

<https://jsras.rcc.edu.ly/>
 Vol.1 No.2 (2024), 38-60
 Article history:
 Received: 25 Oct. 2024
 Accepted: 14 Nov. 2024
 Published: 18 Nov. 2024

مجلة البحوث المستدامة في العلوم التطبيقية



المفاضلة بين نموذج الانحدار الخطي البسيط ونماذج السلاسل الزمنية في التنبؤ بعدد الإصابات بمرض السرطان في المعهد القومي لعلاج الأورام – مصراتة خلال الفترة من 2012 إلى 2017

امحمد ميلاد إبراهيم بعوه¹

¹جامعة مصراتة، كلية العلوم، قسم الإحصاء، ليبيا

بريد المؤلف المراسل: emhemmedmilad2017@gmail.com

الملخص:

يهدف هذا البحث إلى المفاضلة بين أسلوب الانحدار الخطي البسيط وأسلوب السلاسل الزمنية واختيار الأسلوب الأفضل من بينها وفقاً للاختبارات الإحصائية وبعض معايير المفاضلة وذلك للتنبؤ المستقبلي لعدد الإصابات بمرض السرطان، وقد أظهرت البيانات أنه يوجد ارتباط طردي قوي بلغ 0.88 بين الزمن وعدد المصابين وأن هناك إمكانية لتمثيل هذه البيانات بخط مستقيم، ومن ثم إيجاد معادلة انحدار خطي بسيط وكان نموذج الانحدار الخطي البسيط على قدر من الجودة في توفيق البيانات، حيث بلغ معامل التحديد 78%، وكذلك كانت سلسلة مرض السرطان غير مستقرة ويوجد بها اتجاه عام وأن النموذج الملائم لها هو نموذج الانحدار الذاتي والمتوسطات المتحركة المتكامل من الدرجة الأولى $ARIMA(1,1,1)$ وقد تم المقارنة والمفاضلة بين هذا النموذج ونموذج الانحدار الخطي البسيط في التنبؤ المستقبلي لعدد الإصابات بمرض السرطان وتبين أن هذا النموذج أفضل من نموذج الانحدار الخطي البسيط، وذلك وفقاً للاختبارات الإحصائية وبعض معايير المفاضلة الأخرى، وهذا ما أكدته القيم المتنبئ بها داخل العينة باستخدام النموذج $ARIMA(1,1,1)$ حيث أظهرت تقارب كبير من القيم الحقيقية عكس نموذج الانحدار الخطي البسيط والتي كانت القيم المتنبئ بها داخل العينة على هيئة خط مستقيم وبالتالي فهي أقل تقارب من القيم الحقيقية.

الكلمات المفتاحية: الانحدار الخطي البسيط، السلاسل الزمنية، منهجية بوكس – جنكز، الأورام السرطانية.

1 المقدمة:

يعد علم الإحصاء أحد العلوم المهمة والحيوية في البحث العلمي التي شهدت تطوراً في العصر الحديث، حيث أنه يعطي العديد من النظريات والطرق والأساليب اللازمة للقيام بالدراسات والبحوث، ومن أهم فوائد تحليل الظواهر هو التنبؤ الذي يعتبر عملية عرض حالي لمعلومات مستقبلية باستخدام معلومات تاريخية بعد دراسة سلوك الطاهر في الماضي.

ومن أهم الأساليب الإحصائية للقيام بعملية التنبؤ هو أسلوب تحليل الانحدار الخطي البسيط، والذي يعتبر من أهم وأقوى أساليب التحليل الإحصائي، حيث يلعب تحليل الانحدار دوراً كبيراً في دراسة الظواهر في مختلف مجالات الحياة، وذلك من خلال تحليل هذه الظواهر والتنبؤ بنتائجها المستقبلية. وكذلك يعتبر أسلوب السلاسل الزمنية من الأساليب الإحصائية الحديثة والهامة التي تدرس التغيرات التي

تحدث على قيم الظاهرة خلال فترة زمنية معينة، وتحدد الأسباب والنتائج وتقيس العلاقة المشاهدة بينها والتنبؤ بقيم الظاهرة في المستقبل على ضوء ما حدث لها في الماضي.

ولقد جاءت هذه الدراسة لغرض المقارنة بين نموذج الانحدار الخطي البسيط ونماذج السلاسل الزمنية في عملية التنبؤ بعدد الإصابات لمرض السرطان، حيث تم الاعتماد على بيانات شهرية واقعية من أعداد المصابين بالأورام السرطانية في المعهد القومي لعلاج الأورام خلال الفترة (2012-2017م)، وذلك من أجل تحديد الأسلوب الأفضل من بينها في عملية التنبؤ لغرض تقليل حجم الخطأ إلى أقل حد ممكن وبالتالي يمكن تقدير أعداد الإصابات بمرض السرطان في المستقبل بحيث يكون قريب من الأعداد الحقيقية حتى يتمكن المسئولون من اتخاذ القرارات الصحيحة.

أهمية وأهداف البحث:

يعد التنبؤ من الموضوعات التي لها أهمية كبيرة والتي من خلالها يتمكن المسئولين وأصحاب القرار من اتخاذ القرارات الصحيحة في كافة المجالات، ويعد التنبؤ بأعداد الإصابات بالسرطان لفترات لاحقة ذو أهمية كبيرة، وذلك لغرض الاستعداد لهذه الظاهرة من حيث توفير الإمكانات اللازمة وتهيئة الظروف المناسبة لاستقبال الأعداد المتزايدة من هذا المرض.

وتكمن أهمية هذا البحث في تحديد الأسلوب الإحصائي المناسب في عملية التنبؤ وتوضيح مبررات ودواعي استخدام كل أسلوب واختيار الأسلوب الأكثر دقة وكفاءة حتى يعطي صورة واضحة للمختصين والمهتمين بتحليل مثل هذه البيانات، وبالتالي تقليل الخطأ إلى أقل حد ممكن مما يساعد متخذي القرارات في اتخاذ القرارات الصحيحة دون الوقوع في أخطاء كبيرة.

الدراسات السابقة:

- رسالة ماجستير بعنوان: " المقارنة بين نموذج السلاسل الزمنية والانحدار الخطي البسيط في التنبؤ بحجم المبيعات - دراسة حالة مؤسسة مطاحن الأبييض الحديثة من الفترة (1995-2013) "، إعداد الطالب: أبو ذر إسماعيل مقرح إسماعيل في العام 2014 م: حيث هدفت الدراسة إلى المقارنة بين نموذج الانحدار الخطي البسيط ونماذج السلاسل الزمنية في التنبؤ بمبيعات شركة مطاحن الأبييض الحديثة في الفترة (1995-2013م) والوصول للنموذج الأكثر قوة في التنبؤ والتفسير. توصلت هذه الدراسة إلى أن البيانات لم تتخذ اتجاه معين مما أدى إلى ضعف القوة التفسيرية لنموذج الانحدار حيث بلغت قيمة معامل التحديد $(R^2 = 0.35)$. أما نماذج السلاسل الزمنية فقد كان النموذج المناسب هو النموذج $ARIMA(0,1,1)$ وهو يفوق نموذج الانحدار في قيمته التفسيرية حيث بلغت الفرق الأول أصبحت السلسلة مستقرة. أن تذبذب السلسلة وعدم اتخاذها خط معين في تحديد الاتجاه ناتج من الارتفاع والانخفاض في المبيعات مع قلة حجم العينة المتوفرة.

- رسالة ماجستير بعنوان: " المفاضلة بين نموذج السلاسل الزمنية ونموذج الانحدار الخطي البسيط في التنبؤ بحجم المبيعات في المؤسسة الاقتصادية ":

هدفت الدراسة إلى محاولة إيضاح وتفسير النماذج والأساليب الكمية الفعالة في تسيير وإدارة المبيعات انطلاقاً من التنبؤ. وكذلك تحسين تقديرات المبيعات من خلال تحديد النموذج المناسب مما يضمن التقليل من الخطأ في تقدير حجم المبيعات. وتوصلت الدراسة إلى أن تطبيق معايير المفاضلة بين النموذجين في التنبؤ بحجم المبيعات اتضح أفضلية نموذج السلاسل الزمنية على نموذج الانحدار الخطي

البسيط ويرجع ذلك لتوفر معطيات حول حجم المبيعات لفترات سابقة عديدة ونقص بعض المعطيات الكافية لنموذج الانحدار الخطي البسيط وصعوبة تحديد أهم العوامل المؤثرة بصورة مباشرة في المبيعات.

1 الجانب النظري:

يتناول هذا البند استعراضاً لبعض المفاهيم العامة للارتباط والانحدار الخطي البسيط، وطريقة إيجاد نموذج انحدار خطي بسيط يوافق البيانات المتاحة وكذلك عرض بعض المفاهيم الخاصة بالسلاسل الزمنية وكيفية بناء نموذج السلاسل الزمنية، واعتماد تحليلها على (منهجية بوكس-جنكنز).

1.1 الارتباط (Correlation):

يبحث تحليل الارتباط في دراسة العلاقة بين متغيرين (بسيط) أو أكثر (متعدد) لمعرفة إذا كان أحدهما أو مجموعة منها مرتبطاً بمتغير آخر، وبعبارة أخرى فإن تحليل الارتباط يهتم بدراسة قوة واتجاه العلاقة بين متغيرين أو أكثر دون التعرض لدراسة العلاقة السببية ونتيجتها، ويمكن تقسيم الارتباط إلى عدة أنواع كما يلي:

1.1.1 من حيث قوة الارتباط:

أ. الارتباط التام (Perfect correlation): ويظهر هذا النوع من الارتباط في العلوم التطبيقية كالعلاقة بين مساحة الدائرة ونصف قطرها، ويندر في العلوم الاقتصادية والإدارية والتربوية.

ب. ارتباط غير تام (Imperfect correlation): وفي هذا النوع يمكن الجزم بوجود علاقة بين متغيرين ولكن يصعب تفسير التغير في إحدى المتغيرين كلياً بالتغير في المتغير الثاني فمثلاً مبيعات سلعة معينة يتأثر بسعر هذه السلعة ولكن هناك متغيرات أخرى تؤثر على مبيعات السلعة كالدعاية وذوق المستهلك.

2.1.1 من حيث عدد المتغيرات:

أ. ارتباط بسيط (Simple correlation): ويدرس العلاقة بين متغيرين فقط، كالعلاقة بين الطلب والعرض.

ب. ارتباط متعدد (Multiple correlation): ويدرس العلاقة بين أكثر من متغيرين كالارتباط بين كمية الإنتاج ونوع التربة وكمية المياه وكمية الأسمدة.

3.1.1 من حيث شكل العلاقة:

أ. ارتباط خطي (Linear correlation): حيث يمكن تمثيل العلاقة بين متغيرين أو أكثر بخط مستقيم وتعني هذه العلاقة بأن التغير في أحد المتغيرين يكون ثابتاً إذا زاد المتغير الآخر بمقدار ثابت.

ب. ارتباط غير خطي (Non-linear correlation): حيث يمكن تمثيل العلاقة بين المتغيرين أو أكثر بنموذج غير خطي تعني هذه العلاقة أن التغير في إحدى المتغيرين إذا زاد المتغير الآخر بمقدار معين غير ثابت.

2.1 قياس الارتباط: يمكن قياس الارتباط بين متغيرين كما يلي:

1.2.1 رسم شكل الانتشار (Scatter Diagram):

يقدم شكل الانتشار صورة سريعة لطبيعة العلاقة بين المتغيرين ومدى قوتها واتجاهها فإننا نستطيع بمجرد النظر إلى الشكل أن نحكم بوجود أو عدم وجود علاقة بين المتغيرين.

2.2.1 القياس الكمي للارتباط:

إن التعرف على حقيقة قوة الارتباط بين المتغيرين بشكل واضح من خلال شكل الانتشار غير كافي ولا بد من التعرف عن هذه العلاقة بشكل رقمي ويوجد عدة طرق لقياس العلاقة كمياً منها معامل ارتباط بيرسون.

3.1 الانحدار (Regression):

بعد الجزم بوجود علاقة بين متغيرين أو أكثر، والتعرف على طبيعة وقوة هذه العلاقة، يوجه اهتمامنا إلى صياغة النموذج الذي يمثل هذه العلاقة باستخدام الطرق الإحصائية والرياضية ويتطلب صياغة النموذج إلى تحديد المتغير التابع (Dependent Variable) ونرمز له بالرمز Y والمتغيرات المستقلة (Independent Variables) ويرمز لها ب X_1, X_2, \dots, X_k ويكون النموذج خطياً (Liner model) أو غير خطي (Non-Liner model) وأن دراسة الانحدار تعني صياغة النموذج وتقدير معالمه واختبار الفرضيات المتعلقة بها وتحليلها واستخدامها في التنبؤ بقيم المتغير التابع عند مستويات محددة.

1.3.1 النموذج الخطي البسيط (Simple Linear Model):

إن أبسط علاقة دالية تربط بين متغيرين Y, X هي $Y = f(x)$ حيث أن $f(x)$ ترمز إلى أن Y دالة في (X) ، أي أن قيم (Y) تتغير تبعاً لقيم (X) ولكن هذه العلاقة لا تتحدد بشكل العلاقة بين المتغيرين فقط، تكون خطية أو منحنى إلخ ويعتمد تحديد صيغة الدالة على الباحث.

فمثلاً: قد نفرض وجود علاقة خطية ولهذا التقريب يكون شكل المعادلة صيغة الدالة الخطية كما يلي:

$$Y_i = \beta_0 + \beta_1(x_i) + e_i$$

حيث أن Y_i : المتغير التابع أو المفسر.

X_i : المتغير المستقل وسمى بالمتغير التفسيري (Explanatory) لأنه يفسر ما يحدث في المتغير التابع Y_i .

β_0 : معلمة النموذج المجهولة وهي المعامل الثابت أو الجزء المقطوع من محور الصادات.

β_1 : معلمة النموذج المجهولة وهي معامل الانحدار وهي ميل خط الانحدار.

e_i : متغير عشوائي غير مشاهد يأخذ قيمة موجبة وسالبة ويطلق عليه (الخطأ العشوائي).

وتسمى المعادلة السابقة بمعادلة الانحدار الخطي للمجتمع بما أن β_0 و β_1 مجهولتان فإن الهدف الحصول على قيمة تقديرية لها على أساس مشاهدات عينة مسحوبة من مجتمع الدراسة.

وتقدير معالم نموذج الانحدار الخطي البسيط (β_1, β_0) بطريقة المربعات الصغرى وهي من أهم الطرق المتعلقة بتقدير المعالم

المجهولة (β_1, β_0) وذلك باستخدام بيانات العينة وذلك لنحصل على المقدرات $(\hat{\beta}_1, \hat{\beta}_0)$ التي تجعل مجموع مربعات الأخطاء

$$\sum_{i=1}^n e_i^2 \text{ أقل ما يمكن أي نهاية صغرى، وبالتالي فإن}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\left(\sum_{i=1}^n x_i y_i \right)^2 - n \bar{x}^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

2.3.1 اختبارات الجودة الإحصائية لنموذج الانحدار الخطي البسيط:

أ- اختبار النموذج ككل بواسطة إحصائية (F):

يتم اختبار ما إذا كان التباين المفسر أمبر من التباين الغير مفسر وذلك عن طريق اختبار الفرض التالي:

$$H_0 : \beta_1 = 0 , \quad H_1 : \beta_1 \neq 0$$

لإجراء هذا الاختبار نستخدم إحصائية F

$$F = \frac{MSR}{MSE} = \frac{\text{متوسط مربعات الانحدار}}{\text{متوسط مجموع مربعات البواقي}}$$

أولاً: تحديد القيمة الجدولية $F_{1, n-2, \alpha}$ فإذا كانت F المحسوبة أكبر من F الجدولية نرفض فرض العدم ونقبل البديل، أي قيمة المعلمة β_1 تختلف عن الصفر ونحكم بوجود علاقة خطية بين المتغيرين.

ثانياً: إذا كان F المحسوبة أصغر من F الجدولية ليس لدينا دليل الكافي لرفض فرض العدم H_0 أي نحكم بعدم وجود علاقة خطية بين المتغيرين Y , X ذات دلالة إحصائية أن قيمة $\hat{Y} = \hat{\beta}_0$ أي قيمة Y ثابتة ولا يكون لقيم X أي تأثير خطي على قيم Y.

ب- اختبار جودة التوفيق باستخدام معامل التحديد:

يستخدم معامل التحديد لتوضيح مدى جودة توفيق خط الانحدار المقدر لبيانات العينة المستخدمة في تقديره ويرمز له بالرمز (r^2) وهو يمثل نسبة التباين التي تم تفسيرها عن طريق نموذج الانحدار إلى نسبة التباين الكلي في قيم المتغير التابع التي يمكن تفسيرها عن طريق نموذج الانحدار.

$$SST = SSE + SSR$$

$$\frac{SST}{SST} = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$1 = r^2 + \frac{SSE}{SST}$$

$$r^2 = 1 - \frac{SSE}{SST}$$

وعليه فإن معامل التحديد

$$r^2 = \frac{SSR}{SST}$$

إذا كانت القيمة العظمى لمعامل التحديد $r^2=1$ وذلك يعني عندما تكون $SSR=SST$ و $SSE=0$ أي أن القيمة الفعلية تساوى القيم المقدرة، أي العلاقة تامة.

القيمة الصغرى لمعامل التحديد $r^2=0$ وذلك يعني عندما تكون $SSR=0$ و $SST=SSE$ التباين الكلي = التباين غير المفسر، أي لا توجد علاقة.

4.1 السلاسل الزمنية (Time series):

هي مجموعة من المشاهدات أو القياسات لقيم ظاهرة معينة مأخوذة على فترات زمنية محددة، ويمكن تمثيل السلسلة الإحصائية

$$z_t = f(t) + e_t$$

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

e_t : الجزء العشوائي وقد يسمى بالضجيج (التشويش).

وهناك نوعان من السلاسل الزمنية:

1. السلاسل الزمنية المستقرة.

2. السلاسل الزمنية غير المستقرة.

- الاستقرار في المتوسط (Stationary in mean)

وهي الحالة التي لا تظهر فيها السلسلة اتجاهاً عاماً ويمكن تحويلها إلى مستقرة بأخذ الفروق، وغالباً ما تستقر بعد أخذ الفرق الأول والثاني.

- الاستقرار في التباين (Stationary in variance):

وهي الحالة التي لا تظهر فيها السلسلة تذبذبات (تعرجات) متباينة في شكلها، ويمكن تحويلها إلى مستقرة بتثبيت التباين بأخذ اللوغاريتم الطبيعي أو الجذر التربيعي أو المقلوب لمشاهدات السلسلة الأصلية.

- الارتباط الذاتي (Autocorrelation (AC)

وهي عبارة عن مؤشر يوضح درجة العلاقة بين قيم المتغير نفسه عند فترات إزاحة (k) مختلفة ويميز له بالرمز (\hat{P}_k) وتتراوح قيمته بين (1, -1) ويقدر من العلاقة:

$$\hat{P}_k = \frac{\sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}$$

حيث Z_t مشاهدات السلسلة.

\bar{z} : الوسط الحسابي للسلسلة، n : عدد المشاهدات.

ويطلق على الرسم البياني لمعاملات الارتباط الذاتي (\hat{P}_k) مقابل فترات إزاحة مختلفة بدالة الارتباط الذاتي ويرمز لها بالرمز (ACF).

الارتباط الذاتي الجزئي (Partial Auto correlation (PAC)

وهو مؤشر يقيس درجة الارتباط الخطي بين متغيرين يبعدان عن بعضهما البعض وحدتان زمنيتان بعد حذف تأثير المتغير الذي يقع بينهما ويحسب من العلاقة:

$$\hat{\phi}_{kn} = \frac{\hat{P}_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \hat{P}_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \hat{P}_j}$$

ويطلق على الرسم البياني لمعاملات الارتباط الذاتي الجزئي مقابل فترات إزاحة مختلفة بدالة الارتباط الذاتي الجزئي ويرمز لها بالرمز (PACF).

ويستخدم دالة الارتباط الذاتي والذاتي الجزئي في تحديد الاستقرار وكذلك تحديد نوع ورتبة النموذج وفحص ملائمة النموذج للبيانات.

1.4.1 نماذج بوكس - جنكنر للسلاسل الزمنية:

أ) نماذج الانحدار الذاتي (AR) Auto Regressive Models:

ويعبر في نماذج الانحدار الذاتي عن قيمة السلسلة الحالية Z_t بدلالة قيم السلسلة السابقة $(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p})$ وقيمة الخطأ e_t ويمكن التعبير عن هذه النماذج من الرتبة (P) كما يلي:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + e_t$$

ϕ_0 : الحد الثابت

$\phi_1, \phi_2, \dots, \phi_p$ معالم نموذج الانحدار الذاتي

e_t : الأخطاء العشوائية

ويمكن كتابة هذا النموذج بدلالة الارتداد الخلفي (B) كما يلي:

$$\phi_p(B) Z_t = \phi_0 + e_t$$

$\phi_p(B) Z_t$ متعدد حدود في (B) لمعالم نموذج الانحدار الذاتي $(\phi_1, \phi_2, \dots, \phi_p)$ وكحالة خاصة عندما $(P=1)$ فإن معادلة نموذج الانحدار الذاتي:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + e_t$$

وهي تمثل نموذج انحدار ذاتي من الدرجة الأولى (AR(1) إن شرط الاستقرار في هذا النموذج يتطلب أن تكون جذور المعادلة $\phi(B) Z_t = 1 - \phi B = 0$ خارج دائرة الوحدة و $-1 < \phi_1 < 1$.

وعندما $(p = 2)$ فإننا نحصل على نموذج انحدار ذاتي من الرتبة الثانية (AR(2) وتكون معادلتها على الصيغة التالية:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t$$

ولكي يكون النموذج (AR(2) مستقر فإنه يجب أن تقع جذور المعادلة $1 - \phi_1 B - \phi_2 B^2 = 0$ خارج دائرة الوحدة، أي يجب أن تحقق المعلمتين (ϕ_1, ϕ_2) للشروط التالية:

$$1. \quad \phi_1 + \phi_2 < 1$$

$$2. \quad \phi_2 - \phi_1 < 1$$

$$3. \quad -1 < \phi_2 < 1$$

ب) نماذج المتوسطات المتحركة (MA) Moving Average Models:

يتم التعبير في نموذج المتوسطات المتحركة عن قيم السلسلة الحالية Z_t بدلالة القيم الحالية للأخطاء (e_t) وبدلالة القيم السابقة للأخطاء $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$ ويمكن التعبير عن هذه النماذج من الرتبة (q) كما يلي:

$$Z_t = \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} + \dots - \theta_q e_{t-q}$$

حيث Z_t مشاهدات السلسلة.

θ_0 الحد الثابت

$\theta_1, \theta_2, \dots, \theta_q$ معاملات النموذج.

$(e_t, e_{t-1}, \dots, e_{t-q})$ الأخطاء العشوائية.

ويمكن كتابة هذا النموذج بدلالة الارتداد الخلي (B) كما يلي:

$$z_t = \theta_0 + \theta_q (B) e_t$$

حيث $\theta_q (B) e_t$ متعدد حدود في (B) لمعالم نموذج المتوسطات المتحركة $(\theta_1, \theta_2, \dots, \theta_q)$

ج) نماذج الانحدار الذاتي والمتوسطات المتحركة (ARMA)

أحياناً لا يمكن التعبير عن السلسلة الزمنية بصيغة الانحدار الذاتي (AR(P) فقط أو صيغة المتوسطات المتحركة (MA(q) فقط، ولكن

يمكن التعبير عنها بنموذج مختلط انحدار ذاتي ومتوسطات متحركة وهو نموذج مركب يحتوي على خصائص نموذج الانحدار الذاتي

والمتوسطات المتحركة وبذلك يمكن التعبير في هذا النموذج عن القيم الحالية للسلسلة Z_t بدلالة القيم السابقة $(z_{t-1}, z_{t-2}, \dots, z_{t-q})$

وكذلك القيم الحالية للأخطاء e_t بدلالة القيم السابقة $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$ ونرمز له بالرمز ARMA(p,q) ويمكن كتابة هذا النموذج

على النحو :

$$z_t = \phi_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

$$\phi_p (B) z_t = \theta_q (B) e_t$$

حيث $\phi_p (B) z_t$ متعدد حدود في (B) لمعالم نموذج المتوسطات المتحركة $(\theta_1, \theta_2, \dots, \theta_q)$.

د) نماذج الانحدار الذاتي والمتوسطات المتحركة التكاملية:

Auto Rgressive Integrated Moving Average Models (ARIMA)

يوجد بعض نماذج السلاسل الزمنية تكون غير مستقرة، ولكي تكون مستقرة يجب إجراء بعض التحويلات أو الفروق عليها إن مثل هذه

النماذج تختلف عن النماذج الأصلية وذلك لاحتوائها على هذه التحويلات التكاملية ويرمز لها بالرمز ARIMA(p,d,q).

p : رتبة نموذج الانحدار الذاتي.

q : رتبة المتوسطات المتحركة.

d : عدد الفروق اللازمة كي تكون السلسلة مستقرة.

ويمكن كتابة النموذج بدلالة الارتداد الخلفي (B) كما يلي:

$$\begin{aligned} \phi(B) = \phi(B)(1-B)^d \quad z_t &= \theta_0 + \theta_q (B) e_t \\ &= (1 - \phi_1 B - \dots - \phi_p B^p) \end{aligned}$$

$$\phi(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

وبفرض أن $\nabla^d z_t = z_t$

2.4.1 مراحل بناء نموذج السلاسل الزمنية:

يتم بناء نموذج السلسلة الزمنية وفقاً منهجية بوكس – جنكنر عبر أربعة مراحل هي:

مرحلة التعرف (Identification):

يقصد بمرحلة التعرف هي اختيار رتب النموذج (p , d , q) ويجب أن يسبق هذه المرحلة مرحلة تهيئة البيانات، فإذا كانت البيانات مستقرة تنتقل إلى مرحلة التعرف وإذا كانت غير مستقرة فيتم معالجتها ومن ثم الدخول في مرحلة التعرف أن الأدوات المستخدمة في تحديد نوع ورتبة النموذج هما دالة الارتباط الذاتي (ACF) ودالة الارتباط الذاتي الجزئي (PACF) ومن ثم مطابقة معاملات الارتباط الذاتي والذاتي الجزئي مع السلوك النظري لدالة الارتباط الذاتي (ACF) والذاتي الجزئي (PACF).

مرحلة التقدير (Estimation)

وهي المرحلة الثانية من مراحل تحليل السلاسل الزمنية ولكي يحقق النموذج الهدف الأساسي من بنائه وهو التنبؤ فيجب علينا أن نضمن جودة تقدير معالمه وهناك عدة طرق لتقدير معالم النموذج من أهمها:

أ. طريقة المربعات الصغرى الاعتيادية (Method of ordinary-Least squares).

ب. طريقة الإمكان الأعظم (Maximum likelihood method).

مرحلة التشخيص (Diagnostic)

ويتم في هذه المرحلة فحص النموذج وذلك للتأكد من أن مرحلة التقدير ومخرجاتها تتوافق مع الفروض النظرية الخاصة بالعملية العشوائية التي ولدت البيانات، أو على الأقل لا تظهر خلافاً في هذه الفروض وفيما يلي بعض الاختبارات والفحوص التي تجري على النموذج لاختبار مدى ملائمته وصلاحيته لتمثيل البيانات هي:

أ. اختبار معنوية معالم النموذج وذلك باستخدام اختبار (t-student) فإذا كانت غير معنوية لا بد من استبعاد إحدى رتب

(AR)، (MA).

ب. تحليل البواقي:

1. رسم البواقي: يجب أن تكون البواقي خالية من كل الانماط والتحركات المنتظمة وأن تتأرجح بتشتت ثابت حول الصفر وأن

يبدو شكلها عشوائياً.

2. تحليل دالة الارتباط الذاتي البواقي (e_t) ويتم ذلك بطريقتين:

الطريقة الأولى: وتعتمد هذه الطريقة على اختبار (Box-Ljung) وذلك لاختبار الفرضية الآتية:

$$H_0 : P_1 = P_2 = \dots = P_k = 0$$

$$H_1 : P_1 = P_2 = \dots = P_k \neq 0$$

معتمدين على الارتباط الذاتي للبواقي وإحصاء الاختبار

$$Q = n(n + 2) \sum_{k=1}^m \frac{\hat{P}_k^2}{n - k}$$

n : عدد المشاهدات.

m : أكبر إزاحة ل (k) وهي مساوية $\left(\frac{k}{4}\right)$.

ونقارن إحصاء الاختبار (Q) بالقيمة الجدولية α ، حيث X_{m-f}^2 عدد المعالم المقدرة فإذا كانت:

(1) $Q < X_{m-f}^2$ ، α تقبل فرض العدم، أي أن معاملات الارتباط الذاتي للأخطاء لها توزيع عشوائي وبذلك يكون النموذج ملائم وجيد.

(2) $Q > X_{m-f}^2$ ، α فهذا يعني أن النموذج غير ملائم ولا بد من إعادة مرحلة التعرف واختيار نموذج آخر.

الطريقة الثانية: وتعتمد هذه الطريقة على حدود فترة الثقة لمعاملات الارتباط الذاتي للبواقي المقدرة $(\hat{\epsilon}_t)$ والتي يجب أن تقع بين الحدين $(\pm Z_{(\alpha/2)})/\sqrt{n}$ فإذا تحقق هذا فهذا يدل على أن البواقي تتوزع عشوائياً، وإن النموذج ملائم وجيد للبيانات ويمكن استخدامه في التنبؤ وهناك عدة معايير أخرى للمفاضلة بين مجموعة من النماذج منها (معياري اكيك (AIC) ، معيار بيزيان (BIC)، متوسط مربعات الأخطاء (MSE).

مرحلة التنبؤ (Forecasting):

وهي المرحلة الأخيرة من مراحل تحليل السلاسل الزمنية وهي عادة ما تكون الهدف من الدراسة ولا يتم الانتقال إلى هذه المرحلة إلا بعد أن يتجاوز النموذج المبدئي كافة الفحوصات والاختبارات وإلا فيجب العودة إلى مرحلة التعرف وتكرار هذه العملية حتى الحصول على نموذج ذو كفاءة عالية.

2 الجانب التطبيقي:

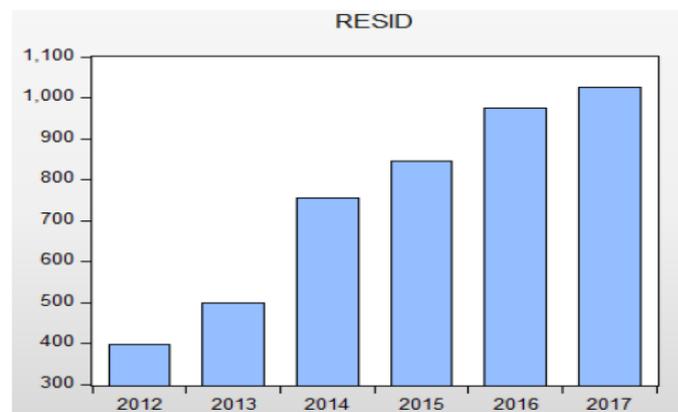
1.2 جمع البيانات:

جمعت هذه البيانات التي تتألف من (72) مشاهدة تمثل أعداد المصابين الشهرية بمرض السرطان في المعهد القومي لعلاج الأورام وذلك خلال الفترة من (2012/1/1) وحتى (2017/12/31) وتم تلخيص وتبويب هذه البيانات وحساب بعض المقاييس الإحصائية لما هو موضح بالجدول (1).

جدول (1) وصف أعداد المصابين بمرض السرطان في المعهد القومي لعلاج الأورام خلال الفترة من (2012-2017)

إجمالي المشاهدات	أقل قيمة	أكبر قيمة	المتوسط	الانحراف المعياري
72	27	118	62	22

وزيادة التوضيح تم تمثيل أعداد المصابين بالسرطان بالأعمدة البيانية كما هو موضح بالشكل رقم (1).



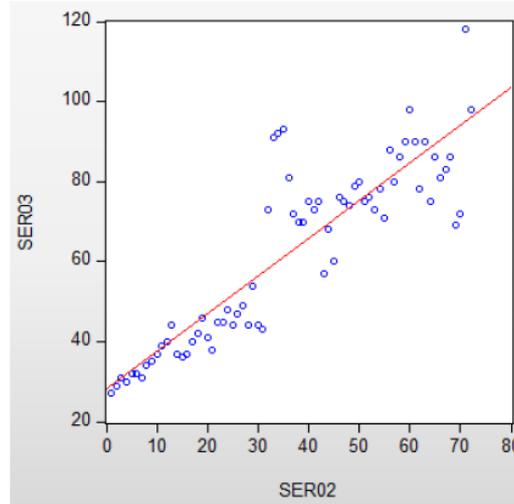
شكل رقم (1) التمثيل بالأعمدة البيانية لأعداد المصابين بمرض السرطان في معهد القومي لعلاج الأورام خلال الفترة (2012-2017)

نلاحظ أن من الشكل (1) أن المرض يزداد بزيادة الزمن حيث كان في سنة (2012) عدد المصابين (397) وأصبح في (2017) عدد المصابين (1026).

2.2 تحليل البيانات باستخدام الانحدار الخطي البسيط وأسلوب (بوكس – جنلر) لتحليل السلاسل الزمنية:

1.2.2 أولاً: تحليل البيانات باستخدام نموذج الانحدار الخطي البسيط:

يتم في البداية رسم شكل الانتشار بين المتغيرين (عدد المصابين، الزمن) لإعطاء صورة سريعة لطبيعة العلاقة بين المتغيرين ومدى قوتها واتجاهها كما هو موضح بالشكل (2).



الشكل (2): انتشار البيانات خلال الفترة (2012-2017)

من الشكل السابق نلاحظ أن بيانات أعداد المصابين تزداد بزيادة الزمن، أي أن هناك اتجاه عام متزايد ونستطيع تقريباً تمثيلها بخط مستقيم فنقول أن هناك علاقة خطية طردية.

وبهذا التقارب يكون شكل المعادلة صيغة الدالة الخطية كما يلي:

$$Y_i = a + B X_j$$

Y_i : المتغير التابع.

X_i : المتغير المستقل.

A , B معالم النموذج المجهولة، وقد تم تقدير معالم النموذج من بيانات الدراسة حسب الشكل (2) وبذلك أصبح شكل النموذج:

$$\hat{Y} = \hat{a} + \hat{B} x_i$$

أولاً: تقدير معالم نموذج الانحدار الخطي البسيط:

تم تقدير معالم نموذج الانحدار الخطي البسيط بناء على البيانات المتاحة وكان النموذج المقدر على الصورة:

$$\hat{Y} = 28.07942 + 0.9415 t_i$$

حيث $\hat{a} = 28.074$ تمثل متوسط الإصابات عند ثبات المتغيرات الأخرى خلال الفترة (2012-2017).

\hat{B} تعني عند التغير في فترة زمنية نتوقع زيادة مقدارها (0.9415) في عدد الإصابات وهذا يعني وجود علاقة طردية بين الإصابات والزمن.

2.2.2 الاختبارات الإحصائية لمعنى معالم النموذج:

عن طريق الخطأ المعياري:

أ- بالنسبة للمعلمة (\hat{a}) :

الخطأ المعياري ل (\hat{a}) يساوي (2.492746) وحيث أن نصف قيمة المعلمة $\frac{\hat{a}}{2} = \frac{28.07942}{2} = 14.03771$ أكبر من الخطأ

المعياري لها فإن هذا يدل على صغر الخطأ المعياري لتقدير المعلمة وبالتالي معنويتها.

ب- بالنسبة للمعلمة (\hat{B}) :

الخطأ المعياري ل (\hat{B}) يساوي (0.059343) ونصف قيمة المعلمة $\frac{\hat{B}}{2} = \frac{0.941057}{2} = 0.4707535$ وبما أن نصف قيمة

المعلمة أكبر من الخطأ المعياري لها فإن هذا يدل على صغر الخطأ المعياري لتقدير المعلمة وبالتالي فإن المعلمة معنوية إحصائياً.

اختبارات معنوية المعالم بواسطة اختبار (t):

أ- بالنسبة للمعلمة (\hat{a}) :

$H_0 : \hat{a} = 0$ (المعلمة لا تختلف عن الصفر)

$H_1 : \hat{a} \neq 0$ (المعلمة تختلف عن الصفر)

حيث بلغ قيمة (t = 11.26445) بقيمة احتمالية (0.000) وهي أقل من مستوى معنوي (0.05) ولذلك يتم قبول H_1 ورفض H_0 وبهذا المعلمة المعنوية.

ب- بالنسبة للمعلمة (\hat{B}) :

$H_0 : \hat{B} = 0$

$H_1 : \hat{B} \neq 0$

حيث بلغ قيمة (t = 15.86410) بقيمة احتمالية (0.000) وهي أقل من مستوى معنوي (0.05) وبذلك يتم قبول (H_1) ، ورفض (H_0) وبهذا نقر بوجود علاقة إحصائية بين الزمن وعدد الإصابات (علاقة طردية) وتكون المعلمة (\hat{B}) قادرة على تمثيل المجتمع تمثيلاً صحيحاً ويمكن الوثوق بها كمعلمة للتقدير في المجتمع.

3.2.2 اختبارات الجودة الإحصائية:

أ- اختبار النموذج ككل بواسطة إحصائية (F):

• لا توجد علاقة خطية بين عدد الإصابات والزمن $H_0 : (a, B) = 0$

• توجد علاقة خطية بين عدد الإصابات والزمن $H_1 : (a, B) \neq 0$

بما أن قيمة (F) بلغت (251.6698) وهي أكبر من قيمة (F) الجدولية وهي $F_{70,0,05} = 3.92$ وكذلك قيمتها الاحتمالية (0.000) وهي أقل من مستوى معنوية (0.05) لذلك نرفض H_0 ونقبل H_1 ونحكم بوجود علاقة ذات دلالة إحصائية بين عدد الإصابات والزمن وبذلك يمكن الحكم بأن النموذج ككل معنوي.

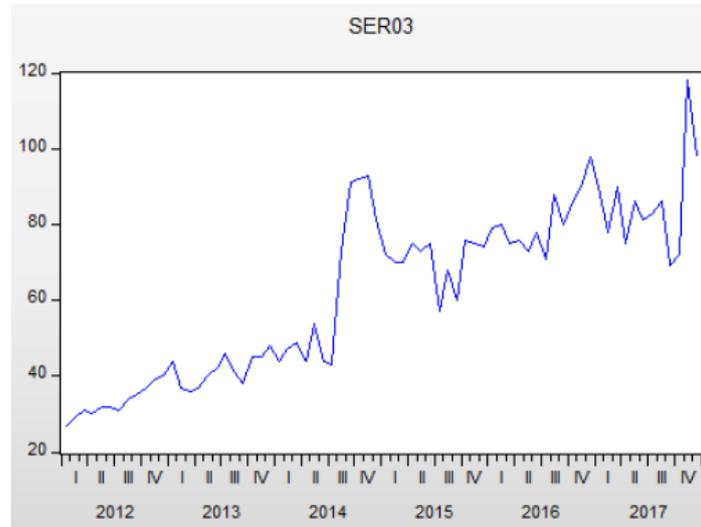
ب- اختبارات جودة التوفيق باستخدام معامل التحديد:

من خلال قيمة معامل التحديد ($R^2 = 0.78$) يظهر جودة توفيق عالية لنموذج الانحدار الخطي البسيط، وهذا يعني أن حوالي (0.78) من التغيرات في عدد الإصابات يمكن تفسيرها عن طريق التغيرات في الزمن، ولقد تم حساب معامل الارتباط من معامل التحديد $r = \pm\sqrt{R^2}$ وحيث أن إشارة ميل الانحدار موجبة بالتالي ($r = 0.884$) وهذا يعني وجود ارتباط طردي قوي بين الزمن وعدد الإصابات.

3.2 تحليل البيانات باستخدام السلاسل الزمنية:

1.3.2 مرحلة تهيئة البيانات:

يتم في هذه المرحلة رسم المنحنى البياني لسلسلة المصابين بمرض السرطان، وكذلك رسم حدود الثقة لدالة الارتباط الذاتي والذاتي الجزئي لمعرفة سلوك البيانات، فمن خلال الشكل (3) نلاحظ أن متوسط السلسلة غير ثابت بمرور الزمن، مما يدل على عدم استقرار السلسلة ولتأكيد ذلك نقوم باختبار معنوية معاملات الكلية لدالة الارتباط الذاتي باستخدام اختبار (Box-Jung) كما في الشكل (4).



شكل (3) المنحنى البياني لأعداد المصابين بمرض السرطان في المعهد القومي لعلاج الأورام خلال الفترة (2012-2017)

من الشكل (4) نلاحظ أن الإحصاءة ($Q = 455.06$) أكبر من القيمة الجدولية لـ ($X^2_{0.05,32} = 43.773$) وهذا ما يؤكد عدم استقرار السلسلة ولزيادة التأكيد قمنا باختبار ديكيير فولر الموسع (ADF) واختبار فيبس بيرون (P-P) والجدول رقم (2) يوضح بقية هذه الاختبارات.

Date: 01/28/24 Time: 20:58
 Sample: 2012M01 2017M12
 Included observations: 72

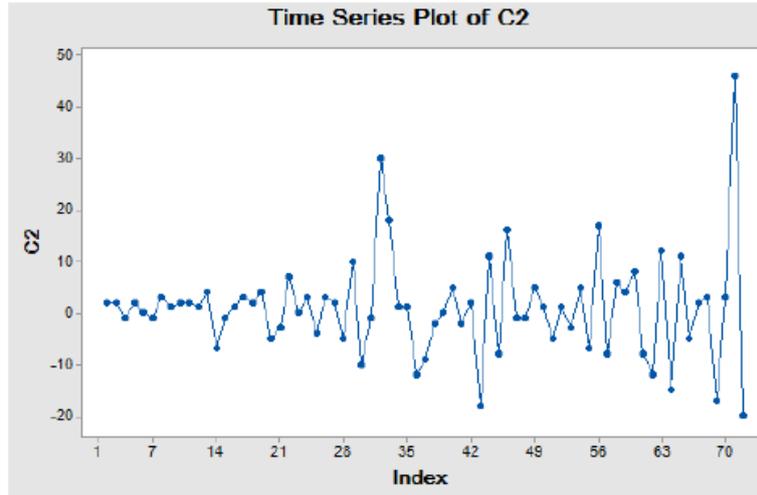
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.867	0.867	56.347	0.000
		2	0.764	0.053	100.78	0.000
		3	0.719	0.184	140.66	0.000
		4	0.678	0.046	176.72	0.000
		5	0.641	0.051	209.41	0.000
		6	0.614	0.053	239.84	0.000
		7	0.577	-0.019	267.15	0.000
		8	0.542	0.008	291.61	0.000
		9	0.504	-0.033	313.05	0.000
		10	0.474	0.016	332.33	0.000
		11	0.452	0.064	351.01	0.000
		12	0.439	-0.026	368.15	0.000
		13	0.402	-0.045	382.71	0.000
		14	0.373	-0.000	395.50	0.000
		15	0.337	-0.059	406.10	0.000
		16	0.287	-0.082	413.95	0.000
		17	0.247	-0.035	419.84	0.000
		18	0.208	-0.052	424.09	0.000
		19	0.187	0.045	427.61	0.000
		20	0.158	-0.046	430.15	0.000
		21	0.146	0.079	432.39	0.000
		22	0.125	-0.044	434.04	0.000
		23	0.096	-0.023	435.05	0.000
		24	0.073	-0.001	435.65	0.000
		25	0.039	-0.087	435.82	0.000
		26	-0.013	-0.110	435.83	0.000
		27	-0.049	-0.023	436.12	0.000
		28	-0.109	-0.164	437.56	0.000
		29	-0.136	0.071	439.87	0.000
		30	-0.162	-0.057	443.20	0.000
		31	-0.201	-0.048	448.47	0.000
		32	-0.222	0.032	455.06	0.000

شكل (4) دالة الارتباط الذاتي والذاتي الجزئي واختبار (Box-Ljung) لسلسلة المصابين بمرض السرطان

جدول رقم (2) نتائج اختبارات جذور الوحدة (ADF , P-P) لسلسلة المصابين بمرض السرطان

نوع الاختبار	فرع النموذج	القيمة المحسوبة	قيمة الجدولية عند مستوى معنوية 5%	القرار
اختبار ADF للسلسلة غير مستقرة: H_0	النموذج (1)	0.717732	-1.945525	نقبل H_0
اختبار ADF للسلسلة غير مستقرة: H_0	النموذج (2)	-1.883296	-2.902953	(السلسلة غير مستقرة)
اختبار (p-p) للسلسلة غير مستقرة: H_0	النموذج (1)	1.029924	-1.9454556	نقبل H_0
اختبار (p-p) للسلسلة غير مستقرة: H_0	النموذج (2)	-1.587581	-2.902953	(السلسلة غير مستقرة)

حيث يشير النموذج (1) إلى أن نموذج بدون ثابت، نموذج (2) إلى أنه نموذج بوجود ثابت نلاحظ أن اختبائي ديكي فولر الموسع وفليس بيرون أن القيمة المطلقة للإحصاء المحسوبة للنموذج الأول أصغر من القيمة الجدولية المطلقة عند مستوى معنوية 5% وبالتالي وجود جذر الوحدة، وعليه فإن السلسلة غير مستقرة ولمعالجة ذلك نقوم بأخذ الفرق الأول للسلسلة، فنحصل على سلسلة الفروقات من الدرجة الأولى لأعداد المصابين بمرض السرطان كما هو موضح بالشكل (5).



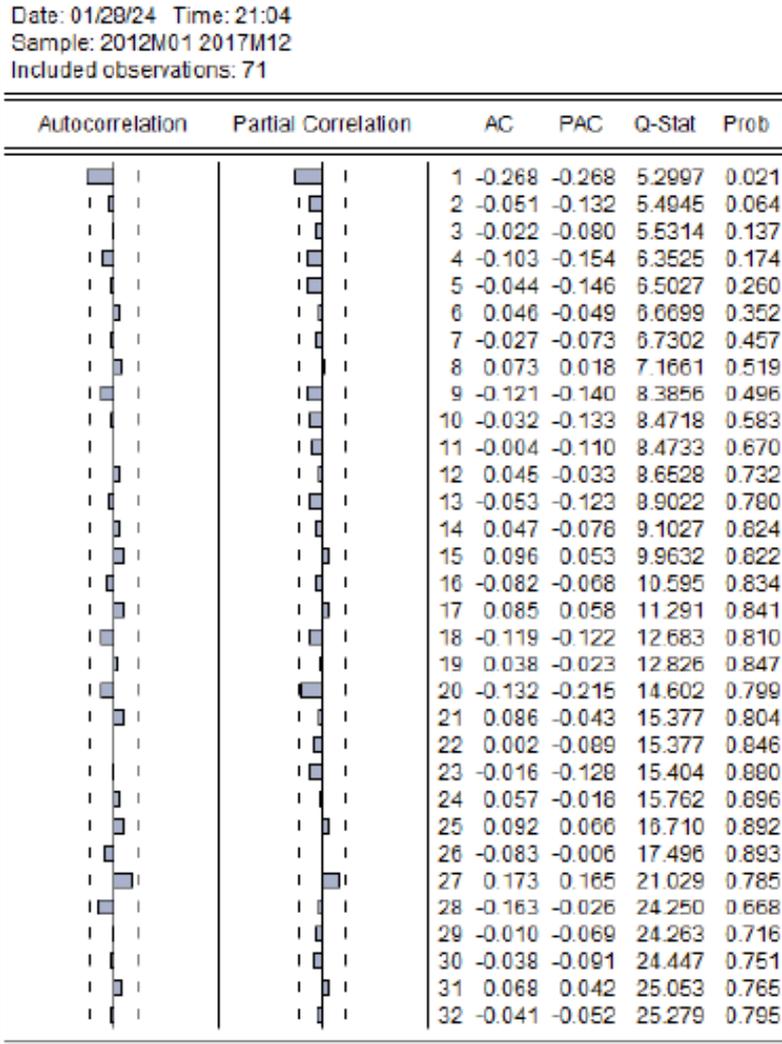
شكل (5) شكل المنحنى البياني لسلسلة الفروقات من الدرجة الأولى لأعداد المصابين بمرض السرطان

من خلال الشكل (5) أن المنحنى لسلسلة الفروقات لأعداد المصابين يتذبذب حول الصفر، وهذا يمكن أن يدل على أن السلسلة مستقرة ولتأكيد ذلك نعيد إجراء اختبارات (Box-Ljung) واختبارات جذر الوحدة كما هو موضح بالجدول رقم (3) والشكل (6).

جدول رقم (3) نتائج اختبارات جذر الوحدة (ADF, P-P) لسلسلة الفروقات من الدرجة الأولى لأعداد المصابين بمرض السرطان

القرار	قيمة الجدولية عند مستوى معنوية 5%	القيمة المحسوبة	فرع النموذج	نوع الاختبار
نرفض H_0	-1.945525	-10.6593	النموذج (1)	اختبار ADF للسلسلة
(السلسلة مستقرة)	-2.902953	-10.74700	النموذج (2)	غير مستقرة: $H_0 = 0$
نرفض H_0	-1.945525	-11.03752	النموذج (1)	اختبار (p-p) للسلسلة
(السلسلة مستقرة)	-2.903566	-11.89975	النموذج (2)	غير مستقرة: H_0

من خلال الجدول (3) نلاحظ أن القيمة المطلقة للإحصاء المطلوبة للنموذجين أكبر من القيمة المطلقة الجدولية، وبالتالي السلسلة ليس لها جذر وحدة وهي مستقرة ومن الشكل رقم (6) نلاحظ أن دالة الارتباط الذاتي والذاتي الجزئي أن معظم معاملات الارتباطات الذاتية والذاتية الجزئية داخل حدود فترة الثقة وأنها غير معنوية بأعداد الإزاحة الأولى هذا ما أكدته اختبار (Box-Ljung) حيث أن $(X^2_{0.05,32} = 43.773 > Q = 25.297)$ وعليه فإن السلسلة مستقرة تقريباً.



شكل رقم (6) دالة الارتباط الذاتي والذاتي الجزئي واختبار (Box-Ljung) لسلسلة الفروقات لأعداد المصابين بمرض السرطان

2.3.2 مرحلة التعرف:

بالنظر إلى دالتي الارتباط الذاتي والذاتي الجزئي ومطابقة قيم معاملات الارتباط الذاتي و الذاتي الجزئي بعد أخذ الفرق الأول نستنتج أن النموذج الملائم هو نموذج $ARIMA(1,1,1)$ لزيادة التأكيد تم اقتراح نموذجين $ARIMA(0,1,1)$ و $ARIMA(1,1,0)$.

جدول رقم (4) نتائج تطبيق مجموعة من المعايير على النماذج المقترحة لتحديد الأفضل من بينها

Akaike	MAPE	RMSE	R-Sugared	النموذج
7.61032	8.743	9.484	0.82	ARIMA(1,1,0)
7.58132	8.807	9.404	0.823	ARIMA(0,1,1)
7.4614	8.452	8.828	0.846	ARIMA(1,1,1)

من خلال الجدول رقم (4) نلاحظ أن النموذج $ARIMA(1,1,1)$ هو النموذج الأفضل من بين النماذج المقترحة حيث له أُل قيمة وفق كل معيار وكذلك أكبر قيمة وفق $(R - sugared = 0.846)$ وبالتالي فهو النموذج الأفضل. بعد التعرف على النموذج الأفضل (الأكثر ملائمة) للبيانات تقوم بتقدير معالمه كما هو موضح بالجدول (5).

جدول رقم (5) تقدير معالم النموذج $ARIMA(1,1,1)$

Type	Coef	SE coef	T-Value	P-Value
AR	0.966961	0.049361	19.58976	0.000
MA	0.282545	0.100325	2.816285	0.000
Constant	63.07631	17.83565	3.536529	0.007

3.3.2 مرحلة التشخيص:

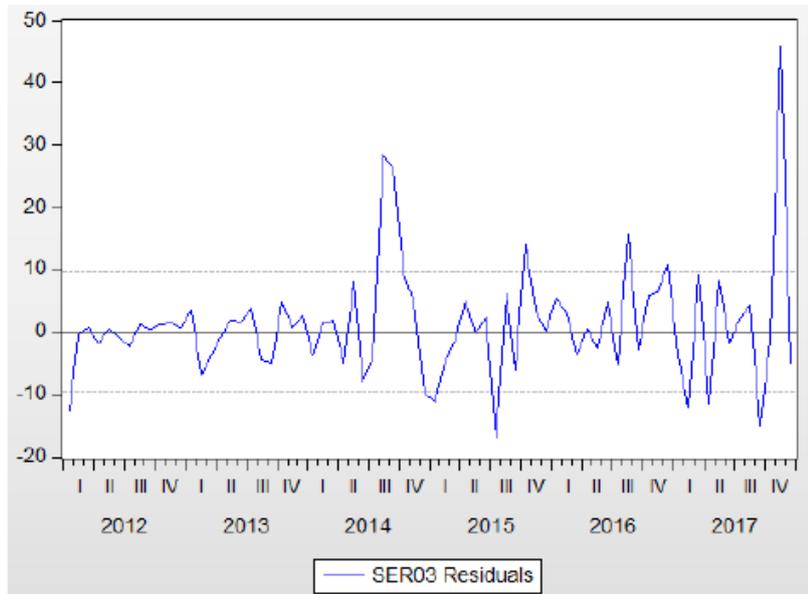
لكي يكون النموذج جيد وملائم للبيانات لابد من إجراء بعض الاختبارات عليه كما يلي:

أ- اختبار معنوية معامل النموذج:

نلاحظ من الجدول (5) أن قيمة (P-Value) لكل المعالم تساوي الصفر مما يؤكد معنوية تلك المعالم.

2- تحليل البواقي

أ- رسم البواقي:

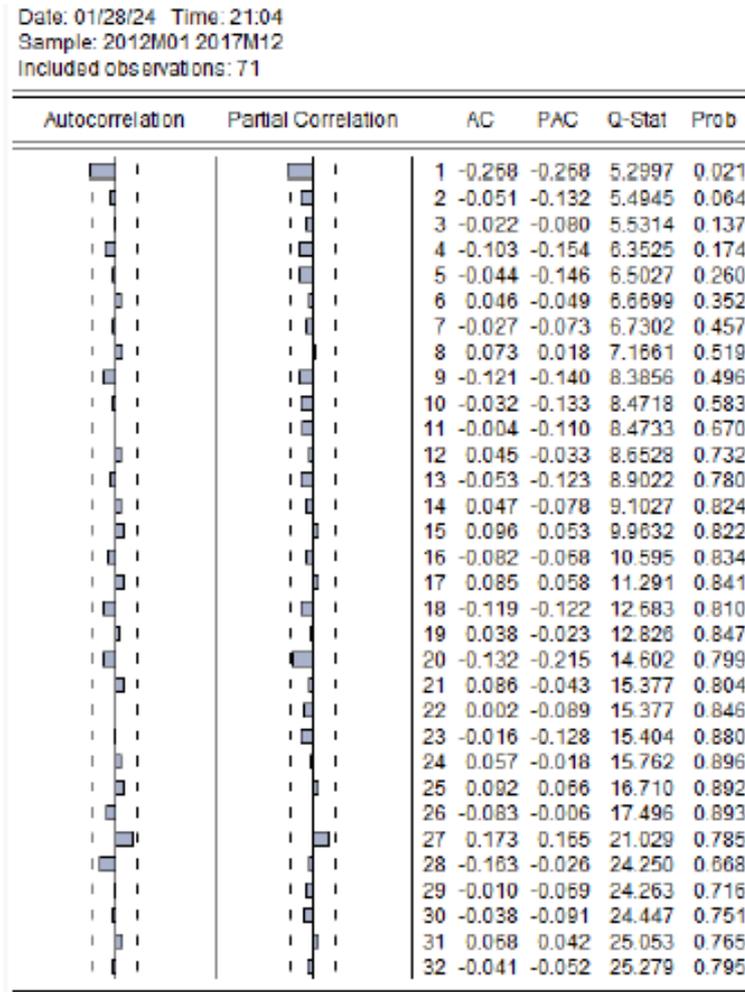


شكل (7) بواقي النموذج $ARIMA(1,1,1)$

من خلال الشكل (7) نلاحظ أن البواقي بتاريخ بشكل باق جدول الصفر وهذا دليل على أنها عشوائية ومستقرة

ب- تحليل الارتباط الذاتي للبواقي:

لاختيار عشوائية سلسلة البواقي تم استخراج معاملات الارتباط الذاتي والذاتي الجزئي للبواقي المقدر كما موضح بالشكل (8).



شكل (8) يوضح معاملات الارتباط الذاتي والارتباط الذاتي الجزئي

نلاحظ من الشكل (8) أن سلسلة البواقي مستقرة، حيث أن معاملات الارتباط الذاتي والذاتي الجزئي للبواقي تقع داخل حدود فترة الثقة، والفرض التأكيد من ملائمة النموذج تم استخدام إحصاءة (Bok-Ljung) حيث كانت (Q=18.477) وهي أصغر من القيمة الجدولية لـ $(X^2_{0.05,32} = 43.773)$ وهذا يدل على عشوائية البواقي (white noise) وعليه فإن النموذج ARIMA(1,1,1) ملائم ويمكن استخدامه في التنبؤ.

4.3.2 اختبارات جودة التوفيق لنموذج السلاسل الزمنية:

أ- بواسطة إحصائية (F):

حيث بلغت قيمة (f=102.766) وقيمتها الاحتمالية (0.000) وهي أقل من (0.05) وبالتالي فإن النموذج جيد ككل.

ب- باستخدام معامل التحديد:

يظهر معامل التحديد جودة وتوفيق عالية لنموذج السلسلة الزمنية $ARIMA(1,1,1)$ حيث بلغت قيمته (0.84) وهذا يعني أن (0.84) من المتغيرات في عدد الإصابات يمكن تفسيرها عن طريق التغيرات في الزمن .

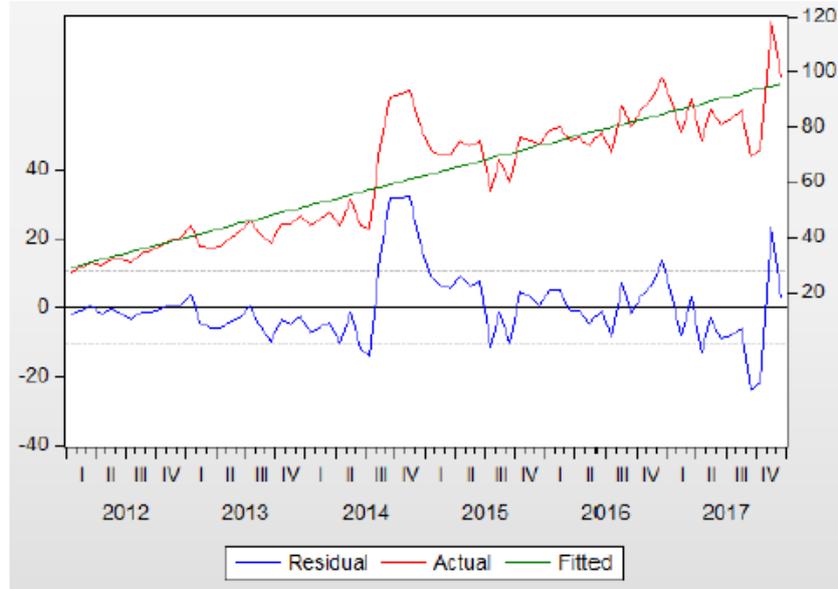
المفاضلة بين نموذج الانحدار الخطي البسيط والنموذج $ARIMA(1,1,1)$:

قبل الحكم على أي من النموذجين أفضل في عملية التنبؤ بعدد الإصابات لمرض السرطان في المستقبل يجب التطرق للمفاضلة بينهما عن طريق الاختبارات الإحصائية ونتائج اختبارات الجودة وكذلك بعض المعايير الأخرى لكي نموذج كما هو موضح بالجدول (6).

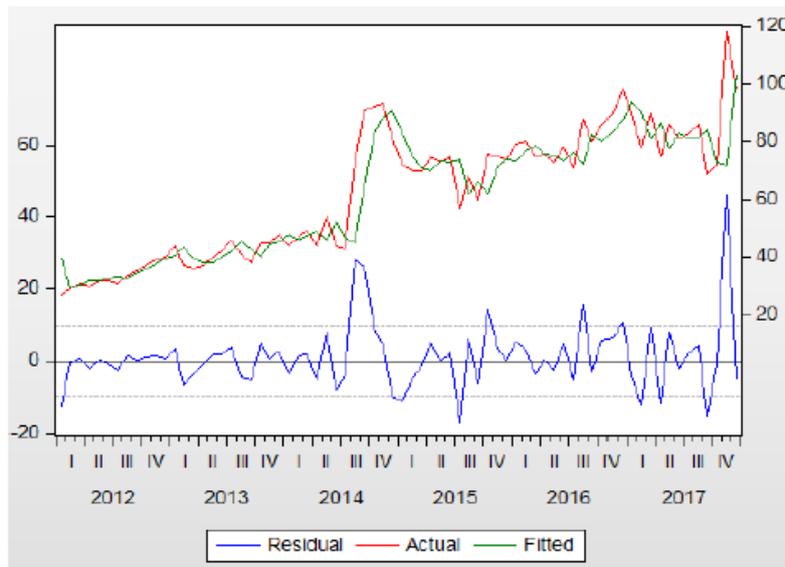
جدول رقم (6) ملخص لمختلف نتائج الاختبارات الإحصائية واختبارات جودة التوفيق

الاختبارات	نموذج الانحدار الخطي البسيط	نموذج السلسلة الزمنية $ARIMA(1,1,1)$
1- معنوية المعالم	$\hat{Y} = 28.079 + 0.9415(t_i)$	$z_t = 63.076 + 0.96677z_{t-1} + \epsilon_t - 0.282z_{t-2}$
2- اختبارات الجودة	المعالم معنوية	المعالم معنوية
أ- بولسطة R^2	$R^2 = 0.78$	$R^2 = 0.84$
ب- بولسطة (F)	معنوية النموذج كل	معنوية النموذج ككل
3- MSE	109.534	8.937
4- MAD	7.172	5.659
5- MAPE	11.062	8.452
6- معيار AKIK(AIC)	7.561	7.461467
7- (BIC)	7.624737	7.587947
8- Hamma Quine	7.586678	7.511818

وبعد المعايير الأخرى لكلا النموذجين من خلال الجدول (6) نلاحظ أن مختلف الاختبارات الإحصائية تشير إلى وجود معنوية إحصائية لكلا النموذجين وكذلك جودة التوفيق لكل النموذجين إلا أن نموذج السلسلة الزمنية يفسر التغير الحادث في عدد الإصابات أكثر من نموذج الانحدار الخطي البسيط حيث كان معامل التجديد لنموذج السلاسل الزمنية (0.84) وهو أكبر من معامل التحديد لنموذج الانحدار الخطي البسيط (0.78). وأئن النموذج $ARIMA(1,1,1)$ يتمتع بأقل قيمة لكل معيار من معايير المفاضلة عن نموذج الانحدار الخطي البسيط مما يجعلنا نفضل استخدام نموذج $ARIMA(1,1,1)$ على نموذج الانحدار الخطي البسيط في التنبؤ بعدد الإصابات المستقبلية لمرضى السرطان وهذا ما يؤكد التنبؤ داخل العينة حيث أظهر نموذج السلاسل الزمنية أن البيانات التقديرية قريبة جداً من البيانات الحقيقية وهذا دليل على أن النموذج أخذ في الاعتبار التغيرات (التقلبات) التي حدثت في الماضي للسلسلة على العكس من نموذج الانحدار الخطي البسيط حيث كانت البيانات التقديرية على هيئة خط مستقيم وبالتالي لم يأخذ التغيرات (التقلبات) الحادثة في الماضي والشكلين (8)، و (9) وكذلك الشكلين (10) و (11) يوضحان التنبؤ داخل العينة للنموذجين.



شكل (9) التنبؤ داخل العينة لنموذج باستخدام انحدار الخطي البسيط خلال الفترة (2012-2014)



شكل (10) التنبؤ داخل العينة باستخدام نموذج $ARIMA(1,1,1)$ الانحدار الخطي البسيط خلال الفترة (2012-2014)

3 النتائج والتوصيات:

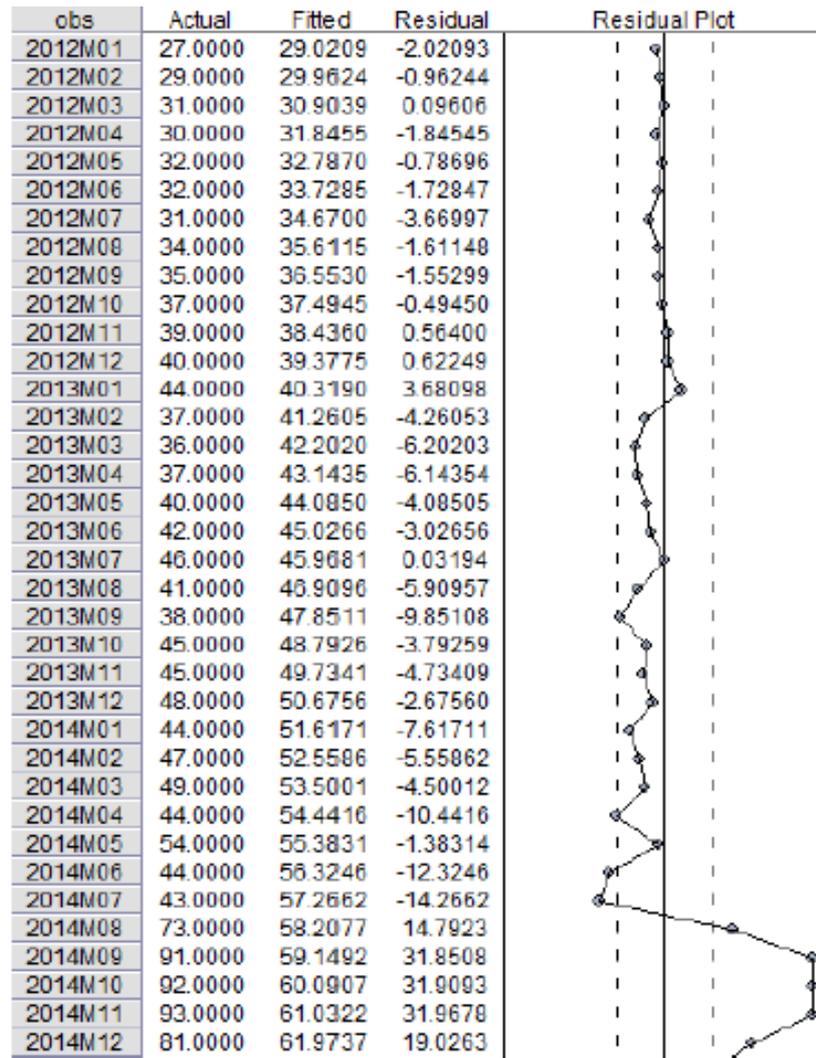
النتائج:

1. سلسلة أعداد المصابين بمرض السرطان في المعهد القومي لعلاج الأورام غير مستقرة في المتوسط وأن هناك اتجاه عام إلى الأعلى وذلك خلال الفترة (2012-2017) وهذا يعني أن انتشار المرض يزداد بمرور الزمن ولقد تم تحويلها إلى سلسلة مستقرة بعد أخذ الفرق الأول.

obs	obs	Actual	Fitted	Residual	Residual Plot
2012M01	2012M01	27.0000	39.5922	-12.5922	
2012M02	2012M02	29.0000	29.4193	-0.41932	
2012M03	2012M03	31.0000	30.2424	0.75765	
2012M04	2012M04	30.0000	31.8459	-1.84587	
2012M05	2012M05	32.0000	31.6142	0.38575	
2012M06	2012M06	32.0000	32.9177	-0.91774	
2012M07	2012M07	31.0000	33.2860	-2.28603	
2012M08	2012M08	34.0000	32.7057	1.29433	
2012M09	2012M09	35.0000	34.5949	0.40506	
2012M10	2012M10	37.0000	35.8132	1.18683	
2012M11	2012M11	39.0000	37.5262	1.47380	
2012M12	2012M12	40.0000	39.3790	0.62096	
2013M01	2013M01	44.0000	40.5870	3.41303	
2013M02	2013M02	37.0000	43.6659	-6.66593	
2013M03	2013M03	36.0000	39.7450	-3.74495	
2013M04	2013M04	37.0000	37.9527	-0.95269	
2013M05	2013M05	40.0000	38.1307	1.86929	
2013M06	2013M06	42.0000	40.2343	1.76574	
2013M07	2013M07	46.0000	42.1974	3.80256	
2013M08	2013M08	41.0000	45.4898	-4.48979	
2013M09	2013M09	38.0000	42.9979	-4.99794	
2013M10	2013M10	45.0000	40.2406	4.75936	
2013M11	2013M11	45.0000	44.2525	0.74751	
2013M12	2013M12	48.0000	45.3860	2.61398	
2014M01	2014M01	44.0000	47.7595	-3.75954	
2014M02	2014M02	47.0000	45.6925	1.30750	
2014M03	2014M03	49.0000	47.1617	1.83828	
2014M04	2014M04	44.0000	48.9457	-4.94567	
2014M05	2014M05	54.0000	46.0276	7.97237	
2014M06	2014M06	44.0000	52.0473	-8.04732	
2014M07	2014M07	43.0000	46.9040	-3.90399	
2014M08	2014M08	73.0000	44.7664	28.2336	
2014M09	2014M09	91.0000	64.6949	26.3051	
2014M10	2014M10	92.0000	82.6451	9.35494	
2014M11	2014M11	93.0000	88.4012	4.59879	
2014M12	2014M12	81.0000	90.7120	-9.71199	

جدول (11) التنبؤ داخل العينة باستخدام نموذج $ARIMA(1,1,1)$ الانحدار الخطي البسيط خلال الفترة (2012-2014)

- النموذج الملائم للتنبؤ بعدد الإصابات بمرض السرطان هو النموذج $ARIMA(1,1,1)$ وهو أفضل من نموذج الانحدار الخطي البسيط وذلك وفقاً للاختبارات الإحصائية ومعايير المفاضلة.
- نظراً لوجود مجموعة من المتغيرات المؤثرة في عدد الإصابات بمرض السرطان ولكن صعوبة تحديدها والضبط العددي لبعض منها وكذلك صعوبة الحصول على معلومات تاريخية حولها الأمر الذي أدى إلى نقص كفاءة نموذج الانحدار الخطي البسيط الذي تم بناءه مقارنة بنموذج السلسلة الزمنية.
- عند غياب العلاقات السببية بين المتغيرات أو عدم توفر المعلومات الكافية حول المتغيرات التفسيرية فإن أسلوب السلاسل الزمنية يكون الأكثر دقة وجودة في عملية التنبؤ.
- النموذج الذي تم التوصل إليه من بين نماذج السلاسل الزمنية جيد ويمكن الاعتماد عليه في عملية التنبؤ، حيث أظهرت القيم التنبؤية داخل العينة تقارب كبير مع القيم الحقيقية.



شكل (12) التنبؤ داخل العينة لنموذج باستخدام انحدار الخطي البسيط خلال الفترة (2012-2014)

التوصيات:

1. الأخذ بنتائج هذا البحث الذي يظهر تزايد في أعداد المصابين بمرض السرطان بمرور الزمن مما يقتضي اتخاذ التدابير اللازمة من قبل الجهات المختصة للحد من هذه الظاهرة.
2. نوصي بدراسة مستقبلية إن أمكن يكون فيها إدخال بعض المتغيرات إلى جانب الزمن لتصبح العلاقة متعددة مع عدد الإصابات.
3. الاهتمام بقسم الإحصاء بالمعهد القومي لعلاج الأورام وتوفير الإمكانيات اللازمة وزيادة تفعيل دورة في تسجيل وترصيد وتوثيق كل البيانات الخاصة بالمرضى مثل (الجنس، العمر، الشكل، نوع الورم، ... إلخ)، حتى تتاح الفرصة أمام الباحثين من دراسة تأثير هذه المتغيرات في أعداد المصابين في المستقبل.

قائمة المراجع:

- [1] إسماعيل محمد عبد الرحمن. تحليل الانحدار الخطي. الرياض، المملكة العربية السعودية: مكتبة الملك فهد الوطنية؛ 2001.
- [2] بريء عدنان ماجد عبد الرحمن. طرق التنبؤ الإحصائي. الطبعة الأولى. المملكة العربية السعودية: مطابع الملك سعود؛ 2002.
- [3] عطية عبد القادر محمد محمد عبد القادر. الحديث في الاقتصاد القياسي بين النظرية والتطبيق. الإسكندرية، مصر: الدار الجامعية؛ 2005.
- [4] شعراوي سمير مصطفى. مقدمة في التحليل الحديث للسلاسل الزمنية. الطبعة الأولى. المملكة العربية السعودية: مركز النشر العلمي؛ 2005.
- [5] الخضيرى مجد قدوري. دراسة مقارنة لطرق التقدير والتنبؤ لبعض نماذج بوكس- جنكز الموسمية [رسالة ماجستير]. بغداد، العراق: كلية الإدارة والاقتصاد، جامعة بغداد؛ 1996.
- [6] 1. Box G, Pierce D. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc. 1970;65(332):1509-26.
- [7] 2. Montgomery DC, Contreras LE. A note on forecasting with adaptive filtering. J Oper Res Soc. 1977;28(1):87-91.
- [8] 3. Jenkins GM, Box GE. Time series analysis: forecasting and control. (No Title). 1976.