حزم البرامج الجاهزة والتحليل الإحصائي للبيانات

الدكتور / جمال أحمد الشوادفي(*)

ملخص البحث

هذا البحث يستعرض أهم حزم البرامج الإحصائية الجاهزة التي يمكن استخدامها في تحليل البيانات، ومزايا وعيوب كل منها، ومواقع هذه البرامج على شبكة الانترنت. وكذلك يستعرض إمكانية تحليل البيانات عن بعد عبر شبكة الانترنت وأهم المواقع والبرامج التي تدعم ذلك. ونظرا لتوافر قواعد بيانات ضخمة - كتلك التي تجمعها الأقمار الاصطناعية بطريقة آلية وشبه آلية - وتعجز البرامج التقليدية عن تحليلها، فإن البحث يناقش مشكلة تحليل هذه البيانات، والبرامج المناسبة لتحليلها. ثم يبين كيف يمكن حل بعض المشاكل الإحصائية التي يتعذر حلها مباشرة باستخدام أي من الوسائل الثلاث السابقة (حزم البرامج الجاهزة - التحليل عن بعد - برامج تحليل البيانات الضخمة). وسنناقش أحد هذه المشاكل وهي مشكلة استخدام طريقة جديدة للتنبؤ بالسلاسل الزمنية باستخدام تقنيات الذكاء الاصطناعي والمنشورة في بحث Alshawadfi (2003)، واختبار كفاءة هذه الطريقة ومقارنة أدائها مع طريقة بوكس جينكنز الشهيرة، وتوضيح قدرتها على تحليل بيانات واقعية.

وقد استخلص البحث أنه يمكن تحليل البيانات المتاحة لدينا باستخدام برنامج مناسب من حزم البرامج التقليدية وفقا لظروف البحث والبرنامج والباحث. كما يمكن تحليل البيانات عن بعد عبر الانترنت، وفي حالة البيانات الضخمة نستخدم حزم برامج التحليل الخاصة بها، وهي قليلة العدد وبعضها قيد التطوير. أما البيانات التي تتطلب طرق حديثة لمعالجتها ويصعب معالجتها باستخدام أي من الوسائل السابقة فيجب تصميم برامج مناسبة لتحليلها وفق طرق التحليل المطلوبة. وقد أوضحنا ذلك بحل مشكلة استخدام بعض طرق الذكاء الاصطناعي في تحليل

(*) أستاذ بقسم الإحصاء ووكيل كلية التجارة – جامعة الأزهر «بنين» مدينة نصر - القاهرة

السلاسل الزمنية بتصميم برنامج مناسب MATLAB Toolbox ـ مرفق في ملحق A ـ لتوليد ١٦٠٠٠ سلسلة زمنية من نماذج ARIMA مختلفة والتنبؤ بقيمها المستقبلية بطريقتي الذكاء الاصطناعي وبوكس جينكنز مع المقارنة بين الطريقتين وتلخيص النتائج، كما تم تصميم برنامج أخر ـ مرفق في ملحق B ـ للتنبؤ بالسلاسل الزمنية باستخدام طريقة الذكاء الاصطناعي وتم استخدامه في تحليل بيانات واقعية.

[١] مقدمة

لقد كان للانتشار الواسع لاستخدام الكمبيوتر أثراً واضحاً في إنجاز البحوث والدراسات ودعم صناعة القرار. فلقد قلل الكمبيوتر وقت إجراء الحسابات الشاقة عند تحليل بيانات البحوث. كما أنه شجع الكثير من الباحثين وصناع القرار على استخدام طرق علمية وبحثية متقدمة، والتي أصبحت متاحة من خلال حزم البرامج الجاهزة، وتحليل البيانات عن بعد. توجد حاجة كبيرة في مجال البحث العلمي وإدارة الأعمال إلى حزم برامج جاهزة تمكن غير المتخصصين من تحليل البيانات بدقة وكفاءة. وهذه البيانات التي يتم تحليلها قد تكون تاريخية أو ميدانية أو خليط منهما، والتجميع والتقييم والاستنتاج الذكي للمعلومات من هذه البيانات باستخدام طرق مقبولة علمياً سيكون حاسماً لفهم فعال ومؤثر في صناعة القرار. وبالطبع توجد مجموعات كبيرة من البرامج لتحليل البيانات وإنشاء الرسوم البيانية، ولكن ما هو البرنامج الذي نختاره لتحليل البيانات؟.

ولنبدأ أولا بتقديم تعاريف أساسية لبرنامج الكمبيوتر والفرق بين برامج النظام وبرامج التطبيقات والبيانات والمعلومات.

برنامج الكمبيوتر: عبارة عن مجموعة مرتبة ومتتابعة من التعليمات تحدد للكمبيوتر خطوات تنفيذ عملية معينة. وتتضمن هذه التعليمات تعليمات حسابية وتعليمات مقارنة وتعليمات تفرع بالإضافة إلي تعليمات الإدخال والإخراج. تمثل برامج الكمبيوتر شريان الحياة بالنسبة لهذا الجهاز، وعلى قدر توفرها وتنوعها يمكن الاستفادة من إمكانيات جهاز الكمبيوتر. وتنقسم برامج الكمبيوتر عموما

إلي نوعين: برامج للنظام وبرامج للتطبيقات وفيما يلي نقدم فكرة موجزة عن كل نوع.

١ــ برامج النظام : System Software programs

وهى برامج سابقة التجهيز تم تصميمها لتنفيذ الوظائف الأساسية للكمبيوتر مثل مراقبة ودعم أجهزة الحاسب وأنشطة معالجة البيانات التي يقوم الحاسب بتنفيذها. وفى أجهزة التطبيقات تكون بعض هذه البرامج مكتوبة ومخزنة في الذاكرة الدائمة (الروم) ويطلق عليها مجموعة برامج النظام الأساسي للإدخال والإخراج BIOS وهي اختصار العبارة Basic Input/Output System ومن هذه البرامج أيضاً البرامج الخاصة بأنظمة التشغيل مثل نظام التشغيل Windows

٢ــ برامج التطبيقات: Application Programs

يكتب برنامج الكمبيوتر عادة باستخدام لغة للتفاهم مع الكمبيوتر، والكمبيوتر يستخدم ما يسمى بلغة الآلة Machine Language والتي ابتكرها العالم الرياضي الشهير فون نويمان وهى تتضمن فقط الأرقام الثنائية bits والمكونة من صفر أو واحد، ونظراً لصعوبة الكتابة بهذه اللغة فقد تم ابتكار لغة تسمى لغة التجميع Assembly Language وتستخدم الحروف والأرقام للتعبير عن العمليات على أن يتم ترجمتها بعد ذلك إلي لغة الآلة بواسطة ما يسمى بالمجمع Assembler حتى يمكن للكمبيوتر فهمها. ثم ظهرت بعد ذلك لغات المستوى العالي High Level Languages وهذه اللغات تستخدم الكلمات والحروف الإنجليزية العادية بالإضافة إلى الأرقام والرموز الخاصة، ومن أشهر هذه اللغات لغة البيزك ولغة سي++ والأوراكل Oracle وغيرها.

وبناءً على ذلك فإن برامج التطبيقات تكتب عادة بواحدة من لغات المستوى العالي مثل لغة البيزك المرئي أو لغة سى++ أو غيرها. وتستخدم حزم البرامج الجاهزة هذه في حل المشاكل العلمية مثل التحليل الإحصائي والتحليل العددي والبرمجة الخطية والنماذج الرياضية، ومن أمثلتها حزم برامج MINITAB, SAS SPSS، QSB، MATLAB. وتستخدم بعض برامج التطبيقات كذلك في حل

المشاكل التجارية مثل برامج معالجة الأجور ومراقبة المخزون ومراقبة جودة الإنتاج وأعمال البنوك. كما توجد برامج تطبيقات عديدة في مجالات الطب والتعليم والقانون والفنون ... وغيرها.

ويفضل عادة استخدام برامج للحماية من فيروسات الكمبيوتر مثل برنامج NORTON Antivirus ، أو برنامج Mcafee. وفيروس الكمبيوتر عبارة عن برنامج صغير يتم كتابته بواسطة أحد مبرمجي الكمبيوتر المحترفين غالبا ربما على سبيل الدعابة والتسلية أو لأغراض أخرى. وينتقل من جهاز إلى جهاز أخر عبر شبكة الانترنت أو عبر الأقراص المدمجة أو الفلاشات.

البيانات DATA: هي عبارة عن مجموعة من المشاهدات أو الملاحظات أو القياسات والتي قد تكون في صورة كمية (أعداد) أو في صورة وصفية تصف الظاهرة محل الدراسة.

وتستخدم المنظمات التقليدية البيانات كوسيلة للمساعدة في إدارة عملياتها، في حين أن المنظمات القيادية تستخدم البيانات استراتيجياً لتوسيع أعمالها وتحسين الربحية وتقليل التكاليف والتسويق بفعالية أكثر. وجودة أساليب ووسائل تحليل البيانات تسهم بفعالية كبيرة في إنجاز الأهداف الاستراتيجية هذه.

المعلومات Information: هي النتائج التي يمكن الحصول عليها بعد معالجة البيانات وهى تزيد من مستوى المعرفة لدى الأفراد أو الهيئات ومن ثم تساعد على اتخاذ القرارات بطريقة صحيحة بعبارة أخرى البيانات هي المادة الخام للمعلومات .

علم الإحصاء Statistics: هو علم يهتم بجمع البيانات الضرورية ومعالجتها بأساليب علمية من خلال تبويبها وتلخيصها وعرضها وتحليلها للحصول منها على المعلومات اللازمة لاتخاذ القرارات المناسبة . في العصر الحالي تستخدم عادة طرق معالجة إلكترونية Electronic Processing أي باستخدام الكمبيوتر لإجراء تحليل إحصائي للبيانات بدلاً من طرق المعالجة اليدوية Manual processing.

مما سبق يتبين لنا أن هناك الكثير من الدراسات التي يتحتم فيها استخدام علم الإحصاء والأساليب الإحصائية العلمية لتحليل البيانات. ويتم انجاز ذلك بمعالجة

البيانات الكترونيا عبر حزم برامج إحصائية جاهزة. ومن ثم نحصل على المعلومات. ولا شك أن المعلومات أصبحت اليوم سلعة رابحة وبالغة الأهمية وخاصة مع تطور الحاسبات الإلكترونية ووسائل الاتصالات الحديثة عبر الأقمار الاصطناعية، وتؤثر المعلومات على سلوك الفرد والمنشأة في اتخاذ القرارات المختلفة، فقد انتقل العالم اليوم من عصر الثورة الصناعية إلى عصر ثورة المعلومات، غير أن الحصول على المعلومات يستلزم توافر بيانات دقيقة وممثلة لواقع الظاهرة محل الدراسة، ثم استخدام أسلوب علمي في معالجة هذه البيانات للحصول على معلومات كافية كما وكيفاً لاحتياجات الباحثين وصانعي القرارات.

هدف البحث:

هذا البحث يستعرض أهم حزم البرامج الإحصائية الجاهزة التي يمكن استخدامها في تحليل البيانات، ومزايا وعيوب كل منها، ومواقع هذه البرامج على شبكة الانترنت. وكذلك يناقش إمكانية تحليل البيانات عن بعد عبر شبكة الانترنت وأهم المواقع والبرامج التي تدعم ذلك. ونظراً لتوافر قواعد بيانات ضخمة في العصر الحالي تعجز البرامج التقليدية عن تحليلها، فإن البحث يناقش مشكلة تحليل البيانات الضخمة، والبرامج المناسبة لتحليلها. ثم يبين كيف يمكن حل بعض المشاكل الإحصائية التي يتعذر حلها مباشرة باستخدام أي من الوسائل الثلاث السابقة (حزم البرامج الجاهزة - التحليل عن بعد - برامج تحليل البيانات الضخمة). وكتطبيق عملي على ذلك سنقوم بدراسة أحد هذه المشاكل وهي مشكلة استخدام طريقة جديدة للتنبؤ بالسلاسل الزمنية باستخدام تقنيات الذكاء الاصطناعي والمنشورة في بحث Alshawadfi (2003)، واختبار كفاءة هذه الطريقة ومقارنة أدائها مع طريقة بوكس جينكنز الشهيرة، وتوضيح قدرتها على تحليل بيانات واقعية.

المبحث التالي يستعرض أهم حزم برامج تحليل البيانات ومزايا وعيوب كل منها، ومواقع هذه البرامج على شبكة الانترنت. أما المبحث الثالث فيناقش إمكانية تحليل البيانات عن بعد عبر الانترنت وأهم المواقع والبرامج التي تدعم ذلك، أما

المبحث الرابع فيناقش مشكلة تحليل البيانات الضخمة والتي تعجز برامج التحليل التقليدية عن إنجازها ، المبحث الخامس يقدم بعض الأمثلة لمشاكل تحليل البيانات وكيفية التغلب عليها عندما يتعذر التعامل معها مباشرة باستخدام حزم البرامج الجاهزة أو التحليل عن بعد أو برامج تحليل البيانات الضخمة . المبحث الأخير يقدم الخلاصة والتوصيات .

[٢] برامج تحليل البيانات

حزم البرامج الإحصائية عبارة عن مجموعة برامج كمبيوتر مكتوبة لإنجاز أهداف تتضمن عادة تحليل البيانات . هذه الأهداف تشمل إدخال البيانات وجدولتها وتمثيلها بيانياً وعمل الحسابات والتحليلات الإحصائية اللازمة . وبالإضافة إلى إمكانية قيام الباحث بكتابة برامج تحليل إحصائي مناسبة بأحد لغات الكمبيوتر مثل لغة سى أو غيرها ، فإنه توجد مئات من حزم البرامج الإحصائية الجاهزة Packages والتي تقدم معظم طرق التحليل الإحصائي الوصفي والاستدلالي اللازمة للباحثين .

توجد ثلاث مجموعات من البرامج المتاحة تستخدم أشكالاً مختلفة لواجهة المستخدم . وهذه الأشكال الثلاث هي برامج قواعد بيانات منطقية مثل Oracle و Sybase وبرامج تحليل رياضي وإحصائي مثل MATLAB . تتميز البرامج الرياضية بتقديم إمكانيات تحليلية عبر برمجة تخطيط المصفوفات ، كما أنها تشمل تخطيط الأمثلية ودوال أخرى مفيدة لأغراض إحصائية . كما أن البرامج المرئية مفيدة في تحليل بيانات وصفية .

تتدرج واجهة المستخدم من سطور الأوامر إلى الواجهة الرسومية للمستخدم ، وواجهات نظام السحب والإفلات .

حزم البرامج الإحصائية الأكثر انتشاراً تشمل حزم برامج : SPSS , SAS , MINITAB وهذه برامج كبيرة تقوم بمعظم عمليات التحليل الإحصائي الوصفي والاستدلالي بالإضافة إلي مجموعة كبيرة من حزم البرامج الأخرى مثل : SYSTAT , SIGMASTAT, Datadesk .

يمكن استخدام برامج ليست في مقدمة برامج تحليل البيانات نظرا لتوفرها لدى الباحث أو سهولة استخدامها أو اعتياده عليها . وقد راجع Morgan (1998) ثمانية من هذه البرامج ، وقدم توضيحات عن هدف كل منها وإمكانياته وحدوده وواجهته وطريقة الحصول عليه، بما ينير الطريق لمن يرغب في استخدام مثل هذه البرامج . وهذه البرامج هي :

GraphPad, Prism, InStat, ISP, NCSS, SigmaStat, Statistix, Statmost and Winks.

بعض حزم البرامج السابقة تكون مهتمة بجوانب معينة في الإحصاء مثل InStat, ISP, NCSS, Statistix, Statmost, Winks. في حين يهتم بعضها بتحليل البيانات الخاصة بالمجالات الطبية مثل BMDP أو غيرها من المجالات. وبعضها تكون مشتركة بمعنى أنها تخدم مجالات أخرى مثل حزم برامج الاقتصاد القياسي والتي تكون مفيدة جدا في التحليل الإحصائي للسلاسل الزمنية ومنها حزم برامج : RATS – EVIEWS – SHAZAM

هذا وتوجد حزم برامج مساعدة في مجالات الرياضيات وبحوث العمليات والإحصاء مثل :

MATLAB, MATHCAD, QSB, MANAGER, ORSTAT, STORM

كما توجد حزم برامج لإجراء التنبؤ الأوتوماتيكي Automatic Forecasting مثل حزم برامج :

AUTOBOX, AUTOCAST, FORECAST PRO, NCSS, 4CAST/2

كما يمكن الحصول على رسوم بيانية عالية الدقة من خلال حزم برامج مثل Harvard أو excel وغيرها ، والتي يمكن من خلالها الحصول على أعمدة ودوائر ومنحنيات بدقة عالية وأشكال وألوان مختلفة. ولمزيد من التفاصيل عن حزم البرامج السابقة أنظر على سبيل المثال : Aghadazeh and Romal (1992) , Tashman and Leach (1991), Small (1997) Rahlf,(1994) , Ouster (1998) , Ord and Lowe (1996).

وتختلف البرامج الإحصائية فيما بينها من حيث متطلبات الإدخال والإخراج وأداء الحسابات المطلوبة ولكل منها مزاياه وعيوبه، وعلى أية حال، فمن المهم أن نكون على دراية بالبرامج المتاحة على جهاز الكمبيوتر الذي نستخدمه، وطريقة إدخال البيانات إليها، ثم بعد إدخال البيانات بطريقة صحيحة وتشغيلها سنحصل على النتائج المطلوبة وهذه تتطلب منا دراية بالمفاهيم الإحصائية الأساسية حتى نستطيع فهم وتفسير النتائج بطريقة صحيحة .

على أية حال فحزم البرامج الإحصائية الجاهزة يمكنها عمل الحسابات لمعظم البحوث، كما أنها تسمح بكتابة برامج مناسبة داخلها لإجراء الحسابات اللازمة بطريقة آلية كما هو الحال في حزم برامج Minitab . وسنناقش هنا أهم حزم البرامج هذه بشيء من التفصيل :

برنامج ساس The SAS® System for Statistical Analysis

ساس هي اختصار لعبارة Statistical Analysis System بمعنى نظام تحليل إحصائي، وقد بدأ عام ١٩٦٠ في مشروع الإحصاءات التجريبية بجامعة ولاية كارولينا الشمالية ثم أنشئ معهد ساس لاحقا في عام ١٩٧٦م . ومنذ ذلك التاريخ ونظام ساس يتطور كنظام لإدارة وتحليل كامل للبيانات . من بين منتجات نظام ساس إدارة قواعد البيانات الضخمة وتحليل إحصائي للسلاسل الزمنية وتحليل إحصائي لمعظم المشاكل التقليدية في الإحصاء تشمل تحليل متغيرات متعددة ونماذج خطية أو خطية معممة والتصنيف والرسوم البيانية ونظم المعلومات الجغرافية.

وهذا النظام متاح للاستخدام مع حاسبات شخصية ومتوسطة أو كبيرة الحجم وباستخدام أنظمة تشغيل مختلفة. يمكن استخدام نظام ساس كذلك لإجراء دراسات محاكاة لتوزيعات مختلفة بمولدات أرقام عشوائية لتوزيعات مختلفة . كما أن التحكم في قواعد البيانات الضخمة سهل نسبيا، إذ يمكن من خلاله تقسيم أو دمج أو إعادة ترتيب قواعد البيانات بسهولة. بالإضافة إلى ذلك، فانه يمكن باستخدام تقنيات قواعد البيانات الحديثة تداول البيانات بسهولة تامة وسرعة فائقة.

أما برامج المستخدم المكتوبة بلغة C أو PL/1 فيمكن إدخالها لنظام ساس من خلال برنامج SAS/Base ، خاصة عند رغبة المستخدم في عمل تحويلات بيانات معقدة. كما يمكن كتابة برامج تطبيقات باستخدام لغة مصفوفات من خلال برنامج SAS/IML في نظام ساس .

يمكن تصدير واستيراد البيانات من وإلى ساس من معظم البرامج التجارية عبر برنامج SAS/ACCESS، كما أن برنامج ساس للحاسبات الشخصية يسمح باستيراد وتصدير البيانات من وإلى أوراق العمل التقليدية وبرامج قواعد البيانات.

يسمح برنامج SAS/STAT بتحليل أشكال عديدة من البيانات والنماذج الخطية وتحليل الانحدار والتباين والتغاير والنماذج الخطية العامة والانحدار اللوجستي وانحدار بواسون بالإضافة إلى طرق تحليل المتغيرات المتعددة مثل: MANOVA, Canonical Correlation, Discriminant Analysis, Factor Analysis, Clustering , Categorical Data Analysis (Including Log-Linear Models), . بالإضافة إلى كل طرق التحليل الوصفي والاستدلالي .

كما أن منتج .SAS/GRAPH يسمح بعمل رسوم بيانية مثل q-q plots والبواقي ودوال أخرى ورسوم ثلاثية الأبعاد . ويتم انجاز تحليل السلاسل الزمنية المتقدم عبر ساس من خلال برنامج SAS/ETS والذي يسمح بتحليل سلاسل زمنية متعددة ونماذج ARIMA ونماذج فضاء الحالة State Space Models والتحليل الطيفي للسلاسل الزمنية.

أما برنامج SAS/INSIGHT فيقوم بالتحليل الاستكشافي للبيانات مع تقديم عروض بيانية للنتائج، في حين أن معظم البرامج الأخرى مرتبطة بعروض مرئية.

بالنسبة لنظم المعلومات الجغرافية فإنها متاحة في ساس عبر برنامج SAS/GIS .

و يمثل برنامج ساس لفترة طويلة حزم البرامج الإحصائية القياسية. كما أن المهارات التي يكتسبها المستخدمون لبرنامج ساس قد تكون أكبر منها عند استخدام حزم برامج أخرى .

والموقع الرسمي لبرنامج ساس على الإنترنت هو : http://www.sas.com/

كما أن موقع http://is.rice.edu/~radam/prog.html يحوي أفكاراً مطورة لاستعمال ساس. وباستخدام محركات البحث المعروفة يمكن الحصول على آلاف المواقع التي تشير إلى حزم برامج ساس.

هذا وتوجد نظم برامج إحصائية أخرى مماثلة لبرنامج ساس، وقد بدأت العمل علي أجهزة كمبيوتر كبيرة ثم شملت الحاسبات الشخصية لاحقا من هذه البرامج : MINITAB, BMDP,SPSS

برنامج ميني تاب MINITAB

بدأ برنامج ميني تاب في الظهور قبل أكثر من ٢٠ عاماً ليقدم تحليل البيانات عبر مستويات متنوعة للعلماء وللمستخدمين في المجال الصناعي والإداري وكذلك للطلاب الجامعين. بالرغم من أنه أنشئ أساساً للمساعدة في تعليم أساسيات الإحصاء للطلاب، ثم وسعت الشركة مجال البرنامج، بعد ذلك، ليشمل التحكم الكمي وتصميم التجارب ومنظومة من الإحصاءات العامة.

البرنامج متاح عبر نظم تشغيل مختلفة تشمل أنظمة تشغيل ويندوز وماكنتوش و.OpenVMS-Unix، كما توجد نسخة للطالب من البرنامج تستخدم على نطاق كبير لطلاب الجامعات والمعاهد العليا. والبرنامج سهل تعلمه واستخدامه كما يمكن تبادل البيانات مع نفس البرنامج في نظام ماكنتوش وكذلك مع برامج أخرى كثيرة تشمل:

Lotus, Excel, Symphony, Quattro Pro, dBase and Text (ASCII) Files.

يمكن تصميم برامج قصيرة (Macro) داخل برنامج ميني تاب لتنفيذ عمليات معينة .

موقع البرنامج على الانترنت هو : http://www.minitab.com/. بالإضافة إلى آلاف المواقع التي تشير إلى البرنامج.

وإذا كان برنامج ميني تاب يستخدم أساسا في المجال التعليمي، فإن برنامج BMDP يستخدم في المجال الطبي ، أما برنامج SPSS فمناسب للعلوم الاجتماعية. والبعض يستخدم كذلك برنامج SAS بكثرة، من ناحية أخرى فإن برنامج S-PLUS يستخدم أكثر في العمليات الحسابية الإحصائية.

برنامج بي أم دي بي BMDP

ترجع جذور هذا البرنامج إلى حزم برامج التحليل الطبي الحيوي في أواخر الستينات وتأتي الإصدارات الحالية منه في أشكال مختلفة تتضمن :

BMDP New System Personal Edition, The BMDP Classic for PCs – Release 7 and The BMDP New System Professional Edition.

الإصدارات الحديثة من برنامج BMDP لها واجهة مستخدم سهلة الاستعمال وتمكن من ملئ البيانات المطلوبة بسهولة.كما توجد نوافذ تبرز عند الطلب ومربعات حوار تبرز للمستخدم لتساعده في إكمال تحليل البيانات.

تجمع نسخة المحترفين من هذا البرنامج بين الإصدار رقم ٧ للحاسبات الشخصية وبين ميزة التحكم واستكشاف البيانات في النظام الجديد لبرنامج BMDP .

يعتبر برنامج BMDP التقليدي قياسياً لتحليلات إحصائية متخصصة إذ أنه يضم مكتبة برامج شاملة بها أكثر من ٤٠ برنامج فرعي. كل برنامج من هذه البرامج الفرعية قد تم اختباره وتجربته ومصمم بطرق برمجة أكثر حداثة. وتتضمن الأدوات التصويرية إمكانيات رسوم تفصيلية والتي يمكن ربطها مع صفحة البيانات والمخرجات الإحصائية وعدد من الرسوم القياسية والخرائط.

كما يتضمن النظام الجديد إحصاءات وصفية واختبارات معلمية واختبارات لامعلمية وتحليل التباين وتحليل متغيرات متعددة. ونسخة المحترفين تضيف إلى ما سبق توصيف متعدد الاتجاهات للبيانات والنمذجة الخطية اللوغاريتمية والانحدار

للمكونات الرئيسية وانحدار القمة Ridge Regression وتحليل التناظر والانحدار المتدرج وانحدار غير خطي وتحليل التمايز لحالات أو متغيرات أو قطاعات وتحليل السلاسل الزمنية . موقع البرنامج على الانترنت هو :
http://www.ppgsoft.com/bmdp00.html
بالإضافة إلى مواقع عديدة تشير إلى هذا البرنامج .

برنامج أس بي ا س اس SPSS

يعتبر برنامج SPSS برنامج لشركة متعددة الجنسيات ويقدم حلولاً إحصائية لبحوث التسويق والمبيعات وتحسين الجودة والبحوث العلمية والتعليم والبحوث الحكومية . والبرنامج يعمل مع جميع نماذج الكمبيوتر وصمم البرنامج في أواخر الستينات (عام ١٩٦٨م) . وفي عام ١٩٧٥م تم تكوين مركز رئيسي للشركة في شيكاغو وفي أغسطس من عام ١٩٩٣م تم تسويق البرنامج تجارياً . أما قبل ذلك فكان مصمماً للعمل مع أجهزة كبيرة الحجم ويتم تسويقه لمعاهد ومؤسسات أكاديمية . وقد ظهرت نسخة من البرنامج وهي SPSS/PC+ في ١٩٨٤م لتعمل في بيئة مايكروسوفت دوس، كما صمم للعمل مع نظام يونكس عام ١٩٨٨م ولنظام ماكنتوش عام ١٩٩٠م . والبرنامج مجهز كأداة كاملة للتحليل الإحصائي والرسوم البيانية وإعداد التقارير ، تشمل منتجات اس بي اس اس برامج : SPSS Professional Statistics, SPSS Advanced Statistics, SPSS Tables, SPSS Trends, SPSS Categories, SPSS CHAID, SPSS LISREL 7, SPSS Developer's Kit, SPSS Exact Tests, Teleform, and MapInfo.
موقع البرنامج على الانترنت هو .http://www.spss.com/ كما توجد آلاف المواقع على شبكة الانترنت التي تشير إلى البرنامج .

برنامج اس بلس S-PLUS

يعتبر برنامج S-PLUS أحد البرامج التي تقدم مرونة أكبر في توظيف دوال المستخدم . يمكن اعتبار S-PLUS بأنه أحد لغات المستوى العالي التي صممت لإجراء توظيف سهل للدوال الإحصائية . بالإضافة إلى ذلك فهذه اللغة تدعم الرسومات وإمكانيات الطباعة . وهي تمثل تطور داعم للغة التحليل الإحصائي اس .

وكما هو معلوم فإن لغة اس طورت أساساً بواسطة معامل أيه تي وبيل ويثا وانقسمت لاحقاً إلى معامل AT&T، ومعامل لوسنت بيل بواسطة فريق من الباحثين. الوصف الأصلي للغة أس كتبه شامبر وويكس ١٩٨٨ . برنامج S-PLUS يعمل على أجهزة الحاسبات الشخصية بنظام تشغيل ويندوز وكذلك أجهزة يونكس. كما يمكن استدعاء S-PLUS في لغات سي وفورتران والعكس استدعاء دوال من سي وفورتران في S-PLUS . تضاف برامج مستمرة لدعم S-PLUS .

موقع S-PLUS على الانترنت هو : http://www.mathsoft.com/.

توجد حزم برامج إحصائية أخرى منها حزم برامج : SYSTAT, DataDesk, and JMP.

برنامج Systat تابع الآن للشركة المالكة لبرنامج SPSS، ويمكن الحصول على معلومات إضافية عنه في الموقع : http://www.spss.com/

أما برنامج DataDesk فمصمم للعمل مع أجهزة ماكنتوش، وله واجهة مستخدم رسومية تتيح العديد من الرسومات وطرق التحليل الإحصائي، وموقع البرنامج على الانترنت هو : http://www.lightlink.com/datadesk/

يعتبر برنامج JMP أحد منتجات ساس الذي له واجهة رسومية يعمل مع أجهزة بي سي وماكنتوش . يشبه برنامج DATADESK وموقعه على الانترنت هو : http://www.sas.com

والوصف السابق لحزم البرامج الإحصائية يغطي معظم حزم البرامج الإحصائية المتاحة.

حزم البرامج الرياضية تتضمن كذلك إمكانات إحصائية خاصة في المجالات التي تشمل تطبيقات إحصائية هندسية أو في مجال العلوم الأساسية. من أهم حزم البرامج الرياضية نجد حزم برامج ماتلاب MATLAB . وهذا البرنامج له عدة أشكال مثل لغة أيه بي أل APL الشهيرة في السبعينيات. والبرنامج له بيئة حسابية نشطة تستخدم لعرض وتحليل بيانات علمية وإحصائية، وهو مشابه لبرنامج آيه دي ال ADL المستخدم في مجال الذرة. يتم التعامل مع البيانات في

البرنامج في شكل مصفوفات، يمكن للمستخدم إجراء تحليل عددي، عمليات إشارة، وتحليلات إحصائية باستخدام المصفوفات، بما يعفي المستخدم من اعتبارات البرمجة في لغات سي وفورتران وغيرها. توجد نسخ متنوعة من البرنامج تعمل في بيئة يونكس بي سي وماكنتوش.

يتضمن البرنامج العديد من دوال التحليل الأساسية والرسومات ومعظمها مكتوب كدوال ملف ام M File، وهي عبارة عن ملفات نصية يمكن للمستخدم قراءتها واستعمالها في مجالات أخرى.

وللمستخدم كذلك أن يصمم ملفات ام خاصة له بما يجعل MATLAB بمثابة لغة برمجة .

كما أن إضافة مترجم لغة سي ومكتبة سي الرياضية إلى البرنامج يسمح للمستخدم كتابة ملف تنفيذي باستخدام دوال مكتبة MATLAB وبتنفيذ أسرع عدة مرات وتطبيقات فريدة.

وللباحثين الراغبين في دوال تخصصية أكثر يقدم MATLAB برامج عديدة في شكل صندوق أدوات (TOOLBOX)، وهي عادة عبارة عن ملفات ام توظف طرقاً ودوالاً متخصصة بفروع علمية معينة. من هذه التخصصات: الإحصاء وغيرها. وموقع برنامج MATLAB على الانترنت هو:

http://www.mathworks.com/

هناك حزم برامج رياضية أخرى منها Mathematica وMaple وكلاهما يدعم واجهات رسومية وعددية ولكن صعب تعلمهما. توجد تفاصيل أكثر عن برنامج Mathematica متاحة في الموقع :

http://www.wolfram.com/

أما معلومات برنامج Maple فمتاحة في الموقع :

http://www.maplesoft.com/

أحد حزم البرامج الرياضية الأخرى هو برنامج MathCad وهي حزمة برامج تجمع واجهات عددية ورسومية وإشارية معا ، وتوجد معلومات عن البرنامج متاحة

في نفس موقع برنامج اس بلس S-plus لأن كلاهما من شركة واحدة وهي شركة Mathsoft . وعنوان هذا الموقع هو : http://www.mathsoft.com/.

تجمع حزمة برامج اي ام اس ال IMSL مكتبة برامج فرعية لطرق الرياضيات والإحصاء ، معظم العلماء يعرفون هذه المكتبة، ولمعلومات إضافية عنها راجع الموقع : http://www.vni.com/

كما توجد ملاحظات كثيرة عنها في الموقع :
http://www.c8.lanl.gov/dist_comp2/MATH/Imsl/imsl_keyword.html

[٣] تحليل البيانات عن بعد عبر الانترنت

الإنترنت هي عبارة عن مجموعة من آلاف الشبكات المتصلة ببعضها البعض حول العالم، بحيث تتضمن مجموعة من أجهزة الكمبيوتر التي تتصل ببعضها البعض للمشاركة في المعلومات حول العالم . ومجموعة الأجهزة المتصلة في شبكة قد تشكل شبكة محلية (وتسمى إنترانت) . وتتولى كل منشأه أو جهة متصلة بالإنترنت مسئولية صيانة الشبكة الخاصة بها . وقد أنشئت الشبكة العالمية العنكبوتية World Wide Web(WWW) في أوائل التسعينات بمعرفة المعمل الأوربي للفيزياء الجزئية بغرض السماح للباحثين للعمل معا في مشروعات وجعل الوصول للمعلومات سهلاً . وقد أنشئ أول موقع يمكن للعامة الوصول إليه عام ١٩٩٣م والآن توجد آلاف المواقع التجارية والتعليمية والترفيهية وغيرها على الشبكة وذلك لبيع وعرض معلومات عن منتجاتها وذلك منذ منتصف التسعينات، ويوجد حالياً ملايين المشتركين في شبكة الانترنت حول العالم . تنقل المعلومات عبر الشبكة في شكل كتل Packets وكل كتلة تنقل بصورة مستقلة ويتم إعادة تجميع الكتل بعد وصولها والتحقق من وصولها بسلام . وتستخدم أجهزة الكمبيوتر لذلك بروتوكول خاص يسمى بروتوكول TCP/IP وهى تعنى بروتوكول التحكم بالنقل عبر الإنترنت وبرنامج هذا البروتوكول موجود ضمن برنامج نظام التشغيل النوافذ windows ، وتطور بعض الشركات حالياً جهاز كمبيوتر صغير جداً يسمح باستخدام أجهزة التليفزيون للوصول إلي المعلومات عبر الإنترنت . وتتضمن الإنترنت خدمات عديدة من أهمها أربعة مجموعات هي :

١- الشبكة العنكبوتية العالمية WWW

وتتضمن ملايين المواقع في شتى العلوم وتستخدم برامج تسمى المتصفحات للتعامل معها مثل برنامج إنترنت أكسبلورر Internet Explorer ـ من شركة مايكروسوفت.

ونود أن تشير هنا إلى إمكانية الاستفادة من شبكة الانترنت في مجالات علمية كثيرة، مثل تحميل برامج كمبيوتر مجانية أو تجارية ، وكذلك الإطلاع على أحدث الأبحاث في المجالات المختلفة، وفي النقاش والمشاركات والمراسلات العلمية. كما توجد بعض المواقع التي تقدم معلومات وبيانات محلية أو دولية في مجالات مختلفة، وبعض المواقع تقدم المساعدة عن بعد مثل خدمة التحليل الإحصائي للبيانات عن بعد بحيث يمكن إدخال البيانات إليها لنحصل على نتائج تحليلها مباشرة. مثال ذلك موقع برنامج Webstatعلى شبكة الانترنت وهو :

www.crunch .com

٢ـ إمكانية جلب وإرسال الملفات Downloading and uploading :

يرمز لهذه الخدمة بالرمز FTP وتعنى بروتوكول جلب الملفات حيث يمكن للمستخدم جلب ملفات متاحة Downloading على كمبيوتر آخر أو إرسال ملفات uploading إلي كمبيوتر آخر متصل بهذه الشبكة حول العالم.

٣ـ البريد الإلكتروني Electronic Mail

يمكن باستخدام البريد الإلكتروني تبادل رسائل تتضمن صوراً ووثائق وبرامج وكذلك رسائل صوتية مسجلة بين مستخدمي الإنترنت بسهولة وسرعة كبيرة جداً وعنوان البريد الإلكتروني يأخذ الشكل التالي :

Whose @ where / what

حيث يوضع عادة مكان Whose اسم صاحب البريد ، ومكان Where اسم الشركة المقدمة لخدمة البريد الالكتروني، أما مكان What فيوضع نشاط الشركة أو الجهة المقدمة لخدمة البريد الالكتروني، فقد تكون مثلاً تجارية نضع الاختصار COM، أو تعليمية EDU أو حكومية GOV أو عسكرية Mil، ويمكن بعد ذلك

وضع اسم الدولة الموجودة فيها الشركة أو الجهة المقدمة للخدمة. ويلاحظ أنه يمكن كذلك الاتصال بين الأشخاص بالصوت والصورة عبر شبكة الإنترنت. وكذلك عقد اللقاءات والندوات والمؤتمرات والاجتماعات والبث المباشر للمحاضرات عبر شبكة الإنترنت.

٤ـ المجموعات الإخبارية:

يمكن لمستخدم الانترنت ومن خلال برامج التصفح الشهيرة الإطلاع والمشاركة بلغات مختلفة في العديد من المجموعات الإخبارية.

لغة جافا JAVA

تعتبر لغة جافا أحد امتدادات إمكانيات الشبكة العالمية. الوثائق الأساسية على الشبكة العالمية تصمم باستخدام لغة تسمى لغة اتش تي ام ال HTML، وهي اختصار عبارة:

Hyper Text Markup Language ، وتمثل HTML إضافة بسيطة لنص الآسكي تسمح بتضمين أوامر تنسيق بسيطة والتي يمكن تفسيرها باستخدام مستعرض شبكة انترنت ، إن لغة جافا JAVA مشابهة للغة سي بلس بلس C^{++} ، كما أنها سهلة التعلم، ويترجم البرنامج المصدر إلى ما يسمي بايت كود ، ويتم قراءته بسهولة في كل ماكينة افتراضية لجافا JAVA Virtual Machine (JVM). هذا وتوجد معلومات عن لغة جافا في موقع:

http://java.sun.com/

ويمكن إجراء تحليل للبيانات عن بعد عبر شبكة الإنترنت خاصة عند عدم توافر أحد حزم برامج التحليل الإحصائي. إذ يمكن فتح أحد المواقع الخاصة بذلك على الإنترنت وإدخال البيانات المراد تحليلها وإرسالها ليتم تحليلها عبر برنامج على هذا الموقع وإعادة نتائج تحليلها مباشرة. مثال ذلك موقع برنامج Webstatعلى شبكة الإنترنت وهو: www.statcrunch.com ولمزيد من التفاصيل عن تحليل البيانات عن بعد عبر الإنترنت انظر .West and Ogden (1997) .

[٤] برامج تحليل البيانات الضخمة

Massive Datasets and SOFTWARE

حزم البرامج السابق عرضها في هذا البحث تعكس التفكير التقليدي لمجموعات البيانات وتحليلها، نقصد من حيث الحجم وأبعاد البيانات. في عام ١٩٩٥م نشر Wegman بحوثا عن الصعوبات الحسابية مع أحجام بيانات وحدود رؤية البيانات وتمثيلها. وقد ناقش Wegman أحجام بيانات مختلفة، من حجم ١٠٠ بايت إلى ١٠١٠ بايت (أي ما يقارب ١٠٠ميجا بايت والبايت يمثل حرفاً أو رقماً)، وحتى فيما هو أكبر من عدة مليارات من البايتات .وأوضح أن كبر أحجام البيانات، والتي تتراكم بطريقة آلية أو شبه آلية، يتعدى الحدود الحسابية وحدود الرؤية للبرامج التقليدية. وقد تبنت وكالة ناسا NASA مشروعاً للتعامل مع مجموعات ضخمة جداً من البيانات والتي تظهر مثلاً في قواعد البيانات المنقولة عبر الأقمار الاصطناعية حول العالم ومن الفضاء الخارجي والمتراكمة مع مرور الزمن .

في هذا البحث نقدم تصور لتحليل البيانات الضخمة، بسهولة سيكون متاحاً للأفراد المشاركة والبحث واستدعاء قواعد بيانات رقمية ووصفية عبر محركات بحث مختلفة على الإنترنت، تماماً كما هو الحال حالياً بالنسبة لبيانات النصوص والصور والصوت. بمعنى أنه سيكون متاحاً مستقبلاً مكتبة بيانات افتراضية عبر الإنترنت. بعض هذه المكتبات موجودة بالفعل ومثال ذلك قواعد البيانات الطبية.

برنامجي METANET and VDADC يقدمان حلولاً لقبول وتوزيع البيانات الضخمة التي تأتي من نظم مراقبة الأرض. أحد البرامج التي تهتم بالتعامل مع البيانات الضخمة هو برنامج VDADC.

برنامج VDADC:

وهو اختصار عبارة **The Virtual Domain Application Data Centre** والتي تعني مركز بيانات تطبيقات المدى الافتراضي، والذي طورته وكالة ناسا للفضاء. هذا البرنامج يقترح طريقة لتنظيم البيانات الضخمة لتقديم طريقة بحث مثلى لأي مجموعة من المستخدمين. البرنامج يعرض عالم البيانات لديه كصورة

مجموعة أشبه بأشجار متفرقة، قد لا يربطها رابط، وهو ما يسمى بقواعد بيانات. ويقوم البرنامج بتحديد البيانات الأقرب للباحث حسب ما يطلبه الباحث.

برنامج METANET

Automated Generation of Metadata

بصورة عامة سنفترض أنه توجد بيانات خام غير مرئية تتضمن ملفاً ومتغيراً ومنظمة. على سبيل المثال نظام الأقمار الاصطناعية يجمع آلاف المجموعات من الصور بنفس شكل الملف وبنفس الوسيلة ولنفس المنطقة الجغرافية. نحتاج بعد ذلك لتصنيف هذه البيانات واستخلاص مجموعة منها تهم العلماء في مجال معين. عملية التصنيف قد تتم بأكثر من طريقة، إما باستخدام طريقة تجريبية أو طريقة إحصائية، أو طريقة مبنية على نماذج معينة. توجد مواقع إنترنت أخرى مفيدة للإحصاء منها مثلا موقع لمكتبة إحصائية وعنوانه هو:

http://lib.stat.cmu.edu

وهو موقع شامل للبيانات والبرامج ومعلومات أخرى عن الإحصاء.

كما أن موقع جامعة Cornell يشير إلى برامج مختلفة في هذا المجال وعنوان هذا الموقع هو:

http://www.stat.cornell.edu/compsites.html

[٥] البرمجة للتغلب على مشاكل تحليل البيانات وتطبيقاتها باستخدام بيانات واقعية:

بعض المشاكل الإحصائية يتعذر حلها مباشرة باستخدام أي من الوسائل السابقة، وهذا المبحث يقدم تطبيقا عمليا لحل بعض مشاكل التحليل الإحصائي من خلال معالجة عدة أهداف من أهمها:

(١) كيف يمكن حل بعض المشاكل الإحصائية التي يتعذر حلها مباشرة باستخدام أي من الوسائل السابقة (حزم البرامج الجاهزة ـ التحليل عن بعد ـ برامج تحليل البيانات الضخمة).

(٢) توضيح كيف يمكن الاستخدام العملي لأحد طرق الذكاء الاصطناعي للتنبؤ بالسلاسل الزمنية وهذه الطريقة سبق نشرها في بحث Alshawadfi(2003)

(٣) المقارنة بين الطريقة المقترحة وطريقة بوكس جينكنز.

(٤) استخدام طريقة الذكاء الاصطناعي المقترحة للتنبؤ ببيانات واقعية.

من المعلوم أن التنبؤ بالقيم المستقبلية لسلسلة زمنية هو أحد أهم مراحل تحليل السلسلة الزمنية. والتنبؤ لسلسلة زمنية يقصد به تقدير القيم المستقبلية لهذه السلسلة بنقطة و/أو بفترة ثقة باستخدام البيانات المتاحة من هذه السلسلة. توجد أكثر من طريقة للتنبؤ بالقيم المستقبلية لسلسلة زمنية من أهمها طريقة بوكس - جينكنز. وطريقة بوكس - جينكنز تتضمن أربع مراحل هي توصيف نموذج مناسب من بين مجموعة نماذج ARIMA ثم تقدير معالم النموذج والتحقق من صلاحيته وأخيرا استخدام النموذج في التنبؤ و/أو التحكم. وهذه الطريقة تستلزم دراية بطرق تحليل انحدار وتحليل إحصائي بصفة عامة، كما تتضمن صعوبات في العمليات الحسابية وفي تفسير النتائج. ونموذج ARIMA يمكن التعبير عنه كما يلي:

$$\phi(B)\,\nabla^d y_t = \theta(B)\,\varepsilon_t \qquad \ldots\ldots\ldots (5\text{-}1)$$

حيث: $\theta(B)$ ، $\phi(B)$ تمثل كثيرات حدود بالشكل التالي:

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) \qquad \ldots\ldots\ldots (5\text{-}2)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta q B^q) \qquad \ldots\ldots\ldots (5\text{-}3)$$

$$= 1 - B \qquad \ldots\ldots\ldots (5\text{-}4)$$

B معامل إزاحة الزمن للخلف حيث:

$$B^d y_t = y_{t-d} \quad , d = 0,1,2,.. , \qquad \ldots\ldots\ldots (5\text{-}5)$$

قد تكون d=0 ، وقد نستبعد $\phi(B)$ بمعنى أن p = 0 وهذا يتضمن أن لدينا نموذج متوسطات متحركة فقط. أما إذا استبعدنا $\theta(B)$ أو وضعنا q = 0 فإننا نحصل على نموذج انحدار ذاتي فقط. ويلاحظ أن قيم كل من p ,d , q في حالات عملية كثيرة لا تتعدى ٢ (انظر Box-Jenkins(1994) صفحة رقم ١١).

ويلاحظ أنه بوضع $z_t = \nabla^d y_t$ في معادلة (٥- ١) والتعويض بالرموز الواردة في المعادلات (٥- ٢) إلى (٥- ٥) يمكن تبسيط نموذج ARIMA في شكل نموذج ARMA(p,q) كما يلي :

$$z_t = \phi_1 z_{t-1} + \ldots\ldots + \phi_p z_{t-p} - \theta_1 \varepsilon_{t-1} - \ldots\ldots - \theta q\, \varepsilon_{t-q} + \varepsilon_t \quad \ldots\ldots (5\text{-}6)$$

حيث :

z_t المشاهدة عند الزمن t أو أي دالة مناسبة فيها . ε_t الخطأ العشوائي

$i = 1,2,3,\ldots,p$, $j=1,2,3,\ldots,q$ ، ϕ_i 's and θ_j's تمثل معالم النموذج

تتطلب عملية التنبؤ باستخدام طريقة بوكس جينكنز ٤٠ مشاهدة على الأقل موزعة على فترات زمنية متساوية .(انظر Pankratz(1983) , p. 297.)، ويلاحظ أنه باستخدام المعادلات (٥- ١) إلى (٥- ٥) يمكن التعبير عن نموذج بوكس جينكنز بالصورة التالية :

$$z_{t+1} = \sum_{j=1}^{\infty} \pi_j \, z_{t+1-j} + \varepsilon_{t+1} \quad \ldots\ldots\ldots (5-7)$$

حيث :

$$\sum_{j=1}^{\infty} \pi_j = 1 \quad \ldots\ldots\ldots\ldots (5-8)$$

ويتم الحصول على الأوزان π_j بمساواة معاملات B المتماثلة في طرفي المعادلة :

$$\phi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots)\ \theta(B) \quad \ldots\ldots\ldots (5\text{-}9)$$

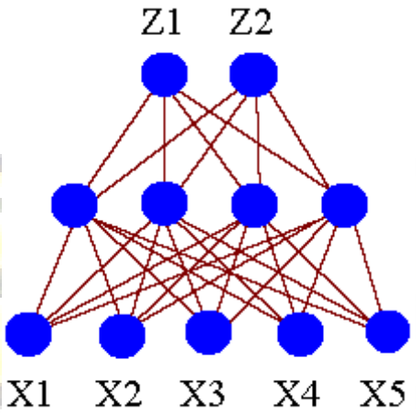
حيث $\theta(B), \phi(B)$ كثيرات حدود سبق تعريفهما في المعادلات (٢-٥)، (٣-٥) . ولمزيد من التفاصيل عن طريقة بوكس- جينكنز انظر على سبيل المثال :
Box and Jenkins (1994) , Lutkepohi(1993) and Harvey(1990) .

من ناحية أخرى، يمكن استخدام تقنيات الذكاء الاصطناعي في التحليل الإحصائي للبيانات. والذكاء الاصطناعي يقصد به طرق لجعل الكمبيوتر يقلد سلوك الذكاء البشري من حيث تفسير الأسباب والتفكير والتعلم. ويتضمن فروعاً كثيرة من أهمها الشبكات العصبية Artificial Neural Networks ـ التنقيب في البيانات DATA MINING ـ الجينات الوراثية GENATIC ALGORITMES ـ النظم الخبيرة EXPERT SYSTEMS . وفي بحث Alshawadfi(2003) تم استخدام تقنية الشبكات العصبية الاصطناعية (ANN) لتقديم طريقة للتنبؤ بالسلاسل الزمنية المولدة من نماذج ARIMA .

والشبكات العصبية تمثل مجموعة نماذج غير خطية مرنة يمكنها التعلم من البيانات، بمعنى أنها يمكنها توصيف العلاقات غير الخطية ـ مهما كانت درجة تعقيدها ـ بين مجموعات بيانات المدخلات والمخرجات للنظام. ويتكون بناء الشبكة العصبية من طبقة مدخلات وطبقة مخفية (أو أكثر) تربط بعلاقة خطية و/أو غير خطية سلسلة المدخلات $x_t , x_{t-1},\dots ,\dots ,x_{t-p}$ مع متجه معالم أو أوزان W من خلال دالة تحويل Y ، والمخرجات تمثل مدخلات لدالة تحويل أخرى للوصول إلى التنبؤات المطلوبة Z's في طبقة المخرجات. كما يتضح في الشكل التالي :

شكل (٥-١)

التنبؤ لسلسلة زمنية باستخدام الشبكات العصبية



تفاصيل كثيرة عن الشبكات العصبية الاصطناعية وطريقة عملها موجودة في مراجع كثيرة منها على سبيل المثال : Arminger and Enache (1995)

يمكن تقريب نموذج (٧ـ٥) للتنبؤ باستخدام طريقة الشبكات العصبية الاصطناعية في شكل نموذج انحدار متعدد غير خطي على الصورة :

$$Z_{t+1} = g(x,w) + \varepsilon_{t+1} \quad \ldots\ldots\ldots\ (5 - 10)$$

حيث :

$z = (z_{t+1}, z_{t+2}, \ldots, \ldots, \ldots, z_{t+h})$ تمثل مخرجات الشبكة أو التنبؤات المطلوبة .

$X = (x_t, x_{t-1}, \ldots, \ldots, \ldots, x_{t-p})$ هي مدخلات الشبكة وهي هنا تمثل المشاهدات السابقة و/أو الأخطاء السابقة .

W هي الأوزان أو الترجيحات وتقابل معالم نموذج بوكس جينكنز

ويمكن النظر إلى نموذج بوكس جينكنز كحالة خاصة من نموذج الشبكات العصبية السابق . وعند التنبؤ بالمتغير z باستخدام المعادلة السابقة ، نتعرض لنوعين من الخطأ هما الخطأ العشوائي ε وخطأ التقريب $\{Z_{t+1} - g(x,w)\}$ ، وهذا الأخير

مساو لخطأ التوصيف في نماذج السلاسل الزمنية . توزيع الخطأ قد لا يمكن تعيينه لنموذج إحصائي محدد ، والتوصيف المعلمي للتقريب g(x,w) يسمى network architecture وهو يمثل توليفة خطية و/أو غير خطية من الدوال . وعادة يمكن تمثيله برسم بياني في شكل دوائر تمثل الدوال الفرعية وأسهم تمثل تحويلات بين هذه الدوال، كما بشكل (٥-١) السابق .

باستخدام أحد تقنيات الذكاء الاصطناعي Artificial intelligent وهي تقنية الشبكات العصبية الاصطناعية Artificial Neural Networks أمكن تقديم إستراتيجية للتنبؤ بالسلاسل الزمنية المولدة من نماذج ARIMA ، وهي منشورة في بحث Alshawadfi (2003) . ولدراسة كفاءة هذه الطريقة في التنبؤ بالسلاسل الزمنية تم مقارنها بطريقة بوكس - جينكنز الشهيرة من خلال دراسة محاكاة تضمنت توليد ١٦٠٠٠ عينة بأحجام ومعالم مختلفة من نماذج ARMA(p,q) وتم التنبؤ لكل عينة بالقيم المستقبلية الثلاث الأولى بطريقتين: طريقة بوكس - جينكنز وطريقة الشبكات العصبية الاصطناعية .

وللتنبؤ بطريقة الشبكات العصبية الاصطناعية تم تدريب الشبكة باستخدام ١٦٠٠٠ عينة سلسلة زمنية مولدة من نماذج انحدار ذاتي مع متوسطات متحركة بمعلمات مختلفة. كما تم اختبار مدى قدرة الشبكة على التنبؤ بالقيم المستقبلية لسلسلة زمنية معطاة باستخدام ١٦٠٠٠ سلسلة زمنية مولدة من نماذج ARMA(p, q) بقيم معالم مختلفة. وقد تم تصميم برنامج كمبيوتر مناسب (MATLAB Toolbox) من خلال حزم برامج MATLAB لإنجاز عملية تدريب الشبكة واختبارها وحساب تنبؤات طريقة بوكس - جينكنز مع المقارنة بين الطريقتين وهذا البرنامج مرفق في ملحق A .

وقد أظهرت النتائج إمكانية استخدام طريقة الشبكات العصبية المقترحة للتنبؤ بالقيم المستقبلية للسلاسل الزمنية، حيث أظهرت الشبكة قدرة كبيرة على التنبؤ بالقيم المستقبلية للسلاسل الزمنية المعطاة وبطريقة أوتوماتيكية وأظهرت نتائج المقارنة أن متوسط مربعات أخطاء التنبؤ (NNF MSD)، وكذلك متوسط

القيمة المطلقة لخطأ التنبؤ (NNF MAD)، أقل من نظيرتها في أسلوب بوكس - جينكنز (BJ MAD , BJ MSD)، وذلك بالنسبة لكل نموذج وللمشاهدات المستقبلية الثلاث الأولى، كما يتضح من جدول (٥-١) التالي:

جدول رقم (٥-١)

p,q Method	1 ,0	2 ,0	0,1	0,2	1 , 1	1, 2	2 ,1	2 , 2	Main average
NNF MSD	3.770	3.102	2.930	3.176	1.886	2.788	4.604	3.829	3.261
BJ MSD	5.000	3.676	2.972	3.234	2.254	3.153	5.234	4.299	3.7280
NNF MAD	2.103	1.957	1.909	1.987	1.523	1.844	2.305	2.143	1.971
BJ MAD	2.363	2.134	1.937	2.013	1.682	1.967	2.465	2.273	2.104
MPE Ratio	3.119	1.124	8.109	9.107	8.130	5.118	4.121	5.118	118. 8

وفي السطر الأخير من جدول (٥-١) تم حساب النسبة MPE Ratio وهي تمثل عدد الحالات التي تكون فيها تنبؤات طريقة الشبكات العصبية أقرب إلى المشاهدة الحقيقية مقسوما على عدد الحالات التي تكون فيها تنبؤات طريقة بوكس - جينكنز أقرب إلى المشاهدة الحقيقية، وهذه النسبة تبين أن تنبؤات طريقة الشبكات العصبية أفضل من تنبؤات طريقة بوكس جينكنز بنسبة تتراوح ما بين ٩ر٧٪ في نماذج ARMA(0,2) إلى ٨ر٣٠٪ في نماذج ARMA(1,1) أو بمتوسط عام للنسبة هو ٨ر١٨٪ مما يشير إلى تفوق التنبؤات بهذه الطريقة على التنبؤات بطريقة بوكس جينكنز.

وفي هذا البحث تم استخدام أحد حزم البرامج الجاهزة وهي حزم برامج MATLAB في تحليل خمسة سلاسل زمنية واقعية والتنبؤ بقيمها المستقبلية. واستخدمنا لذلك طريقتين مختلفتين الطريقة الأولى باستخدام الذكاء الاصطناعي، والثانية طريقة بوكس جينكنز الشهيرة لتحليل نماذج ARIMA. وبيانات السلاسل الخمس موجودة في ملحق C بأخر هذا البحث وهي مأخوذة من المرجع Pankratz (1983) وتتضمن هذه السلاسل بيانات مختلفة في مجال الاقتصاد والأعمال. السلسلة الأولى تمثل معدل الادخار للفرد كنسبة من دخله، أما السلسلة الثانية فتمثل إنتاج الفحم، والثالثة تمثل الرقم القياسي للوحدات السكنية الخاصة

الجديدة مرجحاً بعدد الرخص المحلية الممنوحة، والسلسلة الرابعة تمثل أحجام الشحنات المنقولة بالسكك الحديدية بالبليون طن للميل. وأخيرا السلسلة الخامسة تمثل أسعار الإغلاق بالبورصة لأسهم الشركة الأمريكية للهاتف والتلغراف.

وقد وجد أن نماذج ARIMA المقترحة بطريقة بوكس جينكنز لسلاسل البيانات هذه هي على الترتيب: , ARMA(1,1) , ARMA(2,0), ARMA(1,0) ARMA(2,1) and ARMA(2,2) .

وقد استخدمنا حجم عينة n-3 للتنبؤ بالمشاهدات الثلاث الأخيرة من كل سلسلة، وحسبنا متوسط مجموع مربعات أخطاء التنبؤ وكذلك متوسط مربعات الخطأ المطلق للتنبؤ، كما هي موضحة بالجدول رقم (٥-٢). وفي هذا الجدول يلاحظ أنه ولكل النماذج فإن متوسط مربعات الأخطاء وكذلك متوسط الأخطاء المطلقة أقل دائماً في طريقة الذكاء الاصطناعي عنها في طريقة بوكس جينكنز فيما عدا النموذج ARMA(1,0) .

جدول رقم (٥-٢)

السلسلة الزمنية	النموذج	الطريقة	متوسط مربعات الأخطاء	متوسط الانحرافات المطلقة للأخطاء
معدل الادخار	ARMA(1,0)	NNF	٠.٢٧٢٤	٠.٩٦٤٧
		BJF	٠.٢٦٤٣	٠.٠٩٦٣
إنتاج الفحم	ARMA(2,0)	NNF	١.٤٢٩٢	٢.١٣٤٢
		BJF	١.٧٦١٧	٣.٢٤٢٢
رخص المباني	ARMA(1,1)	NNF	٨.٥١٥٢	٠.١٠٢٦
		BJF	٢٦.٢٦٩٩	١.٢٥٨١
حمولات الشحن بالقطارات	ARMA(2,1)	NNF	٠.٥٧٢٩	٠.٣٣٣٧
		BJF	١.٥٧٢٥	٢.٦٠٨٠
أسعار أسهم شركة AT&T	ARMA(2,2)	NNF	٨.٩٣١٥	٠.٠٩٦٠
		BJF	٥٠.٤٤٤٦	٢.٩٥٧٩

وقد تم تصميم برنامج MATLAB للتنبؤ بالقيم المستقبلية لأي سلسلة زمنية باستخدام طريقة الشبكات العصبية المقترحة بصورة آلية و دون الحاجة إلى خبرة كبيرة أو تفهّم أساليب تحليل رياضية وإحصائية معقدة. وهذا البرنامج مرفق في ملحق B.

[٦] الخلاصة و الاستنتاجات

في هذا البحث استعرضنا أهم حزم البرامج الجاهزة، والتي يمكن استخدامها في تحليل البيانات ، ومزايا وعيوب كل منها ، وأهم المواقع على شبكة الانترنت لحزم البرامج الإحصائية الجاهزة . وكذلك ناقشنا إمكانية تحليل البيانات عن بعد عبر شبكة الانترنت وأهم المواقع والبرامج التي تدعم ذلك. ونظراً لتوافر قواعد بيانات ضخمة – كتلك التي تجمعها الأقمار الاصطناعية بطرية آلية أو شبه آلية - وتعجز البرامج التقليدية عن تحليلها ، فقد ناقشنا مشكلة تحليل مثل هذه البيانات الضخمة، والبرامج المناسبة لتحليلها . كما قدمنا بعض الأمثلة التطبيقية للمشاكل الإحصائية التي يتعذر حلها مباشرة باستخدام أي من الوسائل الثلاث السابقة (حزم البرامج الجاهزة - التحليل عن بعد - برامج تحليل البيانات الضخمة) . وهذه المشاكل تتضمن - وعلى سبيل المثال - مشكلة استخدام طريقة جديدة للتنبؤ بالسلاسل الزمنية باستخدام تقنيات الذكاء الاصطناعي والمنشورة في بحث Alshawadfi(2003) ، واختبار كفاءة هذه الطريقة ومقارنة أدائها مع طريقة بوكس جينكنز الشهيرة، وتوضيح قدرتها على تحليل بيانات واقعية.

وقد استخلص البحث ما يلي :

١. يمكن تحليل البيانات المتاحة لدينا باستخدام برنامج مناسب من حزم البرامج التقليدية وفقا لظروف البحث والبرنامج والباحث .

والبرنامج المناسب لتحليل البيانات يتوقف على عدة عوامل، فالعوامل التي تتعلق بالبحث تشمل طبيعة البيانات وهدف البحث وأما العوامل التي تتعلق بالبرنامج فمنها مدى توافر البرنامج وسهولة استخدامه ووضوح مخرجاته وإمكانيات استيراد البيانات وإمكانيات البرنامج الإحصائية وواجهة المستخدم.

وعوامل متعلقة بالباحث مثل درايته وخبرته بالبرنامج، ومدى استعداده لاستخدام تقنية حديثة واتخاذ قراراته بطرق علمية.

٢. يمكن تحليل البيانات عن بعد عبر شبكة الانترنت من خلال مواقع مختلفة منها مثلاً موقع: www.statcrunch.com

٣. البيانات الضخمة ـ كتلك التي تجمعها الأقمار الاصطناعية بطريقة آلية أو شبه آلية ـ يمكن تحليلها باستخدام حزم برامج تحليل البيانات الضخمة، وهي قليلة العدد وبعضها قيد التطوير. من برامج تحليل البيانات الضخمة برنامجي VDADC و METANET وهذه البرامج متاحة من وكالة NASA الأمريكية لأبحاث الفضاء.

٤. البيانات التي تتطلب طرق حديثة لمعالجتها ويصعب معالجتها باستخدام أي من الوسائل السابقة تتطلب تصميم برامج مناسبة لتحليلها وفق طرق التحليل المطلوبة. ومن خلال الأمثلة التطبيقية، أوضحنا كيف يمكن حل مثل هذه المشاكل الإحصائية التي يتعذر حلها مباشرة باستخدام أي من الوسائل السابقة (حزم البرامج الجاهزة ـ التحليل عن بعد ـ برامج تحليل البيانات الضخمة).

٥. وقد أوضحنا ذلك بحل مشكلة استخدام بعض طرق الذكاء الاصطناعي في تحليل السلاسل الزمنية بتصميم برنامج مناسب MATLAB Toolbox ـ مرفق في ملحق A ـ لتوليد ١٦٠٠٠ سلسلة زمنية من نماذج مختلفة والتنبؤ بقيمها المستقبلية بطريقتي الذكاء الاصطناعي و بوكس ـ جينكنز مع المقارنة بين الطريقتين وتلخيص النتائج.

٦. أظهرت الطريقة المقترحة نتائج منافسة لطريقة بوكس ـ جينكنز للتنبؤ بسلاسل زمنية مولدة من نماذج ARIMA. إذ أن متوسط مربعات الأخطاء، والذي يقيس تباين البواقي، وكذلك متوسط الانحرافات المطلقة أقل بالنسبة لطريقة الشبكات العصبية عنها لطريقة بوكس ـ جينكنز. وكذلك فإن نسبة التنبؤات الأقرب للمشاهدات الحقيقية بطريقة الشبكات العصبية أعلى منها لطريقة بوكس ـ جينكنز (انظر جدول (٥ـ١) في مبحث ٥ السابق.

٧. تم تصميم برنامج أخر ـ مرفق في ملحق B ـ للتنبؤ بالسلاسل الزمنية باستخدام طريقة الذكاء الاصطناعي وتم استخدامه في تحليل بيانات واقعية.

٨. برنامج MATLAB يسمح باستخدام طريقتي الشبكات العصبية وبوكس جينكنز للتنبؤ بالسلاسل الزمنية، في حين أن البرامج الأخرى لا تسمح بذلك حالياً.

والخلاصة أننا نقوم بتحليل البيانات المتاحة لدينا باستخدام حزم البرامج التقليدية أو عن بعد عبر الانترنت، وفي حالة البيانات الضخمة ـ كتلك التي تجمعها الأقمار الاصطناعية ـ نستخدم حزم برامج تحليل البيانات الضخمة. أما البيانات التي تتطلب طرق حديثة لمعالجتها ويصعب معالجتها باستخدام أي من الوسائل السابقة فيجب تصميم برامج مناسبة لتحليلها وفق طرق التحليل المطلوبة.

REFERENCES

1) **Alshawadfi, G. (2003), «A comparison between Neural Network and Box-Jenkins Forecasting Techniques With Application to Real Data», Research Center, Deanship of Scientific Research, King Saud University, King Fahd National Library Cataloging –in – Publication Data, ISBN : 9960-37-522-6 K.S.A.**
2) **Arminger , G. and Enache, D. (1995), «Statistical Models and Artificial Neural Networks», Proceedings of the 19th annuel conference of the Gesellschaft fur classification e.V., University of Basel, March 8-10, 1995, H.-H. Bock .W.Polasek Editors, Springers, Germany.**
3) **Aghadazeh S. M. and Romal, J. B. (1992), «A Directory of 66 Packages for Forecasting and Statistical Analysis», Journal of business forecasting Methods and systems, 8, No.2, 14-20, U.S.A.**
4) **Box, G. E. P. and Jenkins, G. M. (1994) , Time Series Analysis Forecasting and Control , Holden Day, San Francisco, U.S.A.**
5) **Cheng Y. , Karjala , T.W. and Himmelblau , D. M. (1997), «Closed Loop Non-linear Process Forecasting Using Internally Recurrent Nets», Neural Networks, Vol. 10, No. 3, pp. 573-586. Concepts and cases John Wily & Sons New York U.S.A. Day, San Francisco, U.S.A.**
6) **Harvey, A. C. (1990), The Econometric Analysis of Time Series, Philip Allan Publishers Limited, Market Place, Deddincton Oxford Ox 545 E, Great Britain.**
7) **Lutkepohi, H. (1993) , Introduction to Multiple Time Series analysis, second edition, Springer Verlag, Berlin, Heidelberg, Germany.**
8) **Minitab User's Guide Manual (2000), Release 13.3, for windows, Pennsylvania State College, PA 16801: Minitab Inc., 3081 enterprise Dr. U.S.A.**
9) **Morgan, W. T. (1998), «A Review of Eight Statistics Software Packages for General Use», Journal of the American Statistical Association, February, 998 Vol. 52 No.1. U.S.A.**
10) **Ord, K. and Lowe S. (1996), «Automatic Forecasting», Journal of the American Statistical Association, February 1996, Vol. 50 No.1. U.S.A.**
11) **Ouster, R. A. (1998), «An Examination of Five Statistical Software Packages for Epidemiology», Journal of the American Statistical Association, August 1998, Vol. 52. No. 3, U.S.A.**
12) **Pankratz, A. (1983) , Forecasting with univariate Box-Jenkins Models, Publishers Limited, Market Place, Deddincton Oxford Ox 545 E, Great**
13) **Rahlf, T. (1994), «PC Programs for Time Series Analysis, Data Management, Graphics, and univariate Analysis process (SPSS, SYSTAT, STATISTICA, Microtsp, Mesosaur)», Historical – Social Research, 1994, 19, 3(69), 78-123, Germany.**
14) **Small, J. (1997), «SHAZAM 8.0», Journal of Econometric Surveys, vol. 11, no 4., Blackwell, Publishers Ltd., 1997, 108 Cowley rd., Oxford ox4 1JF, U.K. and 350 Main Street, Boston, MA 02148, U.S.A.**
15) **SPSS User's Guide Manual (2008), Release 17.0.0, for windows.**

16) Swanson, N. R. and White, H. (1995), «A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks», The Pennsylvania State University, Department of Economics, University Park, PA, 16802 and Research Group for Econometric Analysis, University of California, San Diego, La Jolla, California 92093

17) Tashman, l. J. and Leach, M. L. (1991), «Automatic Forecasting Software: A Survey and Evaluation», International Journal of Forecasting, 7, pp.209-230, U.S.A.

18) Tsoi , A.C. and Tan , S. (1997), «Recurrent neural networks: A constructive algorithm and its properties», Neurocomputing, 15, pp.309-326. Vol.16 May. 1997, pp.147-163.

19) West, R. W. and Ogden, R. T. (1997), «Statistical Analysis with Webstat, a Java Applet for the World Wide Web», Journal of Statistical Software, 2, 3, U.S.A.

20) Zhang , G. P. (2001), «An investigation of neural networks for linear time series forecasting», Computers and Operations Research 28, 2001, pp1183-1202, Elsevier Science Ltd.

21) Zhang G. P. , Patuwo B. and HU M.(2001), «A simulation study of artificial neural networks for nonlinear time-series forecasting», Computers And Operations Research Vol. 28 (4) pp. 381-396 , Elsevier Science Ltd ,

عناوين مواقع على الانترنت لأهم حزم البرامج الإحصائية

	عنوان الموقع على شبكة الانترنت	حزمة البرامج
22)	http://www.sas.com	ساس SAS
23)	http://is.rice.edu/~radam/prog.html	موقع إضافي لبرنامج ساس
24)	http://www.spss.com	SPSS
25)	http://www.minitab.com	MINITAB
26)	http://www.mathsoft.com	S-plus, MathCAD
27)	http://www.wolfram.com	Mathematica
28)	http://www.maplesoft.com	Maple
29)	http://www.vni.com	IMSL (مكتبة برامج فرعية لطرق الرياضيات والإحصاء)
30)	http://www-c8.lanl.gov/dist_comp2/ MATH/Imsl/imsl_keyword.html	موقع لملاحظات كثيرة عن IMSL
31)	http://www.lightlink.com/datadesk	DataDesk
32)	http://www.mathworks.com	Mat Lab
33)	http://www.sas.com	JMP من شركة ساس

34)	http://www.stat.cornell.edu/compsites.html	موقع جامعة Cornell(برامج مختلفة) :
35)	www.crunch .com	Webstat
36)	http://www.spss.com http://www.ppgsoft.com/bmdp00.html	Systat
37)	http://java.sun.com	موقع عن لغة جافا
38)	http://lib.stat.cmu.edu	مكتبة إحصائية

APPENDIX A
MATLAB MACRO COMPUTER PROGRAM FOR TIME SERIES FORECASTING TRAINING & TESTING USING NEURAL NETWORK TECHNIQUE

```
% ... MATLAB Program for comparison between NNF and BJF.
%... ...file name: train135 ... output file out135.mat..., outf.mat ...
diary ('outf135')
clc;
clear all;
tic
load 'train131'
mu=0 ;  sigma =1; mm=60 ; m= mm-10 ; n=1 ;  m0 =500 ; n1=32 ; n2=8 ; n3=4 ;
h=3; ss01(2*n2,h)=0 ; ss02(2*n2,h)=0 ; ss(n2,h)=0; sb(n2,h)=0; ss03(1,h)=0;
ss04(2*n2,1)=0; ss05(1,h)=0; ss06(2*n2,1)=0; ss3(1,h)=0; ss4(n2,1)= 0;
sb3(1,h)=0; sb4(n2,1)=0 ;  s1=0 ; s2=0 ; s3=0 ; s4=0 ;
p=[1 0;1 0;1 0;1 0;2 0;2 0;2 0;2 0;0 1;0 1;0 1;0 1;0 2;0 2;0 2;0 2;1  1;
1 1;1 1;1 1;1 2;1 2;1 2;1 2;2 1;2 1;2 1;2 1;2 2;2 2;2 2;2 2;2 2];
a=[.3 .5 .7 .9  .3 .3  .5  .7  0  0  0  0    0  0  0    0 .3 .5 .7 .9  .3 .5  .7 .9  .3 .3  .5  .7  .3
.5 .5  .7;
   0  0  0  0 -.5 .5 -.7 -.5  0  0  0  0    0  0  0    0  0  0  0  0  0  0  0  0 -.5 .5 -.7 -.5 -.5
-.7 .3 -5;
   0  0  0  0   0  0   0   0 .3 .5 .7 .9   .3  .3  .5   .7 .5 .3 .5 .7  .3 .3  .5 .5   0  0   0   0  .3
.3 .5  .5;
   0  0  0  0   0  0   0   0  0  0  0  0  -.5  .5 -.7  -.5  0  0  0  0 -.5 .5 -.7 .3  .3 .5  .7  .5  .3
.3 .5  5];
%1... ... ... ... generating DATA ... ... ... ...
for I =  1: m0
z(mm,n1) = 0;
E(mm,1)  = 0.0 ;
E0(m,1)   = 0.0;
E            =  normrnd(mu,sigma,mm,n);
z(1,:)       =  E(1)*ones(1,n1) ;
z(2,:)       = E(2)*ones(1,n1) + a(1,:).*z(1,:)- a(3,:)*E(1);
  for i1    = 3 :mm
z(i1,:) = E(i1)*ones(1,n1) + a(1,:).*z(i1-1,:)+a(2,:).* z(i1-2,:) -a(3,:)*E(i1-1) -
a(4,:)*E(i1-2);
  end
  z00  = z(11:mm,:) ;
  y0   = z00(1:(m-h),:); y1=z00((m-h+1):m ,:) ;
  z001 = 0.8*(z00-ones(m,1)*min(z00))./(ones(m,1)*(max(z00)-min(z00)))+ 0.1 ;
  z0   = z001(1:(m-h),:);
```

```
z1=z001((m-h+1):m ,:) ;
  if  i==1
z01=z0;z02=z1 ;
z2=z00;
y01=y0;y02=y1 ;
  else
    z01=[z01 z0] ;
z02=[z02 z1] ;
z2=[z2 z00] ;
y01=[y01 y0] ;
y02=[y02 y1] ;
  end
end
%2... ... ... ...comparison and testing phase ... ... ... ... ... ...
j00=0
for j       = 1:m0
  for j0    = 1:n1
    j00    =  j00+1
    j1     =  fix((j0-1)/n3)+1;
     s     =  simuff(z01(:,j00),w1,b1,'logsig',w2,b2,'logsig');
    TH     =  ARMAX(y01(:,j00),[p(j0,:)]);
    YP     =  PREDICT([y01(:,j00);y02(:,j00)] , TH , m-h)
    y      = (s - 0.1)*((max(z2(:,j00))- min(z2(:,j00)))/0.8) + min(z2(:,j00));
    for j3 = 1:h
      s01 = abs(y02(j3,j00)-y(j3));
      s02 = (s01)^2;
      b01 = abs(y02(j3,j00)-YP(m-h+j3));
      b02 = (b01)^2;
      ss01((2*j1-1),j3) = ss01((2*j1-1),j3)+ s01;
      ss01((2*j1  ),j3)  = ss01((2*j1  ),j3)+ b01;
      ss02((2*j1-1),j3) = ss02((2*j1-1),j3)+ s02;
      ss02((2*j1 ) ,j3)  = ss02((2*j1  ),j3)+ b02;
      if s01 < b01   ;   ss(j1    ,j3) = ss(j1,j3)+1 ;else;sb(j1 ,j3)= sb(j1,j3)+1 ;end
    end
  end
end
ss03 = (ones(1,2*n2)* ss01)   /(n2*n3*m0);
ss04 = (ss01      * ones(h,1))/(h *n3*m0);
ss05 = (ones(1,2*n2)* ss02)   /(n2*n3*m0);
ss06 = (ss02      * ones(h,1))/(h *n3*m0);
```

```
ss3   = (ones(1,n2)* ss)       /(n2*n3*m0);
ss4   = (ss      * ones(h,1))  /(h *n3*m0);
sb3   = (ones(1,n2)* sb)       /(n2*n3*m0);
sb4   = (sb      * ones(h,1))  /(h *n3*m0);
s1    = sum(ss03)/h;
s2    = sum(ss05)/h;
s3    = sum(ss3) /h;
s4    = sum(sb3) /h;
%3...Results ... ...
disp ' MAE RESULTS'
Mabs = [ss01,ss04;[ss03,s1]]
disp ' MSE RESULTS'
Mmse = [ss02,ss06;[ss05,s2]]
disp ' NNF RATIOS RESULTS'
Mnnf = [ss/(n3*m0),ss4;[ss3,s3]]
disp ' BOX-JENKINS RATIOS RESULTS '
Mbox = [sb/(n3*m0),sb4;[sb3,s4]]
save out135
diary off  ;
toc
```

APPENDIX B

MATLAB MACRO COMPUTER PROGRAM FOR TIME SERIES FORECASTING USING NEURAL NETWORK & B-J TECHNIQUES

```
%... ...file name : train131 ... output file out13.mat...
clc
clear all
load 'train131'
diary ('outf13')
tic
m0 =1;n1=1;n2=1;n3=1;h=3;j00=1;j1=1;j0=5
ss01(2*n2,h)=0;ss02(2*n2,h)=
0;ss(n2,h)=0;sb(n2,h)=0;ss03(1,h)=0;ss04(2*n2,1)=0;
ss05(1,h)=0;ss06(2*n2,1)=0;ss3(1,h)=0;ss4(n2,1)=
0;sb3(1,h)=0;sb4(n2,1)=0;s1=0;s2=0;s3=0;s4=0;
p=[1 0;2 0;0 1;0 2;1 1;1 2;2 1;2 2];
%1... ... ... ... PREPARING THE DATA ... ... ... ...
%      … DATA1     SAVING RATE
 % … DATA2     COAL PRODUCTION
 % … DATA3     HOUSING PERMITS
 % … DATA4     RAIL FREIGHT
 % … DATA5     AT & T STOCK PRICE
  z00 = data1;
  m1  = length(z00)
  if m1 >50
     m=50;y0 = z00(m1-m+1:(m1-h),:); y1=z00((m1-h+1):m1,:);
  else
     m=m1;
     y0   = z00(1:(m-h),:); y1=z00((m-h+1):m ,:) ;
 end
  z0 = 0.8*(y0 - min(y0))/(max(y0)-min(y0))+ 0.1 ;
%2... FORECASTING WITH NUERAL & BJ  METHODS ...
      s   = simuff(z0,w1,b1,'logsig',w2,b2,'logsig');
     TH  = ARMAX(y0,[p(j0,:)]);
     YP  = PREDICT([y0;y1] , TH , m-h)
     y   =(s - 0.1)*((max(y0)- min(y0))/0.8) + min(y0);
     plot(1:m1,z00)
     acf=xcorr(z00,'coeff')
     acf1=acf(m+1:m+25)
```

```
    newplot
    plot(1:25,acf1)
    ac=xcorr(y0,'coeff')
    ac1=ac(m+1:m+25)
    hold on
    plot(1:25,ac1)
    for j3=1:h
      s01 = abs(y1(j3)-y(j3));
      s02 = (s01)^2;
      b01 = abs(y1(j3)-YP(m-h+j3));
      b02 = (b01)^2;
      ss01(1,j3)= ss01(1,j3)+ s01;
      ss01(2,j3)= ss01(2,j3)+ b01;
      ss02(1,j3)= ss02(1,j3)+ s02;
      ss02(2,j3)= ss02(2,j3)+ b02;
      if s01 < b01
                ss(1,j3)= ss(1,j3)+1 ;
              else
              sb(1,j3)= sb(1,j3)+1 ;
              end
    end
  end
end
ss04 = (ss01     * ones(h,1))/(h *n3*m0);
ss06 = (ss02     * ones(h,1))/(h *n3*m0);
ss4  = (ss     * ones(h,1))  /(h *n3*m0);
sb4  = (sb     * ones(h,1))  /(h *n3*m0);
 %3...Results ... ...
disp ' MAE RESULTS'
Mabs = [ss01,ss04]
disp ' MSE RESULTS'
Mmse = [ss02,ss06]
disp ' NNF RATIOS RESULTS'
Mnnf = [ss/(n3*m0),ss4]
disp ' BOX-JENKINS RATIOS RESULTS '
Mbox = [sb/(n3*m0),sb4]
save out136;toc;;diary off
```

APPENDIX C بيانات واقعية لخمس سلاسل زمنية مختلفة		
NO.	series	DATA
1	معدل الادخار	7; 7.6 ;7.2;7.5;7.8;7.2;7.5;5.6;5.7;4.9;5.1;6.2;6;6.1; 7.5;7.8;8;8;8.1;7.6;7.1;6.6;5.6;5.9;6.6;6.8;7.8;7.9;8.7;... 7.7;7.3;6.7;7.5;6.4;9.7;7.5;7.1;6.4;6;5.7;5;4.2;5.1;... 5.4;5.1;5.3;5;4.8;4.7;5;5.4;4.3;3.5
2	إنتاج الفحم	47730;46704;41535;41319;36962;32558;31995;32993;44834;29883;... 39611;40099;38051;36927;37272;39457;38097;40226;43589;39088;... 39409;37226;34421;34975;32710;31885;32106;30029;29501;31620;... 34205;32153;32764;33230;35636;35550;34529;37498;37229;36021;38281;... 36676;44541;40850;38404;37575;41476;42267;43062;45036;43769;42298;... 44412;40498;37830;42294;38330;43554;42579;36911;42541;42430;43465;... 44468;43597;40774;42573;41635;39030;41572;37027;34732;36817;34295;... 33218;32034;31417;35719;30001;33096;35196;36550;33463;37195;34748;... 36461;35754;36943;35854;37912;30095;28931;31020;31746;34613;37901
3	رخص المباني	83.3;83.2;105.3;117.7;104.6;108.8;93.9;86.1;83;102.4;119.6;141.4; 158.6;161.3;158.2;136.1;121.9;97.7;103.3;92.7;106.8;102.1;110.3; 114.1;109.1;105.4;97.6;100.7;102.7;110.9;120.2;131.3;138.9;130.9; 123.1;110.8;108.8;103.8;97;93.2;89.7;89.9;90.2;89.6;85.8;96.9;112.7; 122.7;119.8;117.4;111.9;104.7;98.3;94.9;93.3;90.9;91.9;97.2;104.7; 107.7;108.2;110.7;113.2;114.6;122.2;120.2;122.1;126.6;122.3;115.9; 116.9;110.1;110.4;108.9;112.1;117.6;112.2;96;78;66.9;83.5;95.8; 107.7;113.7
4	السكة الحديد	166.8;172.8;178.3;180.3;182.6;184.2;188.9;184.4;181.7;178.5; 177.6;181;186.5;185.7;186.4;186.3;189.3;190.6;191.7;196.1;189.3; 192.6;192.1;189.4;189.7;191.9;182;175.7;192;192.8;193.3;200.2; 208.8;211.4;214.4;216.3;221.8;217.1;214;202.4;191.7;183.9;185.2; 194.5;195.8;198;200.9;199;200.6;209.5;208.4;206.7;193.3;197.3; 213.7;225.1
5	اسعار أسهم شركة AT&A	61;61.625;61;64;63.75;63.375;63.875;61.875;61.5;61.625;62.125; 61.625;61;61.875;61.625;59.625;58.75;58.75;58.25;58.5;57.75; 57.125;57.75;58.875;58;57.875;58;57.125;57.25;57.375;57.125;57.5; 58.375;58.125;56.625;56.25;56.25;55.125;55;55.125;53;52.375; 52.875;53.5;53.375;53.375;53.5;53.75;54;53.125;51.875;52.25