

## استخدام أسلوب بوكس- جنكز في تحليل السلاسل الزمنية للتنبؤ بعدد الإصابات بمرض سرطان الرئة في المعهد القومي لعلاج الأورام مصراتة خلال الفترة (2012-2021)"

امحمد ميلاد إبراهيم بعوه  
جامعة مصراتة، قسم الإحصاء، ليبيا  
[Emhemmedmilad2017@gmail.com](mailto:Emhemmedmilad2017@gmail.com)

### المخلص

تضمن أهمية تحليل السلاسل الزمنية في الحصول على وصف دقيق للسلسلة الزمنية وبناء نموذج مناسب للتنبؤ المستقبلي، وذلك لاتخاذ القرارات الصحيحة في كافة المجالات. ويهدف هذا البحث إلى تحليل السلاسل الزمنية باستخدام طريقة (Box- Jenkins) لإيجاد أفضل وأكفأ نموذج توصّل إليه الباحث للتنبؤ بعدد مرضى سرطان الرئة في المعهد القومي لعلاج الأورام بالاعتماد على البيانات الشهرية للفترة (2012-2021). وقد أظهرت سلسلة مرضى سرطان الرئة بأنها غير مستقرة ويوجد بها اتجاه عام وتم تحويلها إلى سلسلة مستقرة بعد أخذ اللوغاريتم لبيانات السلسلة وذلك لتثبيت التباين ومن ثم أخذ الفرق الأول وذلك لتثبيت المتوسط، وأن النموذج الملائم لها هو نموذج المتوسطات المتحركة المتكامل من الدرجة الأولى ARIMA(0,1,1)، وبالاعتماد على هذا النموذج تم التنبؤ بأعداد مرضى سرطان الرئة شهرياً لسنتين قادمتين (2022، 2023)، وكانت القيم التنبؤية متناسقة في الاتجاه مع قيم السلسلة الأصلية مما يدل على كفاءة النموذج.

استلمت الورقة بتاريخ  
2025/02/28، وقبلت  
بتاريخ  
2025/03/15  
ونشرت  
بتاريخ  
2025/03/18

### الكلمات المفتاحية:

السلاسل الزمنية، منهجية  
بوكس- جنكز، الأورام  
السرطانية.

### 1. المقدمة Introduction

لقد أصبح الاهتمام كبيراً باستخدام طرق القياس والأساليب الإحصائية في البحوث والدراسات، وذلك لغرض تحديد الاتجاهات العامة للظواهر وتحليل العلاقات المتشابكة بينها على أساس موضوعي غير متحيز. وتعتبر السلاسل الزمنية من أهم الأساليب الإحصائية الحديثة التي تدرس التغيرات التي تطرأ على قيم الظاهرة خلال الزمن وتحديد الأسباب والنتائج وتفسير العلاقات المشاهدة بينها والتنبؤ بما سيحدث من تغير على قيم الظاهرة في المستقبل على ضوء ما حدث لها في الماضي. ونتيجة للظروف التي مرت بها بلادنا خلال السنوات الأخيرة من حروب أدت إلى تلوث الهواء والماء، وكذلك المواد الكيميائية والإشعاعية الناتجة من مخلفات المصانع والآلات والسيارات وانتشار ظواهر التدخين، الأمر الذي أدى إلى التأثير على جوانب الحياة الخاصة بالإنسان، حيث أن من أساسيات الصحة هي الحد من الإصابة بالأمراض خاصة المستعصية منها التي تؤدي إلى نسب عالية من الوفيات. وقد اعتمدت هذه الدراسة على البيانات الشهرية لأعداد الإصابات بسرطان الرئة خلال الفترة (2012-2021) كسلسلة زمنية وذلك لغرض تحليلها والوصول إلى أفضل نموذج للتنبؤ بأعداد الإصابات بمرض سرطان الرئة لفترات لاحقة لاتخاذ التدابير اللازمة ووضع الخطط للحد من هذه الظاهرة مستقبلاً.

### 2. أهمية وأهداف البحث

يعد التنبؤ من الموضوعات التي لها أهمية كبيرة والتي من خلالها يتمكن المسؤولون وأصحاب القرار من اتخاذ القرارات الصحيحة في كافة المجالات، ويهدف هذا البحث إلى تحليل السلاسل الزمنية باستخدام طريقة بوكس- جنكز، لتحديد أفضل وأكفأ نموذج إحصائي للتنبؤ بعدد مرضى سرطان الرئة في المعهد القومي لعلاج الأورام مصراتة بالاعتماد على البيانات الشهرية خلال الفترة (2012-2021). وذلك لغرض التنبؤ بعدد مرضى سرطان الرئة لفترات لاحقة حتى يتمكن المسؤولين ومتخذي القرار من توفير الإمكانيات اللازمة واتخاذ التدابير اللازمة ووضع الخطط لاستيعاب الأعداد المتزايدة لمرضى سرطان الرئة في المستقبل.

### 3. الدراسات السابقة

- رسالة ماجستير بعنوان: "تطبيق بوكس- جنكز للسلاسل الزمنية على إصابات مرضى سرطان السودان في الفترة (يناير 2002- ديسمبر 2004)" حيث هدفت الدراسة إلى إيجاد نموذج قياسي مقترح يحدد العلاقة بين متغيرات الدراسة ومعرفة أهميتها ومدى تأثيرها على إصابات السرطان في السودان، وتشمل هذه المتغيرات (نوع المصاب، عمره، القبيلة التي ينتمي إليها). [6]

توصلت هذه الدراسة إلى أن:

1. ينتشر مرض السرطان في الإناث أكثر من الذكور.
2. تعتبر قبيلة الجعلين أكثر القبائل السودانية إصابة بمرض السرطان.
3. النموذج الملائم لإصابات مرض السرطان بالسودان خلال فترة الدراسة هو نموذج المتوسطات المتحركة من الرتبة الأولى.

- بحث بعنوان: "تحليل السلاسل الزمنية للتنبؤ بأعداد المصابات بمرض سرطان الثدي في محافظة كربلاء بالعراق في الفترة (من 2012- 2016): حيث هدفت الدراسة لمعرفة طبيعة التغيرات التي تطرأ على الظاهرة خلال هذه الفترة وتحليلها للوصول إلى أفضل نموذج إحصائي للتنبؤ بأعداد النساء المصابات بهذا المرض لفترات قادمة. توصلت هذه الدراسة إلى:

1. سلسلة أعداد المصابات بمرض سرطان الثدي في كربلاء غير مستقرة في المتوسط، وأن هناك اتجاها عاما متزايدا مع مرور الزمن.

2. بعد تحقيق الاستقرار في السلسلة ومطابقة معاملات الارتباط الذاتي والذاتي الجزئي للسلسلة مع السلوك النظري للدالتين وجد أن النموذج الملائم للبيانات هو نموذج الانحدار الذاتي من الرتبة الأولى. [7] من خلال الدراسات السابقة يمكن استنتاج أن مرض السرطان يزداد انتشاراً بمرور الزمن، وأن السلسلة الزمنية للمرضى كانت غير مستقرة في المتوسط وبها اتجاه عام متزايد وهذا ما أكدته هذه الدراسة أيضاً، حيث كانت سلسلة أعداد مرضى سرطان الرئة تتزايد بمرور الزمن وتم تثبيتها بأخذ الفروق الأولى.

#### 4. الجانب النظري

يتناول هذا البند استعراضاً لبعض المفاهيم العامة، وعرض مراحل بناء نموذج السلاسل الزمنية، واعتماد تحليل السلاسل الزمنية على أسلوب (بوكس-جنكيز).

### أ. مفاهيم رئيسية (Fundamental Concepts)

#### - السلسلة الزمنية (Time series)

هي مجموعة من المشاهدات أو القياسات لقيم ظاهرة معينة (اقتصادية، اجتماعية، طبية،... إلخ) مأخوذة على فترات زمنية محددة (يوم، أسبوع، شهر،... إلخ)، وقد تكون هذه الفترات متساوية أو غير متساوية وغالباً ما تكون متساوية، ويمكن تمثيل السلسلة الإحصائية  $Z_t = f(t) + e_t$  ،  $t = 0, \pm 1, \pm 2, \dots$  ،  $f(t)$ : تمثل الجزء المنتظم الذي يعبر عنه بدالة رياضية.

$e_t$ : الجزء العشوائي وقد يسمى بالضجيج (التشويش). [3]

#### - السكون (الاستقرار): (Stationarity)

يقال أن السلسلة الزمنية ساكنة (مستقرة) إذا كانت الخصائص الإحصائية لهذه السلسلة عبر الزمن لا تتغير بالإزاحة إلى الأمام أو للخلف بعدد من الوحدات الزمنية، والخصائص الإحصائية يمكن وضعها بشكل كامل ومؤكد عن طريق دالة التوزيع التراكمي، ويمكن وضعها بشكل جزئي عن طريق بعض المؤشرات الإحصائية كالتوقع، التباين، التغاير، ومن هنا يمكن أن نستنتج نوعين من السكون هما:

#### 1. السكون التام (المؤكد): (Strictly Stationarity)

يقال أن السلسلة الزمنية مؤكدة السكون أو تامة السكون إذا كان التوزيع الاحتمالي المشترك لأي مجموعة جزئية من المتغيرات التي تتكون منها السلسلة لا تتأثر بالإزاحة إلى الأمام أو للخلف بعدد من الوحدات الزمنية، وهذا يؤدي إلى ثبات التوقع والتباين للعملية العشوائية، ويكون التغاير بين أي متغيرين يعتمد فقط على الفجوة الزمنية بين هذين المتغيرين.

#### 2. السكون الضعيف: (Weak Stationarity)

يقال أن السلسلة الزمنية ساكنة (مستقرة) سكوناً ضعيفاً إذا كان العزم من الرتبة الثانية موجود وتحقق الشروط الآتية:

1. توقع العملية العشوائية لا يعتمد على الزمن

$$E(z_t) = \mu \quad , \quad t = 0, \pm 1, \pm 2, \dots$$

2. تباين العملية العشوائية  $\sigma^2$  لا يعتمد على الزمن

$$V(z_t) = \sigma^2 \quad , \quad t = 0, \pm 1, \pm 2, \dots$$

3. التغاير بين أي متغيرين في السلسلة يعتمد على الفجوة الزمنية بين هذين المتغيرين وليس على الزمن الذي حُسب

[3]

التغاير

عنده

$$. COV(z_{t-k}, z_t) = rk \quad , \quad t = \pm 1, \pm 2, \dots$$

### - الارتباط الذاتي : Autocorrelation(AC)

هو عبارة عن مؤشر يوضح درجة العلاقة بين قيم المتغير نفسه عند فترات إزاحة (k) مختلفة، وتتراوح قيمة دالة الارتباط الذاتي بين [-1,1] أي أن:  $(-1 < r_k < 1)$  ويرمز له بالرمز  $\hat{r}_k$  ويقدر حسب الصيغة التالية [9]  $\hat{r}_k =$

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

حيث أن :

$\hat{r}_k$ : تقيس الارتباط بين المشاهدات المتعاقبة للسلسلة.

$z_t$ : هي قيم مشاهدات السلسلة.

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$$
 ويحسب الحسابي، ويمثل المتوسط الحسابي،

وكذلك التوزيع الاحتمالي لمعاملات الارتباط الذاتي هو تقريباً توزيع طبيعي بمتوسط صفري وتباين  $(\frac{1}{n})$ ، أي أن:

$$\hat{r}_k \sim N(0, \frac{1}{n}) \text{ حيث } (k=1,2,3,\dots), n \text{ عدد المشاهدات. [5]}$$

يمكن حساب معامل الارتباط الذاتي لكل فجوة من الفجوات الزمنية  $k=0, \pm 1, \pm 2, \dots$  فبنشأ لدينا علاقة دالية بين معاملات الارتباط الذاتي  $\hat{r}_k$  والفجوة الزمنية (k) تسمى بدالة الارتباط الذاتي للعملية الساكنة  $z_t$ .

يطلق على الرسم البياني لمعاملات  $r_k$  مقابل فترات الإزاحة (k) حيث  $(k=1,2,3,\dots)$  بدالة الارتباط الذاتي ونرمز لها بالرمز (Autocorrelation Function (ACF)).

وتستخدم دالة الارتباط الذاتي في تحليل السلاسل الزمنية لأنها تعطي معلومات حول سلوك السلسلة ومكوناتها الأساسية، حيث تستخدم في تحديد استقرارية السلسلة. فهي تميل بسرعة نحو الصفر مع ازدياد فترات الإزاحة (k)، أو تنقطع بعد عدد معين من فترات الإزاحة (k=q) في حالة السلسلة مستقرة.

ولا تتحدر بسرعة نحو الصفر ولا تنقطع في حالة السلسلة غير مستقرة بسبب وجود اتجاه نازل أو صاعد في المعدل، كما تستخدم دالة الارتباط الذاتي للبواقي (Residuals Autocorrelation Function (RACF)) لفحص ملائمة النموذج عن طريق اختبار عشوائية أخطاء التنبؤ.

### - الارتباط الذاتي الجزئي : Partial Autocorrelation (PAC)

هو مؤشر يقيس درجة الارتباط الخطي بين متغيرين يبعدان عن بعضهما البعض وحدتان زمنيان بعد حذف تأثير المتغير الذي يقع بينهما، ويحسب معامل الارتباط الذاتي الجزئي من العلاقة التالية: [4]

$$\hat{\phi}_{kk}, \dots, k=2,3, \dots = \frac{\hat{r}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{r}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{r}_j}$$

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-1}$$

ويطلق على الرسم البياني لمعاملات الارتباط الذاتي الجزئي مقابل فترات إزاحة مختلفة k بدالة الارتباط الذاتي الجزئي ونرمز لها بالرمز (PACF)، وتستخدم دالة الارتباط الذاتي الجزئي كأداة لتشخيص النموذج المناسب وتحديد درجته وفحص ملائمته للبيانات من خلال اختبار عشوائية أخطاء التنبؤ.

إن دالة الارتباط الذاتي الجزئي (PACF) للسلسلة الزمنية المستقرة تميل للانحدار بسرعة نحو الصفر مع ازدياد فترات الإزاحة، أو تنقطع بعد عدد معين من فترات الإزاحة، والعكس صحيح في حالة السلاسل الغير مستقرة. [10]

### 5. نماذج بوكس - جنكز (B-J) للسلاسل الزمنية:

#### - نماذج الانحدار الذاتي: Autoregressive Models (AR)

يعبر في نماذج الانحدار الذاتي عن قيمة السلسلة الحالية  $z_t$  بدلالة القيم السابقة  $z_{t-1}, z_{t-2}, \dots, z_{t-p}$  وقيمة الخطأ  $e_t$  أي أن قيمة السلسلة الحالية دالة في قيمتها السابقة إضافة إلى بعض الأخطاء ويمكن أن نعبر عن النموذج من الرتبة (p) كما يلي:

$$z_t = \phi_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} \dots + \phi_p z_{t-p} + e_t.$$

حيث:

$z_t$ : قيم مشاهدات السلسلة.

$\phi_0$ : الحد الثابت.

$(\phi_1, \phi_2, \dots, \phi_p)$ : معالم نموذج الانحدار الذاتي.

$e_t$ : الأخطاء العشوائية التي تتوزع توزيعاً طبيعياً بمتوسط صفري وتباين  $\sigma^2$ .

p: درجة النموذج. [5]  
وتستخدم نماذج الانحدار الذاتي لتمثيل السلاسل الزمنية المستقرة وغير المستقرة.

#### - نماذج المتوسطات المتحركة: Moving Average Models (MA)

يتم التعبير في نموذج المتوسطات المتحركة عن قيم السلسلة الحالية  $Z_t$  بدلالة القيم الحالية للأخطاء  $e_t$  ودلالة القيم السابقة للأخطاء  $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$ ، وأن الصيغة الرياضية لهذا النموذج من الرتبة (q) والتي يرمز لها بالرمز MA(q) هي:

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

حيث أن :

$Z_t$ : قيم مشاهدات السلسلة.

$\theta_1, \theta_2, \dots, \theta_q$ : معالم نموذج المتوسطات المتحركة.

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

q: درجة النموذج.

ودالة الارتباط الذاتي (ACF) تنقطع أو تقترب من الصفر بعد الإزاحة q في حين أن دالة الارتباط الذاتي الجزئي (PACF) متناقصة بشكل أسي. [4]

#### - نماذج الانحدار الذاتي والمتوسطات المتحركة: Autoregressive Moving Average Model (ARMA)

هو نموذج مركب يحتوي على خصائص نموذج الانحدار الذاتي والمتوسطات المتحركة. لذلك يمكن التعبير في هذا النموذج عن القيم الحالية للسلسلة  $Z_t$  بالقيم السابقة  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$  وكذلك بالقيم الحالية للأخطاء  $e_t$  بدلالة القيم السابقة للأخطاء  $e_{t-1}, e_{t-2}, \dots, e_{t-q}$  ومعلمات النموذج ويرمز له بالرمز ARMA(p,q).

حيث أن :

p: رتبة الانحدار الذاتي.

q: رتبة المتوسطات المتحركة.

ويمكن كتابة النموذج على النحو التالي:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \dots - \theta_q e_{t-q}$$

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

#### - نماذج الانحدار الذاتي و المتوسطات المتحركة التكاملية: Autoregressive Integrated Moving

#### Average (ARIMA)

يوجد بعض نماذج السلاسل الزمنية تكون غير مستقرة، ولتكون مستقرة يجب إجراء بعض التحويلات أو الفروق عليها، إن مثل هذه النماذج تختلف عن النماذج الأصلية وذلك لاحتوائها على هذه التحويلات أو الفروق، وأن هذه النماذج المستقرة تدعى بالنماذج المختلطة التكاملية، وتتكون هذه النماذج من ثلاثة أجزاء: [8]

الجزء الأول: نموذج انحدار ذاتي، ويستخدم عادة في عملية التنبؤات في السلاسل الزمنية.

الجزء الثاني: نموذج أوساط متحركة MA(q).

الجزء الثالث: (d) هو عدد الفروق التي تحتاج إليها السلسلة الزمنية لكي تكون مستقرة ونرمز لهذه النماذج

ARIMA(p,d,q).

P: رتبة نموذج الانحدار الذاتي AR(p).

q: رتبة نموذج المتوسطات المتحركة MA(q).

d: عدد الفروق التي تتطلبها السلسلة لكي تستمر.

والصيغة العامة لهذه النماذج هي كالتالي:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \dots + d Z_{t-p-d} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

ويمكن اعتبار نماذج (ARMA) هي نماذج (ARIMA) إذا كانت (d=0).

## 6. مراحل بناء نموذج السلاسل الزمنية:

### أ. مرحلة التعرف (Identification):

تعتبر هذه المرحلة من أهم مراحل بناء نماذج السلاسل الزمنية حيث يتم خلالها تحديد رتبة النموذج، ويجب أن يسبق هذه المرحلة تهيئة البيانات، فإذا كانت مستقرة تبدأ عملية تحديد النموذج المناسب لبيانات السلسلة ودرجته باستخدام دالة الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي (PACF).

أما إذا كانت غير مستقرة فنحدد عدم الاستقرارية ناتج من عدم ثبات المتوسط أو التباين و يتم معالجة عدم استقرار المتوسط بأخذ الفرق الأول أو الفرق الثاني، وغالباً ما تستقر بعد أخذ الفرق الأول أو الثاني.

أما عدم استقرارية التباين فيتم معالجته من خلال بعض التحويلات الرياضية على بيانات السلسلة مثل (اللوغاريتم، الجذر التربيعي، المقلوب).

إن الدالتين المستخدمتين في تحديد النموذج و درجته هما دالة الارتباط الذاتي (ACF) والذاتي الجزئي (PACF) وذلك من خلال الرسم البياني لهاتين الدالتين ومطابقة معاملات الارتباط الذاتي والذاتي الجزئي مع السلوك النظري للدالة (ACF) وكذلك الدالة (PACF) فإذا كانت:

- الدالة (ACF) تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب وكانت الدالة (PACF) تنقطع بعد الإزاحة (P) فإن النموذج الملائم للبيانات هو نموذج (AP(p)).

- إذا كانت الدالة (PACF) تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب وكانت (ACF) تنقطع بعد الإزاحة (q) فإن النموذج الملائم هو نموذج MA(q).

- إذا كانت الدالتين ACF، PACF تتناقص تدريجياً بشكل أسي أو سلوك دالة الجيب المتضائلة فإن النموذج الملائم هو ARMA(p,q) [2].

### ب. مرحلة التقدير: (Estimation)

إن عملية تقدير النموذج هي المرحلة الثانية من مراحل بناء السلاسل الزمنية وتحليلها، فبعد التعرف على النموذج الملائم للسلسلة الزمنية يتم تقدير معالمه، و لكي يحقق النموذج الهدف الأساسي من بنائه وهو التنبؤ يجب أن نضمن جودة تقدير معالمه وهناك عدة طرق أهمها:

#### 1. طريقة المربعات الصغرى: (Method of ordinary least squares)

تعتمد هذه الطريقة على تقليص مجموعة مربعات خطأ التقدير وجعله في نهايته الصغرى وذلك من اشتقاق المعادلة:

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

بالنسبة لـ  $(\theta, \phi)$  ومساواة التفاضلات بالصفر فنحصل على المقدرات  $\hat{\theta}, \hat{\phi}$  على التوالي.

#### 2. طريقة الإمكان الأعظم: (Maximum Likelihood Method)

وتعتمد هذه الطريقة على تعظيم دالة الإمكان (أي إيجاد التقدير للمعلمة التي تجعل دالة الإمكان أكبر ما يمكن)، ويتم ذلك كما يلي:

$$L(\phi, \theta, \sigma^2 | z_t) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^n e^2_t}$$

$$\ln(L(\phi, \theta, \sigma^2 | z_t)) = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n e^2_t$$

وبأخذ التفاضل الجزئي للمعادلة السابقة بالنسبة لـ  $\phi, \theta, \sigma^2$  ومساواة التفاضلات بالصفر نحصل على المقدرات  $\hat{\sigma}^2, \hat{\phi}, \hat{\theta}$  على التوالي.

#### 3. مرحلة التشخيص: (Diagnostic)

هي المرحلة الثالثة من مراحل تحليل السلاسل الزمنية، فبعد التعرف على النموذج المبدئي وتقدير معالمه، يجب فحص النموذج والتأكد من مدى ملاءمته وصلاحيته لتمثيل بيانات السلسلة الزمنية ومن ثم إمكانية استخدامه في التنبؤ.

حيث نقوم في هذه المرحلة بإخضاع النموذج لعدد من الاختبارات الإحصائية فإذا اجتاز هذه الاختبارات يكون صالح للاستخدام، وإذا كان غير ملائم نعود إلى المرحلة الأولى، ويتم اختبار نموذج آخر و نكرر هذه العملية حتى الحصول على أفضل وأكفأ نموذج لتمثيل البيانات.

ويتم اختبار ملاءمة النموذج و مدى صلاحيته من خلال اختبار معنوية معالم النموذج وتحليل البواقي والتمثلة في رسم البواقي، حيث أن البواقي يجب أن تتأرجح بنشنت ثابت حول الصفر، كما يجب أن يكون شكل البواقي عشوائياً خالياً من أي أنماط منتظمة.

وكذلك تحليل الارتباطات الذاتية للبواقي عن طريق اختبار (بوكس - لنق) وكذلك طريقة حدود الثقة لمعاملات الارتباطات الذاتية للبواقي والتي يجب أن تقع بين الحدين  $\left(\pm \frac{z_{\alpha/2}}{\sqrt{n}}\right)$  باحتمال  $(1-\alpha)$ ، وإذا تحقق ذلك فإن البواقي تتوزع عشوائيا و أن النموذج ملائم و جيد ويمكن استخدامه في التنبؤ. ولغرض تحديد النموذج الأفضل من بين مجموعة من النماذج توجد عدة معايير منها:

- **معياري أكاي (AIC):** ويستخدم لاختبار النموذج الأفضل حيث يقل تباينه بزيادة عدد المعالم المقدره وصيغته الرياضية:

$$.AIC = -2 \ln(L) + 2k$$

حيث:

L: دالة الإمكان.

k: عدد المعالم المقدره.

- **معياري بيزيان (BIC):** وهو طريقة أخرى لتحديد رتبة نموذج ARIMA حيث يكون النموذج الأفضل والذي يجعل قيمة (BIC) أقل ما يمكن وصيغته الرياضية:

$$.BIC = -2 \ln(L) + k \ln(n)$$

حيث:

L: دالة الإمكان.

k: عدد المعالم المقدره.

- **متوسط مربعات الخطأ:** وهو عبارة عن مجموع مربعات الأخطاء مقسوما على عدد القيم و يرمز له بالرمز (MSE) وصيغته الرياضية:

$$MSE = \frac{1}{n} \sum_{i=1}^n (z_t - \hat{z}_t)^2$$

- **متوسط نسب الأخطاء المطلقة:** وهو عبارة عن متوسط القيمة المطلقة لكل خطأ مقسوما على عدد القيم:

$$MAPE = \frac{1}{n} \sum_{i=1}^n (|z_t - \hat{z}_t| / z_t)$$

والنموذج الأفضل هو النموذج الذي يحقق أقل قيمة وفق كل معيار. [1]

#### 4. مرحلة التنبؤ: (Forecasting)

وهي آخر مرحلة من مراحل تحليل السلاسل الزمنية، وهي عادة تكون الهدف النهائي من تحليل السلاسل الزمنية، ولا يتم الانتقال إلى هذه المرحلة إلا بعد أن يتجاوز النموذج المبدئي كافة الفحوص والاختبارات التشخيصية وإلا يجب العودة إلى مرحلة التعرف و تكرار هذه العملية حتى نتحصل على نموذج ذو كفاءة عالية و يتجاوز جميع الاختبارات و الفروض. [11]

#### 7. الجانب التطبيقي

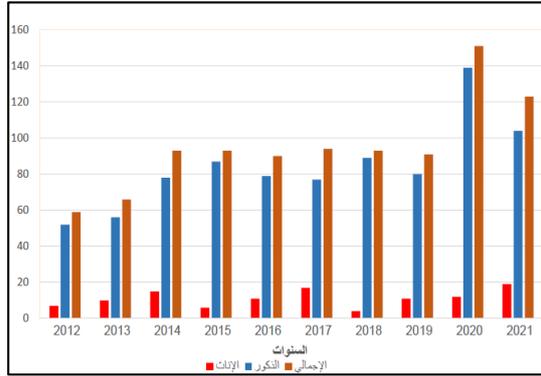
##### أ. جمع البيانات:

تم جمع البيانات والتي تتألف من (120) مشاهدة، وهذه البيانات في الفترة من (2012/01/01) وحتى (2021/12/31)، وتمثل هذه البيانات أعداد النساء والذكور المصابين بمرض سرطان الرئة في المعهد القومي لعلاج الأورام، وتم تلخيص وتبويب هذه البيانات كما هو موضح بالجدول (1).

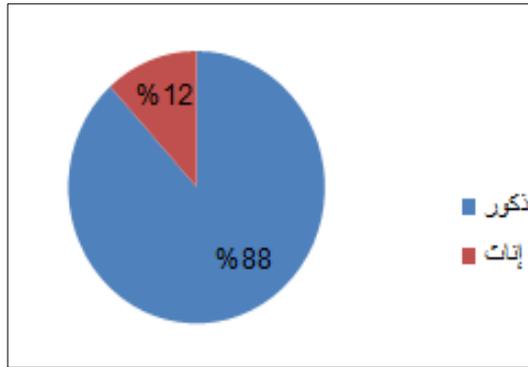
جدول (1) وصف أعداد المصابين بسرطان الرئة في المعهد القومي لعلاج الأورام خلال الفترة (2012- 2021)

الجنس	المقياس			
	أقل قيمة	أكبر قيمة	المتوسط	الانحراف المعياري
ذكور	52	139	84	24
إناث	4	19	11	5
الإجمالي	59	151	95	26

والشكل (1) يوضح تمثيل أعداد المصابين بسرطان الرئة، والشكل (2) يوضح تمثيل المصابين من الذكور والإناث بالنسبة المئوية.

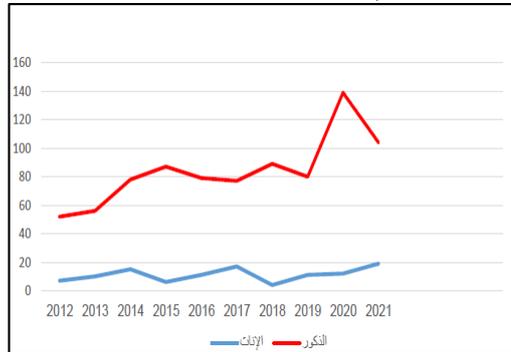


شكل (1): التمثيل البياني لأعداد المصابين بسرطان الرئة في المعهد القومي لعلاج الأورام خلال الفترة (2021/12/31 - 2012/01/01)



شكل (2) النسبة المئوية لأعداد المصابين بسرطان الرئة من الذكور والإناث في المعهد القومي لعلاج الأورام خلال الفترة (2021/12/31 - 2012/01/01)

والشكل (3) يوضح المنحنى البياني لأعداد المصابين بسرطان الرئة من الذكور والإناث في المعهد القومي لعلاج الأورام خلال الفترة (2021/12/31 - 2012/01/01).



شكل (3): المنحنى البياني لأعداد المصابين بسرطان الرئة من الذكور والإناث في المعهد القومي لعلاج الأورام خلال الفترة (2021/12/31 - 2012/01/01)

من الجدول والرسومات البيانية أعلاه نلاحظ أن:

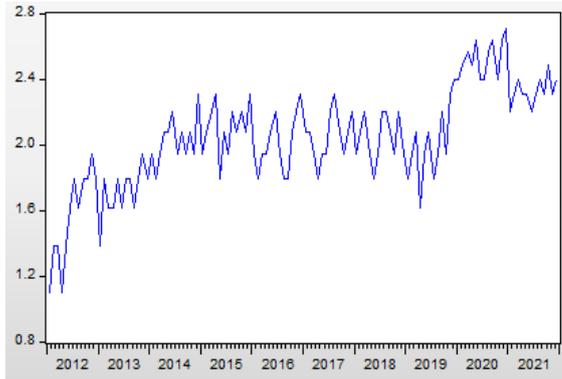
1. أقل عدد للإصابات للإناث كان (4) حالات إصابة في سنة (2018)، وأكبر عدد كان (19) حالة في سنة (2021)، بالنسبة للذكور أقل عدد لحالات الإصابة (52) حالة في سنة (2012)، وأكبر عدد كان (139) حالة في سنة (2020).

مما سبق نستخلص أن المرض يزداد بمرور الزمن.

2. إن المتوسط والانحراف المعياري وكذلك معامل الاختلاف لعدد الإصابات في الذكور أكثر منه في الإناث مما يدل على أن المرض ينتشر في الذكور أكثر من الإناث، وهذا ما تؤيده الأشكال (1) و(2) و(3)، حيث توضح الأعمدة البيانية مقارنة أعداد الإناث بالذكور عبر الزمن، وتوضح القطاعات الدائرية نسبة الإناث ونسبة الذكور مقارنة بالعدد الكلي لمرضى سرطان الرئة خلال الفترة من (2012- 2021). أما المنحنيات فتظهر مدى التطور عبر الزمن لسلسلتي الإناث والذكور خلال الفترة من (2012- 2021).

#### ب. مرحلة تهيئة البيانات:

يتم في هذه المرحلة تحضير البيانات من خلال رسم المنحنى التاريخي لسلسلة المصابين بسرطان الرئة، واستخراج معاملات الارتباط الذاتي والجزئي، وكذلك رسم حدود الثقة لدالة الارتباط الذاتي والجزئي للبيانات الأصلية لمعرفة سلوك هذه البيانات.



شكل (4): المنحنى البياني لأعداد المصابين بسرطان الرئة في المعهد القومي لعلاج الأورام خلال الفترة (2012/01/01 - 2021/12/31)

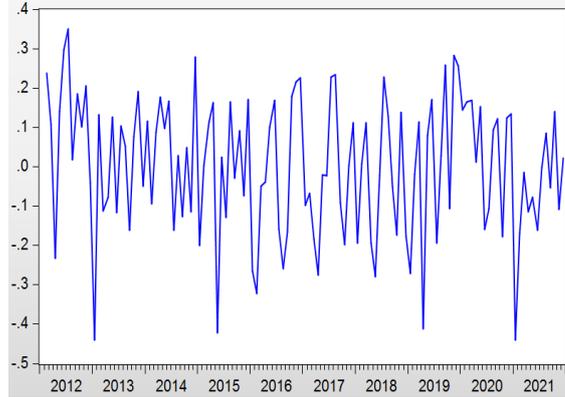
فمن خلال الشكل (4) نلاحظ وجود اتجاه عام متزايد مع الزمن وأن التباين أيضاً لا يميل إلى الثبات مما يدل على عدم استقرارية السلسلة في المتوسط والتباين، وقد أكدت ذلك معاملات الارتباط الذاتي والجزئي كما هو موضح في الشكل (5).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.746	0.746	68.392	0.000	
2	0.672	0.261	124.39	0.000	
3	0.649	0.203	177.07	0.000	
4	0.570	-0.013	218.02	0.000	
5	0.486	-0.071	248.12	0.000	
6	0.503	0.154	280.67	0.000	
7	0.462	0.017	308.35	0.000	
8	0.453	0.091	335.14	0.000	
9	0.399	-0.094	356.10	0.000	
10	0.340	-0.097	371.49	0.000	
11	0.345	0.099	387.52	0.000	
12	0.373	0.155	406.34	0.000	
13	0.276	-0.158	416.73	0.000	
14	0.282	0.000	427.69	0.000	
15	0.251	-0.085	436.48	0.000	
16	0.170	-0.089	440.54	0.000	
17	0.146	0.045	443.59	0.000	
18	0.147	0.013	446.70	0.000	
19	0.109	-0.018	448.41	0.000	
20	0.112	-0.005	450.24	0.000	
21	0.101	0.017	451.75	0.000	
22	0.044	-0.076	452.04	0.000	
23	0.015	-0.070	452.08	0.000	
24	0.002	-0.003	452.08	0.000	
25	0.001	0.109	452.08	0.000	
26	-0.049	-0.165	452.45	0.000	
27	-0.048	0.027	452.81	0.000	
28	-0.025	0.123	452.91	0.000	
29	-0.020	0.045	452.97	0.000	
30	-0.028	0.034	453.10	0.000	
31	-0.034	-0.066	453.28	0.000	
32	-0.014	0.002	453.31	0.000	
33	-0.004	0.042	453.32	0.000	
34	-0.039	-0.013	453.57	0.000	
35	-0.022	0.043	453.65	0.000	
36	0.028	0.084	453.78	0.000	

شكل (5): دالتي الارتباط الذاتي والجزئي لسلسلة المصابين بسرطان الرئة في المعهد القومي لعلاج الأورام خلال الفترة (2012/01/01 - 2021/12/31)

من الشكل (5) نلاحظ أن دالتي الارتباط الذاتي والجزئي أظهرت أن قيمة معاملات الارتباط الذاتي تختلف معنوياً عن الصفر، ولكي تكون السلسلة مستقرة لا بد من دخول جميع معاملات الارتباط الذاتي للعينة ضمن حدود فترة الثقة ما عدا الإزاحة الأولى أو الثانية فيمكن أن تكون خارج حدود الثقة، حيث كانت حدود الثقة للبيانات عند مستوى دقة (95%)

هي (0.18، -0.18). ولاختبار معنوية المعاملات الكلية لدالة الارتباط الذاتي باستخدام اختبار (Box- Lung) فكانت قيمتها ( $Q - stat = 453.78 > X_{0.05,36}^2$ )، وهذا ما يؤكد عدم استقرار السلسلة. ولمعالجة ذلك تم أخذ اللوغاريتم للبيانات الأصلية لغرض تثبيت التباين وأخذ الفروق الأولى لغرض تثبيت المتوسط كما هو موضح بالشكل (6).



شكل (6): المنحنى البياني لسلسلة الفروق اللوغاريتمية للمصابين بسرطان الرئة في المعهد القومي لعلاج الأورام خلال الفترة (2021/12/31 - 2012/01/01)

نلاحظ من الشكل (6) أن السلسلة فقدت الاتجاه العام في سلوكه وثبات التباين مما يدل على استقرار السلسلة في المتوسط والتباين.

ولتأكيد ذلك تم رسم دالة الارتباط الذاتي والذاتي الجزئي لسلسلة الفروق اللوغاريتمية كما هو موضح بالشكل (7).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
█	█	1 -0.370	-0.370	16.700	0.000
█	█	2 -0.139	-0.319	19.072	0.000
█	█	3 0.088	-0.126	20.033	0.000
█	█	4 0.058	0.007	20.460	0.000
█	█	5 -0.190	-0.185	25.039	0.000
█	█	6 0.184	0.059	29.347	0.000
█	█	7 -0.123	-0.118	31.283	0.000
█	█	8 0.142	0.136	33.904	0.000
█	█	9 0.002	0.120	33.905	0.000
█	█	10 -0.116	-0.046	35.685	0.000
█	█	11 -0.076	-0.128	36.454	0.000
█	█	12 0.173	-0.016	40.473	0.000
█	█	13 -0.163	-0.111	44.078	0.000
█	█	14 0.068	-0.024	44.721	0.000
█	█	15 0.100	0.081	46.108	0.000
█	█	16 -0.099	-0.042	47.488	0.000
█	█	17 -0.084	-0.096	48.481	0.000
█	█	18 0.130	-0.007	50.903	0.000
█	█	19 -0.122	-0.060	53.054	0.000
█	█	20 0.006	-0.080	53.058	0.000
█	█	21 0.139	0.071	55.891	0.000
█	█	22 -0.030	0.066	56.020	0.000
█	█	23 -0.067	-0.002	56.700	0.000
█	█	24 0.002	-0.083	56.701	0.000
█	█	25 0.093	0.154	58.028	0.000
█	█	26 -0.065	0.028	58.686	0.000
█	█	27 -0.045	-0.080	58.999	0.000
█	█	28 0.060	0.018	59.568	0.000
█	█	29 0.069	0.040	60.336	0.001
█	█	30 -0.080	-0.042	61.390	0.001
█	█	31 -0.024	-0.012	61.476	0.001
█	█	32 0.024	0.032	61.571	0.001
█	█	33 0.094	0.076	63.064	0.001
█	█	34 -0.141	-0.071	66.412	0.001
█	█	35 0.007	-0.035	66.421	0.001
█	█	36 0.042	-0.050	66.731	0.001

شكل (7) دالتي الارتباط الذاتي والذاتي الجزئي لسلسلة الفروق اللوغاريتمية للمصابين بسرطان الرئة خلال الفترة (2021/12/31 - 2012/01/01)

من الشكل (7) نلاحظ أن جميع الارتباطات الذاتية والذاتية الجزئية للعينة داخل حدود فترة الثقة وأنها غير معنوية، أي لا تختلف عن الصفر ما عدا الفترتين الأولى والثانية، وهذا يؤكد استقرار السلسلة وبذلك أصبحت البيانات جاهزة لعملية التحليل.

### ج. التعرف والتقدير:

يتم في هذه المرحلة تحديد رتبة ونوع النموذج وذلك بالاعتماد على كل من دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي لسلسلة الفروق من الدرجة الأولى اللوغاريتمية، حيث نلاحظ من الشكل (7) أن كل من الدالتين يسلك سلوك

التناقص نحو الصفر، وأن دالة الارتباط الذاتي تنقطع بعد الإزاحة الأولى وهي توافق النموذج MA(1)، وكذلك دالة الارتباط الذاتي الجزئي تنقطع بعد الإزاحة الثانية وهي توافق النموذج AR(2)، وبناء على ما سبق يمكن اقتراح ستة نماذج وهي:

ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(1,1,2), ARIMA(2,1,1), ARIMA(2,1,2).

ولاختيار النموذج الأفضل (الأكثر ملاءمة) من بين هذه النماذج تم استخدام بعض المعايير كما هو موضح بالجدول (2).

جدول (2): تطبيق مجموعة من المعايير على مجموعة من النماذج لتحديد الأفضل بينها

المعيار				النموذج
Schwarz	Akaike	MAP E	RMS E	
0.40644 5	0.336383	6.901	0.168	ARIMA (0,1,1)
0.60946 8	0.539406	7.649	0.180	ARIMA (1,1,0)
0.52149 8	0.164914	6.971	0.168	ARIMA (1,1,1)
0.44173 7	0.348322	6.919	0.169	ARIMA (1,1,2)
0.52828 4	0.621700	6.952	0.169	ARIMA (2,1,1)
0.26953 0	0.362947	6.947	0.169	ARIMA (2,1,2)

من خلال الجدول (2) نلاحظ أن النموذج ARIMA(0,1,1) هو النموذج الأفضل من بين النماذج المقترحة، حيث له أقل قيمة في أغلب المعايير، وبالتالي فهو النموذج الأكثر ملاءمة للبيانات. بعد أن تم التعرف على النموذج الأفضل (الأكثر ملاءمة) للبيانات تأتي المرحلة الثانية من مراحل تحليل السلاسل الزمنية وهي مرحلة التقدير، ولكي يحقق النموذج الهدف الأساسي من بنائه وهو التنبؤ فيجب علينا أن نضمن جودة تقدير معالمه، ولقد تم تقدير معالم النموذج ARIMA(0,1,1) كما هو موضح بالجدول (3).

جدول (3): تقدير معالم النموذج ARIMA(0,1,1)

Type	Coef	SECoef	T-value	P- value
MA	0.6027	0.0731	8.31	0.000
Constant	0.0943	0.00612	15	0.000

#### د. مرحلة التشخيص:

لكي يكون النموذج جيداً وملائماً للبيانات لابد من إجراء الاختبارات عليه، وهي كما يلي:

##### 1. اختبار معنوية المعالم:

نلاحظ من الجدول (3) معنوية معالم النموذج ARIMA(0,1,1) أي أنها تختلف جوهرياً عن الصفر وبالتالي فهو نموذج جيد ويمكن الاعتماد عليه في التنبؤ.

##### 2. تحليل البواقي:

##### أ. دالة الارتباط الذاتي والذاتي الجزئي للبواقي:

لاختبار عشوائية البواقي نقوم برسم دالة الارتباط الذاتي والذاتي الجزئي للبواقي كما هو موضح بالشكل (8).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.046	0.046	0.2539		
2	-0.089	-0.092	1.2348	0.266	
3	0.051	0.061	1.5630	0.458	
4	0.022	0.008	1.6257	0.654	
5	-0.122	-0.116	3.5208	0.475	
6	0.117	0.134	5.2792	0.383	
7	-0.014	-0.055	5.3043	0.505	
8	0.121	0.168	7.2165	0.407	
9	-0.019	-0.059	7.2650	0.508	
10	-0.169	-0.164	11.041	0.273	
11	-0.118	-0.085	12.889	0.230	
12	0.067	0.021	13.490	0.263	
13	-0.106	-0.070	15.002	0.241	
14	0.047	0.050	15.298	0.289	
15	0.065	0.019	15.875	0.321	
16	-0.119	-0.129	17.898	0.270	
17	-0.119	-0.063	19.871	0.226	
18	0.037	0.023	20.063	0.271	
19	-0.085	-0.046	21.100	0.274	
20	0.035	0.022	21.277	0.322	
21	0.157	0.116	24.899	0.205	
22	0.018	-0.001	24.947	0.249	
23	-0.053	-0.040	25.362	0.280	
24	0.011	0.003	25.380	0.331	
25	0.073	0.141	26.186	0.344	
26	-0.045	-0.102	26.503	0.381	
27	-0.028	-0.059	26.627	0.429	
28	0.079	0.070	27.611	0.431	
29	0.076	0.005	28.545	0.436	
30	-0.060	-0.048	29.120	0.459	
31	-0.037	0.033	29.341	0.500	
32	0.028	0.035	29.475	0.545	
33	0.052	-0.005	29.925	0.572	
34	-0.118	-0.102	32.290	0.502	
35	-0.010	0.029	32.306	0.551	
36	0.076	0.031	33.301	0.550	

شكل (8): دالة الارتباط الذاتي والذاتي الجزئي للبواقي لنموذج ARIMA(0,1,1)

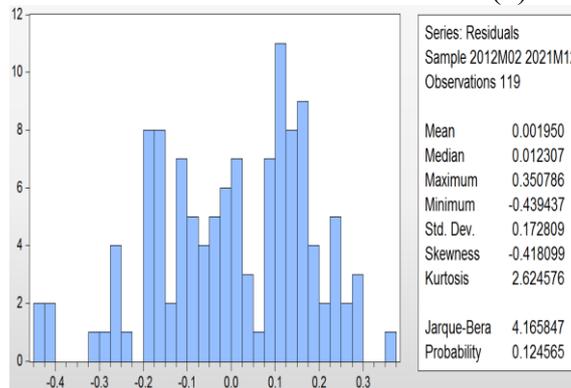
نلاحظ من الشكل (8) أن سلسلة البواقي مستقرة حيث أن جميع معاملات الارتباط الذاتي والذاتي الجزئي للبواقي تقع داخل حدود فترة الثقة وهذا يدل على عشوائية البواقي، وأن النموذج جيد.

**ب. اختبار بوكس- جنكز:**

لزيادة التأكيد تم استخدام اختبار (بوكس- لينج) حيث كانت قيمة الإحصاءة ( $Q=33.301$ ) كما بالشكل (8.)، وهي أصغر من القيمة الجدولية لـ ( $X^2_{0.05,30} = 43.7731$ )، وهذا يدل على أنه لا يوجد أنماط في سلسلة البواقي وبالتالي فهي عشوائية، وعليه فإن النموذج جيد.

**ج. اختبارات التوزيع الطبيعي:**

تم التأكد من أن البواقي تخضع للتوزيع الطبيعي بمتوسط صفري وتباين ثابت ( $\sigma^2$ ) عن طريق الاستعانة بمعاملات التفلطح والتناظر كما هو موضح بالشكل (9).



شكل (9): اختبار طبيعية البواقي باستخدام إحصائية (جوكو- بيررا)

نلاحظ من الشكل (9) أن سلسلة البواقي لها توزيع طبيعي تقريباً وذلك لأن إحصاءة (Jaque- Bera=4.165847) وهي أصغر من القيمة الجدولية لـ ( $X^2_{0.05,2} = 5.99$ ). بالإضافة إلى أن مؤشر التفلطح ( $Kurtosis=2.624576$ ) وهي قريبة من (3)، وكذلك مؤشر الالتواء ( $Skewness=-0.418099$ ) وهي قريبة جداً من الصفر مما يوحي بتناظر التوزيع.

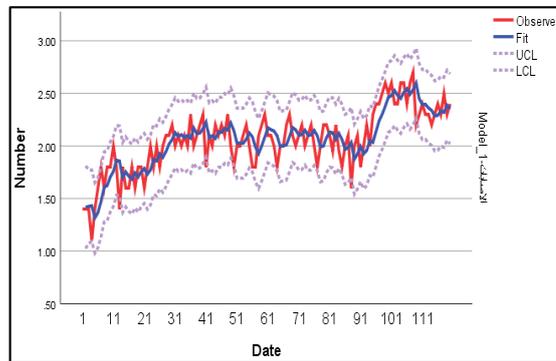
#### هـ. مرحلة التنبؤ:

بعد اختيار النموذج وتقدير معالمه واختبار مدى جودته يمكننا التنبؤ لسنتين قادمتين (2022، 2023) بما يعرف بالتنبؤ خارج العينة وبناء فترة ثقة (95%) لهذه القيم كما هو موضح بالجدول (4).

جدول (4): التنبؤ بعدد الإصابات لمرضى سرطان الرئة في المعهد الوطني لعلاج الأورام لسنتي (2022، 2023) وفق نموذج  $ARIMA(0,1,1)$

السنوات						
فترة الثقة 95%		التنبؤ لسنة 2023	فترة الثقة 95%		التنبؤ لسنة 2022	الشهر
الحد الأعلى	الحد الأدنى		الحد الأعلى	الحد الأدنى		
22	7	12	15	8	11	يناير
22	7	12	16	8	11	فبراير
23	7	13	16	8	11	مارس
23	7	13	17	8	11	أبريل
24	7	13	18	7	12	مايو
24	7	13	18	7	12	يونيو
25	7	13	18	7	12	يوليو
25	7	13	19	7	12	أغسطس
26	7	13	19	7	12	سبتمبر
27	7	13	20	7	12	أكتوبر
27	7	14	20	7	12	نوفمبر
28	7	14	21	7	12	ديسمبر
		156			140	الإجمالي

ويمكن إثبات مدى جودة وكفاءة النموذج من خلال مدى تقارب القيم المتنبئ بها داخل العينة مع القيم الأصلية بما يعرف بالتنبؤ داخل العينة كما هو موضح بالشكل (10).



شكل (10): القيمة الأصلية لسلسلة مرضى سرطان الرئة والقيم المتنبئ بها داخل العينة المقابلة لها خلال الفترة (2012- 2021)

نلاحظ مدى تقارب القيم المتنبئ بها داخل العينة مع القيم الأصلية، وهذا يدل على مدى كفاءة وجوده النموذج في عملية التنبؤ.

### 8. النتائج:

1. نلاحظ من خلال دراسة سلسلة أعداد المصابين بسرطان الرئة في المعهد القومي لعلاج الأورام بمصر أنها غير مستقرة في المتوسط والتباين وأن هناك اتجاهاً عاماً في السلسلة، وهذا يعني أن انتشار المرض يزداد مع مرور الزمن.
2. إن المتوسط والانحراف المعياري ومعامل الاختلاف لعدد إصابات الإناث على التوالي هي (11، 5، 45%) خلال الفترة (2012/01/01 - 2021/12/31) أقل من المتوسط والانحراف المعياري ومعامل الاختلاف لعدد إصابات الذكور في نفس الفترة (84، 24، 29%) مما يدل على أن المرض ينتشر في الذكور أكثر من الإناث.
3. من أسباب زيادة انتشار المرض في الذكور أكثر من الإناث أن نسبة الذكور المدخنين في مجتمعنا أكبر منه في النساء.
4. تم تحقيق الاستقرار لسلسلة مرض سرطان الرئة بعد أخذ الفروق للسلسلة اللوغاريتمية.
5. النموذج الملائم للتنبؤ بعدد الإصابات بمرض سرطان الرئة هو  $ARIMA(0,1,1)$  وتم التوصل إليها باستخدام معايير المفاضلة من بين عدة نماذج.
6. النموذج الذي تم التوصل إليه ملائم وجيد ويمكن الاعتماد عليه في التنبؤ، حيث أظهرت القيم التنبؤية خارج العينة تناسقاً مع القيم الأصلية للسلسلة في الاتجاه.

### 9. التوصيات:

- من خلال النتائج التي تم التوصل إليها نوصي بما يلي:
1. الأخذ بنتائج هذا البحث الذي يظهر تزايداً في أعداد المصابين بمرض سرطان الرئة بمرور الزمن مما يقتضي اتخاذ التدابير اللازمة من قبل الجهات المختصة والكفيلة للحد من هذه الظاهرة.
  2. نوصي بالإقلاع عن ظاهرة التدخين واتباع نظام غذائي صحي وممارسة الرياضة.
  3. نوصي باتباع الخطوات السليمة لصحة الرئتين والمواظبة على الفحص الدوري للتمكن من اكتشاف المرض في مراحله الأولى للاستفادة من عدة خيارات لعلاج.

### 10. المراجع:

1. عبد القادر محمد عبد القادر عطيه، 2000، الحديث في الاقتصاد القياسي، الدار الجامعية، الإسكندرية الطبعة الثانية.
2. عدنان ماجد عبد الرحمن بري، 2002، طرق التنبؤ الإحصائي، مطابع الملك سعود، المملكة العربية السعودية، الطبعة الأولى.
3. سمير مصطفى شعراوي، 2005، مقدمة في التحليل الحديث للسلاسل الزمنية، مركز النشر العلمي، المملكة العربية السعودية، الطبعة الأولى.
4. محمد قدوري الخضير، 1996، دراسة مقارنة لطرق التقدير والتنبؤ لبعض نماذج بوكس- جنكز الموسمية، رسالة ماجستير في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.
5. أحمد شاكر محمد طاهر المتولي، 1989، استخدام تحليل التدخل في السلسلة الزمنية وتطبيقاتها في البيانات البيئية، رسالة ماجستير في الإحصاء، كلية الإدارة والاقتصاد، جامعة صلاح الدين.
6. أكرم عبد الدائم محمد، 2006، تطبيق بوكس- جنكز للسلاسل الزمنية على إصابات مرضى السرطان بالسودان في الفترة (يناير 2002 - ديسمبر 2004).
7. زهراء هادي عبد راهي، ميعاد على كطران، 2017، تحليل السلاسل الزمنية للتنبؤ بأعداد المصابات بمرض سرطان الثدي في محافظة كربلاء بالعراق في الفترة (2012-2016).
8. Box, G.E.P & Jenkins, G.M, (1976), "Time Series Analysis Forecasting and Control", Holden day, London.
9. Box, G.E.P & Pierce, D.A, (1970), "Distribution of the Residual Autocorrelation Autoregressive\_ integrated Moving Average Time series Model", JASA, VOL. 65.
10. Anderson, R.L(1942), "Distribution of the series Analysis correlation Coefficient", Ann. Mat Statistic, Vol. B P-P, ( 113- 129).
11. Douglas, C.M & Contreras, J.G, (1976), "Note On Forecasting With Adaptive Filtering. O.P.Q, VOL. 24, NO.4.

## Using Box- Jenkins Approach in Time- series Analysis forecast the number of lung Cancer infections at the National cancer Institute, Misurata (2012- 2021)

Emhemed Milad Baawa

Statistics Department, Science Collage, Misurata University, Misurata, Libya

[emhemmedmilad2017@gmail.com](mailto:emhemmedmilad2017@gmail.com)

### Article information

### Abstract

#### Key words

Time- series, Box- Jenkins approach, Cancer tumors.

Received 28 02 2025, Accepted 15 03 2025, Available online 18 03 2025

The important of time- series analysis is to get an accurate description of the time- series and to create an appropriate Model for future forecasting, in order to take the right decisions in all fields.

This research aims to analyze the time- series by using the (Box- Jenkins) approach to find the best and most efficient model the researcher reached for predicting the number of lung cancer at the National cancer Institute based on monthly data for the period (2012 to 2021).

The lung cancer series has shown to be unstable and has a general trend in these conditions, the appropriate.

Development for it, is the model of first- class integrated moving averages. (ARIMA) (0,1,1), and based on this model, the numbers of lung cancer patients were predicted monthly for the next two years (2022- 2023) and the predicted values were consistent with red values in the direction indicating the model's efficiency.