

نمذجة متسلسلة اعداد المصابين بالتهاب اللوزتين في محافظة نينوى للمدة من 2004-2009

أ.د. ظافر رمضان البدراني / كلية علوم الحاسوب والرياضيات / جامعة الموصل
م.م. عمر سالم ابراهيم كلية علوم الحاسوب والرياضيات / جامعة الموصل
م.احصائي نورا سهيل اسماعيل / كلية علوم الحاسوب والرياضيات / جامعة الموصل

المستخلص

في هذا البحث سيتم التعامل مع ظاهرة صحية لها اثر كبير على مختلف الفئات العمرية في المجتمع الا وهي ظاهرة التهاب اللوزتين حيث ستتم نمذجة اعداد المصابين بهذه الظاهرة كاتنموذج انحدار حركي من خلال توفيق أنموذج انحدار ذاتي ومتوسطات متحركة موسمي ARMA Seasonal باستخدام منهجية بوكس جنكنز لاعداد المصابين بالتهاب اللوزتين في مدينة الموصل وللفترة 2004-2009 مع التنبؤ بهذه الظاهرة لاثنى عشر شهرا قادما ، وقد وجد ان الانموذج الافضل في تمثيل بيانات الظاهرة هو SARMA $(2,1)_{12}^*(1,1)$ من جانب اخر تمت النمذجة باستخدام المتغيرات التوضيحية المؤثرة على ظاهرة البحث وهي درجة الحرارة العظمى ودرجة الحرارة الصغرى والسطوع الشمسي والتبخر ومعدل الامطار وسرعة الرياح، الضغط الجوي، ولنفس المدة السابقة حيث امكن نمذجة اعداد المصابين بالتهاب اللوزتين باعتماد طريقة الانحدار الخطي المتعدد ، وقد اعطت منهجية بوكس جنكنز نتائج افضل في التنبؤ من الانحدار الخطي المتعدد بالاعتماد على معيار متوسط الاخطاء المطلقة MAE، وهذه النتائج يمكن ان تساعد متخذي القرار على وضع الاستراتيجيات المناسبة في التعامل مع هذا المرض لتقليل اعداد المصابين وتهيئة مستلزمات العلاج الملانمة.

المصطلحات الرئيسية للبحث / ARIMA ، بوكس جنكنز والانحدار الخطي المتعدد والمكونات الرئيسية.



1- المقدمة Introduction

لقد خصص الاحصاء مجالا منفردا لتحليل البيانات الزمنية يعرف بمجال المتسلسلات الزمنية والذي تطور تطورا كبيرا اعتماداً على المنهجية الحديثة التي قدمها الباحثان بوكس وجنكنز في سنة 1976 والتي يمكن عدّها بحق البداية الحقيقية لتحليل المتسلسلات الزمنية في المجال الزمني Time Domain. حيث يمكن رصد المتسلسلات الزمنية في شتى انواع المعرفة وميادين التطبيق المختلفة مثل الاقتصاد والتعليم و الارصاد الجوية والبيئة و الطب وغيرها .

تعرف العملية التصادفية stochastic process بانها عائلة من المتغيرات العشوائية $\{y(t), t \in T\}$ حيث t هو الدليل $(t \in T)$ و T مجموعة دليلية (index set) . اذا كانت المجموعة الدللية T تمثل الزمن فان العملية التصادفية تسمى عندئذ متسلسلة زمنية Time Series ، وان T اما ان تأخذ فيما مستمرة بحيث ان $T = \{t; -\infty \leq t \leq \infty\}$ فيطلق على المتسلسلة الزمنية بانها " مستمرة الزمن " ويرمز لها عادة $\{y(t); -\infty \leq t \leq \infty\}$ او للاختصار $y(t)$. وامثلة على ذلك المتسلسلات الزمنية للموجات الصوتية والكهربائية وغيرها . اما اذا كانت T تأخذ قيما متقطعة $T = \{t; 0, \pm 1, \pm 2, \dots\}$ فان المتسلسلة الزمنية تسمى عندئذ " متقطعة الزمن " ويرمز لها عادة $\{y(t); t = 0, \pm 1, \dots\}$ او للاختصار $\{y_t\}$ وامثلة على ذلك المتسلسلات الزمنية المناخية والاقتصادية وغيرها . انظر (الخياط ، 1991).

ويعد التنبؤ بالقيم المستقبلية للمتسلسلة الزمنية احد الاهداف الرئيسية لتحليل المتسلسلات الزمنية، ولقد تم تطوير العديد من طرائق التنبؤ وتقييم كفاءتها (cordeivo and neves, 2009). وتعد نماذج ARIMA اداة تنبؤية مهمة ، وهي اساس للعديد من الافكار الرئيسية في تحليل المتسلسلات الزمنية . ان مختصر ARIMA يعبر عن نماذج الانحدار الذاتي والمتوسطات المتحركة التكاملية ويسمى احيانا نماذج بوكس جنكنز ، ولكن المعروف تقنياً منهجية ARIMA (mahsin et al., 2012) . يجزا انموذج المتسلسلة الزمنية الى ثلاث مكونات او عناصر تسمى مرشحات خطية اولها مرشح السكون (التكامل) (I) Autoregressive Filter (AR) ومرشح المتوسطات المتحركة (MA) Moving Average Filter (MA) والهدف من استخدام هذه المرشحات الثلاثة هو الحصول في النهاية على التغيرات العشوائية البحتة التي لا يمكن التنبؤ بها (فاندل، 1983). لذلك فان الانموذج العام المختلط من المرتبة العليا لمتسلسلة غير موسمية ARIMA (p,d,q) يكتب على وفق المعادلة الاتية

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \dots (1)$$

حيث ان p : تمثل مرتبة الانحدار الذاتي ، q : مرتبة المتوسطات المتحركة ، I : تمثل درجة الفروق وباستخدام مؤثر الازاحة للخلف (backwardshift operator) فان المعادلة (1) تكتب بالصيغة الاتية :

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad \dots (2)$$

تتميز منهجية بوكس- جنكنز بقابليتها على التعامل بشكل سلس مع المكون الموسمي الذي تتميز به العديد من الظواهر في الحياة ويشار للنماذج في هذه الحالة SARMA ، علاوة على ذلك فان العديد من الظواهر يتعدى تمثيلها الا من خلال النماذج المنضربة التي تتعامل مع المكون الموسمي وغير الموسمي في الوقت نفسه ويشار لها بـ (P,D,Q)S ARIMA (p,d,q) ، حيث P مرتبة AR الموسمي ، Q مرتبة MA الموسمي ، D درجة الفرق الموسمي ، S طول المدة الموسمية (Makridakis et al.,1998) وباستخدام مؤثر الازاحة للخلف فان النموذج يكتب بالصيغة الاتية :

$$\phi(B)\phi_s(B^s)(1-B)^d(1-B^s)^D y_t = \theta(B)\theta_s(B^s)a_t \quad \dots (3)$$

ان نماذج بوكس - جنكنز تسمى ايضا نماذج الانحدار الحركي Dynamic Regression (Pankeatz,1991) . اما الطريقة الثانية في التنبؤ هي تحليل الانحدار وهو عبارة عن اسلوب تنبؤ مبني على علاقة سببية، وهذا يعني انه يجب تحديد تلك المتغيرات التي تؤثر في المتغير المراد التنبؤ له . ان احدى الفوائد المهمة للانحدار المتعدد بالنسبة للأساليب الاخرى هي انه بالامكان بناء عدد من العلاقات السببية المختلفة التي يمكن اقتراحها واختيارها لغرض استخدامها في التنبؤ (الوردي ،1990).

2- هدف البحث

يهدف البحث الى نمذجة المتسلسلة الزمنية الخاصة باعداد المصابين بالتهاب اللوزتين في محافظة نينوى من خلال منهجية بوكس - جنكنز (الانحدار الحركي) اولا والتي من خلالها يتم بناء انموذج يعد ممثلا جيدا لبيانات الظاهرة محل الدراسة ويتم استخدامه للتنبؤ لاثنا عشر شهرا قادمة، كذلك تتم النمذجة باستخدام طريقة الانحدار الخطي المتعدد اعتمادا على علاقات السبب والتأثير وبمقارنة النتائج التنبؤية التي يتم الحصول عليها من خلال الطريقتين فان ذلك يمكن ان يساعد متخذ القرار على وضع الاستراتيجيات المناسبة في التعامل مع هذا المرض لتقليل اعداد المصابين وتهينة مستلزمات العلاج الملائمة في الفترات الزمنية المناسبة.

3- عينة البحث

لقد تم الحصول على البيانات الخاصة بأعداد المصابين بمرض التهاب اللوزتين شهريا والخاصة بمستشفيات الطوارئ الموجودة في محافظة نينوى للفترة 2004-2009 (النابل،2011) وتم استبعاد (12) مشاهدة لعام 2009 من عملية النمذجة لمقارنتها مع القيم المتنبأ بها وحيث ان مرض التهاب اللوزتين يتأثر بمتغيرات اخرى مثل درجة الحرارة العظمى (م) ودرجة الحرارة الصغرى (م) والسطوع الشمس (ساعة) ومعدل الامطار (m.m) والرطوبة النسبية (%) ومعدل التبخر (m.m) والضغط الجوي (باسكال) (pa) وسرعة الرياح (م/ثا) فقد تم الحصول على بيانات المعدلات الشهرية لهذه المتغيرات وللمدة نفسها من هيئة الانواء الجوية/الرشيدية في محافظة نينوى.

3- منهجية بوكس جنكنز The Box –Jenkins Methodology

ان المراحل الاساسية في منهجية بوكس جنكنز هي التعرف Identification والتقدير estimation والفحوص التشخيصية Diagnostic Checking ثم التنبؤ Prediction. ان المرحلة الاولى والاكثر اهمية في بناء الانموذج هي مرحلة التعرف على الانموذج الملائم لتمثيل المتسلسلة الزمنية بشكل اولي بعد التحقق من كون المتسلسلة الزمنية مراوحة في الوسط والتباين سواء لمكونها الموسمي او غير الموسمي ، وتعد دالة الارتباط الذاتي ACF مقياس لمقدار الاعتماد الخطي Linear dependence بين مشاهدات المتسلسلة الزمنية والتي يفصل بينها تخلف مقداره k كما ان الارتباط الذاتي الجزئي PACF التي تمثل مقياس لدرجة الاقتران بين Y_t و Y_{t+k} عندما تكون التأثيرات للمتغيرات Y والمتمثلة بـ $1,2,3,\dots,k-1$ مستبعدة ، حيث رسمها يساعد على تحديد عدد حدود الانحدار الذاتي الضرورية لظهور واحد او اكثر من المميزات الاتية : التخلفات الزمنية التي تظهر ارتباطات مرتفعة وكذلك التخلفات الزمنية الموسمية للمتسلسلة، وايضا الاتجاه العام اما في مستوى المتوسط اوفي تباين للمتسلسلة (mahsin et al.,2012)

ان كلتا الدالتين هما ادوات مهمة في تحديد الانموذج الملائم والذي يعد في هذه المرحلة انموذجا اوليا ولغرض سرعة تحديد الانموذج الاكثر ملائمة يمكن توفير العديد من النماذج القريبة من الانموذج الاولي وحساب قيمة احد معايير المعلومات واختيار الانموذج الذي يمتلك اقل قيمة لمعيار المعلومات المستخدم بوصفه الانموذج الاقرب الى تمثيل حقيقة الظاهرة المدروسة . ان بعض المعايير شائعة الاستخدام مثل معيار معلومات اكاكي **Akaike Information Criteria (AIC)** ومعيار شوارز-بييز **schwarz-bayes (BIC)** ومعيار هانان كوين **Hannan Quinn (HQ)** والتي في حالة المعادلة المنفردة تشترك بتركيب بسيط $C(m) = \ln(\sigma^2) + f(m, n)$ حيث σ^2 مقدر الامكان الاعظم لتباين الخطا وكل من $(2\ln(Ln(n))m/n, (2m/n), \ln(n)m/n)$ هي حد الجزاء $f(m, n)$ لكل من **AIC** و **HQ** و **BIC** على التوالي ، حيث m عدد المعلمات و n حجم العينة المستخدم ، وفي بحثنا هذا سيتم استخدام معيار شوارز-بييز **BIC** والذي يعرف ايضا بالرمز **SIC** (Balcombe, 2005) . ان المرحلة الثانية هي تقدير معلمات الانموذج الاولي والفحوص التشخيصية ، ويتم تقدير معالم نماذج **ARIMA** باستخدام طريقة المربعات الصغرى . والطريقة الاخرى التي تستعمل كثيرا هي الامكان الاعظم **Maximum likelihood** ، حيث الامكان لمجموعة من البيانات يشار له بالرمز L وهو التناسب لاحتمال الحصول على البيانات معطى النموذج ، وبذلك فهو مقياس لقابلية امكان ملاحظة عينة المشاهدات معطى المجموعة الخاصة من قيم المعلمات ، ان طريقة الامكان الاعظم تجد قيم المعلمات التي تعظم الامكان ، وهذه الطريقة هي المفضلة من قبل الاحصائيين لامتلاكها بعض الخصائص الاحصائية المرغوبة (Makridakis et al.,1998)(mahsin et al.,2012) .

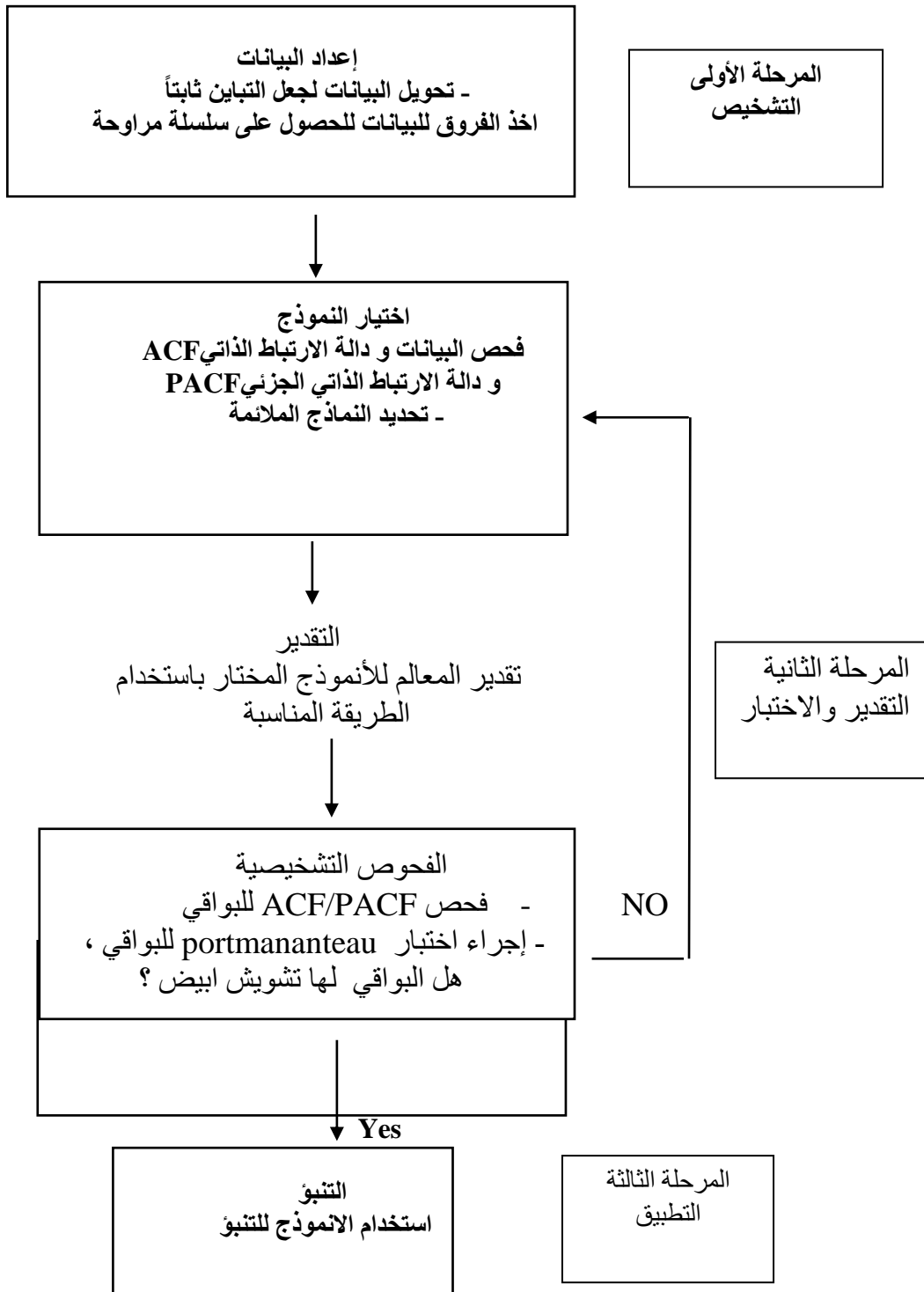
الخطوة التالية بعد التعرف وتقدير المعلمات هي الفحوص التشخيصية ويتم ذلك باختبار معنوية معلمات الانموذج حيث ان المعلمة تعد معنوية اذا كانت القيمة المطلقة لها اكبر من ضعف الخطأ المعياري او ان قيمة **P-value** للمعلمة اقل من مستوى المعنوية المعتمد في الدراسة ومن الفحوص التشخيصية الاساسية اختبار عشوائية البواقي باستخدام اختبار **portmananteau** والذي صمم من قبل الباحثين (Liung and Box,1978) حيث تأخذ الاحصاء الصيغة التالية :

$$Q^* = n(n+2) \sum_{k=1}^h \frac{r_k^2}{n-k} \quad \dots (4)$$

التي تمتلك توزيع يقترب من توزيع مربع كاي بدرجة حرية $(h-m)$ ، حيث h اعلى تخلف تتم دراسته ، n عدد مشاهدات المتسلسلة ، r_k الارتباط الذاتي عند التخلف k ، m عدد معلمات الانموذج الملاءم لتمثيل البيانات ، فاذا كانت قيمة Q اقل من قيمة مربع كاي الجدولية عند درجة حرية $h-m$ ومستوى معنوية α عند ذلك تقبل فرضية عدم بان البواقي عشوائية والعكس صحيح .

وفي حالة فشل الانموذج الاولي الذي تم ترشيحه باجتياز هذه الفحوص فيتم الرجوع الى المرحلة الاولي وترشيح انموذج اخر ، وفي هذه الحالة يمكن الاستفادة من دراسة نمط البواقي للانموذج الاولي في تحديد الانموذج التالي (فاندل، 1983). وان الانموذج الذي يجتاز الفحوص التشخيصية يمكن اعتباره انموذجا ملائما لتمثيل الظاهرة المدروسة .

المرحلة الثالثة التنبؤ وهي احد اسباب شيوع استخدام نماذج **ARIMA** لنجاحها في التنبؤ للعديد من الحالات والتطبيقات ، حيث انها تعطي نتائج تنبؤ جيدة ومتطابقة مع البيانات الاصلية (Gujarati,2004)(Makridakis et al.,1998) والمخطط التالي يوضح منهجية بوكس-جنكنز بمراحلها الثلاثة



الشكل (1) منهجية بوكس جنكيز

4- الانحدار الخطي المتعدد multiple linear regression

ان تحليل الانحدار عبارة عن وسيلة احصائية يستخدم لتحليل العلاقة بين متغير توضيحي واحد او اكثر (explondrtay variables) ومتغير تابع او الاستجابة (dependent variable) ويعتبر تحليل الانحدار من اكثر الطرق الاحصائية استعمالا في مختلف العلوم لانه يصف العلاقة بين المتغيرات على هيئة معادلة فالمعادلة التي تضم عدة متغيرات توضيحية تسمى معادلة الانحدار الخطي المتعدد multiple linear regression equation (الراوي، 1987) وتكتب كالآتي :

$$Y_i = B_o + B_1 X_{i1} + B_2 X_{i2} + \dots + B_m X_{im} + e_i \quad i=1,2,3 \dots n \quad (5)$$

حيث ان Y_i : تمثل قيمة متغير الاستجابة response variable في المشاهدة i $X_{i1}, X_{i2}, \dots, X_{im}$ هي قيم ثابتة لـ m من المتغيرات التوضيحية e_i : يمثل حد الخطأ error term في المشاهدة i ، β_o ثابت احصائي يمثل موقع تقاطع مستوى الانحدار مع المحور الصادي، β_i معامل الانحدار الجزئي لـ y على X_i عند جعل قيمة بقية المتغيرات التوضيحية ثابتة.

ان تحليل الانحدار هو عبارة عن اسلوب تنبؤ مبني على علاقة سببية بمعنى يجب تحديد تلك المتغيرات التي ويعتقد بانها تؤثر في متغير الاستجابة المراد التنبؤ له ، وان احدى الفوائد المهمة للانحدار المتعدد بالنسبة للأساليب الاخرى هي انه بالإمكان بناء عدد من العلاقات السببية المختلفة التي يمكن افتراضها واختبارها واختيار افضلها لغرض استخدامها في التنبؤ. (الوردي، 1990). ويمكن اجراء التنبؤات في جميع المجالات اليومية ومن الامثلة على ذلك التنبؤ بالطقس وتوقعات المبيعات وتوقعات المحاصيل والتنبؤ بسعر الوقود ، اذ ان بعض التوقعات اكثر دقة من غيرها تبعا لقوة العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة (Bluman,2009).

يتم تقدير معلمات النموذج الانحدار الخطي المتعدد باستخدام طريقة مربعات الصغرى التي تتميز بانها تختار احسن نموذج مطابق للبيانات بحيث تجعل مجموع مربعات الخطأ او المتبقي اقل ما يمكن ، ويعطي جدول تحليل التباين ANOVA امكانية اختبار فرض العدم القائل بان المتغيرات التوضيحية ليس لها تاثير على متغير الاستجابة وذلك من خلال مقارنة قيمة F المحسوبة مع F الجدولية عند مستوى معنوية محدد ودرجة حرية معينة $F(m,n-m-1, \infty)$. من جانب اخر فان قيمة معامل التحديد المتعدد (multiple coefficient of determination) والتي يرمز لها R^2 تمثل تلك النسبة من مجموع التباين او الاختلافات المفسرة من قبل المتغيرات التوضيحية تحت الدراسة وان قيمة R^2 تتراوح بين الصفر والواحد ، وعندما يكون النموذج مطابقا للبيانات فان قيمة R^2 تقترب من الواحد اي ان القيم المشاهدة y_i والمتوقعة \hat{y}_i تكون متقاربة جداً (الراوي، 1987).

5- تحليل المكونات الرئيسية principal component analysis

يهدف التحليل الى ايجاد عوامل factors او توليفات خطية linear combination بعدد المتغيرات) يسمى بالمكونات الرئيسية (يقوم الباحث باخذ m منها بحيث $m < n$ وتكون m مؤهلة لتفسير معظم التباين الكلي للقيم الاصلية وتكون هذه المكونات الرئيسية متعامدة اي لا يوجد ارتباط فيما بينها وتكتب المكونات الرئيسية pc_{ij} كالآتي

$$pc_{ij} = a_{1j} X_{i1} + a_{2j} X_{i2} + \dots + a_{nj} X_{in} \quad \dots(6)$$

حيث ان pc_{ij} : القيمة (i) للمكون الرئيسي (j)

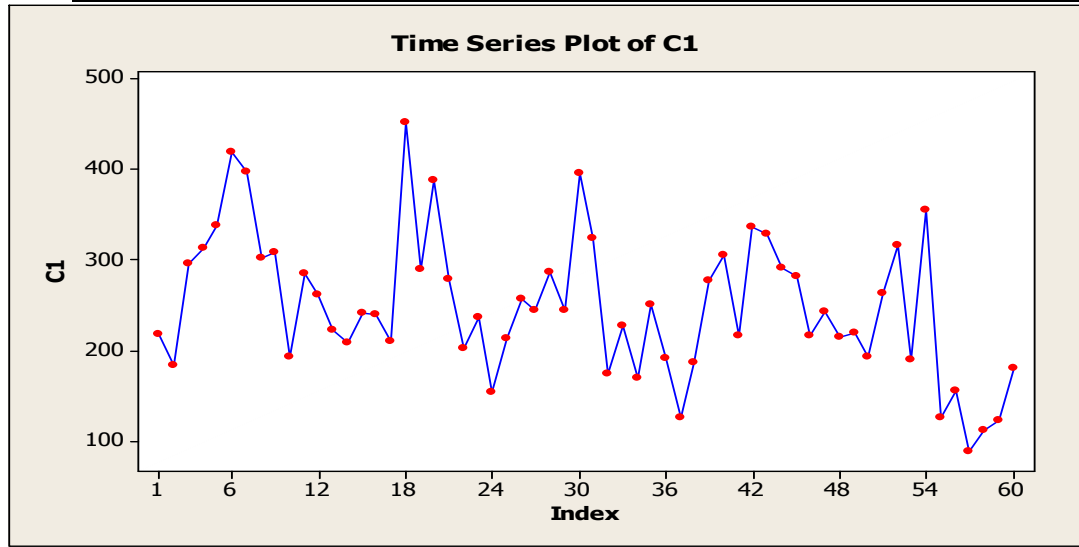
ان ajz تنتخب بحيث تجعل تباين المكون الرئيسي الاول الى مجموع التباين الكلي اكبر ما يمكن ، وبعبارة اخرى ان تباين المكون الرئيسي الاول pc_1 هو اكبر من تباين اي مكون رئيسي اخر وان دالة المكون الرئيسي الثاني له تبايناً اقل من تباين المكون الرئيسي الاول ولكن اكبر من تباين اي مكون رئيسي اخر وهكذا فاذا كان لدينا n من المتغيرات x فان المكون الرئيسي pc_1 هو توليفة خطية من n من المتغيرات الاصلية وله معاملات ajz مساوية لقيم المتجه المميز القياسي التابع لأكبر جذر مميز (λ_1) وان المكون الرئيسي الثاني pc_2 هو أيضاً توليفة خطية من n من المتغيرات الاصلية ومعاملاتها تساوي قيم المتجه المميز القياسي التابع لثاني اعلى من جذر مميز (λ_2) .

وهكذا نلاحظ بان $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$ ان كل مكون رئيسي له تباين مساو لقيمة الجذر المميز التابع له اي ان
..... $(7) S^2(p c_i) = \lambda_i$

ونلاحظ انه من الممكن ايجاد مكونات رئيسية بعدد المتغيرات الاصلية X ولكن علمياً فان معظم التباين الكلي يفسر من قبل مكونات رئيسية قليلة (الراوي، 1987)

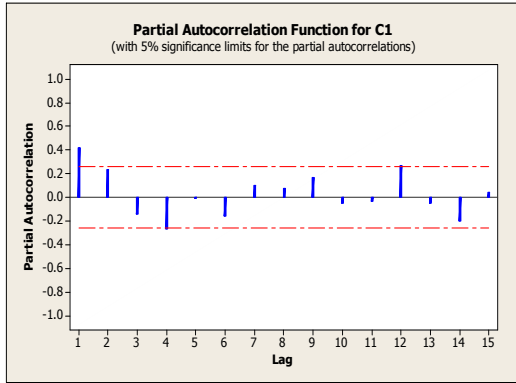
5- مناقشة النتائج

ب. نمذجة والتنبؤ باعداد المصابين بالتهاب اللوزتين باستخدام منهجية بوكس جنكينز

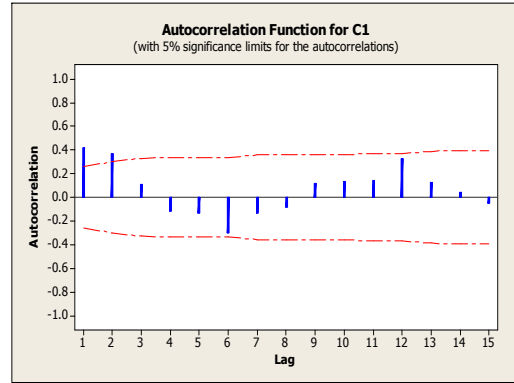


شكل (2) تحليل الاعداد الشهرية للمصابين بالتهاب اللوزتين وللفترة من من (2004-2008)

يتضح من الشكل (2) اعلاه عدم وجود تغير وتقلبات عبر الزمن وبذلك فان المتسلسلة مراوحة في الوسط ويظهر ايضا انها مراوحة في التباين. ويؤكد هذا الاستنتاج الشكل (3) الذي يظهر رسم كل من ACF و PACF للبيانات الاصلية وواضح ايضا وجود اثر موسمي عند التخلف (12)



A



B

شكل (3) : A رسم ACF و b PACF: لاعداد الشهرية للمصابين بالتهاب اللوزتين من الشكل (3) اعلاه يمكن مبدئيا ترشيح الانموذج $ARMA(1,2)(1,0)_{12}$ لتمثيل بيانات الظاهرة المدروسة، اضافة الى ذلك فقد تم توفيق عدة نماذج وحساب قيمة معيار شوارز-بيز (BIC) والانموذج الذي يمتلك اقل قيمة BIC يعتبر الانموذج المرشح لتمثيل البيانات، ويمكن كتابة صيغة هذا المعيار كما يلي (konishi and kitagawa,2008).

$$BIC = n \log(\sigma^2) + m \log(n) \quad \dots (8)$$

حيث σ^2 تمثل تباين البواقي و n عدد المشاهدات المستخدمة في نمذجة المتسلسلة الزمنية و m=p+q+P+Q عدد المعلمات المقدرة في الانموذج والجدول رقم (1) يوضح النماذج المرشحة مع قيمة BIC لكل منها، ويتضح من الجدول ان الانموذج $ARMA(1,1)(2,1)_{12}$ يمتلك اقل BIC لذا يعد هو الانموذج المرشح لتمثيل الظاهرة المدروسة.

جدول (1) بعض نماذج ARMA المرشحة لتمثيل ظاهرة الدراسة مع قيمة BIC لكل منها

ARMA model	BIC	ARMA model	BIC
AR(1)	533.918	ARMA(0,1)(0,1) ₁₂	575.719
MA(1)	620.298	ARMA(1,1)(0,1) ₁₂	519.844
ARMA(1,1)	526.030	ARMA(1,1)(2,1) ₁₂	481.337
ARMA(1,1)(1,0) ₁₂	508.713	ARMA(1,2)(1,0) ₁₂	513.885
ARMA(1,0)(1,0) ₁₂	510.836	ARMA(0,2)(0,1) ₁₂	547.273
ARMA(1,0)(1,1) ₁₂	491.022	ARMA(0,2)(1,0) ₁₂	498.818
ARMA(1,1)(1,1) ₁₂	490.705	ARMA(1,2)(1,1) ₁₂	492.933
ARMA(0,1)(1,1) ₁₂	499.124		

ادناه نتائج تقدير الانموذج $ARMA(11,)(2,1)_{12}$ ويتضح جليا معنوية معلمات الانموذج حيث ان قيمة P-Value لكل معلمة اقل من 0.05، ويظهر ايضا من خلال قيم P-Value لاختبار (Ljung-Box) عشوائية بواقي الانموذج حيث ان هذه القيم اكبر من مستوى المعنوية 0.05.

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.8374	0.1327	6.31	0.000
SAR 12	0.4652	0.1603	2.90	0.005
SAR 24	0.5346	0.1588	3.37	0.001
MA 1	0.4299	0.2179	2.01	0.009
SMA 12	0.8172	0.1541	5.30	0.000

Number of observations: 60

Residuals: SS = 119204 (backforecasts excluded)

MS = 2167 DF = 55

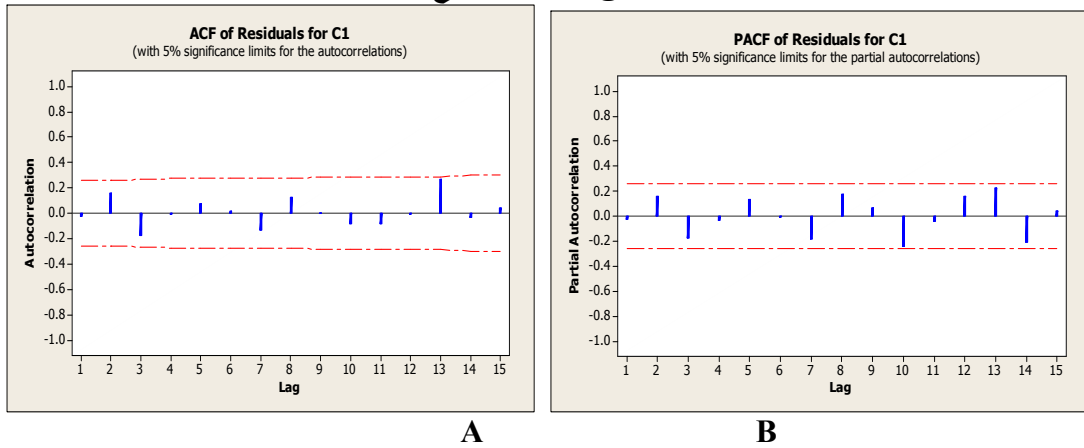
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	7.3	22.5	29.2	48.4
DF	7	19	31	43
P-Value	0.400	0.261	0.560	0.265

من النتائج اعلاه يمكن كتابة معادلة الانموذج كما يلي

$$y_t = 0.8374 y_{t-1} + 0.4652 y_{t-12} - 0.3895 y_{t-13} + 0.5346 y_{t-24} - 0.4476 y_{t-25} + 0.4299 a_{t-1} + 0.8172 a_{t-12} + 0.3455 a_{t-13} + a_t$$

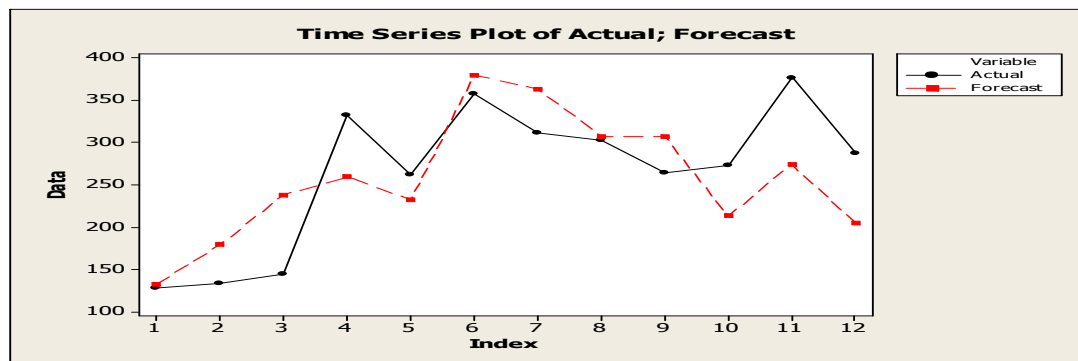
الشكل (4) يبين رسم ACF و PACF لبواقي الانموذج المرشح ونلاحظ ان البواقي غير معنوية أي تشويش ابيض وبذلك يتحقق اجتياز الانموذج للفحوص التشخيصية



شكل (4) A: ACF ، B: PACF لبواقي الانموذج $ARMA(1,0)(ARMA 1,1)_{11}(1,1)(2,1)_{12}$

$ARMA(1,0)(1,1)_{11}$

المرحلة الاخيرة هي التنبؤ لـ اثنا عشر شهراً باستخدام الانموذج $ARMA(1,1)(2,1)_{12}$ ويظهر الشكل (5) القيم المتنبأ بها مقارنة مع القيم الاصلية .



شكل (5): القيم المتنبأ بها مقارنة مع القيم الحقيقية نلاحظ من شكل (5) اعلاه ان التنبؤات تسلك نفس سلوك القيم الاصلية

جـ . نمذجة اعداد المصابين بالتهاب اللوزتين باستخدام الانحدار الخطي المتعدد مع التنبؤ: من خلال تطبيق انحدار المتغير المعتمد (الاعداد الشهرية للمصابين بالتهاب اللوزتين) على المتغيرات التوضيحية .



نمذجة متسلسلة اعداد المصابين بالتهاب اللوزتين
في محافظة نينوى للفترة من 2004-2009

X_1 : يمثل المعدلات الشهرية للأمطار ، X_2 : درجة الحرارة العظمى، X_3 : درجة الحرارة الصغرى، X_4 :
معدلات الرطوبة ، X_5 : سرعة الرياح، X_6 : الضغط الجوي، X_7 : السطوع الشمسي، X_8 : معدل التبخر

The regression equation is

$$Y = 10006 + 0.378 X_1 - 2.80 X_2 - 5.29 X_3 - 0.54 X_4 + 60.2 X_5 - 9.78 X_6 + 38.6 X_7 - 0.551 X_8$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	10006	5430	1.84	0.071	
X1	0.3775	0.4057	0.93	0.356	3.910
X2	-2.796	8.050	-0.35	0.730	151.763
X3	-5.291	9.160	-0.58	0.566	118.076
X4	-0.536	1.542	-0.35	0.730	15.796
X5	60.17	20.09	3.00	0.004	1.431
X6	-9.776	5.269	-1.86	0.069	31.738
X7	38.58	11.49	3.36	0.001	14.838
X8	-0.5513	0.3833	-1.44	0.156	36.570

يلاحظ من خلال النتائج اعلاه ظهور مشكلة تعدد العلاقة الخطية حيث كانت قيمة عامل تضخم التباين (VIF) اكبر من 10 ولاكثر من متغير لذلك سيتم تحويل المتغيرات الى مكونات رئيسية للحصول على عدد قليل من المتغيرات غير المرتبطة ومن خلال تطبيق تحليل المكونات الرئيسية على المتغيرات التوضيحية ظهرت النتائج ادناه

Eigenanalysis of the Correlation Matrix

Eigenvalue	6.2722	1.0149	0.4522	0.0993	0.0895	0.0435	0.0232	0.0052
Proportion	0.784	0.127	0.057	0.012	0.011	0.005	0.003	0.001
Cumulative	0.784	0.911	0.967	0.980	0.991	0.996	0.999	1.000
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
C8	-0.287	-0.446	-0.774	-0.338	-0.044	-0.019	-0.029	0.055
C9	0.393	0.030	-0.151	0.164	-0.269	-0.396	-0.155	0.736
C10	0.386	-0.061	-0.288	0.371	-0.220	-0.345	-0.188	-0.651
C11	-0.383	-0.133	0.042	0.383	0.644	-0.427	-0.291	0.078
C12	0.147	-0.871	0.451	0.009	-0.122	0.020	-0.016	0.010
C13	-0.384	0.125	0.231	-0.254	-0.451	-0.113	-0.705	-0.070
C14	0.383	0.057	0.119	-0.713	0.369	-0.415	-0.059	-0.125
C15	0.390	-0.032	-0.150	0.042	0.329	0.596	-0.595	0.069

حيث لوحظ ان التباين للمكون الرئيسي الاول يساوي 6.2722 والتباين للمكون الرئيسي الثاني 1.0149 وكلاهما اكبر من 1 اضافة الى ذلك فان نسبة المساهمات التجميعية للمكونات الرئيسية تبين ان المكون الاول والثاني يفسران سوية مانسبته 91% من التباين الكلي لذلك يعتمد على المكون الاول والثاني وترك بقية المكونات. ولتحديد المتغيرات المتشعبة على المكونات المختارة يمكن تطبيق العلاقة $\frac{0.5}{\sqrt{\text{var}(pci)}}$

(Afif & clark, 1984) على المكون الاول والثاني حيث تبين ان جميع المتغيرات متشعبة على المكون الاول في حين ان متغير سرعة الرياح متشعب على المكون الثاني وبتطبيق معادلة الانحدار الخطي المتعدد على المكونات الرئيسية الاولى والثانية تم الحصول على النتائج التالية

The regression equation is

$$y = 247 + 14.4pc_1 - 27.0pc_2$$

Predictor	Coef	SE Coef	T	P	VIF
-----------	------	---------	---	---	-----

Constant 247.409 8.412 29.41 0.000
 pc_1 14.415 3.351 4.30 0.000 1.001
 pc_2 -27.037 8.093 -3.34 0.001 1.001
 S = 64.9826 R-Sq = 33.6% R-Sq(adj) = 31.3%

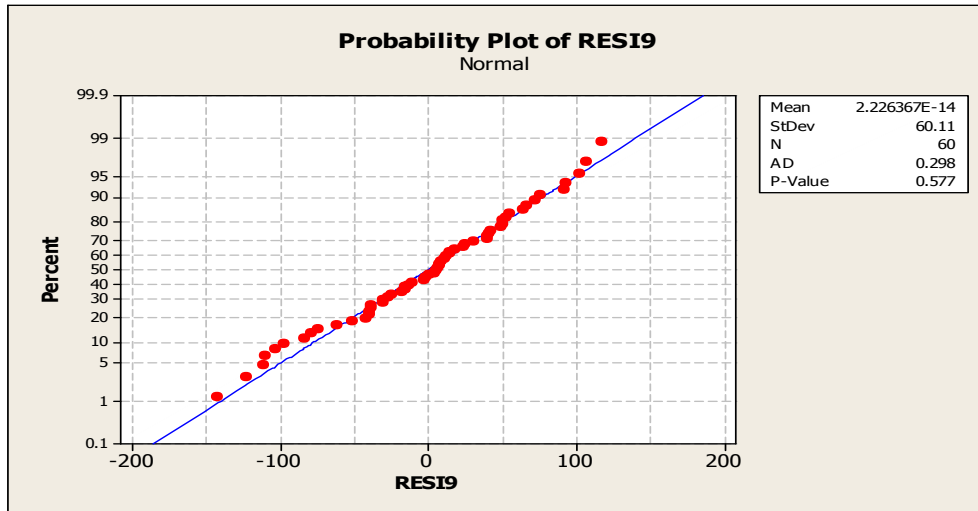
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	121886	60943	14.43	0.000
Residual Error	57	240696	4223		
Total	59	362582			

Durbin-Watson statistic = 1.78558

Predicted Values for New Observations

حيث نلاحظ من خلال النتائج اعلاه معنوية المتغيرات والحد الثابت حيث ان المعلمة اكبر من ضعف الخطأ المعياري وقيمة p-value اقل من مستوى المعنوية 0.05 وهذا يدل على معنوية تاثير المكونات المستخدمة على المتغير المعتمد، ومن خلال جدول تحليل التباين وملاحظة ان قيمة الاحتمال p-value بانها اقل من مستوى المعنوية 0.05 فان هذا يدل على معنوية نموذج الانحدار. والشكل (6) ادناه تظهر اختبار اعتدالية التوزيع الاحتمالي للبقايا حيث ان قيمة p-value تساوي 0.577 وهي اكبر من مستوى معنوية 0.05 وبالتالي فان شرط اعتدالية التوزيع الاحتمالي للبقايا متوفر لغاية 0.577.



شكل (6) رسم اعتدالية التوزيع الاحتمالي للبقايا

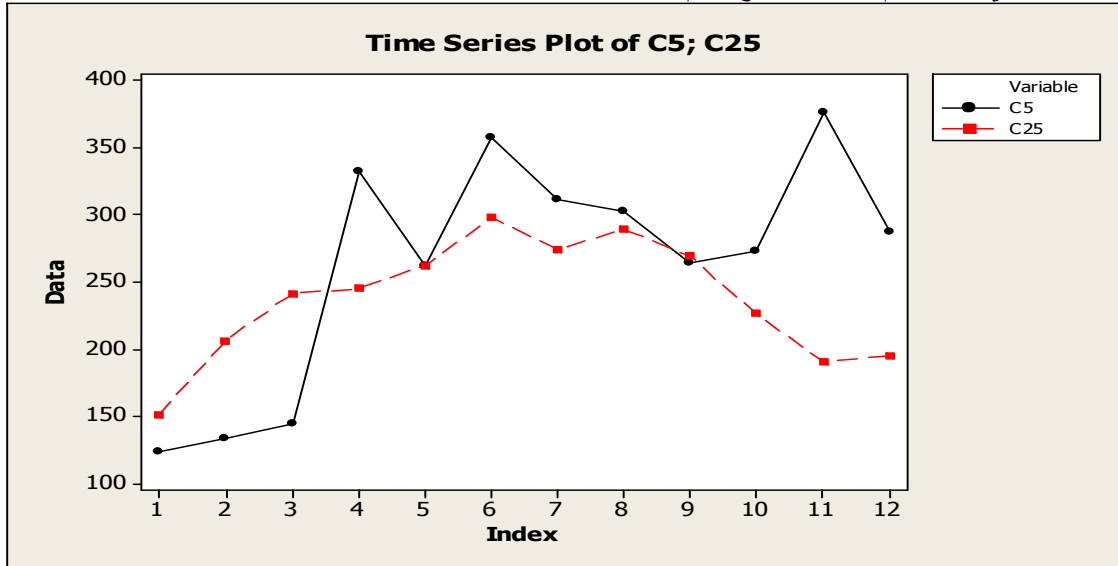
ان الخطوة الاخيرة هي التنبؤ بأعداد المصابين بالتهاب اللوزتين لاشهر عام 2009 من خلال معادلة الانحدار ومقارنتها بالقيم الحقيقية ، والنتائج التالية تبين القيم المتنبأ بها وحدود الثقة

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	151.35	21.63	(108.03; 194.67)	(14.20; 288.50)
2	206.38	11.70	(182.96; 229.81)	(74.16; 338.60)
3	241.56	10.10	(221.33; 261.78)	(109.87; 373.24)
4	245.39	8.53	(228.31; 262.48)	(114.15; 376.64)
5	262.56	10.74	(241.05; 284.06)	(130.67; 394.45)
6	298.51	13.20	(272.07; 324.95)	(165.73; 431.29)
7	274.62	15.98	(242.62; 306.63)	(140.62; 408.63)

8	289.13	13.33	(262.44; 315.81)	(156.29; 421.96)
9	269.27	10.26	(248.72; 289.83)	(137.54; 401.01)
10	226.57	11.24	(204.06; 249.07)	(94.51; 358.62)
11	191.12	13.83	(163.43; 218.80)	(58.08; 324.15)
12	195.03	14.93	(165.13; 224.92)	(61.51; 328.54)

والشكل التالي يبين القيم المتنبأ بها مع القيم الحقيقية:



شكل (7) رسم القيم المتنبأ بها باستخدام الانحدار الخطي المتعدد مقارنة مع القيم الحقيقية ولغرض مقارنة دقة التنبؤ بين النموذجين المستخدمين في الدراسة فقد تم استخدام معيار متوسط الأخطاء المطلقة (Mean Absolute Errors) MAE: والذي يقيس دقة القيم المتنبأ بها ويعبر عنه بالصيغة التالية

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n}$$

حيث أن Y_t هي القيمة الفعلية عند الزمن t ، \hat{Y}_t هي القيمة المتنبأ بها، n عدد المشاهدات للعينة البعدية المتنبأ بها (Liu, 2006).

الطريقة	MAE
منهجية بوكس جنكز	51.0763
الانحدار الخطي المتعدد	60.5625

يتبين من خلال معيار MAE ان منهجية بوكس جنكز تعطي نتائج تنبؤ افضل من طريقة الانحدار الخطي المتعدد.

4- الاستنتاجات

1. لوحظ ان المتسلسلة الزمنية لاعداد المصابين بالتهاب اللوزتين مراوحة في الوسط والتباين ولها اثر موسمي عند التخلف (12) ومن ثم فان اعداد المصابين تتاثر معنويا نتيجة التغير المناخي من الصيف الى الشتاء .
2. ان الانموذج الملائم لظاهرة الدراسة هو $ARMA(1,1)(2,1)_{12}$ وقد اجتاز النموذج الفحوص التشخيصية وقدم تنبؤات ملائمة ويمكن الاستفادة منه في التعامل مع ظاهرة الدراسة واخذ الاحتياطات المناسبة في التفاعل مع حالات الاصابة لكل شهر من اشهر السنة في هذا المجال.
3. من خلال المقارنة بين نتائج التنبؤ لكل من الانحدار الخطي المتعدد واسلوب بوكس جنكنز (الانحدار الحركي) بالاعتماد على معيار MAE فقد اظهر جليا تفوق اسلوب بوكس جنكنز في التنبؤ. وهذا يدل على فعالية الانحدار الحركي في التغير عن ظاهرة الدراسة بصورة افضل من الانحدار الخطي المتعدد (غير الحركي).

5 - المصادر

- 1 الخياط ، باسل يونس ذنون (1991): " الاحتمالية والمتغيرات العشوائية " دار الكتب للطباعة والنشر، جامعة الموصل – العراق.
2. الراوي ، خاشع محمود (1987): المدخل الى تحليل الانحدار ، دار الكتب للطباعة والنشر وجامعة الموصل .
3. فاندل، والتر، (1983): "السلاسل الزمنية من الوجة التطبيقية ونماذج بوكس-جنكنز"، تعريب ومراجعة عزام عبدالمرضي (1992)، دار المريخ للنشر، الرياض، المملكة العربية السعودية.
4. الناييل ، حامد محمد خلف، (2011): "دراسة استقرارية بعض نماذج المتسلسلات الزمنية النسبية غير الخطية مع التطبيق " ، رسالة ماجستير غير منشورة ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل.
5. الوردى ، عدنان هاشم، (1990): "أساليب التنبؤ الإحصائي طرق وتطبيقات "، مطابع دار الحكمة ،العراق.
- 6 Balcombe, k.G (2005): Model selection using Information criteria and Genetic Algorithms. Computational Economics, 25:207-228.
- 7 Bluman , G Allan(2009). "Elementary statistics: A step by step Approach" , seventh ed. McGraw-Hill , New york.
- 8 Cordeiro, c. and Never, M.M.(2009) : Forecasting Time series with boot. Expos procedure. Statistical Journal , vol.7 , No .2, 153-149.
- 9 Gujarati D.N., (2004), "Basic econometric", 4rd ed., Mc Graw-hill New York.
- 10 Afifi .A.A And Clark, V. (1984). "Computer Aided Multivariate Analysis A Division Of Wadsworth , Inc . London
- 11 Konishi, S. and kitagawa, G.(2008), Information criterion and statistical Modeling , springer Science+ Business Media ,LLc .
- 12 Liu,L.-M.,(2006), "Time Series Analysis and Forecasting " , 2nd ed., Scientific Computing Associates Crop., Illinois, USA .
- 13 Liung , G.M .and box ,G.E.P.(1978) . " on a measure of lack of fit in time series models " Biometrika 65 ,297-303 .
- 14 Mahsin, MD., Akhter, yesmin, and Begum Monira, (2012), " modeling rainfall in Dhaka division of Bangladesh using time series analysis" , journal of mathematical modeling and application, vol.1, no.5.
- 15 Makridakis, S., Wheelwright, S.C. and Hyndman, R., (1998), "Forecasting Methods and Applications", 3rd ed., John Wiley and Sons, New York.
- 16 Pankratz , a (1991) . " Forecasting With Dynamic Regression Models " John wiley and sone .U.S.A .



Modeling sequential preparation with rheumatoid tonsils in Nineveh for the period 2004-2009

Abstract

In this research will be treated with a healthy phenomenon has a significant impact on different age groups in the community, but a phenomenon tonsillitis where they will be first Tawfiq model slope self moving averages seasonal ARMA Seasonal through systematic Xbox Cengzla counter with rheumatoid tonsils in the city of Mosul, and for the period 2004-2009 with prediction of these numbers coming twelve months, has found that the specimen is the best representation of the data model is the phenomenon SARMA (1,1) * (2,1) 12 from the other side and explanatory variables using a maximum temperature and minimum temperature, solar brightness , evaporation, rainfall, wind speed, air pressure, for the same period the previous possible modeling preparation with rheumatoid tonsils and in the method of multiple regression, and then compare the results to predict the two models where given methodology X box Jenkins better results in the prediction of the linear regression multi-depending on the standard of the average errors absolute MAE, and these results can help decision makers to develop appropriate strategies to deal with this disease to reduce the preparation of the injured and provide the prerequisites for appropriate treatment.

Keywords \ARIMA, box Jenkins, multiple linear regression , principal component .