

تحليل النماذج العشوائية للسلاسل الزمنية غير المستقرة

سرطان الدم (اللوكيميا) في محافظة البصرة .

. . . دريد حسين بدر .

في هذا البحث تمت المحاولة إلى دراسة أحد الامراض السرطانية وهو مرض سرطان الدم (اللوكيميا) في محافظة البصرة ، اذ تم أخذ النماذج العشوائية للسلاسل الزمنية غير المستقرة وتحليلها والتعرف على اهم الخواص في بناء النماذج المناسبة للظاهرة المدروسة اذ تم اعتماد المراحل الأساسية في بناء النماذج الخاصة بالسلاسل الزمنية ابتداء من عملية التشخيص وحتى عملية وضع النموذج المناسب والتنبؤ بالظاهرة المدروسة وتم استخدام طريقة Box-Jenkins في بناء النموذج وكذلك اختبار هذه النماذج باستخدام اختبار (Box-Ljung) إذ تم استخدام البيانات الحقيقية لمرض سرطان الدم (اللوكيميا) في محافظة البصرة ، وتم استخدام النموذج المشخص وهو $ARIMA(0,1,2)$ في عملية التنبؤ للأعوام 2018-2024 .

الكلمات الأفتتاحية : السلسلة الزمنية ، النماذج العشوائية للسلاسل الزمنية غير المستقرة ، أنموذج الانحدار الذاتي التكاملي ، أنموذج الاوساط المتحركة التكاملي ، طريقة بوكس – جينكنز ، معيار معلومات بيز ، طريقة الأماكن الأعظم ، طريقة المربعات الصغرى ، طريقة العزوم .

Assistant Lecture: Duraid Hussein Badr

Abstract

In this research ,I've tried to study one of the blood cancer diseases which we called 'Leukemia ' in Basra governorate . In this study I've taken the unstable stochastic models, through which an analysis has been made to identity the most significant properties in the process of constructing the suitable models for the intended case study , in such a way that I have depended on the basic stages in building the specified time series models beginning from the diagnose process up to the stage of producing the appropriate model ,and predicting the given phenomenon using Box-Jenkins way that's used to reconstruct the given model . Moreover, this would help in the process of analyzing these models by using (Box _ Ljung) test in which realistic data of blood cancer disease are used in the research, Finally ,it is worth to mention that the diagnosed model which put to the process of analysis is $ARIMA(0,1,2)$ in the predicting process for years 2018 -2024 .

Keyword : Time series , Stochastic models for non- stationary time series , $ARI(p,d)$ Autoregressive integrated model , $IMA(d,q)$ Integrated moving average model , Box – Jenkins method , Bayesian criterion , Maximum Likelihood Method , Least squares method , Moment method .

(Introduction)

1-1. تمهيد :

من المعروف أن من أهم عمليات تطور البلدان هي عملية التخطيط ووضع الخطط المفصلة المستقبلية وهذا يتطلب اعتماد الأساليب الإحصائية المتقدمة ومن أهم هذه الأساليب طرائق تحليل السلاسل الزمنية التي تأخذ بالحسبان التغيرات الزمنية في دراسة الظواهر وعلى هذا الأساس تم التركيز في هذه الدراسة على مرحلة تشخيص واختبار النماذج في السلاسل الزمنية غير المستقرة لكونها المرحلة الأصعب والأهم في عملية التحليل، فقد تم تناول بعض المفاهيم الأساسية للسلسلة الزمنية والنماذج غير المستقرة الخاصة بطريقة (Box & Jenkins) وكذلك الأساليب الأساسية المستخدمة في تشخيص رتبة النماذج الإحصائية فضلاً عن استخدام الاختبارات المهمة في تشخيص هذه النماذج وأختيار الأنموذج الملائم لهذه الأمراض باستخدام بيانات حقيقية لمرض سرطان الدم (اللوكيميا) في محافظة البصرة.

2-1. _____ :

أن الأمراض السرطانية من الأمراض التي تتسبب في معاناة المرضى وعدم إمكانية القيام بالأعمال الاعتيادية التي تكفل لهم العيش مما يتسبب بخسائر مادية كبيرة للمرضى وبالتالي للبلد خاصة (وانطلاقاً من أهمية دراسة الجوانب الطبية وبخاصة الأمراض السرطانية في العراق، فقد تم استخدام النماذج الإحصائية للسلاسل الزمنية في دراسة لبعض الأمراض السرطانية مثل أمراض سرطان الدم (اللوكيميا)) ، ونتيجة لخطورة هذا المرض فقد تم تناوله في الدراسة باستخدام طريقة (Box & Jenkins) للنماذج العشوائية للسلاسل الزمنية غير المستقرة وتحليلها فضلاً عن استخدام الاختبارات المهمة في تشخيص هذه النماذج ، واختيار الأنموذج الملائم لهذه الأمراض باستخدام بيانات حقيقية لمرض سرطان الدم (اللوكيميا) في محافظة البصرة .

(Objective of research)

3-1 . هدف البحث :

تهدف الدراسة الى :

اولاً: دراسة تحليل (تشخيص واختبار) النماذج المختلفة للسلسلة الزمنية ذات الرتب المختلفة ، وذلك باستخدام طريقة (Box & Jenkins) ومعيار (BIC) وتحديد الأنموذج الامثل للتنبؤ بالبيانات الخاصة بامراض سرطان الدم (اللوكيميا) في محافظة البصرة .

ثانياً: توفيق الأنموذج الامثل للإصابة بأمراض سرطان الدم (اللوكيميا) في محافظة البصرة اعتماداً على البيانات للمرض ومعيار بيز (BIC) ومتوسط مربعات خطأ التنبؤ اقل ما يمكن (MSE) ومن ثم التنبؤ بالقيم المستقبلية.

2 . الجانب النظري .

بعض التعاريف والمفاهيم الاساسية في تحليل السلاسل الزمنية:

1.2 السلسلة الزمنية Time series (2):

السلسلة الزمنية هي مجموعة من المشاهدات المرتبطة مع بعضها يتم تسجيلها في فترات زمنية متعاقبة لظاهرة معينة وتكون على نوعين سلاسل زمنية متقطعة (Discrete time series) وسلاسل زمنية مستمرة (Continuous time series) ولكن السلاسل الزمنية الاكثر استخداماً في المجال التطبيقي هي السلاسل الزمنية المتقطعة التي تكون المدة الزمنية بين مشاهدة واخرى متساوية وهذه يمكن الحصول عليها إما عن طريق تسجيل قيم الظاهرة في ازمنة ثابتة او عن طريق تجميع قيم الظاهرة لمدة زمنية ثابتة. واذا استطعنا من المشاهدات السابقة للسلسلة الزمنية التنبؤ بصورة دقيقة عن السلوك المستقبلي للظاهرة التي تمثلها السلسلة الزمنية فإننا نسميها سلسلة زمنية غير العشوائية (non stochastic time series)، اما اذا استطعنا التعرف على الهيكل الاحتمالي للسلوك المستقبلي للظاهرة فقط فتسمى في هذه الحالة بالسلسلة الزمنية العشوائية (stochastic time series). إنّ التغير في قيم السلسلة الزمنية من مدة الى اخرى يمكن ان يعزى الى تأثير عدد من المركبات اي ان قيمة السلسلة الزمنية في مدة زمنية معينة تتحدد نتيجة تأثير المركبات وهي : مركبة الاتجاه ومركبة الموسمية و مركبة الدورية ومركبة العشوائية .

2.2 الاستقرارية Stationary⁽¹⁾:

المقصود بأستقرارية السلسلة الزمنية إنَّ مشاهداتها تتذبذب بشكل عشوائي حول المتوسط ،
وتكون على نوعين:

1-2-2. الاستقرارية التامة Strictly stationary :

يقال للسلسلة الزمنية $(Y_t ; t=1,2,\dots,n)$ بانها مستقرة استقرارية تامة إذا كان
التوزيع المشترك لاية مجموعة من المشاهدات لايتأثر بازاحة كل المدة الزمنية للمشاهدات
الى الامام او الى الخلف بأية كمية صحيحة

2-2-2. الأستقرارية الضعيفة او من الدرجة الثانية weakly or second – order stationary:

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

$$1) E(Y_t) = \sim \dots\dots\dots(2.1)$$

إذ ان \sim هو متوسط العملية العشوائية ويكون ثابتاً ولا يعتمد على قيم t والذي يقدر من
مشاهدات السلسلة الزمنية كما هو آت :

$$\hat{\sim} = \bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t \dots\dots\dots(2.2)$$

(2) وتباين السلسلة الزمنية يكون ثابتاً

$$\text{var}(Y_t) = E(Y_t - \sim)^2 = \dagger_y^2 = x_o \dots\dots\dots(2.3)$$

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

$$\hat{x}_o = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2 \dots\dots\dots(2.4)$$

(3) التباين المشترك للسلسلة الزمنية او (التغاير الذاتي)

$$\text{cov}(Y_t, Y_{t+k}) = E(Y_t - \sim)(Y_{t+k} - \sim) = x_k \dots\dots\dots(2.5)$$

إذ إن X_k هو التغير الذاتي Auto covariance للعملية العشوائية عند الازاحة k (Lag k) ويكون ثابتاً لا يعتمد على قيم t لجميع القيم الصحيحة الى k والذي يقدر من مشاهدات السلسلة الزمنية كما هو آت:

$$\hat{X}_k = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) \quad \dots\dots\dots(2.6)$$

$$\bar{Y} = \frac{\sum_{t=1}^T Y_t}{T} \quad \text{حيث}$$

إن المجموعة $(X_k ; k = 0, \pm 1, \dots)$ تسمى بدالة التغير الذاتي Auto covariance function. وقد سميت السلسلة ذات الاستقرار الضعيفة بالسلسلة ذات الاستقرار من الدرجة الثانية، لأن كلاً من العزمين الأول والثاني يكونا موجودين وثابتين مع الزمن وإن التغير الذاتي يعتمد على الازاحة k فقط .

3.2 عدم الاستقرارية Non Stationary :

عندما تتغير خصائص السلسلة الزمنية بتغير الزمن تكون السلسلة غير مستقرة، وعدم وجود الاستقرار في السلاسل الزمنية يعني حدوث تغييرات نظامية حول الوسط والتباين وقد يؤدي وجود تغييرات دورية في السلسلة احياناً الى عدم استقراريتها (5) ، وقبل تحليل السلسلة الزمنية يجب ازالة عدم استقراريتها، لأن وجوده سيعطي نتائجاً مضللة. وهناك نوعان من عدم الاستقرارية:

1.3.2 عدم الاستقرارية حول التباين Non Stationary around the variance :

عندما تنذبذ السلسلة الزمنية حول تباين غير ثابت فإن هذا يعني عدم الاستقرارية حول التباين.

عن طريق استخدام التحويل اللوغاريتمي او الأسّي (تحويلات القوى) (Power transformation) تتحول السلسلة غير المستقرة حول التباين الى المستقرة وتعد تحويلات القوى المعرفة ابسط مجموعة تحويلات وهي كالاتي :

$$Y_t^{(3)} = \left\{ \begin{array}{ll} Y_t & \text{if } \{ \neq 0 \} \\ \ln Y_t & \text{if } \{ = 0 \} \end{array} \right\} \dots\dots\dots(2.7)$$

$Y_t^{(3)}$: السلسلة المحولة عند الزمن t.

Y_t : قيمة السلسلة غير المحولة.

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

2.3.2 عدم الاستقرار حول الوسط Non stationary around the mean

عدم تذبذب السلسلة الزمنية حول وسط ثابت يعني عدم استقرارية الوسط ويمكن أزالته بأخذ الفروق (differences) المناسبة، وهذه الطريقة تعد من الطرائق الكفوءة والسهلة نسبياً وتتناسب مع نماذج السلسلة الزمنية.

وتتلخص طريقة اخذ الفروق فيما يأتي:

تؤخذ الفروق بين المدة المتعاقبة لتكوين سلسلة جديدة للوصول الى الاستقرارية.

$$\nabla^1 Y_t = Y_{t+1} - Y_t \dots\dots\dots(2.8)$$

أن ∇ : رمز لمعامل الفروق.

وتضم السلسلة الجديدة (n-1) من القيم وستكون مستقرة اذا كانت المشاهدات الاصلية خطية من الدرجة الاولى، واذا كانت الارتباطات الذاتية للبيانات التي أخذ لها الفرق الاول لا تهبط الى الصفر بعد الفترة الثانية او الثالثة فيجب أخذ الفروق الاولى للبيانات التي تم أخذ الفرق الاول لها ويكون ذلك كما يأتي:

$$\nabla^2 Y_t = \nabla^1 Y_{t+1} - \nabla^1 Y_t \dots\dots\dots(2.9)$$

إذ يشكل $\nabla^2 Y_t$ سلسلة الفروق الثانية.

وبتعويض معادلة رقم (2.8) في المعادلة (2.9) نحصل على:

$$\begin{aligned} \nabla^2 Y_t &= (Y_{t+2} - Y_{t+1}) - (Y_{t+1} - Y_t) \\ &= Y_{t+2} - 2Y_{t+1} + Y_t \end{aligned}$$

وهكذا الحال فيما يخص الفروقات الاخرى.

فبعد اخذ عدد مناسب من الفروقات للسلسلة الزمنية الاصلية Y_t ولتكن d وحينما يكون $d=0,1,2,\dots$ نحصل على سلسلة زمنية جديدة W_t مستقرة.

$$W_t = \nabla^d Y_t$$

عن طريق ذلك نستطيع ايجاد قيم السلسلة الزمنية الاصلية Y_t بعد معرفة قيم W_t وذلك من خلال استخدام عامل الجمع \sum إذ :

$$Y_t = \sum_{t=1}^d W_t$$

$$= W_1 + W_2 + \dots + W_d$$

إنّ عملية الفروقات مهمة ميكانيكية متعاقبة حتى تهبط الارتباطات الذاتية الى الصفر عن طريق فترتين او ثلاث فترات تخلف زمنية (6).

4.2 النماذج العشوائية وغير العشوائية

:Stochastic and non stochastic models

إذا استطعنا تمثيل ظاهرة معينة بأنموذج معين بحيث يمكننا من حساب قيمة الظاهرة بشكل دقيق عند أية مدة زمنية نسمي مثل هذا الأنموذج بالأنموذج غير العشوائي، أمّا إذا استطعنا تمثيل الظاهرة بأنموذج يمكننا من حساب الاحتمالية للقيمة المستقبلية للظاهرة واقعة بين قيمتين محددة نسمي مثل هذا الأنموذج بالأنموذج العشوائي. وسأتناول في دراستي هذه النماذج العشوائية للسلاسل الزمنية غير المستقرة.

5.2 النماذج العشوائية للسلاسل الزمنية غير المستقرة

Stochastic models for non- stationary time series

إن معظم السلاسل الزمنية التي تواجهنا في التطبيقات العملية تكون غير مستقرة، إذ يتمّ التعرف عليها عن طريق دالة الارتباطات الذاتية إذ لاتؤول قيمها للصفر بعد الازاحة الثانية او الثالثة وإنّما تبقى قيمها كبيرة لعدد من الازاحات (7).

ولغرض معالجة هذا النوع من السلاسل يجب تحويلها الى سلاسل مستقرة ويتم ذلك باستخدام عامل الفروق الخلفية (Backward difference operator) والذي يعرف بالشكل الاتي :

$$\nabla Y_t = (1-s)Y_t = Y_t - Y_{t-1} \quad \dots\dots\dots(2.10)$$

وبشكل عام والى d من الفروق غير الموسمية يعرف بالشكل الاتي:

$$\nabla^d Y_t = (1-s)^d Y_t \quad \dots\dots\dots(2.11)$$

فاذا عوضنا عن Z_t بـ $\nabla^d Z_t$ في المعادلة التالية

$$w(s)Z_t = {}_n(s)u_t \quad \dots\dots\dots(2.12)$$

نحصل على انموذج جديد يستطيع معالجة نوع معين من السلاسل الزمنية غير المستقرة التي تسمى المتجانسة (Homogeneous) والذي يطلق عليها اسم الانموذج المختلط غير المستقر والذي يرمز اليه بـ ARIMA(p,d,q) وصيغته:

$$w(s)\nabla^d Z_t = w(s)(1-s)^d Z_t = {}_n(s)u_t \quad \dots\dots\dots(2.13)$$

واذا عوضنا عن $\nabla^d Z_t$ بـ (W_t) فإنّ الانموذج ARIMA(p,d,q) للسلسلة (Z_t) يتحول الى الانموذج ARMA(p,q) للسلسلة (W_t) اي ان

$$w(s)W_t = {}_n(s)u_t \quad \dots\dots\dots(2.14)$$

لذلك فإنّ المفاهيم النظرية كافة لنماذج ARMA(p,q) نستطيع تطبيقها على نماذج ARIMA(p,d,q).

السلسلة الزمنية التي ليس لها وسط ثابت عبر الزمن فإنها غير مستقرة ففي الانموذج الاتي:

$$Y_t = \tilde{\mu}_t + X_t$$

$\tilde{\mu}_t$: متوسط غير ثابت (non constant mean) لذلك فان السلسلة Y_t تعد غير مستقرة. ونادراً ما توجد سلسلة زمنية مستقرة في التطبيقات العملية إذ غالباً ما تحتوي السلسلة عامل الاتجاه ويتم تحويل السلسلة غير المستقرة الى سلسلة مستقرة بواسطة أخذ عدد معين من الفروق للسلسلة الاصلية :

$$W_t = \nabla^d Y_t$$

إذ إن:

W_t : السلسلة المستقرة.

d : درجة الفرق المأخوذ للسلسلة الزمنية.

Y_t : السلسلة الاصلية (غير المستقرة).

∇ : عامل الفروق (Difference operator).

بعد اخذ الفرق او الفروق تتم اضافة كلمة (Integrated) الى اسم الانموذج للتمييز بين

النوعين من النماذج ومن الانواع الشائعة للسلسلة غير المستقرة.

1.5.2 أنموذج الانحدار الذاتي التكاملی ARI(p,d)

يرمز للسلسلة الزمنية من نوع AR بعد أخذ الفروق المناسبة لتحويلها الى سلسلة

مستقرة بـ $ARI(p,d)$ وذلك أن:

p : رتبة الانحدار الذاتي (AR).

d : درجة الفرق غير الموسمي.

ويوجد كثير من نماذج الانحدار الذاتي فمثلاً الصيغة العامة للأنموذج $ARI(1,1)$ هي:

$$Y_t - Y_{t-1} = w(Y_{t-1} - Y_{t-2}) + u_t$$

والسلسلة هي :

$$Y_t = (1+w)Y_{t-1} - wY_{t-2} + u_t \quad \dots\dots\dots(2.15)$$

إذ إن: $|w| < 1$

2.5.2 أنموذج الاوساط المتحركة التكاملی IMA(d,q)

يرمز للسلسلة الزمنية من نوع MA بعد أخذ الفروق المناسبة لتحويلها الى سلسلة

مستقرة بـ $IMA(d,q)$.

إذ إن:

d : رتبة الفرق المأخوذ للسلسلة.

q : رتبة الاوساط المتحركة غير الموسمية.

مثال على هذا الانموذج، أنموذج الاوساط المتحركة $IMA(1,1)$.

$$Y_t = Y_{t-1} + u_t - \alpha u_{t-1} \quad \dots\dots\dots(2.16)$$

$$= u_t + (1 - \alpha) u_{t-1} + (1 - \alpha) u_{t-2} + \dots + (1 - \alpha) u_{-m} - \alpha u_{-m-1}$$

إذ ان :

m : الزمن قبل (t=1) وإن m < 1 ، عندما t < 0.

6.2 التنبؤ Forecasting :

يعرف التنبؤ بأنه معرفة السلوكية المستقبلية لعملية معينة بأقل خطأ ممكن ، عند مقارنتها بالواقع ومن الاهداف الرئيسية لبناء النموذج في السلاسل الزمنية إمكانية استخدام تلك النماذج للتنبؤ بالقيم المستقبلية للسلسلة، أحصائياً فاذا كانت لدينا سلسلة زمنية معينة مثل $(Y_t, Y_{t-1}, \dots, Y_1)$ فإننا نتنبأ بـ $(Y_{t+\ell})$ للسلسلة إذ أن:

t : الزمن الاصيلي او الحاضر للسلسلة (Origin time).

t + ℓ : الزمن المستقبلي (الزمن القادم) (Lead time).

$$\hat{Y}_t(\ell) = E(Y_{t+\ell} / Y_t, Y_{t-1}, \dots, Y_1) \quad \dots\dots\dots(2.17)$$

إن عملية التنبؤ، تكون عملية ناجحة إذا تم الحصول على قيمة لـ $\{\hat{Y}_t(\ell)\}$ بحيث تكون لها أقل تباين للخطأ وهناك طرائق عدة للتنبؤ في السلاسل الزمنية.

7.2 طريقة بوكس – جينكنز Box – Jenkins method (3) (4) :

إن عملية بناء انموذج ARIMA(p,d,q) لتمثيل بيانات السلسلة الزمنية واستخدامه لأغراض التنبؤ تسمى بطريقة بوكس – جينكنز إذ تعد هذه الطريقة واحدة من الطرائق العامة للتنبؤ بمختلف انواع السلاسل الزمنية (المستقرة وغير المستقرة، الموسمية وغير الموسمية)، لأنها لا تفترض وجود نمط معين في بيانات السلسلة قبل تطبيقها كما هو الحال بالنسبة لطرائق التنبؤ الاخرى مثل (طرائق التمهيد الاسي) وإنما تبدأ بنموذج تجريبي يحدد بالاعتماد على دالتي (ACF) و (PACF) وبعد ذلك يتم تقدير معالمه بالاعتماد على

مشاهدات السلسلة الزمنية بحيث تجعل اخطاء التنبؤ اقل ما يمكن، ويستخدم في هذه الطريقة عدد من المؤشرات بحيث تجعل الباحث قادراً على الحكم فيما اذا كان الانموذج ملائماً ام لا. وهناك أسباب تقف وراء أهمية استخدام هذه الطريقة وهي:

1. تستخدم هذه الطريقة في حالات السلسلة الزمنية المستقرة وكذلك السلسلة الزمنية غير المستقرة بعد أن يتم تحويلها الى السلسلة المستقرة.

2. لاتعالج هذه الطريقة النماذج ذات المتغير الواحد (univariate) فحسب، بل يمكنها معالجة النماذج ذات المتغيرات المتعددة.

3. تناسب هذه الطريقة السلاسل الزمنية المعقدة وحالات التنبؤ التي توجد فيها أنماط مختلفة في آن واحد، مما يؤدي الى اكتشاف أنموذج ملائم للبيانات ذات العلاقة مع تقليل الاخطاء الى ادنى حد ممكن.

4. إنّ القوة الحقيقية لهذه الطريقة تكمن في قدرتها على معالجة الانماط المعقدة باستعمال مجموعة من القواعد بشكل جيد نسبياً.

5. الدقة الكبيرة التي تحتويها هذه الطريقة ، لكونها تمرّ بمراحل مهمة بغية أستخلاص انموذج التنبؤ المناسب.

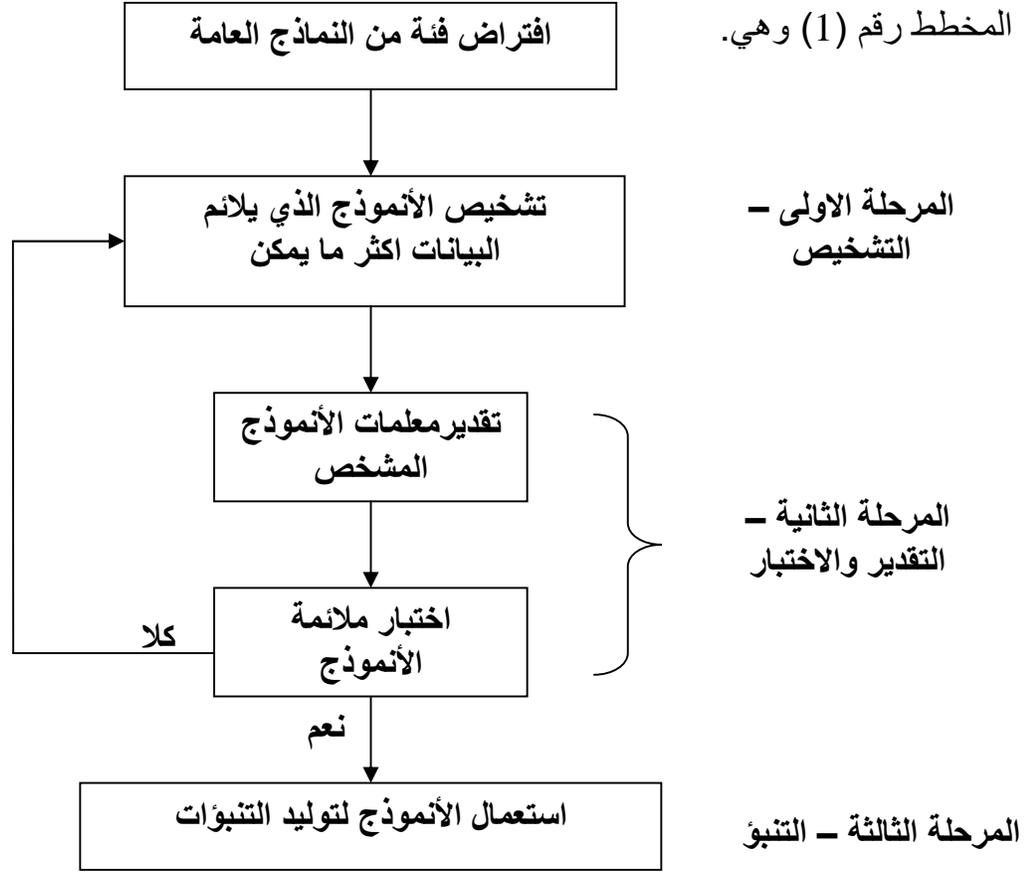
6. هذه الطريقة تتعامل مع ظروف أكثر عمومية وشمولية، مقارنة بالطرائق الأخرى للتنبؤ.

7. بُعد الافق الزمني المطلوب للتنبؤ بسبب قلة خطأ التنبؤ.

8. تتعامل هذه الطريقة مع حالات متعددة ومتنوعة للسلاسل الزمنية أدى الى زيادة درجة تعقيدها قياساً بطرائق التنبؤ الأخرى. وتتطلب طريقة بوكس – جينكنز أستعمال الحاسبة الألكترونية مما يجعل الطريقة أكثر دقة من طرائق التنبؤ الأخرى ، عندما يتم بناء أنموذج مناسب للبيانات قيد الدراسة يكون الحصول على التنبؤات بواسطة الحاسوب سهلاً جداً واذا تبدل لسبب ما نمط المشاهدات فسيكون من السهل تعديل الأنموذج طبقاً لهذه التبدلات.

1.7.2 مراحل بناء الأنموذج بحسب طريقة بوكس – جينكنز :

إنَّ مراحل بناء الأنموذج بحسب طريقة بوكس – جينكنز في التنبؤ هي كما هو موضح في



المخطط رقم (1) Box G, E.P & Jenkins (1976)

المخطط يوضح مراحل طريقة (Box – Jenkins)

1.1.7.2 المرحلة الاولى / Phase I :

التشخيص Identification :

تعد مرحلة تشخيص الأنموذج من المراحل المهمة والصعبة للوصول للأنموذج الملائم، إذ الخطوة الاولى لهذه المرحلة، هي رسم السلسلة الزمنية الاصلية للتعرف على خصائصها (الاتجاه والموسمية والدورية)، ثم اختبار السلسلة الزمنية لمعرفة مدى حاجتها الى التحويل لعدم استقراريتها حول الوسط والتباين، ثم تحقيق الاستقرارية حول الوسط وحول التباين اذا كان ضرورياً على وفق الطريقة الخاصة التي تم ذكرها في الفقرة (1.3.2) و (2.3.2).

إن المهمة الرئيسية لهذه المرحلة هي التشخيص واختبار الأنموذج الافضل، الذي يمثل السلسلة الزمنية من النماذج العامة ARMA(p,q) او ARIMA(p,d,q) بإستخدام معاملات الارتباط الذاتي والارتباط الذاتي الجزئي.

فإذا كانت معاملات دالة الارتباط الذاتي تتناقص أسياً وكانت معاملات الارتباط الذاتي الجزئي تنقطع بعد المدة الفاصلة (p) فإنّ الأنموذج هو AR(p). أما إذا تناقصت معاملات الارتباط الذاتي الجزئي أسياً وانقطعت معاملات الارتباط الذاتي بعد الفترة الفاصلة (q) فهذا يعني وجود الأنموذج MA(q). أمّا في حالة ARMA(p,q) فإنّ معاملات الارتباط الذاتي ومعاملات الارتباط الذاتي الجزئي تنحدران أسياً.

وهناك طرائق اخرى لتشخيص النموذج ومنها طريقة معيار معلومات BIC.

1. معيار معلومات بيز Bayesian information criterion :

في عام 1978 صار المعيار الجديد هو BIC وصيغته هي :

$$BIC = 2 \ln\left(\frac{1}{u}\right) + M \ln(N) \quad \dots\dots(2.18)$$

إذ إن N و M هي عدد مشاهدات السلسلة والعدد الكلي لمعلمات الأنموذج على التوالي. وعلى وفق هذا المعيار فإنّ الأنموذج الافضل هو الأنموذج الذي يعطي اقل قيمة من معيار BIC(M).

2.1.7.2 المرحلة الثانية / Phase II :

وتتكون هذه المرحلة من جزأين هما:

1.2.1.7.2 التقدير Estimation :

بعد التشخيص وتحديد رتبة الأنموذج يتمّ تقدير معلمات الأنموذج وتوجد طرائق متعددة تستخدم في التقدير منها:

1. طريقة العزوم (معادلات Yule – Walker) :

إنّ هذه الطريقة تعتمد على دالة الارتباط الذاتي للأنموذج مثلاً في حالة النموذج AR(p)

$$\dots_k = W_1 \dots_{k-1} + W_2 \dots_{k-2} + \dots + W_p \dots_{k-p} \quad \dots\dots(2.19)$$

وبالتعويض عن قيم $(k=1, \dots, p)$ نحصل على p من المعادلات الخطية التي تحتوي على p من المجاهيل.

$$\dots_1 = W_1 + W_2 \dots_1 + \dots + W_p \dots_{p-1}$$

$$\dots_2 = W_1 \dots_1 + \dots + W_p \dots_{p-2}$$

.

.

$$\dots_p = W_1 \dots_{p-1} + \dots + W_p$$

وبصيغة المصفوفات :

$$W_k = \dots^{-1} \dots_k \quad \dots\dots\dots(2.20)$$

وتقديرات Yule-Walker كمعاملات W_k نحصل عليها عن طريق التعويض عن معامل الارتباط الذاتي النظري \dots_k بمعامل الارتباط الذاتي التقديري $\hat{\rho}_k$ كالآتي:

$$W_k = R^{-1} r_k \quad \dots\dots\dots(2.21)$$

إذ ان R تمثل مصفوفة الارتباط الذاتي التقديري :

$$R = \begin{bmatrix} 1 & r_1 & r_2 & \cdot & \cdot & r_{p-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{p-1} & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}, \quad r_k = \begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ r_p \end{bmatrix}$$

فمثلاً لتقدير معاملات النموذج MA بهذه الطريقة نحصل على q من المعادلات غير الخطية المحتوية على q من المجاهيل بعدها يتم حلّ هذه المعادلات باحدى الطرائق الخاصة بحلّ المعادلات غير الخطية.

2. طريقة الامكان الاعظم Maximum likelihood :

لتقدير معاملات الأنموذج (ARMA) نستخدم الامكان الاعظم والدالة التجميعية للمعاملات بثبوت المشاهدات وهي:

$$L(w, \sigma, \sigma_u^2 / Y_t) = (2\Pi)^{-N/2} (\sigma_u^2)^{-N/2} \exp\left(\frac{-1}{2\sigma_u^2} S(w, \sigma)\right) \dots\dots\dots(2.22)$$

$$\text{Log } L(w, \sigma, \sigma_u^2) = -(N/2) \text{Ln } \sigma_u^2 - \frac{S(w, \sigma)}{2\sigma_u^2} - (N/2) \text{Ln}(2\Pi)$$

لتعظيم الدالة يجب تقليل مجموع مربعات الخطأ بحيث يكون أقل ما يمكن بالتفاضل الجزئي للمعلمات فنحصل على التقديرات.

3. طريقة المربعات الصغرى Least squares :

الفكرة الأساسية لطريقة المربعات الصغرى هي جعل مجموع مربعات الأخطاء اقل ما يمكن (3).

2.2.1.7.2 التحقق من الانموذج Model diagnostic checking:

إن الخطوة التالية لتقدير معلمات الانموذج المشخص هي اجراء تدقيق ، فيما إذا كان تحديد رتبة ونوع الانموذج ملائم ام لا. ويتم ذلك بطرائق مختلفة منها:

1. اختبار الارتباط الذاتي للبواقي (الأخطاء) Auto correlation test :

عند ايجاد دالة الارتباط الذاتي للسلسلة المولدة (Simulated series) ومقارنتها بدالة الارتباط الذاتي للعينة للسلسلة الأصلية (sample auto correlation function) توجد أختلافات كبيرة بين الدالتين يدل ذلك على عدم كفاءة الانموذج المحدد او المشخص، اذا كانت دالتا الارتباط الذاتي لاتختلفان بشكل مميز، فيجب اجراء تحليل كمي للأخطاء المتكونة من الانموذج ويفترض أن هذه البواقي تتوزع توزيعاً طبيعياً ومستقلاً بعضها عن بعض، ومتى ما كان الانموذج المشخص صحيحاً فإن \hat{u}_t يجب ان يتصف تقريباً بصفات u_t أي أنها دالة الخطأ العشوائي ويفترض عدم ترابط البواقي ببعضها الاخر (nearly uncorrelated) ، لذلك فإن الارتباط الذاتي العيني للبواقي يقترب من الصفر عندما $k > I$ وصيغة النموذج المختلط هي كما يأتي:

$$u_t = \left(\hat{\sigma}_u^{-1} / s\right) \hat{W}(s) \hat{W}_t \dots\dots\dots(2.23)$$

إذا كانت السلسلة كبيرة فإن \hat{u}_t يقترب من u_t فدراسة البواقي \hat{u}_t تساعد على التحقيق من دقة وملاءمة الانموذج. إن الارتباطات الذاتية المقدره اي (العينية) $r_k(u)$ للنموذج المختلط، تتوزع بصورة تقريبية توزيعاً طبيعياً بمتوسط صفراً وتباين $\left(\frac{1}{\sqrt{N}}\right)$ اي:

$$r_k(u) \sim N\left(0, \frac{1}{\sqrt{N}}\right)$$

وتستخدم هذه المعلومة في تحديد نسبة المعنوية الاحصائية (أختبار المعنوية) لانحرافات الارتباطات الذاتية المقدره عن قيمتها الصفرية النظرية وبما أن هذه المعلمات غير معروفة في التطبيق، فلجأ الى تقديرها واحتماب بواقبها عن طريق المعادلة رقم (2.23) ثم حساب الارتباط الذاتي التقديري $r_k(\hat{u})$ وإذا كانت معاملات الارتباط الذاتي للأخطاء واقعة في حدود الثقة باحتمال 95% فهذا يعني أن الأخطاء غير نظامية اي عشوائية وبالتالي يعني أن النموذج المشخص ملائم وحدود الثقة هي:

$$-1.96 \frac{1}{\sqrt{N}} \leq r_k(\hat{u}) \leq 1.96 \frac{1}{\sqrt{N}} \quad \dots\dots\dots(2.24)$$

إن :

$$r_k(\hat{u}) = \frac{\sum_{t=1}^{N-k} \hat{u}_t \hat{u}_{t+k}}{\sum_{t=1}^N \hat{u}_t^2} \quad \dots\dots\dots (2.25)$$

N : هو حجم السلسلة الزمنية.

$r_k(\hat{u})$: الارتباطات الذاتية المقدره للأخطاء.

2. أختبار حسن المطابقة Goodness of fit test:

في عام 1970 توصل كلُّ من (Box and Pierce) للنموذج ARIMA(p,d,q) وبافتراض

أن يكون لدينا m من الارتباطات الذاتية المقدره للبواقي $r_k(\hat{u})$.

إذا ان $(k=1, \dots, m)$ وإن:

$$Q = n \sum_{k=1}^m r_k^2(\hat{u}) \sim t_{(m-p-q)}^2 \quad \dots\dots\dots(2.26)$$

أذ ان :

n : يمثل عدد المشاهدات للنموذج المشخص.
 N : عدد المشاهدات الاصلية للسلسلة الزمنية.
 d : يمثل الفروق المأخوذة لتحقيق الاستقرارية للسلسلة حول متوسطها.
 وحيث ان : $n = N - d$.

ثم تتم مقارنة Q المحسوبة بحسب المعادلة رقم (2.26) بـ X^2 الجدولية بدرجة حرية $(m-p-q)$ فاذا كان Q المحسوبة اصغر من الجدولية فالنموذج ملائم وجيد أما اذا كان اكبر فالنموذج غير ملائم وفي هذه الحالة لابد من إعادة المرحلة الاولى اي تشخيص أنموذج آخر لتمثيل السلسلة وتقدير معالمها ثم يتم التحقق من صحتها لقد قام كل من (Ljung,G.M&Box,G.E.P) بتعديل صيغة اختبار Q الاصلية التي اقترحها كل من (Box & Pierce) بالشكل الآتي :

$$Q' = n(n+2) \sum_{k=1}^m (n-k) \bar{r}_k^2(\hat{u}) \dots\dots\dots(2.27)$$

وقد اثبتنا ان لها الافضلية في الاستخدام ، لأن مستوياتها المعنوية قريبة من القيم المتوقعة.

3.1.7.2 المرحلة الثالثة Phase III : التنبؤ Forecasting

بعد الحصول على أنموذج (ARIMA) الملائم لتمثيل بيانات السلسلة الزمنية فإن هذا الأنموذج يستخدم للتنبؤ بالقيم المستقبلية للسلسلة الزمنية على النحو الآتي:
 اذا رمزنا للقيمة الحالية للسلسلة الزمنية بـ Y_t وارادنا التنبؤ بقيمة السلسلة الزمنية في الفترة $t+L$ وفرضنا أن $(\hat{Y}_t(L))$ تمثل هذه القيمة عند زمن الاصل (t) ، فإننا نستطيع الحصول على التنبؤات بأخذ التوقع الشرطي عند زمن الاصل (t) للأنموذج وبعد كتابته عند الفترة $(t+L)$ اي $(E_t(Y_{t+L} / Y_t, Y_{t-1}, \dots))$ إذ باستخدام التوقع الشرطي نحصل على التنبؤات $\hat{Y}_t(L)$ بمتوسط مربعات خطأ التنبؤ (MSE) اقل ما يمكن.

فاذا فرضنا أن الأنموذج الملائم الذي تمّ التوصل اليه عن طريق المراحل السابقة هو $ARIMA(p,d,q)$ فإننا نستطيع التنبؤ بالقيم المستقبلية الى $(L=1, 2, \dots)$ فترة قادمة وباستخدام صيغة معادلة الفروق Difference equation form لهذا الأنموذج وكما يأتي:

$$W(s)Z_t = {}_n(s)u_t \quad \dots\dots(2.28)$$

حيث $Z_t = Y_t - \dots - w_{p+d} s^{p+d}$ و $w_{(p+d)}(s) = w(s)(1-s)^{-1} = (1-w_1s - \dots - w_{p+d}s^{p+d})$

ويسمى بمعامل الانحدار الذاتي غير المستقر من الدرجة $(p+d)$.

ويمكن كتابة المعادلة (2.28) على النحو الآتي:

$$Z_t = w_1 Z_{t-1} + \dots + w_{p+d} Z_{t-p-d} - {}_n u_{t-1} - \dots - {}_n u_{t-q} + u_t \quad \dots\dots(2.29)$$

وبكتابة المعادلة (2.29) عند الفترة $(t+L)$ واخذ التوقع الشرطي عند الفترة (t) نحصل على

معادلة التنبؤ عند زمن الاصل (t) والى $(L=1, 2, \dots)$ فترة تنبؤية. أي أنّ

$$E_t(Z_{t+L}) = \hat{Z}_t(L) \\ = w_1 E_t(Z_{t+L-1}) + \dots + w_{p+d} E_t(Z_{t+L-p-d}) - {}_n E_t(u_{t-1}) - \dots - {}_n E_t(u_{t+L-q}) + E_t(u_{t+L}) \quad \dots\dots(2.30)$$

وبعد احتساب التنبؤات يكون من الضروري احتساب دقة النتائج ويتم هذا عن طريق احتساب

الحدود الاحتمالية العليا والدنيا لكل قيمة تنبؤية ولأي مستوى مناسب من المستوى المعنوي

(99% و 95% و 90%) ويحسب بحسب الصيغة الآتية وهي حدود الثقة للقيمة $Y_{t+\ell}$

المستقبلية و باحتمال $(1 - \alpha)$

$$Y_{t+\ell}(\pm) = \hat{Y}_t(\ell) \pm U_{\alpha/2} \left(1 + \sum_{j=1}^{\ell-1} \Psi_j^2 \right)^{\frac{1}{2}} S_a \quad \dots\dots(2.31)$$

$$S_a^2 = \frac{S(\hat{w}, {}_n)}{n - p - q} \\ = \frac{\sum_{t=1}^n \hat{u}_t^2}{n - p - q}$$

S_a : هو تقدير لـ σ_u^2 من المشاهدات

$U_{\alpha/2}$: قيمة التوزيع الطبيعي القياسي اي القيمة الجدولية المستخرجة من جداول التوزيع

الطبيعي باحتمال $(1 - \alpha)$ والشيء المتوقع ان حدود الثقة للتنبؤ ستكون اوسع كلما صار

كبيراً، لأن التباين لخطأ التنبؤ سيزداد كلما كبر (ℓ) .

3 . الجانب التطبيقي :

1.3 مقدمة :

يحتوي هذا الجانب على المجرىات العملية كافة الخاصة بتطبيق طريقة – Box Jenkins إذ يشتمل على جمع البيانات وتحليلها باستخدام الحاسب الالكتروني وبالاعتماد على الاصدار الاخير للبرنامج الجاهز 22-sps ، وبرنامج 7-Eviews إذ تمّ تحليل السلاسل الزمنية (عينة البحث) وتشخيص الانموذج الملائم لها ثم التنبؤ بقيم السلسلة .

2.3 جمع البيانات :

لقد تمّ اعتماد السجلات الخاصة بوزارة الصحة – المجلس الاعلى للبحوث السرطانية والتي تتعلق باعداد المصابين بامراض سرطان الدم (اللوكيميا) في محافظة البصرة وللمدة من 2004-2017.

3.3 تحليل السلاسل الزمنية :

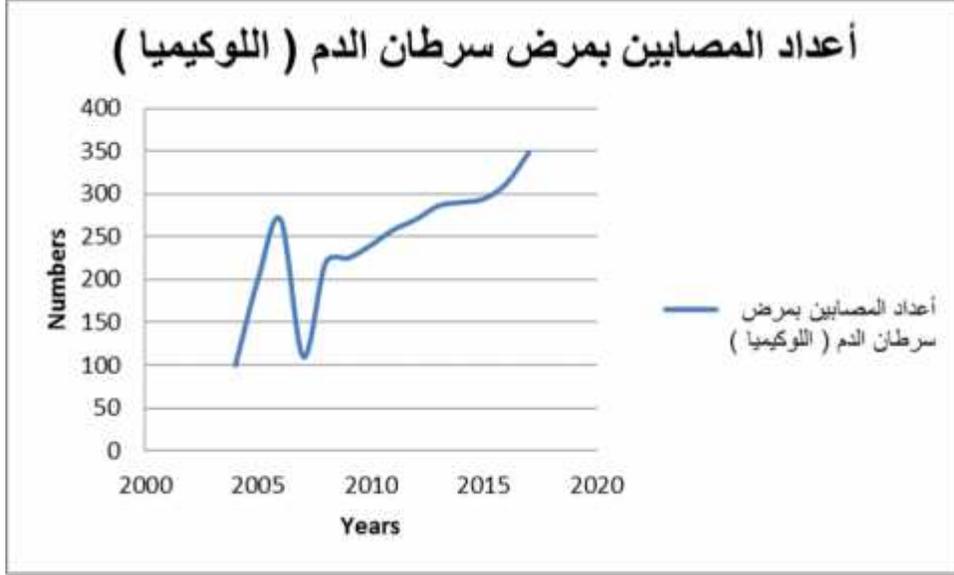
اعتماداً على ماجاء في الفقرة (7.2)، سيتم تحليل السلاسل الزمنية للاصابة بامراض سرطان الدم (اللوكيميا):

1.3.3 تحليل السلسلة الزمنية للأصابة بمرض سرطان الدم (اللوكيميا):

تمّ رسم مشاهدات السلسلة الزمنية لمعرفة الاتجاه العام لها، والشكل (1.3) يمثل رسم السلسلة الزمنية سرطان الدم (اللوكيميا).

(1.3)

السلسلة الزمنية للاصابة بمرض سرطان الدم (اللويميا)



المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

اماً قيم معاملات الارتباط الذاتي ACF والارتباط الذاتي الجزئي PACF فهي كما في الجدولين (1) و (2) على التوالي:

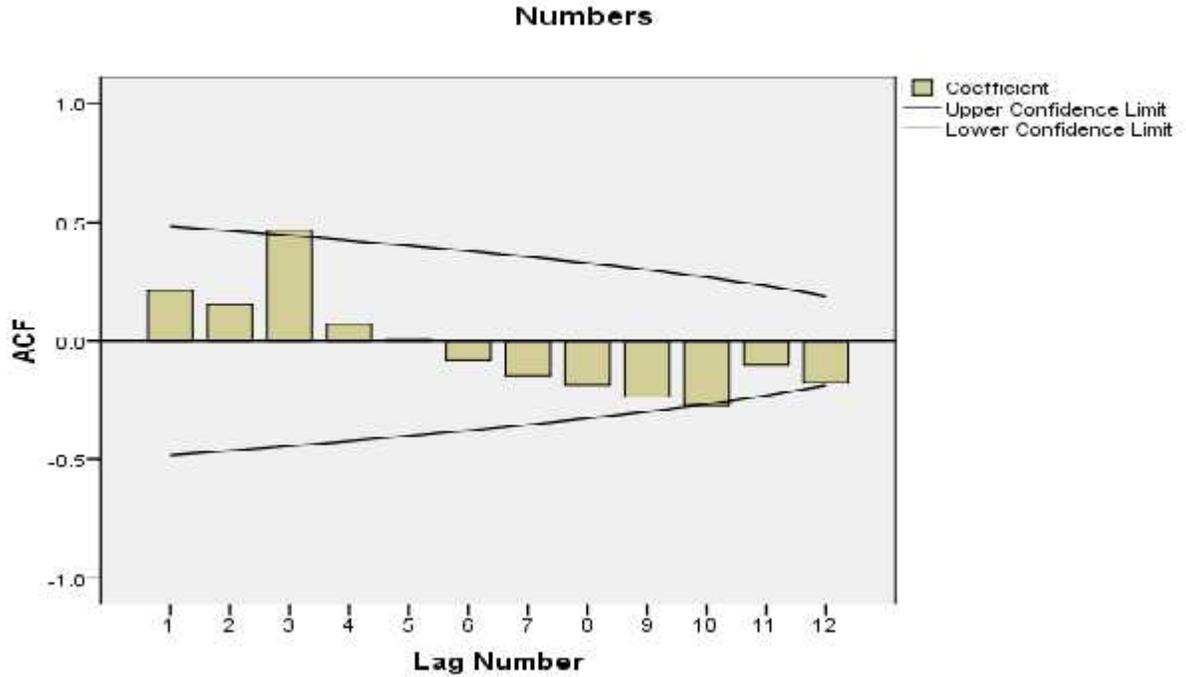
(1)

قيم دالة الارتباط الذاتي لسلسلة سرطان الدم (اللويميا)

Lag.	Auto - corr.	Stand. Err.
1	0.212	0.241
2	0.152	0.231
3	0.465	0.222
4	0.070	0.211
5	0.004	0.200
6	-0.082	0.189
7	-0.145	0.177
8	-0.188	0.164
9	-0.238	0.149
10	-0.275	0.134
11	-0.104	0.116
12	-0.177	0.094

المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

(2.3)
دالة الارتباط الذاتي لسلسلة سرطان الدم (اللويميا)



المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

(2)

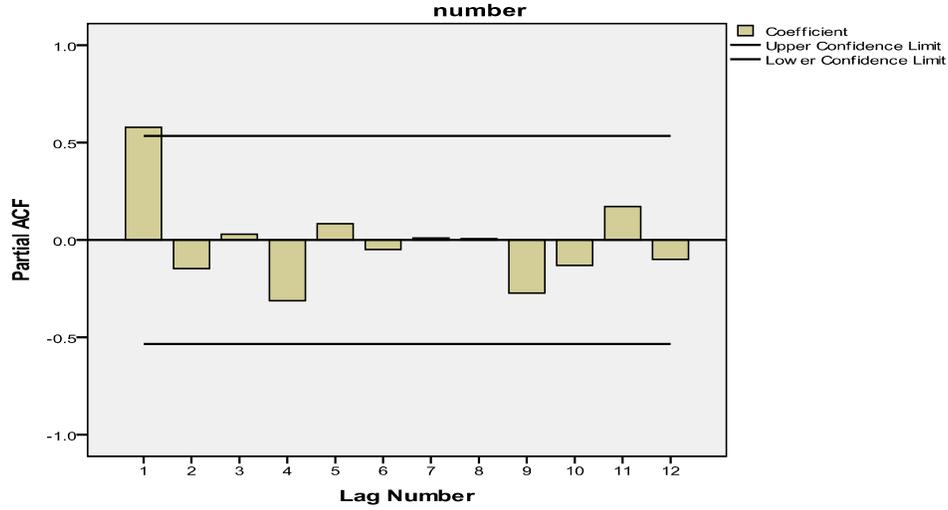
قيم دالة الارتباط الذاتي الجزئي سرطان الدم (اللويميا)

Lag.	Partial. Auto - corr.	Stand. Err.
1	0.212	0.267
2	0.112	0.267
3	0.438	0.267
4	-0.120	0.267
5	-0.091	0.267
6	-0.357	0.267
7	-0.096	0.267
8	-0.135	0.267
9	0.038	0.267
10	-0.137	0.267
11	0.191	0.267
12	-0.117	0.267

المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

(3.3)

سرطان الدم (اللوكيميا)



عن طريق رسم دالة الارتباط الذاتي ACF الشكل (2.3) للسلسلة الاصلية يتبين هناك عدم استقرارية لبيانات السلسلة الاصلية إذ هناك قيم هي خارج حدود المدى المطلوب، اي أنّ السلسلة غير مستقرة، لذلك يتم أخذ الفرق الاول (first difference) لهذه البيانات بعد أخذ الفرق الاول تم استخراج دالة الارتباط الذاتي ACF والارتباط الذاتي الجزئي PACF لبيانات السلسلة بحسب الجدول (3) و (4).

(3)

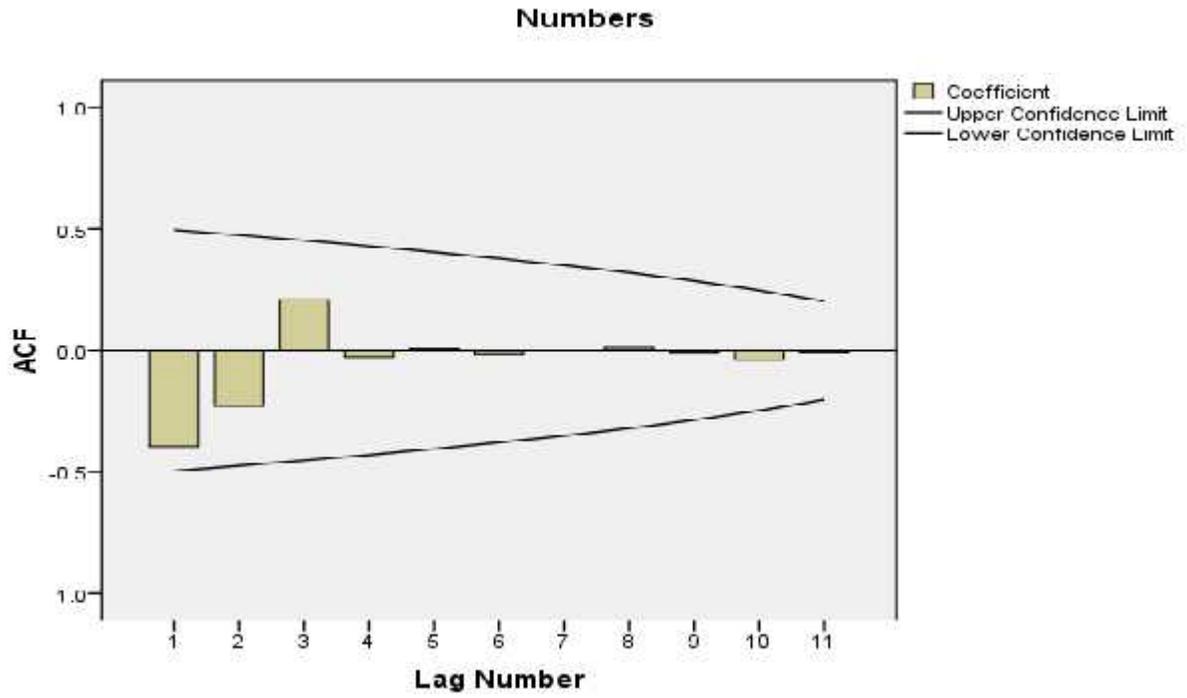
قيم دالة الارتباط الذاتي لسلسلة سرطان الدم (اللوكيميا) بعد اخذ الفرق الاول

Lag.	Auto - corr.	Stand. Err.
1	-0.396	0.248
2	-0.228	0.238
3	0.209	0.226
4	-0.027	0.215
5	0.007	0.203
6	-0.016	0.189
7	-0.002	0.175
8	0.011	0.160
9	-0.012	0.143
10	-0.039	0.124
11	-0.011	0.101

المصدر من اعداد الباحث بالاعتماد على مخرجات البرنامج

(4.3)

دالة الارتباط الذاتي لسلسلة سرطان الدم (اللوكيميا) بعد اخذ الفرق الاول



المصدر من اعداد الباحث بالاعتماد على مخرجات البرنامج

(4)

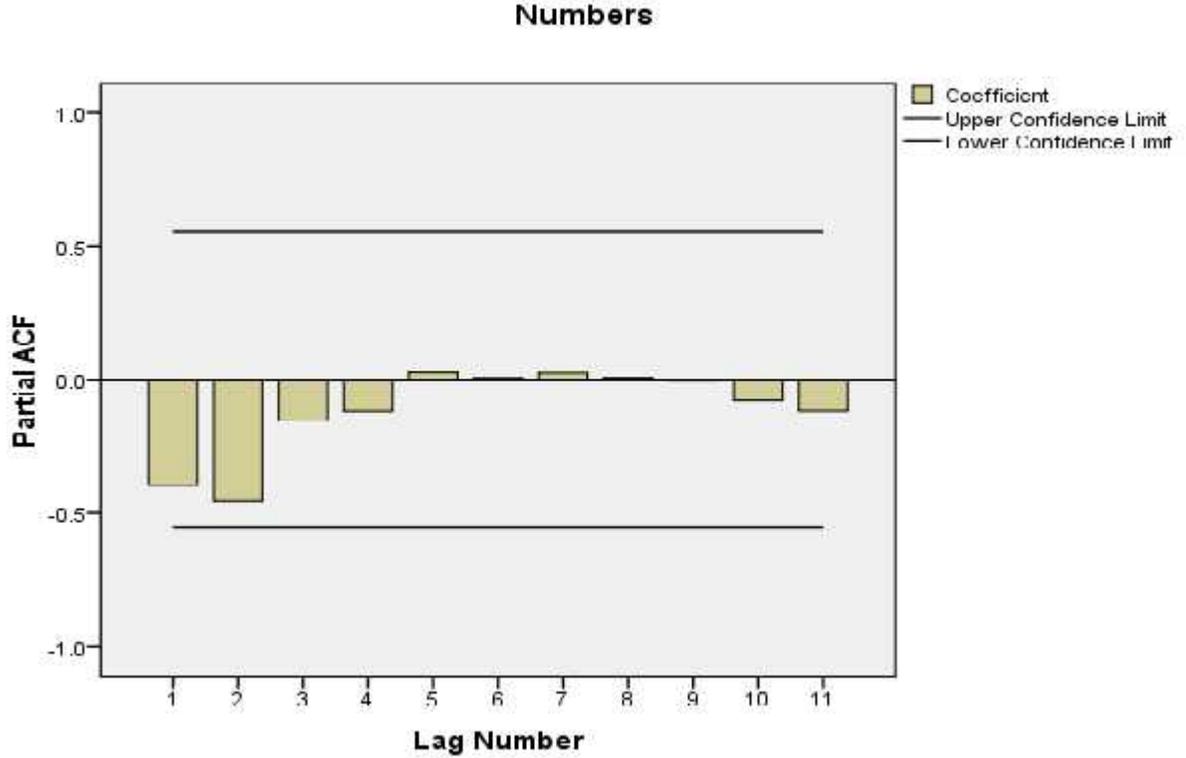
قيم دالة لارتباط الذاتي الجزئي لسلسلة سرطان الدم (اللوكيميا) بعد اخذ الفرق الاول

Lag.	Partial. Auto - corr.	Stand. Err.
1	-0.396	0.277
2	-0.456	0.277
3	-0.154	0.277
4	-0.119	0.277
5	0.026	0.277
6	0.007	0.277
7	0.025	0.277
8	0.007	0.277
9	-0.006	0.277
10	-0.077	0.277
11	-0.199	0.277

المصدر من اعداد الباحث بالاعتماد على مخرجات البرنامج

شكل (5.3)

دالة الارتباط الذاتي الجزئي لسلسلة سرطان الدم (اللوكيميا) بعد اخذ الفرق الاول



المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

عن طريق الشكل (4.3) نلاحظ ان جميع قيم ACF هي في المدى المطلوب وبذلك تم تحقيق الاستقرارية للسلسلة.

2.3.3 تحليل السلسلة الزمنية لسرطان الدم (اللوكيميا) بعد أخذ الفرق الاول :

عن طريق رسم دالة الارتباط الذاتي ACF وكذلك رسم دالة الارتباط الذاتي الجزئي PACF تبين أن السلسلة تسلك سلوك النموذج $ARIMA(0,1,2)$ ، وأن $P=0$ حيث P يمثل درجة نموذج الأنحدار الذاتي وأن $d=1$ يمثل درجة الفرق الاول للسلسلة الزمنية وأن $q=2$ حيث إن q يمثل درجة نموذج الاوساط المتحركة . وبعد تقدير النموذج كانت قيمة المعلمة المقدرة $\hat{\rho}$ تساوي (0.082). كما موضح في الجدول (5) :

جدول (5)

(ARIMA (0,1,2))

يمثل تحليل

(الوكيميا)

ARIMA (0,1,2) Model Parameters							
			Estimate	SE	t	Sig.	
number-Model_1	number	Natural Log	Constant	.082	.055	1.488	.168
			Difference	1			
		MA	Lag 1	.001	9.144	5.700E-5	1.000
			Lag 2	.985	10.308	.096	.926

المصدر من اعداد الباحث بالاعتماد على مخرجات البرنامج

وقد تمّ اختبار النموذج للتأكد من كفاءته عن طريق أحصاءة اختبار (Box-Ljung) المعروفة صيغته في المعادلة (2.27). وكما مبين في الجدول (6).

(6)

يمثل اختبار الأنموذج المشخص

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig.
1	-0.396	0.248	2.544	1	0.111
2	-0.228	0.238	3.468	2	0.177
3	0.209	0.226	4.322	3	0.229
4	-0.027	0.215	4.338	4	0.362
5	0.007	0.203	4.340	5	0.502
6	-0.016	0.189	4.347	6	0.630
7	-0.002	0.175	4.347	7	0.739
8	0.011	0.160	4.351	8	0.824
9	-0.012	0.143	4.359	9	0.886
10	-0.039	0.124	4.460	10	0.924
11	-0.011	0.101	4.471	11	0.954

المصدر من اعداد الباحث بالاعتماد على مخرجات البرنامج

وبمقارنة قيمتي الاختبار مع قيمة t^2 الجدولية نلاحظ عدم معنويتها مما يؤكد ملاءمة النموذج للبيانات علماً أنّ قيمة t^2 الجدولية عند مستوى معنوية (0.05) ودرجة حرية (13) بلغت (22.36).

ولمزيد من الدقة في توفيق افضل أنموذج من نماذج Box – Jenkins تمّ أستخراج المعيار BIC الموضح في الفقرة (2.2.1.7.2)، وتمّ تطبيق النماذج الممكنة لـ ARIMA(p,d,q) كافة وحساب لكل منها MSE و BIC كما مبين في الجدول (7).

جدول (7)

يمثل مقارنة لمجموعة من الأنموذجات لقيم MSE، BIC

النماذج	AR	MA	MSE	BIC
0,0,1	0	1	9.495	4.879
0,0,2	0	2	9.669	5.103
0,0,3	0	3	10.096	5.378
0,0,4	0	4	10.328	5.612
0,1,1	0	1	10.253	5.050
0,1,2	0	2	8.684	4.315
0,1,3	0	3	9.293	5.248
0,1,4	0	4	9.866	5.565
1,0,0	1	0	10.144	5.011
1,0,1	1	1	9.652	5.100
1,0,2	1	2	10.080	5.375
1,0,3	1	3	10.983	5.735
1,0,4	1	4	11.156	5.955
2,0,0	2	0	10.317	5.233
2,0,1	2	1	10.068	5.373
2,0,2	2	2	10.571	5.659
2,0,3	2	3	10.953	5.918
2,0,4	2	4	12.037	6.296

المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

عن طريق نتائج مقارنة MSE و BIC لجميع النماذج المفترضة تبين أنّ افضل أنموذج ملائم لبيانات السلسلة هو ARIMA(0,1,2) مما يؤكد ملاءمة البيانات للنموذج المختار، بعد التأكد من ملاءمة البيانات للنموذج ولغرض المقارنة وضعت النتائج للقيم الحقيقية مع التقديرية والأخطاء للمدة من (2017-2004) وكما مبين في الجدول (8) ، ثم بعد ذلك أخذت القيم التنبؤية للمدة من (2024-2018) وكما مبين في الجدول (9) .

جدول (8)

يمثل القيم الاصلية والقيم التنبؤية والاختلاف لسلسلة سرطان الدم (اللوكيميا)

Period	Data	Forecast	Residual
1	100.00	100.00	00.00
2	200.00	190.00	10.00
3	270.00	265.00	5.00
4	110.00	124.00	-14.00
5	220.00	209.00	11.00
6	225.00	225.00	00.00
7	240.00	252.00	-12.00
8	258.00	260.00	-2.00
9	270.00	266.00	4.00
10	286.00	268.00	18.00
11	290.00	276.00	14.00
12	294.00	298.00	-4.00
13	312.00	319.00	-7.00
14	348.00	341.00	7.00

المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

جدول (9)

يمثل القيم التنبؤية للاعوام 2024 -2018

Period	Forecast	Lower 95.0% Limit	Upper 95.0% limit
1	400.00	111.10	310.10
2	400.00	111.10	310.10
3	420.00	107.10	350.10
4	425.00	103.90	370.10
5	411.00	109.12	320.10
6	422.00	107.90	360.10
7	422.00	107.90	360.10

المصدر من اعداد الباحث بالاعتماد على محرجات البرنامج

4. _____:

في ظل تحليل نتائج الجانب التطبيقي تم التوصل إلى الاستنتاجات الآتية :

1. أتضح عن طريق هذه الدراسة والتحليل لجميع سلسلة مرض سرطان الدم (اللويميا) بانها غير مستقرة بالنسبة للبيانات الاصلية وأصبحت مستقرة بعد أخذ الفرق الاول لها.
2. استنتج الباحث أن السلسلة الزمنية لمرض سرطان الدم (اللويميا) لها أنموذج من نوع $ARIMA(0,1,2)$ وهو افضل أنموذج تم اختياره في ضوء البيانات الحقيقية بغض النظر عن طول السلسلة الزمنية .
3. كان توفيق النموذج المعتمد $ARIMA(0,1,2)$ توفيقاً جيداً بعد تطبيق أحصاء أختبار (Box-Ljung) .
4. استنتج الباحث عن طريق مقارنة القيم التنبؤية باستخدام النموذج المعتمد $ARIMA(0,1,2)$ مع قيم المشاهدة ولجميع السلسلة الحقيقية بان النموذج المقدر هو افضل أنموذج لاستخدامه في التنبؤ المستقبلي لقيم الظاهرة المدروسة وذلك بسبب الفروق الضئيلة جداً بين القيم التنبؤية وقيم المشاهدة.

5. التوصيات:

1. يوصي الباحث اجراء دراسات مشتركة بين وزارتي الصحة والتعليم العالي للوقوف على اهمية مثل هذه الدراسات والأخذ بها من قبل وزارة الصحة أو العاملين في حقل أمراض السرطان في العراق.
2. حفظ البيانات بالوسائل الآلية مثل الحاسبة الالكترونية حفاظاً عليها من التلف او التخريب مع مرور الزمن .
3. التوعية والفحص المبكر ذلك محاولة للسيطرة على المرض قدر المستطاع .

المصادر References

المصادر العربية

1. الحسن، اياد جواد، (2002). "استخدام المحاكاة للتحري عن حصانة معيار معلومات اكيائي لتحديد درجة عملية الانحدار الذاتي"، رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
2. شمو، خالد داود، (1987). "أستخدام اسلوب السلاسل الزمنية للتنبؤ بكمية الودائع الشهرية لدى مصرف الرافدين"، رسالة ماجستير، كلية الادارة والاقتصاد، جامعة بغداد.
3. عبد الحافظ ، علي سيف الدين (2006) "تحليل السلاسل الزمنية لبعض الأمراض السرطانية في العراق " رسالة ماجستير ، كلية الإدارة والاقتصاد ، الجامعة المستنصرية.

المصادر الانجليزية

4. Box G, E.P & Jenkins, G. M., "Time series analysis forecasting and control sanfrancisco Holden-day", 1976.
5. Box, G. E. and Tiaa, "Distribution of Residual Autocorrelation in Multiple Autoregressive", JASA, Vol. (74), pp. (928-934), 1979.
6. Ljung G. M. "Diagnostic testing of univariate time series model identification", Biometrika, vol. (72), No. (2), pp. (299-315), 1985.
7. Tsay R. S. & Tiao G.C. (1985) "Use of canonical analysis in time series model identification", Biometrika, vol. (72), No. (2), pp. (299-315).