



Correlation & Regression

تحليل الارتباط والانحدار

Dr. Eihab Damman

الارتباط والانحدار

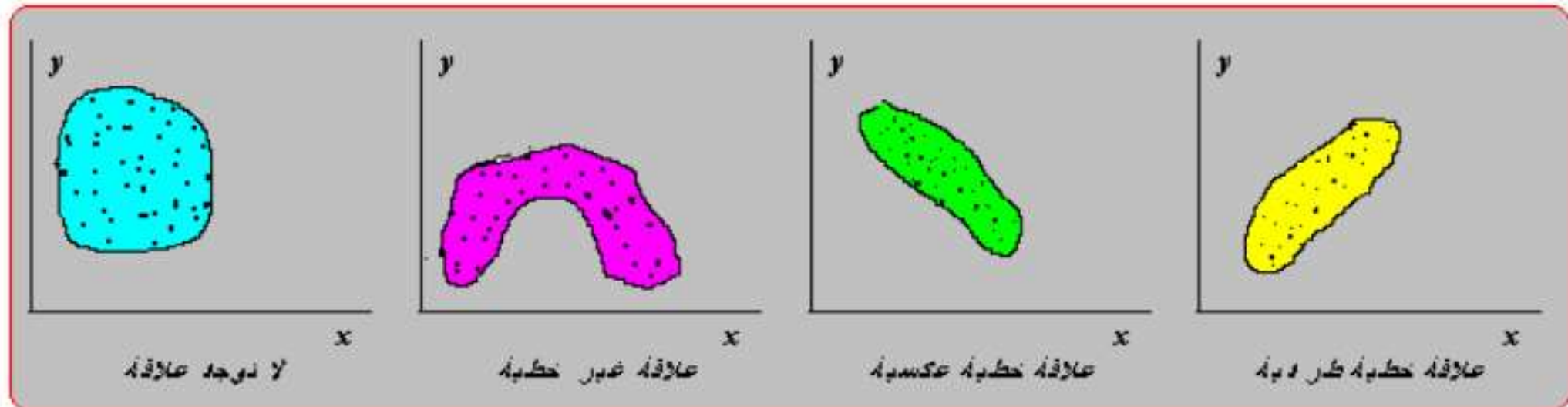
ننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار، ومن الأمثلة على ذلك:

- ١- الإنفاق، والدخل العائلي.
- ٢- سعر السلعة، والكمية المطلوبة منها.
- ٣- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.
- ٤- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكولسترول في الدم.
- ٥- وزن الجسم، وضغط الدم.

الارتباط والانحدار

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين (y, x) ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل الانتشار لبيان نوع العلاقة بين y, x



الارتباط البسيط Simple correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار.

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r ، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط،

الارتباط البسيط Simple correlation

يركز هذا التحليل على نقطتين

- نوع العلاقة:- وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:
 - ١- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.
 - ٢- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .
 - ٣- إذا كان معامل الارتباط قيمته صفراً ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.
- قوة العلاقة:- ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى ($-1 < r < 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

الارتباط البسيط Simple correlation

ارتباط عكسي					ارتباط طردي					
شوي جدا	شوي	متوسط	ضعيف	متوسط جدا	متوسط جدا	ضعيف	متوسط	شوي	شوي جدا	
-1	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9	1
نام				نام						نام

في حالة جمع بيانات عن متغيرين كميين (y, x) ، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" **Pearson**، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بين الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة. ولحساب معامل الارتباط في العينة، تستخدم صيغة "بيرسون" التالية:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

حساب معامل الارتباط Correlation باستخدام Excel

- بعد إدخال البيانات في صفحة إكسل (استخدم بيانات الملف Examples2 الورقة مثال 1) نستطيع تقدير معامل الارتباط بيرسون بين متغيرين وذلك بالذهاب إلى تبويب بيانات Data ثم اختيار الأمر Data analysis
- فتظهر لدينا النافذة التالية:

Data Analysis



Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram


OK

Cancel

Help

Correlation ? X


Input

Input Range: 

Grouped By: Columns
 Rows

Labels in first row

Output options

Output Range: 

New Worksheet Ply:

New Workbook

OK **Cancel** **Help**



Microsoft Excel interface showing the DATA tab ribbon and a spreadsheet with a linear programming problem.

Ribbon: DATA

- Get External Data: From Access, From Web, From Text, From Other Sources, Existing Connections
- Connections: Refresh All, Properties, Edit Links
- Sort & Filter: Sort (AZ, ZA), Filter, Clear, Reapply, Advanced

Spreadsheet Data:

	A	B	C	D	E	F	G	H
1			سعر القطعة	عدد ساعات العمل				
2			Y	X				موارات
3			150	15				
4			190	20			Y	X
5			175	19		Y	1	
6			325	22		X	0.911258	1
7			210	18				
8			160	13				
9			240	20				
10			400	30				
11			350	28				
12			310	21				
13								

The value 0.911258 in cell G6 is highlighted with a red dashed box.

إجراء تحليل الانحدار Regression باستخدام Excel

- يعد الانحدار أحد أهم الأساليب الإحصائية. يختص بقياس العلاقة بين متغير يسمى المتغير التابع، ومتغير آخر أو مجموعة متغيرات تسمى المتغيرات المستقلة.

- إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:
- دراسة أثر الإنتاج على التكلفة.
 - دراسة أثر كمية البروتين التي يتناولها الإنسان على الزيادة في الوزن.
 - دراسة أثر الدخل على الإنفاق الاستهلاكي.
 - دراسة أثر حجم القروض على عدد الأنشطة.
 - وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

Simple Linear Regression الانحدار الخطي البسيط

- يعتبر أبسط أنواع نماذج الانحدار

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$Y = \beta_0 + \beta_1 X$$

Y : هو المتغير التابع (الذي يتأثر)

X : هو المتغير المستقل (الذي يؤثر)

β_0 : هو الجزء المقطوع من المحور الرأسي Y ، وهو يعكس قيمة المتغير التابع في

حالة انعدام قيمة المتغير المستقل X ، أي في حالة $X = 0$

β_1 : ميل الخط المستقيم $(\beta_0 + \beta_1 X)$ ، ويعكس مقدار التغير في Y إذا تغيرت X

بوحددة واحدة.

Data Analysis



Analysis Tools

- Histogram
- Moving Average
- Random Number Generation
- Rank and Percentile
- Regression**
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two Sample for Means

OK

Cancel

Help

Regression

Input

Input Y Range:

Input X Range:

Labels Constant is Zero

Confidence Level: %

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

Residuals Residual Plots

Standardized Residuals Line Fit Plots

Normal Probability

Normal Probability Plots

OK
Cancel
Help



D	E	F	G	H	I	J	K	L	M	N
15										
20										
19		SUMMARY OUTPUT								
22										
18		<i>Regression Statistics</i>								
13		Multiple R	0.911258257							
20		R Square	0.830391612							
30		Adjusted R Square	0.809190563							
28		Standard Error	38.726763							
21		Observations	10							
		ANOVA								
			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
		Regression	1	58741.90262	58741.90262	39.16748	0.000243486			
		Residual	8	11998.09738	1499.762173					
		Total	9	70740						
		<i>Coefficients</i>	<i>Standard Error</i>		<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
		Intercept	-68.36743044	52.47917766	-1.302753463	0.228906	-189.3846311	52.64977	-189.385	52.64977
		X	15.50327332	2.477197294	6.258392643	0.000243	9.790846118	21.2157	9.790846	21.2157

الجذر التربيعي (كلما اقترب من الواحد كان النموذج أفضل)

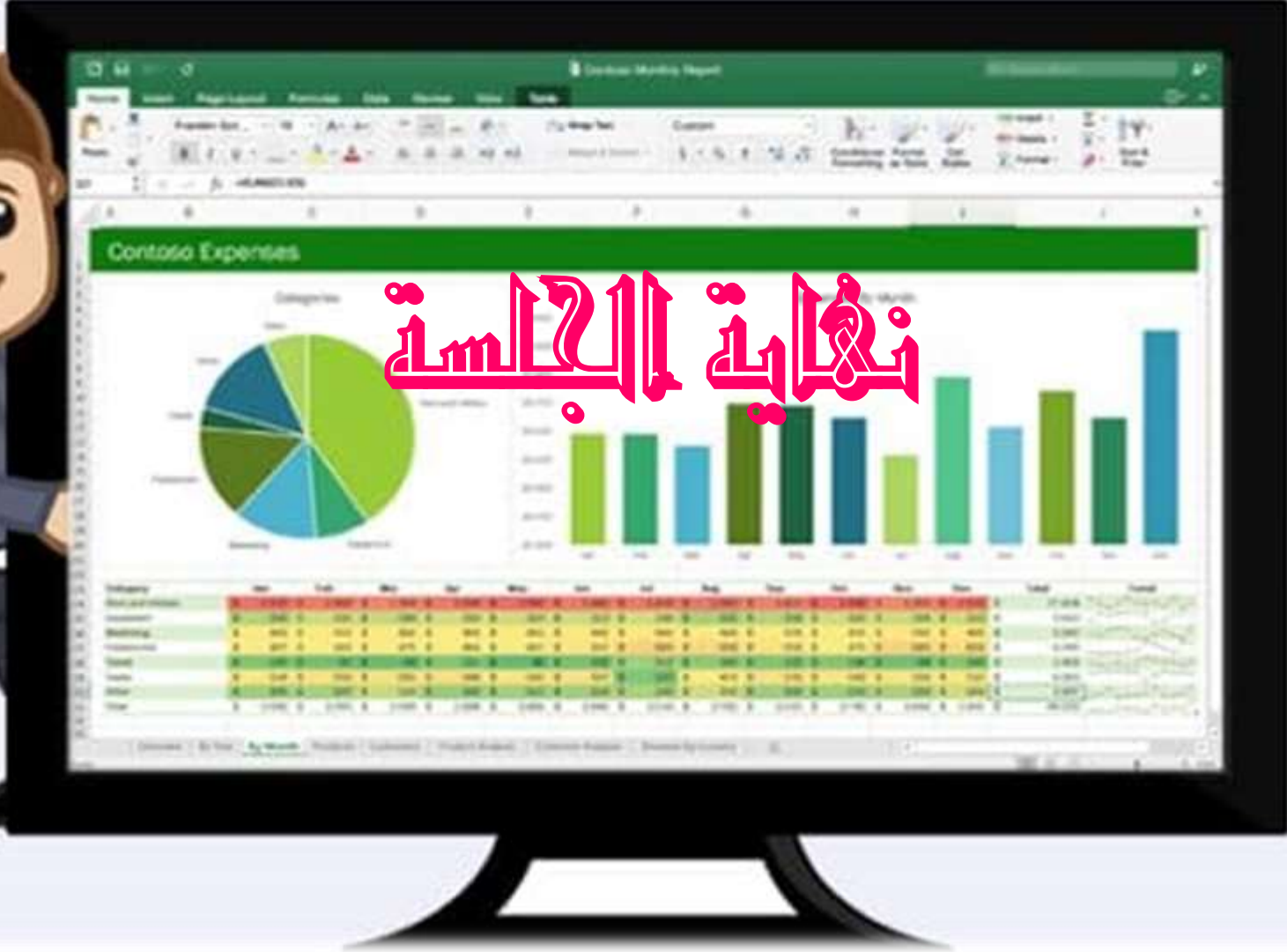
اختبار F (يجب أن تكون معنويتها أقل من 0.05)

قيم ألفا وبيتا للمعادلة حيث تكتب المعادلة الناتجة:

$$Y = -68.37 + 15.5 X$$

اختبار الخطأ القياسي
لمعلمت النموذج

معنوية اختبار T لمعلمت النموذج
إذا كانت أقل من 0.05 تكون معنوية
أقل من 0.01 معنوية جداً



نهاية الجلسة