المحاضرة الأولى (01)

الأهداف: في نهاية هذه المحاضرة يكون الطالب قادرا على:

- **أنواع البيانات وتعريف السلسلة الزمنية.**
- **التمييز بين السلاسل الزمنية المتصلة والمتقطعة.**
- **تنظيف البيانات.**
- **معرفة اهداف دراسة وتحليل السلاسل الزمنية.**
- **تعريف أخطاء التنبؤ وقياس حجمها.**
- **معرفة معايير إختيار أسلوب التنبؤ المناسب.**

ماهية السلاسل الزمنية

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. ويطلق على البيانات الزمنية بالسلسلة الزمنية وهي التي نحن بصدد تعريفها كما يلي:

تعريف: السلسة الزمنية هي مجموعة من المشاهدات او القياسات التي تأخذ على إحدى الظواهر (الاقتصادية – الاجتماعية – الطبية – الطبيعية..) على فترات زمنية متتابعة (ساعة، يوم، شهر، أسبوع، سنة...) عادة ما تكون متساوية الطول.

ويمكن رصد السلاسل الزمنية في شتى أنواع المعرفة وميادين التطبيق المختلفة ففي الاقتصاد يمكن رصد بيانات الدخل الوطني والنمو الاقتصادي ومعدلات التضخم والبطالة السنوية وغيرها من البيانات، وفي علم الاجتماع يمكن رصد عدد الجرائم أسبوعيا وعدد حالات الطلاق او الزواج السنوي وغيرها. وفي مجال التعليم يمكن رصد السلاسل الزمنية الخاصة بتطور أعداد الطلبة واعداد المدرسين في الكليات وغيرها. وفي مجال الطب يمكن رصد السلاسل الزمنية الخاصة بتطور الأمراض المختلفة ومدى تزايد او التناقص في الإصابة بهذه الأمراض مثل التطور التاريخي لنسبة المصابين بالأورام الخبيثة او الكورونا أو السلاسل الزمنية الخاصة برسم الدماغ أو القلب. وفي مجال الأرصاد الجوية يمكن رصد السلاسل الزمنية الخاصة بسرعة الرياح ونسبة الرطوبة ودرجات الحرارة وغيرها.

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

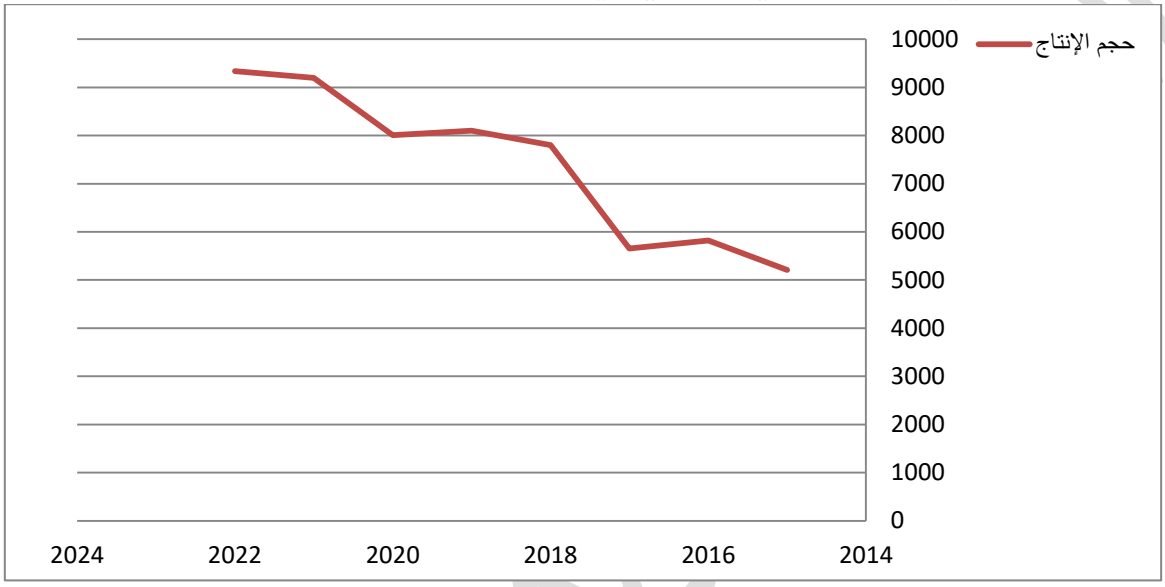
1. **تأخذ بيانات السلسلة الزمنية على فترة طويلة نسبيا يعتقد انها تؤثر على الظاهرة أو المتغير موضع الدراسة، بينما تأخذ البيانات التجريبية أو بيانات الحصر عند نقطة زمنية معينة أو على الأكثر في فترة زمنية قصيرة يعتقد انها لا تؤثر بشكل معنوي على الظاهرة أو المتغير موضع الدراسة وعادة ماتسمى هذه البيانات بالبيانات 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.**

وتعرض السلاسل الزمنية عادة في صورة جدول أو خط أو منحنى بياني يعرف بالخط التاريخي أو المنحنى الزمني (time series plot) **كما في الأمثلة التالية:**

يوضح الجدول التالي حجم الإنتاج السنوي للبترول بالمليون متر مكعب في إحدى الدول من سنة 2015 إلى سنة 2023

السنة	2015	2016	2017	2018	2019	2020	2021	2022
حجم الإنتاج	5210	5820	5655	7800	8100	8010	9200	9335

ويمكن عرض هذه البيانات في شكل خط بياني (تاريخي) في شكل رقم (1):

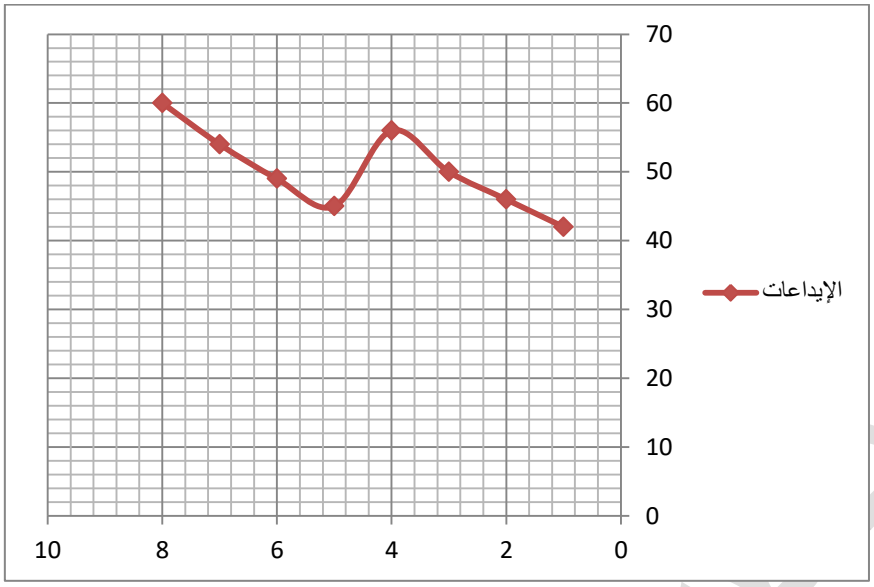


مثال 02:

يوضح الجدول الآتي تطور قيمة الإيداعات ربع سنوية بالمليون دولار في أحد البنوك في سنتي 2020 و 2021

السنة	الموسم (الفصل)	الإيداعات
2020	01	42
	02	46
	03	50
	04	56
2021	01	45
	02	49
	03	54
	04	60

ويمكن عرض هذه البيانات في الشكل رقم (02):



أنواع السلاسل الزمنية

عند دراسة السلاسل الزمنية لبعض الظواهر قد يكون من الممكن أخذ قياسات أو قراءات عند كل لحظة زمنية، ويقال لهذه السلاسل الزمنية بأنها سلسلة متصلة (continuous)، ومن أمثلة هذه السلاسل درجات الحرارة ورسم القلب والدماغ.

أما معظم السلاسل الزمنية التي تنشأ في الواقع فتتكون من قراءات أو مشاهدات مأخوذة عند فترات زمنية محددة مسبقا. وقد تكون هذه الفترات دقائق أو ساعات أو أيام أو أسابيع أو شهور أو سنوات. وتعرف هذه السلاسل بالسلاسل المتقطعة (discret time series) بغض النظر عن طبيعة الظاهرة أو المتغير موضع الدراسة، ومن أمثلة هذه السلاسل الدخل الوطني وسعر الإقفال اليومي لأحد الأسهم في بورصة الأوراق المالية وعدد الحوادث الأسبوعية التي تحدث على أحد الطرق وعدد خريجي الجامعة السنوي وكمية الأمطار الشهرية. والسلاسل الزمنية المتقطعة هي السلاسل التي سنتعامل معها فقط في هذا المقياس، أي اننا سنفترض دائما أن السلسلة متاحة فقط عند نقاط زمنية كمتقطعة تبعد عن بعضها فجوات زمنية متساوية الطول.

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.

أهداف دراسة السلسلة الزمنية

تدرس السلاسل الزمنية عادة لتحقيق عدد من الأهداف، وهي:

1. وصف وتصوير المعلومات المتاحة عن فترة زمنية توضح تطور الظاهرة المدروسة أي وصف الملامح والسمات الرئيسية للسلسلة، ويساعد وصف السلسلة إلى حد كبير في تحديد النموذج الذي يمكن ان يكون مناسبا لتحقيق الأهداف الأخرى والتعرف على حركات الصعود والهبوط في السلسلة الزمنية والتعرف على المكونات الرئيسية مثل الاتجاه العام والتغيرات الموسمية كما سنوضحه في المحاضرة القادمة.
2. التفسير ويقصد به توضيح وشرح التغيرات التي تحدث في الظاهرة باستخدام السلاسل الزمنية الأخرى التي ترتبط بها أو باستخدام عوامل البيئة المحيطة بالظاهرة ومثال ذلك تفسير التغيرات التي تحدث في

سلسلة المبيعات الخاصة بإحدى السلع باستخدام السلسلة الخاصة بتغيرات أسعار هذه السلعة أو بمعرفة القرارات الاقتصادية التي اتخذت وكانت لها علاقة مباشرة على التطور التاريخي للظاهرة. ان ربط التغيرات التي تحدث في السلسلة موضع الدراسة بالتغيرات التي تحدث في السلاسل والعوامل المحيطة من شأنه فهم آلية عمل السلسلة وتفسير الأنماط والتغيرات المنتظمة وغير المنتظمة التي تتعرض لها الظاهرة موضع الدراسة ومدى تأثير كل منها عليها.

3. الرقابة والتحكم، فقد تستخدم الخرائط الزمنية في مراقبة جودة الإنتاج وذلك من أجل التحكم في مستوى كفاءة العملية الإنتاجية وذلك باتخاذ القرارات المناسبة من وقف العملية الإنتاجية وتعديل مسارها أو استمرارها.

4. التنبؤ بالمشاهدات المستقبلية والذي عادة ما يمثل الهدف النهائي من تحليل السلاسل الزمنية. وهذا الهدف هو أوضح الأهداف وأكثرها شعبية بالنسبة لدارس الإحصاء أو مستخدمه، فتحليل السلاسل الزمنية يبدأ عادة بالتعرف على النمط المناسب لشرح آلية تطور هذه السلسلة واستكمال هذا النمط مستقبلا. والفرض الأساسي في أساليب التنبؤ المستخدمة هو أن هذا النمط الذي تم التعرف عليه سيستمر في المستقبل القريب، وتجدر الإشارة إلى أنه لا يمكن لأي أسلوب تنبؤ أن يعطي نتائج جيدة إذا استمر هذا النمط، ولذلك فإنه ينصح دائما بالتنبؤ بالقيم المستقبلية القريبة وتحديثها بمجرد الحصول على أي مشاهدة جديدة.

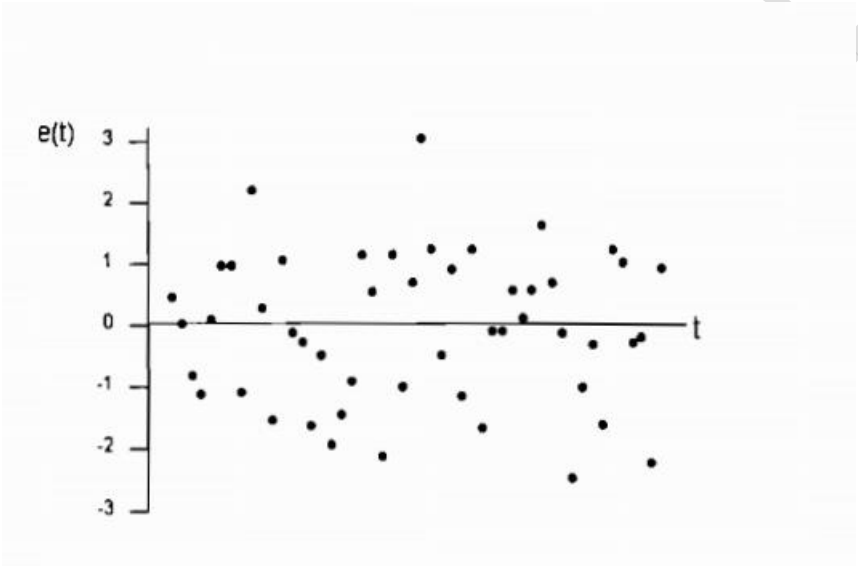
قياس أخطاء التنبؤ

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:

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

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.

إختيار أسلوب التنبؤ المناسب

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.

المحاضرة الأولى أ.د عماري زهير