

الانحدار الخطي البسيط

Simple Linear Regression

يُعتبر الانحدار الخطي البسيط أحد الأدوات الأساسية في التحليل الإحصائي، يهدف إلى دراسة العلاقة بين متغيرين، متغير مفسّر أو كما يعرف بالمتغير التابع ومتغير مفسّر أو مستقل. يتيح هذا النموذج فهم كيفية تأثير المتغير المستقل على المتغير التابع، مع تقديم تقديرات للعلاقة بينهما.

1. الصياغة العامة للنموذج

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

حيث:

Y_i : المتغير التابع *Dependent Variable*.

X_i : المتغير المستقل *Independent Variable*.

β_0 : القاطع (الثابت) *Constant*.

β_1 : معامل ميل الانحدار (الميل) *X Coefficient*.

ε_i : الأخطاء العشوائية (البواقي) *Residuals*.

$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i), \quad \forall i = \overline{1, n}$$

ويرجع وجود الأخطاء خلال تقدير نماذج الانحدار إلى:

- إهمال بعض المتغيرات المستقلة المهمة.
- الصياغة الرياضية الخاطئة للنموذج.
- أخطاء القياس أثناء تجميع البيانات.

2. فرضيات النموذج

في تحليل الانحدار الخطي البسيط، توجد مجموعة من الفرضيات الأساسية للنموذج التي يجب تحققها لضمان تقدير معاملات صحيحة وغير متحيزة وقابلة للاختبار الإحصائي.

- الخطية: وجود علاقة خطية بين المتغير التابع والمتغير المستقل.
- الوسط معدوم: الأمل الرياضي للأخطاء معدوم $E(\varepsilon_i) = 0$.
- التوزيع الطبيعي: الأخطاء تتبع توزيعاً طبيعياً $\varepsilon_i \sim N(0, \sigma^2)$.
- التجانس: ثبات تباين الأخطاء (*Homoscedasticity*)، أي تشتتها حول المتوسط ثابت.

$$V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2, \quad \forall i = \overline{1, n}$$

- **الاستقلالية:** عدم وجود ارتباط ذاتي بين الأخطاء، أي أن التباينات المشتركة لأخطاء المشاهدات المختلفة معدومة، وكذلك استقلالية الأخطاء عن المتغير المستقل.

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0, \quad \forall i \neq j, \quad i, j = \overline{1, n}$$

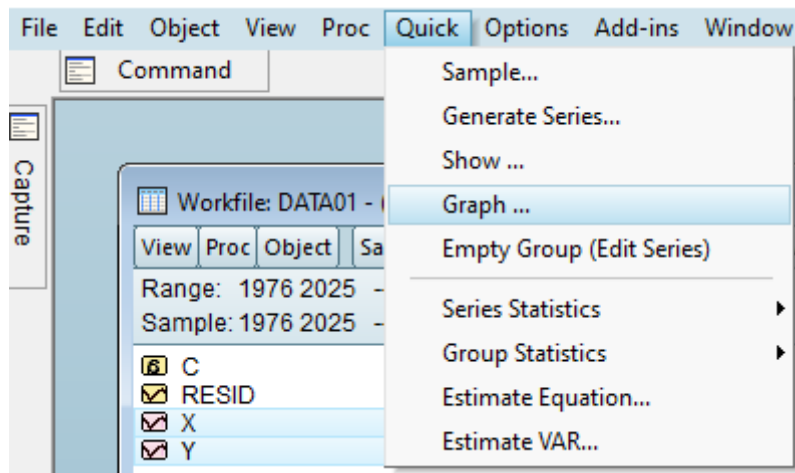
$$Cov(X_i, \varepsilon_i) = 0, \quad \forall i = \overline{1, n}$$

3. تقدير معلمات النموذج

يعتبر برنامج **EViews** أداة قوية وبسيطة تُستخدم لتطبيق الانحدار الخطي البسيط وتحليل نتائجه بفعالية. من خلال واجهته المرنة وأدواته المتقدمة. من أجل تطبيق الانحدار الخطي البسيط في **EViews** بالاعتماد على طريقة المربعات الصغرى **Least Squares**، نتبع الخطوات التالية:

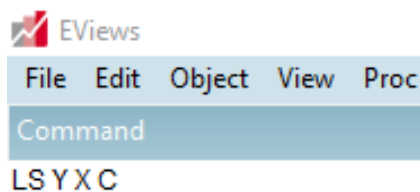
- إدخال البيانات يدوياً أو نسخها أو استيرادها.
- معالجة البيانات وتنظيمها.
- التحقق من وجود علاقة خطية بين المتغيرين وذلك بتمثيل سحابة الانتشار للمتغير التابع بدلالة المتغير المستقل من خلال اختيار المتغير المستقل ثم التابع ثم اتباع الخطوات التالية:

Quick → Graph



- كتابة الصيغة التالية في نافذة الأوامر:

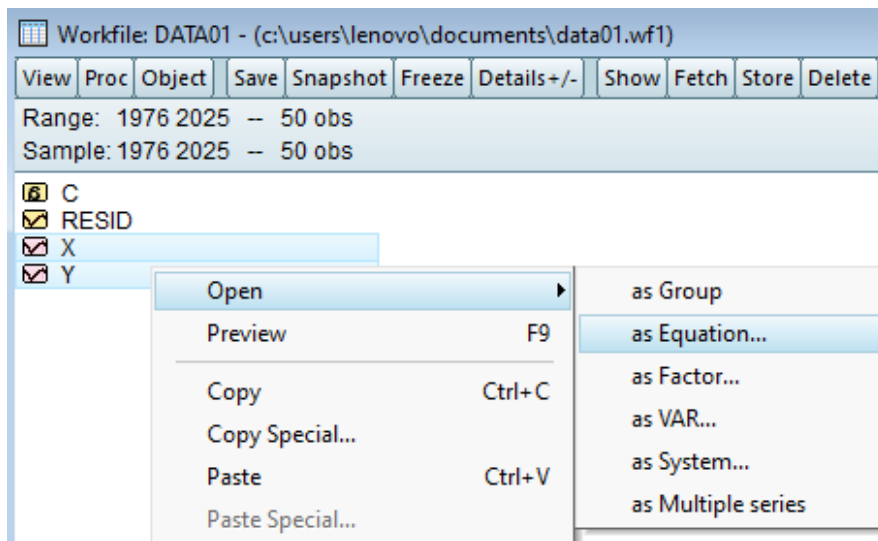
LS Y X C



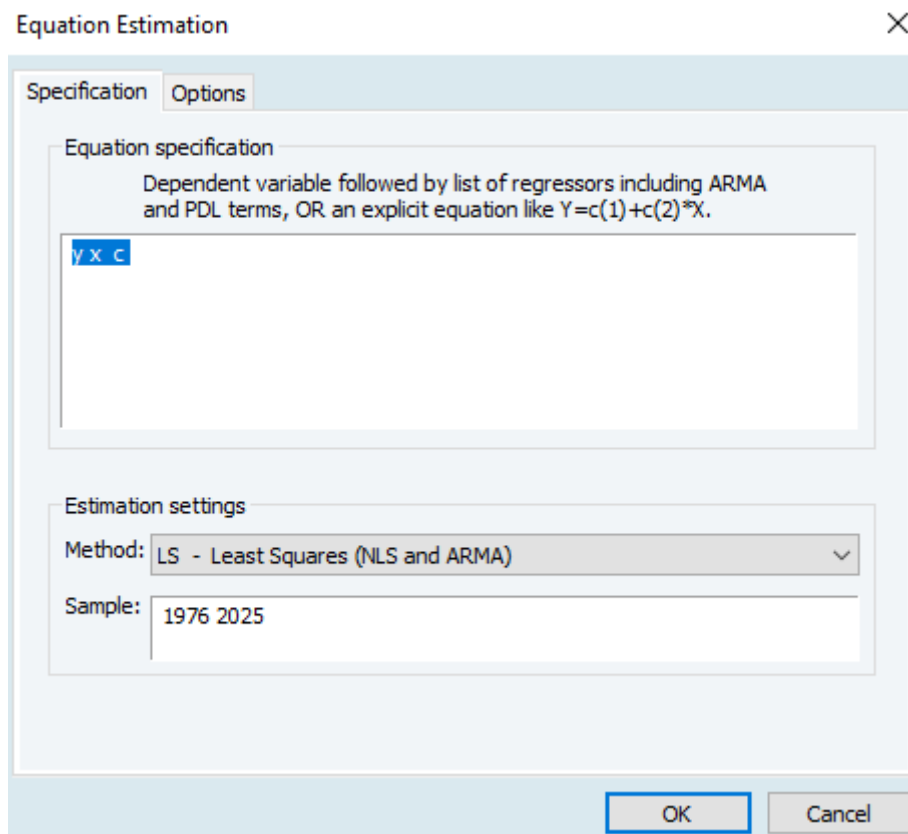
ثم الضغط على **Enter** من لوحة المفاتيح.

أو تظليل المتغير التابع ثم المتغير المستقل على التوالي ثم النقر على يمين الفأرة واختيار:

Open → as Equation



فتظهر النافذة التالية:



نتأكد من ترتيب المتغيرات (المتغير التابع ثم المتغير المستقل ثم الثابت)، مع اختيار طريقة التقدير **Method** هنا نعتمد على طريقة المربعات الصغرى **LS – Least Squares** وأخير نضغط على **OK**.

يظهر جدول نتائج الانحدار كما يلي:

Equation: UNTITLED Workfile: DATA01::Data01\				
View	Proc	Object	Print	Name
Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: Y				
Method: Least Squares				
Date: 11/03/25 Time: 15:01				
Sample: 1976 2025				
Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	-32.16585	5.731608	-5.612011	0.0000
C	367.3765	43.17035	8.509928	0.0000
R-squared	0.396186	Mean dependent var	125.5826	
Adjusted R-squared	0.383606	S.D. dependent var	24.42584	
S.E. of regression	19.17691	Akaike info criterion	8.784469	
Sum squared resid	17652.18	Schwarz criterion	8.860949	
Log likelihood	-217.6117	Hannan-Quinn criter.	8.813593	
F-statistic	31.49467	Durbin-Watson stat	1.528258	
Prob(F-statistic)	0.000001			

يتضح من خلال الجدول أعلاه معلومات عامة حول نموذج الانحدار الخطي البسيط المقدر، تتمثل في:

- المتغير التابع **Dependent Variable**.
- طريقة التقدير (طريقة المربعات الصغرى) **Method Least Squares**.
- تاريخ وساعة تقدير النموذج **Date and Time**.
- المجال الزمني للبيانات **Sample**.
- عدد المشاهدات (n) أو العينة **Included Observations**.

بعد المعلومات العامة يأتي جدول المعاملات، ويتضمن كل من:

- **Coefficient**: معاملات النموذج β_i ، حيث تحسب كما يلي:

$$\hat{\beta}_1 = \frac{Cov(X, Y)}{Var(X)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

ويكتب نموذج انحدار Y على X من الشكل:

$$\hat{Y} = 367.38 - 32.17X_i$$

- **Std. Error**: الخطأ المعياري للمعاملات أي الفرق بين المعاملات الحقيقية والمقدرة، وتحسب كما يلي:

$$S.E(\hat{\beta}_1) = \hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

$$S.E(\hat{\beta}_0) = \hat{\sigma}_{\hat{\beta}_0} = \sqrt{\hat{\sigma}_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)}$$

الانحراف المعياري للأخطاء العشوائية يحسب كالتالي:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{n - k - 1} = \frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n - k - 1}$$

حيث:

k : عدد المتغيرات المستقلة.

n : عدد المشاهدات.

• **t. Statistic**: القيمة الإحصائية لستيوذنت t المحسوبة، وتحسب بالعلاقة التالية:

$$t_{cal} = t. Statistic = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} = \frac{Coefficient}{Std. Error}$$

يستخدم اختبار ستيوذنت لتحديد المعنوية الفردية للمعاملات، ويتم بناء على الفرضية الصفرية H_0 ، حيث:

$$\begin{cases} H_0: \beta_i = 0 & (\text{غير معنوي}) \\ H_1: \beta_i \neq 0 & (\text{معنوي}) \end{cases}$$

ترفض الفرضية الصفرية إذا كانت t المحسوبة أكبر من t الجدولة (الدرجة)،

أي $(t_{cal} > t_{tab})$ ، عند مستوى المعنوية α .

• **Prob**: القيمة الاحتمالية الموافقة لـ t المحسوبة.

كما يمكن الاعتماد على القيم الاحتمالية في اختبارات معنوية المعاملات، حيث ترفض الفرضية الصفرية

إذا كانت القيمة الاحتمالية الموافقة لاختبار t أصغر من مستوى الدلالة α ، أي $(Prob_{t-Statistic} < \alpha)$.

ووفقاً لنتائج الجدول أعلاه فإن القيمة الاحتمالية لاختبار معنوية β_1 أصغر من مستوى المعنوية

0.05 ($0.05 < \alpha = 0.05$)، وعليه فإن β_1 معنوي، أي أن X يؤثر على Y . وكلما

تغير X بوحدة واحدة يؤدي بدوره إلى التغير في Y بـ 32.17 وحدة.

القيمة الاحتمالية للثابت C كانت أصغر من مستوى المعنوية 0.05 ، وعليه فإن الثابت β_0 معنوي،

أي أن قيمة المتغير التابع Y تساوي قيمة الثابت 367.38 عندما يكون X معدوم.

كما نجد في أسفل الجدول ما يلي:

• **$R - Squared$** : معامل التحديد R^2 ، يعبر عن نسبة التباين المفسر في البيانات. ويحسب بالعلاقة التالية:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

حيث:

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 \text{ :التباين المفسر}$$

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \text{ :التباين غير المفسر}$$

$$SST = \sum (Y_i - \bar{Y})^2 \text{ :التباين الإجمالي}$$

يظهر من خلال الجدول أعلاه أن قيمة معامل التحديد تساوي $R^2 = 0.396$ ، هذا يعني أن المتغير المستقل X يفسر 39.6% من تباين المتغير التابع Y ، وتبقى نسبة 60.4% من التغيرات في Y تفسرها متغيرات أخرى لم يتم إدراجها في النموذج.

• **$Adjusted R - Squared$** : معامل التحديد المعدل (المصحح)، يحسب بالعلاقة التالية:

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

حيث:

R^2 : معامل التحديد.

k : عدد المتغيرات المستقلة.

n : عدد المشاهدات.

يفسر معامل التحديد المعدل مثل معامل التحديد تماماً، الفرق بينهما يكمن في كون معامل التحديد المعدل يأخذ في الحسبان عدد المتغيرات المستقلة وحجم العينة مما يجعله أكثر دقة من معامل التحديد، حيث يستعمل في حالة العينات الصغيرة.

• **$S.E. of regression$** : الخطأ المعياري للانحدار وهو مقياس يعبر عن مدى تشتت القيم الفعلية Y حول

القيم المتوقعة $\hat{Y}$ ، وهو يمثل الانحراف المعياري للبواقي $\hat{\sigma}_\varepsilon$ ، يحسب بالعلاقة التالية:

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n - k - 1}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n - k - 1}}$$

قدر الانحراف المعياري للبواقي في النموذج أعلاه بـ: $\hat{\sigma}_\varepsilon = 19.17691$

- **Sum squared resid**: مجموع مربعات البواقي $\sum \varepsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2$ وقد قدرت في النموذج أعلاه بـ: $\sum \varepsilon_i^2 = 17652.18$.

- **Ln(L) Log likelihood**: دالة تعبر عن مدى قدرة النموذج على تفسير البيانات الفعلية، حيث كلما كانت توقعات النموذج قريبة من القيم الحقيقية $\leftarrow$ **Log likelihood** أكبر $\leftarrow$ النموذج أفضل. وكلما كانت التوقعات بعيدة عن القيم الحقيقية $\leftarrow$ **Log likelihood** أصغر $\leftarrow$ النموذج أضعف، تحسب بالعلاقة التالية:

$$Ln(L) = -\frac{n}{2}Ln(2\pi) - \frac{n}{2}Ln(\hat{\sigma}_\varepsilon^2) - \frac{\sum \varepsilon_i^2}{2\hat{\sigma}_\varepsilon^2}$$

حيث:

n: عدد المشاهدات.

π : 3.1416.

ε_i^2 : مجموع مربعات البواقي.

$\hat{\sigma}_\varepsilon^2$: تباين البواقي.

- **F.Statistic**: القيمة الإحصائية لفisher **F** المحسوبة، وتحسب بالعلاقة التالية:

$$F_{cal} = F.Statistic = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}}$$

حيث:

SSR: التباين المفسر $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$.

SSE: التباين غير المفسر $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$.

n: عدد المشاهدات.

k: عدد المتغيرات المستقلة.

يستخدم اختبار **F** لتحديد المعنوية الإجمالية للنموذج، ويتم بناء على الفرضية الصفرية H_0 ، حيث:

$$\begin{cases} H_0: \beta_0 = \beta_1 = 0 & (\text{غير معنوي}) \\ H_1: \beta_0 \neq \beta_1 \neq 0 & (\text{معنوي}) \end{cases}$$

ترفض الفرضية الصفرية إذا كانت **F** المحسوبة أكبر من **F** المجدولة (الدرجة)، عند مستوى المعنوية α .

• **Prob**: القيمة الاحتمالية الموافقة لـ **F** المحسوبة.

كما يمكن الاعتماد على القيمة الاحتمالية لاختبار صلاحية النموذج (معنوية النموذج بالإجمال)، حيث ترفض الفرضية الصفرية إذا كانت القيمة الاحتمالية الموافقة لاختبار **F** أصغر من مستوى الدلالة α ، أي:
 $(Prob_{F-Statistic} < \alpha)$.

ووفقا لنتائج الجدول أعلاه فإن القيمة الاحتمالية الموافقة لاختبار فيشر $(Prob(F - Statistic))$ كانت أصغر من مستوى المعنوية ($\alpha = 0.05 > Prob_{F-Statistic} = 0.000001$)، وعليه فإنه توجد على الأقل معلمة معنوية ونعتبر أن النموذج معنوي إجمالاً.

• **Mean dependent var**: المتوسط الحسابي للمتغير التابع، يحسب بالعلاقة التالية:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

قدر المتوسط الحسابي للمتغير التابع بـ: $\bar{Y} = 125.5826$.

• **S.D. dependent var**: الانحراف المعياري للمتغير التابع يقيس مدى تشتت القيم الفعلية للمتغير التابع حول متوسطها، يحسب بالعلاقة التالية:

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

قدر الانحراف المعياري للمتغير التابع بـ: $S_Y = 24.42584$.

• **Akaike info criterion**: مقياس يستخدم لتقييم مدى جودة النموذج الانحداري، مع أخذ عدد المعاملات في الحسبان، يحسب بالعلاقة التالية:

$$AIC = \frac{2(k + 1) - 2\ln(l)}{n}$$

حيث:

n: عدد المشاهدات.

k: عدد المتغيرات المستقلة.

ln(l): Log likelihood.

قدر معيار **Akaike** بـ: $AIC = 8.784469$.

- **Schwarz criterion**: مقياس يستخدم لتقييم جودة النموذج مثل **AIC**، لكنه يعاقب النماذج المعقدة بشدة أكبر (أين أنه يجعل النماذج التي تحتوي على عدد كبير من المعاملات أقل تفضيلاً)، يحسب بالعلاقة التالية:

$$BIC = \frac{(k + 1)\ln(n) - 2\ln(l)}{n}$$

قدر معيار **Schwarz** بـ: **BIC = 8.860949**

- **Hannan – Quinn criter**: مقياس لتقييم جودة النموذج يأخذ بعين الاعتبار ملائمة النموذج للبيانات وعدد المعاملات، يحسب بالعلاقة التالية:

$$HQC = \frac{2(k + 1)\ln(\ln(n)) - 2\ln(l)}{n}$$

قدر معيار **Hannan – Quinn** بـ: **HQC = 8.513593**

- **Durbin – Watson stat**: اختبار إحصائي يستخدم للتحقق من وجود ارتباط ذاتي (**Autocorrelation**) من الدرجة الأولى بين البواقي (**Residuals**) في نموذج الانحدار الخطي. ويبنى على أساس الفرضية الصفرية:

$$\begin{cases} H_0: \rho = 0 & (\text{لا يوجد ارتباط ذاتي بين البواقي}) \\ H_1: \rho \neq 0 & (\text{يوجد ارتباط ذاتي بين البواقي}) \end{cases}$$

الصيغة الرياضية لإحصائية ديربن واتسون **DW** المحسوبة:

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$$

تتم مقارنة إحصائية ديربن واتسون المحسوبة مع القيم الجدولية الدنيا **dl** والقيم الجدولية العليا **du** لتحديد وجود أو عدم وجود ارتباط ذاتي من الدرجة الأولى بين البواقي، وذلك كما يوضحه الجدول التالي:

0	<i>dl</i>	<i>du</i>	2	4 – <i>du</i>	4 – <i>dl</i>	4
<i>RH</i>₀	!	<i>RH</i>₀	!	<i>RH</i>₀	!	<i>RH</i>₀
يوجد ارتباط ذاتي بين البواقي	لا يمكن الجزم	لا يوجد ارتباط ذاتي بين البواقي	لا يمكن الجزم	يوجد ارتباط ذاتي بين البواقي	لا يمكن الجزم	لا يوجد ارتباط ذاتي بين البواقي

حيث:

dl : (lower bound) الحد الأدنى من القيم الحرجة التي يقدمها جدول ديربن واتسون.

du : (upper bound) الحد الأعلى من القيم الحرجة التي يقدمها جدول ديربن واتسون.

RH_0 : رفض الفرضية الصفرية.

$\bar{RH}_0$: قبول الفرضية الصفرية.

قدرت قيمة ديربن واتسون المحسوبة بـ $DW = 1.528258$ ، أما القيم الحرجة الدنيا والعليا فقد قدرتا بـ:

$dl = 1.503$, $du = 1.585$ ، إذا $DW = 1.528258 \in [1.503, 1.585]$ ، وعليه لا يمكن الجزم

بوجود أو عدم وجود ارتباط ذاتي ويجب اجراء اختبارات لاتخاذ القرار المناسب.

0	$dl = 1.503$	$du = 1.585$	2	$4 - du = 2.415$	$4 - dl = 2.497$	4
RH_0	!	$\bar{RH}_0$		!	RH_0	
يوجد ارتباط ذاتي بين البواقي	لا يمكن الجزم	لا يوجد ارتباط ذاتي بين للبواقي		لا يمكن الجزم	يوجد ارتباط ذاتي بين البواقي	