مقياس

تحليل السلاسل الزمنية

حل السلسلة الأولى

الأستاذ الدكتور

عماري زهير

اقتصاد تطبيقي

جامعة المسيلة

zohir.ammari@univ-msila.dz

التمرين 01

أجب على الأسئلة التالية:

- بماذا تتميز البيانات الزمنية عن البيانات التجريبية وبيانات المسح، مستشهدا بمثال يوضح ذلك؟
- أذكر أهم الحالات التي نلجأ إليها لتنظيف البيانات الخام وكيف يكون ذلك مستشهدا بمثال؟
- ماذا نقصد بخطأ التنبؤ؟ ماهي أهم طرق قياسه وأفضل طريقة ولماذا؟
- حدد معايير اختيار أسلوب التنبؤ؟
- أذكر أنواع المعادلات؟ حدد طرق التنبؤ وأساليبه؟
- متى نلجأ إلى التنبؤ باستخدام أسلوب تحليل السلاسل الزمنية؟
- ميز بين نماذج السلاسل الزمنية للتنبؤ؟

التمرين (02)

الجدول الآتي يوضح قيمة المبيعات السنوية من إحدى السلع بالملايين الدولارات خلال الفترة (2022/2014)

السنة	2014	2015	2016	2017	2018	2019	2020	2021	2022
قيمة المبيعات	9	11	10	12	11	9	13	11	9

- استخدم طريقة المتوسطات المتحركة البسيطة في إيجاد جميع التنبؤات الممكنة مرة باستخدام k=2 ومرة باستخدام k=3 ، وأوجد متوسط مربعات الأخطاء المناظرة في كل حالة؟

التمرين (03)

ليكن لديك الجدول التالي الذي يوضح عدد الأجهزة المباعة (بالمئة) التي سجلت شهريا في دفاتر إحدى الشركات

الشهر	1	2	3	4	5	6	7	8	9	10	11	12	13
النسبة	11	12	12	14	13	15	14	15	13	17	16	14	16

- قدر القيمة الابتدائية $\widehat{y_0}$ باستخدام الوسط الحسابي لقيم السلسلة ثم استخدم هذه القيمة لإيجاد التنبؤات المناظرة مرة باستخدام w=0.7 ومرة باستخدام w=0.9 أي التنبؤات أفضل؟ برر اجابتك؟

الإجابة على التمرين 01

تتميز البيانات الزمنية عن البيانات التجريبية وبيانات المسح، مستشهدا بمثال يوضح ذلك.

في أدبيات الإحصاء يمكن التعرف على ثلاث أنواع مختلفة من البيانات، وهي البيانات التجريبية وبيانات الحصر (المسح) والبيانات الزمنية. وتعتمد الفلسفة الخاصة **بالبيانات التجريبية** على الأسلوب التجريبي والذي يبدأ بتحديد العوامل الهامة والتي يعتقد الباحث أن لها تثير معنوي على الظاهرة أو المتغير موضع الدراسة، ثم يبدأ بالحصول على البيانات من خلال تصميم تجربة –تعتمد على مبدأ العشوائية– تسمح بقياس تأثير أحد أو بعض العوامل على الظاهرة او المتغير موضع الدراسة في ظل ثبات العوامل الأخرى، ويمكن توضيح ذلك من خلال المثال التالي:

لنفترض أن أحد الباحثين يريد معرفة تأثير ضوء الشمس على نمو النبات . في هذه التجربة، يتم تعريض مجموعة من النباتات –تعتمد على مبدأ العشوائية في الإختيار– لأشعة الشمس، ويتم إبقاء مجموعة أخرى في الظلام .وبعد شهر، يتم تسجيل ارتفاعات النباتات، ويتم استنتاج ما إذا كان ضوء الشمس (متغير واحد) ضروريًا لنمو النبات (الظاهرة).

وتعتمد الفلسفة الخاصة **ببيانات الحصر أو المسح او التعداد** على مبدأ الحصول على البيانات (وحدات المجتمع) عن طريق حصر أو مسح الوضع القائم للظواهر موضع الدراسة في نقطة زمنية واحدة فقط كما هو دون محاولة التحكم في العوامل المختلفة التي قد تؤدي إلى الحالة التي توجد عليها الظواهر.

أما **البيانات الزمنية** فيتم الحصول عليها من خلال رصد البيانات او القيم التي تعبر عن الظاهرة أو المتغير موضع الدراسة على فترات زمنية متتالية بهدف تحقيق عدة أهداف أهمها اكتشاف نمط التطور التاريخي للظاهرة أو المتغير موضع الدراسة وكيفية الاستفادة من هذا النمط في التنبؤ بهذه الظاهرة في المستقبل

وتختلف السلاسل الزمنية عن البيانات التجريبية وبيانات الحصر في ثلاث نقاط أساسية وهي:

1. تأخذ بيانات السلسلة الزمنية على فترة طويلة نسبيا يعتقد انها تؤثر على الظاهرة أو المتغير موضع الدراسة، بينما تأخذ البيانات التجريبية أو بيانات الحصر عند نقطة زمنية معينة أو على الأكثر في فترة زمنية قصيرة يعتقد انها لا تؤثر بشكل معنوي على الظاهرة أو المتغير موضع الدراسة وعادة ماتسمى هذه البيانات بالبيانات المقطعية (cross sectional data).
2. يتم دراسة السلسلة الزمنية عادة بمعزل عن العوامل الأخرى – بخلاف الزمن– التي قد تؤثر عليها وعن الظواهر الأخرى التي قد ترتبط معها في علاقة إحصائية.
3. عادة ما تكون بيانات أو مشاهدات السلسة الزمنية مرتبطة ببعضها البعض، ويأخذ الارتباط بين هذه المشاهدات أشكالا وأنماطا عديدة تختلف باختلاف طبيعة الظاهرة، ومن ثم فإن ترتيب المشاهدات في السلاسل الزمنية ذو أهمية خاصة ولذلك فإن معظم الأساليب التي تستخدم في تحليل البيانات التجريبية أو بيانات الحصر لا تكون صالحة لتحليل السلاسل الزمنية وبالتالي كان لابد من ابتكار وتطوير أدوات وأساليب خاصة لتحليل السلاسل الزمنية.

ذكر أهم الحالات التي نلجأ إليها لتنظيف البيانات الخام وكيف يكون ذلك مستشهدا بمثال

في كثير من الحالات تظهر البيانات آثار غير مرغوب فيها، وتعتبر الفروق بين أطوال الشهور مثالا على ذلك فقد يحتوي الشهر على أربعة أو خمسة إجازات (عطل) لنهاية الأسبوع مما يؤثر عادة على البيانات المشاهدة. كذلك فإن الأعياد والإجازات المتحركة زمنيا تساهم بدورها في إظهار مثل هذه الآثار، مثل عيد الفطر الذي يأتي إما في الربع الأول أو الثاني من السنة الميلادية، كل هذه الاضطرابات تؤدي إلى حدوث بعض التضليل والإنحراف من بيانات السلسلة.

وفي كثير من المواقف يتم تنظيف البيانات الخام قبل استخدامها في التحليل وفيما يلي بعض الأمثلة:

1. لتصحيح القصور الناتج من اختلاف أطوال الأشهر في بيانات الإنتاج الصناعي (مثلا) الشهرية، يمكن ضرب الإنتاج الشهري في 30 وقسمة الناتج على عدد أيام الشهر. بذلك يصبح الإنتاج مناظرا لشهر طوله 30 يوما، وبالتالي يمكن مقارنة البيانات مع بعضها البعض. مثلا يضرب انتاج شهر جانفي في 30 ويقسم على 31 لنحصل على إنتاج مناظر لثلاثين يوم.
2. يمكن أحيانا حذف أثر التغيرات قصيرة المدى بزيادة طول الفترة الزمنية، فمثلا، لإزالة الأثر الناتج من اختلاف مواعيد عيد الإفطار بالنسبة للسنة الميلادية يمكن جمع البيانات لفترات نصف سنوية (إذا ذلك لعلاج الموقف) بدلا من ربع سنوية.
3. البيانات المسجلة بالقيمة الاسمية للنقود يمكن قسمتها على بعض الأرقام القياسية التي تعكس التغيرات في قيمة النقود حتى نحصل على بيانات ثابتة القيمة، أي حتى نحصل على بيانات بالقيم الحقيقية.
4. في بعض الحالات الأخرى، قد يواجه الباحث بيانات شاذة يؤدي وجودها إلى افساد تحليل بيانات السلسلة الزمنية، مما يستوجب إزالتها أو التضحية بها.

ماذا نقصد بخطأ التنبؤ؟ ماهي أهم طرق قياسه وأفضل طريقة ولماذا؟

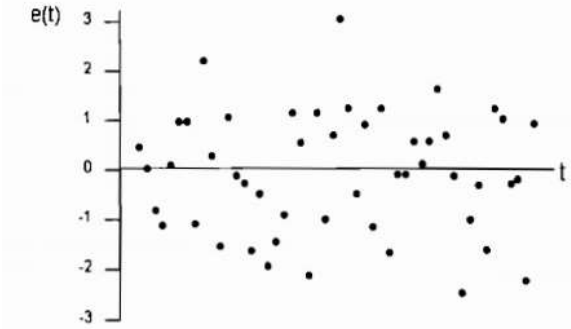
عادة ما تدرس السلسلة الزمنية بغرض اكتشاف نمط التطور التاريخي للظاهرة واستغلال هذا النمط في التنبؤ بالقيم المستقبلية. وأي تنبؤ مستقبلي لأي ظاهرة لا بد أن يحتوي على قدر معين من عدم التأكد، ويمكن ترجمة هذه الحقيقة بإدراج مركبة الخطأ (error component) في نموذج التنبؤ. ومركبة الخطأ هي المركبة غير النمطية التي تعبر عن العوامل التي لا يمكن شرحها باستخدام التغيرات النمطية أو المنتظمة في السلسلة، وكلما كانت هذه المركبة صغيرة زادت قدرتنا على التنبؤ والعكس صحيح. إذا افترضنا أن قيمة الظاهرة موضع الدراسة عند الزمن t هي y_t وأن التنبؤ بالظاهرة عند الزمن t هو $\hat{y}$ فإن الخطأ في التنبؤ عند الزمن t يعرف كالآتي:

$$e_t = y_t - \hat{y_t} \qquad , \quad t = 1,2,3 \dots \dots n$$

حيث يرمز n إلى عدد المشاهدات في السلسلة الزمنية.

وفحص أخطاء التنبؤ المتتالية e_t يوضح مدى ملاءمة أسلوب التنبؤ المستخدم، فكما هو معروف من دراسة الإنحدار أن أسلوب التنبؤ الملائم لابد ان ينتج أخطاء تتصف بطابع العشوائية أي أخطاء خالية من أي تغيرات منتظمة كما في الشكل 03 بالإضافة

إلى بعض الشروط الأخرى . وإذا كانت هذه الأخطاء محتملة بحيث يمكن اعتبار أسلوب التنبؤ ملائم فإنه يجب قياس حجم هذه الأخطاء لتقدير دقة التنبؤ.



ولقد عرف الفكر الاحصائي طرقا عديدة لقياس حجم الأخطاء أهمها ما يلي:

1. مجموع الأخطاء (sum of errors) ويرمز له عادة بالرمز SE ويعرف على الصورة التالية:

$$SE = \sum_{t=1}^{n} e_t = \sum_{t=1}^{n} (y_t - \hat{y}_t)$$

وهذا المقياس لا يفيد كثيرا حيث أنه من المعروف أنه إذا كانت الأخطاء عشوائية فإن هذا المجموع عادة ما يكون قريبا جدا من الصفر بغض النظر عن حجم هذه الأخطاء.

2. متوسط الإنحرافات المطلقة (mean absolute deviation) والذي يرمز له عادة بالرمز MAD ويعرف على الصورة التالية:

$$MAD = \frac{1}{N}\sum_{t=1}^{n} |e_t| = \frac{1}{N}\sum_{t=1}^{n} |(y_t - \hat{y}_t)|$$

بالرغم من معقولية هذا المقياس إلا أنه لا يستخدم كثيرا في مجال السلاسل الزمنية نظرا لصعوبة خصائصه الإحصائية.

3. متوسط مربعات الأخطاء (mean squared error) المعروفة في مجالات الانحدار والسلاسل الزمنية تعتمد على تصغير مجموع مربعات الأخطاء SSE أو تصغير متوسط مربعات الأخطاء MSE وذلك لأن المقام n والذي يمثل عدد الوحدات الزمنية المتاحة (عدد المشاهدات) هو مقدار ثابت. وبصفة عامة يمكن القول بأن خصائص هذا المقياس الاحصائي أسهل بكثير من خصائص متوسط الأخطاء المطلقة MAD.

تحديد معايير اختيار أسلوب التنبؤ

من أهم عناصر تحليل السلاسل الزمنية اختيار أسلوب التنبؤ المناسب. واختيار أسلوب التنبؤ المناسب ليس بالعمل الهين وإنما هو عمل صعب وشاق ويحتاج من الإحصائي ومتخذ القرارات التحلي بالصبر وعدم اليأس بالإضافة إلى مقومات العمل الأساسية من علم وخبرة ومهارة. ويعتمد الإحصائي أو متخذ القرارات بصفة عامة في اختياره لأسلوب التنبؤ المناسب على بعض المعايير أو العوامل العامة أهمها ما يلي:

1. تصغير حجم أخطاء التنبؤ باستخدام أحد المقاييس الثلاثة التي سبق ذكرها.
2. نوعية التنبؤ المطلوب، فإذا كان تنبؤ النقطة (point forecast) هو المطلوب من الدراسة، فإن استخدام أحد هذه الأساليب أو النماذج التقليدية البسيطة قد يكون كافيا لتحقيق هذا الهدف. وفي الكثير من الدراسات قد يكون تنبؤ بالفترة (Interval forecast) هام وكذلك اختبارات الفروض ، وفي مثل هذه الحالات لابد من استخدام أسلوب تنبؤ حديث أكثر دقة وتنظيما مثل أسلوب بوكس جينكنز.
3. عدد المشاهدات المتاحة، فإذا كان عدد المشاهدات صغيرا فإن استخدام أحد الأساليب الحديثة ليس له ما يبرره ويفضل استخدام أحد الأساليب التقليدية.
4. مدى تحقق الفروض النظرية التي تعتمد عليها أسلوب أو نموذج التنبؤ المناسب وهو أهم المعايير التي يجب أن تؤخذ في الاعتبار عند اختيار مثل هذا الأسلوب.

مما سبق يتضح للطالب بأن أفضل أسلوب للتنبؤ ليس بالضرورة هو الأسلوب الذي يحقق أعلى دقة أو أصغر حجم أخطاء ممكن، فقد يستخدم أحد الأساليب بسبب نوعية التنبؤ المطلوب، وقد يستخدم أسلوب آخر بسبب صغر عدد المشاهدات المتاحة، وقد يستخدم أسلوب ثالث بسبب انخفاض تكاليفه، وقد يستخدم أسلوب رابع بسبب سهولة عملياته الإحصائية والحسابية، وقد يستخدم أسلوب خامس لأن الفروض النظرية التي يعتمد عليها تتوافق مع بيانات السلسلة المتاحة. وعادة ما يعتمد أسلوب التنبؤ المستخدم على قدرة الاحصائي أو متخذ القرارات في تحقيق التوازن لكل هذه المعايير. وبصفة عامة يمكن القول بأن طريقة التنبؤ التي يجب استخدامها هي أسهل وأبسط طريقة يمكن تنفيذها في الزمن المتاح والتي تفي باحتياجات وظروف التنبؤ بأقل تكاليف ممكنة.

أذكر أنواع المعادلات؟ حدد طرق التنبؤ وأساليبه؟

هناك ثلاث أنواع من المعادلات:

1. معادلات تعريفية.
2. معادلات سلوكية.
3. معادلات توازنية.

يمكن تجميع طرق التنبؤ الكمية المعروفة في أدبيات السلاسل الزمنية في أسلوبين أساسيين هما:

أسلوب الانحدار: يعتمد على تحديد المتغيرات الأخرى التي قد ترتبط بعلاقة سببية بالظاهرة او المتغير موضع الدراسة الذي يراد التنبؤ به يعرف بالمتغير التابع مع تحديد العلاقة الدالية الملاءمة التي توضح الكيفية التي يرتبط بها هذا المتغير بالمتغيرات الأخرى والتي تسمى بالمتغيرات المستقلة او المفسرة. وباستخدام هذا النموذج يمكن التنبؤ بالمتغير التابع موضع الدراسة في المستقبل إذا أمكن تحديد القيم المستقبلية للمتغيرات المفسرة.

عادة ما يتوافر لدى الباحث مشاهدات تاريخية عن المتغير موضع الدراسة فقط ويريد التنبؤ بالمشاهدات المستقبيلة لهذا المتغير بالاعتماد فقط على هذه المشاهدات. في مثل هذه الحالات يستخدم الأسلوب الثاني للتنبؤ التالي.

تحليل السلاسل الزمنية: والذي يضم تحت مضلته ما يعرف بنماذج السلاسل الزمنية. ويعتمد على تحليل البيانات التاريخية التي أخذت عن الظاهرة او المتغير موضع الدراسة وذلك بغرض تحديد نمط البيانات.

والسؤال الأن هو كيف يمكن التنبؤ بالظاهرة أو المتغير موضع الدراسة باستخدام أسلوب السلاسل الزمنية دون اللجوء إلى متغيرات أخرى مفسرة؟ للإجابة عن هذا السؤال يمكن القول بأن أدبيات السلاسل الزمنية

.

نلجأ إلى التنبؤ باستخدام أسلوب تحليل السلاسل الزمنية.

بافتراض أن هذا النمط سيستمر في المستقبل يستكمل هذا النمط لإعطاء التنبؤات المطلوبة، مثلا إذا كان الهدف من الدراسة التنبؤ بمعدلات النمو الاقتصادي لبد ما، فقد يستطيع الباحث تفسير سلوك هذا المتغير جزئيا بواسطة عدد العمال ورأس المال والصادرات، ولكن جزء من تطور النمو الاقتصادي قد يعود لبعض العوامل الأخرى التي لا يمكن أخذها بعين الاعتبار بسهولة مثل سعر الصرف الاستهلاك العائلي وغيرها من العوامل التي يكون من الصعب أو المستحيل إدراجها في النماذج السببية (الانحدارية) وفي هذه الحالة قد يفضل دراسة التطور التاريخي للنمو الاقتصادي بمعزل عن جميع العوامل المفسرة الأخرى واكتشاف الكيفية التي يتطور بها النمو الاقتصادي واستخدام احد نماذج السلاسل الزمنية لاستكمال هذه السلسلة في المستقبل.

التمييز بين نماذج السلاسل الزمنية للتنبؤ

عرفت العديد من الطرق ونماذج السلاسل الزمنية والتي يمكن تقسيمها إلى ثلاثة أنواع رئيسية هي:

1. النماذج المحددة (غير العشوائية).
2. الطرق الحسية (ad hoc) .
3. نماذج السلاسل الزمنية العشوائية.

الأستاذ: أ. عماري زهير

الإجابة على التمرين 02

		K=2			K=3		
السنة	قيمة المبيعات	المتوسط	الانحراف	مربع الانحراف	المتوسط	الانحراف	مربع الانحراف
2014	9	-	-	-	-	-	-
2015	11	-	-	-	-	-	-
2016	10	10	0	0	-	-	-
2017	12	10,5	1,5	2,25	10	2	4
2018	11	11	0	0	11	0	0
2019	9	11,5	-2,5	6,25	11	-2	4
2020	13	10	3	9	10,67	2,33	5,44
2021	11	11	0	0	11	0	0
2022	9	12	-3	9	11	-2	4
2023	**11**		**المتوسط**	**3,79**		**المتوسط**	**2,91**

أي التنبؤات أفضل؟ برر اجابتك؟

تعتبر تنبؤات k=3 أفضل من تنبؤات k=2 وبالتالي يمكن ان نستعين k=3 للتنبؤ بـ 2023 والتي قيمته تساوي 11.

المبرر: لأن متوسط مربعات الأخطاء k=3 وتساوي 2.91 وهي أقل من متوسط ومربعات الأخطاء k=2

الإجابة على التمرين 03

تقدير القيمة الابتدائية $\widehat{y_1}$ باستخدام الوسط الحسابي لقيم السلسلة

$$\widehat{y_1} = \bar{y} = \frac{\sum y}{n} = \frac{11 + 12 + \cdots .16}{13} = 14$$

استخدام القيمة الابتدائية لإيجاد التنبؤات المناظرة مرة باستخدام w=0.7 ومرة باستخدام w=0.9

بالتعويض في معادلة التمهيد الأسي التالية:

$$\hat{y}_{t+1} = (1 - w)y_t + w\hat{y}_t$$

نتحصل على النتائج وفق الجدول التالي

الأشهر	عدد الأجهزة المباعة بالنسبة المئوية	التنبؤات داخل السلسلة		مربعات الأخطاء	
		W = 0,7	w = 0,9	w = 0,7	w = 0,9
1	11	14	14	9	9
2	12	13,10	13,70	1,21	2,89
3	12	12,77	13,53	0,59	2,34
4	14	12,54	13,38	2,13	0,39
5	13	12,98	13,44	0,00	0,19
6	15	12,98	13,40	4,06	2,57
7	14	13,59	13,56	0,17	0,20
8	15	13,71	13,60	1,66	1,96
9	13	14,10	13,74	1,21	0,55
10	17	13,77	13,67	10,44	11,11
11	16	14,74	14,00	1,59	4,00
12	14	15,12	14,20	1,25	0,04
13	16	14,78	14,18	1,48	3,31
14	?				
		المجموع		34,80	38,56
		متوسط مربعات الأخطاء		**2,68**	**2,97**

للتنبؤ بنسبة الأجهزة المباعة في الشهر الرابع عشر 14 نستخدم الوزن الترجيحي w = 0.7 لأن هذا الوزن يناظر القيمة الأقل لمتوسط مربعات الأخطاء لنموذج التمهيد الأسي وبالتالي تكون القيمة التنبئية $\hat{y}_{14}$ كما يلي:

$$\hat{y}_{14} = (1 - 0.7)16 + (0.7)(14.78) = 15.15\%$$