Sandvik 1982)، إلى أول اختبار حصين يتم تطبيقه في حالة تساوي التباينات للبيانات المزدوجة من خلال تطبيق أهم الاختبارات اللامعلمية لمعالجة حالة ابتعاد البيانات عن التوزيع الطبيعي والتي تخص المقارنة بين موقعين متغيرين مرتبطين باستعمال تطبيق اختبار (Wilcoxons Signed Rank – Test). كما وقدم الباحثان (1988 Merran & Maxewll) ، بحثا تضمن بعض اختبارات عدم تجانس التباين في نموذج الانحدار وكذلك درسا أيضا القيم الحرجة للاختبارات ومن خلال مقارنة هذه الاختبارات التي تم دراستها في البحث مع اختبارات أخرى باستعمال تجارب مونتي كارلو للمحاكاة .كما وقدم الباحث (1999 القيسي) ، بحثا تضمن أسلوب بيز وبيز التجريبي للتشخيص ومعالجة مشكلة عدم تجانس التباين ، بالإضافة إلى ذلك قام الباحث أيضا بالمقارنة بين الطرائق التقليدية مع الطرائق البيزية ولكي يقدم أفضل الطرائق قام الباحث باستعمال أسلوب المحاكاة للمقارنة بين الطرائق . وأوصى الباحثان (Bootsrap & Forythe 2004) باستعمال الوسيط "Median" في صيغة اختبار(Levene) في إحصاءه اختبار (Brown-Forsythe) حيث تم الحصول على (P-value) عن طريق (bootstrop) التي تكون أكثر قوة من صيغة توزيع (F). كما واجرى كل من الباحثان (Lemeshko and Gorbuneva 2012) دراسة لتطبيق الاختبارات المعلمية لتجانس التباين في ظل انتهاك الافتراضات الأساسية لتحليل التباين. و قدم الباحثين (Violeta De La Huerta Contrerasa, Humberto Vaquera Huertaa & Barry C. Arnoldb 2014) ، بحثا اقترحا فيه اختبارا لتجانس التباين (type II right-censored) ويستند هذا الاختبار على إحصاءه اختبار بارتليت (Bartlett's) ويتوزع مربع كأي بدرجة حرية (1) وباستخدام تجارب المحاكاة من إحصاءه (Bartlett's) المعدلة لاختبار تجانس التباين تم الحصول على الفروق لأثنين من المجموعات في حالة عدم تساوي العينات . كما ودرس الباحث (2014 فرحان)، اختبار فرضيات تجانس تباين الأخطاء في حالة عدم توفر الشروط والفرضيات الخاصة بها من خلال دراسة نوعين من الاختبارات والمقارنة فيما بينهما، النوع الأول يمثل الاختبارات المعلمية إذ تم دراسة سلوك هذه الاختبارات مع أربع حالات من توزيع البيانات ولغرض تحقيق المقارنة بين الاختبارات المعلمية تم استعمال أسلوب المحاكاة لاختبار فرضية تساوي تباين الخطأ من خلال احتساب الخطأ من النوع الأول وقوة الاختبار لهذه الاختبارات. كما واجرى كل من الباحثان (Lemeshko and Sataeva 2016) ، لدراسة خواص وقوة الاختبارات الأحصائيه الكلاسيكية لتجانس التباينات في ظل غياب بعض الافتراضات الأساسية المهمة الخاصة بتحليل البيانات .

2- الجانب النظري

سيتم في هذا المبحث معرفة الافتراضات الأساسية لأنموذج الانحدار الخطي المتعدد وكذلك معرفة ماهي القيم الشاذة وطريقة التعامل معها واخيراً عرضت اختبارات تجانس التباين المعدلة.

2-1مشكلة عدم تجانس تباين الخطأ في نموذج الانحدار [1][4]

ان مشكلة عدم تجانس التباين تعتبر من اهم المشاكل التي تؤدي الى الحصول على نتائج غير دقيقة ومظللة لأي نموذج احصائي ووجود هذه المشكلة بالتحديد يخلق كثير من المشاكل التي تواجه الباحثين وتؤثر على مختلف طرق التقديرات الكلاسيكية للانحدار الخطي وتحدث هذه المشكلة عندما يكون هناك تفاوت كبير بدرجة التباين وخاصة عندما يتم الاعتماد على البيانات المقطعية في تقدير معالم الأنموذج ، لذلك من المهم جدا التأكد من عدم وجود هذه المشكلة من

خلال الاختبارات التي تهتم بدارسة تجانس التباينات ، ومن اجل دراسة هذه الظاهرة نفترض إن لدينا (X, Y) متغيرات لأنموذج الانحدار الخطي المتعدد الذي يمثل بالصيغة الآتية:

$$\underline{Y} = X\underline{\beta} + \underline{U} \qquad \dots\dots (3)$$

إذ إن :

$\underline{Y}$: متجه عمودي من المشاهدات المتغير المعتمد ذو بعد $(n\times1)$.

$\underline{\beta}$: متجه عمودي من المعلمات ذو بعد $1\times(r+1)$.

X : مصفوفة المعلومات ذو بعد $n\times(r+1)$ من مشاهدات المتغيرات التوضيحية .

$\underline{U}$: موجه عمودي ذو بعد $(n\times1)$ من قيم الأخطاء العشوائية .

ولكي نتمكن من تقدير متجه معالم النموذج $(\underline{\beta})$ يجب أن تتحقق الفرضيات الأساسية ول (r) من المتغيرات التوضيحية إذ يستند الأنموذج أعلاه الى الفرضيات الأساسية لخواص الخطأ العشوائي $(\underline{U})$ في العلاقة المدروسة:-

$$E(\underline{U}) = 0$$

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

$$\left[E(\underline{U}\acute{U}) = \sigma^2 I_n\right] = var(\underline{u})$$

وذلك يعني بأن قيم تباين $(\underline{U})$ حول متوسطها تكون مستقرة، علما إن (I_n) تكون مصفوفة خاضعة للعديد من لشروط وهي مربعة ذو بعد $(n\times n)$، وأحاديه (Unit Matrix)، بالإضافة الى شرط أخير وهو يجب إلا تكون بين المتغيرات المستقلة مشكلة التعدد الخطي.

ويمكن تلخيص نقاط الفرضيات الأساسية بالشكل الآتي:

$$\underline{U} \sim i.i.d(\underline{0}, \sigma^2 I_n) \qquad \dots\dots (4)$$

وعلى أساس الافتراضات الأساسية لنموذج الانحدار الخطي المتعدد يكون تباين حد الخطأ العشوائي بالصيغة الآتية:

$$\sigma_1^2 = \sigma_2^2 = \dots\dots\dots\dots = \sigma_n^2$$

وتعرف الفرضية أعلاه بفرضية تجانس تباين الخطأ (Error Homogeneity of variance Assumption) ولكن في اغلب الأحيان لا تتحقق هذه الفرضية ، وإذا لم تتحقق فرضية تجانس التباين فان ذلك يؤدي إلى عدم ثبوت تباين الخطأ بين المشاهدات .

$$\sigma_1^2 \neq \sigma_2^2 \neq \dots\dots\dots \neq \sigma_n^2$$

وهذا يؤدي إلى ظهور مشكلة عدم تجانس تباين الخطأ (Heteroscedasticity)

2-2-القيم الشاذة [5] **outlier values**

يمكن تعريف القيم الشاذة : بانها المشاهدة التي تكون بعيدة جدا عن بقية المشاهدات الأخرى ضمن البيانات المستخدمة تحت الدراسة إي ضمن العينة الواحدة بمعنى أخر تكون هذه المشاهدات غير متسقة (inconsistent) مع المشاهدات الأخرى التي تكون تحت الدراسة مما تثير الشكوك حولها وتسمى القيم الشاذة ، وذكر الباحثون انه لا يمكن التعامل مع هذه القيم بواسطة الطرق التقليدية الكلاسيكية التي تستخدم في تقدير معالم الانحدار لأنها تكون غير كفوءة في الحصول على مقدرات جيدة ولذلك كان لابد من إيجاد طرائق بديلة لا تتأثر بهذه القيم وتكون اقل حساسية تجاه هذه القيم ولقد اثبتت الدراسات والبحوث التي قدمت في هذا المجال إن المقدرات الحصينة قادرة على التعامل مع وجود قيم الشواذ في البيانات ، كما نعلم هناك العديد من الطرائق الحصينة التي يمكن استخدامها إلا إن اغلب هذه الطرائق تشترك في نقطة أساسية هي : إعطاء وزن اقل للمشاهدة التي تبتعد (المشاهدة الشاذة) عن المشاهدات الأخرى وإن الغرض من إعطاء وزن هو لكي نقلل من تأثير هذه المشاهدة الشاذة. وهناك العديد من الطرائق لكشف الشواذ بعضها يكون بياني كطريقة (Boxplot)، وبعد اكتشاف الشواذ في البيانات هناك طريقتان للتعامل مع البيانات هما:

الطريقة الأولى: هي حذف أو التخلص من المشاهدات الشاذة وإذا تم حذف هذه المشاهدات يعتبر تحريف أو خلل في البيانات.

الطريقة الثانية: هي ان يتم التعامل مع هذه المشاهدات من خلال الاعتماد على المقدرات الحصينة بدلا عن المقدرات العادية التي تستخدم في الطرق الكلاسيكية التي تم ذكرها سابقا، وهناك العديد من الطرائق الحصينة وتكون المفاضلة بين هذه الطرائق.

2-3اختبارات عدم تجانس التباين في حالة وجود القيم الشاذة:

هناك العديد من الاختبارات التي تستخدم للكشف عن مشكلة عدم تجانس التباين ولكن تكون هذه الاختبارات غير فعالة إي بمعنى تكون غير كفوءة عندما تحتوي البيانات على قيم شاذة لذا سوف يتم عرض بعض اختبارات تجانس التباين المعدلة حيث يتم أولا اكتشاف مشكلة عدم تجانس تباين الخطأ بواسطة الاختبارات الاعتيادية ثم بعد ذلك يتم عرض عمل الاختبارات المعدلة وبعد ذلك عمل المقارنة بين هذه الاختبارات لمعرفة إي الاختبارات المستخدمة هي الأفضل والأكفاء من بين الاختبارات المستخدمة في هذه الدراسة في ظل وجود مشكلة عدم تجانس التباين في حالة القيم الشاذة .

2-3-1اختبار Z- للتباين المعدل [6][10][9] **modified Z-variance test**

هو احد اختبارات تجانس التباين قدم من قبل الباحثان Overall and Woodward عام 1974 يستخدم للكشف عن مشكلة عدم تجانس التباين وان هذا الاختبار له اداء جيد للغاية في حالة البيانات التي يتم توزيعها بشكل طبيعي ، ولكنه

ينـتج خطـأ مـن النـوع الاول عنـدما تكـون توزيعـات العينـات منحرفـة او عنـد احتـواء هـذه العينات علي قيم شاذة ، وان فرضية الاختبار هي :-

$$H_0: \sigma_1^2 = \cdots = \sigma_K^2$$

$$H_1: \sigma_1^2 \neq \cdots \neq \sigma_K^2$$

وان صيغة اختبار Z -للتباين تتمثل بالصيغة التالية:

$$Z = \frac{1}{m-1}\sum_{i=1}^{m} {z_i}^2 \qquad \ldots (5)$$

$$z_i = \sqrt{\frac{c_i(n_i-1){s_i}^2}{MSE}} - \sqrt{c_i(n_i-1) - \frac{c_i}{2}} \qquad \ldots (6)$$

$$c_i = 2 + \frac{1}{n_i}$$

$$MSE = \frac{1}{N-m}\sum_{i=1}^{m}\sum_{j=1}^{ni}(x_{ij} - \overline{x}_i)^2$$

m : عدد العينات

n_i : حجم العينة لـ i-th من المشاهدات

${s_i}^2$: مقدر غير متحيز لتباينات العينة

$\overline{x}_i$: متوسط العينة i ، $N = \sum_{i=1}^{m} n_i$

وكما ذكرنا ان الاختبارات الاعتيادية تتأثر بوجود القيم شاذة او بانحراف البيانات عن التوزيع الطبيعي لذلك في عام 1976قام العالمان Overall and Woodward بدراسات جديدة لتحديد قيمة ci حتى تصبح التباينات z_i مستقرة وثابته وقد وجدوا ان قيمة c تعتمد على حجم العينة، وان قيمةc هي التي تحدد معامل القياس الذي يؤثر على التغير في قيم Z_i وبالتالي يعمل على معالجة مشكلة عدم تجانس التباينات، وان الصيغة الجديدة الى c_i هي [11] :

$$c_i = 2.0\left(\frac{2.9 + 2/n_i}{K}\right) \qquad \ldots\ldots (8)$$

وأن المؤشر الذي استخدمه العالمان Overall and Woodward هو:

$$K = \frac{\sum {Z_{ij}}^4}{n_i - 2} \qquad \ldots\ldots (9)$$

$$Z_{ij} = \frac{X_{ij} - \overline{X}_i}{\sqrt{\frac{n_i - 1}{n_i} {S_i}^2}} \qquad \dots\dots (10)$$

وبعدها استخراج قيمة c_i الجديدة، والمتمثلة بالصيغة الاتية:

$$c_i = 2.0 \left[\frac{1}{K_i}\left(2.9 + \frac{0.2}{n_i}\right)\right]^{\frac{1.6(n_i - 1.8K_i + 14.7)}{n_i}} \qquad \dots (11)$$

ومن ثم تعويض قيمةc_i بالقانون الاتي:

$$Z_i = \sqrt{\frac{c_i(n_i - 1){S_i}^2}{MSE}} - \sqrt{c_i(n_i - 1) - \frac{c_i}{2}} \qquad \dots (12)$$

وبعدها نطبق قانون Z – للتباين الأصلي:

$$F = \frac{\sum_{i=1}^{k} {Z_i}^2}{K - 1} \qquad \dots (13)$$

وتقارن قيمة احصاءة الاختبار مع قيمة F الجدولية بدرجة حرية $(k - 1\ ,\ \alpha)$ وبمستوى معنوية (0.05) فاذا كانت قيمة احصاءة الاختبار اكبر من قيمة F الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة الي تنص على عدم تساوي التباينات اي ان التباينات غير متجانسة أما اذا كانت احصاءة الاختبار اقل من قيمة F فهذا يعني قبول فرضية العدم ورفض الفرضية البديلة اي ان التباينات متساوية .

2-3-2اختبارليفين المعدل [6][7] Modified Levene test

هو اختبار إحصائي يستخدم في الإحصاء الاستدلالي لتقيم المساواة بين الفروق لمتغير يحسب لمجموعتين أو أكثر بشرط إن تكون متساوية حيث ناقش الباحث (Howard Levene) عام 1960م مشكلة اختبار فرضية التجانس لـ(k) من المتغيرات المستقلة التي يبتعد جزء أو كل منها عن التوزيع الطبيعي ويعد اختبار ليفين اقل حساسية في حالة خروج العينة من الافتراضات الأساسية إلا انه اقل قوة، وان فرضية الاختبار هي:

$$H_0: \sigma_1^2 = \cdots = \sigma_K^2$$
$$H_1: \sigma_1^2 \neq \cdots \neq \sigma_K^2$$

المتمثل بالصيغة الآتية:

$$W = \frac{(N - K)\sum_{i=1}^{K} N_i(\overline{Z}_{i.} - \overline{Z}_{..})^2}{(K - 1)\sum_{i=1}^{K}\sum_{j=1}^{ni}(\overline{Z}_{ij} - \overline{Z}_{i.})^2} \qquad \dots\dots (14)$$

ni : عدد القيم في المجموعة (i) N :العدد الكلي لكل المجموعات .

k: يمثل عدد المجموعات.

$Z_{ij} = | Y_{ij} - \overline{Y}_{i.} |$ ، و ان $\overline{Y}_{i.}$: متوسط المجموعة i .

$\overline{Z}_{..} = \frac{1}{N}\sum_{i=1}^{n1}\sum_{j=1}^{n2} Z_{ij}$ ، $\overline{Z}_{i.} = \frac{1}{N}\sum_{i=1}^{n1} Z_{ij}$

إما في هذا البحث تم الاعتماد على الوسيط (median) بدلا من (mean) المستخدم في الاختبار (Levene) الاعتيادي في حساب Z_{ij} ، وهي الصيغة التي توصل إليها الباحثان Brown & (Forsyth) في عام 1974م وعادةً ما يسمى هذا الاختبار باختبار (Brown & Forsyth)، وتكون صيغة الاختبار النهائية على النحو الاتي:

$$H_0: \sigma_1{}^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

$$H_1: \sigma_1^2 \neq \ldots \neq \sigma_k^2$$

$$ML = \frac{(N-k)\sum_{i=1}^{k} Ni\,(\bar{z}i. - \bar{z}..)^2}{(k-1)\sum_{i=1}^{ki}\sum_{j=1}^{ni.}(zij - \bar{z}i.)^2} \qquad \ldots\ldots (14)$$

$$\text{Where} \qquad Z_{ij} = \left| y_{ij} - \tilde{y}_{i.} \right|$$

حيث إن:

K: هو عدد اختلاف المجموعات التي شملتها العينة.

N: المجموع الكلي للعينة.

ni: حجم العينة في المجموعة i .

Yij: متجه قياس المتغير j في المجموعة i .

.ỹi: الوسيط (median) للمجموعة الفرعية i .

.Zi: المتوسط (mean) المتوسط ل(zij) للمجموعة الفرعية (i) .

$\bar{z}$: المتوسط العام لكل (zij) .

وبالتالي يمكن تكوين قاعدة القرار هي رفض فرضية العدم وقبول الفرضية البديلة في حالة:

$$ML > F(\alpha, K-1, N-K) \qquad \ldots (15)$$

حيث إن $F(\alpha, K-1, N-K)$ هي القيمة الحرجة العليا لتوزيع (F) مع درجة حرية(k-1) و (N-k) عند مستوى دلالة (α) ، إما في حالة عدم تحقق الصيغة (23) نقبل فرضية العدم ونرفض الفرضية البديلة هذا يعني ان التباينات غير متساوية وهناك فروق معنوية بين المتغيرات .

3-الجانب التجريبي /وصف تجربة المحاكاة [2]

لأجل الوصول الى الهدف المتوخى من الدراسة وهو المقارنة بين طرائق اختبار فرضية عدم تجانس التباين تم اللجوء الى استخدام المحاكاة لتوليد البيانات لغرض الوصول الى الطريقة المثلى للكشف عن مشكلة عدم تجانس التباين وقد تم استعمال اسلوب مونت كارلو وبالاعتماد على نموذج الانحدار الخطي المتعدد وسيتم وصف مراحل تجربة المحاكاة من خلال الخطوات الآتية:

- تم توليد عينات بأحجام مختلفة (14، 20، 60 ،100) للمتغير المعتمد (y_i) على وفق نموذج الانحدار الخطي المتعدد:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \quad \dots (16)$$

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

حيث إن المتغير العشوائي (u_i) يتبع التوزيع الطبيعي ويتم توليدة بمتوسط $(0, \sigma^2)$.

- بعد إكمال المرحلة الاولى تم كتابة فرضيات عدم تجانس التباين التي على ضوئها يتم رفض وقبول فرضية العدم والفرضية البديلة.
- إن القيم التي تم توليدها كانت تحتوي على ثلاثة مراحل في المرحلة الأولى كانت نسبة القيم الشاذة (10%) وفي المرحلة الثانية كانت نسبة القيم الشاذة (20%) اما في المرحلة الثالثة فقد كانت نسبة القيم الشاذة فيها هي (30%) من البيانات.
- تم كتابة القيم الجدولية التي استخرجت من الجداول الخاصة بإحصاءات الاختبار.
- وتم تعيين قيم افتراضية للمعالم وتعتبر من الخطوات المهمة للحصول على نموذج الانحدار الخطي وهذه المعالم موضحه كما يأتي:

$$(\beta_0 = 3, \beta_1 = -1.5, \beta_2 = -1, \ \beta_3 = 1.5)$$

$$(\beta_0 = 4, \beta_1 = -2.5, \beta_2 = -2.5, \ \beta_3 = 3)$$

- تم تقدير معلمات الأنموذج بطريقة المربعات الصغرى الاعتيادية (OLS).
- وتم توليد الخطأ العشوائي في نموذج الانحدار الخطي بالاعتماد على نسبة القيم الشاذة المفترضة مسبقا.
- وبعد ان تم استخراج نموذج الانحدار المتعدد واحتساب معدل الخطأ العشوائي تمت كتابة اختبارات تجانس التباين التي سيتم المقارنة فيما بينها وهي اختبار (Z المعدل - للتباين، ليفين المعدل)
- تم تكرار التجربة (1000) مرة.
- تم احتساب الاختبار الأفضل من بين الاختبارات المذكورة عن طريق المقارنة بالاعتماد على معيار قوة الاختبار $\hat{\pi}$ والتي يتم ايجادها باستعمال الصيغة الأتية: -

$$\hat{\pi}(T) = \frac{K(T)}{R}$$

وان:

$K(T)$:هو عدد مرات رفض احصاءة الاختبار (T) للفرضية (H_0)الخاطئة.

R :عدد مرات تكرار التجربة.

عرض وتحليل نتائج تجربة المحاكاة

بعد كتابة برنامج المحاكاة وتحديد قيم معالم نموذج الانحدار الخطي الافتراضية سيتم عرض نتائج المحاكاة وقوة الاختبار وتحليل نتائجها في حالة وجود مشكلة عدم تجانس التباينات والقيم الشاذة كانت النتائج كما يأتي: -

الحالة الاولى

عندما تكون نسبة القيم الشاذة في البيانات (10%) وتقديرات معالم الأنموذج هي

$$(\beta_0 = 3, \beta_1 = -1.5, \beta_2 = -1,\ \beta_3 = 1.5)$$

من خلال النظر الى الجدول (1) نلاحظ ان

- اختبار z المعدل – للتباين كان يمتلك قوة اختبار أفضل بالمقارنة مع اختبار ليفين المعدل.
- هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.
- في حالة حجم عينة 14 نلاحظ وجود اقل قوة اختبار اي ان الاختبارات تكون ضعيفة في حجوم العينات الصغيرة.
- في حالة حجوم العينات الكبيرة نلاحظ وجود اعلى قوة الاختبار اي ان الاختبار يعمل بشكل اكفئ وأفضل مع العينات الكبيرة.
- كانت قيمة اختبار ليفين المعدل تساوي صفر في حالة حجم العينة 14 اي ان هذا الاختبار يكون ضعيف جدا مع العينات الصغيرة.

جدول (1) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل – للتباين	0.0930	0.1250	0.2950	0.3360
ليفين المعدل	0	0.05181	0.0930	0.1380

عندما تكون نسبة القيم الشاذة في البيانات (10%) وبتغيير معالم الانموذج الى: -

$$\beta_0 = 4, \beta_1 = -2.5, \beta_2 = -2.5, \quad \beta_3 = 3$$

بعد تغيير معالم نموذج الانحدار الخطي المتعدد وعرض النتائج في جدول (2) نلاحظ ان: -

- ♦ اختبار z المعدل – للتباين كان يمتلك قوة اختبار أفضل بالمقارنة مع اختبار ليفين المعدل.
- ♦ هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.
- ♦ اختبار ليفين كان الأضعف وبالتحديد في حالة حجوم العينات صغيرة الحجم.
- ♦ في حالة حجوم العينات الكبيرة نلاحظ وجود اعلى قوة الاختبار اي ان الاختبار يعمل بشكل اكفئ وأفضل مع العينات الكبيرة.
- ♦ نلاحظ تغيير نتائج قوة الاختبار بعد تغيير معالم نموذج الانحدار الخطي.

جدول (2) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل – للتباين	0.1480	0.2430	0.3280	0.4050
ليفين المعدل	0	0.0790	0.1020	0.1490

الحالة الثانية

عندما تكون نسبة القيم الشاذة في البيانات (20%) وتقديرات معالم الأنموذج هي

$$(\beta_0 = 3, \beta_1 = -1.5, \beta_2 = -1, \ \beta_3 = 1.5)$$

من خلال النظر الى الجدول (3) نلاحظ ان: -

- ♦ اختبار z المعدل – للتباين كان يمتلك قوة اختبار افضل بالمقارنة مع اختبار ليفين المعدل .
- ♦ هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.
- ♦ هناك زيادة نسبية في قوة الاختبار بعد زيادة القطع بنسبة 20%
- ♦ في حالة حجم عينة 14 نلاحظ وجود اقل قوة اختبار اي ان الاختبارات تكون ضعيفة في حجوم العينات الصغيرة
- ♦ تكون قوة الاختبار الأفضل مع احجام العينات الكبيرة.

جدول (3) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل – للتباين	0.1270	0.4460	0.5390	0.6210
ليفين المعدل	0.0667	0.0955	0.1179	0.1640

عندما تكون نسبة القيم الشاذة في البيانات (20%) وبتغيير معالم الأنموذج الى: -

$$\beta_0 = 4, \beta_1 = -2.5, \beta_2 = -2.5, \qquad \beta_3 = 3$$

بعد تغيير معالم نموذج الانحدار الخطي المتعدد وعرض النتائج في جدول (4) نلاحظ ان: -

- ♦ هناك تفوق واضح في اختبار z المعدل – للتباين فقد كان يمتلك قوة اختبار أفضل بالمقارنة مع اختبار ليفين المعدل.
- ♦ هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.
- ♦ اختبار ليفين كان الأضعف وبالتحديد في حالة حجوم العينات صغيرة الحجم.
- ♦ في حالة حجوم العينات الكبيرة نلاحظ وجود اعلى قوة الاختبار اي ان الاختبار يعمل بشكل اكفئ وأفضل مع العينات الكبيرة.
- ♦ وايضا هناك زيادة نسبية في قوة الاختبار عند ادخال قيم شاذة للبيانات بنسبة 20% عنها في حالة ادخال قيم شاذة بنسبة 10%.
- ♦ نلاحظ تغيير نتائج قوة الاختبار بعد تغيير معالم نموذج الانحدار الخطي.

جدول (4) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل– للتباين	0.1910	0.3150	0.6560	0.7010
ليفين المعدل	0.0960	0.1940	0.2179	0.2640

الحالة الثالثة

عندما تكون نسبة القيم الشاذة في البيانات (30%) وتقديرات معالم الأنموذج هي

$$(\beta_0 = 3, \beta_1 = -1.5, \beta_2 = -1,\ \beta_3 = 1.5)$$

من خلال النظر الى الجدول (5) نلاحظ ان: -

♦ اختبار z المعدل – للتباين كان يمتلك قوة اختبار افضل بالمقارنة مع اختبار ليفين المعدل فقد حصل على اعلى قوة اختبار مقدارها (0.8830) عند حجم 100، وكانت اقل قوة اختبار حصل عليها هنا كان مقدارها (0.2560) عند حجم 14 .

♦ هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.

♦ هناك زيادة نسبية في قوة الاختبار بعد زيادة القطع بنسبة 30%

♦ في حالة حجم عينة 14 نلاحظ وجود اقل قوة اختبار فقد كان مقدارها (0.1420) اي ان الاختبارات تكون ضعيفة في حجوم العينات الصغيرة

♦ تكون قوة الاختبار الأفضل مع احجام العينات الكبيرة.

جدول (5) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل – للتباين	0.2560	0.5570	0.7590	0.8830
ليفين المعدل	0.1420	0.2685	0.3045	0.4210

عندما تكون نسبة القيم الشاذة في البيانات (30%) وبتغيير معالم الأنموذج الى: -

$$\beta_0 = 4, \beta_1 = -2.5, \beta_2 = -2.5,\ \beta_3 = 3$$

بعد تغيير معالم نموذج الانحدار الخطي المتعدد وعرض النتائج في جدول (6) نلاحظ ان: -

♦ هناك تفوق واضح في اختبار z المعدل – للتباين فقد كان يمتلك قوة اختبار مقدارها (0.9040) عند حجم عينه 100 بالمقارنة مع اختبار ليفين المعدل الذي حصل على قوة اختبار مقدارها (0.4580) عند حجم عينة 100.

♦ هنالك علاقة طردية مع قوة الاختبار وحجوم العينات فكلما زاد حجم العينة كلما ازدادت قوة الاختبار.

♦ اختبار ليفين كان الأضعف وبالتحديد في حالة حجوم العينات صغيرة الحجم.

♦ وايضا هناك زيادة نسبية في قوة الاختبار عند ادخال قيم شاذة للبيانات بنسبة 20% عنها في حالة ادخال قيم شاذة بنسبة 10%.

♦ نلاحظ تغيير نتائج قوة الاختبار بعد تغيير معالم انموذج الانحدار الخطي.

جدول (6) نتائج قوة الاختبار في حالة مستوى معنوية (0.05)

حجم العينة / الاختبار	N=14	N=20	N=60	N=100
Z المعدل – للتباين	0.2170	0.4890	0.8220	0.9040
ليفين المعدل	0.1770	0.3180	0.3540	0.4580

الاستنتاجات: -

1- من دراسة نتائج الجانب التجريبي يمكن القول بان اختبار Z المعدل – للتباين كان الأفضل لأنه يقبل فرضية العدم والتي تنص على ان التباينات متساوية وبالتالي اثبت عدم تأثره بالقيم الشاذة.
2- بعد تطبيق اختبار ليفين المعدل لتجانس التباين قد تبين انه اختبار ضعيف في حالة استعماله مع حجوم العينات الصغيرة.
3- تبين ان هناك تأثيرا معنويا من قبل القيم الشاذة على معلمات نموذج الانحدار الخطي المتعدد حيث ان القيم الشاذة تعمل على تضخيم تباين الخطأ MSEوتقلل من قيمة R^2 وقيمة F .
4- وجود علاقة طردية بين قوة الاختبار وحجوم العينات فكلما يكبر حجم العينة كلما تكون قوة الاختبار أفضل.
5- وجود علاقة طردية بين القيم الشاذة وقوة الاختبار فكلما تكون نسبة القيم الشاذة في البيانات اعلى كلما ازدادت قوة الاختبار.

التوصيات: -

1- عند وجود التطرف في قيم المتغير (المتغيرات) التوضيحية أو في المتغير المعتمد أو في كليهما يفضل استخدام طريقة Z المعدل - للتباين.
2- ان الاستنتاجات التي تم التوصل اليها محدده فقط بالحالات المدروسة ولا يمكن تعميمها على حالات اخرى فمثلا عندما تكون عدد المشاهدات غير متساوية عند مختلف العينات فربما تكون النتائج التي يتوصل اليها مختلفة عن نتائج هذه الدراسة.
3- ضرورة تشخيص وتقليل أثر القيم الشاذة عند تطبيق نموذج تحليل الانحدار المتعدد للحصول على نموذج أفضل.

المصادر

1–القيسي ، باسم شليبة مسلم ،1999،"اسلوب بيز في اختبار ومعالجة مشكلة عدم التجانس التباين في النماذج الخطية" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد ، جامعة بغداد.

2– بتال ، احمد حسين & عصام كامل احمد &البراءة عبد الوهاب خضير ،(2008)،" استخدام المحاكاة في تدريس الانحدار الخطي البسيط"، كلية الادارة والاقتصاد ، مجلة جامعة الأنبار التعليم والصيرفة العدد الثالث ، المجلد الثاني .

3– رشيد، هدى عبدالله، 1997، " مقارنة بين الاختبارات الحصينة لعد التجانس في البيانات المزدوجة"، رسالة ماجستير في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

4–فرحان ، عقيل حميد ،(2014)، " بعض الاختبارات المعلمية والاستكشافية في تحديد مشكلة عدم التجانس في انموذج الانحدار الخطي المتعدد " ، رسالة ماجستير في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد .

5– يوسف ، عصام الدين يوسف عبد الله ، (2015) ،"تأثير القيم الشاذة في معلمات نموذج تحليل الأنحدار الخطي المتعدد "اطروحة دكتوراه في الأحصاء ، كلية الدراسات العليا ، جامعة السودان للعلوم والتكنولوجيا .

6- Lee H.B., Katz G.S. and Restori A.F(2010),"A Monte carlo Study of seven homogeneity of variance tests", Journal of Mathematics and Statistics, vol.6(3) , pp. 359-366 .

7- Lemeshko B.Yu. and Lemeshko S.B.(2008)," Power and robustness of criteria used to verify the homogeneity of means", Measurement Techniques, 51, 9, pp 950–959.

8-Lemeshko,B.Y.and Gorbuneva,A.A.(2012),"Appllication of paramet-eric homogeneity of variance test under violation of classical assumption ",proceedings , 2 nd stochastic modeling techniques and data analysis international conferece 5-8 , Chaina Crete Greece.

9- Lemeshko,B.Y.and Sataeva.T.S.(2016),"A bout properties and power of classical tests of homogeneity of variances" Novosibirsk State Technical University Novosibirsk, Russia.

10-Overall, J.E. & Woodward, J.A. (1974)." A simple test for homogeneity of variance in complex factorial design". Psychometrika, 39, 311- 318.

11-Overall, J.E. & Woodward, J.A. (1976). "A robust and powerful test for heterogeneity of variance". Galveston, TX: University of Texas Medical Branch Psychometric Laboratory. Report No. 57, 12 pp .

12-Potter, k.p.(2006)."Methods for Presenting Statistical Information: The Box Plot " University of Utah School of Computing Salt Lake City.