



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية علوم الحاسوب وتكنولوجيا المعلومات
قسم الإحصاء والمعلوماتية

استخدام تحليل السلاسل الزمنية لتنبؤ بأعداد المصابات بمرض سرطان الثدي في محافظة كربلاء

بحث مقدم

الى مجلس كلية علوم الحاسوب وتكنولوجيا المعلومات \ قسم الإحصاء والمعلوماتية
وهو جزء من متطلبات نيل درجة البكالوريوس في علوم الإحصاء والمعلوماتية

من قبل الطالبتان:-

زهراء هادي عبد راهي

ميعاد علي كطران

وبإشراف:-

أ.م.د. مهند فائز السعدون

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

يُخَوِّفُ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ
أُوتُوا الْعِلْمَ سَرَّ جَاهٍ

...

صدق الله العظيم

سورة المجادلة

آية (١١)

الإمام

بأسم الله الخالق الواحد الأحد الذي اضاء الكون بنوره نحمده لمنه وفضله علينا في
أتمام هذا الجهد..

الى.....

من اكرمهم الله تعالى وخلق كل شئ حياً لهم أهل بيت النبوة ومعدن الرساله ومهبط
الوحي عباده الصالحين المخلصين نبينا محمد ،وعلي وفاطمة والحسن والحسين
والتسعة المعصومين من أبنائهما الطاهرين (عليهم السلام)

الى.....

من جرع الكأس فارغاً ليسقيني قطرة حب ...

الى من تأكلت أنامله ليقدّم لنا لحظات من السعادة اللاحدود لها ..

الى من ازاح الأشواك عن دربي ليمهد لي الى ذلك النور طريقاً للعلم ...

الى قلبي الكبير (والدي العزيز)

الى ... من حملتني وهنا على وهن وارضعتني حبا وحنانا مستمر ... اية انزلت بها
جنة تحت اقدامها

الى من وضع الله الجنه تحت اقدامها الى رمز الحب وبلسم الشفاء الى القلب

الناصع بالبياض(والدتي الحبيبة)

الى.....

القلوب الطاهرة الرقيقة والنفوس البريئة الى رياحين حياتي (أخوتي)

الى.....

الشموع التي ذابت في كبرياء لتتير كل خطوة في دربنا لتذلل كل عائق امامنا

فكانوا رسلاً للعلم والأخلاق الذين صاغوا من علمهم حروفاً من ذهب ومن

فكرهم منارة تنير لنا سيرة العلم والنجاح الى اساتذتي الكرام .

الشكر والنعمة

لا بد لنا ونحن نخطوا خطواتنا الاخيرة في الحياة الجامعية من وقفه نعود الى اعوام قضيناها في رحاب الجامعة مع اساتذتنا الكرام الذين قدموا لنا الكثير باذلين بذلك جهود كبيرة في بناء جيل العز لتبعث الامه من جديد ...

وقبل ان نمضي نقدم اسمى آيات الشكر والامتنان والتقدير والمحبة الى اللذين حملوا اقدس رسالة في الحياة...

الى اللذين مهدوا لنا طريق العلم والمعرفة ...

الى جميع اساتذتنا الافاضل في (قسم الإحصاء)...

كما نتقدم بالشكر والامتنان الى استاذنا الفاضل الدكتور " مهند فائز السعدون " الذي تفضل بالإشراف على هذا البحث ولما قدمه من توجيهات قيمة وسديدة كان لها الاثر الاكبر في اتمام هذا العمل .

فهرست المحتويات

الصفحة	الموضوع
١	الفصل الأول (المقدمة، هدف البحث، منهجية البحث)
٢	١-١ المقدمة
٣	١-٢ هدف البحث
٣	١-٣ منهجية البحث
٤	الفصل الثاني (الجانب النظري)
٥	المقدمة
٥	٢-١ المفاهيم الأساسية
٩	٢-٢ نماذج بوكس وجينكنز
١٣	٢-٣ بناء نموذج السلاسل الزمنية
١٧	الفصل الثالث (الجانب التطبيقي)
٢٠	٣-١ جمع البيانات
٢٢	٣-٢ مرحلة تهيئة البيانات
٢٩	الفصل الرابع (الأستنتاجات والتوصيات)
٣٠	٤-١ الأستنتاجات
٣١	٤-٢ التوصيات
٣٢	المصادر

فهرست الاشكال

الصفحة	عنوان الشكل	رقم الشكل
٢٢	يمثل اعداد النساء المصابات بأورام الثدي للفترة (2012-2016)	شكل (١)
٢٣	معاملات الارتباط الذاتي للسلسلة الزمنية	شكل (٢)
٢٣	معاملات الارتباط الذاتي الجزئي للسلسلة الزمنية	شكل (٣)
٢٤	منحنى السلسلة الزمنية بعد اخذ الفرق الاول	شكل (٤)
٢٤	معاملات الارتباط الذاتي بعد اخذ الفرق الاول للسلسلة الزمنية	شكل (٥)
٢٥	معاملات الارتباط الذاتي الجزئي بعد اخذ الفرق الاول للسلسلة الزمنية	شكل (٦)
٢٧	منحنى القيم التنبؤية لسلسلة اعداد المصابات بأورام الثدي للفترة (كانون الثاني-تشرين الاول\2017)	شكل (٧)

فهرست الجداول

الصفحة	عنوان الجدول	رقم الجدول
٢١	اعداد النساء المصابات بسرطان الثدي في مدينة كربلاء	جدول رقم (١)
٢٦	يوضح رتب النموذج مع تقدير معالمه باستخدام دالة الامكان الاعظم	جدول رقم (٢)
٢٨	اعداد المصابات بأورام الثدي المتنبأ بها في محافظة كربلاء للفترة (كانون الثاني-تشرين الاول\2017)	جدول رقم (٣)

الفصل الأول

المقدمة وهدف البحث

(١)

١-١ المقدمة :

نتيجة للظروف التي مر بها البلد خلال السنوات السابقة من تلوث عام في البيئة وهذا ما يعود الى العديد من الاسباب ومنها الحروب التي اصابت البلد و دمرت البنى التحتية له ولوثت مائه وهوائه وكذلك المواد الكيميائية والمواد الأشعاعية وملوثات طبيعية ناتجة عن ظواهر طبيعية اسهمت في التأثير على الكثير من جوانب الحياة الخاصة بالانسان مما تعيقه على ممارسة حياة بشكل صحيح ، وقد أجريت هذه الدراسة لتوعية الفرد بسبب الاهمال الصحي وللد من هذه الضواهر ولما اصاب البلد من تلوث وامراض وآفات ، ومن الأساسيات التي تبني عليها الصحة هي الحد من الاصابة بالأمراض التي تهدد حياة الانسان ومنها الأمراض الخبيثة التي تؤدي في النهاية الى نسب عالية من الوفيات مقارنة بالأمراض الأخرى.

وجاء هذا البحث ليتناول الجانب الصحي ولأهميته في التقليل من الأصابات والتي ازدادت في الأونه الأخيرة بداخل محافظات العراق وخصوصا في مدينة كربلاء المقدسة

وقد اعتمدت هذه الدراسة على البيانات الشهرية لأعداد النساء المصابات بمرض سرطان الثدي للفترة من (٢٠١٢ _ ٢٠١٦) كسلسلة زمنية حيث تعتبر من بين اهم الأساليب الأحصائية الحديثه والتي تمكنت من خلالها معرفة طبيعة التغيرات التي تطرأ على القيم التي تظهر مع الزمن لغرض تحليلها والوصول لأفضل نموذج بالتنبؤ بأعداد النساء المصابات بهذا المرض لفترات لاحقة وذلك لأتخاذ التدابير اللازمة للحد من هذه الظاهرة مستقبلاً.

٢-١ هدف البحث:

يهدف البحث الى تحليل السلاسل الزمنية باستخدام طريقة (Box and Jenkins) في التحليل (التشخيص، التقدير، اختبار ملائمة النموذج، التنبؤ) لأيجاد افضل نموذج احصائي لتنبؤ بأعداد النساء المصابات بسرطان الثدي في محافظة كربلاء، وذلك بالإعتماد على البيانات التي تمثل اعداد المصابات بمرض السرطان شهرياً للفترة (٢٠١٢ - ٢٠١٦) .

٣-١ منهجية البحث:

اعتمد البحث في منهجته على الجانب النظري الذي تناول خارطة

(Box and Jenkins) في تحليل السلاسل الزمنية وهي:(التشخيص ،التقدير ،اختبار ملائمة النموذج ،التنبؤ المستقبلي) وتعتبر منهجية بوكس وجينكيز الأكثر شيوعاً في الأوساط العلمية النظرية والتطبيقية خاصة في العالم المتقدم والمشكاة الأساسية التي أنبثقت منها سائر المنهجيات الحديثة والمرجعية الرئيسية على جودة وملائمة الكثير من هذه المنهجيات فقد اثبتت هذه المنهجية كفاءة عالية في النمذجة والتنبؤ بالسلاسل الزمنية التي تنشأ في مجالات المعرفة المختلفة مثل الأقتصاد وإدارة الاعمال والبيئة والكيمياء والهندسة وغيرها . ودعم مجريات الجانب النظري بالجانب التطبيقي الذي اعتمد على بيانات واقعية من اعداد النساء المصابات بأورام الثدي في محافظة كربلاء للوصول الى افضل نموذج رياضي للتنبؤ بأعداد النساء المصابات بهذا المرض لفترات اللاحقة

الفصل الثاني

الجانب النظري

مقدمة:

يتناول هذا الفصل بعض المفاهيم العامة، والتعرف على مراحل بناء السلاسل الزمنية، ويعتمد بناء السلاسل الزمنية على الخارطة التي رسمها الباحثان Box and Jenkins التي تبدأ بالمرحلة الأولى وهي تشخيص النموذج الملائم للبيانات، تليها مرحلة تقدير معاملات النموذج المشخص، ثم تأتي مرحلة فحص ملائمة النموذج المشخص، فإذا كان النموذج ملائماً تأتي المرحلة الأخيرة وهي مرحلة التنبؤ المستقبلي.

١-٢ المفاهيم الأساسية

السلسلة الزمنية : Time Series

هي مجموعة من المشاهدات مرتبة وفق حدوثها في الزمن لقيم ظاهرة ما تكون مأخوذة في اوقات زمنية محددة وكذلك هي القياسات التي تأخذ على احدى الظواهر (الاقتصادية-الاجتماعية-الطبية-الطبيعية....) على فترات زمنية متتابعة عادة ما تكون متساوية بالطول

فإذا كانت متساوية فيعبر عن $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ عند الفترات الزمنية t_1, t_2, \dots, t_n

حيث أن n تمثل عدد القيم المشاهدة وتكتب هذه المشاهدات كالاتي :

X_1, X_2, \dots, X_n حيث ان X_1 تعني قيمة المشاهدة التي وقعت في الزمن t_1 ويمكن تمثيل السلسلة الاحصائية بالشكل التالي (الكاطع، 2007):

$$X_t = f(t) + a_t$$

حيث ان:

$f(t)$: يمثل الجزء المنتظم الذي يعبر عنه بدالة رياضية .

a_t : يمثل الجزء العشوائي.

ويمكن ان تكون السلسلة الزمنية من النوع المحدد (Deterministic) ومثال على ذلك :

$$X_t = \cos 2\pi f(t) \quad t=0, \pm 1, \pm 2, \dots$$

اي لا تحتوي على الجزء العشوائي. وفي هذا النوع من السلاسل الزمنية يمكن تحديد السلوك المستقبلي لها

او قد تكون السلسلة الزمنية دورية (periodic) مثال ذلك البيانات التي تظهر بشكل جيبي ويمكن تمثيلها بالصيغة :

$$\forall t \quad t = 0, \pm 1, \pm 2, \dots \quad X_t = X_{t+s}$$

حيث ان S: هي دورة السلسلة

وهناك نوعين من السلاسل الزمنية هما :

السلسلة الزمنية المستقرة والسلسلة الزمنية غير المستقرة وهناك نوعان من الاستقرارية

هما الاستقرارية في المتوسط (Stationary in Mean) والاستقرارية في التباين (Stationary in Variance) تكون السلسلة الزمنية غير مستقرة في التباين

في حالة لا تظهر تذبذبات متباينة في شكل السلسلة ويمكن تثبيت التباين بالحصول على اللوغاريتم الطبيعي او الجذر التربيعي او المقلوبات لبيانات السلسلة مما يجعل لها عدة اوساط تتذبذب حولها البيانات حتى عندما تكون السلسلة متجانسة فهذه النماذج توصف بأن لها سلوك غير مستقر ومتجانس Non stationary (Homogenous) اما اذا كانت الأستقرارية في

المتوسط هي حالة السلسلة عندما لا تظهر اتجاهها عاما ويمكن تحويلها الى مستقرة باستخدام الفروقات.

الأرتباط الذاتي (AC): Autocorrelation

هو عبارة عن مؤشر يوضح درجة العلاقة بين قيم نفس المتغير عند فترات ازاحة (K) مختلفة ، وتتراوح قيمته بين (-1, 1) اي $-1 \leq \rho_k \leq 1$ ويقدر حسب الصيغة التالية (Box&Pierce,1970,1509-1519)

$$\widehat{\rho}_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (2-1)$$

حيث أن

ρ_k : تقيس الارتباط بين المشاهدات المتعاقبة للسلسلة

X_t : يمثل قيم مشاهدات السلسلة

$$\bar{X}: \text{هو الوسط الحسابي حيث يمثل } \frac{x_1}{n} \sum_{t=1}^n$$

حيث تمثل دالة الارتباط الذاتي وترمز لها (ACF) حيث تستخدم في تحليل السلسلة الزمنية لأنها تعطي معلومات حول سلوك السلسلة وعن مكوناتها الأساسية كما تساعد على تحديد استقرار السلسلة. وبصفة عامة فإن دالة الارتباط الذاتي (ACF) للسلسلة المستقرة تتنازل (تتناقص) بسرعة وتكون قريبة من الصفر كلما زادت درجات الابطاء. وبما ان دالة الارتباط الذاتي للعينة هي فقط تقديرات للارتباطات الذاتية فإن قيمها من المحتمل ان تكون صغيرة وليس صفر أي أن:

$$r_k \neq 0 \quad \forall k > q$$

اما اذا كانت السلسلة الزمنية غير مستقرة بسبب وجود اتجاه صاعد أو نازل في المعدل فإن دالة (ACF) للعينة لا تنقطع ولا تتحدر ببطء تجاه الصفر وذلك لكون المشاهدات تميل لأن تكون على نفس اتجاه الوسط الحسابي للسلسلة الزمنية لفترات زمنية عديدة وكنتيجه لذلك نحصل على ارتباطات ذاتية كبيرة عند فترات ازاحة طويلة .

كما تستخدم دالة الارتباط الذاتي للبواقي (RACF) لفحص ملائمة النموذج عن طريق اختبار عشوائية اخطاء التنبؤ .

وحيث ان توزيع الاحصائي لمعاملات الارتباط الذاتي هو توزيع طبيعي بوسط حسابي صفر وتباين $1/N$ حيث N حجم العينة، (المتولي، 1989)

$$p_k \sim N(0, \frac{1}{N})$$

الارتباط الذاتي الجزئي: Partial Autocorrelation(PAC)

هو مؤشر يقيس العلاقة بين المشاهدين X_t و X_{t-k} لنفس السلسلة مع افتراض ثبوت قيم السلسلة الزمنية ويعرف على انه الحد الأخير من نموذج الانحدار الذاتي من الدرجة $AR(P)$ ، ويمكن ايجاد معامل الارتباط الذاتي الجزئي وذلك عن طريق دالة الارتباط الذاتي وحسب الصيغة التالية (الخصيري، 1996)

$$\phi_{k+1,k+1} = \frac{\rho_{k+1} - \sum_{j=1}^k \phi_{kj} \rho_{k+1-j}}{1 - \sum_{j=1}^k \phi_{kj} \rho_j} \quad (2-2)$$

أن دالة الارتباط الذاتي الجزئي (PACF) لا تقل أهمية عن الارتباط الذاتي (ACF) فهي ايضا اداة مهمة في تحليل السلاسل الزمنية وتستخدم ايضا في تشخيص النموذج وتحديد درجته وفي فحص ملائمة النموذج من خلال اختبار عشوائية اخطاء التنبؤ (البواقي) .

ان دالة الارتباط الذاتي الجزئي (PACF) للسلسلة الزمنية المستقرة تميل للانحدار بسرعة نحو الصفر مع ازدياد فترات الأزاحة أو تنقطع بعد عدد معين من فترات الأزاحة (Anderson, 1942)

٢-٢ نماذج بوكس و جينكنز (B-I) للسلاسل الزمنية :

نموذج الأنحدار الذاتي (AR) Autoregressive Model

حيث ان الصيغة العامة لنموذج الأنحدار الذاتي من الدرجة (P) تكون بالشكل التالي (الجبوري، 2010):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + a_t \quad (2-3)$$

حيث أن:

X_t : تمثل قيم المشاهدات للسلسلة

Φ_i : معالم النموذج $i=1,2,3,\dots,P$

Φ_0 : الحد الثابت

P: درجة النموذج

a_t : تمثل الأخطاء العشوائية التي تتوزع طبيعياً بمتوسط صفر وتباين σ^2

وكذلك يستخدم نموذج الأنحدار الذاتي لتمثيل السلاسل الزمنية المستقرة وغير المستقرة ولكي تتحقق الاستقرارية يشترط أن تكون جذور المعادلة $\Phi_p(B) = 0$

خارج دائرة الوحدة أي أنه لكي يكون النموذج مستقراً

$$-1 < \Phi_t < 1$$

حيث أن B: عامل الأرتداد الخلفي (Back Shift Operator) ويعرف بالشكل التالي:

$$B^k X_t = X_{t-k} \quad \forall k = 1, 2, \dots$$

أي أن دالة الارتباط الذاتي للنموذج AR(p) تتناقص أسياً ،

في حين أن دالة الارتباط الذاتي الجزئي تنقطع بعد الفترة الفاصلة

وهناك حالتان خاصتان لصيغة العامة للانحدار الذاتي $AR(p)$ وهما نموذجا
 الانحدار الذاتي من الدرجة الأولى $AR(1)$ ومن الدرجة الثانية $AR(2)$
 اللذان يعتبران من النماذج الشائعة الاستخدام لتمثيل معظم السلاسل الزمنية
 حيث أن في حالة كون $p=1$ تصبح المعادلة رقم (1-3) كالآتي :

$$X_t = \Phi_0 + \Phi_1 X_t + a_t \quad (2-4)$$

حيث تعتبر المعادلة أعلاه نموذج انحدار ذاتي من الدرجة الأولى $AR(1)$.
 وعند تحقيق شروط الاستقرار في النموذج تتطلب ان تكون جذور المعادلة
 خارج حدود الوحدة أي ان :

$$1 < \Phi_1 < 1$$

وأن دالة الارتباط الذاتي (ACF) للنموذج تكون :

$$\rho_k = \Phi_1 \rho_{k-1}$$

فإذا كان $\rho_0 = 0$ يمكن الحصول على

$$\rho_k = \Phi_1^k \quad k=0,1,2,\dots$$

لذلك فإن دالة الارتباط الذاتي (ACF) لنموذج الانحدار الذاتي من الدرجة الأولى
 $AR(1)$ تنحدر بصورة أسية عندما تكون Φ_1 موجبة وتنحدر بصورة أسية متناوبة
 في الإشارة عندما تكون Φ_1 سالبة .

اما دالة الارتباط الذاتي الجزئي (PACF) تنقطع بعد الأزاحة الأولى ($K>1$)
 وفي حالة كون $p=2$ في المعادلة (1-3) فنحصل على نموذج انحدار ذاتي من
 الدرجة الثانية $AR(2)$ فنكون الصيغة بالشكل الآتي (المتولي، 1989):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + a_t \quad (2-5)$$

لكي يكون نموذج الأنحدار الذاتي AR(2) مستقراً يجب أن تقع جذور المعادلة خارج حدود الوحدة اي يجب أن تحقق المعلمتين (Φ_1, Φ_2) الشروط (B-J,1976):

$$(\Phi_1 + \Phi_2 < 1), (\Phi_1 - \Phi_2 < 1)$$

وأن دالة الارتباط الذاتي (ACF) تكون كالاتي:

$$\rho_k = \Phi_1 \rho_k + \Phi_2 \rho_{k-2} \quad k > 0$$

وأن دالة الارتباط الذاتي الجزئي (PACF) لنموذج AR(2) تنقطع بعد الأزاحة الثانية اي $K > 2$

وتتمثل بالصيغة التالية :

$$\rho_{11} = \frac{\Phi_1}{1 - \Phi_1}, \rho_{22} = \Phi_2, \rho_{kk} = 0$$

نموذج الأوساط المتحركة (MA) Moving Average Model

يمكن تمثيل نموذج الأوساط المتحركة من الدرجة (q) بالصيغة التالية (الجبوري، 2010):

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2-6)$$

حيث أن θ_i : معالم نموذج الأوساط المتحركة

$$-1 < \theta_i < 1 \quad i=1,2,3,\dots,q$$

q : درجة النموذج .

حيث أن دالة الارتباط الذاتي (ACF) تنقطع أو تقترب من الصفر بعد الأزاحة

(q) في حين تتضاءل دالة الارتباط الذاتي الجزئي (PACF) بشكل أسي

النموذج المختلط (الانحدار الذاتي – الأوساط المتحركة)

Mixed Autoregressive Moving Average Model (ARMA)

يمكن كتابة النموذج بالصيغة العامة من الدرجة (p,q) بالشكل الآتي (B-J,1976):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2-7)$$

بستخدام عامل الارتداد الخلفي (B)

$$\Phi_p(B)Z_t = \Phi_0 + \theta_q(B)a_t$$

ولكي تتوفر الاستقرار في هذا النموذج يجب ان تكون جذور المعادلة $(\Phi_p(B) = 0)$ خارج حدود دائرة الوحدة. وكذلك بالنسبة لجذور المعادلة

$$\theta_q(B) = 0 \text{ (الجبري، 2010).}$$

النموذج المختلط المتكامل

Autoregressive Integrated Moving Average Model

تعد نماذج (ARIMA) أكثر نماذج السلاسل الزمنية استخداماً إذ يمكن من خلالها

اشتقاق جميع النماذج منها سواء الانحدار الذاتي AR(P) أو الأوساط المتحركة

او المختلطة .

والصيغة العامة لهذه النماذج هي كالتالي:

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \dots + dX_{t-p-d} + a_t - \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (2-8)$$

وتتكون هذه النماذج من ثلاثة اجزاء يمثل الجزء الاول منها نموذج انحدار

ذاتي $AR(p)$ وهو الجزء الذي يستخدم عادة في التنبؤات للسلسلة الزمنية
اما الجزء الآخر فيمثل الاوساط $MA(q)$ ، ويمثل الجزء الثالث $I(d)$
الفروق التي تتطلبها السلسلة من اجل ان تكون مستقرة
ولذلك يمكن اعتبار نماذج $ARIMA$ هي نماذج $ARMA$ مستقرة مع اختلاف
الرتبة.

٢-٣ بناء نموذج السلاسل الزمنية

يتم بناء نموذج السلاسل الزمنية من خلال اربعة مراحل هي :
أولاً. تشخيص النموذج الملائم للبيانات
ثانياً. تقدير معلمات النموذج المشخص
ثالثاً. اختبار ملائمة النموذج المشخص
رابعاً. التنبؤ المستقبلي.

أولاً. تشخيص النموذج Identification

أن مرحلة تشخيص نماذج السلاسل الزمنية تعد اهم مراحل بناء نماذج السلاسل
الزمنية حيث يتم من خلالها تحديد رتبة النموذج. ويجب ان تسبق مرحلة التشخيص
مرحلة تهيئة البيانات فإذا كانت البيانات مستقرة حيث ان بعد تحقيق الاستقرار
في السلسلة الزمنية تبدأ عملية تحديد النموذج المناسب لتمثيل السلسلة ودرجته
بأستخدام دالتي الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي $(PACF)$ فتكون
البيانات مهيأة للتشخيص.

أما اذا كانت السلسلة غير مستقرة في الوسط والتباين، فإنه يتم معالجة عدم
الاستقرارية في الوسط بأخذ الفرق الاول فإذا لم تستقر نأخذ الفرق الثاني، وغالبا ما
تستقر بعد الفرق الاول او الثاني. اما عدم الاستقرار في التباين، فيتم معالجتها من
خلال اجراء التحويل المناسب للبيانات، فبعد تحقيق استقرار السلسلة الزمنية تبدأ
عملية تحديد النموذج.

حيث أن الهدف من عملية التشخيص هو الحصول على قيمة p و q و d التي سوف نحتاجها في النموذج العام ARIMA ثم الحصول على تقديرات للمعاملات حيث ان الدالتين المستخدمتين في تحديد النموذج ودرجته هما دالة الارتباط الذاتي (ACF) ودالة الارتباط الذاتي الجزئي (PACF) من خلال الرسم البياني لهذين الدالتين وبعدها يتم مطابقة معاملات الارتباط الذاتي والجزئي مع السلوك النظري لدالة الارتباط الذاتي ACF ودالة الارتباط الذاتي الجزئي PACF فإذا كانت :

-دالة (ACF) تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب المتضائلة وكانت دالة (PACF) تنقطع بعد الأزاحة (P) فإن النموذج الملائم للبيانات هو نموذج $AR(P)$.

-وإذا كانت دالة (PACF) تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب المتضائلة وكانت دالة (ACF) تنقطع بعد الأزاحة (q) فإن النموذج الملائم للبيانات هو نموذج $MA(q)$.

- وإذا كانت الدالة ACF و PACF تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب المتضائلة فإن النموذج الملائم هو $ARMA(P,q)$.

ثانياً . التقدير Estimation

أن عملية تقدير النموذج هي المرحلة الثانية من مراحل بناء السلاسل الزمنية وتحليلها ، فبعد تحديد وتشخيص النموذج الملائم للسلسلة الزمنية يتم تقدير معالمه باستخدام إحدى الطرق المعروفة في نظرية الأحصاء من أهمها :

أ- طريقة الأماكن الأعظم Maximum Likelihood Method :
وتتلخص هذه الطريقة في ان قيام مصفوفة معالم النموذج المراد تقديرها يتم اختيارها وفقاً لمبدأ تعظيم دالة الأماكن .

ب- طريقة المربعات الصغرى Method of Ordinary Least Squares

(O.L.S.E) تقوم هذه الطريقة على مبدأ تقليص مجموع مربعات خطأ

التقدير، وجعله في نهايته الصغرى

(١٤)

ثالثاً . اختبار دقة النموذج Diagnostic Checking of

Method

المرحلة الثالثة في منهجية بوكس - جنكنز بعد التعرف على النماذج الملائمة وتقدير معلوماتها هي فحص النموذج ومدى صلاحيته لتمثيل بيانات السلسلة الزمنية وذلك لمعرفة كل نموذج و كفاءته ومدى امكانية تطويره أو تحسينه . حيث نقوم في هذه المرحلة بإخضاع النموذج لعدد من الاختبارات الأحصائية فأذا اجتاز هذا النموذج الاختبارات يكون صالح للاستخدام وعندما يكون غير ملائم نعود الى المرحلة الثانية. ويتم اختبار ملائمة النموذج ومدى صلاحيته في تمثيل السلسلة الزمنية من خلال اختبار معنوية معالم النموذج وتحليل الارتباطات الذاتية للبواقي وكذلك تم وضع معايير التقييم لتحديد رتب نموذج ARIMA التي تساعد في المفاضلة بين النماذج المرشحة حيث يتم اختيار النموذج الافضل الذي يملك اقل قيمة لهذه المعايير ومن هذه المعايير الاحصائية :

- معيار المعلومات أكايك (AIC): ويستخدم معيار أكايك لاختيار النموذج الافضل حيث يقل تباينه بزيادة المعالم المقدره ومجموع مربعات البواقي قليل ويعرف رياضياً (S.Makridakis,1998):

$$AIC = \log(\sigma^2) + \frac{2(p+q)}{n} \quad (2-9)$$

حيث ان

σ^2 : تمثل تباين النموذج ، وان (p+q) يمثل عدد المعالم المقدره

- معيار معلومات البيزي (BIC): وهو طريقة اخرى لتحديد رتب

نموذج ARIMA حيث يكون النموذج الأفضل هو الذي يجعل قيمة BIC

اقل ما يمكن ويعرف رياضياً (S.Makridakis,1998):

$$BIC = \log(\sigma^2) + \left(\frac{p+q}{n}\right) \log(n) \quad (2-10)$$

رابعاً . التنبؤ Forecasting :

هو الخطوة الأخيرة من خطوات دراسة وتحليل نماذج السلاسل الزمنية وهو الهدف الأساسي من الدراسة ، بعد تحديد النموذج الملائم للبيانات يتم استخدامه لمعرفة قيمة الظاهرة المستقبلية ولفترات (L) ويمكن حساب التنبؤ بعدد خطوات (L) حسب

الصيغة (Douglas,1977):

$$\hat{X}_{t+L} = E[X_{t+L} | X_t, X_{t-1}, X_{t-2}, \dots] \quad \text{for } L \geq 1 \quad (2 - 11)$$

فإذا كان النموذج AR(1) فإن أفضل تنبؤ هو:

$$\hat{X}_{t+L} = \Phi^L X_{t-1+L} \quad (2 - 12)$$

أما إذا كان النموذج AR(2) فإن أفضل تنبؤ هو:

$$\hat{X}_{t+L} = \Phi_1^L X_{t-1+L} + \Phi_2^L X_{t-2+L} \quad L \geq 1 \quad (2 - 13)$$

وفي حالة الأوساط المتحركة MA(q) فإن أفضل تنبؤ:

$$\hat{X}_{t+L} = a_{t+L} - \Phi_1^L a_{t-1+L} - \Phi_2^L a_{t-2+L} - \dots - \Phi_q^L a_{t-q+L} \quad (2-14)$$

وفي حالة النموذج المختلط ARIMA فإن أفضل تنبؤ هو:

$$\hat{X}_{t+L} = \Phi_1^L X_{t-1+L} + \Phi_2^L X_{t-2+L} \quad (2-15)$$

الفصل الثالث

الجانب التطبيقي

مفهوم سرطان الثدي :

سرطان الثدي (Breast cancer) هو نوع من أنواع السرطان يظهر في أنسجة الثدي ويعرف : بأنه نمو غير طبيعي للخلايا المبطنة لقنوات الحليب أو لفصوص الثدي، غالباً ما يتكون الورم السرطاني في قنوات نقل الحليب و احياناً في الفصوص وجزء بسيط جداً في بقية الأنسجة .

أما أنواع الأورام التي تتكون في الثدي بشكل كتله فهي نوعين :

الأورام الحميدة: وهي ليست أوراماً سرطانية ويمكن أزلتها ، وفي أكثر الحالات لا تعود للظهور وأهم ما في هذه الأورام أنها لا تنتشر الى اماكن اخرى من الجسم، لذا هي لا تهدد حياة الإنسان فمثلا ورم الثدي الحميد من أعراضه ازدياد في حجم الثدي يصاحبه انتفاخ ، وان الأورام الحميدة في الثدي تعتبر مشكلة بسيطة حيث يمكن أزلتها بعملية جراحية غير معقدة وبمنتهى السهولة.

الأورام الخبيثة: الخلايا الخبيثة تنقسم بسرعة ولا تموت حسب النظام العام للخلايا وتسمى بالسرطان لأن بإمكانها غزو الخلايا المجاورة وباقي أعضاء الجسم ، كذلك يمكن لهذه الأورام أن تتفكك وتدخل في مجرى الدم أو الجهاز اللمفاوي وبهذه الطريقة ينتشر السرطان ليكون أوراماً ثانوية في أجزاء من الجسم مثل: العظام والكبد والرئه لذا فهي تهدد حياة الإنسان

ظهور أحد هذه الأعراض أو العلامات التالية يدل على بداية الإصابة بسرطان الثدي :

١. ألم موضعي في الثدي أم تحت الأبط (رغم أن معظم الأورام الخبيثة غير مصحوبة بالألم)
٢. وجود كتله أو غلاظه تحت الثدي أم الأبط ، تغير في شكل أو حجم الثدي ،
٣. إفرازات دموية أو غير دموية من الحلمه تغير في مظهر أو لون الحلمه (انقلاب الحلمه للداخل بشكل مستمر ،تغير في المكان أو الهيئة)
٤. الشعور بتغيرات في الجلد أو الحلمه من حيث المظهر (تشققات ، تهيج ، انكماش، شد للداخل) أو من حيث الأحساس.

علاج سرطان الثدي:

إذا كانت اورام الثدي حميدة فتعالج عادة بالاستئصال الجراحي ومن ثم متابعة المريضة حسب تعليمات الطبيب ولاتحتاج المريضة الى علاجات مكمله اما بالنسبة الى الأورام الخبيثة (السرطانيه) هناك طرق علاجية مختلفه ، فهي تعتمد على حجم الورم وموقعه في الثدي وكذلك على نتائج الفحوصات المختبرية للخلايا السرطانية ومرحلة المرض ، اذ يعتمد علاج سرطان الثدي على :

- ١- مرحلة المرض
- ٢- نوعية الخلايا السرطانية
- ٣- رغبة المريض

هناك عدة طرق مختلفة لعلاج أورام الثدي وهي:

- الجراحه
- العلاج بأشعاع
- العلاج الكيميائي
- العلاج الهرموني
- العلاج الموجه
- العلاج المناعي
- العلاج البيولوجي
- العلاج البديل والعلاج المكمل
- العلاج النفسي والدعم العاطفي

١-٣ جمع البيانات:

جمعت البيانات والتي تتألف من (٦٠) مشاهدة، والتي تعود الى المدة من كانون الثاني عام ٢٠١٢ ولغاية كانون الأول عام ٢٠١٦ ، وان هذه البيانات تمثل اعداد النساء المصابات بسرطان الثدي في مدينة كربلاء المقدسة والمأخوذة من سجلات مستشفى الحسين (ع) التعليمي مركز أورام الدم، شعبة الإحصاء. حيث تم استخدام البرنامج الإحصائي R في تحليل النتائج.

جدول رقم (١)

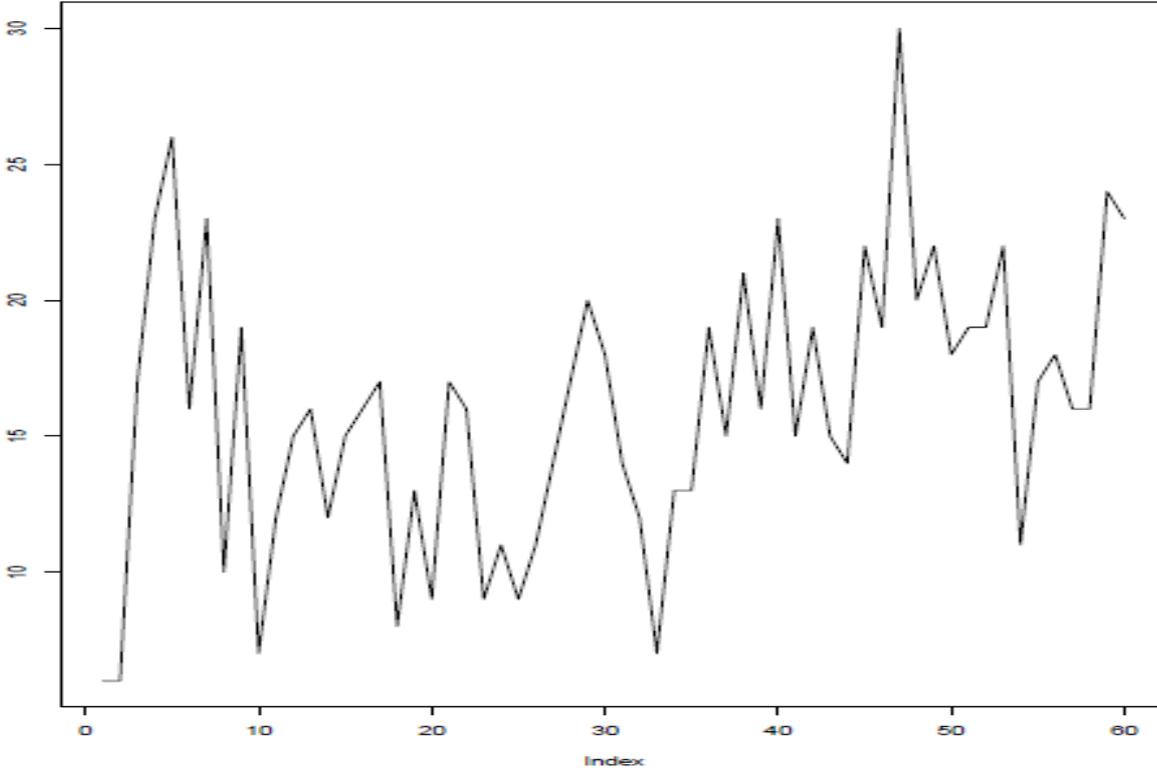
اعداد النساء المصابات بسرطان الثدي في مدينة كربلاء

Year Month	2012	2013	2014	2015	2016
January	6	16	9	15	22
February	6	12	11	21	18
March	17	15	14	16	19
April	23	16	17	23	19
May	26	17	20	15	22
June	16	8	18	19	11
July	23	13	14	15	17
August	10	9	12	14	18
September	19	17	7	22	16
October	7	16	13	19	16
November	12	9	13	30	24
December	15	11	19	20	23

٢-٣ مرحلة تهيئة البيانات:

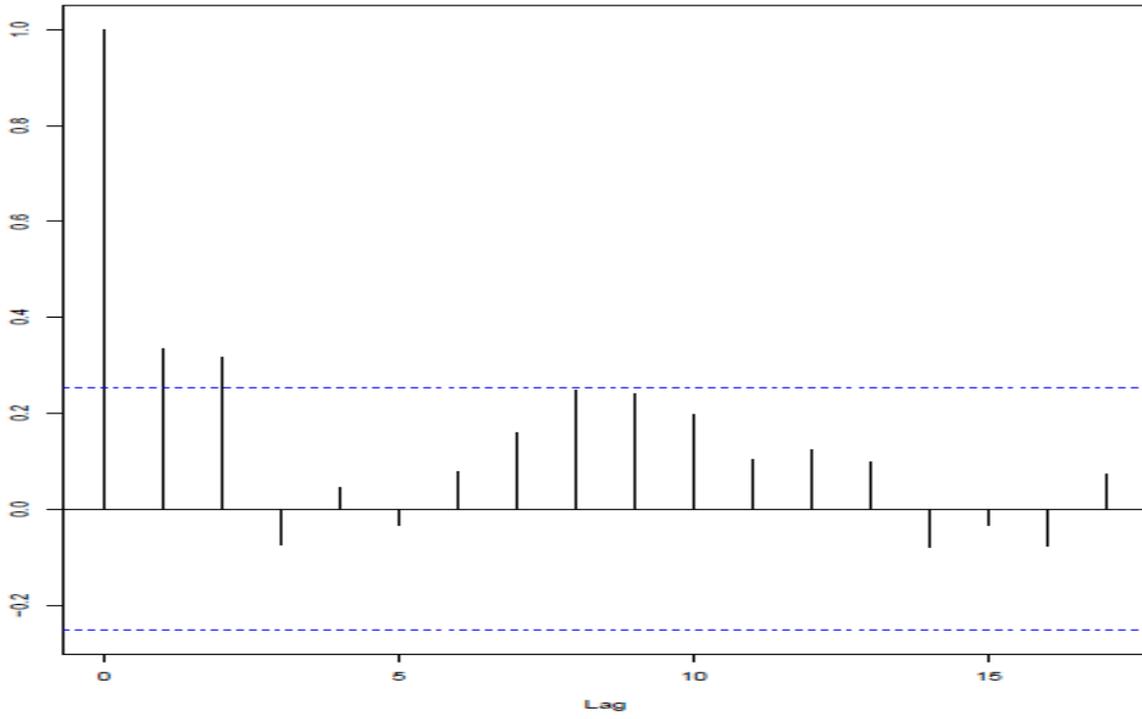
يتم في هذه المرحلة تحضير البيانات من خلال رسم شكل الانتشار، واستخراج معاملات الارتباط الذاتي والجزئي، فمن خلال شكل رقم (1) نلاحظ أن البيانات تميل إلى الثبات إلا أننا نلاحظ وجود اتجاه عام متزايد مع الزمن مما يدل على عدم استقراره بيانات السلسلة في المتوسط وقد أكدت ذلك قيم معاملات الارتباط الذاتي والجزئي كما في الشكل رقم (2)، والتي أظهرت فيه قيم معاملات الارتباط الذاتي، وبعد اخذ الفرق الأول نلاحظ فقدان الاتجاه العام في سلوكه، مما يدل على استقراره السلسلة في المتوسط، ومن خلال رسم السلسلة بعد اخذ الفرق الأول في الشكل رقم (4) والارتباط الذاتي والجزئي في الشكل (5)، (6) يوضح ان السلسلة مستقرة في المتوسط.

The observations of Breast cancer as Time series

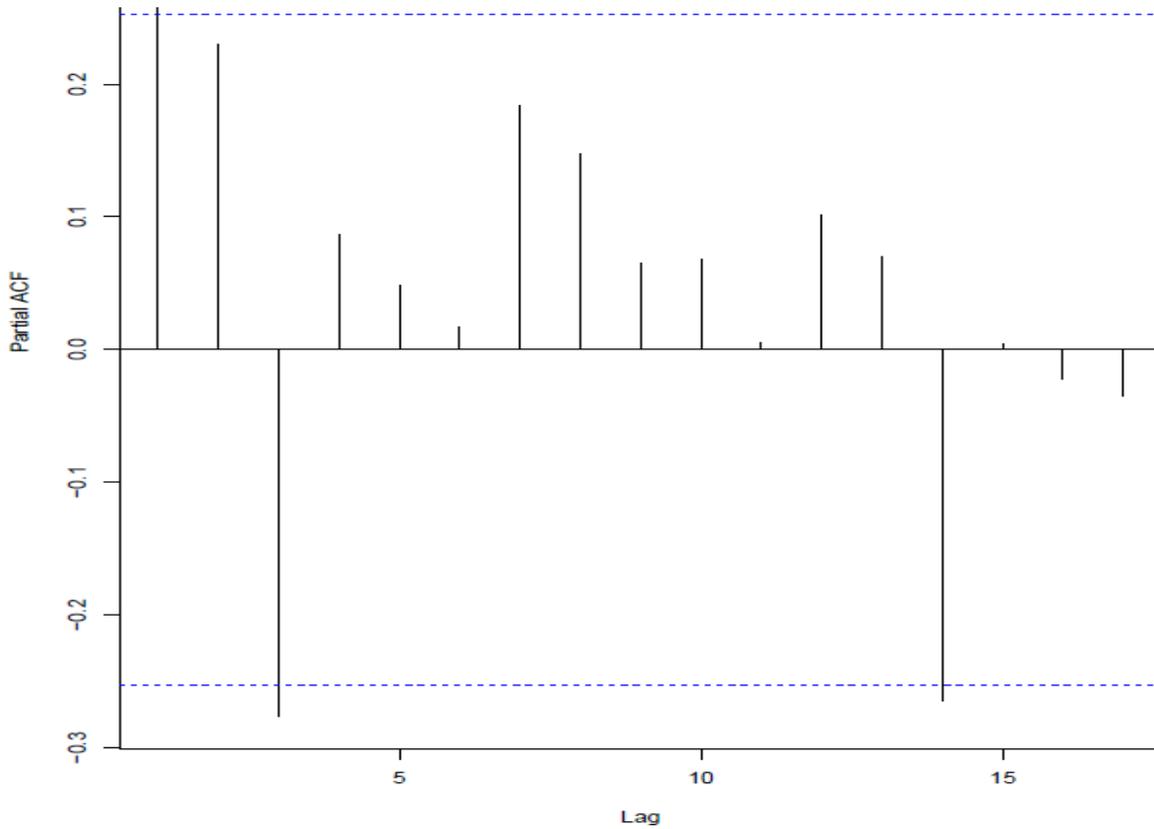


شكل رقم (1) يمثل اعداد النساء المصابات بأورام الثدي للفترة (2012-2016)

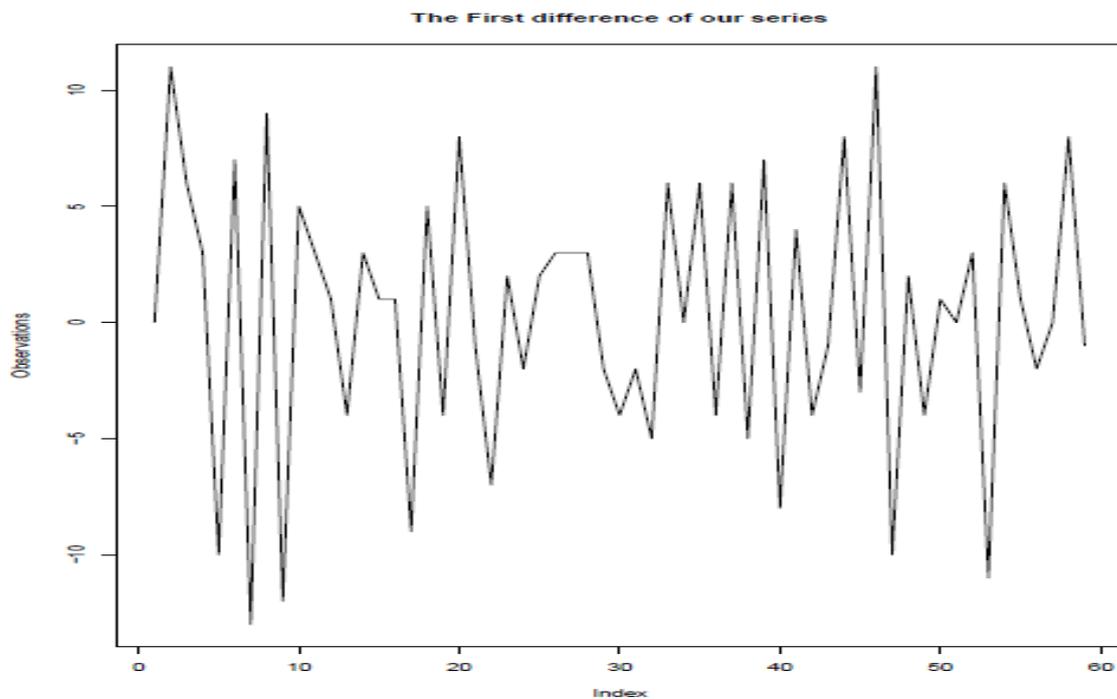
Auto-Correlation Function of our series



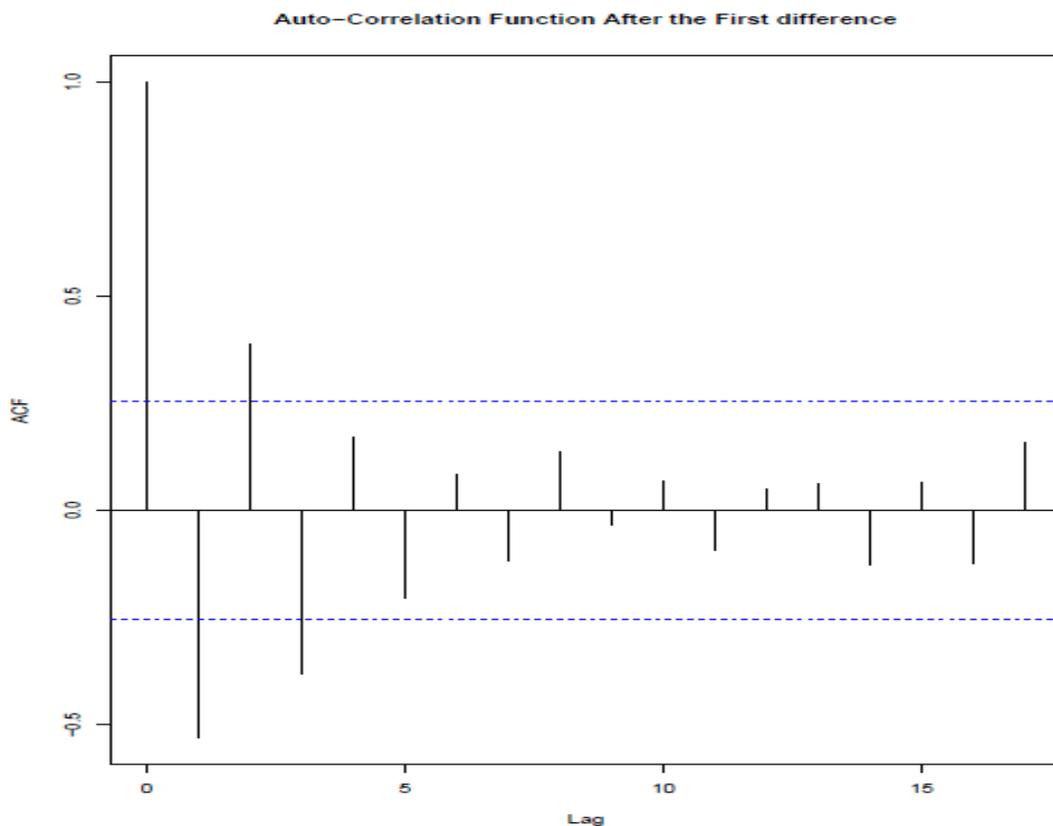
الشكل رقم (2) معاملات الارتباط الذاتي للسلسلة الزمنية



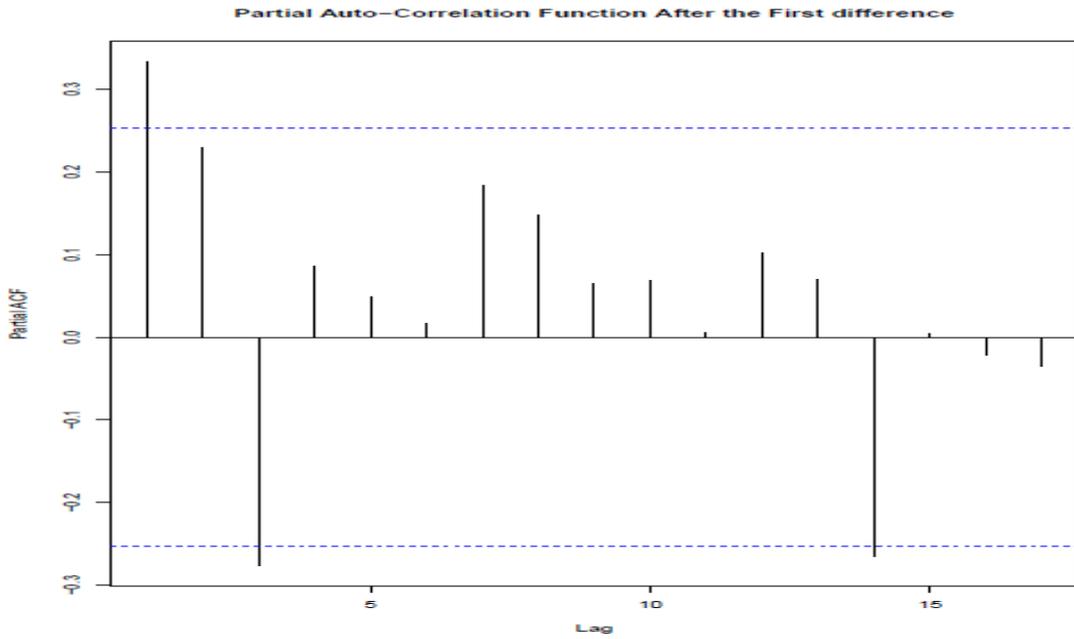
الشكل رقم (3) معاملات الارتباط الذاتي الجزئي للسلسلة الزمنية



شكل رقم(4) منحنى السلسلة الزمنية بعد اخذ الفرق الاول



الشكل رقم (5) معاملات الارتباط الذاتي بعد اخذ الفرق الاول للسلسلة الزمنية



الشكل رقم(6) معاملات الارتباط الذاتي الجزئي بعد اخذ الفرق الاول للسلسلة الزمنية

٣-٣-٣ التشخيص والتقدير:

ان الخطوة الأولى من مراحل بناء نموذج السلاسل الزمنية هي تشخيص النموذج (Identification)، وقد تم تطبيق معايير التشخيص التي تعتمد على شكل منحنى دالة الارتباط الذاتي للعينه (ACF)، وشكل منحنى دالة الارتباط الذاتي الجزئي (PACF) وعند مطابقة قيم معاملات الارتباط الذاتي والجزئي للسلسلة الزمنية بعد اخذ الفرق الاول مع السلوك النظري لها كما في الشكل(4) للاحظ ان النموذج هو $ARIMA(1,1,0)$ ومن خلال الاعتماد على رسم الارتباط الذاتي والارتباط الجزئي للسلسلة بعد اخذ الفرق الأول تبين ما يلي:

جدول رقم (2) يوضح رتب النموذج مع تقدير معالمه باستخدام دالة الامكان الاعظم

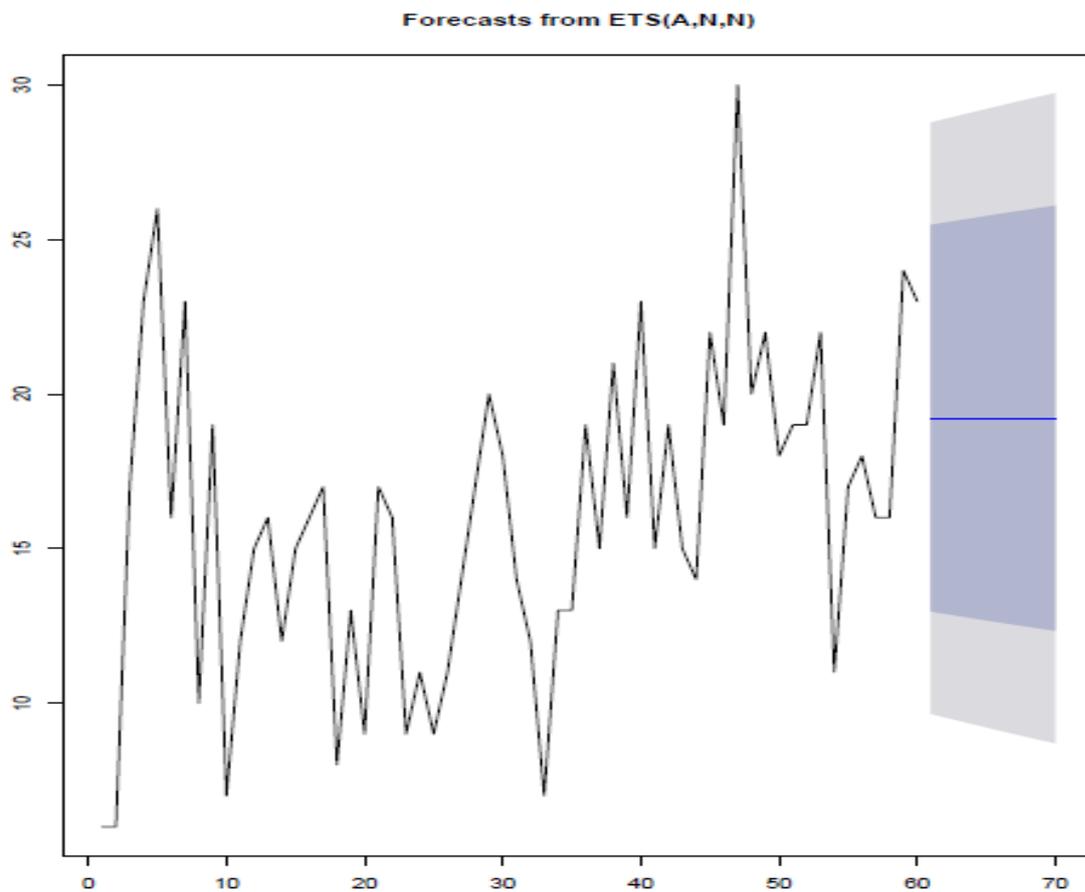
ARIMA(1,1,0)	Model order(p)
Coefficients	AR(1) 0.5191
S.e	0.1091
sigma ² estimated	24.15
Log. Likelihood	-177.31
AIC	358.61
AIC c	358.82
BIC	362.77

٣-٤ التنبؤ Forecasts:

تم استخدام النموذج للتنبؤ بأعداد المصابات بمرض سرطان الثدي في محافظة كربلاء من (شهر كانون الثاني-شهر تشرين الاول 2017) والتي عرضت نتائجه في جدول رقم (2)

وتم تمثيل السلسلة الزمنية لهذه التنبؤات في

الشكل رقم (7) والتي اظهرت تتبع نفس سلوك السلسلة الأصلية..



الشكل رقم (7) منحنى القيم التنبؤية لسلسلة اعداد المصابات بأورام الثدي
للفترة (كانون الثاني – تشرين الاول\2017)

جدول رقم (3) اعداد المصابات بأورام الثدي المتنبأ بها في محافظة كربلاء
للفترة (كانون الثاني-تشرين الاول\2017)

2017 month	Forecasts value	Low	High	Low	High
January	28.7852 9	9.65372 0	25.4742 3	12.96477	19.2195
February	28.8959 2	9.54308 0	25.5465 7	12.89243	19.2195
March	29.0053 0	9.43370 2	25.6180 9	12.82091	19.2195
April	29.1134 8	9.32552 7	25.6888 3	12.75018	19.2195
May	29.2204 8	9.21852 2	25.7587 9	12.68021	19.2195
June	29.3263 6	9.11265 1	25.8280 2	12.61099	19.2195
July	29.4311 3	9.00787 6	25.8965 3	12.54248	19.2195
August	29.5348 4	8.90416 6	25.9643 4	12.47467	19.2195
September	29.6375 2	8.80148 9	26.0314 8	12.40753	19.2195
October	29.7391 9	8.69981 3	26.0979 6	12.34105	19.2195

الفصل الرابع

الاستنتاجات والتوصيات

١-٤ الاستنتاجات:

١- نلاحظ من خلال دراسة سلسلة اعداد المصابات بأورام الثدي في محافظة كربلاء انها غير مستقرة في المتوسط وان هناك اتجاه عام واضح في السلسلة حيث ان ظهور تأثير الحروب التي اصابت البلد وكذلك المواد الكيميائية والمواد الاشعاعية اضافة الى الملوثات الطبيعية زاد من عددا لاصابات

٢- لقد تم تحقيق استقراريه السلسلة الزمنية بعد اخذ الفرق الاول للبيانات وبعد مطابقة معاملات الارتباط الذاتي والجزئي للسلسلة الزمنية مع السلوك النظري لدالتي الارتباط الذاتي والجزئي فقد اتضح ان دالة الارتباط الذاتي تتناقص تدريجيا مع زيادة فترات الازاحة k

٣- باستخدام معايير المفاضلة بين عدة نماذج وهي (معيار معلومات أكايك (AIC) ، معيار معلومات البيزي (BIC) ، وجد ان النموذج الملائم للبيانات هو نموذج الانحدار الذاتي من الرتبة الاولى $ARIMA(1,1,0)$

٤- باستخدام هذا النموذج للتنبؤ بأعداد المصابات بأورام الثدي في محافظة كربلاء للفترة (كانون الثاني-تشرين الاول\2017) اظهرت القيم التنبؤية تناسقا مع القيم الاصلية للسلسلة.

٢-٤ التوصيات:

من خلال النتائج التي تم التوصل اليها نوصي بما يأتي:

١- الأخذ بنتائج هذا البحث الذي يظهر تزايد في اعداد الأصابات بأورام الثدي بمرور الزمن مما يقتضي اتخاذ التدابير اللازمة من قبل الجهات المختصة والكفيلة بالحد من هذه الظاهرة

٢- نوصي النساء باتباع الخطوات السليمة لصحة الثدي والمواظبة على الفحص الذاتي شهريا بعد تخطي سن العشرين فاذا تم اكتشاف سرطان الثدي مبكرا فسيكون للمرأة خيارات عديدة لعلاجها وأمل اكبر في شفائها .

المصادر

١. الجبوري ، وليد دهان صليبي ،(2010) "التنبؤ بمستوى التضخم في اسعار المستهلك الشهرية في العراق باستخدام السلاسل الزمنية ثنائية المتغيرات" رسالة ماجستير في الاحصاء ، كلية الإدارة والاقتصاد ،جامعة المستنصرية
٢. الخضيرى، محمد قدوري عبد، (1996) "دراسة مقارنة لطرائق التقدير والتنبؤ لبعض نماذج بوكس وجينكنز الموسمية" ، رسالة ماجستير في الإحصاء ، جامعة بغداد، كلية الإدارة والاقتصاد..
٣. الصراف، نزار مصطفى،(1981) "تحليل السلاسل الزمنية باستخدام التقنية الاحصائية للتنبؤات الاقتصادية في العراق"، رسالة ماجستير في الإحصاء ، كلية الإدارة والاقتصاد، جامعة بغداد..
٤. الكاطع ، احلام حنش،(2007) "اختبارات التكامل الكسري في نماذج ARIMA" رسالة ماجستير في الإحصاء، جامعة بغداد ، كلية الإدارة والاقتصاد..
٥. المتولي ، احمد شاكر محمد طاهر،(1989)"استخدام تحليل التدخل في السلاسل الزمنية وتطبيقاتها في البيانات البيئية" ، رسالة ماجستير في الإحصاء ، جامعة صلاح الدين، كلية الإدارة والاقتصاد..
٦. الأنعيمي، محمد عبد العال،ومحمد حبيب الشاروط،(2000)"استخدام نموذج التدخل في السلاسل الزمنية لتقدير عدد المصابين بالأورام الخبيثة في محافظة القادسية"، مجلة القادسية للعلوم الإدارية والاقتصادية ،المجلد (3) العدد 1..

7. Ahtola, J. and Tiao, G. C. (1987), "A note on Asymptotic Inference in Autoregressive Models with Roots on the unit circle", Journal of time series analysis, Vol. 8 No. 1.
8. Anderson, R. L. (1942), "Distribution of the series Analysis Correlation Coefficient", Ann. Mat. Statistic, Vol. 13.
9. Box, G.E.P. and Jenkins, G.M. (1976), "Time Series Analysis Forecasting and control", Holden day, London.
10. Box, G.E.P. & Pierce, D.A., (1970), "Distribution of the Residual Autocorrelation in Autoregressive-integrated moving Average Time Series Model", JASA, VOL. 65
11. Douglas, C.M. & Contreas, J.G., (1976), "Note On Forecasting With Adaptive Filtering", O.P.Q, Vol. 24, No. 4.
12. S. Makridakis, S.C. Wheelwright and R.J. Hyndman (1998), "Forecasting Methods and Applications", Third Edition, John Wiley & Sons, Inc. Publication, New York, USA.

