المحور الخامس: تحليل الانحدار الخطي Linear Regression Analysis

1- نموذج الانحدار الخطي البسيط

2- نموذج الانحدار الخطي المتعدد

إعداد الدكتور: إبراهيم رحيم-2023-2024

تمهيد:

تعتمد نماذج الانحدار على مقاييس العلاقة وتضفي عليها معنى أكثر عمقا في دراسة الظواهر المختلفة، وذلك بالوصول إلى معادلات (نماذج مقدرة) يمكن استخدامها في عمليات التقدير والتنبؤ.

تعريف الانحدار:

هو دراسة اعتماد المتغير التابع على متغير مستقل أو أكثر، وذلك لتقدير قيمة المتغير التابع من خلال الاستفادة من البيانات المتاحة التي تقدمها قيم المتغير المستقل أو المتغيرات المستقلة. ويؤكد هذا التعريف على أن نموذج الانحدار لا يكتفي بإيضاح العلاقة فقط، بل يسعى لكشف الاعتماد بين المتغيرين، بمعنى: أن أحدهما يؤثر في الآخر.

أهداف تحليل الانحدار:

يمكن القول بأن تحليل الانحدار يسعى إلى تحقيق هدف أو أكثر من الأهداف التالية:

- **الوصف**(Descriptive): من خلال تحديد شكل العلاقة بين المتغير التابع والمتغير أو المتغيرات المستقلة.
- **التقدير**(Estimation): من خلال تحديد أي من المتغيرات المستقلة كان ذا أثر أكبر في قيم المتغير التابع.
- **التنبؤ**(Prediction): تمهد المعطيات التي تقدمها قيم المتغير أو المتغيرات المستقلة للوصول إلى قيم المعلمات التي تستخدم في معادلة التنبؤ بقيم المتغير التابع للظاهرة محل الدراسة.
- **التحكم**(Controlling): وهو الإفادة من الأهداف السابقة في تعديل قيم المتغير المستقل بالزيادة أو النقص، بالشكل الذي يحقق الهدف المنشود في قيم المتغير التابع.

وفي هذا البرنامج سنتناول نماذج الانحدار التالية:

- ✓ الانحدار الخطي البسيط (Simple Linear Regression)،
- ✓ الانحدار الخطي المتعدد (Multiple Linear Regression).

1- نموذج الانحدار الخطي البسيط (Simple Linear Regression)

(تحديد قيمة كل معلمة للنموذج، اختبار الموثوقية، التنبؤ)

إعداد الدكتور: إبراهيم رحيم-2023-2024

1–نموذج الانحدار الخطي البسيط (Simple Linear Regression):

الانحدار الخطي البسيط هو من الأساليب المعتمدة في قياس العلاقات الاقتصادية، يهتم بدراسة وتحليل أثر متغير مستقل واحد على متغير تابع. ويسمى بالخطي لأن الصيغة الممثلة للعلاقة خطية، ووصف بأنه بسيط لأن عدد المتغيرات المستقلة محل الدراسة متغير واحد فقط.

1|2 معادلة النموذج الخطي البسيط

يتناول هذا النوع من النماذج العلاقة بين متغير واحد تابع (Y) ومتغير واحد مستقل (X)، ويأخذ نموذج الانحدار الخطي البسيط الشكل التالي:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i : \; i = \overline{1, n} \dots [1]}$$

Y_i: قيم المشاهدات في المتغير التابع،

X_i: قيم المشاهدات في المتغير المستقل

β_0: ثابت النموذج وهي نقطة التقاطع مع المحور العمودي،

β_1: الميل الحدي للانحدار،

ε_i: الخطأ العشوائي

<u>الشكل المصفوفي للنموذج</u>: لدينا

$$[1] \Leftrightarrow \begin{cases} Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \;\vdots \quad \vdots \quad \vdots \quad \vdots \\ Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n \end{cases}$$

ويمكن تحويل هذه الجملة إلى الشكل المصفوفي التالي:

$$\underset{[n\times1]}{Y} = \underset{[n\times2]}{X}\underset{[2\times1]}{\beta} + \underset{[n\times1]}{\varepsilon} \Leftrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \dots [1]$$

<u>2|2 الفرضيات الأساسية للنموذج</u>

(1) التوقع الرياضي للأخطاء معدوم، بمعنى: $\forall i = \overline{1, n} : \; E(\varepsilon_i) = 0$

والتي تكافئ الفرضية التالية: $\hat{Y} = E(Y_i) = \hat{\alpha} + \hat{\beta} X_i$

(2) تباين الخطأ العشوائي ثابت، بمعنى: $\forall i = \overline{1, n} : \; Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 = Var(Y_i)$

(3) استقلالية الأخطاء العشوائية، بمعنى: $\forall i \neq j: Cov(\varepsilon_i, \; \varepsilon_j) = Cov(Y_i, \; Y_j) = 0$

(4) المتغير X_i هو متغير غير عشوائي، بمعنى: $\forall i = \overline{1, n}: Cov(X_i, \; \varepsilon_i) = 0$

(5) قيم المتغير العشوائي (ε_i) تتوزع توزيعا طبيعيا حول وسطها الحسابي: $\varepsilon_i \sim N(0, 1)$

(6) أن يكون حجم العينة أكبر بكثير من عدد المعلمات المقدرة، بمعنى: $n \gg 2$

3|2 تقدير معلمات النموذج الخطي البسيط

بعد اختيار النموذج المناسب، نقوم بتقدير معلمتي هذا النموذج β_0 و β_1، واختبار معنويتها احصائيا. ومن أجل ذلك نختار الطريقة المناسبة والأكفأ لعملية التقدير وهي طريقة: "المربعات الصغرى العادية OLS " Ordenary Least Squered

1|3|2 تقدير المعلمات بطريقة المربعات الصغرى العادية

عند تمثيل المشاهدات (X.Y) في معلم متعامد تظهر لنا سحابة من النقط، ونحاول في الخطوة الموالية تقدير خط الانحدار الذي يشمل أكبر عدد من النقط ويمثل العلاقة بين X و Y أحسن تمثيل من خلال تهنئة مجموع مربعات الأخطاء العشوائية $\hat{\varepsilon}_i$. والشكل التالي سوف يوضح كيفية عمل طريقة المربعات الصغرى.

ويمكن الحصول على قيم β_0 و β_1، باستخدام طريقة المربعات الصغرى على النحو التالي:

نعلم أن: $\forall i = \overline{1,n} : \ \hat{\varepsilon}_i = Y_i - \widehat{\mathrm{Y}}_{\mathrm{i}}$

ولتقدير معالم النموذج $\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i}$ نقوم بتصغير مجموع مربعات البواقي:

أي أن: $Min(SSE) = \ Min \sum \left(Y_i - \hat{Y}_i\right)^2$

واعتمادا على الاشتقاق الجزئي للدالة $L\left(\widehat{\beta_0}, \widehat{\beta_1}\right) = \sum \left(Y_i - \hat{Y}_i\right)^2$ بالنسبة لكل من $\widehat{\beta_0}, \widehat{\beta_1}$ وإعدام هذه المشتقات الجزئية، وهو الشرط اللازم لبلوغ الدالة قيد الدراسة قيمتها الصغرى، يتم الحصول على الصيغ الرياضية للمقدرتين $\widehat{\beta_0}, \widehat{\beta_1}$:

على الشكل التالي:

$$\boxed{\begin{cases} \boldsymbol{\widehat{\beta_1} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \dfrac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum {X_i}^2 - n\overline{X}^2}} \\ \boldsymbol{\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\overline{X}} \end{cases}}$$

$\overline{X}$: المتوسط الحسابي للمشاهدات (X) ، **$\overline{Y}$: المتوسط الحسابي للمشاهدات (Y)**

أما الشرط الكافي هو: أن يكون محدد مصفوفة المشتقات الجزئية من الرتبة الثانية بالنسبة للمعلمات المقدرة موجبا.

بمعنى:

$$\mathrm{H} = \begin{vmatrix} \dfrac{\partial^2 \mathrm{L}}{\partial \widehat{\beta_0}^2} & \dfrac{\partial^2 \mathrm{L}}{\partial \widehat{\beta_1}\, \partial \widehat{\beta_0}} \\ \dfrac{\partial^2 \mathrm{L}}{\partial \widehat{\beta_0}\, \partial \widehat{\beta_1}} & \dfrac{\partial^2 \mathrm{L}}{\partial \widehat{\beta_1}^2} \end{vmatrix} = \begin{vmatrix} 2n & 2\sum X_i \\ 2\sum X_i & 2\sum X_i^2 \end{vmatrix} = 4n \sum X_i^2 - 4(\sum \mathrm{X_i})^2 > 0$$

2− <u>خطوات توفيق نموذج الانحدار الخطي البسيط:</u>

للحكم على صلاحية نموذج الانحدار الذي تم توفيقه للعلاقة بين المتغير التابع والمتغير المستقل، لابد وأن يتوفر في هذا النموذج مجموعة من الشروط، يمكن تقسيمها إلى:

أولا: الشروط النظرية (الاقتصادية)

1) **اتفاق (أو منطقية) إشارات وقيم معاملات الانحدار:**

يجب أن تكون إشارات وقيم المعالم منطقية مع الأساس النظري الذي يحكم الظاهرة محل الدراسة.

مثلا: نموذج انحدار العلاقة بين الدخل والاستهلاك: $C = a + bY$.

الشرطان المفروضان على معلمي النموذج، وفقا للنظرية الاقتصادية، هما:

- أن يكون $0 < b < 1$ حيث يمثل b الميل الحدي للاستهلاك.
- أن يكون $a > 0$ ويمثل a الجزء الثابت من الاستهلاك (حالة الدخل معدوم: $Y = 0$).

فعدم توافر هذين الشرطين يجعل نموذج الانحدار الذي تم توفيقه غير سليم من الناحية النظرية.

2) قوة الارتباط والقدرة التفسيرية للنموذج

أ) قوة الارتباط بين متغيري النموذج:

يستخدم تحليل الارتباط في تقدير درجة الارتباط الخطي بين متغيرين، وتحديد اتجاه هذه العلاقة.

وتقاس قوة الارتباط بين المتغيرين اعتمادا على معامل الارتباط (r) المعرف كما يلي: $r = \mp\sqrt{R^2}$

- تنتمي قيمة معامل الارتباط (r)إلى المجال $[-1,+1]$؛
- فإذا كان: $r > 0$ فإن العلاقة بين المتغيرين طردية؛
- وإذا كان: $r < 0$ فإن العلاقة بين المتغيرين عكسية

ملاحظات:

1. إذا كان $r = 0$، لا توجد علاقة خطية بين المتغيرين.
2. إذا كان $r = \mp 1$، هناك ارتباط خطي تام بين المتغيرين.
3. هناك فرق جوهري بين معامل الارتباط (r) ومعامل التحديد (R^2) حيث:

يقيس (r)العلاقة بين المتغيرين X_i و Y_i، دون تحديد السببية.

ب) **القدرة التفسيرية للنموذج:** هي مدى قدرة المتغير المستقل في النموذج على تفسير التغيرات التي تحدث في المتغير التابع. بمعنى آخر هي نسبة التغيرات التي تحدث في المتغير التابع، وتعزى إلى المتغير المستقل.

ويعرف أيضا أنه: " نسبة تأثير المتغير المستقل في المتغير التابع.

تعريف: معامل التحديد R^2 هو نسبة مجموع المربعات المقدرة على مجموع المربعات الكلية.

ويقيس هذا المؤشر جودة التوفيق والارتباط بين المتغير المستقل X_i والمتغير التابع Y_i .

وهو أيضا، يعبر عن القدرة التفسيرية للنموذج، ومن الواضح أن: $0 \leq R^2 \leq 1$

ملاحظات:

- إذا كان $R^2 = 1$، فهناك جودة في التوفيق والارتباط بين المتغيرين X_i و Y_i، وأن القدرة التفسيرية للنموذج عالية؛
- إذا كان $R^2 = 0$، فإما أنه لا توجد قدرة تفسيرية في للنموذج، أو توجد قدرة تفسيرية، لكن النموذج غير خطي.

ثانيا: الشروط الرياضية: وتتضمن:

1) اختبار الفرضيات

يحتاج الاقتصادي في كثير كم الأحيان إلى اختبار الفرضيات لاتخاذ القرارات المناسبة التي تتعلق بالظاهرة المدروسة. ويعتمد اختبار الفرضيات على التوزيعات الاحتمالية: توزيعات فيشر، ستودنت، الطبيعي...الخ.

- **اختبار المعنوية الكلية للنموذج:**

وهي اختبار النموذج بين المتغيرين التابع والمستقل في نموذج الانحدار باستخدام اختبار فيشر(**F–test**). ويقصد باختبار المعنوية الكلية، الإجابة عن السؤال التالي: " هل النموذج الخطي مقبول لتمثيل العلاقة بين المتغيرين؟ "

ولاختبار مدى قبول هذا النموذج إحصائيا نعتمد على إحصاءة فيشر المعرفة على النحو التالي:

$$F_C = \frac{MES}{MRS} = \left(\frac{R^2}{1-R^2}\right)(n-2)$$

وهو ما يوضحه جدول تحليل التباين التالي:

Analysis of Variance table "ANOVA"

Model	Sum of Squares	df	Mean Square	F	Sig
Regression	ESS	1	MES = ESS/1	MES/MRS=$\left(\frac{R^2}{1-R^2}\right)(n-2)$	؟
Residual	RSS	n−2	MRS = RSS/(n−2)		
Total	SST	n−1			

الفرضية الصفرية: $\mathrm{H}_0 : \hat{\beta}_0 = \widehat{\beta_1} = 0$

الفرضية البديلة: $\mathrm{H}_1 : \widehat{\beta_1} \neq 0$

ومن ثم: $F_c \backsim F(1, n-2, \alpha)$

<u>اتخاذ القرار</u>:

بفرض p هي القيمة الاحتمالية لاختبار فيشر، و α هو مستوى الدلالة المطلوب،

- **إذا اعتمادنا الاحتمال p: نرفض الفرضية الصفرية H_0 إذا كان $p \leq \alpha$، ونقبل الفرضية البديلة H_1**
- **إذا اعتمادنا الإحصاءة F : نرفض الفرضية الصفرية H_0 إذا كان $F_c \leq F_\alpha$، ونقبل الفرضية البديلة H_1**

- **<u>المعنوية الجزئية لمعالم النموذج</u>:**

وهي اختبار معنوية معامل الانحدار للمتغير المستقل وثابت الانحدار، من خلال اختبار ستودنت **(t−test)**.

الفرضية الصفرية: $\mathrm{H}_0 : \widehat{\beta_0} = \widehat{\beta_1} = 0$

الفرضية البديلة: $\mathrm{H}_1 : \widehat{\beta_0} \neq \widehat{\beta_1}$

<u>اتخاذ القرار</u>:

بفرض p هي القيمة الاحتمالية لاختبار ستودنت، و α هو مستوى الدلالة المطلوب،

- **إذا اعتمادنا الاحتمال p: نرفض الفرضية الصفرية H_0 إذا كان $p \leq \alpha$، ونقبل الفرضية البديلة H_1**
- **إذا اعتمادنا الإحصاءة t : نرفض الفرضية الصفرية H_0 إذا كان $|t_c| \geq t_{\alpha/2}$، ونقبل الفرضية البديلة H_1**

2) شروط طريقة المربعات الصغرى العادية في تقدير النموذج:

تعتبر طريقة المربعات الصغرى العادية (OLS) أشهر الطرق المستخدمة، وتتمثل أهم شروط هذه الطريقة في التالي:

أ) اعتدالية التوزيع الاحتمالي للبواقي Normality test:

حتى يمكن استخدام كل من **(F−test)** و **(t−test)**، عند اختبار المعنوية الكلية أو الجزئية لنموذج الانحدار، يلزم توفر شرط اعتدالية التوزيع الاحتمالي للبواقي. وللإشارة فإن التقيد بهذا الشرط يكون في حالة العينات الصغيرة، أما في حالة العينات الكبيرة $(n \geq 30)$ فيمكن التخلي عنه وفقا لنظرية النهاية المركزية حيث تؤول التوزيعات الاحتمالية إلى التوزيع الطبيعي.

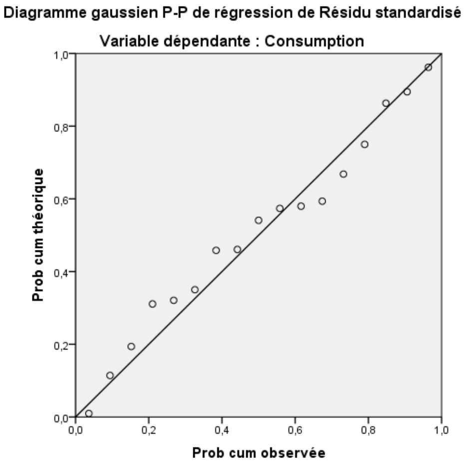
- **الفروض الإحصائية:**

الفرضية الصفرية(H_0) : تتوزع البواقي توزيعا طبيعيا

الفرضية البديلة (H_1) : لا تتوزع البواقي توزيعا طبيعيا

وتتم دراسة اعتدالية التوزيع الاحتمالي للبواقي بطريقتين (حسب برنامج SPSS):

- **الطريقة البيانية:** من خلال الشكل البياني للعلاقة بين الاحتمال التجميعي المشاهد والاحتمال التجميعي المتوقع للبواقي المعيارية بحيث:
  - إذا كانت النقط تقع بشكل متقارب جدا من الخط الواصل بين الركن الأيمن العلوي والركن الأيسر السفلي، أو تتوزع بشكل عشوائي على جانبي هذا الخط يقال أن الأخطاء تتوزع توزيعا طبيعيا.
  - أما إذا تم رصد نمط معين لتوزيع هذه النقط، فيقال أن الأخطاء لا تتوزع توزيعا طبيعيا."



- **الطريقة الحسابية:** يوفر برنامج SPSS اختبارين لفحص النموذج من خلال تحليل البواقي وهما:

- اختبار " كولموغوروف– سميرنوف " Kolmogorov– Smirnov
- اختبار " شابيرو – وايلك " Shapiro– Wilk

حيث يمكن إنجاز الاختبارين في آن واحد أو الاكتفاء بأحدهما.

ب) عدم الارتباط الذاتي بين البواقي:

تعود أهمية دراسة الارتباط الذاتي للبواقي في تحليل الانحدار، إلى أن وجود هذا الارتباط من شأنه أن يجعل قيمة تباين الخطأ مقدرة بأقل من قيمته الحقيقية. وبالتالي فإن قيم إحصاءات الاختبارات التي تعتمد على هذا التباين، مثل (t) و (F) و (R^2) تكون أكبر من قيمها الحقيقية، مما يجعل القرار الخاص بجودة توفيق النموذج قرار مشكوك في صحته.

نفترض أن الفرضية الثالثة للنموذج غير محققة، بمعنى: $\exists i \neq j: Cov(\varepsilon_i,\ \varepsilon_j) \neq 0$

أي هناك ارتباط ذاتي بين الأخطاء، ونطرح الأسئلة الآتية:

1) ما هي الأسباب التي تؤدي إلى وجود ارتباط ذاتي بين الأخطاء؟
2) كيف يمكن الكشف عن وجود ارتباط ذاتي بين الأخطاء؟
3) كيفية النمذجة في ظل اختلال فرضية عدم الارتباط الذاتي بين البواقي؟

الإجابة:

1) **الأسباب المؤدية إلى وجود ارتباط ذاتي بين الأخطاء، منها:**

- غياب متغير مستقل في النموذج قد يؤدي إلى اختلال الفرضية المذكورة؛
- احتمال أن تكون النمذجة غير صحيحة، كوجود علاقة غير خطية بين المتغيرين التابع والمستقل.

ومن نتائج اختلال هذه الفرضية هو: الحصول على مقدرات متحيزة وغير متسقة.

2) **الكشف عن وجود ارتباط ذاتي بين الأخطاء:** يمكن الاعتماد على الأسلوب الإحصائي التالي:

اختبار درين–واتسون Durbin–Watson

اقترح $(D.W)$ اختبارا يعتمد على سيرورة الانحدار الذاتي من الرتبة الأولى حيث: $\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \end{cases}$

ρ : هو معامل الارتباط الذاتي بين الأخطاء، حيث: $|\rho| \leq 1$

η_t : هو حد الخطأ العشوائي في نموذج الانحدار الذاتي $AR(1)$.

أ) **الفروض الإحصائية:**

الفرضية الصفرية: للاستقلال (عدم الارتباط) الذاتي بين الأخطاء، بمعنى: $H_0 : \rho = 0$

الفرضية البديلة: الارتباط الذاتي بين الأخطاء $H_1 : \rho \neq 0$

ب) **العلاقة بين $\widehat{DW}_c$ و $\hat{\rho}$:** $\boxed{\hat{\rho} = 1 - \frac{\widehat{DW}}{2} \;:\; \widehat{DW} \in [0,\ 4]}$

تحتاج الإحصاءة $\widehat{DW}_c$ إلى قيمتين حرجتين dl و du، اللتان تعتمدان على حجم العينة T، وعدد المتغيرات المستقلة في النموذج k، ونسبة الدلالة α. ويتم الحصول على القيمتين الحرجتين من خلال " جدول دربن-واتسون".

<u>اتخاذ القرار</u>:

0 dl	du	2	$4-du$	$4-dl$ 4
$\hat{\rho} > 0$ ارتباط ذاتي موجب نرفض H_0	منطقة شك (فشل الاختبار)	$\hat{\rho} = 0$ استقلالية الأخطاء نقبل H_0	منطقة شك (فشل الاختبار)	$\hat{\rho} < 0$ ارتباط ذاتي سالب نرفض H_0

3) **النمذجة في ظل اختلال فرضية الاستقلال الذاتي بين الأخطاء**

نستخدم شبه الفروقات للمتغيرين X و Y والمتمثلة في: $X_i - \rho X_{i-1}$ و $Y_i - \rho Y_{i-1}$، وبالتالي نفقد المشاهدة الأولى لكل متغير. ولتجنب ضياعها نضع: $Y_1^* = Y_1\sqrt{1-\rho^2}$ ، $X_1^* = X_1\sqrt{1-\rho^2}$

نحصل على النموذج المصحح التالي:

$$\mathbf{Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(1-\rho)(X_t - \rho X_{1-t}) + (\varepsilon_t - \rho\varepsilon_{t-1})}$$

إذن يأخذ النموذج المصحح الشكل التالي: $Y_t^* = \beta_0{}^* + \beta_1{}^* X_t^* + u_t$

ويكون النموذج المقدر بطريقة (OLS) هو: $\hat{Y}_t^* = E(Y_t^*) = \widehat{\beta_0}^* + \widehat{\beta_1}^* X_t^*$

حيث: $\widehat{\beta_0}^* = \widehat{\boldsymbol{\beta_0}}(\mathbf{1} - \hat{\boldsymbol{\rho}})$ و $\widehat{\beta_1}^* = \widehat{\boldsymbol{\beta_1}}(\mathbf{1} - \hat{\boldsymbol{\rho}})$

<u>سؤال</u>: كيف يتم تقدير معامل الارتباط الذاتي (ρ)؟

<u>الجواب</u>: يتم تقدير معامل الارتباط الذاتي (ρ) بعدة طرق، منها طريقة $D.W$ المباشرة: $\hat{\rho} = 1 - \widehat{DW}/2$

أما $\widehat{\beta_0}$ ، $\widehat{\beta_1}$ ، $\widehat{DW}$ فهي تحدد من النموذج الأصلي المقدر: $\hat{Y}_t = \widehat{\beta_0} + \widehat{\beta_1} X_t$

ت) **اختبار التجانس (ثبات التباين) Homoscedasticity:**

يترتب عن عدم ثبات التباين في نموذج الانحدار نفس الآثار في حالة وجود ارتباط ذاتي بين البواقي، حيث تكون الأخطاء المعيارية مقدرة بأقل من قيمتها الحقيقية. وبالتالي تصبح هذه المقدرات متحيزة **(Biased)**، مما يجعل نتائج الاستدلال الإحصائي مشكوك في صحتها.

نفترض أن الفرضية الثانية للنموذج غير محققة، بمعنى: $\exists i: Var(\varepsilon_i) \neq \sigma_\varepsilon^2$

أي أن تباين الأخطاء غير ثابت (يتغير بتغير الزمن)، ونطرح الأسئلة الآتية:

1) ما هي الأسباب التي تؤدي إلى عدم التجانس (عدم ثبات التباين)؟

2) كيف يمكن الكشف عن عدم التجانس؟

3) كيفية النمذجة في ظل اختلال فرضية عدم التجانس؟

الإجابة:

1) **الأسباب المؤدية إلى عدم تجانس البواقي، منها:**

- مشكل في النمذجة (احتواء النموذج على متغير مبطأ بالنسبة للمتغير التابع)؛ فقد تختل الفرضية المذكورة؛
- تحسن أساليب تجميع البيانات، وهذه يقلل من الأخطاء المرتكبة في القياس، ومن ثم سوف يقل تباين حد الخطأ.

ومن نتائج اختلال هذه الفرضية هو: الحصول على مقدرات متحيزة وغير متسقة.

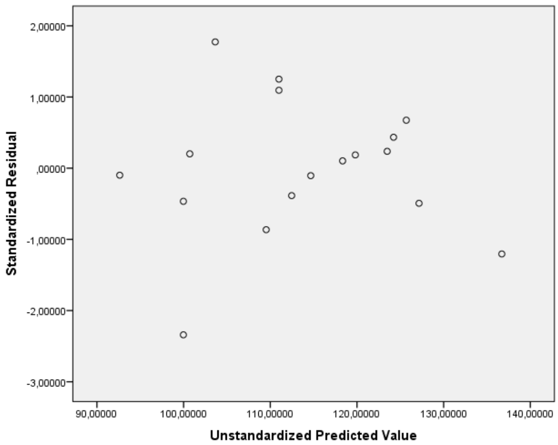
2) **الكشف عن مدى التجانس:**

- **الفروض الإحصائية:**

الفرضية الصفرية: تجانس البواقي (ثبات التباين)

الفرضية البديلة: عدم تجانس البواقي (عدم تساوي التباينات)

يمكن الاعتماد على الأسلوب البياني أولا ثم على الأسلوب الإحصائي التي يوفرها برنامج SPSSعلى النحو التالي:



- **طريقة الرسم البياني لاختبار ثبات التباين**

يلاحظ أن انتشار وتوزيع البواقي يأخذ شكلا عشوائيا على جانبي الخط الذي يمثل الصفر (الخط الفاصل بين البواقي السالبة والبواقي الموجبة)، حيث لا يمكننا رصد شكل معين لتباين هذه البواقي، وهو ما يعني أن هناك تجانس أوثبات في تباين الأخطاء.

- **الطريقة الحسابية (Goldfield– Quandt) لاختبار ثبات التباين: وتتم وفق الخطوات التالية:**

أ- ترتيب المشاهدات- **تنازليا**- وفقا للمتغير المستقل

ب- استبعاد 20% من المشاهدات في المنتصف، لكل من المتغيرين(X) و (Y) فنحصل على سلسلتين؛

ت- حساب مجموع مربعات $(\mathrm{SSE})_{(1)}$ و $(\mathrm{SSE})_{(2)}$ للسلسلتين (1) و(2)، من خلال جدول تحليل التباين لمعادلة الانحدار لكل سلسلة.

ث- تنفيذ خطوات إيجاد نموذجي انحدار (Y)على$(X'S)$ للسلسلتين (1) و(2). ومن مخرجات SPSS نهتم بجدول تحليل التباين للحصول على كل من $(\mathrm{SSE})_{(1)}$ و $(\mathrm{SSE})_{(2)}$.

ج- حساب النسبة التالية: $F_c = \frac{(\mathrm{SSE})_{(2)}}{(\mathrm{SSE})_{(1)}}$

ح- المقارنة بين F_c المحسوبة و $F(df_1, df_2, \alpha)$ المجدولة، لرفض أو قبول الفرضية H_0.

حيث: يمكن حساب درجتي الحرية كما يلي: $\boldsymbol{df_1 = \mathrm{df}_2 = \frac{n-2-m}{2}}$ ، m : عدد المشاهدات المستبعدة.

4) النمذجة في ظل اختلال فرضية التجانس

للتخلص من مشكل عدم التجانس، نقوم بقسمة طرفي النموذج على $\hat{\rho}_{\varepsilon}$ فنحصل على النموذج المصحح الذي ينبغي تقديره بطريقة (OLS)، والمتمثل فيما يلي: $\frac{Y_t}{\hat{\sigma}_{\varepsilon}} = \frac{\alpha}{\hat{\sigma}_{\varepsilon}} + \frac{\beta}{\hat{\sigma}_{\varepsilon}} X_t + \frac{1}{\hat{\sigma}_{\varepsilon}} \varepsilon_t$

والهدف من عملية القسمة على " $\hat{\sigma}_{\varepsilon}$ " هو إقصاء المتغير المستقل المتسبب في عدم التجانس.

وبصفة عامة: يمكن قسمة طرفي النموذج على مقادير أخرى، (محمد شيخي،2011، ص ص 117−118).

2- تحليل الانحدار الخطي العام (النموذج الخطي المتعدد)

1.2-الكتابة الرياضية للنموذج:

يستند النموذج الخطي المتعدد في التحليل الساكن على افتراض وجود علاقة خطية بين متغير داخلي Y وعدد k $(k \geq 2)$ من المتغيرات الخارجية المستقلة $X_1, X_2, X_3, ..., X_k$.

نعتبر النموذج التالي الذي يكتب بشكل مختصر كما يلي: $Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i : i = \overline{1,n}$

$\beta_0, \beta_1, \beta_2, ..., \beta_k$ هي معالم (وسائط) النموذج يطلب تقديرها.

ε_i هي الأخطاء العشوائية (الظاهرة غير المشاهدة).

2.2- الشكل المصفوفي لنموذج الانحدار الخطي المتعدد:

$$(Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i : i = \overline{1,n}) \Leftrightarrow \begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_k X_{1k} + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + \varepsilon_2 \\ .... \\ .... \\ Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \varepsilon_n \end{cases}$$

ويمكن تحويل هذه الجملة إلى الشكل المصفوفي التالي:

$$\underset{[n\times1]}{Y} = \underset{[n\times(k+1)]}{X}\ \underset{[(k+1)\times1]}{\beta} + \underset{[n\times1]}{\varepsilon} \Leftrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} \cdots\cdots & X_{1k} \\ 1 & X_{21} & X_{22} \cdots\cdots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \cdots\cdots & X_{nk} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} .....[1]$$

3.2-الفرضيات الأساسية لطريقة المربعات الصغرى (OLS):

لطريقة المربعات الصغرى العادية (Ordenary least squares method) عدة فرضيات أساسية، والإخلال بإحداها قد يؤدي إلى نتائج مظللة، خاصة عند دراسة صلاحية النموذج. وسنتطرق إلى هذه الفرضيات، كما سنرى كيفية معالجة بعض المشاكل الناتجة عن الإخلال بإحداهن. وفيما يلي عرض لهذه الفرضيات[1] :

الفرضية الأولى: خطية العلاقة بين المتغير التابع والمتغيرات المستقلة.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_k X_{ik} + \varepsilon_i : i = \overline{1,n}$$

الفرضية الثانية: التوقع الرياضي للمتغيرات المفسرة المهملة معدوم، بمعنى: $E(\varepsilon^2) = 0$

نستنتج أن التوقع الرياضي لـ : Y هو: $\hat{Y} = E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k$

الفرضية الثالثة : تباين الأخطاء متجانس، بمعنى أن: $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 : i = \overline{1,n}$

[1]- جلاطو جيلا لي، **الإحصاء التطبيقي**، دار الخلدونية للنشر والتوزيع، القبة القديمة، الجزائر، الطبعة الأول ى2007، ص16-17.

الفرضية الرابعة : استقلالية الأخطاء، بمعنى أن: $Cov(\varepsilon_i,\varepsilon_j)=0 : i\neq j$

الفرضية الخامسة : عدم الارتباط بين المتغيرات المستقلة والبواقي: $Cov(X_i,\varepsilon_i)=E(X_i\varepsilon_i)=X_iE(\varepsilon_i)=0$

الفرضية السادسة :المتغير العشوائي يتبع التوزيع الطبيعي $\varepsilon \rightarrow N(0,\sigma_\varepsilon^2 I_n)$

الفرضية السابعة : أن يكون حجم العينة أكبر بكثير من عدد المعالم المقدرة في النموذج: $n\rangle\rangle(k+1)$.

4.2-تقدير **شعاع المعالم** β بالطريقة (OLS) :

أ) تقدير **مركبات شعاع المعالم** β:

ولتقدير معالم النموذج $Y=X\beta+\varepsilon......[1]$ بمفهوم المربعات الصغرى العادية، نقوم بتدنئة مجموع مربعات البواقي الذي يكتب من الشكل: $\hat{\varepsilon}'\hat{\varepsilon}=\sum_{i=1}^{n}\hat{\varepsilon}_i^2$ حيث $\hat{\varepsilon}'$ هو منقول الشعاع $\hat{\varepsilon}$.

وبالتالي:
$$Min\sum_{i=1}^{n}\varepsilon_i^2=Min\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$$

وعندئذ: يمكن الحصول على الشعاع المقدر لمعلمات النموذج بالصيغة التالية:

$$\boxed{\hat{\beta}=(X'X)^{-1}X'Y..................[1]}$$

5.2-تحليل القدرة التفسيرية للنموذج:

جدول تحليل التباين $(ANOVA)$

المتوسط	درجة الحرية	مجموع المربعات	مصدر التغير
SSR/k	k	$SSR=\sum(\hat{Y}-\bar{Y})^2$	$X_1,X_2,...,X_k$
$SSE/(n-k-1)$	$n-k-1$	$SSE=\sum\hat{\varepsilon}_i^2$	البواقي
	$n-1$	$SST=SSR+SSE$	المجموع

ب) معامل التحديد المتعدد(Multiple Coefficient of Determination):

هو نسبة تأثير كل المتغيرات المفسرة (المستقلة) على المتغير المفسر(التابع). ويعرف على انه[2] "نسبة التباين المفسر إلى التباين الكلي، وهو مؤشر على النسبة المئوية من التغير الكلي في المتغير التابع التي يمكن تفسيرها بدلالة المتغيرات المستقلة المدرجة في النموذج. ونرمز لهذا المعامل بالرمز R^2".

وهو يقيس جودة التوفيق والارتباط بين محددات الظاهرة المدروسة والظاهرة نفسها.

[2] شعوبي محمود فوزي، مرجع سابق، ص20.

- **خواص معامل التحديد** R^2:

1) من معادلة تحليل التباين يكون لدينا: $0 \le R^2 \le 1$.

2) يمكن التمييز بين حالتين كما يلي:

أ-إذا اقتربت قيمة R^2 من الصفر فهو دلالة على ضعف العلاقة بين المتغير التابع والمتغيرات المستقلة، وإذا كان $R^2 = 0$ فلا وجود للعلاقة بين المتغير التابع والمتغيرات المستقلة الداخلة في النموذج.

ب-وإذا كان $R^2 = 1$ فان ذلك يشير إلى أن (100%) من التغير في المتغير التابع يعود إلى التغير في المتغيرات المستقلة المدرجة في النموذج. لذلك كلما اقتربت قيمة R^2 من الواحد الصحيح، كلما زادت درجة الثقة في النموذج المقدر.

3) يعاب على المعامل R^2 انه لا يأخذ في الاعتبار عدد المشاهدات ولا عدد المتغيرات.

ج) معامل التحديد المعدل $(\overline{R}^2)$: إن إضافة متغير مستقل أو أكثر إلى النموذج المقدر يؤدي إلى زيادة في قيمة R^2 نتيجة لزيادة قيمة البسط في معادلة R^2 وبقاء المقام على حاله دون تغيير. ، وبالتالي قيمة R^2 لا تعبر عن القيمة الحقيقية له، ولتفادي هذا الإشكال أُقتُرِح مؤشرٌ آخر يأخذ بعين الاعتبار درجة الحرية في النموذج ، ويسمى معامل التحديد المعدل " $\overline{R}^2$ " حيث :

$$\boxed{\overline{R}^2 = 1-\left(\frac{n-1}{n-k-1}\right)\left(1-R^2\right)}$$

1. في العينات الصغيرة، $k \rangle 1$ فان: $\overline{R}^2 \le R^2$ أما في العينات الكبيرة فإن: $\lim\limits_{n\to+\infty}\left(\frac{n-1}{n-k-1}\right) = 1 \Rightarrow \overline{R}^2 \approx R^2$
2. يمكن أن تنخفض قيمة R^2 وذلك بإدخال متغيرة خارجية جديدة في النموذج.
3. يمكن أن يأخذ $\overline{R}^2$ قيما سالبة.

د) العلاقة بين R^2 واختبار F : يمكن تحديد طبيعة العلاقة بين R^2 و F من خلال الجدول التالي:

جدول تحليل التباين $(ANOVA)$ ، والعلاقة بين R^2 واختبار F.

Source	SS	$d.f$	MS	F_C
Regression	$SSR = \hat{Y}'\hat{Y}$	k	$R^2Y'Y/(k)$	$F_C = \left(\frac{R^2}{1-R^2}\right)\left(\frac{n-k-1}{k}\right)$
Residuel	$SSE = \hat{\varepsilon}'\hat{\varepsilon}$	$n-k-1$	$(1-R^2)Y'Y/(n-k-1)$	
Total	$SST = Y'Y$	$n-1$		

6.2-الاختبارات الإحصائية لدراسة مدى صلاحية النموذج المقدر:

تناولنا فيما سبق استخدام أسلوب المربعات الصغرى لتقدير معلمات النموذج. إلا أن ذلك يتطلب إجراء بعض الاختبارات الإحصائية حول قيم المعلمات التي تم تقديرها للتأكد من معنويتها الإحصائية.

أ) <u>اختبار المعنوية الكلية للنموذج</u> $(F-Fisher)$:

- **الفروض الإحصائية:**

الفرضية الصفرية هي: $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$

والفرضية البديلة هي: $H_1 : \exists \beta_i \neq 0; i = \overline{1,k}$

*__القرار__: يمكن إجراء هذا الاختبار باستخدام جدول تحليل التباين $(ANOVA)$.

- فإذا كان $F_C \rangle F_\alpha$ نرفض الفرضية H_O ، بمعنى أن للنموذج ككل معنوية إحصائية بنسبة دلالة α .
- أما إذا كان $F_C \leq F_\alpha$ نقبل الفرضية H_O ، بمعنى أن النموذج غير مقبول إحصائيا عند مستوى الدلالة α .
ويمكن اختبار المعنوية الكلية للنموذج اعتمادا على مفهوم P-value حسب القاعدة التالية[3]: " نرفض الفرضية الصفرية إذا كان $p \leq \alpha$ ".

ب) اختبار المعنوية الإحصائية للمعالم:

وتصاغ الفرضية الصفرية والبديلة لها كما يلي: $\begin{cases} H_0 : \forall i = 1,n : \beta_i = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$ [4].

* __القرار__: تتم مقارنة قيمة t_C المحسوبة بقيمة $t_{\alpha/2}$ المجدولة.

- فإذا كان: $|t_c| \geq t_{\alpha/2}$ ، نرفض الفرضية H_O بمعنى أن المعلم β_i يختلف معنويا عن الصفر بنسبة دلالة α .
- وإذا كان: $|t_c| \langle t_{\alpha/2}$ ، نقبل الفرضية H_O . أي أن المعلم β_i ليس له معنوية إحصائية عند مستوى الدلالة α .
كما يمكن اختبار المعنوية الإحصائية للمعلم β_i اعتمادا على مفهوم P-value حسب القاعدة التالية: " نرفض الفرضية H_O إذا كان $p \leq \alpha$ ".

7.2- شروط طريقة المربعات الصغرى العادية في تقدير النموذج:

يتم تطبيق نفس خطوات وشروط نموذج الانحدار الخطي البسيط، مع إضافة شرط رابع إلى الشروط الثلاثة السابقة (**1**- اعتدالية توزيع البواقي، **2**-عدم الارتباط الذاتي بين البواقي، **3**-عدم ثبات تباينات البواقي، التي تم تناولها سابقا) وهو: شرط عدم التعدد الخطي بين المتغيرات المستقلة.

3- التعدد الخطي بين المتغيرات المستقلة Multicollinearity

بافتراض أن النموذج هو: $$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i : i = \overline{1,n}$$

فإنه في أغلب الأحيان نواجه مشكلة التعدد الخطي بين المتغيرات المستقلة، بمعنى: $\exists i \neq j: Cov(X_i,\ X_j) \neq 0$، مما يؤدي إلى عدم استقرار تباين المقدرات ويكون هناك فائض في التقدير، وهو ما يجعل المصفوفة $X'X$ غير قابلة للقلب.

◊ **P-value** : هو احتمال البيانات التي توضح عزوفا عن الفرضية الصفرية، عندما تكون هذه الفرضية صحيحة.
[3] David R. Anderson et autres, p692.
[17]David R. Anderson et autres, p694.

أ) أسباب وجود تعدد خطي بين المتغيرات المستقلة

- تقل قيمة R، والسبب أن المتغيرات المستقلة تتشارك في نفس تباين المتغير التابع؛
- صعوبة تحديد الأهمية النسبية لكل متغير من المتغيرات المستقلة بالنسبة للمتغير التابع؛
- زيادة تباين المعاملات في الانحدار المتعدد، فكلما زاد تباين المعاملات، قل ثبات معادلة التنبؤ.

ب) الكشف عن وجود التعدد الخطي بين المتغيرات المستقلة

لتشخيص مشكلة الارتباطات المتداخلة بين المتغيرات المستقلة بشكل دقيق، يمكن استخدام برنامج SPSS بأكثر من طريقة (أو مقياس) منها:

- **مصفوفة الارتباط**$(\boldsymbol{Correlation\ Matrix})$: فحص مصفوفة الارتباط بين المتغيرات المستقلة، بحيث يمكن الحكم بعدم وجود تعدد خطي في الحالة التي تتراوح قيم معاملات الارتباط بين $(\mathbf{-0.7})$ و $(\mathbf{+0.7})$.
- **مؤشر الحالة (Condition Index)**: ويستخرج مؤشر الحالة من قيم الجذور الكامنة (القيم الذاتية) المدرجة في العمود الثاني من الجدول " Collinearity Diagnostic " وذلك وفق العلاقة الرياضية: $\boldsymbol{C} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$

حيث: $\boldsymbol{\lambda_{max}}$ الجذر الكامن الأكبر، $\boldsymbol{\lambda_{min}}$ الجذر الكامن الأصغر

- يرى جونسون أنه إذا كان: $\boldsymbol{C > 20}$، فهذا دليل على وجود تعدد خطي مرتفع؛
- ويرى بيلسلي وزملاؤه أنه إذا كان: $\boldsymbol{C > 30}$ فهو دلالة على وجود تعدد خطي مرتفع جدا بين المتغيرات المستقلة.

- **معامل التحميل**$(\boldsymbol{Tolerance})$: يعد هذا المعامل من المؤشرات التي تكشف عن مدى وجود تعدد خطي بين المتغيرات المستقلة، ويمكن حسابه من خلال الصيغة الرياضية: $(\boldsymbol{Tolerance}) = \mathbf{1} - \boldsymbol{R^2}$

حيث: $(\boldsymbol{R^2})$ هو مربع معامل الارتباط بين كل متغيرين مستقلين.

فإذا كانت قيمة معامل التحمل أصغر من $(\mathbf{0.20})$ فإنها تشير إلى احتدام مشكلة التعدد الخطي بين المتغيرات المستقلة. ووفقا للمعادلة أعلاه فإنه كلما ارتفعت قيمة $(\boldsymbol{R^2})$، اقتربت قيمة معامل التحمل من الصفر.

- **معامل تضخم التباين** $(\boldsymbol{Variance\ Inflation\ Factor})$: يعد هذا المعامل من المؤشرات التي تكشف عن مدى وجود تعدد خطي بين المتغيرات المستقلة. ففي حالة وجود تعدد خطي تام، يأخذ $\boldsymbol{VIF}$ قيمة لانهائية. وفي حالة عدم وجود تعدد خطي فإن $\boldsymbol{VIF = 1}$.

أما إذا تجاوزت قيمة $\boldsymbol{VIF}$ الأربعة (4) دل ذلك على احتدام مشكلة التعدد الخطي، فكلما ارتفعت قيمته تسبب في عدم استقرار قيمة المعامل $\left(\textbf{بيتا}\right)$ من جراء ارتفاع الخطأ المعياري. ويبدي بعض المختصين تسامحا في هذا الأمر حيث يرون أن نقطة القطع (المحك) هي خمسة (5).

ويمكن حساب معامل تضخم التباين من خلال الصيغة الرياضية: $(\boldsymbol{VIF}) = \frac{\mathbf{1}}{\boldsymbol{Tolerance}}$

ت) النمذجة في ظل وجود تعدد خطي بين المتغيرات المستقلة

من الإجراءات التي يمكن اتخاذها في هذا الإطار ما يلي:

- دمج المتغيرات المستقلة المرتبطة مع بعضها ارتباطا قويا؛
- استخدام التحليل العاملي في حال وجود عدد كبير من المتغيرات للتخلص من بعضها؛
- استخدام طريقة الاختبار$(\boldsymbol{Stepwise})$ التي يتيحها برنامج SPSS؛
- استخدام القيم المعيارية للمتغيرات المستقلة بدلا من قيمها العادية.