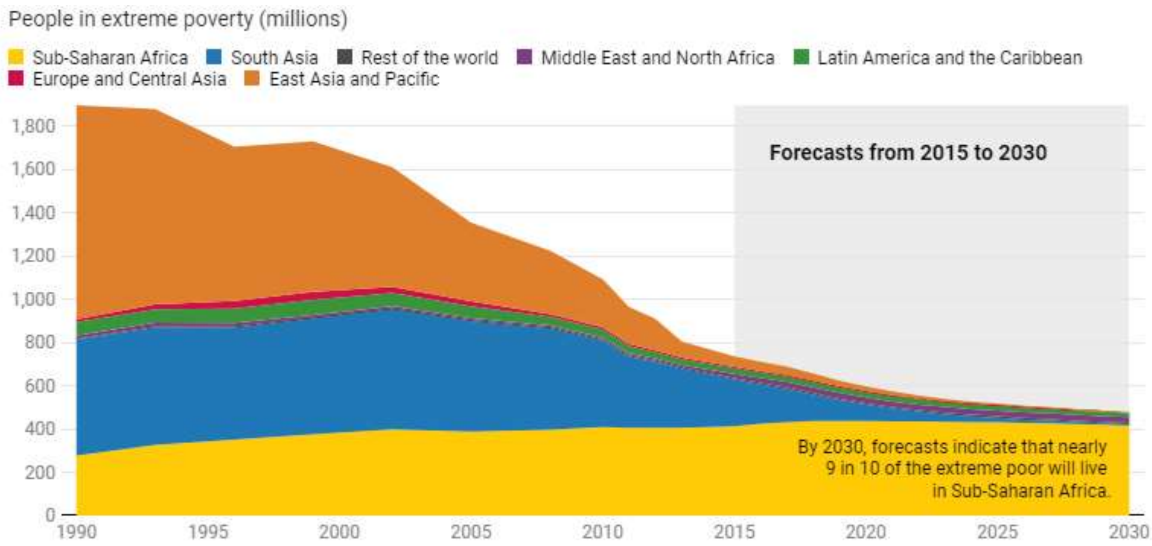
محتويات الفصل 4. تحليل التوجه

فصل 4. تحليل التوجه

معادلة التوجه الخطي– التوجه متعدد الحدود – خلاصة – ملحق – تمارين

توطئة. في المثال التالي[1] يقدر البنك الدولي من توجه السلاسل الزمنية لعدد السكان في فئة الفقر الشديد في العالم أن هذه الظاهرة في انخفاض في مختلف مناطق العالم، إلا في منطقة "إفريقيا جنوب الصحراء" حيث التوجه صاعد، ويتوقع أن الفقر سيتركز في هذه المنطقة بنسبة 9 من كل 10 فقراء في 2030.



صورة 1. التوجه العام لظاهرة الفقر الشديد في العالم (أقل من 1.9 دولار يوميا) هو الانخفاض ما عدا في إفريقيا جنوب الصحراء حيث التوجه صاعد. إذا استمر هذا التوجه، سيتركز 9 من 10 من السكان شديدي الفقر في هذه المنطقة في 2030.

المصدر : البنك الدولي.

يدرس هذا الفصل تحليل التوجه في السلسلة الزمنية وأهميته بينها المثال. ندرس أولا "المؤشر" و"التفريق"، وهي مفاهيم تعطي المحلل نظرة عن التوجه وتسمح بإبرازه أو سحبه من السلسلة لتفحص المكونات الأخرى... المبحث الثاني يخص كيفية استخراج دالة التوجه الخطي واستخدامها للتنبؤ. في المبحث الثالث نتطرق للتوجه متعدد الحدود، وفي الملحق نعرج على بعض دوال التوجه الأخرى. كل ذلك مع التطرق لكيفية الحساب بالبرمجيتين إكسال و R. الغرض من هذا الفصل هو أن يصبح الطالب قادرا على استكشاف التوجه في البيانات وعزله عن السلسلة باستخدام التفريق والتعبير عنه بدالة واستخدامها للتنبؤ مع القدرة على تقييم صلاحية النموذج المستخدم، وذلك يدويا وباستخدام الحاسوب.

---

[1] https://www.worldbank.org/en/news/feature/2018/12/21/year-in-review-2018-in-14-charts
Retrieved: 01-01-2019.

1. معادلة التوجه الخطي

التنبؤ باستخدام خط التوجه

طرق لتخفيف الحسابات

اختبار التوجه

تقييم جودة تمثيل البيانات بخط التوجه

إستخدام Excel

إستخدام R

1-1. التنبؤ باستخدام دالة المربعات الصغرى

ليكن T **أصل التنبؤ** أي اللحظة التي نريد منها تكوين التنبؤ، انطلاقا من بيانات سابقة $\{y_1, \ldots, y_T\}$.

إذا كانت الظاهرة **لا تتضمن دورة ولا موسمية**، يمكن التنبؤ للأفق h، أي للتاريخ (T + h) باستخدام مكون التوجه (f)، حيث تكون المتغيرة Y دالة في الزمن و في مكون الخطأ: $y_t = f(t, e_t)$، حيث e_t هو متغيرة مركزية غير مرتبطة تمثل الفرق بين القيمة الحقيقية والنموذج، هذا الأخير يكتب كما يلي:

$$\hat{y}_T(h) = f(T + h)$$

الدالة التي نريد استخراجها هي كما يلي (نفترض أننا ننطلق من اللحظة 0 وبالتالي الأفق هو t نفسها):

$$\hat{y}(t) = a(t) + b$$

لحساب معاملات دالة التوجه نحتاج إلى التعبير عن الزمن ب $(t = 1, 2, 3, \ldots, n)$، بشرط أن تكون الفترة بين التواريخ متساوية. ميل الدالة، أي المعامل (a)، هو نسبة التباين المشترك إلى تباين t. يحسب الميل (أو معامل الانحدار) والثابت في الدالة بطريقة المربعات الصغرى كما يلي:

$$a = \frac{S_{ty}}{S_t^2} = \frac{\sum_t [(t - m_t) \times (y_t - m_y)]}{\sum_t (t - m_t)^2};$$

$$b = m_y - am_t;$$

$$m_t = (n + 1)/2\,; \quad m_y = \frac{1}{n}\sum_{t=1}^{n} y_t$$

تفسير الميل (a): هو **متوسط تغير** y عند انتقال t بوحدة زمنية واحدة (نفس وحدة t)، وهو أيضا **تغير** y^ من فترة لفترة موالية.

تفسير الثابت (b): هو القيمة التي تأخذها المتغيرة y^ في اللحظة 0، أي نقطة تقاطع خط الدالة مع المحور العمودي.

سؤال: ما هي وحدة الميل a، وما هي وحدة الثابت b (الجواب في التعريف).

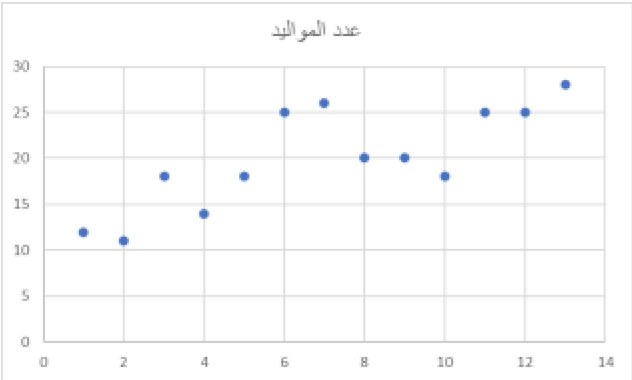
عيب الطريقة أنه كلما طال أفق التنبؤ كلما زاد احتمال دخول تغييرات على التوجه.

مثلا: الجدول التالي يعطي عدد المواليد في عيادة أمومة ما على مدى 13 أسبوعيا.

- إستخدم التمثيل البياني للنظر هل تصلح دالة التوجه الخطي للتنبؤ.
- أحسب معاملات دالة التوجه العام وأكتب الدالة وفسر قيم المعاملين.
- أحسب العدد المتوقع للحوادث خلال الشهرين 14 و15.

الأسبوع	1	2	3	4	5	6	7	8	9	10	11	12	13
عدد المواليد	12	11	18	14	18	25	26	20	20	18	25	25	28

الحل:



رسم بياني 1. تمثيل بياني من نوع سحابة النقاط للتحقق مما إذا كان هناك توجه خطي.

يمكن القول أن الرسم يظهر توجها عاما خطيا، ولذلك يمكن استخدام دالة توجه خطي للتنبؤ.

لاحظ. يمكن أيضا أن يجادل البعض بأن التوجه غير خطي، وبالتالي تكون دالة غير خطية تعطي تمثيلا أفضل للبيانات وقدرة أفضل على التنبؤ، ولكن هذا موضوع المبحث الموالي.

t	y_t	$(t - m_t)(y - m_y)$	$(t - m_t)^2$
1	12	(1 – 7)(12 – 20) = 48	$(1-7)^2=6^2=36$
2	11	(2 – 7)(11 – 20) = 45	$(2-7)^2=5^2=25$
3	18	(3 – 7)(18 – 20) = 8	$(3-7)^2=4^2=16$
4	14	(4 – 7)(14 – 20) = 18	9
5	18	10	4
6	25	-5	1
7	26	0	0
8	20	0	1
9	20	2	4
10	18	21	9
11	25	20	16
12	25	25	25
13	28	0	36
m_t (13+1)/2 = 7	**m_y = 20**	**205**	**182**

$$a = \frac{205}{182} = 1.13; \qquad b = \mathrm{m_y} - a\mathrm{m_t} = 20 - 1.13(7) = 12.12$$

$$\widehat{y(t)} = \mathbf{1.13(}\boldsymbol{t}\mathbf{) + 12.12}$$

تفسير الميل (a):

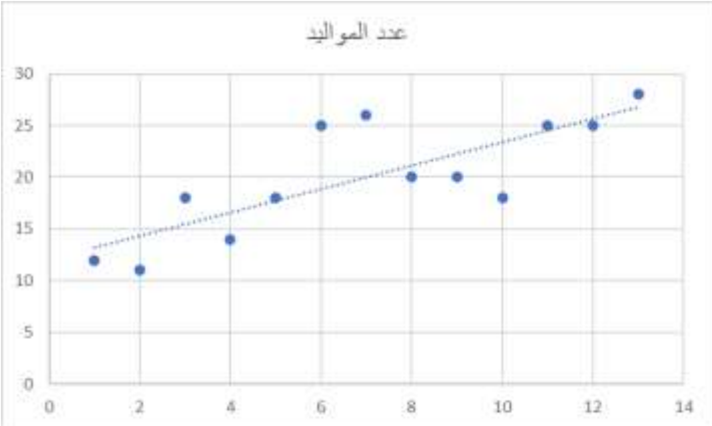
في المتوسط يزيد عدد المواليد أسبوعيا ب 1.13، مثلا تحقق من أن الفرق بين عدد المواليد المتوقع للأسبوعين 14 و15 هو 1.13، وكذلك بين الأسبوعين 15 و16...

تفسير الثابت (b):

الثابت هنا له تفسير حسابي بحت، وهو نقطة نقاطع خط الدالة مع المحور العمودي. ليس هناك دائما تفسير اقتصادي للثابت. مثلا: قم بتعويض t ب 0 في الدالة وتحقق أن النتيجة هي 12.12.

للتنبؤ بقيم السلسلة في تواريخ لاحقة (عدد المواليد)، يكفي أن نعوض في الدالة. مثلا للتنبؤ للأسبوعين 14 و 15:

$$\hat{y}_{14} = 1.13(14) + 12.12 = 27.94;\ \hat{y}_{15} = 1.13(15) + 12.12 = 29.07;$$



رسم بياني 2. التمثيل البياني لخط التوجه. لفعل ذلك رسم نقطتين بتعويض قيمتين ل t في الدالة ثم الربط بينهما ثم تمديد الخط.

2-1. طرق لتخفيف حسابات معاملي الدالة

1- استخدام المتتالية

لأن t هي متتالية حسابية من 1 إلى n يمكن استخدام عدد من الخصائص مثل:

$$m_t = \frac{n+1}{2}; V(t) = \frac{n^2-1}{12};\ \sum t^2 = \frac{n(n+1)(2n+1)}{6} = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}$$

مثلا يمكن استخدام الصيغتين أعلاه لحساب المتوسط والتباين ل t عند حساب الميل:

$$a = \frac{\frac{1}{n}\sum ty_t - m_t m_y}{V(t)} = \frac{\frac{1}{n}\sum ty_t - \frac{n+1}{2} m_y}{\frac{n^2-1}{12}}$$

بهذه الصيغة يحتاج حساب الميل إلى عمود واحد فقط هو ty_t.

مثال. إستخدم الطريقة المختصرة أعلاه لحساب الميل والثابت في المثال السابق.

t	y_t	$t*y_t$
1	12	12
2	11	22
3	18	54
4	14	56
5	18	90
6	25	150
7	26	182
8	20	160
9	20	180
10	18	180
11	25	275
12	25	300
13	28	364
m_t (13+1)/2 = 7	m_y = 20	(1/n)∑ $t*y_t$ =155,7692

$$a=\frac{\frac{1}{n}\sum ty_t-\frac{n+1}{2}m_y}{\frac{n^2-1}{12}}=\frac{155.7692-7(20)}{(13^2-1)/12}=1.13;$$

$$b=20-1.13(7)=12.12$$

$$\hat{y}_t=1.13t+12.12$$

2-طريقة إعدام المتوسط

هناك أيضا طريقة مختصرة للحساب لكن أراها أقل جدوى من التي ذكرنا للتو، وإنما نذكرها للفائدة، وتتمثل في التعبير عن الزمن بطريقة تجعل متوسطه معدوما $m_t = 0$، وذلك بطرح m_t من قيم t. بهذه الطريقة نحتاج إلى عمود واحد (t’y) حيث $t' = t - m_t$ بالإضافة طبعا لحساب قيم t’ (Droesbeck, 1997:)

$$a=\frac{\sum t'y}{\sum t'^2}=\frac{\sum t'y}{\frac{n'(n'+1)(2n'+1)}{3}};b=m_y$$

3- دالة ماير Mayer

تعتمد طريقة ماير على استخراج دالة خطية من خلال تقسيم بيانات السلسلة إلى سحابتين، ومن ثم استخرج مركز كل سحابة (إحداثيات كل مركز هي متوسط قيم t مع متوسط قيم y_t المقابلة لها في نفس السحابة)، ثم التوصيل بين المركزين واستخراج دالة هذا الخط الذي يربط بينهما من خلال حل جملة معادلتين.

مثال: قم بالخطوات السابقة لاستخراج دالة ماير على المثال السابق.

استخراج مركزي السحابتين:

t		y_t	
1	(1+2+3+4+5+6+7)/7=4	12	(12+11+18+14+18+25+26)/7=17.71
2		11	
3		18	
4		14	
5		18	
6		25	
7		26	
8	(8+9+10+11+12+13)/6=10.5	20	(20+20+18+25+25+28)/6=22.67
9		20	
10		18	
11		25	
12		25	
13		28	

استخراج دالة ماير باستخدام مركزي السحابتين:

$$17.71 = a(4) + b$$

$$22.67 = a(10.5) + b$$

$$22.67 - 17.71 = a(10.5-4) = 6.5a \Rightarrow a = 4.96/6.5 = 0.76$$

$$b = 17.71 - 0.76(4) = 14.66 \Rightarrow \hat{y} = 0.76\,t + 14.66$$

4- اختصار البيانات

يمكن لتخفيف الحسابات في حالة كون البيانات كثيرة (مثلا مئات المشاهدات) الاقتصار على عدد محدود من القيم، كأن نقتصر على المتوسطات السنوية في بيانات شهرية أو ثلاثية، وهكذا. اليوم مع استخدام الاعلام الآلي لم يعد هذا التسهيل مهما.

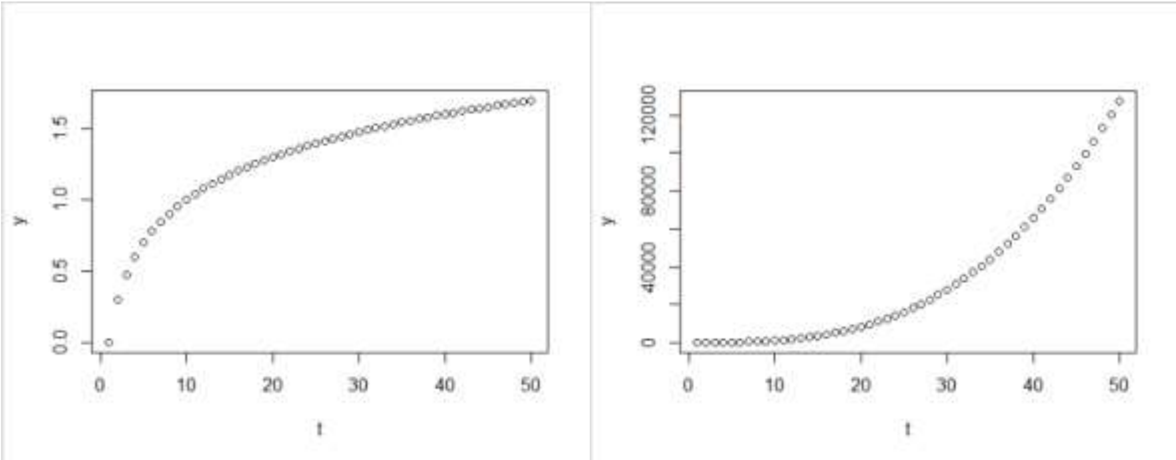
1-3. إختبار دانيال للتوجه

يمكن أن يكون للسلسلة توجه غير ظاهر من الوهلة الأولى. في مثل هذه الحالة يمكن للتحقق من وجود أو عدم وجود توجه في السلسلة استخدام اختبار إحصائي للتوجه، مثل إختبار دانيال.

اختبار دانيال للتوجه هو اختبار "غير معلمي" لا يتطلب طبيعية التوزيع الاحتمالي للمتغيرة. يعتمد اختبار دانيال على إحصائية هي ارتباط الرتب لسبيرمان.

قيمة معامل سبيرمان مثل معامل بيرسون هو قيمة بين -1 و 1، وكلما اقترب من 1 دل على علاقة طردية قوية، وكلما اقترب من -1 دل على علاقة عكسية قوية، وكلما اقترب من 0 دل علاقة ضعيفة أو منعدمة. لكن على خلاف معامل بيرسون الذي يقيس العلاقة الخطية، ميزة معامل بيرسون أنه يقيس العلاقة النمطية (monotonic trend)؛ أي التي لا تغير اتجاهها من طردية إلى عكسية أو العكس. كلما كانت العلاقة نمطية أكثر كلما اقترب معامل بيرسون من 1 أو -1 حتى لو لم تكن العلاقة خطية،

لكن يمكن أن تكون العلاقة تامة لكن المعامل يأتي قريبا من الصفر إذا كانت هذه العلاقة غير نمطية. ميزة معامل سبيرمان أيضا أنه لا يتأثر بالقيم الشاردة.



رسم توضيحي 1. علاقة نمطية تامة في الحالتين غير خطية. معامل سبيرمان = 1 معامل بيرسون أقل من 1.

نكتب الفرضيات في حالة الاختبار الثنائي – وهو الغالب – كما يلي:

H_0: لا توجد علاقة نمطية في السلسلة.

H_1: هناك علاقة نمطية في السلسلة (توجه صاعد أو نازلا).

الإحصائية المستخدمة في اختبار سبيرمان هي معامل سبيرمان وتكتب كالتالي:

$$r_S = 1 - \frac{6\sum_{t=1}^{n} d_t^2}{n(n^2-1)}$$

حيث n حجم السلسلة، و $d_t = (t - y'_t)$، و y'_t هي الترتيب التصاعدي ل y_t.

لحساب rs نستخرج رتب y_t وتمسى 'y، ثم نستخرج قيم d^2_t في عمود ثاني، ونستخدم مجموعه لحساب المعامل rs حسب الصيغة أعلاه.

تقارن الإحصائية rs في حالة العينة الصغيرة (وعدم وجود قيم مكررة كثيرة) مع قيمة جدولية تستخرج من جداول تعطي الارتباط بحسب قيمة n ومستوى المعنوية α. في حالة الاختبار الثنائي نستخدم القيمة المطلقة للإحصائية، وتكتب قاعدة القرار كما يلي:

$$RH_0 \; if \; |r_s| > r_{s_{n;\,\alpha/2}}$$

مثال[1]: لدينا في العمودين على اليسار كمية الأمطار خلال 19 سنة، ونريد التحقق من وجود أو عدم وجود توجه عام للسلسلة.

---

[1] Conover, W. J., Practical Nonparametric Statistics, 3rd edition, John Wiley & Sons, 1999. P 3234.

year	rainfall	t	y'	$d_t^2 = (t - y')^2$
1950	101,11	1	15	196
1951	75,56	2	7	25
1952	54,44	3	3	0
1953	111,11	4	19	225
1954	81,11	5	9	16
1955	66,67	6	4,5	2,25
1956	73,33	7	6	1
1957	105,56	8	16,5	72,25
1958	51,11	9	2	49
1959	48,89	10	1	81
1960	66,67	11	4,5	42,25
1961	105,56	12	16,5	20,25
1962	93,33	13	13	0
1963	82,22	14	10	16
1964	86,67	15	11	16
1965	106,67	16	18	4
1966	96,67	17	14	9
1967	77,78	18	8	100
1968	87,78	19	12	49
			$\sum d_t^2$	**924**

عدد القيم المتكرر قليل، نستخدم الصيغة المذكورة بدون تعديل.

$$r_S = 1 - \frac{6\sum_{t=1}^{n} d_t^2}{n(n^2-1)} = 1 - \frac{6(924)}{19(19^2-1)} = 0.189$$

القيمة الجدولية نستخرجها من الجداول الإحصائية ل rs بحسب حجم العينة (19) ومستوى المعنوية (0.05)، في اختبار ثنائي وهي في هذه الحالة 0.4579.

بما أن القيمة المطلقة ل rs لا تتجاوز القيمة الجدولية، لا يمكن رفض الفرضية الصفرية وبالتالي نقول أنه لا دليل على وجود توجه.

في حالة العينة الكبيرة (تساوي أو تتجاوز 30) يمكن استخدام التوزيع الطبيعي بحساب الإحصائية التالية ومقارنتها مع قيم جدولية من التوزيع الطبيعي:

$$\boldsymbol{r_s\sqrt{n-1} \approx N(0,1)}$$

في هذه الحالة نكتب قاعدة القرار (للاختبار الثنائي) كما يلي:

$$RH_0 \; if \; \left|r_s\sqrt{n-1}\right| > z_{\alpha/2}$$

على سبيل المثال، لاختبار ثنائي للتوجه بمستوى معنوية 0.05 في حالة العينة أكبر من 30 نقارن القيمة المطلقة للإحصائية $r_S\sqrt{n-1}$ مع القيمة الجدولية 1.96. إذا كان الاختبار أحاديا (وجود توجه صاعد/نازل) نقارن قيمة الإحصائية $r_S\sqrt{n-1}$ ب 1.64 لحالة الشك في وجود توجه صاعد، و$-$1.64 في حالة الشك في وجود توجه نازل.

لاحظ.

- الاختبار أحادي الاتجاه يكون في حالة الاشتباه في نوع محدد من التوجه (صاعد أو نازل)، في هذه الحالة ما يتغير هو الفرضية البديلة والقيمة الجدولية (تتضاعف المساحة وراءها: rs_{α} بدل $rs_{\alpha/2}$)، الفرضية الصفرية وقيمة الإحصائية تبقيان على حالهما.
- في حالة كثرة حالات تساوي الرتب (ties) في قيم y يتعين القيام بتعديل معين للإحصائية rs يأخذ في الاعتبار تلك الحالات؛ أنظر مثلا (Conover, 1999). يمكن أيضا العودة إلى الصيغة الأصلية لمعامل بيرسون وهي ارتباط بيرسون بين رتب المتغيرتين $rs = Cov(t,y')/S_tS_{y'}$ حيث الإشارة أعلى الحرف تعني متغيرة الترتيب.
- يمكن أن تشهد السلسلة تغيرا للمتوسط (أنظر مثال مؤشر بورصة الجزائر)، وهذا يمكن اختباره إحصائيا وتحديد موضع الانتقال بالتحديد.
- هناك عدة بدائل لاختبار دانيال للتوجه، منها اختبار ستيودنت للميل في دالة التوجه الخطي، وهو اختبار معلمي يتطلب طبيعية التوزيع الاحتمالي وهذا شرط يتعين التحقق منه بالرسم البياني مثلا. من البدائل غير المعلمية اختبار مان–كيندل (Test Mann–Kendall) للتوجه واختبار كوكس–ستيوارت (Cox–Stuart) واختبار التوالي (runs test). تتفاوت هذه الاختبارات عن بعضها البعض من حيث القوة (القدرة على كشف التوجه إن وجد من أجل نفس حجم العينة) والمرونة (القابلية لكشف التوجه الدوري مثلا).

4–1. تقييم تمثيل خط دالة التوجه الخطي لبيانات السلسلة من خلال R^2

دقة الدالة ^y في التنبؤ بقيمة y يتوقف على قوة العلاقة بين y و t. وجود علاقة قوية يجعل التقدير باستخدام دالة التوجه أدق، بمعنى أن القيمة المطلقة للفرق المتوقع بين y و^y تكون ضئيلة. والعكس، ضعف العلاقة يجعل استخدام الدالة (النموذج) غير مبرر أصلا، واستخدامها يعطي تقديرات غير دقيقة، أي أن الخطأ المتوقع (القيمة المطلقة للفرق المتوقع بين القيمة الحقيقية ل y والقيمة المتوقعه لها) يكون كبيرا. من أجل ذلك نتفقد العلاقة ونتحقق من انها كافية لتبرير صياغة دالة واسخدامها في التنبؤ.

تظهر قوة العلاقة في الرسم البياني من **سحابة النقاط** (scatter dotes): ارتصاف نقاط السحابة وقربها من خط الدالة يعني علاقة خطية قوية؛ والعكس، يدل تشرذم نقاط السحابة وبعدها عن خط الدالة على ضعف العلاقة أو انعدامها. الرسم مهم أيضا لإظهار أن كانت العلاقة خطية أم لا وإن كانت هناك قيم متطرفة قد تأثر على معاملات الدالة.

يمكن أيضا تقييم قوة العلاقة (وبالتالي جودة الدالة) بحساب **معامل التحديد** R^2. يعطي هذا المعامل **نسبة تباين المتغيرة y_t المفسر بالزمن t. تتراوح قيمة R^2 بين 0 و 1 وكلما اقتربت قيمته من 1 دل على**

علاقة أقوى وكلما اقتربت من 0 دل ذلك على ضعف أو انعدام العلاقة. عموما تعتبر العلاقة قوية كفاية إذا كانت t تفسر نسبة من تباين y قريبة أو أكبر من 50 بالمائة، أي R^2 قريب أو أكبر من 0.5.

في حالة العلاقة الخطية يحسب R^2 بتربيع معامل ارتباط بيرسون بين y وt، وتمثل R^2 نسبة التباين المشترك إلى جداء الانحرافين المعياريين، كما توضح الصيغة أدناه:

$$\boldsymbol{R^2} = \left[\frac{\boldsymbol{S_{ty}}}{\boldsymbol{S_t S_y}}\right]^2 = \frac{\left(\sum_i[(\boldsymbol{t_i} - \boldsymbol{m_t})(\boldsymbol{y_i} - \boldsymbol{m_y})]\right)^2}{\sum_i(\boldsymbol{t_i} - \boldsymbol{m_t})^2 \sum_i(\boldsymbol{y_i} - \boldsymbol{m_y})^2}$$

أو يمكن أيضا حساب R^2 من خلال استخدام العلاقة بين الارتباط والميل a:

$$\boldsymbol{R^2} = \left(\frac{\boldsymbol{aS_t}}{\boldsymbol{S_y}}\right)^2 = \boldsymbol{a}^2 \frac{\sum_i(\boldsymbol{t_i} - \boldsymbol{m_t})^2}{\sum_i(\boldsymbol{y_i} - \boldsymbol{m_y})^2}$$

حيث S_t الانحراف المعياري ل t و S_y الانحراف المعياري ل y.

مثال: أحسب معامل التحديد لبيانات العلاقة المعبر عنها في المثال السابق، وفسر النتيجة.

t	y_t	$(t - m_t)^2$	$(y - m_y)^2$
1	12	$(1 - 7)^2 = 36$	$(12 - 20)^2 = 64$
2	11	$(2 - 7)^2 = 25$	$(11 - 20)^2 = 81$
3	18	$(3 - 7)^2 = 16$	$(18 - 20)^2 = 4$
4	14	$(4 - 7)^2 = 9$	$(14 - 20)^2 = 36$
5	18	4	25
6	25	1	25
7	26	0	36
8	20	1	0
9	20	4	1
10	18	9	49
11	25	16	25
12	25	25	25
13	28	36	0
m_t (13+1)/2 = 7	**m_y = 20**	**182**	**368**

$$R^2 = a^2 \frac{\sum_i(t_i - m_t)^2}{\sum_i(y_i - m_y)^2} = 1.13^2 \frac{182}{368} = 0.627$$

تفسير R^2: معامل التحديد 0.627 يعني أن الزمن يفسر تقريبا 63 بالمائة من تغاير y (عدد المواليد)، وهي نسبة عالية جدا وبالتالي فإن استخدام دالة التوجه الخطي مبرر ويعطي نتائج دقيقة.

لاحظ.

- تستخدم مؤشرات أخرى لتقييم العلاقة بين المتغيرة Y والزمن t، منها نسبة متوسط التباين البيني إلى متوسط التباين المتبقي:

$$F = MSB/MSR$$

تعطي البرمجيات هذه الإحصائيات، وتعطي نتيجة اختبارها إحصائيا.

5-1. إستخراج دالة التوجه الخطي والتنبؤ في *Excel*

إستخراج دالة التوجه الخطي

يمكن الحصول على الدالة وتمثيلها البياني على سحابة النقاط ل Y. الخطوات هي كالتالي:

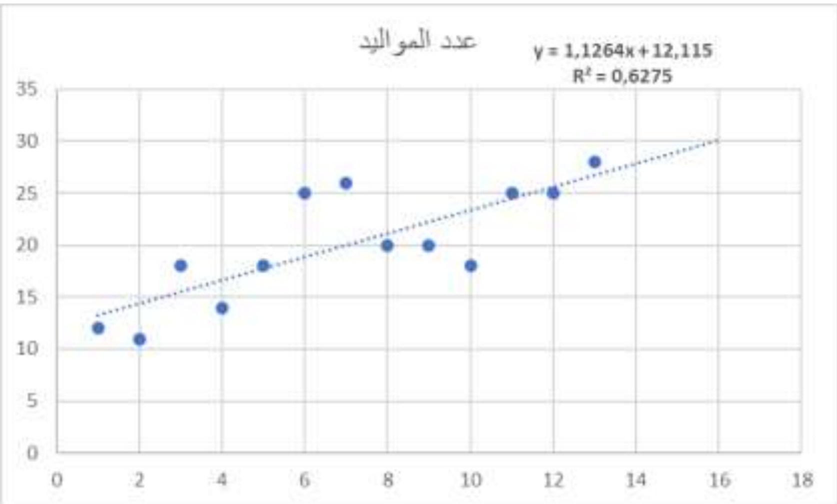
1. تحديد بيانات Y على الورقة ونطلب الرسم البياني من نوع سحابة نقاط (scatter plot).
2. نأشر على السحابة وبالزر الأيمن للفأرة نحصل على قائمة نختار فيها إضافة خط التوجه (Add trend line).
3. نحدد نوع الخط (خطي) ونؤشر لطلب إظهار الدالة (display function) ومعامل التحديد R^2.

Select Data

Insert → Graph → scatter plot.

Select Chart →Add trend line → linear+ display function on chart+ display R^2

بتطبيق هذه الخطوات على المثال السابق يأتي الرسم ما يلي:



رسم بياني 3 التمثيل البياني لخط التوجه الذي تمثله الدالة مع تمديد للفترات اللاحقة وإظهار دالة التوجه ومعامل التحديد R^2. الحرف x في الدالة يمثل المتغير المستقل وهو لدينا الزمن t والحرف y يمثل التابع وهو لدينا ^y.

التنبؤ في Excel

هناك أكثر من طريقة في Excel للتنبؤ باستخدام خط التوجه:

– الدالة **Forecast.linear**: يمكن في Excel 2016 التقدير عن طريق الدالة المذكورة، ويتطلب ذلك تحديد ثلاث بيانات على التوالي (من اليسار إلى اليمين): الزمن المستهدف (x)، بيانات المتغيرة y (known Ys)، بيانات الزمن (Known Xs)، وذلك كما يلي:

=FORECAST.LINEAR(x; known Ys; known Xs)

يمكن بعد ذلك نسخ الخلية إلى الأسفل للتنبؤ بفترات لاحقة مع تثبيت بيانات Ys و Xs.

– الدالة **Trend**: يمكن أيضا التنبؤ لعدد من الفترات دفعة واحدة من خلال الدالة **TREND**:

= TREND (known Ys; known Xs ; new Xs ; true)

– الدالتين **slope** و**intercepte**: تستخدمان لحساب الميل a، والثابت b كما يلي:

=slope(known Ys; known Xs)

=intercepte(known Ys; known Xs)

مثال على استخدام Excel: إستخدم Excel بالطرق المذكورة لاستخراج دالة التوجه وتمثيلها بيانيا واستخراج معامل التحديد ومن ثم التوقع للأشهر الثلاث المقبلة.

الحل. الحسابات للتنبؤ بالطرق الثلاث أعلاه مبينة في الآتي:

	A	B	C	D	E
1	t	y_t			
2	1	12			
3	2	11			
4	3	18			
5	4	14			
6	5	18			
7	6	25			
8	7	26			
9	8	20			Slope
10	9	20			=SLOPE(B2:B14;A2:A14)
11	10	18			
12	11	25			Intercept
13	12	25			=INTERCEPT(B2:B14;A2:A14)
14	13	28	Forecast.linear	Trend	at+b
15	14		=FORECAST.LINEAR(14;B2:B14;A2:A14)	=TREND(B$2:B$14;A$2:A$14;A15:A17)	=E$10*A15+E$13
16	15		=FORECAST.LINEAR(15;B2:B14;A2:A14)	=TREND(B$2:B$14;A$2:A$14;A16:A18)	=E$10*A16+E$13
17	16		=FORECAST.LINEAR(16;B2:B14;A2:A14)	=TREND(B$2:B$14;A$2:A$14;A17:A19)	=E$10*A17+E$13

صورة 2. التقدير في Excel باستخدام الدالة **Forecast.LINEAR** والدالة **TREND** وأيضا باستخدام التعويض في الدالة.

النتيجة تأتي كما يلي:

8	20			Slope
9	20			1,126373626
10	18			
11	25			Intercept
12	25			12,11538462
13	28	Forecast.linear	trend	at+b
14		27,88461538	27,88461538	27,88461538
15		29,01098901	29,01098901	29,01098901
16		30,13736264	30,13736264	30,13736264

يأتي الرسم البياني مع تمديد للأشهر المقبلة والدالة ومعامل التحديد كما هو مبين في الرسم البياني السابق.

6-1. التوجه الخطي في *R*

تستخدم الدالة **lm()** في R لإنشاء نموذج تحليل انحدار. المدخلة الأساسية للدالة هي النموذج formula وهي تحديد المتغيرة التابعة والمتغيرة المستقلة: Y~X. المطة الملتوية تعني أن المتغيرة إلى اليسار هي المتغيرة التابعة.

مثال. نقوم باستخدام الدالة lm() لإنشاء نموذج يربط المبيعات Sales بالزمن t. التعليمات تظهر فيما يلي باللون الأزرق، والنتائج باللون الأسود.

- في البداية ننشئ المتغيرتين sales ، و t وندخل بياناتها.

```
> sales=c(18,30,4,24,30,46,16,44,52,56,30,58,68,72,50,74)
> t=1:16
```

- نطلب تكوين النموذج بالدالة lm() ونسميه lmsales،

```
> lmsales<-lm(sales~t)
```

- ثم نطلب عرض النموذج lmsales

```
> lmsales

Call:
lm(formula = sales ~ t)

Coefficients:
(Intercept)            t
     10.900        3.659
```

- نحسب التوقع للمبيعات في الشهر 13

```
> 10.9+3.659*13
[1] 58.467
```

- نطلب الارتباط بين المتغيرتين،

```
> cor(sales,t)
[1] 0.8249598
```

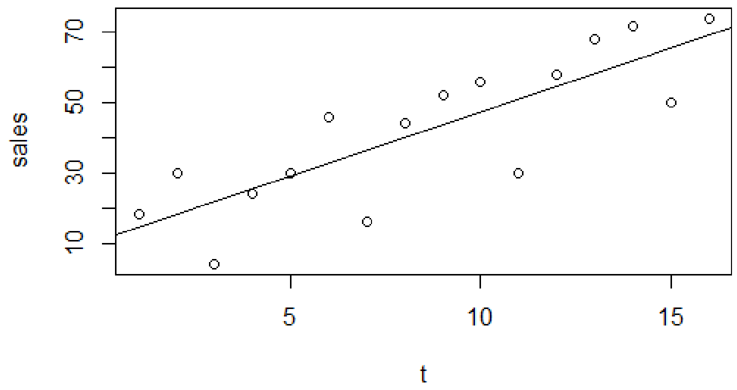
- نعرض مجموعة من الاحصائيات الأخرى للنموذج من خلال الدالة summary

```
> summary(lmsales)
```

```
Call:
lm(formula = sales ~ t)

Residuals:
    Min      1Q  Median      3Q     Max 
-21.147  -5.097   3.635   8.768  13.147 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  10.9000     6.4781   1.683    0.115    
t             3.6588     0.6699   5.461 8.38e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 12.35 on 14 degrees of freedom
Multiple R-squared:  0.6806,    Adjusted R-squared:  0.6577 
F-statistic: 29.83 on 1 and 14 DF,  p-value: 8.383e-05
```

نرسم خط الدالة على سحابة النقاط (t, sales).

```
> plot(t,sales)
> abline(coef(lmsales))
```



للقيام باختبار ارتباط سبيرمان:

```
Cor.test(t,sales,method="spearman")
```

النتيجة تأتي كما يلي:

```
                 Spearman's rank correlation rho

data:  sales and t 
S = 120.35, p-value = 9.009e-05
alternative hypothesis: true rho is not equal to 0 
sample estimates:
      rho 
0.8230124
```

لدينا قيمة معامل سبيرمان (0.823) وهي قيمة موجبة عالية تدل على علاقة طردية قوية نمطية بين الزمن والمبيعات. لدينا أيضا مستوى المعنوية (p-value) قيمة أقل من 0.05 وبالتالي يمكن رفض الفرصية الصفرية وقبول الفرضية البديلة، وهي أن هناك علاقة نمطية بين المبيعات والزمن.

2. دالة التوجه متعددة الحدود

الدالة متعددة الحدود

التوجه غير الخطي في Excel

تقييم النموذج

الدالة متعددة الحدود في R

2-1. الدالة متعددة الحدود *polinomial*

يستخدم التوجه متعدد الحدود عندما يتحول التغير من حيث القوة والاتجاه. تتميز الدالة متعددة الحدود بمرونتها، وهي مرونة تزيد بزيادة درجة القوة التي ترفع إليها المتغيرة، مما يعطيها قدرة أكبر على استيعاب التغيرات في ميل المتغيرة المستهدفة. هذه الميزة لا تتوفر في التنبؤ بطريقة التوجه الخطي البسيط فهذه الدالة تفترض ثبات التوجه. مقارنة مع الدالة الخطية البسيطة التي رأينا سابقا[1]، تضيف الدالة متعددة الحدود حدودا جديدة للنموذج ذات قوة 2 (الدالة التربيعية) أو 2 و3 (الدالة التكعيبية) وهكذا.

الدالة التربيعية:

$$y = b + at + ct^2$$

الدالة التكعيبية:

$$y = b + at + ct^2 + dt^3$$

الدالة التربيعية هي النوع الشائع من الدوال متعددة الحدود. زيادة درجة الدالة يعني مرونة أكبر، لكن عمليا غالبا نبقى في الدرجة الثانية ونادرا ما نذهب أبعد من الدرجة الثالثة، إلا إذا كنا مقتنعين من التوجه الجديد، ولأننا لا نريد نموذجا يتضمن عددا كبيرا من المعالم المقدرة. إستخدام البرمجيات يسمح باختبار عدة نماذج والمفاضلة بينها باستخدام معايير مختلفة منها معامل التحديد R^2 ومستوى معنوية معاملات النموذج، وأيضا من خلال تفحص سحابة المتبقي.

تحل الدالة التربيعية بجملة معادلات كما يلي:

$$bn + a\sum_{i=1}^{n} t_i + c\sum_{i=1}^{n} t_i^{\,2} = \sum_{i=1}^{n} y_i$$

$$b\sum_{i=1}^{n} t_i + a\sum_{i=1}^{n} t_i^{\,2} + c\sum_{i=1}^{n} t_i^{\,3} = \sum_{i=1}^{n} y_i\, t_i$$

$$b\sum_{i=1}^{n} t_i^{\,2} + a\sum_{i=1}^{n} t_i^{\,3} + c\sum_{i=1}^{n} t_i^{\,4} = \sum_{i=1}^{n} y_i t_i^{\,2}$$

[1] الدالة الخطية البسيطة (at+b) تسمى متعددة حدود من الدرجة الأولى فهي حالة خاصة من الدالة متعددة الحدود أين المعاملات الأخرى معدومة.

نحتاج إلى جدول لحساب المجاميع وبالتعويض بقيمها في جملة المعادلات تبقى فقط كمجاهيل المعاملات الثلاث: (a, b, c) ، أي ثلاث مجاهيل في ثلاث معادلات.

الدوال متعددة الحدود من درجات أعلى (تكعيبية فما فوق) يحتاج حلها إلى حسابات طويلة تتولاها البرمجيات، لكن المشكلة في التنبؤ بطريقة الدوال الخطية متعددة الحدود هي في مدى التحقق من التوجه الحديث الطارئ على البيانات.

2-2. تقييم النموذج (أي دالة نختار؟)

لتقييم مدى ملائمة النموذج (الدالة خطية كانت أم غير خطية) نتبع الخطوات التالية:

- تفحص التمثيل البياني لـy مع خط الدالة،
- تفحص التمثيل البياني لسحابة الخطأ.
- تفحص رسم بياني يقيس طبيعية الخطأ.
- تفحص المؤشرات الرقمية (معامل التحديد، الخطأ المعياري، وقيمة F، ومستوى دلالتها)

ما نرمي إليه هو استخراج دالة تمثل البيانات، ويسمى هذا "التعديل" (ajustement)، ونحرص في اختيار الدالة على أن تقلل الخطأ وتحاكي التوجه العام للمتغيرة على المدى البعيد، على أن يكون النموذج مكونا من عدد قليل من المعالم.

يسمح تفحص الرسم البياني للسلسلة وعليها خط التوجه بتفقد ما إذا كانت نقاط السحابة قريبة ومنسجمة حول خط التوجه.

مدى تمثيل الدالة للبيانات يظهر أيضا في عدد من المؤشرات منها: **معامل التحديد**، وقيمة F ومستوى دلالتها، ومستوى دلالة المعاملات.

من أدوات تفحص ملائمة الدالة سحابة المتبقي: يفترض أن تأتي **منتشرة أفقيا حول خط الصفر بالتوازن بين أعلاه وأسفله بدون أي نمط ظاهر** في توزيع النقاط، وبدون قيم شاردة. مثلا، قد تأتي سحابة النقاط أحيانا في شكل مثلث ويدل ذلك على عدم تساوي التباين، أو تأتي في شكل خط منحنى ويعني ذلك أن الدالة الخطية غير ملائمة لتمثيل البيانات. إذا ظهرت هذه الأنماط في سحابة الخطأ فهذا يعني أن الدالة ليست الأفضل للتعبير عن البيانات (حتى لو كان معامل التحديد عاليا).

للحصول على هذه المخرجات يمكن استخدام برمجيات إحصائية مختصة مثل SPSS أو XLSTAT أو أيضا الاعتماد على الإضافة الملحقة ب Excel: محلل البيانات.

3-2. التوجه غير الخطي في *Excel*

نستخرج في Excel الدالة متعددة الحدود من نفس النافذة التي نستخرج منها التوجه الخطي، وفيها نحدد درجة الدالة وإظهارها على الرسم وكذا إظهار معامل التحديد R^2 لتقييم مدى ملائمة خط الدالة للبيانات.

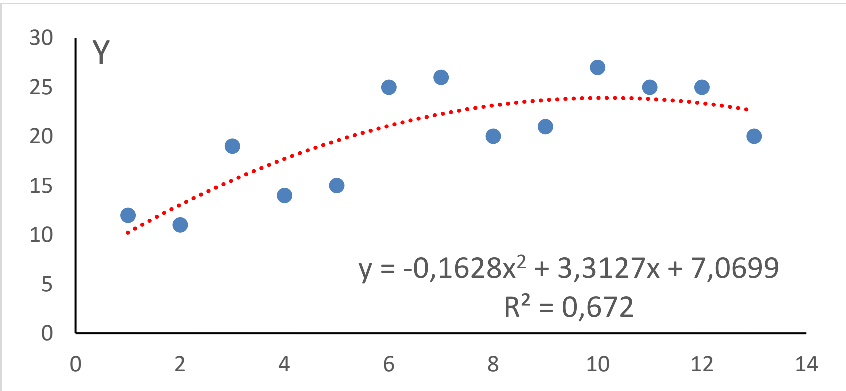
مثلا: لديك بيانات المبيعات للأشهر 13 الماضية.

t	1	2	3	4	5	6	7	8	9	10	11	12	13
y	12	11	19	14	15	25	26	20	21	27	25	25	20

- مثل البيانات في سحابة نقاط في Excel، ثم أضف إليها خط ودالة التوجه من النوع متعدد الحدود مع إظهار R^2 وقارن مع قيمته في حالة استخدام دالة توجه خطي.
- استخدم الدالة للتنبؤ بالمبيعات في الشهر 14.

الحل: تظهر سحابة النقاط هنا أن التوجه الخطي قد لا يكون أحسن تعبير عن توجه عدد الحوادث في المستقبل بما أن السحابة تغير توجهها في نهاية مسارها. من المناسب إذا استخدام دالة متعددة الحدود.

باستخدام إضافة خط التوجه وإظهار الدالة وR^2 نحصل على:



قيمة R^2 هنا قيمة تعتبر جيدة، فهي تدل على أن الزمن يفسر ما نسبته 67.2 بالمائة من عدد الحوادث. الدالة إذن دقيقة إلى حد مقبول جدا في التنبؤ.

$$\hat{y}_{14} = -0.1628(14^2)+3.3127(14)+7.0699 = \dots$$

يمكن استخدام محلل البيانات في Excel للحصول على معاملات الدالة متعددة الحدود وعدد من المؤشرات والرسوم البيانية لتقييم النموذج. لاستخراج الدالة متعددة الحدود نتبع الخطوات التالية.

- ندخل بيانات y والمتغيرات المستقلة (t و t^2 مثلا إذا كانت الدالة تربيعية أو t و t^2 و t^3 إذا أردنا دالة تكعيبية ...) في أعمدة في ورقة Excel
- نطلب محلل البيانات من القائمة الرئيسية Data وفي القائمة التي يعطينا نختار الأداة الإحصائية المناسبة وهي تحليل الانحدار:

Data –> data analysis –> Regression

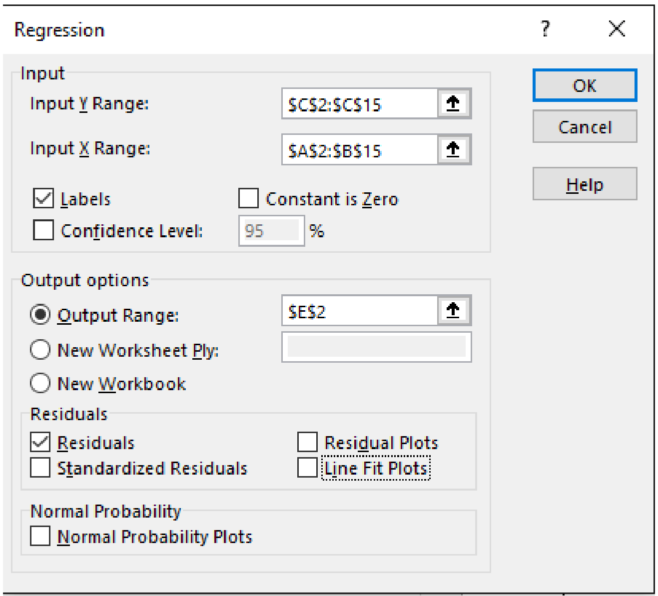
– في النافذة التي تبرز ندخل بيانات المتغيرة التابعة y في الخانة المقررة لها (بكتابة مجال الخلايا أو من خلال تعيين مجال البيانات بالفأرة) ثم المتغيرات المستقلة (مثلا t و t^2).

نحصل من هذا على ثلاثة جداول، حيث نجد في الجدول الثالث الدالة وفي الجدول الأول معامل التحديد (أنظر المثال أدناه). يمكن أيضا طلب إحصائيات أخرى مثل المتبقي Residuals من أجل حساب مؤشرات الدقة MSE وغيره، أو طلب سحابة المتبقي Residuals Plot والرسم البياني لطبيعة الخطأ Normal Probability Plot (أنظر الصورة أدناه).

مثال. في المثال أعلاه استخدم محلل البيانات data analysis لاستخراج دالة انحدار متعدد ل y على t و t^2 (دالة متعددة الحدود تربيعية)، وقيم النموذج.

الحل: بما أن النموذج متعدد الحدود نحتاج أولا لكتابة قيم t^2 في ورقة Excel بجوار t قبل طلب محلل البيانات، لأننا نحتاج إلى إدخالها ضمن المدخلات في نافذة Regression لمحلل البيانات.

	A	B	C
1	Months		Sales
2	t	t^2	y
3	1	1	12
4	2	4	11
5	3	9	19
6	4	16	14
7	5	25	15
8	6	36	25
9	7	49	26
10	8	64	20
11	9	81	21
12	10	100	27
13	11	121	25
14	12	144	25
15	13	169	20



صورة 3. إدخال البيانات والخياريات في مربع الحوار لمحلل البيانات: 1- إدخال البيانات إلى برنامج تحليل الانحدار في محلل البيانات، 2- التاشير على **Labels**، 3- تعيين خلية الإخراج، 4- طلب المتبقي. يمكن أيضا طلب **Normal Probability Plot** و **Residuals Plots** و **Line Fit Plots**.

الصورة أعلاه تبين كيفية تعيين البيانات لمحلل البيانات. لاحظ أن مجال الخلايا الذي يعين المتغيرات المستقلة يتضمن عمودين A وB، وهما الذين وضعنا فيهما بيانات t و t^2.

المخرجات تأتي كما يلي:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,8197
R Square	0,6720
Adjusted R Square	0,6064
Standard Error	3,4744
Observations	13

ANOVA

	Df	SS	MS	F	Significance F
Regression	2	247,28	123,64	10,242	0,0038
Residual	10	120,72	12,072		
Total	12	368			

MODEL

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	7,0699	3,401	2,079	0,064	-0,507	14,647
t	3,3127	1,117	2,965	0,014	0,823	5,802
t^2	-0,1628	0,078	-2,097	0,062	-0,336	0,010

يمكن إذن كتابة الدالة من الجدول الأخير (عمود المعاملات Coefficients)

$$\hat{y}_t = 7.0699 + 3.3127t - 0.1628t^2$$

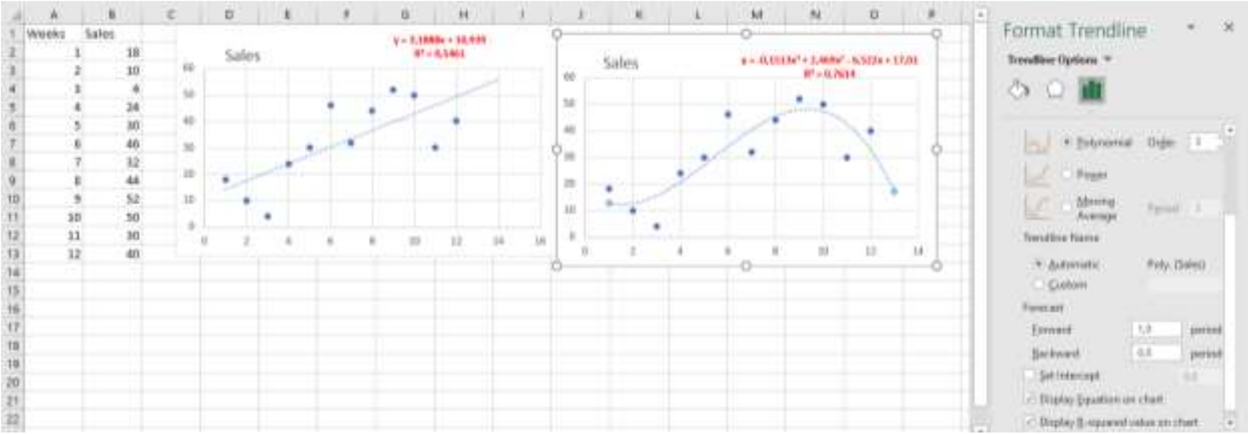
قيمة R^2 تستخرج من الجدول الأول $R^2=0,6720$

جدول ANOVA يبين أن النموذج دال ($\text{Significance F} = 0.0038 < 0.05$)، أي أنه أفضل من استخدام متوسط y في التنبؤ. الجدول الثالث يعطي بالإضافة لمعاملات الدالة مستوى دلالة كل معامل، حيث نعتبر المتغيرة دالة (لها تأثير على التابع) إذا جاء مستوى دلالة معاملها (P–value) أقل من 0.05.

لاحظ. يمكن أيضا إضافة حدود إضافية t^4 ; t^3 لجعل الدالة متعددة الحدود تتمتع بمرونة تمكنها من التكيف لموافقة تغيرات توجه البيانات. لكن لهذه المرونة ثمنها، فمن المهم عدم المبالغة في إضافة متغيرات إلى النموذج، تجنبا للملائمة الزائدة (overfitting).

المثال التالي يقارن بين دالة خطية وأخرى تكعيبية، ويبين كيف تحسن النموذج (قدرة تفسيرية أكبر) باستخدام الدالة متعددة الحدود من الدرجة الثالثة سواء من خلال الرسم (تتبع المنحنى لسحابة النقاط) أو قيمة معامل التحديد. يظهر المثال أيضا الاختلاف بين الدالتين في التنبؤ، فبينما تتوقع الدالة الخطية من الدرجة الأولى زيادة في المبيعات تتوقع الدالة متعدد الحدود إنخفاضا.

المفاضلة بين النموذجين لا تعتمد حصريا على R^2 ومعايير إحصائية بحتة، وإنما أيضا على الدراية بواقع المتغيرة والتوقعات لمسارها المستقبلي. مثلا، قد يفضل المحلل النموذج الخطي رغم أن قدرته التفسيرية اقل إذا كان يتوقع أن المتغيرة متجهة صعودا في المستقبل رغم الانخفاض الذي شهدته في الفترات الأخيرة...



صورة 4. إستخدام دالة متعددة الحدود تكعيبية أعطى قدرة تفسيرية أكبر ($R^2 = 0.76$) مقارنة مع النموذج الخطي. التوقع حسب الدالة متعددة الحدود هو الانخفاض عكس الدالة الخطية. لاحظ في النافذة اختيار Polynomial مع درجة order = 3، والتوقع لفترة واحدة (Forward = 1) مع إظهار الدالة وR^2 (Display Equasion on chart, Display R-squared)

4-2. الدالة متعددة الحدود في *R*

لاستخراج دالة متعددة الحدود نستخدم الدالة ()lm مع تحديد المتغيرة التابعة والمتغيرات المستقلة المدرجة، وجدول البيانات. نسمي النموذج في السطر نفسه، ثم في السطر الموالي نطلب إحصائيات النموذج من خلال ()Summary.

نستخدم هنا مثالا لبيانات مشهورة من R وهي mtcars رغم أنها لا تمثل سلسلة زمنية، لأن ذلك لا يغير شيئا في الطريقة ولا في المخرجات. المتغيرتين هما mpg أي عدد الأميال مقابل كل قالون من البنزين، والمتغيرة المستقلة هي wt أي الوزن بالطن.

```
fit2 <- Lm(mpg ~ wt + I(wt^2), data = mtcars)
Summary(fit2)
```

للحصول على دالة تكعيبية ندرج المتغيرة الإضافية بنفس الطريقة:

```
fit3 <- Lm(mpg ~ wt + I(wt^2) + I(wt^3), data = mtcars)
Summary(fit3)
```

يمكن أيضا كتابة الأمر كما يلي:

```
Fit2 <- Lm(mpg ~ poly(wt,2), data = mtcars)
Summary(fit3)
Call:
lm(formula = mpg ~ poly(wt, 2), data = mtcars)

Residuals:
   Min      1Q Median      3Q     Max
-3.483 -1.998 -0.773  1.462  6.238
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)    20.0906     0.4686  42.877  < 2e-16***
poly(wt, 2)1 -29.1157     2.6506 -10.985 7.52e-12***
poly(wt, 2)2   8.6358     2.6506   3.258  0.00286 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.651 on 29 degrees of freedom
Multiple R-squared:  0.8191,  Adjusted R-squared:  0.8066
F-statistic: 65.64 on 2 and 29 DF,  p-value: 1.715e-11
```

تعطي الدالة summary() مخرجات كثيرة:

1- في جدول المتبقي (Residuals) نجد القيمتين الدنيا والقصوى للمتبقي والربيعي الأول والثاني والثالث. نلاحظ أن الوسيط هنا قريب من الصفر والقيم منتشرة حوله بمسافة متقاربة من الأعلى والأسفل، وهو المطلوب.

2- في جدول المعاملات (coefficients) نجد معاملات النموذج مما يسمح بكتابة الدالة كما يلي:

$$\hat{y} = 20.09-29.1157t+8.6358t^2$$

يعطي الجدول ذاته اختبار المعاملات وهي هنا دالة إحصائيا (مستوى المعنوية الفعلي أقل من 0.05).

3- قمية الخطأ المعياري للمتبقي، وتقارن عادة مع نماذج أخرى إن وجدت.

4- معامل التحديد (R^2) وهي هنا قيمة عالية تدل على قوة العلاقة الخطية، وبالتالي جودة النموذج.

5- قيمة F وهي تفيد في اختبار النموذج، وهي هنا دالة إحصائية مما يعني أنه أفضل من استخدام المتوسط (لأن مستوى دلالة F أقل من 5 بالمائة).

5-2. دوال غير خطية

الدالة الأسية

يناسب التعديل الأسي البيانات التي تتزايد أو تتناقص بسرعة متزايدة لكن التزايد أو التناقص هو بنسبة ثابتة (لا بقيمة ثابتة). الدالة هي من الشكل:

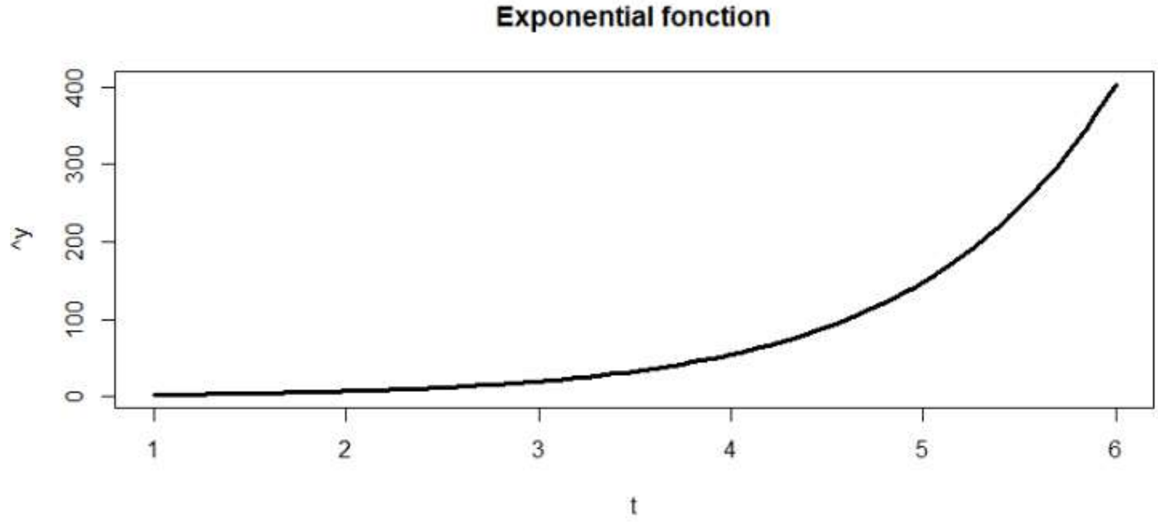
$$\hat{y} = cg^t$$

حيث c و g هي المعالم التي يتعين تقديرها. تقدير هذه المعالم يمكن تسهيله بالتحويل اللغرتمي الطبيعي، للحصول على دالة خطية بسيطة تعطي لغرتم التابع Wing (1974):

$$\ln(\hat{y}) = \ln(c) + \ln(g^t) = \ln(c) + t\ln(g)$$
$$\ln(\hat{y}) = C + Gt$$

حيث C = ln(c) و G = ln(g) .

يكفي إذن أن نحسب C و G بطريقة المربعات الصغرى واستخدامهما لتوقع ln(^y) ثم نزع اللغرتم للحصول على ^y. الرسم التالي يظهر مثالا عن الدالة الأسية حيث التوجه صاعد.



دالة القوة

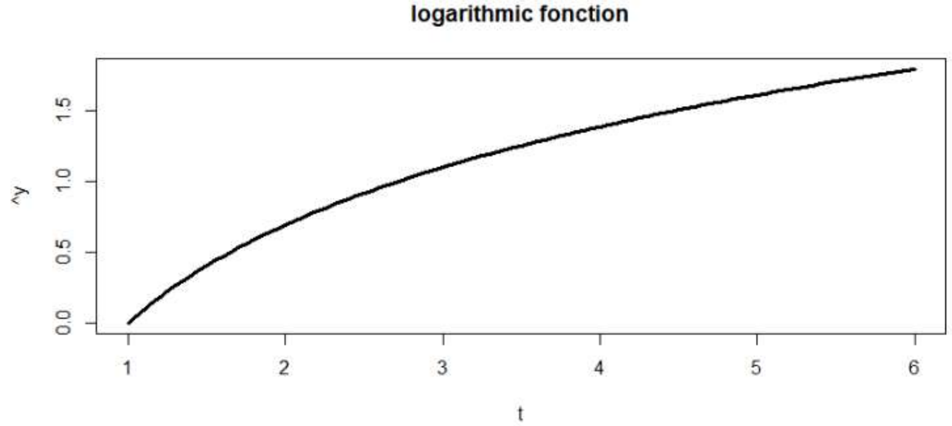
تعدد الاشكال التي يمكن أن تأخذها دالة القوة puissance وتباينها الكبير جعل منها مرشحا جيدا لتمثيل البيانات في العلوم الاقتصادية وميادين أخرى. يحتاج الأمر لأن يكون المحلل على دراية بالأشكال التي تأخذها الدالة في الحالات المختلفة (القوة موجبة، القوة سالبة، تنوع شكل الدالة من أجل القيم المختلفة للقوة a، ...).

$$\hat{y} = ct^a$$

الدالة اللغرتمية

يكون التعديل اللغرتمي ملائما للبيانات عندما تكون البيانات متناقصة أو متزايدة بحدة ثم تميل للاستقرار (بدون أن تغيير الاتجاه، على عكس دالة القوة، أنظر الرسم).

$$\hat{y} = b + a \times ln(t)$$



6-2. التحويل

تحتاج نماذج التنبؤ أحيانا إلى تحقق شروط معينة لكي تعطي تقديرات متقاربة أو دقيقة أو غير متحيزة. عندما لا تتحقق هذه الشروط نلجأ إلى تغيير البيانات بتحويل المتغيرة y (مثلا إلى لغرتم y) على أمل أن هذه المتغيرة الجديدة تحقق الشروط. نستخدم إذا نموذج التنبؤ على المتغيرة المحولة وبعد حساب القيم المتنبأ بها للمتغيرة المحولة نعود ونعكس التحول (back-transform) لنستخرج التنبؤ للمتغيرة الأصلية.

كذلك كثير من الاختبارات الإحصائية مثل اختبار ستيودنت واختبار ANOVA تحتاج إلى أن تكون المتغيرة تتبع التوزيع الطبيعي. عندما تكون المتغيرة بعيدة عن الطبيعية يمكن أن يكون التحويل transformation مفيدا في الحصول على الطبيعية. يستخدم التحويل أيضا أحيانا للسلسلة كحل للتوجه غير الخطي، أو من أجل حل مشكلة التباين غير المنسجم كما قد يستخدم للحصول على استقرار للموسمية. فيما يلي نعرض باختصار نوعين من التحويل مستخدمين بكثرة في السلاسل الزمنية هما تحويل بوكس –كوكس والتحويل اللغرتمي.

التحويل اللغرتمي

يستخدم التحويل اللغرتمي مثلا للسلسلة أحيانا كحل لعدم انسجام التباين أو كحل لعدم خطية العلاقة من أجل تحويلها إلى علاقة خطية، حيث يسمح التحويل اللغرتمي بتحويل التوجه الاسي إلى توجه خطي.

$$z_\lambda(y) = \ln(y)$$

ميزة التعديل اللغرتمي ل y أن القيمة المحولة z تبقى لها علاقة واضحة ب y، وهي أن **انتقال z يساوي تقريبا نسبة انتقال y** ونكتب:

$$z_{t+1} - z_t \approx \frac{y_{t+1}}{y_t}$$

تحويل بوكس – كوكس

يساعد تحويل بوكس–كوكس (George Roy Box and David Roxbee Cox, 1964) في تحويل بيانات غير طبيعية (لا تتبع التوزيع الطبيعي)، مثلا ممتدة من اليمين أو من اليسار، إلى بيانات تقترب من التوزيع الطبيعي. يستخدم هذا التحويل أيضا للحصول على استقرار التباين، حيث يسمح تحويل بوكس كوكس بتخفيض التباين وجعل تذبذب المتغيرة متماثلا أكثر حول المتوسط. استقرار التباين خاصية مطلوبة أحيانا في الاختبارات الإحصائية مثلا. يستخدم تحويل بوكس كوكس خاصة في السلاسل الزمنية للحصول على سلسلة 'مستقرة'، أي لها متوسط وتباين ثابتين، وهذه الخاصية مطلوبة في نماذج التنبؤ، مثلا عند العمل على استخراج نموذج ARIMA، حيث يتم تطبيق التحويل بوكس–كوكس على السلسلة

المفرقة إذا لم يكفي التفريق للحصول على سلسلة مستقرة. عيب تحويل بوكس كوكس أنه يجعل النموذج معقدا للتفسير، على عكس التحويل اللغرتمي مثلا.

يكتب تحويل بوكس–كوكس للبيانات الموجبة[1] كما يلي:

$$z_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & if\ \lambda \neq 0 \\ ln(y) & if\ \lambda = 0 \end{cases}$$

المعلمة Lambda هي عدد حقيقي يتراوح بين -5 و5، ويتم تقديرها من خلال البرمجيات (غالبا بطريقة المعقولية العظمى) حيث تسمح بحساب قيمة لامدا التي تجعل البيانات تقترب من الطبيعية. القيم الموجبة ل لامبدا تضخم القيم الكبيرة للمتغيرة والقيم السالبة تصغرها. كما تدل الدالة فإن التعديل اللغرتمي هو حالة خاصة من تعديل بوكس– كوكس أين لامدا تساوي 0.

في حالة لامدا تساوي 1 فإن التحويل هو طرح واحد من y، وهذا يخفض السلسلة للأسفل لكن لا يغير شكلها، أما للقيم الأخرى للامدا فإن التحويل يغير شكل المنحنى. تسمح البرمجيات بحساب القيمة المثلى للامدا. مثلا في R نستخدم الدالة: BoxCox.lambda()

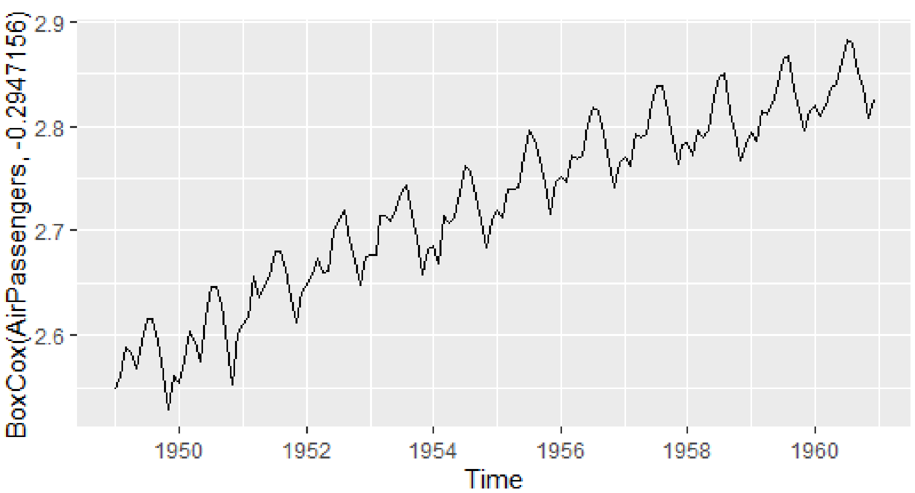
مثال: نعلم أن الموسمية في بيانات AirPassengers غير ثابتة فهي تزداد حدة مع الوقت. للحصول على موسمية مستقرة نستخرج قيمة لامدا:

```
BoxCox.lambda(AirPassengers)
[1] -0.2947156
```

نقوم بالتمثيل البياني للتحقق من جودة قيمة لامدا، نستخدم قيمة هذه الأخيرة في الدالة:

```
autoplot(BoxCox(AirPassengers,-0.2947156))
```

من الواضح أن قيمة تحويل بوكس–كوكس سمح بتثبيت الموسمية، مقارنة مع ما كانت عليه قبل التحويل.



[1] اقترح بوكس وكوكس أيضا صيغة أخرى للبيانات السالبة.

3. خلاصة

التوجه هو مكون لا تكاد تخلو منه سلسلة زمنية خاصة السلاسل التي تمتد للمدى المتوسط والبعيد، لذلك فإن فهم السلسلة الزمنية لا يتأتى بدون فهم هذا المكون الأساسي من مكوناتها.

تطرقنا لكيفية نمذجة التوجه الخطي بمعادلة خطية بطريقة المربعات الصغرى. معاملات الدالة تعطي نظرة عن التوجه الموجود قوته واتجاهه كما تسمح بالتنبؤ لفترات قريبة أو متوسطة المدى. بينا في هذا الفصل إضافة لما سبق بعض الطرق لاختصار الحسابات التي تتطلبها الطريقة، مثل طريقة ماير. كما شرحنا كيفية اختبار التوجه إحصائيا باستخدام اختبار دانيال، وطريقة تستخدم عندما يكون غير واضح ما إذا كانت السلسلة تتضمن مكون التوجه.

يتعرض الفصل أيضا لتقييم النموذج باستخدام معامل التحديد والذي يعطي نسبة التباين المفسر بالنموذج، كما ذكرنا أهمية الرسم البياني لسحابة النقاط في كشف نوع العلاقة إن وجدت وقوتها واتجاهها، وكيفية تفسير الرسم، ومتى يدل على علاقة خطية قوية.

عمليا يتطلب استخراج الدالة حسابات طويلة نسبيا، خاصة إذا كانت البيانات غزيرة، لذلك تطرقنا لكيفية إجراء الحسابات والتمثيل البياني ببرنامجي Excel والذي يعطي إمكانيات متعددة ومفيدة متعلقة بتحلي الانحدار كما تطرقنا لكيفية إجراء الحساب باستخدام R.

في الأخير تطرقنا باختصار لأنواع من التوجه غير الخطي وكيفية تمثيله بدالة وبالرسم البياني.

4. ملحق: استخدام محلل البيانات للنموذج الجمعي

يمكن استخدام محلل البيانات للحصول على نموذج يربط Y بمكونيها: التوجه والموسمية. في حالة وجود موسمية ندرج المواسم (الثلاثيات مثلا) كمتغيرات وهمية (Dummy variables) ثنائية (تأخذ القيمة 0 أو 1)، **وعدد المتغيرات الوهمية يساوي عدد المواسم في النافذة ناقصا واحد، لأن أحد الأنماط يبقى كمرجع.**

في حالة وجود توجه مع الموسمية ندرج مع المتغيرات الوهمية متغيرة تمثل الزمن أي t. فيما يلي نفصل ما سبق من خلال مثالين؛ الأول يتضمن موسمية ولا يتضمن توجه والثاني يتضمن الاثنين.

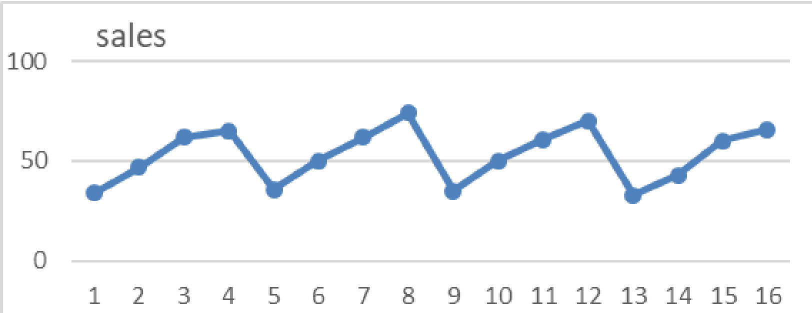
حالة وجود موسمية (مستقرة) مع عدم وجود توجه

مثال: قم بالتمثيل البياني لتفحص المكونات. هل هناك موسمية؟ ما طبيعتها ونافذتها؟ هل هناك توجه؟

Year	2001				2002				2003				2004			
Qter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Sales	34	47	62	65	36	50	62	74	35	50	61	70	33	43	60	66

- قم بإدراج متغيرات ثنائية للتعبير عن الثلاثي.
- استخدم محلل البيانات لاستخراج نموذج يربط y بالثلاثيات.
- فسر المعاملات.
- قم بالتنبؤ للثلاثيات الأربع المقبلة.

الحل. التمثيل البياني يبين وجود موسمية فقط ولا توجه.



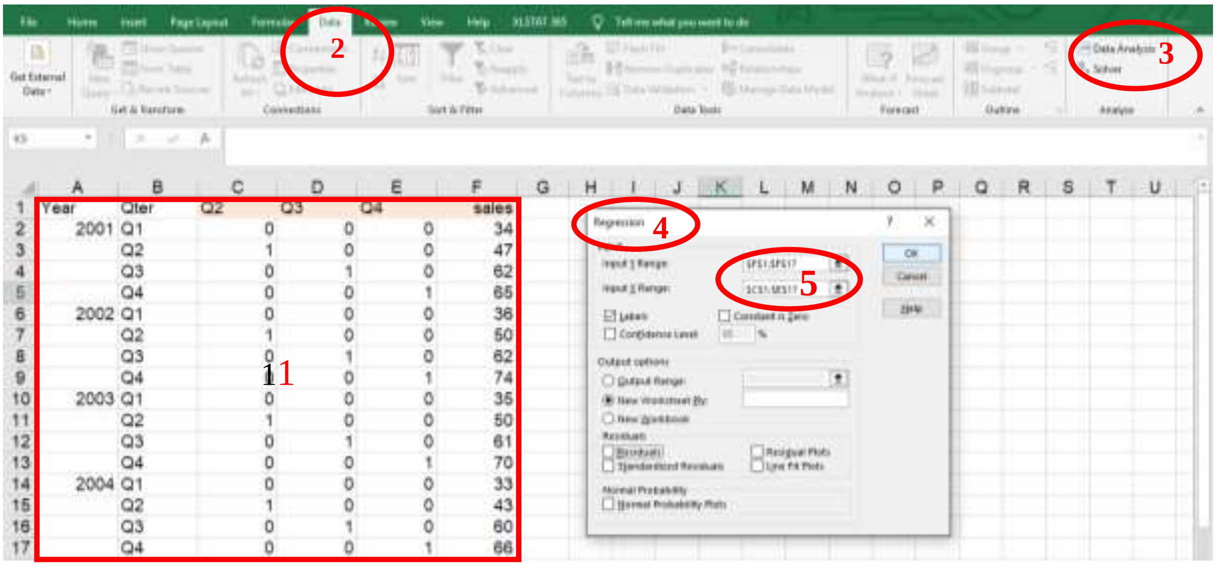
رسم توضيحي 2. سلسلة تتضمن موسمية مستقرة ولا توجه

التعبير عن الثلاثيات من خلال متغيرات ثنائية.

جدول 1 التعبير عن الثلاثي الذي تنتمي إليه كل مشاهدة y من خلال ثلاث متغيرات ثنائية.

Q2	Q3	Q4	Sales
0	0	0	34
1	0	0	47
0	1	0	62
0	0	1	65
0	0	0	36
1	0	0	50
0	1	0	62
0	0	1	74
0	0	0	35
1	0	0	50
0	1	0	61
0	0	1	70
0	0	0	33
1	0	0	43
0	1	0	60
0	0	1	66

استخدام محلل البيانات data analysis لتحليل انحدار Y على الثلاثيات.



صورة 5. التعبير بمتغيرات ثنائية عن الثلاثيات ثم طلب تحليل انحدار متعدد 'Regression' من محلل البيانات 'Data analysis'

بعد إدخال البيانات في المربعين المخصصين (F1 :F17) 'Input Y Range' و 'Input X range' (C1 :E17)

نعلم على Label لأننا ضمنا البيانات عناوين المتغيرات لكي تظهر في المخرجات. من المهم أيضا طلب خطأ التنبؤ بالتأشير على 'Residuals'. يمكن من خلال التمثيل البياني للخطأ معرفة هل النموذج الخطي هو الأنسب؛ يظهر التوجه غير الخطي في نمط ما في سحابة النقاط، إذا كان التقدير جيدا فستأتي النقاط موزعة أفقيا حول خط الصفر، وبشكل عشوائي. أي نمط في توزيع نقاط سحابة الخطأ يعني أن النموذج بحاجة إلى تحسين.

تأتي المخرجات في ثلاث جداول كما يلي:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.983793
R Square	0.96785
Adjusted R Square	0.959812
Standard Error	2.76134
Observations	16

ANOVA

	df	SS	MS	F	Significance F
Regression	3	2754.5	918.1667	120.4153	0.0000
Residual	12	91.5	7.625		
Total	15	2846			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	34.5	1.38067	24.98787	0.0000	31.49177821	37.50822179
Q2	13	1.952562	6.657918	0.0000	8.745731951	17.25426805
Q3	26.75	1.952562	13.69995	0.0000	22.49573195	31.00426805
Q4	34.25	1.952562	17.54105	0.0000	29.99573195	38.50426805

يدل الجدول الأول أن العلاقة قوية بين المتغيرات المستقلة الثلاث $Q_1.Q_2.Q_3$ من جهة والمتغيرة التابعة Y. هذه العلاقة تفسر حوالي 97 بالمئة من تباين Y، وهي نسبة عالية جدا.

يظهر الجدول الثاني أن هذه العلاقة دالة إحصائيا عند مستوى معنوية 5 بالمئة (sig. <0.05)، أي أنها ليست عائدة لمحض الصدفة.

الجدول الثالث تظهر فيه الدالة التي تربط Y بالثلاثيات الأربع وهي كما يلي:

$$\hat{Y} = 34.5 + 13\,(Q2) + 26.75(Q3) + 34.25(Q4)$$

تفسير المعاملات هو كالآتي:

- 34.5 هو الثابت ويمثل متوسط المبيعات في الثلاثي الأول.
- 13 تمثل الزيادة في المبيعات في الثلاثي الثاني مقارنة مع الثلاثي الأول.
- 26.75 تمثل الزيادة في المبيعات في الثلاثي الثالث مقارنة مع الثلاثي الأول.
- 34.25 تمثل الزيادة في المبيعات في الثلاثي الرابع مقارنة مع الثلاثي الأول.

لاحظ أن الدالة لا تتضمن المتغيرة t، فقيمة المبيعات لا تتوقف على الفترة بما انه لا يوجد توجه، وإنما تتوقف فقط على الثلاثي. هذا يعني ان التوقعات تختلف بين ثلاثيات العام الواحد لكنها تتكرر بين السنوات.

للتنبؤ يمكن استخدام الدالة؛ نعوض قيمة الثلاثي المراد التنبؤ له في الدالة. للثلاثي الأول القيمة المتوقعة هي دائما قيمة الثابت.

التوقع للثلاثي الأول:

$$\hat{Y}(17) = 34.5 + 13\,(0) + 26.75(0) + 34.25(0) = 34.5$$

التوقع للثلاثي الثاني:

$$\hat{Y}(18) = 34.5 + 13\,(1) + 26.75(0) + 34.25(0) = 47.5$$

التوقع للثلاثي الثالث:

$$\hat{Y}(19) = 34.5 + 13\,(0) + 26.75(1) + 34.25(0) = 61.25$$

التوقع للثلاثي الرابع:

$$\hat{Y}(20) = 34.5 + 13\,(0) + 26.75(0) + 34.25(1) = 68.75$$

هذه التوقعات لا تتغير أيا كانت السنة، لأننا افترضنا عدم وجود توجه.

حالة وجود موسمية (مستقرة) مع وجود توجه

في هذه الحالة Y ليست مستقلة عن الزمن، لذلك جدول البيانات في هذه الحالة تضاف إليه المتغيرة t. الخطوات الأخرى تبقى تقريبا بدون تغيير ما عدا إدراج t مع المتغيرات المستقلة. اما تفسير المعاملات فيتغير. لشرح هذه الحالة نأخذ المثال التالي.

مثال2. قم بالتمثيل البياني لمبيعات منتج المؤسسة وحدد ما هي المكونات وعلق عليها.

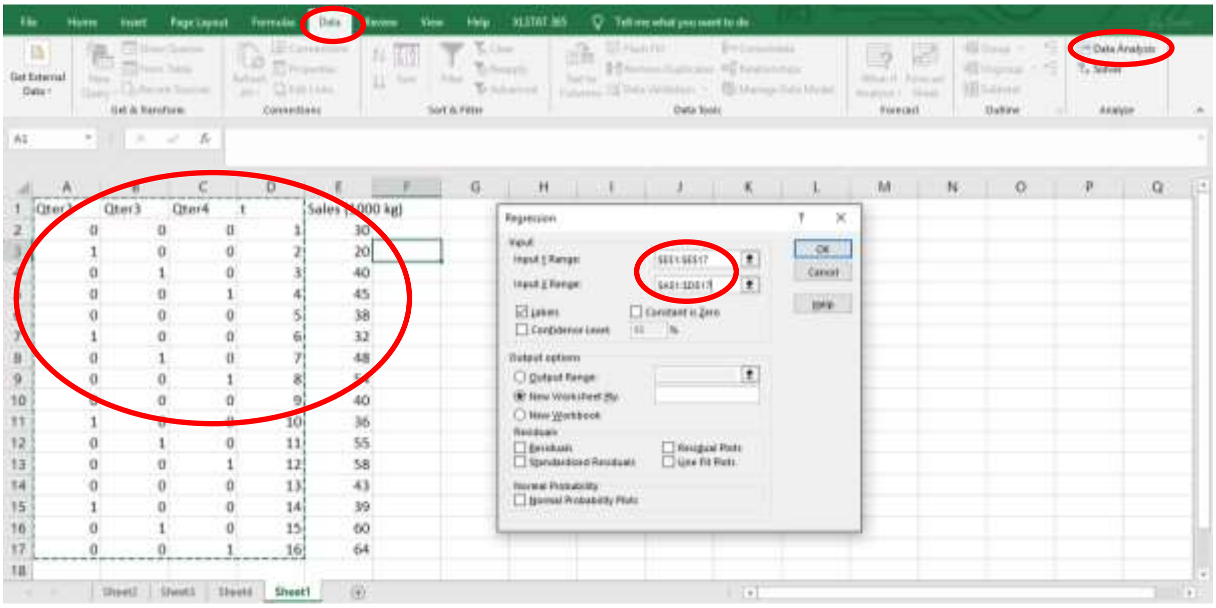
Year	1				2				3				4			
Qter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Sales	30	20	40	45	38	32	48	54	40	36	55	58	43	39	60	64

- استخدم محلل البيانات لاستخراج نموذج يسمح بإبراز تأثر المبيعات بالزمن t.
- استخدم محلل البيانات لاستخراج نموذج يسمح بإبراز تأثر المبيعات بالزمن وبالثلاثي من خلال إدخال متغيرات ثنائية وهمية تمثل الثلاثي.
- قم بالتنبؤ للثلاثيات الأربع المقبلة.

الحل.

نقوم بصياغة ثلاث متغيرات وهمية ثنائية للتعبير عن الثلاثيات الثاني والثالث والرابع (تأخذ 0 أو 1 حسب كون المشاهدة تنتمي أو لا للثلاثي) الثلاثي الأول يترك كمرجع.

بعد ذلك نطلب محلل البيانات من Data وندخل في مربع الحوار بيانات التابع وبيانات المتغيرات المستقلة t و Qter2، Qter3، Qter4. الصورة التالية تبين كيفية إنشاء المتغيرات الوهمية وطلب تحليل الانحدار 'Regression' من محلل البيانات ومن تعبئة البيانات.



صورة **6**. إدخال البيانات في Excel وطلب تحليل انحدار متعدد باستخدام محلل البيانات data analysis من قائمة البيانات data.

تأتي المخرجات في ورقة خاصة كما يلي:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.986564861
R Square	0.973310225
Adjusted R Square	0.963604853
Standard Error	2.291287847
Observations	16

ANOVA

	df	SS	MS	F	Significance F
Regression	4	2106	526.5	100.2857	0.0000
Residual	11	57.75	5.25		
Total	15	2163.75			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	27.6875	1.454787356	19.03199	0.00000	24.4855346	30.8894654
Qter2	-7.4375	1.625240367	-4.57625	0.00080	-11.0146299	-3.86037007
Qter3	10.125	1.64031247	6.172604	0.00007	6.5146966	13.7353034
Qter4	13.1875	1.665129499	7.919804	0.00001	9.52257468	16.8524253
t	1.4375	0.128086885	11.22285	0.00000	1.15558267	1.71941733

الجدول الأول يعطي نسبة التباين المفسر بالنموذج، أي ب t والثلاثيات، وهي هنا نسبة عالية 0.97. النسبة أصبحت أعلى في هذا النموذج لأنه يحتسب مكون إضافي هو التوجه.

الجدول الثاني يبين أن العلاقة دالة إحصائيا (sig. <0.05)، أي أنه يمكن الاستدلال على وجود علاقة بين المتغيرات المستقلة من جهة والمتغيرة التابعة.

الجدول الثالث يبين النموذج، وهو كما يلي:

$$\hat{Sales} = 27.6875 - 7.4375\,Qter2 + 10.125Qter3 + 13.1875Qter4 + 1.4375t$$

معامل t يبين تأثير مكون التوجه بمعزل عن الموسمية (زيادة ب 1.437 طن في المتوسط كل ثلاثي)، ومعاملات الثلاثيات تبين تأثير كل ثلاثي (متوسط الزيادة مقارنة مع الثلاثي المستخدم كمرجع أي الأول). الإشارة الموجبة في الثلاثيين الثالث والرابع تعني أن مبيعاتهما أكبر من متوسط مبيعات الثلاثي الأول الذي استخدم كمرجع. قيمة كل معامل تمثل متوسط الفرق. الإشارة السالبة لمعامل الثلاثي الثاني تعني أن مبيعاته أقل من الثلاثي المرجع (الأول).

للتنبؤ نعوض في الدالة أعلاه[1].

$$\hat{Sales}_{17} = 27.6875 - 7.4375\,(0) + 10.125(0) + 13.1875(0) + 1.4375(17) = 52.125$$
$$\hat{Sales}_{18} = 27.6875 - 7.4375\,(1) + 10.125(0) + 13.1875(0) + 1.4375(18) = 46.125$$
$$\hat{Sales}_{19} = 27.6875 - 7.4375\,(0) + 10.125(1) + 13.1875(0) + 1.4375(19) = 65.125$$
$$\hat{Sales}_{20} = 27.6875 - 7.4375\,(0) + 10.125(0) + 13.1875(1) + 1.4375(20) = 69.625$$

[1] يمكن تسهيل حساب التوقع باستخدام الدوال في Excel. ننقل الثابت والمعاملات، ثم ننقل الثلاثيات 1، 2، 3، 4 بالإضافة إلى t المتنبأ لها في سطر آخر. بعد ذلك نستخدم الدالة: sumproduct= تحتاج هذه الدالة إلى مدخلين: مجال خلايا المعاملات والثابت، ومجال خلايا الثلاثيات و t.

5. سلسلة تمارين

5-1. التمارين

تمرين 1. مراجعة نظرية

- أكتب صيغة ^y في التوجه الخطي (دالة التوجه).
- أكتب صيغة معاملي دالة التوجه بطريقة المربعات الصغرى.
- كيف يفسر الميل في الدالة؟ وماذا يمثل الثابت؟
- كيف تستخدم الدالة للتنبؤ؟
- كيف يتم تقييم مدى جودة تمثيل دالة التوجه للبيانات؟ أكتب صيغة R^2، كيف يفسر؟

تمرين 2. عدد العمال في صناعة النسيج (ثالثا) – دالة التوجه وتقييمها واستخدامها

في بيانات التمرين الأسبق حول عدد العمال في قطاع صناعة الملابس في فرنسا بين 1981 و1988.

- أحسب معاملات دالة التوجه الخطي واكتب الدالة (حول الزمن إلى سلسلة من 1 إلى 8).
- قيم جودة تمثيل الدالة للبيانات باستخدام معامل التحديد.
- قدر باستخدام الدالة عدد العاملين لسنوات 1990 و1991و 1992.

تمرين 3. اختبار دانيال للتوجه

لديك استهلاك الطاقة بالمليون لمؤسسة ما خلال 19 شهرا. اختبر وجود توجه باستخدام اختبار سبيرمان.

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
yt	75	85	83,75	77,5	75	81,25	83,75	92,5	95	88,75	87,5	85	75	85	90	90	91,25	102,5	97,5

تمرين 4. مبيعات الأبقار – التوجه متعدد الحدود (Excel)

لديك مبيعات مزرعة تربية الأبقار خلال الأشهر الماضية.

- مثل البيانات في سحابة نقاط. هل تصلح الدالة الخطية لتمثيل البيانات؟
- أضف خط توجه خطي at+b لسحابة النقاط وعلق على مدى تمثيله للبيانات.
- استخرج معامل التحديد وعلق على قيمته.
- أضف إلى الرسم خط توجه من النوع متعدد الحدود (polynomial) من الدرجة 2 مع إظهار الدالة وR^2. قارن مع نتائج التوجه الخطي. قم بتمديد المنحنى للتنبؤ لثلاث أشهر.

- استخدم دالة التوجه متعدد الحدود لحساب القيمة المتوقعة للأشهر الثلاث المقبلة.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
المبيعات	6	10	7	17	12	20	21	18	22	23	14	21	20	12	14

تمرين 5. حوادث المرور – سحابة الخطأ ودالة التوجه التربيعي

لديك عدد حوادث المرور المسجلة في طريق معين على مدى 13 شهرا.

- استخرج دالة التوجه الخطي وثم بتمثيلها مع سحابة السلسلة. علق.
- استخرج قيم الخطأ $E = (y - \hat{y})$ عن دالة التوجه الخطي ومثلها بيانيا.
- هل تلاحظ نمطا ما في سحابة المتبقي؟ هل تنتشر السحابة في شكل أفقي أم لها توجه ما؟ هل تنتشر في شكل مثلث، هل هناك قيم شاردة؟
- كرر المطلبين السابقين مع دالة التوجه التربيعية.
- استخدم الدالتين لحساب العدد المتوقع للحوادث خلال الشهرين الموالين (14 و15).

الشهر	1	2	3	4	5	6	7	8	9	10	11	12	13
عدد الحوادث	27	25	20	15	21	20	12	13	15	14	19	20	22

تمرين 6. عدد المواليد في مقاطعة ألمانية (من الفصل 1) – إختيار دالة التوجه

لديك البيانات التالية لعدد المواليد في مقاطعة ألمانية خلال الفترة من 2004 إلى 2011.

- أدخل البيانات إلى Excel ومثلها بيانيا ثم علق على المكونات التي يظهرها.
- استخراج دالة غير خطية لتمثيل البيانات. قارن بين الخيارات من خلال الرسم ومعامل التحديد.

السنة / الثلاثي	2004	2005	2006	2007	2008	2009	2010	2011
1	7 684	7 437	7 311	7 221	7 148	7 105	7 067	7 062
2	7 899	7 705	7 616	7 471	7 336	7 189	7 146	7 128
3	7 320	7 208	7 093	7 008	6 970	7 043	6 983	7 008
4	7 683	7 450	7 298	7 184	7 231	7 206	7 185	7088

تمرين 7. (دراسة حالة): التوجه الأسي – أعداد الطلبة في الجامعات الجزائرية

لديك البيانات التالية لعدد خريجي الجامعة الجزائرية (ما عدا الجامعة الليلية) منذ الاستقلال.

- حمل البيانات في ملف Excel ثم قم بالتمثيل البياني واستكشاف نوع التوجه الملاحظ.
- استخرج دالة توجه (اختر الدالة الأنسب) ومثلها بيانيا على الرسم، مع إظهار R^2.
- استخدم الدالة للتوقع لعدد الخريجين في السنة الجامعية 2018–2019. حاول الحصول على الرقم الحقيقي وقارن.
- قم بحساب التفريق بدرجة واحدة ومثله بيانيا ثم علق.

جدول 2. تطور عدد الخريجين من التعليم العالي من 63–1962 إلى 2011–2010 (ما عدا الجامعة الليلية).

Années Scolaires	*Diplômés*	*Années Scolaires*	*Diplômés*
1962-1963	...	**1987-1988**	18 110
1963-1964	180	**1988-1989**	20 493
1964-1965	179	**1989-1990**	22 917
1965-1966	195	**1990-1991**	25 582
1966-1967	378	**1991-1992**	28 182
1967-1968	654	**1992-1993**	29 336
1968-1969	724	**1993-1994**	29 341
1969-1970	817	**1994-1995**	31 970
1970-1971	1 244	**1995-1996**	35 671
1971-1972	1 703	**1996-1997**	37 323
1972-1973	2 355	**1997-1998**	39 521
1973-1974	2 786	**1998-1999**	44 531
1974-1975	2 844	**1999-2000**	52 804
1975-1976	4 661	**2000-2001**	65 192
1976-1977	5 410	**2001-2002**	72 737
1977-1978	5 928	**2002-2003**	77 972
1978-1979	6 046	**2003-2004**	91 828
1979-1980	6 963	**2004-2005**	107 515
1980-1981	7 477	**2005-2006**	112 932
1981-1982	7 800	**2006-2007**	121 905
1982-1983	9 584	**2007-2008**	146 889
1983-1984	10 237	**2008-2009**	150 014
1984-1985	11 713	**2009-2010**	199 767
1985-1986	14 097	**2010-2011**	246 743
1986-1987	16 645		

Source : Ministère de l'Enseignement Supérieur et de la Recherche Scientifique - In Publications : Annuaire Statist
Formation Supérieure en chiffres et Bilan chiffré de la formation supérieure

تمرين 8. (دراسة حالة. بورصة الجزائر: تمرين غير محلول)

لديك البيانات التالية لمؤشر بورصة الجزائر من سبتمبر إلى غاية 17 نوفمبر.

1. مثل البيانات وعلق على المكونات الملاحظة.
2. استخرج دالة التوجه ومثلها بيانيا مع إظهار R^2.
3. استخدم الدالة لتوقع قيمة المؤشر بعد شهر.
4. أدخل إلى موقع بورصة الجزائر واحصل على بيانات سنة 2019 كاملة، واستخرج الدالة وقيم تمثيلها للبيانات. قارن مع الدالة المحسوبة على أساس بيانات 3 أشهر.

سبتمبر		أكتوبر		نوفمبر	
2019-09-02	1527,302	2019-10-01	1533,659	2019-11-03	1551,811
2019-09-03	1527,302	2019-10-03	1533,659	2019-11-05	1543,24
2019-09-05	1527,302	2019-10-06	1533,457	2019-11-07	1545,058
2019-09-08	1527,302	2019-10-08	1534,315	2019-11-10	1545,058
2019-09-11	1527,302	2019-10-10	1534,315	2019-11-12	1545,058
2019-09-12	1527,302	2019-10-13	1534,315	2019-11-14	1541,076
2019-09-15	1530,731	2019-10-15	1535,223	2019-11-17	1541,076
2019-09-17	1530,731	2019-10-17	1535,526		
2019-09-19	1530,932	2019-10-20	1535,526		
2019-09-22	1530,932	2019-10-22	1535,526		
2019-09-24	1530,932	2019-10-24	1535,526		
2019-09-26	1530,932	2019-10-27	1535,526		
2019-09-29	1540,678	2019-10-29	1535,526		
		2019-10-31	1551,811		

تمرين 9. أسئلة للتوسع

- أذكر بعض أنواع التوجه غير الخطي ومتى تستخدم.
- تفترض طريقة المربعات الصغرى استقلال المشاهدات بينما نعلم أن مشاهدات السلسلة الزمنية قد لا تكون مستقلة (مبيعات موسم تتأثر بمبيعات الموسم أو المواسم السابقة ...)، فهل يعني ذلك أن الطريقة غير صالحة إذا لم يتحقق هذا الافتراض؟ أم أن المقدرات تبقى غير متحيزة في حالة وجود ارتباط ذاتي، والذي يتأثر هو الخطأ المعياري؟
- إذا جاءت البيانات محدبة قليلا وليست خطية، فكيف يمكن تمثيل التوجه العام؟ هل نستخدم دالة متعددة الحدود أم نحول البيانات باستخدام اللغرتم ثم نستخرج دالة خطية لهذا الأخير[1]، أم أن كلا الطريقتين صالحتين وتتم المفاضلة بينهما على أساس معامل التحديد.

تمرين 10. استخدام محلل البيانات مع سلسلة موسمية بدون توجه

لديك البيانات التالية للمبيعات الثلاثية خلال 4 سنوات.

- قم بإدخال بيانات المبيعات الثلاثية التالية إلى ورقة Excel،
- من خلال التمثيل البياني تحقق من وجود موسمية،
- استخدم محلل البيانات Data analysis لصياغة دالة تربط المبيعات بالثلاثيات،
- فسر معاملات الدالة،
- أحسب المبيعات الثلاثية المتوقعة للسنة المقبلة.

Year	Qter	t	Sales	Year	Qter	t	Sales
2001	Q1	1	101	2003	Q1	9	93

[1] نفترض أن الدالة هي $z = ae^{bt}$ وهذا يعني أن $\ln(z) = \ln(a) + bt = a_0 + bt$.

	Q2	2	112		Q2	10	104
	Q3	3	117		Q3	11	127
	Q4	4	134		Q4	12	138
2002	Q1	5	105	2004	Q1	13	96
	Q2	6	117		Q2	14	114
	Q3	7	127		Q3	15	116
	Q4	8	124		Q4	16	140

تمرين 11. استخدام محلل البيانات مع سلسلة بها موسمية وتوجه

لديك البيانات التالية للمبيعات الثلاثية خلال 4 سنوات.

- قم بإدخال بيانات المبيعات الثلاثية التالية إلى ورقة Excel،
- من خلال التمثيل البياني تحقق من وجود موسمية وتوجه،
- استخدم محلل البيانات Data analysis لصياغة دالة تربط المبيعات بكل من الزمن والثلاثيات،
- فسر معاملات الدالة،
- أحسب المبيعات الثلاثية المتوقعة للسنة المقبلة.

Year	Qter	Sales	2003	Q1	58
2001	Q1	23		Q2	45
	Q2	46		Q3	76
	Q3	91		Q4	166
	Q4	106	2004	Q1	65
2002	Q1	41		Q2	90
	Q2	58		Q3	184
	Q3	76		Q4	158
	Q4	125	2003	Q1	58

2-5. الحلول

6. مراجع الفصل

Anderson, S. W. (2007). *Statistiques pour l'économie et la gestion* (éd. 2). (A. David R., W. Dennis J., & A. Thomas A., Trads.) Bruxelles: De Booeck.

Boulahlib, S. (2018, 05 26). Année universitaire 2018-2019 : 2 millions d'étudiants sur les bancs de l'université. Algeria. Retrieved 11 10, 2019, from https://www.alg24.net

Droesbeck, J. J. (1997). *Eléments de statistique.* Belgique: Ellips.

Malhotra, N., Décaudin, J.-M., & Bouguerra, A. (2007). *Etude Marketing avec SPSS* (éd. 5). Paris: Pearson.

Meyer, J.-B. (2019, 3 11). Les étudiants, clé du changement en Algérie. *Le Monde*.