إستعمال أنموذج ARIMA للتنبؤ بعدد الإصابات بمرض التدرن في محافظة واسط [+]

علاء حسين عمران انصاف [*]

المستخلص :

يهدف البحث الى ايجاد أنموذج تنبؤي لأعداد المصابين بمرض التدرن في محافظة واسط ، باستعمال منهجية بوكس – جنكينز للفترة من كانون الثاني 2008 الى كانون الاول 2012 . من هذه الدراسة وجد ان الأنموذج الأفضل للتنبؤ بأعداد المرضى هو ARIMA(1,0,0) او AR(1) . وبالاعتماد على هذا الأنموذج ، تم التنبؤ بأعداد المرضى للفترة من كانون الثاني 2013 الى كانون الاول 2014 ، هذه القيم قد أظهرت توافقا مع القيم الأصلية .

USING ARIMA MODEL FOR FORECASTING THE NUMBER OF PATIENTS WITH TUBERCULOSIS DISEASE IN WASSIT PROVINCE

Alaa Husein Omran

Abstract :

This paper aims to modeling of forecasting to the number of patients with Tuberculosis (T.B) disease in Wassit province, using the Box-Jenkins methodology during the period from January 2008 to December 2012. Through this a study, find that the most adequate model for the number of patients is ARIMA (1,0,0) or AR(1). Using the model, we forecast the numbers of patients for two- year period ahead from January 2013 to December 2014, these values has shown a harmonic direction according to the original values.

المقدمة Introduction :

يعد التنبؤ المستقبلي من المواضيع الرئيسة التي تمكن أصحاب القرار من اتخاذ القرارت الصحيحة في كافة المجالات والمستويات سواء كانت في المجالات الصحية والزراعية والتجارية والخدمية وغيرها من المجالات الأخرى . إن اساليب التنبؤ تقسم الى نوعين رئيسين تبعا للمنهجية المعتمدة حيث تقسّم الى اساليب نظامية واساليب غير نظامية، فالأساليب النظامية تعتمد على طرق علمية ولا يدخل العامل الشخصي في التقدير وبالتالي فهي بعيدة عن التحيز اما الأساليب غير النظامية فتعتمد على الخبرة والتجربة والاجتهاد الشخصي . وتعد السلاسل الزمنية من اهم الأساليب النظامية التي يمكن من خلالها التنبؤ بما سيحدث في المستقبل على قيم الظاهرة على ضوء ما حدث لها في الماضي.

وحيث ان الجانب الصحي له اهمية كبيرة في الجانب التنموي لأنه يهتم بحياة الأنسان الذي هو هدف كل دولة تسعى لرفاهية الشعب والرقي بالمستوى الصحي له والخلاص من كافة الأمراض وخاصة مرض التدرن الذي يعتبر من الأمراض

[+] تاريخ إستلام البحث 2013/6/2 ، تاريخ قبول النشر 2013/11/13
[*] مدرس / المعهد التقني / الكوت

الانتقالية المميتة اذا ما تركت دون تخطيط مسبق لمواجهة هذا المرض. لذا فقد أنصب اهتمامنا في هذا البحث على هذا المرض، حيث تم الاعتماد على الأعداد الشهرية من الأشخاص الذين اصيبوا بهذا المرض في محافظة واسط للفترة من كانون الثاني 2008 الى نهاية كانون الاول 2012 لغرض وضع أنموذج للتنبؤ المستقبلي بأعداد المصابين بهذا المرض للسنوات القادمة ووضعها بأيدي الجهات المختصة لغرض اتخاذ التدابير اللازمة لمواجهة هذا المرض.

أهمية البحث :

تكمن أهمية البحث بوضع أنموذج ملائم يستخدم للتنبؤ بأعداد الأشخاص المتوقع اصابتهم بالتدرن، وذلك باستخدام منهج تحليل السلاسل الزمنية المبني على منهجية بوكس– جنكينز(Box–Jenkins) ومن ثم التنبؤ بهذه الأعداد لغاية كانون الاول عام 2014.

أهداف البحث :

يهدف البحث الى:

1. امكانية تطبيق اسلوب بوكس – جنكينز (Box– Jenkins) في التنبؤ بأعداد الأشخاص المتوقع اصابتهم بالتدرن.
2. وضع أنموذج قياسي للتنبؤ بأعداد الأشخاص المتوقع اصابتهم بالتدرن.
3. التنبؤ بأعداد الأشخاص المتوقع اصابتهم بالتدرن لغاية كانون الاول عام 2014.

منهج البحث :

استخدم المنهج الوصفي التحليلي في انجاز هذا البحث من خلال الاطلاع على عدد من المصادر التي تناولت منهجية بوكس – جنكينز في تحليل السلاسل الزمنية ومن ثم الحصول على بيانات السلسلة الزمنية من وزارة الصحة حيث تم تطبيق الحزمة البرامجية Minitab 16 و SPSS18 في تحليلها.

1- الجانب النظري :

1-1 السلسلة الزمنية Time Series :

هي مجموعة من المشاهدات الخاصة بظاهرة معينة مأخوذة في اوقات زمنية ويمكن التعبير عنها بالصيغة الاتية[1]:

$$[X_t]_{t=-\infty}^{\infty} = \ldots, X_{t-1}, X_t, X_{t+1}, \ldots$$

واذا امكن التنبؤ بالقيم المستقبلية للسلسلة بشكل مضبوط باستخدام أنموذج ثابت لا يحتوي على المتغير العشوائي (e_t) يقال عنها حتمية (Deterministic) ولكن أكثر السلاسل الزمنية تمتاز بالتصادفية (Stochastic) أي بمعنى آخر أن

قيمها المستقبلية لها توزيع احتمالي ويتم احتسابها باستخدام أنموذج عشوائي يحتوي على المتغير العشوائي (e_t) . وفي بعض السلاسل الزمنية يمكن مشاهدة قيم المتغير في كل لحظة من الزمن ويطلق عليها هذه بالسلاسل الزمنية المتصلة (Continuous time series) ، ومع ذلك فان معظم السلاسل الزمنية تتضمن مشاهدات لمتغير ما يمكن مشاهدتها عند نقاط زمنية متساوية حيث يطلق عليها بالسلاسل الزمنية المنقطعة (Discrete time series). وهذه يمكن الحصول عليها إما عن طريق تسجيل قيم الظاهرة في أزمنة ثابتة او عن طريق تجميع قيم الظاهرة لفترة زمنية ثابتة.

2-1 دالة الارتباط الذاتي Autocorrelation Function (ACF) [1] :

إن دالة الارتباط الذاتي تستخدم في تحديد درجة وملائمة الأنموذج عن طريق اختبار عشوائية أخطاء البواقي (e_t) ، ولها اهمية كبيرة ؛لأنها احدى اساليب تحديد فيما اذا كانت العملية العشوائية مستقرة أم لا ، فاذا كانت كذلك فيتم اختيار احد النماذج المناسبة من مجموع نماذج العمليات العشوائية المستقرة. اذ يمكن تقدير دالة الارتباط الذاتي بإبطاء k (Autocorrelation Function of lag k) للسلسلة الزمنية $\{X_t ; t\in T\}$ لتكون [1].

$$\rho_k = \frac{\text{cov}\,(X_t, X_{t+k})}{\sqrt{\text{var}\,(X_t)\ \text{var}\,(X_{t+k})}} = \frac{E\,(X_t-\mu)(X_{t+k}-\mu)}{\sqrt{E\,(X_t-\mu)^2\ E\,(X_{t+k}-\mu)^2}}$$

$$= \frac{E\,(X_t-\mu)(X_{t+k}-\mu)}{\sigma_X^2} = \frac{\gamma_k}{\gamma_0} \qquad (1)$$

وعندما تكون العملية العشوائية مستقرة فإنّ المتوسط والتباين يكونان ثابتين ومتساويين لكلّ الفترات، وان $\sigma_{X_t} = \sigma_{X_{t+k}}$.

3-1 دالة الارتباط الذاتي الجزئي Partial Autocorrelation Function (PACF)[2] :

هذه الدالة تسهم ايضا في تشخيص الأنموذج المناسب، وتكشف الارتباط بين X_t , X_{t+k} ، بثبوت المتغيرات الأخرى. ويمكن ايجاد قيم معامل الارتباط الجزئي حسب الصيغة التالية:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \ldots + \phi_{kk}\rho_{j-k} \quad , \quad j=1,2,\ldots,k \qquad (2)$$

حيث أن ϕ_{kj} تمثل معلمة الانحدار

4-1 التشويش الأبيض (White Noise) :

إن العملية العشوائية $\{e_t\}$ تدعى بعملية التشويش الأبيض(White Noise Process) إذا كانت تمثل متتابعة (Sequence) من المتغيرات العشوائية غير المترابطة مع بعضها البعض (Uncorrelated) والتي معدلها يساوي صفر وتباينها محدود[2]،أي أن:

$E(e_t)=0$ (3)

$Var\ (e_t)= \sigma_x^2$ (4)

$Cov\ (e_t,e_{t+k}) = 0\ ,\ \forall k\neq 0.$ (5)

5-1 الإستقرارية وعدم الإستقرارية في السلاسل الزمنية(Stationary and Non Stationary in Time Series) : [3]

هناك مجموعة خاصة من السلاسل الزمنية المستقرة التي تكون مبنية على اساس ان السلسلة الزمنية تمتلك وسطا حسابيا وتباينا ثابتا مع استمرارية الزمن ، وتكون السلسلة مستقرة اذا لم يكن هناك اتجاه الى الأعلى او الى الأسفل في المعدل عبر الزمن . هنالك نوعان من الإستقرارية في السلاسل الزمنية هما:

أولاً- الإستقرارية التامة (Strictly stationary).

يقال للسلسلة الزمنية $(X_t\ ;\ t=1,2,\ldots,n)$ بانها مستقرة استقرارية تامة إذا كان التوزيع المشترك لاية مجموعة من المشاهدات لا يتأثر بإزاحة كل الفترة الزمنية للمشاهدات الى الامام او الى الخلف بأية كمية صحيحة اي ان :

$$Pr\ (X_{t_1}, X_{t_2}, \ldots, X_{t_m}) = Pr\ (X_{t_1+k}, X_{t_2+k}, \ldots, X_{t_m+k}) \quad (6)$$

إذ أن t_m تمثل اية فئة زمنية و k مقداراً ثابتاً. وبمعنى اخر أن تغير الزمن بمقدار (k) ليس له تأثير في التوزيع الاحتمالي المشترك للسلسلة، بل يعتمد التوزيع المشترك على الزمن $(t_1,t_2,\ldots,t_n)$ فقط.

ثانياً – الإستقرارية الضعيفة او من الدرجة الثانية Weakly or Second – order stationary.

يقال للسلسلة الزمنية $(X_t\ ;\ t=1,2,\ldots,n)$ بأنّها ذات استقراريه من الدرجة الثانية اذا تحققت الشروط الآتية :

آ- متوسط السلسلة يكون ثابتاً أي أن:

$$E\ (X_t) = \mu \quad (7)$$

إذ ان μ هو متوسط العملية العشوائية ولا يعتمد على قيم t والذي يقدر من مشاهدات السلسلة الزمنية كالآتي:

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t \quad (8)$$

ب- تباين السلسلة الزمنية يكون ثابتاً

$$var\ (X_t) = E\ (X_t - \mu)^2 = \sigma_x^2 = \gamma_o \quad (9)$$

إذ ان γ_o هو تباين العملية العشوائية ويكون ثابتاً ولا يعتمد على قيم t والذي يقدر من مشاهدات السلسلة الزمنية كما يأتي :

$$\hat{\gamma}_o = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^2 \qquad (10)$$

ج- التباين المشترك للسلسلة الزمنية او (التغاير الذاتي) كالآتي:

$$\text{cov}\ (X_t, X_{t+k}) = \text{E}(X_t - \mu)(X_{t+k} - \mu) = \gamma_k \qquad (11)$$

إذ ان γ_k هو التغاير الذاتي(Auto covariance) للعملية العشوائية عند الازاحة k (Lag k) ويكون ثابتاً لا يعتمد على قيم t لجميع القيم الصحيحة لـ k والذي يقدر من مشاهدات السلسلة الزمنية كما يأتي:

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})(X_{t+k} - \overline{X}) \qquad (12)$$

وقد سميت السلسلة ذات الإستقرارية الضعيفة بالسلسلة ذات الإستقرارية من الدرجة الثانية, لأن كلاً من العزمين الاول والثاني يكونا موجودين وثابتين مع الزمن وإنّ التغاير الذاتي يعتمد على الازاحة (k) فقط.

1-5-1 الاستدلال على استقراريه السلسلة الزمنية [3] :

يمكن الاستدلال حول استقراريه السلسلة عن طريق :

اولاّ- رسم القيم الأصلية فإذا وجد فيها نمو او هبوط في البيانات (اي وجود اتجاه عام) فإنها غير مستقرة .

ثانياّ- رسم معاملات دالة الارتباط الذاتي ACF وكذلك معاملات دالة الارتباط الذاتي الجزئي PACF (حيث يمثل المحور السيني فترة الإبطاء k اما المحور الصادي فيمثل معاملات الارتباط) . فإذا كانت قيم معاملات الارتباط تضمحل (تتناقص) بسرعة وتكون قريبة من الصفر مع ازدياد فترات الإبطاء فهذا دليل على أن السلسلة مستقرة، وخلاف ذلك فإن السلسلة غير مستقرة.

فإذا كانت السلسلة غير مستقرة فيجب ازالة عدم الإستقرارية وكما يأتي:

أ- اذا كانت عدم الإستقرارية ناتجة عن وجود اتجاه عام فقط فيمكن التخلص من ذلك باستخدام طريقة الفروق وبدرجات مختلفة، فمثلا اخذ الفرق الاول يكون بالصيغة الآتية:

$$\nabla^1 X_t = X_t - X_{t-1} \qquad (13)$$

إذ يشكل $\nabla^1 X_t$ سلسلة الفروق الاولى.

وباستخدام عامل الارتداد الزمني للمعادلة السابقة نحصل على الصيغة الأتية:

$$\nabla^1 X_t = X_t - BX_t = (1 - B)X_t \qquad (14)$$

نلاحظ ان الفرق الاول قد مثل بـ (B –1) .
اما للفرق الثاني فيتم أخذ الفروق الاولية للبيانات التي تم أخذ الفرق الاول لها ويكون ذلك كالآتي:

$$\nabla^2 X_t = \nabla^1 X_t - \nabla^1 X_{t-1} \qquad (15)$$

إذ يشكل $\nabla^2 X_t$ سلسلة الفروق الثانية.
وبتعويض معادلة رقم (13) في المعادلة (15) نحصل على:

$$\begin{aligned} \nabla^2 X_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= (1 - 2B + B^2) X_t = (1 - B)^2 X_t \end{aligned} \qquad (16)$$

وهكذا الحال فيما يخص الفروقات الاخرى. فبعد اخذ عدداً مناسباً من الفروقات للسلسلة الزمنية الاصلية X_t ولتكن d نحصل على السلسلة الزمنية المستقرة التالية :

$$\nabla^d X_t = (1 - B)^d X_t \qquad (17)$$

ب- اذا كانت السلسلة فيها تأثيرات موسمية فقط تستخدم طريقة الفرق الموسمي (Seasonal Difference) وذلك بطرح القيم من بعضها البعض حسب فترات الإبطاء lags وهناك، على سبيل المثال:

- الفرق ربع السنوي $X_{1t} = X_t - X_{t-4}$.
- الفرق الشهري $X_{2t} = X_t - X_{t-12}$

ج- اذا كان هناك اتجاه عام و تأثيرات موسمية فيتم اخذ الفروق غير الموسمية والموسمية. حيث يتم اخذ الفرق الاول او الثاني ومن ثم اخذ الفرق الموسمي حسب فترة الإبطاء.

د- إذا كانت السلسلة غير مستقرة بالتباين فيتم استخدام احدى التحويلات باستخدام اللوغاريتم الطبيعي او الجذر التربيعي او مقلوب البيانات او الأسي (تحويلات القوى) Power transformation ، حيث تتحول السلسلة غير المستقرة حول التباين الى مستقرة وتعد تحويلات القوى المعرفة ابسط مجموعة تحويلات وهي كالاتي:

$$X_t^{(\lambda)} = \begin{cases} X_t^{\lambda} & if \quad \lambda \neq 0 \\ Ln\ X_t & if \quad \lambda = 0 \end{cases} \qquad (18)$$

$X_t^{(\lambda)}$: السلسلة المحولة عند الزمن t.

X_t : قيمة السلسلة غير المحولة.

λ : قيمة معلمة التحويل وعادة تتراوح قيمتها بين $\pm 1, \pm \frac{1}{2}, \pm \frac{1}{3}, \pm \frac{1}{4}$ علاوة على الصفر.

6-1 نماذج بوكس–جينكنز للسلاسل الزمنية:

إن نماذج بوكس–جنكينز من الأساليب الإحصائية المهمة لتحليل السلسلة الزمنية، حيث تستخدم هذه النماذج لتمثيل سلسلة زمنية تمثل ظاهرة معينة وفي التنبؤ بقيم الظاهرة في المستقبل ومن هذه النماذج،[4]

1-6-1 أنموذج الانحدار الذاتي Autoregressive Model (AR).

يعد أنموذج الانحدار الذاتي ذو فائدة تطبيقية هامة في تحليل السلاسل الزمنية واكثرها شيوعا ويشار اليه إختصارا بـ AR ويمكن تمثيل هذا الأنموذج بصيغته من الدرجة (P) فيعطى بالصورة الآتية:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + e_t \qquad (19)$$

حيث ان :

e_t : الخطأ العشوائي .

ϕ : معالم الأنموذج.

P : رتبة عمليات الانحدار الذاتي.

وفي الواقع فانه يمكن كتابة الأنموذج اعلاه باستخدام عامل الازاحة الخلفية على النحو الآتي:

$$(1 - \phi_1 B - \phi_2 B^2 - ..... - \phi_p B^P) X_t = e_t$$

$$\phi_p(B) X_t = e_t \qquad \text{حيث أن :}$$

$$1 - \phi_1 B - \phi_2 B^2 - ..... - \phi_p B^P = \phi_p(B) \qquad (20)$$

وعليه فأن:

$$X_t = (\phi_p(B))^{-1} e_t = \sum_{k=0}^{\infty} \Psi_k e_{t-k} \qquad (21)$$

لذا فيمكن ان يقال لعملية الانحدار الذاتي من الدرجة p بانها قابلة للانعكاس (Invertible). وفي الواقع فانه عند وقوع جذور المعادلة $\phi_p(B) = 0$ خارج دائرة الوحدة فهذا يعني بان العملية مستقرة .

2-6-1 أنموذج الاوساط المتحركة Moving Average Model (MA)

إن نماذج الاوساط المتحركة هي عبارة عن ارتباط مشاهدات السلسلة الزمنية الحالية مع خطأ السلسلة نفسها لفترات سابقة ويشار اليه اختصارا بـ MA ويمكن تمثيل هذا الأنموذج بصيغته من الدرجة (q) فيعطى بالصورة التالية:

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots\ldots - \theta_q e_{t-q} \qquad (22)$$

ويرمز لأنموذج الاوساط المتحركة MA(q) ، حيث ان:-

θ : معالم الأنموذج

q : رتبة الاوساط المتحركة

و باستخدام عامل الازاحة للخلف (β) يمكن تحويل المعادلة (22) الى الصيغة التالية:

$$X_t = \theta_{(q)}(B)\, e_t \qquad (23)$$

إذ إنّ

$$\theta_{(q)}(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_q B^q)$$

وهذه هي دالة لمتعدد الحدود (polynomial function) من الدرجة (q في المتغير B).

3-6-1 الأنموذج المختلط (الانحدار الذاتي – الاوساط المتحركة) Mixed Autoregressive Moving Average Model (ARMA)

نماذج الانحدار الذاتي والاوساط المتحركة هي عبارة عن ارتباط قيم السلسلة الزمنية الحالية مع قيم سابقة للسلسلة نفسها وارتباط قيم السلسلة مع خطأ السلسلة نفسها لفترات سابقة، فإذا كانت X_t ; t = 0, ± 1, ±2,. . سلسلة زمنية مستقرة أحادية المتغير فإن الصيغة العامة للأنموذج هي:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots\ldots - \phi_p X_{t-p} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots\ldots - \theta_q e_{t-q} \qquad (24)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots + \phi_p X_{t-p} - \theta_1 e_{t-1} - \ldots\ldots\ldots - \theta_q e_{t-q} + e_t \qquad (25)$$

ويرمز لأنموذج الانحدار الذاتي والاوساط المتحركة بـ ARMA(p,q)

حيث ان (p,q) هما عدد معلمات الأنموذج ، ويمثل الأنموذج اعلاه الجزء غير الموسمي او ما يسمى بالجزء المنتظم للسلسلة الزمنية . ويمكن اعادة كتابة المعادلة (25) باستخدام عامل الازاحة الخلفية(B) لتصبح المعادلة

$$\phi(B) X_t = \theta(B) e_t \qquad (26)$$

إذ أنَّ

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots\ldots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_q B^q)$$

إنَّ الأنموذج يحتوي على (p+q+2) من المعالم وهي ϕ_p ,..., ϕ_2 ,ϕ_1 و μ و σ_e^2 و θ_q ,..., θ_1 التي تقدر من مشاهدات السلسلة الزمنية. وكذلك فإن e_t هي سلسلة الاخطاء العشوائية المستقلة وذات توزيع طبيعي بمتوسط حسابي صفراً وتباين ثابت σ_e^2. إنّ شرطي الإستقرارية والانعكاسية للأنموذج هي إنَّ جذور المعادلتين $\phi(B) = 0$ لنماذج الانحدار الذاتي و $\theta(B) = 0$ لنماذج المتوسطات المتحركة، يجب ان يكونا خارج دائرة الوحدة.

خواص الأنموذج :

أ- إنّ المتوسط للأنموذج هو :

$$E(X_t) = 0 \qquad (27)$$

ب- دالة التغايرات الذاتية هي:

$$\gamma_k = \phi_1 \gamma_{k-1} + ... + \phi_p \gamma_{k-p} \qquad , k > q \qquad (28)$$

ج- دالة الارتباطات الذاتية هي :

$$\rho_k = \phi \rho_{k-1} + ... + \phi_p \rho_{k-p} \qquad , k > q \qquad (29)$$

د- دالة الارتباطات الذاتية الجزئية متناقصة بشكل أسي أو بشكل موجات جيبيه متناقصة معتمدة على أن جذور $(1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q)$ تكون مساوية للصفر.

4-6-1 الأنموذج المختلط المتكامل (Autoregressive Integrated Moving Average)

إن معظم السلاسل الزمنية التي تواجهنا في التطبيقات العملية تكون غير مستقرة، إذ يتمّ التعرف عليها عن طريق دالة الارتباطات الذاتية حيث لا تؤول قيمها للصفر بعد الازاحة الثانية او الثالثة وإنّما تبقى قيمها كبيرة لعدد من الإزاحات. ولغرض استعمال النماذج العشوائية يجب تحويلها الى سلاسل مستقرة ويتم ذلك باستعمال عامل الفروق الخلفية (Backward difference operator) وكما في المعادلة رقم (17).فاذا عوضنا عن X_t بـ $\nabla^d X_t$ في المعادلة (26) نحصل على أنموذج جديد يستطيع معالجة نوعاً معيناً من السلاسل الزمنية غير المستقرة والذي يطلق عليها اسم الأنموذج المختلط غير المستقر والذي يرمز اليه بـ ARIMA(p,d,q) وصيغته:

$$\phi(B)\nabla^d X_t = \phi(B)(1-B)^d X_t = \theta(B) e_t \qquad (30)$$

7-1 بناء أنموذج السلاسل الزمنية باستخدام اسلوب بوكس–جنكينز

تعتمد منهجية بوكس – جنكينز في بناء أنموذج السلسلة الزمنية على توفيق أنموذج الانحدار الذاتي والاوساط المتحركة لمجموعة من البيانات. وتمر بأربع مراحل رئيسية هي ؛ تشخيص الأنموذج، تقدير معلمات الأنموذج ، اختبار دقة الأنموذج ، التنبؤ [5] .

اولا– تشخيص الأنموذج (Model Identification)

في هذه المرحلة وبعد ان يتم التحقق من استقراريه السلسلة الزمنية نقوم بتحديد p و q المطلوبة لتحديد الأنموذج الخطي العام(ARMA) وذلك برسم دالتي الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي (PACF) ومقارنتهما مع السلوك النظري لهما [6] فاذا كانت :

أ. دالة الارتباط الذاتي تتناقص تدريجيا سالكة سلوكا اسيا او سلوك دالة الجيب المتناقصة ودالة الارتباط الذاتي الجزئي تنقطع بعد الفترة p فأن الأنموذج الملائم هو AR(P).

ب. دالة الارتباط الذاتي الجزئي تنقطع بعد الفترة q ودالة الارتباط الذاتي الجزئي تتناقص تدريجيا سالكة سلوكا اسيا او سلوك دالة الجيب المتناقصة فأن الأنموذج الملائم هو MA(q).

ج. دالة الارتباط الذاتي و دالة الارتباط الذاتي الجزئي تتناقص تدريجيا سالكة سلوكا اسيا او سلوك دالة الجيب المتناقصة فأن الأنموذج الملائم للبيانات هو (ARMA(p,q.

ثانيا– تقدير معلمات الأنموذج (Model Parameters Estimation)

إن تقدير المعاملات لأنموذج بوكس– جنكينز هي عملية معقدة وصعبة بعض الشيء خاصة مع السلاسل الزمنية غير الخطية. وتستعمل طرائق عديدة لتقدير معلمات الأنموذج منها [7] :

أ– طريقة المربعات الصغرى Method of ordinary least squares (O.L.S.E)

الفكرة الأساسية لطريقة المربعات الصغرى هي جعل مجموع مربعات الأخطاء اقل ما يمكن. حيث تقوم هذه الطريقة على مبدأ تقليص مجموع مربعات خطأ التقدير وجعله في نهايته الصغرى، وذلك من خلال اشتقاق المعادلة (25) اعلاه بالنسبة لكل من θ , ϕ ومساواة التفاضلات بالصفر نحصل على المقدرات $\hat{\theta}, \hat{\phi}$ على التوالي .

$$S = \sum_{t=P+q+1}^{n} (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots \dots - \phi_p X_{t-p} - e_t + \theta_1 e_{t-1} + \dots \dots + \theta_q e_{t-q})^2 \qquad (31)$$

ب– طريقة الإمكان الأعظم Maximum Likelihood Method)

لتقدير معالم الأنموذج المختلط ARMA تستخدم طريقة الامكان الأعظم فالدالة التجميعية بثبات البيانات هي:

$$L(\theta, \phi, \sigma_e^2 \mid X_t) = (2\pi)^{-\frac{n}{2}} (\sigma_e^2)^{-\frac{n}{2}} Exp \left[-\frac{1}{2\sigma_e^2} S(\theta, \phi) \right]$$

حيث أن $S(\theta,\phi)$ تمثل مجموع مربعات الأخطاء أي:

$$S(\theta,\phi) = \sum_{t=1}^{n} e_t^{\,2}(\theta,\phi)$$

$$LnL\ (\theta,\phi,\sigma_e^{\,2}) = -\tfrac{n}{2} Ln\,(2\pi\sigma_e^{\,2}) - \frac{S(\theta,\phi)}{2\sigma_e^{\,2}} \qquad (32)$$

وبأخذ التفاضل الجزئي للمعادلة (32) بالنسبة لكل من $\sigma_e^{\,2},\theta,\phi$ ومساواة التفاضلات بالصفر نحصل على المقدرات $\hat{\sigma}_e^{\,2},\hat{\theta},\hat{\phi}$ على التوالي.

ثالثا– اختبار دقة الأنموذج (Diagnostic Checking of Model)

يتم ذلك من خلال أساليب عديدة منها :

أ. اختبار معنوية معالم الأنموذج بإستعمال اختبار (t– student) فإذا كانت غير معنوية فلابد من استبعاد احد رتب AR و MA [8] .

ب. اختبار البواقي (Residuals): حيث يجب أن تكون تشويشاً ابيضاً (مستقلة وتتبع التوزيع الطبيعي) . ويُعدّ الأنموذج مناسباً إذا تحقق هذا الشرط، ولكن اذا لم يتحقق ذلك، فيجب إعادة تشخيص الأنموذج ومحاولة الحصول على أنموذج أفضل. ومن هذه الاختبارات ؛

1– اختبار Portmanteau : يستخدم للتحقق من ملائمة الأنموذج المشخص عن طريق اختبار الفرضية الآتية:

$$H_o : \rho_1 = \rho_2 = \ldots\ldots = \rho_K = 0$$

V.S $\quad H_1 : \rho_K \neq 0$

ففي عام 1970 توصل كلٌ من (Box and Pierce) [9] الى إحصاءه لاختبار الفرضية أعلاه وبيان مدى كفاءة الأنموذج المشخص باستخدام معاملات الارتباط الذاتي للبواقي وهي معطاة بالعلاقة التالية

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^{\,2} \qquad (33)$$

حيث أن :

$\hat{\rho}_k$: تمثل مقدار الارتباط الذاتي المقدر من بيانات العينة عند الإزاحة k .

n : يمثل عدد المشاهدات للأنموذج المشخص.

m : أكبر إزاحة لـ K وهي مساوية لـ n/4 .

ثم يتم مقارنة Q المحسوبة بـ χ^2 الجدولية بدرجة حرية (m–p–q) ومستوى دلالة (α) فاذا كان Q المحسوبة اصغر من الجدولية دل ذلك على عدم رفض فرضية العدم ، أي أن الفروقات غير معنوية ,وان معاملات الارتباط الذاتي للأخطاء لها توزيع عشوائي ويكون الأنموذج ملائم وجيد. أمّا اذا كانت اكبر فالأنموذج غير ملائم وفي هذه الحالة لابدّ من إعادة المرحلة الاولى أي تشخيص أنموذج اخراً لتمثيل السلسلة وتقدير معلماتها ثم يتمّ التحقق من صحتها.

في عام 1978 قام كل من (Ljung,G.M&Box,G.E.P) [10] بتعديل صيغة اختبار Q الاصلية التي اقترحها كل من (Box & Pierce) بالصيغة الآتية:

$$Q^{*} = n(n+2)\sum_{k=1}^{m}\frac{\rho\wedge_{k}^{2}}{n-k} \qquad (34)$$

وتقارن قيمة *Q المحسوبة مع قيمة χ^2 الجدولية بدرجة حرية (m–p–q) ومستوى دلالة (α) فاذا كان *Q المحسوبة اصغر من الجدولية دل ذلك على عدم رفض فرضية العدم ، أي أن الفروقات غير معنوية, بمعنى ان معاملات الارتباط الذاتي للأخطاء لها توزيع عشوائي ويكون الأنموذج ملائم وجيد . أمّا اذا كانت اكبر فالأنموذج غير ملائم وفي هذه الحالة لابدّ من إعادة المرحلة الاولى أي تشخيص أنموذج اخراً لتمثيل السلسلة وتقدير معلماتها ثم يتمّ التحقق من صحتها.
إن إحصاءه *Q افضل من Q لان مستوياتها المعنوية قريبة من القيم النظرية باستخدام تقريب توزيع χ^2 كما بينه Ansley و Newbold في عام 1979 [11].

2– استعمال مجال الثقة لمعاملات الارتباط الذاتي للبواقي المقدرة ، ويحسب مجال الثقة كما يلي[8].

$$-1.96/\sqrt{n} \;\le\; \rho_k \le\; 1.96/\sqrt{n} \qquad (35)$$

وحدود الثقة لـ 95% تكون:

$$\Pr\left\{-1.96/\sqrt{n} \;\le\; \rho_k \le 1.96/\sqrt{n}\right\} = 0.95 \qquad (36)$$

اي أنه بحدود الثقة 95% فإن معاملات الارتباط الذاتي للبواقي يجب ان تقع في هذا المجال. فاذا تحقق ذلك فهذا يدل على إن البواقي تتوزع عشوائيا , وان الأنموذج يمكن استخدامه للتنبؤ.

ولغرض تحديد الأنموذج الأفضل، توجد معايير عديدة منها:–
(1) متوسط مربعات الخطأ Mean Square Error (MSE). وهو عبارة عن مجموع مربعات الأخطاء مقسوما على عدد القيم[4] .

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^2 \qquad (37)$$

(2) متوسط نسب الأخطاء المطلقة Mean Absolute Percentage Errors (MAPE). وهو عبارة عن متوسط القيمة المطلقة لكل خطأ مقسوما على عدد القيم وصيغته كما يأتي [4] :

$$\text{MAPE} = \sum_{t=1}^{n}(|e_t|/X_t)/n \qquad (38)$$

(3) معيار معلومات أكياكي Akaike Information Criterion (AIC). في عام 1974 اقترح العالم الياباني Akaike معياراً اسماه اختصاراً بـ (AIC) وصيغته: [12]

$$AIC = n\ Ln\left(\sigma_e^2\right) + 2\ M \qquad (39)$$

إذ أن

M تمثل العدد الكلي لمعلمات الأنموذج المقدرة. فإذا كان الأنموذج يحتوي على الحد الثابت فإن (M= p+q+1)، أما إذا كان خاليا من الحد الثابت فإن (M= p+q).

$\left(\sigma_e^2\right)$تباين البواقي (Residuals Variance).

(4) معيار معلومات بيز – شوارز Bayesian Schwarz Information Criterion (BIC). [13] ،[14].

في عام 1978 قدم الباحث Gideon E. Schwarz معيارا جديدا عرف باسم Schwarz وصيغته هي :

$$BIC = n\ Ln\left(\sigma_e^2\right) + M\ \ Ln\left(n\right) \qquad (40)$$

إذ أن n و M هي عدد مشاهدات السلسلة والعدد الكلي لمعلمات الأنموذج على التوالي.

(5) معيار حنان–كوئين Hannan-Quinn (H-Q).

في عام 1979 اقترح الباحثان Hannan و Quinn معيارا لتحديد رتبة الأنموذج ويرمز له بـ H – Q وصيغته الرياضية هي [15]،[16].

$$H - Q = n\ Ln\left(\sigma_e^2\right) + 2\,M\ (LnLn\ \ (n)) \qquad (41)$$

وجدير بالذكر إن الأنموذج الأفضل هو الذي يحقق اقل قيمة وفق كل معيار.

رابعا– التنبؤ (Forecasting)

عندما يصبح الأنموذج مشخصاً، مقدراً، ومناسباً بتحقق جميع شروط السلسلة الزمنية المستقرة تبدأ المرحلة الأخيرة من مراحل بناء السلسلة الزمنية بالتنبؤ بالقيم المستقبلية $t+L$. فإذا فرضنا أنّ $\left(\left(X_t\left(L\right)\right)\right)$ تمثل هذه القيمة عند الزمن الاصل (t)، فإنّنا نستطيع الحصول على التنبؤات بأخذ التوقع الشرطي عند زمن الاصل *(t)* للأنموذج وبعد كتابته عند الفترة (t+L) أي $E_t(X_{t+L} / X_t, X_{t-1}, \ldots)$ إذ باستخدام التوقع الشرطي نحصل على التنبؤات. فإذا فرضنا أنّ الأنموذج الملائم والذي تمّ التوصل اليه عن طريق المراحل السابقة هو ARIMA(p,d,q) فإننا نستطيع التنبؤ بالقيم المستقبلية الى(L=1, 2, ..) فترة قادمة وباستعمال صيغة معادلة الفروق(Difference equation form) لهذا الأنموذج وكما يأتي [4]:

حيث يمكن كتابة المعادلة(25) على النحو الآتي:

$$X_t = \phi_1 X_{t-1} + \ldots\ldots + \phi_{p+d} X_{t-p-d} - \theta_1 e_{t-1} - \ldots\ldots - \theta_q e_{t-q} + e_t \qquad (42)$$

وبكتابة المعادلة (42) عند الفترة *(t+L)* واخذ التوقع الشرطي عند الفترة *(t)* نحصل على معادلة التنبؤ عند زمن الاصل *(t)* والى *(L=1, 2, …)* فترة تنبؤيه. أي أنّ

$$E(X_{t+l}) = X_t(L)$$
$$= \phi_1 E(X_{t+L-1}) + ... + \phi_{p+d} E(X_{t+L-p-d}) - \theta_1 E(e_{t-1}) - ..... - \theta_q E(e_{t+L-q}) + E(e_{t+L}) \quad (43)$$

2- الجانب التطبيقي :

اعتمد البحث على بيانات حقيقية تمثل أعداد المصابين بمرض التدرن في محافظة واسط للفترة من كانون الثاني 2008 لغاية كانون الاول 2012 مأخوذة من سجلات دائرة صحة واسط وكما في الجدول (1).

جدول (1): أعداد المصابين بمرض التدرن واسط للفترة من(2008– 2012)*

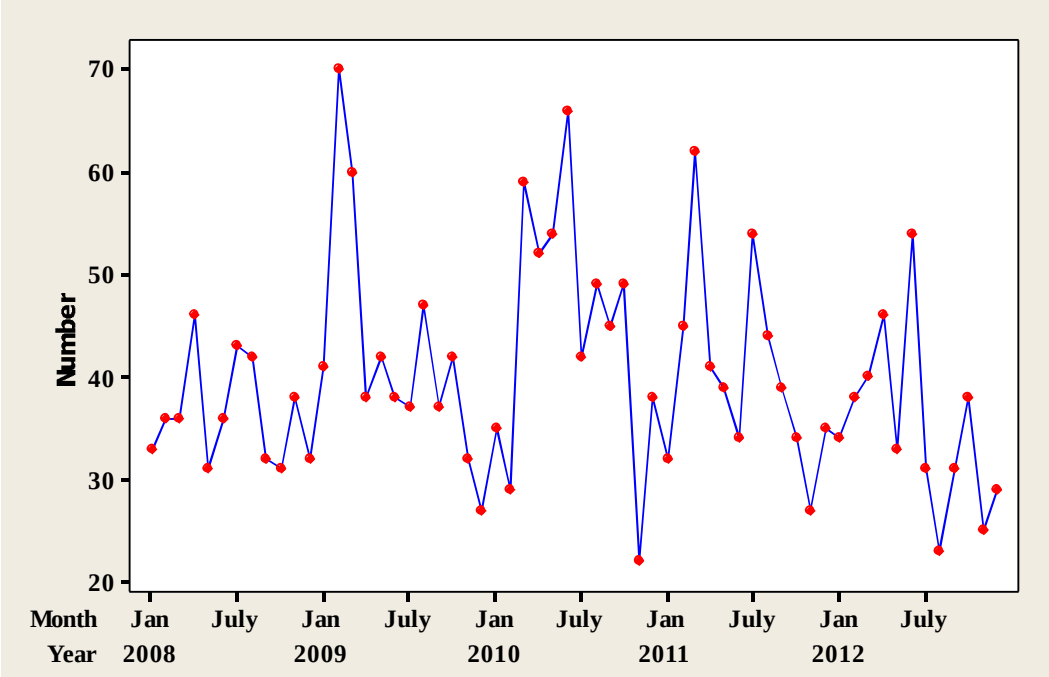
السنـــــــة / الشهر	2008	2009	2010	2011	2012
كانون الثاني	33	41	35	32	34
شباط	36	70	29	45	38
آذار	36	60	59	62	40
نيسان	46	38	52	41	46
أيار	31	42	54	39	33
حزيران	36	38	66	34	54
تموز	43	37	42	54	31
أب	42	47	49	44	23
أيلول	32	37	45	39	31
تشرين الاول	31	42	49	34	38
تشرين الثاني	38	32	22	27	25
كانون الاول	32	27	38	35	29
المجموع	422	486	540	511	436

*المصدر : دائرة صحة محافظة واسط.

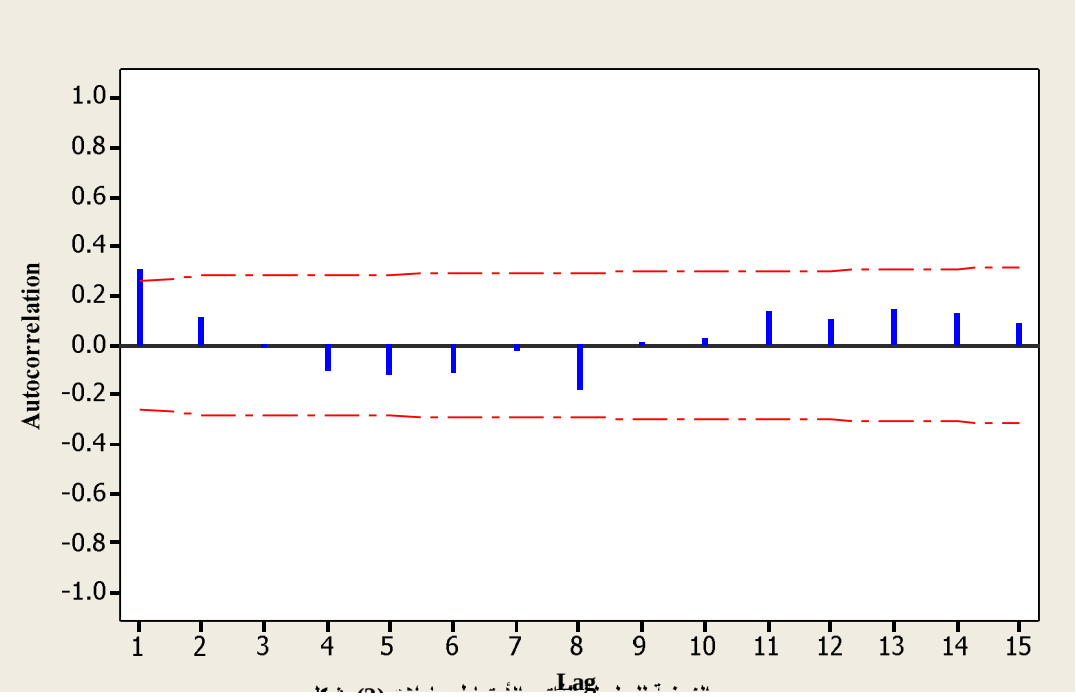
وباستعمال البرامج الإحصائية minitab16 و SPSS 18 تم تحليل السلسلة الزمنية وكما يأتي:

2-1 رسم السلسلة الزمنية

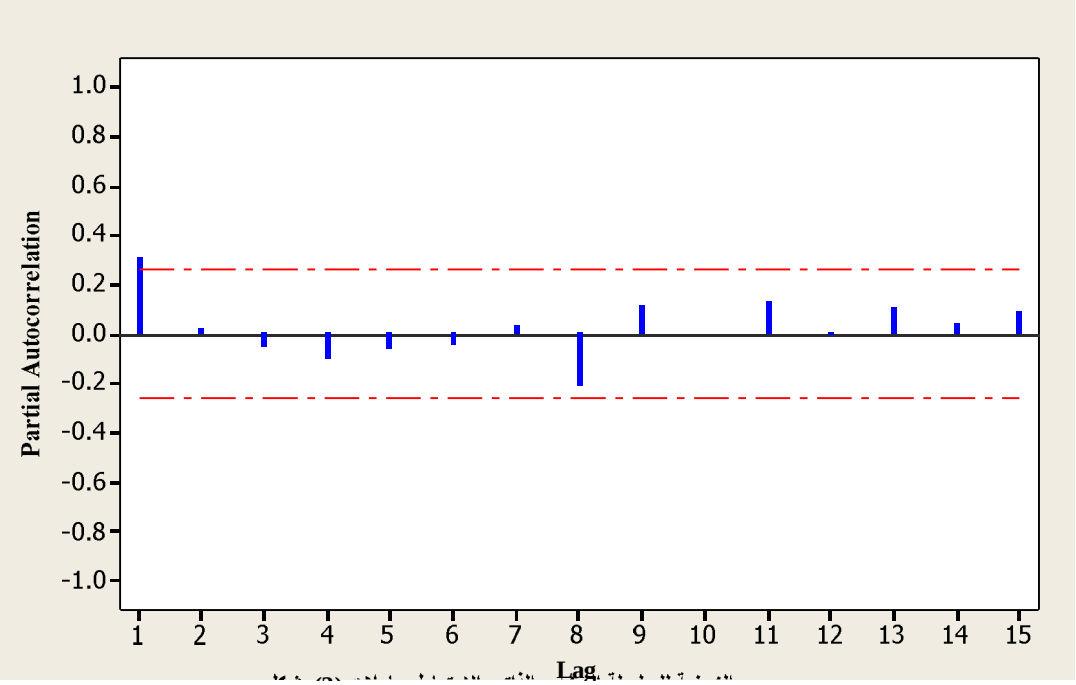
في الشكل (1) رسمت السلسلة الزمنية للتعرف على الخصائص الاولية لها حيث يلاحظ عدم وجود تأثيرات موسمية وإن التباين يميل الى الثبات وعدم وجود اتجاه عام يتزايد مع الزمن مما يدل على استقراريه بيانات السلسلة في المتوسط . ولمزيد من الدقة تم رسم واستخراج قيم معاملات الارتباط الذاتي (ACF) وقيم معاملات الارتباط الذاتي الجزئي (PACF) للسلسلة الزمنية كما في الشكلين (2) و (3) حيث نلاحظ أن جميع الارتباطات الذاتية والجزئية للسلسلة هي داخل حدود الثقة ($-0.255 \leq r_k \leq +0.255$) ، وإنها معنوية عند الإزاحة الاولى فقط .



شكل (1): منحنى أعداد المصابين بالتدرن في الفترة (2008 – 2012)



شكل (2): معاملات الإرتباط الذاتي للسلسلة الزمنية



شكل (3): معاملات الإرتباط الذاتي الجزئي للسلسلة الزمنية

2-2 اختيار الأنموذج الملائم

حيث تم التعرف على الأنموذج المناسب من خلال تحديد رتبة الانحدار الذاتي (AR) والمتوسط المتحرك (MA) بالاعتماد على شكل منحني دالة الارتباط (ACF) ومنحني دالة الارتباط الذاتي الجزئي (PACF) وعند مطابقة سلوك

الدالتين أعلاه وقيمتها مع السلوك النظري لهما حيث وجد أن هناك قطع (Cut off) لدالة الارتباط الذاتي و دالة الارتباط الذاتي الجزئي بعد الإزاحة الاولى. ومن ثمّ فيمكن اقتراح ثلاثة نماذج هي ARIMA(1,0,0) ، ARIMA(0,0,1) ، ARIMA(1,0,1) .

ولغرض تحديد الأنموذج الأفضل تم احتساب ؛ متوسط مربعات الخطأ Mean Square Error (MSE)، متوسط نسب الأخطاء المطلقة Mean Absolute Percentage Errors (MAPE) ، معيار معلومات أكياكي Akaike Criterion Information (AIC) ، معيار معلومات بيز – شوارز Schwarz Information Criterion (BIC)Bayesian ، ومعيار حنان– كوئين Hannan–Quinn (H–Q).وكما موضحة في الجدول (2).

جدول (2): يمثل مقارنة لمجموعة من النماذج باستعمال عدة معايير

الأنموذج	MSE	MAPE	AIC	BIC	H–Q
ARIMA(1,0,0)	97.51	19.2582	441.267	447.500	443.700
ARIMA(0,0,1)	99.00	19.5093	451.335	459.712	454.612
ARIMA(1,0,1)	99.20	19.2921	445.165	455.553	449.220

من نتائج الجدول اعلاه يتبين ان الأنموذج ARIMA(1,0,0) او AR(1) هو الأفضل .

3-2 التقدير

بتطبيق برنامج Mnitab 16 على بيانات السلسلة تم الحصول على النتائج التالية :

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.3174	0.1258	2.52	0.014
Constant	27.154	1.275	21.29	0.000
Mean	39.781	1.868		

Number of observations: 60

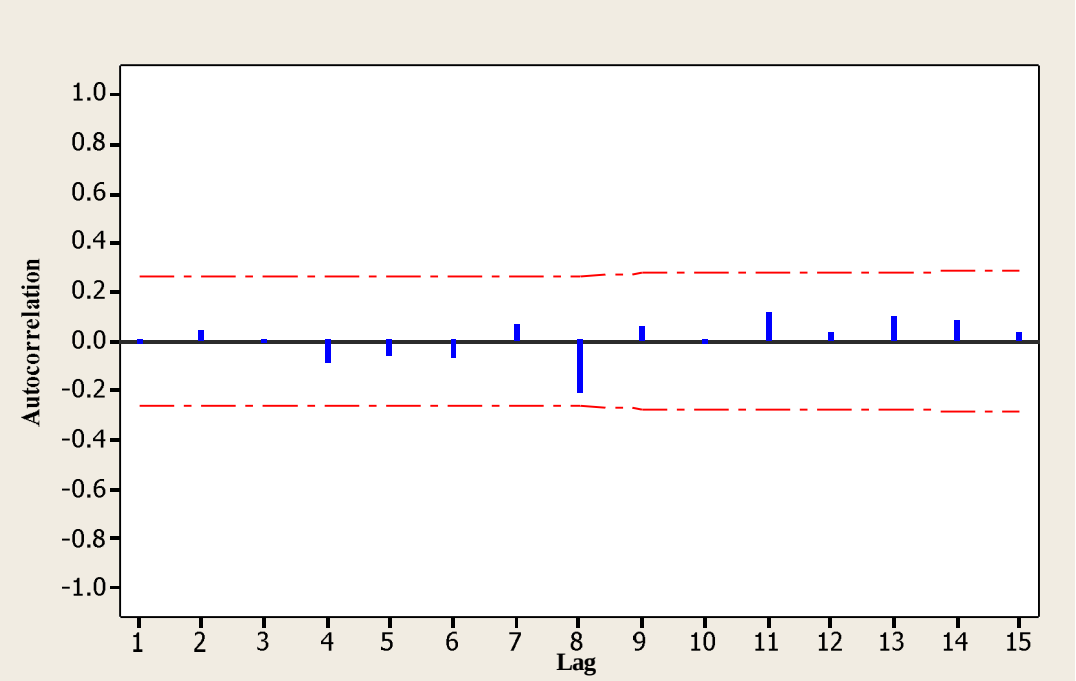
Residuals: SS = 5655.53 (backforecasts excluded)

MS = 97.51 DF = 58

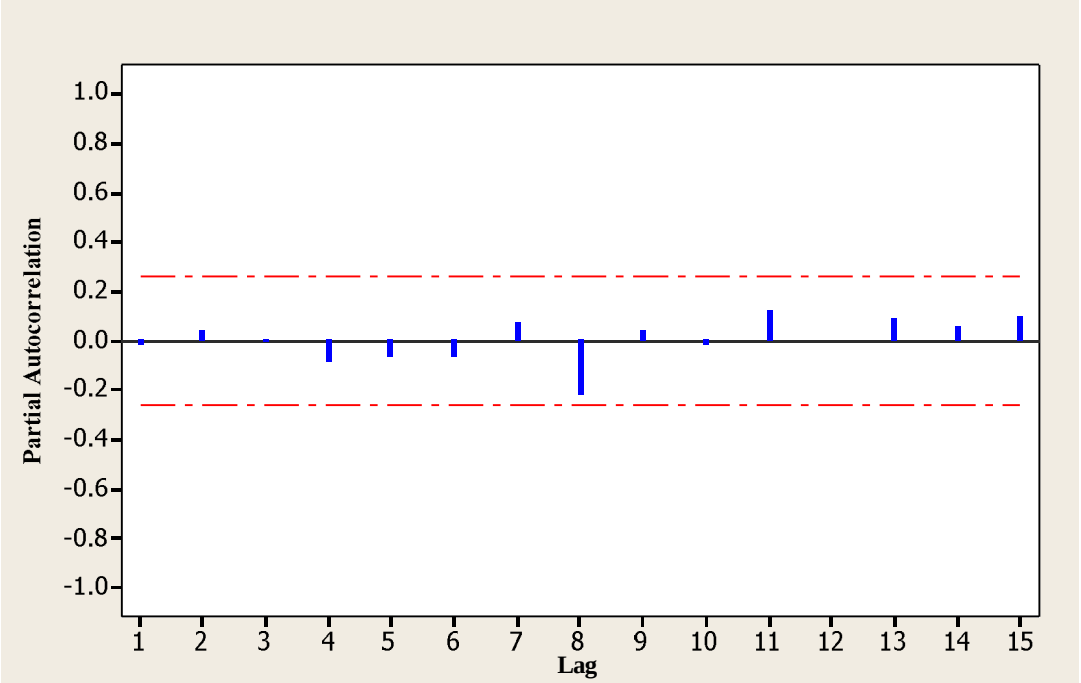
نلاحظ من النتائج أعلاه أن معالم الأنموذج معنوية من الناحية الإحصائية

ولغرض التأكد من معنوية معالم الأنموذج وملائمته وكفائتة للتنبؤ ، تم عمل الآتي :

أولاً – استخراج معاملات الارتباط الذاتي والارتباط الذاتي الجزئي لبواقي الأنموذج المختار وكذلك رسم هذه المعاملات في الشكلين (4) و(5) حيث يلاحظ إن جميع المعاملات تقع ضمن حدود الثقة $-0.253 \leq r_k \leq +0.253$) مما يدل على إن سلسلة البواقي عشوائية وإن أنماط الترابط الذاتي والترابط الجزئي للبواقي تتبع متسلسلة الضجة البيضاء (White Noise) أي أنها غير مترابطة، وبذلك فهي إن كانت طبيعية فهي مستقلة .



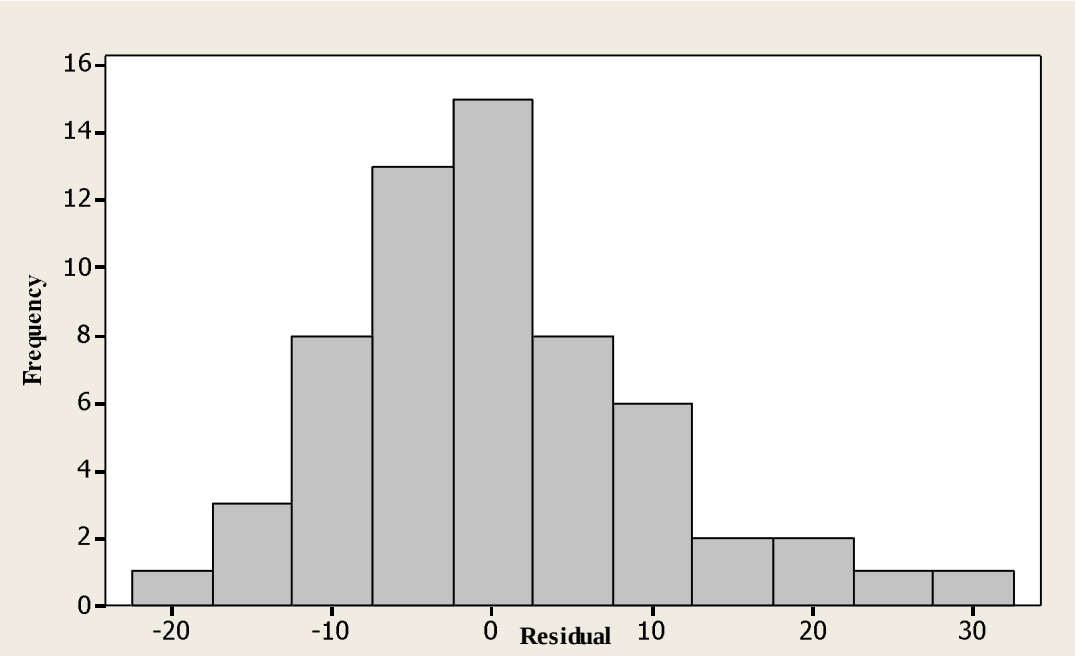
شكل (4): معاملات الإرتباط الذاتي لبواقي الإنموذج المقدر



شكل (5): معاملات الإرتباط الذاتي الجزئي لبواقي الإنموذج المقدر

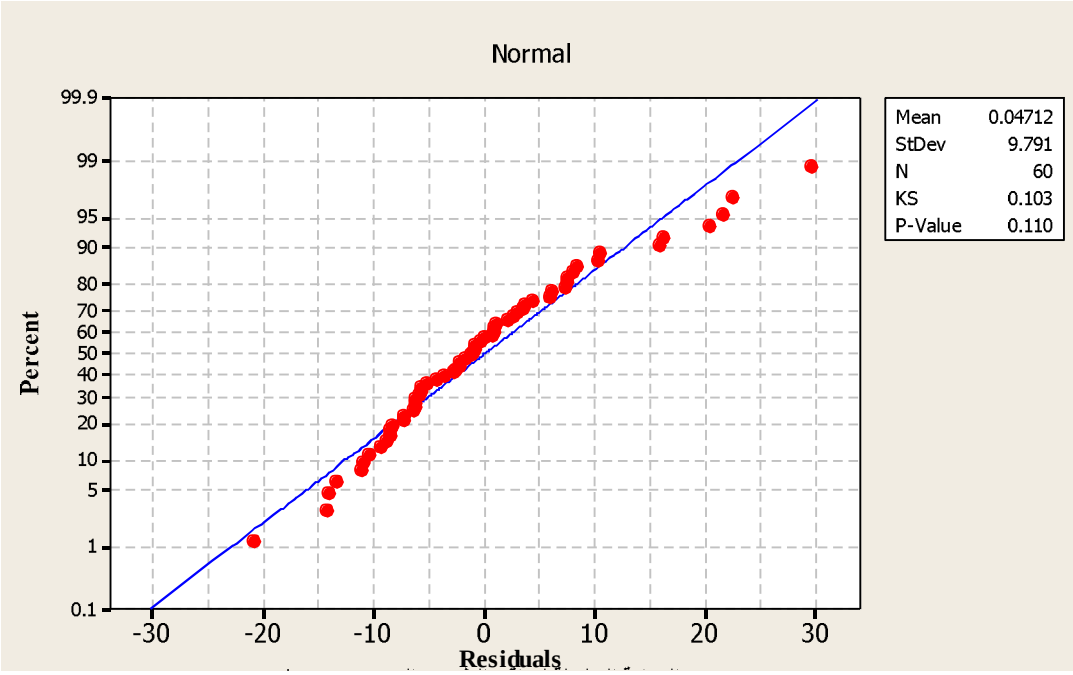
ثانياً – إجراء اختبار Z لإختبار الفرضية الصفرية القائلة بان المتوسط يساوي صفر حيث وجد انه قيمة Z=0.0001 و P-Value= 0.997 وعليه لا يمكن رفض الفرضية الصفرية بان المتوسط يساوي صفرا .

ثالثاً– رسم المدرج التكراري لبواقي الأنموذج المقدر كما موضح في الشكل (6) حيث يلاحظ انه متناظر وله شكل التوزيع الطبيعي تقريبا .



شكل (6): المدرج التكراري لبواقي الإنموذج المقدر

رابعاً– رسم الاحتمال الطبيعي (Normal Probability Plot) كما في الشكل (7) حيث يلاحظ إن البواقي تتوزع توزيعا طبيعيا .



شكل (7): التوزيع الطبيعي لبواقي السلسلة الزمنية

خامساً – إجراء اختبار Ljung- Box لاختبار معنوية الارتباطات الذاتية للسلسلة الزمنية عند الإزاحات (12,24,36,48) حيث عرضت قيم *Q مع درجات الحرية المقابلة لها كما موضحة في أدناه . وبما أن قيم P– value جميعها اكبر من 0.05 ، عليه لا يمكن رفض الفرضية الصفرية القائلة بعدم معنوية دوال الارتباط عند الأزاحات المذكورة. أي أن الإختبار يعزز إختبار إستقلالية البواقي .

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

Lag	12	24	36	48
Chi-Square	6.0	17.9	26.2	34.4
DF	10	22	34	46
P-Value	0.811	0.712	0.829	0.895

كل هذه الإجراءات أدت إلى قبول الأنموذج إحصائيا وأعطت دليلا على أنه جيد وملائم وكفوء للسلسلة الزمنية قيد البحث .

4-2 التنبؤ

باستخدام الأنموذج الذي اختير تم التنبؤ بأعداد المصابين بالتدرن في محافظة واسط في سنتي 2013 و 2014 وكما في الجدول (3).

جدول (3): الأعداد المتنبئ بها للإصابة بمرض التدرن في محافظة واسط في سنتي 2013 و 2014

السنة / الشهر	2013	2014
كانون الثاني	36	40
شباط	38	40
آذار	39	40
نيسان	40	40
أيار	40	40
حزيران	40	40
تموز	40	40
أب	40	40
أيلول	40	40
تشرين الاول	40	40
تشرين الثاني	40	40
كانون الاول	40	40

3-الاستنتاجات والتوصيات

3-1 الاستنتاجات :

من أهم الاستنتاجات التي توصل اليها البحث هي :

أولاً – بينت الاختبارات الإحصائية أن السلسلة الزمنية مستقرة في التباين وعدم وجود اتجاه عام وكذلك عدم احتوائها على المركبة الموسمية.

ثانياً – من خلال اختبار أفضل أنموذج من بين النماذج الممكنة وكذلك عن طريق فحص ملائمة الأنموذج المقترح إحصائيا وجد أن الأنموذج الملائم والكفوء لتمثيل بيانات السلسلة الزمنية هو ARIMA (1,0,0) او AR(1) لحصوله عل اقل قيمة للمعايير MSE , MAPE , AIC, BIC , H-Q .

ثالثاً – إن مجموع اعداد الإصابات بالتدرن المتنبئ بها للأشهر الثلاث الاولى لسنة 2013 وكما في الجدول (3) كانت 113 مصاباً ، حيث أظهرت تناسقا مع الأعداد الفعلية المكتشفة للمصابين التي زودتنا بها دائرة صحة واسط مؤخرا للأشهر الثلاث الأولى لسنة 2013 إذ بلغت مجموع الإصابات 112 مصاباً مما يدل على أن الأنموذج المختار كان هو الأفضل لتمثيل البيانات .

3-2 التوصيات:

من خلال الاستنتاجات التي تم التوصل لها نوصي بالآتي:

أولاً – الأخذ بنتائج هذا البحث والأنموذج الذي اعتمد لغرض التنبؤ بأعداد المصابين بهذا المرض.

ثانياً – إجراء دراسات مماثلة في باقي محافظات العراق .

المصادر :

1. Bisgaard S.and Kulahci M., *Time Series Analysis and Forecasting by Example*, Johin wiley & Sons. Inc, New Jersey, 2011.
2. Voind,H.D., *Time Series analysis*, Economic Fordham University, Bronex, New York,USA, 1999.
3. Joseph George , *The Box-Jenkins forecasting technique*, 2006.
4. Wei, William W. S. , *Time Series Analysis: Univariate and Multivariate Methods , Addison* , Wesley publishing Company, 1990.
5. Makridakis, Spyros and Others, *Forecasting: Methods and Application*, 3rd ed., John Wiley & Sons. Inc, 1998.
6. Enders, W., *Applied Econometric Time Series*, John Wiley and Sons Inc, 1995.
7. Pirece,A.D.,"*Least squares Estimation in regression Model with Autoregression- Moving Average Errors*" , *Biomatrika*, Vol.58, No. 2, PP. 299-321, 1971.
8. Wegman, E.J, *Time Series Analysis – Theory, Data Analysis and Computation*, Addison-Wesley Publishing Company Inc., U.S.A, 2000.
9. Box, G. E. P. and Pierce, David A. , "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models*", JASA*, Vol. 65, No. 332, PP. 1509-1526,1970.
10. Ljung, G. M. and Box, G. E. P. "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, Vol. 65, No. 2, PP. 297-303, 1978.
11. Ansley, Craig F. and Newbold, Paul ,"On the Finite Sample Distribution of Residual Autocorrelations in Autoregressive Moving Average Models" , *Biometrika*, Vol. 66, No. 3, PP. 547-553,1979.
12. Akaike, Hirotugu,*" A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting*", *Biometrika*, Vol. 66, No. 2, PP. 237-242,1979.
13. Schwarz, G.," Estimating the Dimension of a Model", *Annals of Statistics*, Vol. 6, No. 2, PP. 461-464, 1978.
14. Hurvich, Clifford M. and Tsai, Chih-Ling ,*"* A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection", *Journal of Time Series Analysis*, Vol. 14, No. 3, PP. 271-279,1993.
15. Hannan, E. J. and Quinn, B. G. ," The Determination of the Order of Autoregressive", *Journal of Royal Statistical Society*, Ser. B.41,PP.(190-195), 1979
16. Hannan, E. J. ," The Estimation of the Order of an ARMA Process" ,*The Annals of Statistics*., Vol. 8, No. 5, PP. 1071-1081, 1980.