

## التنبؤ بأعداد مرضى السرطان باستخدام منهجية بوكس وجينكنز في محافظات غزة

### Forecasting of the numbers of patients with cancer using Box-Jenkins methodology in Gaza provinces

شريف شحدة سلامة مصلح

المدرسة الوطنية العليا للإحصاء والاقتصاد التطبيقي - الجزائر

Sh.pal2010@gmail.com

2019/2/17

تاريخ القبول

2018/8/13

تاريخ الاستلام

#### ملخص:

يهدف البحث إلى تحليل السلاسل الزمنية باستخدام منهجية (Box-Jenkins) في التحليل لإيجاد أفضل نموذج للتنبؤ بأعداد المصابين بأمراض السرطان في محافظات غزة، وذلك بالاعتماد على البيانات الشهرية للفترة (2010-2016). وقد أظهرت نتائج تحليل البيانات أن النموذج الملائم لها هو نموذج الانحدار الذاتي المتكامل من الدرجة الثالثة  $ARIMA(3,2,0)$ . وبالاعتماد على هذا النموذج تم التنبؤ بأعداد المصابين بأمراض السرطان شهرياً، ولسنتين قادمتين. وقد كانت القيم التنبؤية قريبة من قيم السلسلة الأصلية، وكذلك وفقاً لمعيار جذر متوسط مربعات البواقي لقياس دقة التنبؤ الذي حصل على أقل قيمة بين مجموعة النماذج مما يدل على كفاءة النموذج.

**كلمات مفتاحية:** بوكس وجينكنز، أعداد المصابين بأمراض السرطان، التنبؤ، محافظات غزة.

#### Abstract:

*The aim of this research is to analyze time series using (Box-Jenkins) methodology to find out the best forecasting model of the numbers of patients with cancer in Gaza provinces, by using the monthly data of the period (2010-2016). The results have shown that the most fit model was third degree Integrated Autoregressive model,  $ARIMA(3,2,0)$ . According to this model, the numbers of patients with cancer were predicted monthly as well as in the next two years, so the forecasting values were closed with the original series values, and also according to Root Mean Square Error to*

*measure the prediction accuracy, which has got the lowest value among the models group, indicating the efficiency of the model.*

**Key Words:** Box-Jenkins, numbers of cancer patients , forecasting, Gaza provinces.

## 1. مقدمة

إن التنبؤ من أهم أدوات اتخاذ القرار، وأهم عنصر في عملية التخطيط للمستقبل، وهو أحد أهم القضايا الأساسية المستخدمة في عدة مجالات، ويعتبر تحليل السلاسل الزمنية من الأساليب الإحصائية الشائعة المستخدمة في التنبؤ على نطاق واسع في الكثير من التطبيقات الإحصائية والاقتصادية والصحية وفي مجال علوم الحاسبات والشبكات العصبية وغيرها؛ حيث يتم التنبؤ بسلوك المتغير التابع استناداً إلى سلوك هذا المتغير في الماضي، فمنذ مطلع العقد السابع من القرن العشرين ظهر اهتمام متزايد بتحليل السلاسل الزمنية وطرائق التنبؤ بقيمتها المستقبلية، وفي بداية الثمانينات من القرن العشرين ظهر اهتمام خاص بتحليل السلاسل الزمنية غير الخطية ونمذجتها، ومع بداية العقد الأخير من القرن العشرين ظهرت توجهات لدراسة الخصائص الفوضوية في السلاسل الزمنية، ومع إطلاقة القرن الحادي والعشرين تزايد الاهتمام بدراسة السلاسل الزمنية، وخاصة من خلال علاقتها الوثيقة بالنظم الديناميكية.

يرتكز تحليل السلاسل الزمنية على فرضية رئيسية، هي أن السلسلة الزمنية تتوزع طبيعياً Normal، أي أنها تتوزع طبيعياً، وهذا الافتراض قد تم تخطيه من خلال بعض الدراسات التي ظهرت والمتعلقة بالسلاسل الزمنية غير الطبيعية. وتقوم طريقة تحليل السلاسل الزمنية على فكرة إيجاد نموذج رياضي مناسب لطبيعة البيانات، بحيث يجعل البواقي (الأخطاء) Residuals بين القيم الحقيقية للسلسلة والقيم المقدرة، باستخدام ذلك النموذج الرياضي أقل ما يمكن، ولا يوجد بها أي نوع من الترابط الداخلي فيما بينهما (عكاشة محمود، 2002: 517).

تعرضت محافظات غزة في فلسطين منذ عام 2008م لحروب كارثية طالت موارده المادية والبشرية دمرت البنى التحتية ولوثت البيئة، مما يقتضي إجراء تنمية ونهضة شاملة في شتى المجالات وخاصة قطاع الصحة، وهذا يحصل بتكاتف جهود الباحثين في كافة الاختصاصات؛ لإجراء الدراسات التي من شأنها الحد من التلوث والأمراض التي طالت القطاع الصحي في محافظات غزة لأهميته على الصعيد التنموي لكونه يهتم بالموارد البشرية والتي تقع على عاتقها مسؤولية البناء والإعمار ومواكبة التقدم والتطور الحضاري، ومن مقومات النهوض بالصحة مكافحة جميع الأمراض وخاصة أمراض السرطان، والتي عرفتتها منظمة الصحة العالمية بأنها "مجموعة كبيرة من الأمراض السرطانية التي يمكنها أن تصيب كل أجزاء الجسم، ويشار إليها بالأورام الخبيثة أو السرطان، وهو نمو الخلايا وانتشارها بشكل لا يمكن التحكم فيه، وغالباً ما تغزو خلايا الجسم التي

تحيط بها ويمكنها أن تتسبب في نقائل تظهر في مواضع أخرى بعيدة عن الموضع المصاب" (منظمة الصحة العالمية: 2018)، وهذه الأمراض تسبب أعدادا كبيرة من الوفيات مقارنة بالأمراض الأخرى، ونظراً لازدياد عدد المصابين بهذا المرض في الآونة الأخيرة، فقد جاءت هذه الدراسة من أجل كشف هذه الظاهرة والتي ازدادت بشكل متنام في محافظات غزة التي تأثرت بالأسلحة المحرمة دولياً، بالإضافة إلى النقص المزمن في الرعاية الصحية والعلاجية بسبب الحصار المتواصل الذي فرضه الإحتلال الصهيوني على محافظات غزة ، لذلك سوف نقوم بدراسة هذه الظاهرة بالاعتماد على بيانات السلسلة الزمنية الشهرية للفترة (2010-2016) باستخدام منهجية (Box-Jenkins).

## 2. مشكلة البحث

إن عملية التنبؤ في السلاسل الزمنية تتأثر بشكل مباشر باختيار النموذج المناسب لبيانات السلاسل الزمنية وقد تتأثر دقة التنبؤ بحجم مشاهداتها، وتؤثر هذه الخطوة في دقة التنبؤات المتحصل عليها، لذلك تمحورت مشكلة الدراسة في التساؤل الآتي:

ما النموذج الأمثل والأكثر دقة من بين نماذج (ARIMA) المتعددة في التنبؤ بأعداد مرضى السرطان في محافظات غزة؟

## 3. أهمية البحث

تكمن أهمية هذا البحث في أنه يتكون من جانبين: تطبيقي وإحصائي، فمن الجانب التطبيقي فإن بناء نموذج لأعداد مرضى السرطان في محافظات غزة يحقق الكثير من المنافع الصحية في تلك المحافظات، تتمثل في إمداد أصحاب القرار بمعلومات مستقبلية عن موضوع مهم بأعداد مرضى السرطان، وما يترتب على ذلك من وضع الخطط حسب النتائج المتوقعة لمواجهة المشاكل الصحية المترتبة على التغير في أعداد مرضى السرطان، أما من الجانب الإحصائي؛ فإن استخدام التحليل الإحصائي المعتمد على نماذج وأساليب حديثة ومبتكرة للتنبؤ بالسلسلة الزمنية مثل منهجية بوكس وجينكز، نموذج (ARIMA) يجعل النموذج المستخدم أكثر دقة.

## 4. أهداف البحث

يهدف هذا البحث إلى الآتي:

1. معرفة نماذج (ARIMA) وكيفية بناء النماذج وتحديد النموذج الأمثل.
2. معرفة كفاءة نموذج (ARIMA) بوكس-جينكز في التنبؤ بأعداد مرضى السرطان الشهرية بمحافظات غزة.
3. التنبؤ بأعداد مرضى السرطان في محافظات غزة باستخدام النموذج الأمثل في التنبؤ.

## 5. منهج البحث

استناداً للدراسات السابقة، فإن الباحث سيستخدم في هذا البحث المنهج الإحصائي الوصفي والتحليلي حيث سيستخدم في المنهج التحليلي منهجية بوكس وجينكنز، باستخدام برنامج التحليل الإحصائي SPSS.

## 6. الدراسات السابقة

- تناولت العديد من الدراسات السابقة منهجية بوكس وجينكنز للتنبؤ بالسلاسل الزمنية، أهمها:
1. (طعمه 2012) هدفت في بحثها الى دراسة وتحليل السلاسل الزمنية باستخدام طريقة بوكس وجينكنز للتنبؤ بأعداد المصابين بالأورام الخبيثة لفترة (2006-2010) اعتماداً على البيانات الشهرية. وقد أظهرت نتائج تحليل البيانات أن النموذج الملائم لها هو نموذج الانحدار الذاتي المتكامل من الدرجة الثانية (2.1.0) ARIMA بالاعتماد على هذا النموذج تم التنبؤ بأعداد المصابين بالأورام الخبيثة شهرياً ولستينين قادمين .
  2. (النعمي والشاروط 2000: 4-9) قام الباحثان بتحديد افضل نموذج للتنبؤ بعدد المرضى المصابين بالأورام الخبيثة في محافظة القادسية باستخدام تحليل التدخل المتمثل في الحصار الاقتصادي مجزئاً السلسلة الزمنية الى جزئين من (90-93) كفترة أولى و (94-97) كفترة ثانية، ولاحظا من خلال سلسلة أعداد المصابين بالأورام الخبيثة أنها غير مستقرة في الوسط الحسابي والتباين، وأن هناك اتجاه عام واضح في السلسلة بعد عام 1993.
  3. دراسة (Khan, 2013) " تناولت هذه الدراسة التنبؤ بأسعار الذهب باستخدام منهجية بوكس وجينكنز (ARIMA) لبناء نموذج التنبؤ، وأشارت النتائج إلى أن نموذج (0,1,1) ARIMA هو النموذج المناسب للتنبؤ بأسعار الذهب.

## 7. الإطار النظري

تعددت تعريفات السلسلة الزمنية بحسب طبيعة الغرض من الدراسة والتخصص، ومن أبرزها هو أن "السلسلة الزمنية مجموعة من المشاهدات المرتبطة ببعضها بعضاً، يتم تسجيلها في فترات زمنية متعاقبة لظاهرة معينة (سنوية - ربع سنوية - شهرية - أسبوعية - يومية ... ) وهي إما أن تكون متقطعة أو مستمرة" (Box & Jenkins, 1979).

### 1.7. مركبات السلاسل الزمنية

يقصد بمركبات السلسلة الزمنية العناصر المكونة لها، وهي تفيد في تحديد سلوكها في الماضي وكذلك المستقبل، ويمكن إدراج هذه المركبات في العناصر الآتية (مولود حشمان، 2002: ص12): مركبة الاتجاه العام، والمركبة الفصلية الموسمية، ومركبة الدورات الاقتصادية، والمركبة العشوائية.

## 2.7. خصائص السلسلة الزمنية

### 1.2.7 العشوائية stochastic

تعتبر العشوائية من خصائص السلسلة الزمنية وتتمثل في المركبة العشوائية التي يجب أن تكون قد تولدت عن ظروف عشوائية، وبافتراض أنه لدينا سلسلة  $Y_t$  ذات مركبتين عشوائية واتجاه عام، وبأخذ فروقاتها من الدرجة الأولى فتحصل على سلسلة عشوائية فقط كالآتي (مولود حشمان، 2002)

$$Y_t = Y_{t-1} + \varepsilon_t \quad (7.1) \dots\dots$$

وهذا النموذج (7.1) يسمى بنموذج الانتقال العشوائي (process Random Walk) أو تستطيع تسميته بنموذج الانحدار الذاتي من الدرجة الأولى بمعلمة أحادية (1) AR بتعبير (B-j)، وبالتعويض الخلفي أو التراجعي (Back Substitution) في النموذج (7.1) نحصل على:

$$Y_t = y_{t-j} + \sum_{i=0}^{j-1} \varepsilon_{t-i} \quad (7.2) \dots\dots\dots$$

وبالرجوع الخلفي إلى بداية تكوين هذه السلسلة، ووفق هذا النموذج يمكن كتابة (-2-9) في الشكل الجديد التالي لما  $t=j$

$$Y_t = \sum_{i=0}^{t-1} \varepsilon_{t-i} + y_0 \quad (7.3) \dots\dots\dots$$

(حيث  $y_0$  هي نقطة بداية السلسلة، والجزء الثاني ما هو إلا مجموع عناصر عشوائية.)

### 2.2.7 الاستقرار Stability

يمكن التمييز بين نوعين من السلاسل الزمنية: السلاسل الزمنية المستقرة، والسلاسل الزمنية غير المستقرة؛ حيث إن هناك حالتين من الاستقرار، هما الاستقرار في المتوسط (Stability in mean)، والاستقرار في التباين (Stability in variance).

إن الاستقرار في المتوسط هي حالة السلسلة عندما لا تظهر اتجاهًا عامًا، ويمكن تحويلها إلى مستقرة باستخدام الفروق، أما الاستقرار في التباين فهي حالة السلسلة عندما لا تظهر تذبذبات متباعدة في شكل السلسلة الزمنية، ويمكن تثبيت التباين بالحصول على اللوغاريتم الطبيعي أو الجذر التربيعي أو المقلوبات لبيانات السلسلة (مولود حشمان، 2010: 140).

حيث يوجد نوعان من الاستقرار (فاندلر ولتر، 1992):

#### - الاستقرار التامة Full Stability

يقال للسلسلة الزمنية  $(y_t; t = 1, 2, \dots, n)$  أنها مستقرة استقرارية تامة إذا كان التوزيع المشترك لأي مجموعة من المشاهدات لا يتأثر بإزاحة كل الفترة الزمنية للمشاهدات إلى الأمام، أو إلى الخلف بأي كمية صحيحة، أي أن:  $t_n$  تمثل أي فئة زمنية و  $k$  مقداراً ثابتاً.

- الاستقرار الضعيف Weak Stability

يقال إن السلسلة الزمنية  $(y_t, t = 1, 2, \dots, n)$  بأنها ذات استقرار ضعيف أو من الدرجة الثانية إذا تحقق الاستقرار، ونظراً لصعوبة تحقق شروط الاستقرار التام للسلسلة الزمنية جرت العادة بعدم اشتراط أن تكون السلسلة مؤكدة الاستقرار، ولكن نكتفي فقط بالاستقرار الضعيف (Matroushi, 2011). وتعتبر السلسلة الزمنية مستقرة إذا توفرت الخصائص والشروط الآتية (كاناموس زودون، 2004):

$$1. \text{ ثبات متوسط القيم عبر الزمن } E(Y_t) = \mu$$

إذ إن  $\mu$  هو متوسط العملية العشوائية، ويكون ثابتاً ولا يعتمد على قيم  $t$ ، والذي يقدر من مشاهدات السلسلة على النحو الآتي:

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$$

2. تباين السلسلة الزمنية يكون ثابتاً؛ إذ إن  $y_0$  هو تباين العملية العشوائية ويكون ثابتاً ولا يعتمد على قيم  $(t)$  والذي يقدر من مشاهدات السلسلة الزمنية كما يأتي:

$$\text{var}(y_t) = E(y_t - \mu)^2 = \sigma_y^2 = y_0$$

3. التباين المشترك للسلسلة الزمنية كالتالي:

$$\hat{y}_0 = \frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2$$

3.2.7. دالة الارتباط الذاتي والارتباط الذاتي الجزئي

من المعايير التي تستخدم في اختبار استقرار السلسلة هي:

- دالة الارتباط الذاتي (ACF) Autocorrelation Function

توضح هذه الدالة الارتباط الموجود بين المشاهدات لفترات مختلفة وهي ذات أهمية بالغة في إبراز بعض الخصائص المهمة للسلسلة الزمنية، ومن العملية نقوم بتقدير دالة الارتباط الذاتي للمجتمع بواسطة دالة الارتباط الذاتي للعينة (تومي صالح، 1999: 173)، وحساب دالة الارتباط الذاتي من بيانات عينة كما يأتي (مولود حشمان، 2010: 145-146):

$$\begin{aligned} \rho_k &= \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t) \text{var}(y_{t-k})}} \\ &= \frac{y_k}{\sqrt{\sigma_y^2 \cdot \sigma_y^2}} = \frac{y_k}{\sigma_y^2} \\ \rho_k &= \frac{y_k}{y_0} \end{aligned} \quad (7.4) \dots$$

وتكون مقدرات هذه التباينات والتباينات المشتركة ثم معاملات الارتباط الذاتي الخاص بالعينة كما يأتي:

$$\rho_k = \frac{\hat{Y}_k}{\hat{Y}_0} = r_k$$

#### - دالة الارتباط الذاتي الجزئي (PACF) Partial Autocorrelation function

أمام الوضع الصعب للتعرف على نماذج الانحدار الذاتي (AR (P)، حتى وإن كان من درجة بسيطة من خلال دالة الارتباط الذاتي يجب الاستعانة بدالة الارتباط الذاتي الجزئية (مولود حشمان، 2002: 137). والارتباط الذاتي الجزئي هو مؤشر يقيس العلاقة بين  $Y_t$  و  $Y_{t-k}$  للسلسلة نفسها، مع افتراض ثبات قيم السلسلة الزمنية. ويعرف على أنه الحد الأخير من نموذج الانحدار الذاتي من الدرجة (AR (P) (الخضير، 1996: 8). ويتم إيجاد قيم معامل الارتباط الذاتي الجزئي، وذلك عن طريق دالة الارتباط الذاتي حسب الصيغة الآتية (الطائي والشرابي، 2009: 130):

$$\phi_{11} = \rho_1$$

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j} \quad \dots\dots (7.5)$$

إذ أن:

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} ; \quad j = 1, 2, \dots, k-1$$

### 3.7. نماذج بوكس وجينكنز (B-J) Box & Jenkins لتحليل السلاسل الزمنية

#### 1.3.7. نموذج الانحدار الذاتي العام (AR (P) Autoregressive model

الصيغة العامة لنموذج الانحدار الذاتي من الدرجة (p) ستأخذ الشكل الآتي (مولود حشمان، 2010)

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad \dots (7.6)$$

وبعد إدخال فكرة معامل التأخير (L) lag operator يمكن كتابة النموذج (7.6) كما يأتي:

$$\theta(L) Y_t = \delta + \varepsilon_t \quad \dots (7.7)$$

إذ أن:

$$Y_t = \phi^{-1}(L) \delta + \phi^{-1}(L) \varepsilon_t$$

أي  $\phi^{-1}(L)$  هو مقلوب أو معكوس كثير الحدود  $\phi(L)$ .

إن نموذج الانحدار الذاتي يمكن أن يستخدم لتمثيل السلاسل الزمنية المستقرة وغير المستقرة، وأن شروط تحقيق استقرارية النموذج يجب أن تقع جذور المعادلة  $\phi(p)(L) = 0$  خارج حدود دائرة الوحدة، أي أن تكون  $(-1 < \phi_p < 1)$  (طعمة سعدية، 2012: 376).

### 2.3.7. نموذج المتوسطات المتحركة العام (q) Moving average model MA

يكتب هذا النموذج في شكل خطي عام كما يأتي (مولود حشمان، 2002: 123):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad \dots (7.8)$$

ويمكن تمثيل نموذج المتوسطات المتحركة من الدرجة (q) باستخدام عامل التأخير (L) على النحو الآتي (الجبوري، 2010: 17):

$$y_t = \mu + (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t \quad \dots (7.9)$$

حيث إن  $\theta_i$  معاملات نموذج المتوسطات المتحركة وتكون محصورة ما بين  $-1 < \theta < 1$ .

### 3.3.7. النماذج المختلطة العامة ARMA (p,q)

تشتمل هذه النماذج على القسم الانحداري ذي الدرجة (p) وقسم المتوسطات المتحركة من الدرجة (q) التي تكتب في الشكل العام الآتي (مولود حشمان، 2010: 166-168):

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad \dots (7.10)$$

وبإدخال معامل التأخير تصبح العلاقة كما يأتي:

$$y_t = \phi_1 L y_t + \phi_2 L^2 y_t + \dots + \phi_p L^p y_t + \varepsilon_t + \theta_1 L \varepsilon_t + \dots + \theta_q L^q \varepsilon_t$$

$$y_t - \phi_1 L y_t - \phi_2 L^2 y_t - \dots - \phi_p L^p y_t = \varepsilon_t + \theta_1 L \varepsilon_t + \dots + \theta_q L^q \varepsilon_t$$

ومنه:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t$$

إذ:

$$\phi(L) y_t = \theta(L) \varepsilon_t \quad \dots (7.11)$$

إن شرطي الاستقرار والانعكاسية للنموذج المختلط ARMA (p,q)، هي أن جذور المعادلتين  $\phi(L) = 0$  لنماذج الانحدار الذاتي، و  $\theta(L) = 0$  لنماذج المتوسطات المتحركة، يجب أن يكونا خارج دائرة الوحدة (الجبوري، 2010: 19)، وذلك عندما تكون  $K > q = 1$ .

### 3.3.7. النماذج المختلطة المركبة Autoregressive Integrated Moving

#### . Average Models (ARIMA)

تعتبر نماذج ARIMA أكثر نماذج السلاسل الزمنية استخداماً، وعند إدخال معامل الفروق (d) إلى نموذج ARMA (p,q) يتحول النموذج إلى نموذج ARIMA (p,d,q)، حيث يمكننا اشتقاق جميع النماذج منها سواء نموذج الانحدار الذاتي AR (p) والذي يستخدم عادة في عملية التنبؤات للسلسلة الزمنية، أو المتوسطات المتحركة MA (q) وكذلك النموذج المختلط ARMA (p,q) أما الجزء (d) يمثل الفروق التي تتطلبها السلسلة الزمنية حتى تصبح مستقرة.



$$\phi(L)(1-L)^d y_t = \phi_0 + \theta(L)\varepsilon_t \quad \dots (7.12)$$

حيث إن:

$$\phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$$

$$\theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$$

$$\Delta_d = (1 - L)^d$$

حيث  $\Delta_d$  هو معامل الفروق من الدرجة d

وتكون الصيغة العامة للنموذج المختلط المركب (المتكامل) ARIMA (p,d,q) كما يأتي (Kaiser ,

:2001: 34)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \dots + \Delta_d y_{t-p-d} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

....(7.13)

وعليه يمكن اعتبار نماذج ARIMA هي نماذج ARMA مستقرة مع اختلاف الدرجة.

#### 4.7. مراحل منهجية بوكس وجينكنز

##### 1.4.7. مرحلة تحديد النموذج : Identification

تعد مرحلة تحديد النموذج من أهم مراحل بناء نماذج السلاسل الزمنية Box & Jenkins

عام 1976، ويجب البدء بمرحلة تهيئة البيانات قبل مرحلة تحديد النموذج كما يأتي (Box &

:Jenkins, 1976: 245)

1. فإذا كانت البيانات مستقرة من خلال ملاحظة رسم البيانات الأصلية للسلسلة الزمنية قيد الدراسة، وكذلك الارتباطات الذاتية والجزئية لها فإن البيانات تكون مهيأة للتحديد.

2. أما إذا كانت السلسلة الزمنية غير مستقرة في المتوسط والتباين، فإنه يتم معالجة عدم الاستقرار في الوسط بأخذ الفروق الأول (d=1) فإذا لم تستقر نأخذ الفرق الثاني (d=2) وغالباً ما تستقر بعد أخذ الفرق الأول أو الثاني.

3. أما معالجة عدم الاستقرار في التباين، فيتم من خلال إجراء التحويل المناسب للبيانات.

4. بعد تحقيق استقرار السلسلة الزمنية، تبدأ عملية تحديد النموذج ونقصد بذلك استخدام البيانات أو أية معلومات عن الكيفية التي تتولد بها السلسلة الزمنية، فالهدف هنا هو الحصول على فكرة عن قيمة q,d,p التي نحتاجها في النموذج الخطي العام ARIMA والموضحة صيغته في المعادلة (7.13)، ومن ثم الحصول على تقديرات أولية لمعاملات النموذج. إن المعيارين

المستخدمين لتحديد النموذج ودرجته هما دالة الارتباط الذاتي (ACF) ودالة الارتباط الذاتي الجزئي (PACF)، حيث يتم الرسم البياني للدالتين، وبعد ذلك يتم مطابقة معاملات الارتباط الذاتي والجزئي مع السلوك النظري لدالة الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي لـ (PACF) (مولود حشمان، 2002: 145) و(الصراف، 1981: 16-18) فإن:

1. تكون دالة الارتباط الذاتي (ACF) مؤشراً مهماً لكشف عدم استقرار سلسلة زمنية، فعندما لا تتعدى هذه الدالة بعد الفترة  $(\frac{T}{4})$  أي ربع عدد المشاهدات تكون السلسلة الزمنية غير مستقرة، حيث يجب عملياً أن تقع معاملات هذه الدالة خارج مجال الثقة المحدد إحصائياً، وهنا نكون بصدد دراسة النماذج المركبة.
2. بيان دالة (ACF) تتناقص تدريجياً وبشكل أسي أو سلوك دالة الجيب المتضائلة وبيان دالة (PACF) ينقطع بعد الإزاحة (P) فيكون النموذج الملائم للبيانات هو  $AR(P)$ .
3. بيان دالة (ACF) ينقطع بعد الإزاحة (q) وبيان دالة (PACF) تتناقص تدريجياً وبشكل أسي أو سلوك دالة الجيب المتضائلة، فيكون النموذج الملائم للبيانات هو  $MA(q)$ .
4. بيان الدالة (ACF) و (PACF) تتناقص تدريجياً وبشكل أسي أو سلوك دالة الجيب المتضائلة فيكون النموذج الملائم للبيانات هو  $ARMA(p,q)$ .

#### 2.4.7. مرحلة تقدير معلمات النموذج Estimation

لتحقيق النموذج الهدف الأساسي من بنائه وهو التنبؤ فيجب علينا أن نضمن جودة تقدير معلماته وملائمته للسلسلة الزمنية، ويوجد عدة طرق لتقدير معلمات النموذج من أبرزها:

1. طريقة المربعات الصغرى الاعتيادية (O.L.S.E) Method of ordinary least squares

وتقوم فكرة هذه الطريقة على مبدأ تقليص مجموع مربعات خطأ التقدير، وجعله في نهايته الصغرى، وذلك بين المستوى الفعلي  $y_t$  والمقدرة  $\hat{y}_{t-p}$ .

2- طريقة الإمكان الأعظم أو المعقولة العظمى Maximum likelihood Method

تتلخص هذه الطريقة في أن قيم مصفوفة معلمات النموذج المراد تقديرها يتم اختيارها وفقاً لمبدأ تعظيم دالة الإمكان.

ويتوقف التقدير بهذه الطريقة بالأساس على تحقق فرضية التوزيع الطبيعي التي تعتمد مبدأ تصغير أو تدنية مجموع مربعات البواقي Min RSS، بمعنى اختيار شعاع المعلمات.

#### 3.4.7. مرحلة اختبار دقة النموذج Diagnostic checking of model

بعد التحديد المبدئي للنموذج، وتقدير معلماته تأتي مرحلة اختبار دقة النموذج وهي التحقق من مدى ملائمة النموذج ومدى صلاحيته لتمثيل بيانات السلسلة الزمنية وإمكانية استخدامه للتنبؤ، وهي من

أهم وأدق مراحل التحليل والتي تحدد قبول أو استبدال النموذج بنموذج آخر وذلك من خلال إجراء العديد من الاختبارات أهمها:

#### 1.3.4.7. تحليل الاستقرار

يتم من خلال اختبار تقديرات معلمات الانحدار الذاتي التي تم الحصول عليها في مرحلة التقدير للتأكد من أنها تحقق شروط الاستقرار، وهي أن جميع جذور المعادلة المميزة تقع جميعها خارج دائرة الوحدة  $\theta(B) = 0$  بمعنى إذا كانت القيمة المطلقة لكل جذر من هذه الجذور أكبر من الواحد الصحيح فهذا يدل على استقرار العملية العشوائية.

#### 2.3.4.7. تحليل الانعكاس Invertibility

ويقصد بتحليل الانعكاس إمكانية قلب النموذج، فإذا كانت جذور المعادلة  $\theta(B) = 0$  تقع جميعاً خارج دائرة الوحدة، فإن التقديرات الخاصة بمعلمات المتوسطات المتحركة تحقق شروط الانعكاس، بمعنى إذا كانت القيمة المطلقة لكل جذر من هذه الجذور أكبر من الواحد الصحيح فهذا يدل على انعكاس النموذج الأصلي، وفي حالة كان أحد الجذور قريب من الواحد صحيح فهذا يدل على استخدام فروق غير ضرورية.

#### 3.3.4.7. اختبار معنوية معلمات النموذج

ويتم ذلك باستخدام إحصاء الاختبار ( $t$ - student) من أجل التحقق من معاملات النموذج إحصائياً أي لا تختلف جوهرياً عن الصفر، من خلال قسمة المعلمة المقدرة على انحرافها المعياري، ومن ثم تقارن بالقيمة الجدولية التي توافق مستوى المعنوية  $\alpha$ ، ودرجات الحرية  $(n-k)$  فإذا كانت غير معنوية لا بد من استبعاد أحد درجات  $AR$  أو  $MA$ .

#### 4.3.4.7. تحليل البواقي

بعد التعرف على نموذج مبدئي وتقدير معلماته، يتم تطبيق بعض اختبارات الفحص على البواقي  $e_t$  (مقدرات الأخطاء) لمعرفة مدى مطابقة النموذج لبيانات السلسلة المشاهدة، وبافتراض أن مقدرات البواقي  $(\hat{e}_t)$  أنها مقدرات الأخطاء  $(\varepsilon_t)$ ، بمعنى أن الأخطاء مستقلة فيما بينها، وموزعة طبيعياً بمتوسط صفر وتباين  $\sigma^2$  وتعطي بالعلاقة الآتية:

$$\hat{e}_t = y_t - \hat{y}_t \quad \text{و} \quad t = 1, 2, 3, \dots, n \quad (7) \dots (14).$$

حيث يتم في هذه المرحلة فحص البواقي  $e_t$  وذلك بإجراء بعض الاختبارات لنرى إذا ما كانت تحققت شروط التشويش الأبيض (الضجة البيضاء)  $white\ noise$  فإذا حققت الشروط يعتبر النموذج المطبق تم قبوله وإلا يجب اقتراح نموذج آخر كما يأتي:

1. استقلال البواقي: سيتم استخدام طريقة  $Ljung-Box$  (LB) لاختبار استقلال البواقي، كونها تعطي نتائج أفضل من إحصاء  $Box\ price$ ، فقد قدم  $Ljung-Box$  تعديلاً بسيطاً للإحصاء

$Q_{BP}$  وتكون الصيغة الرياضية لإحصاء الاختبار الاستقلال السلسلي  $Ljung-Box (LB)$  على الشكل التالي (Ljung & Box, 1978: 297-303):

$$Q_{LB} = n(n+2) \sum_{k=1}^m \frac{r_k^2(e_t)}{n-k} \quad \dots (7.15)$$

والتي تتبع أيضاً توزيع  $\chi^2$  بدرجات حرية  $m$  حيث يتم اختبار فرضية العدم الآتية:

$$H_0 = \rho_1 = \rho_2 = \dots = \rho_m = 0$$

معتمدين على الارتباطات الذاتية للبواقي، ويلاحظ أن اختبار  $Q_{LB}$  مشابه كثيراً لاختبار  $Q_{BP}$  والاختلاف بينهما أن  $Q_{LB}$  أعطى وزن قيمة  $\frac{n+2}{n-k}$  لدالة الارتباط لذاتي عند الفجوة الزمنية  $k$  وعندما تزيد قيمة حجم العينة ( $n$ ) سيقترب هذا الوزن من (1) الواحد وبالتالي تصبح قيمة  $Q_{LB}$  و  $Q_{BP}$  تقريباً متساوية بمعنى لا يكون أي اختبار منهما أفضلية على الآخر. واختبار  $BOX PRICE$  صيغته كما يأتي:

$$Q_{BP} = n \sum_{k=1}^m r_k^2 \quad \dots (7.16)$$

1. فالاختبار السابق  $Box Price$  يعمل بشكل أفضل عندما يكون حجم العينة كبيراً معقولاً .
2. طريقة حدود الثقة للارتباطات الذاتية للبواقي  $e_t$  : هي الطريقة التي تعتمد فيها على حدود الثقة للارتباطات الذاتية للبواقي  $\hat{e}_t$  والتي يجب أن تقع بين الحدين  $\left( \pm 1.96 / \sqrt{n} \right)$  باحتمال (0.95)، فإذا تحقق فهذا يدل على أن البواقي ( $Residuals$ ) تتوزع عشوائياً وأن النموذج يقدم تمثيلاً كاملاً للبيانات ويمكن استخدامه للتنبؤ، وأن الارتباطات الذاتية للبواقي تتوزع طبيعياً بوسط حسابي صفر وتباين  $(1/n)$  .
3. اختبار التوزيع الطبيعي للبواقي: يمكننا اختبار التوزيع الطبيعي للبواقي بالاعتماد على شكل المدرج التكراري للبواقي.

#### 5.3.4.7. توفيق النموذج الأعلى مباشرة *over vitting*

يتم رفع كفاءة النموذج الذي تم تشخيصه والتعرف عليه بإضافة معلمة جديدة، حيث يتم توفيق أعلى درجة للبيانات ودراسة التحسن الناتج في نتائج الفحوص الاختبارية، ودراسة معنوية المعلمة المضافة، كذلك دراسة الارتباط بين تقدير المعلمة الجديدة والمعلمت الأساسيّة للنموذج، فإذا وجد أن المعلمة المضافة غير معنوية وأن معامل الارتباط بين تقدير المعلمت قوي فإنه يجب حذف المعلمة المضافة والاكتفاء بالنموذج الأصلي (شعراوي، 2005).

#### 6.3.4.7. توفيق النموذج الأدنى مباشرة *Under fitting*

قد يشتمل النموذج على معلمة ذات درجة عالية غير مهمة، ومن ثم يمكن تبسيط النموذج إلى نموذج الأصغر بحذف هذه المعلمة، ولابد من دراسة اختلاف المعلمة ذات الدرجة العليا معنوياً عن الصفر بمقارنة تقديرها بضعف الخطأ المعياري لهذا التقدير، فإذا كان تقديرها أقل من ضعف الخطأ المعياري فقد يكون من الأفضل حذف هذه المعلمة، ولكن قبل حذفها لابد من دراسة معامل الارتباط بين مقدر هذه المعلمة ومقدر كل معلمة من المعلمات الأخرى. فإذا شاهدنا وجود ارتباط قوي بين مقدر هذه المعلمة ومقدر إحدى المعلمات الأخرى فقد يكون هذا مؤشر جيد على توفيق النموذج الأدنى له مباشرة (شعراوي، 2005).

#### 7-3-4.7. معايير اختيار النموذج

يعتبر استخدام دوال الارتباط الذاتي *ACF* والذاتي الجزئي *PACF* لتحديد درجات نموذج *ARIMA* مهمة صعبة يواجهها الباحث الذي يجد نفسه أمام عدة نماذج مرشحة ولا تكون جميعها مرفوضة، ولهذا المبرر سيتم استخدام أحد المعيارين الإحصائيين لاختيار النموذج الأفضل من بين النماذج المرشحة، والذي لديه أقل قيمة لهذين المعيارين (Matroushi, 2011) كما يأتي:

##### - معيار معلومات أكاي (*Akaike Information Criterion (AIC)*)

$$AIC = -2 \log(\text{maximum likelihood}) + 2k \quad \dots(7.17)$$

حيث  $k$  تمثل عدد معلمات النموذج وتساوي  $(p + q + 1)$  في حالة ما إذا كان النموذج يحتوي على مقدار ثابت، أما إذا كان النموذج بمعلمات  $k=(p+q)$  وفق البيانات ولا يحتوي على مقدار ثابت.

##### - معيار المعلومات البيزي (*Bayesian Information Criterion (BIC)*)

رغبة في تحقيق خصائص تقاربية ولأسباب أخرى اقترح SHWARZ طريقة أخرى معدلة لتحديد درجات نموذج *ARIMA* وهي معيار المعلومات البيزي *BIC* وتعرف العلاقة كالآتي (مولود حشمان، 2010: 199):

$$BIC = \ln \hat{\sigma}^2 + \frac{(p+q)}{T} \ln T \quad \dots(7.18)$$

حيث  $T$  تمثل حجم العينة، ويكون النموذج الأفضل هو الذي يجعل قيمة *BIC* أقل ما يمكن.

#### 4.4.7. مرحلة التنبؤ *Forecasting*

التنبؤ هو المرحلة الأخيرة من مراحل دراسة وتحليل نماذج السلاسل الزمنية، ويعد الهدف الأساسي من الدراسة، فبعد تحديد النموذج الملائم للبيانات يتم استعماله لمعرفة قيم الظاهرة المستقبلية ولفترات (L) (Douglas, 1976: 87).

ففي حالة النموذج المختلط المركب (p,d,q) ARIMA، وبافتراض أنه النموذج الملائم والذي تم التوصل إليه فإننا نستطيع التنبؤ بالقيم المستقبلية إلى  $L=1,2,\dots$  فترة قادمة وباستخدام صيغة معادلة الفروق لهذا النموذج وكما يأتي:

$$\phi(B)y_t = \theta(b)\varepsilon_t$$

حيث أن:

$$\phi_{p+d}(B) = \phi(B)(1-B)^{-1} = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p+d} B^{p+d})$$

من الدرجة p+d ويمكن كتابة المعادلة  $\phi(B)y_t = \theta(b)\varepsilon_t$  على الشكل الآتي:

$$y_t = \varepsilon_t + \phi_1 y_{t-1} + \dots + \phi_{p+d} y_{t-p-d} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

وبكتابة المعادلة السابقة عند الفترة الزمنية  $t+L$  وأخذ التوقع الشرطي عند الفترة  $t$  نحصل على أفضل نموذج تنبؤ بعدد خطوات ( $L$ ) أي أن:

$$E_t(y_{t+1}) = \hat{y}_{t+1}$$

$$\hat{\phi}_1 y_{t-1+L} + \dots + \hat{\phi}_{p+d} y_{t-p+L-d} - \hat{\theta}_1 \varepsilon_{t-1+L} - \dots - \theta_q \varepsilon_{t-q+L} + \varepsilon_{t+L} \dots (7.19)$$

#### 1.4.4.7. قياس دقة التنبؤ

تعتبر هذه المرحلة مهمة في تقييم النموذج للأهداف التنبؤية ومن المعايير المستخدمة في هذا المجال ما يأتي:

- متوسط مربعات الخطأ (*Mean Square Error (MSE)*)

ويعبر هذا المقياس على متوسط الفرق بين المشاهدة والتنبؤ لنفس الفترة الزمنية، ويعطى رياضياً في الصيغة الآتية (Makridakis & Others, 1998):

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \frac{1}{n} \sum_{t=1}^n e_t^2 \dots (7.20)$$

- متوسط مربعات الخطأ النسبي (PMSE)

$$PMSE = \frac{1}{n} \sum_{t=1}^n \left( \frac{y_t - \hat{y}_t}{y_t} \right) \dots (7.21)$$

- جذر متوسطات مربعات البواقي RMSE

جاء هذا المعيار كبديل للمعيارين السابقين MSE، PMSE ويعطى بالعلاقة الآتية (مولود حشمان، 2002: 182):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \dots (7.22)$$

التنبؤ بأعداد مرضى السرطان باستخدام منهجية بوكس وجينكز في محافظات غزة

ويعتبر معيار  $RMSE$  من أهم المعايير المستخدمة في المفاضلة بين مجموعة من النماذج على أساس أصغر  $RMSE$ .

#### 8. تطبيق منهجية Box-Jenkins للتنبؤ بأعداد مرضى السرطان بمحافظات غزة

##### 1.8. وصف البيانات

يتم جمع البيانات والتي تتألف من سلسلة زمنية شهرية تتكون من (84) مشاهدة، وتعود إلى الفترة من يناير عام 2010 إلى ديسمبر عام 2016، وأن هذه البيانات تمثل أعداد مرضى السرطان بمحافظات غزة، والمأخوذة من سجلات وزارة الصحة الفلسطينية "وحدة نظم المعلومات" بمركز المعلومات الصحية، وكما موضحة في جدول رقم (1.8).

جدول رقم (1.8): أعداد مرضى السرطان بمحافظات غزة \*

Year/ Month	2010	2011	2012	2013	2014	2015	2016
January	86	98	88	96	126	160	115
February	74	90	73	100	137	130	120
March	78	95	82	103	105	138	123
April	79	100	93	107	135	142	145
May	84	118	99	110	110	118	149
June	73	98	102	120	116	122	153
July	69	90	105	123	112	128	157
August	79	69	112	125	114	133	165
September	76	87	115	128	115	140	165
October	83	73	118	130	162	139	160
November	74	72	121	132	142	134	140
December	75	47	123	140	128	142	110
Sum	940	1037	1231	1414	1502	1626	1702

\* المصدر: وزارة الصحة الفلسطينية، مركز المعلومات الصحية - وحدة نظم المعلومات.

### شريف شحدة مصلح

وتعتبر عملية جمع البيانات وتبويبها من أهم وظائف علم الإحصاء، حيث إنه لا يمكن الاستفادة من هذه البيانات إلا إذا تم جمع البيانات وعرضها على شكل جداول، وحساب بعض المؤشرات الإحصائية البسيطة التي تدلنا على طبيعة البيانات، ولأخذ فكرة عامة عن البيانات سيتم عرض بعض الإحصاءات الوصفية لها كما يوضح الجدول الآتي:

جدول (2.8): الإحصاءات الوصفية للبيانات

القيمة	الإحصاءات
84	حجم المجتمع (N)
47	أقل قيمة
165	أكبر قيمة
112.40	الوسط
115	الوسيط
9442	مجموع المشاهدات

### 2.8. مرحلة تهيئة البيانات

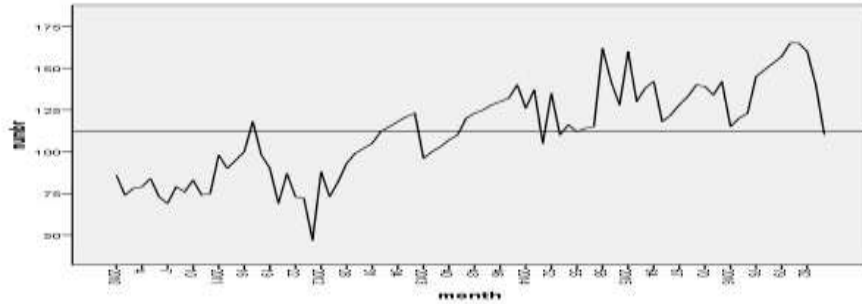
في هذه المرحلة نحضر البيانات من خلال رسم شكل الانتشار، واستخراج معاملات الارتباط الذاتي والجزئي وكذلك رسم حدود الثقة لدالة الارتباط الذاتي للبيانات الأصلية لمعرفة سلوك هذه البيانات وذلك باستخدام البرنامج الإحصائي (SPSS)، فمن خلال شكل رقم (1.8) نلاحظ تغير التشتت حول مستوى السلسلة وبذلك تكون غير مستقرة في التباين، كما أننا نلاحظ وجود اتجاه عام متزايد مع الزمن مما يدل على عدم استقراره بيانات السلسلة في المتوسط. وقد أكدت ذلك معاملات الارتباط الذاتي والجزئي في جدول (3.8) و (4.8) وحدود الثقة كما في الشكل (2.8)، والتي أظهرت فيه قيم معاملات الارتباط الذاتي حتى الفجوة (16) مختلفة معنوياً عن الصفر، ولكي تكون السلسلة مستقرة لابد من دخول جميع قيم معاملات الارتباط الذاتي للعينة ضمن حدود الثقة ما عدا عند الإزاحة الأولى أو الثانية، فمن الممكن أن تقع خارج حدود الثقة، حيث أن حدود الثقة للبيانات بمستوى دقة (95%) هي:  $(-0.21 \leq rk \leq 0.21)$ ، واختبار معنوية المعاملات الكلية لدالة الارتباط الذاتي باستخدام  $Ljung \& Box (Q.stat)$  فكانت قيمتها:  $Q.stat = 439.807 > \chi^2(16, 0.05) = 26.296$

وهذا ما يؤكد عدم استقرارية السلسلة الزمنية في المتوسط، لذلك نرفض فرضية العدم التي تشير تساوي معاملات الارتباط الذاتي مع بعضها ومساواتها بالصفر ونقبل الفرضية البديلة وهذا يعني أن السلسلة الزمنية غير مستقرة، ولتثبيت التباين نأخذ اللوغاريتم الطبيعي للسلسلة الزمنية وبعد أخذها نلاحظ أن التباين يميل إلى الثبات كما في الشكل (3.8)، ولمعالجة عدم الاستقرار تم أخذ الفرق



#### التنبؤ بأعداد مرضى السرطان باستخدام منهجية بوكس وجينكنز في محافظات غزة

الأول للبيانات اللوغاريتمية للسلسلة  $Y_t = Y_t - Y_{t-1}$ ، والشكل رقم (4.8) يبين منحنى السلسلة الزمنية  $[Y^*_t]$  بعد أخذ الفرق الأول للبيانات اللوغاريتمية للسلسلة الزمنية، إذ نلاحظ عدم استقرار السلسلة الزمنية، فتم أخذ الفروق الثانية لها فلاحظنا فقدان الاتجاه العام في سلوكها، مما يدل على استقرار السلسلة الزمنية في المتوسط كما في الشكل (5.8)، ولتأكيد ذلك تم رسم حدود الثقة لدالة الارتباط الذاتي والجزئي لمجتمع الدراسة، والموضحة في الشكل (6.8) و (7.8)، حيث نلاحظ أن

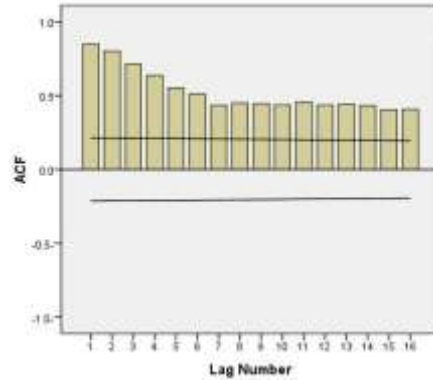


شكل (1.8): منحنى أعداد مرضى السرطان للفترة (2010-2016)

جميع الارتباطات الذاتية هي داخل حدود الثقة وأنها معنوية فقط عند الفترة الأولى والثانية، وهذا يؤكد استقرار السلسلة الزمنية في المتوسط، ويمكننا ملاحظة أيضاً عدم وجود تأثيرات موسمية في السلسلة الزمنية، وبذلك أصبحت البيانات جاهزة لتطبيق المرحلة الأولى من مراحل منهجية (Box&Jenkins, 1976) لدراسة السلاسل الزمنية وتحليلها.

Series number	Lag	Partial Autocorrelation	Std. Error
1	1	.846	.109
2	2	.294	.109
3	3	-.058	.109
4	4	-.056	.109
5	5	-.067	.109
6	6	.092	.109
7	7	-.070	.109
8	8	.238	.109
9	9	.135	.109
10	10	-.046	.109
11	11	.084	.109
12	12	-.093	.109
13	13	.005	.109
14	14	.010	.109
15	15	-.034	.109
16	16	.147	.109

جدول (4.8): قيم معاملات الارتباط الذاتي الجزئي

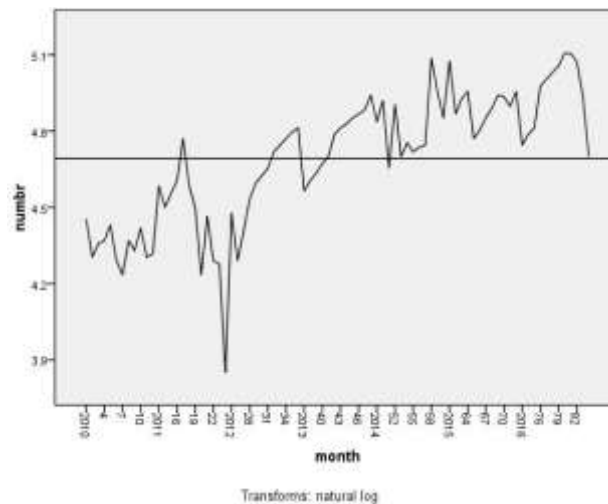


شكل (2.8): حدود الثقة لدالة الارتباط الذاتي للسلسلة الأصلية

**Autocorrelations**

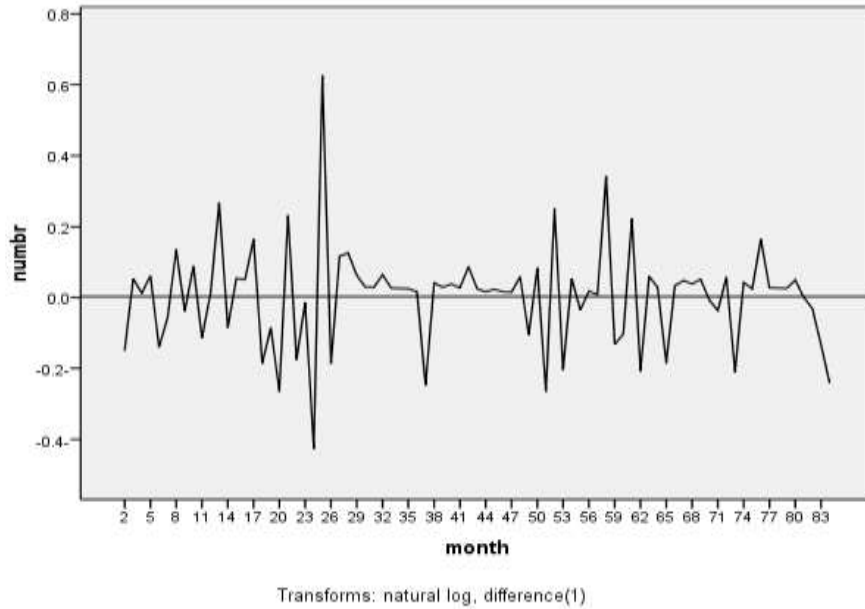
Series: numbr					
Lag	Autocorrelation	Std. Error <sup>a</sup>	Box-Ljung Statistic	Value	Sig. <sup>a</sup>
1	.848	.107	62.329	1	.000
2	.800	.107	118.670	2	.000
3	.711	.106	163.818	3	.000
4	.637	.105	200.422	4	.000
5	.553	.105	228.347	5	.000
6	.510	.104	252.452	6	.000
7	.433	.103	270.053	7	.000
8	.448	.103	289.161	8	.000
9	.445	.102	308.196	9	.000
10	.436	.101	326.713	10	.000
11	.454	.101	347.102	11	.000
12	.435	.100	365.058	12	.000
13	.442	.099	385.967	13	.000
14	.432	.098	405.194	14	.000
15	.404	.098	422.273	15	.000
16	.406	.097	439.807	16	.000

جدول (3.8): قيم معاملات الارتباط الذاتي للسلسلة الأصلية

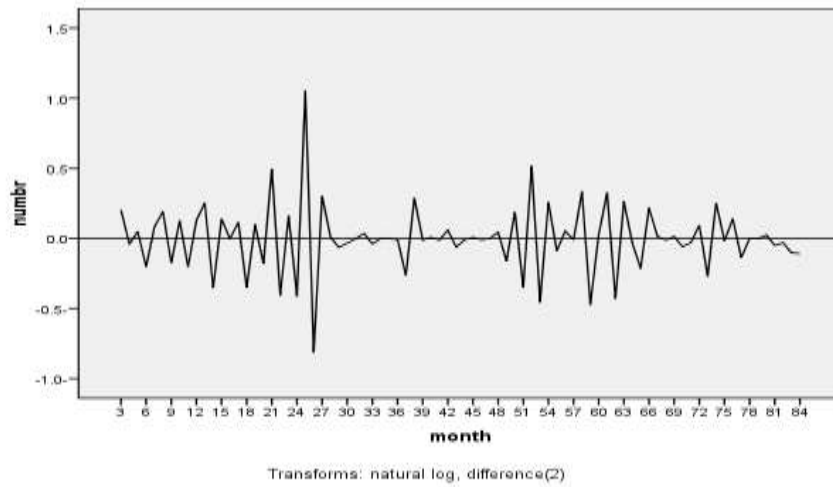


شكل (3.8): منحنى السلسلة الزمنية بعد أخذ اللوغاريتم الطبيعي

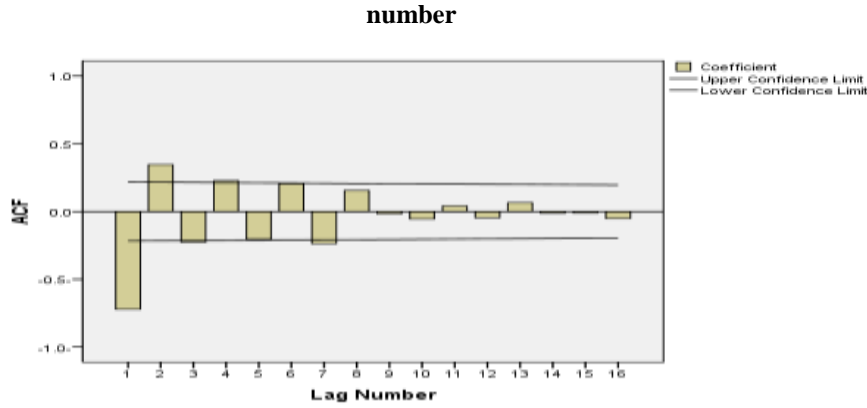
التنبؤ بأعداد مرضى السرطان باستخدام منهجية بوكس وجينكنز في محافظات غزة



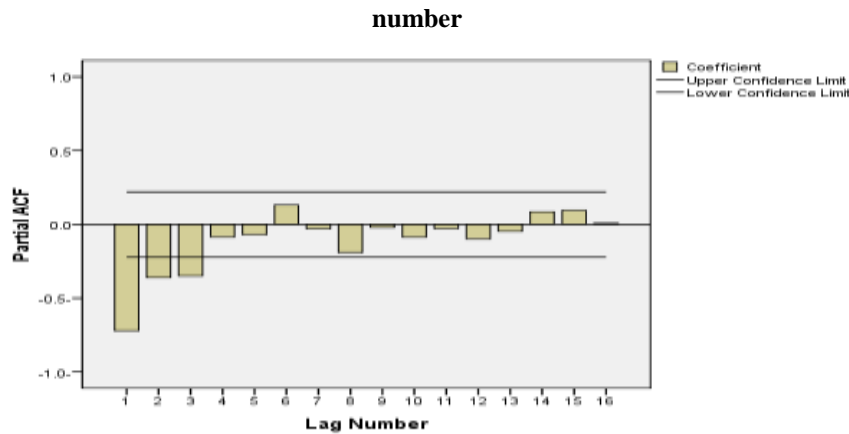
شكل (4.8): منحنى السلسلة الزمنية بعد أخذ الفرق الأول للبيانات اللوغاريتمية



شكل (5.8): منحنى السلسلة الزمنية بعد أخذ الفرق الثاني للبيانات اللوغاريتمية



شكل (6.8): حدود الثقة لدالة الارتباط الذاتي بعد أخذ الفرق الثاني للبيانات اللوغاريتمية



شكل (7.8): حدود الثقة لدالة الارتباط الجزئي بعد أخذ الفرق الثاني للبيانات اللوغاريتمية

### 3.8. مرحلة تحديد النموذج

بعد أن استقرت السلسلة الزمنية في التباين والمتوسط، نأتي بالخطوة الأولى من مراحل بناء نموذج السلسلة الزمنية وهي التعرف على النموذج المبدئي (Identification)، من خلال تحديد درجة القسم الخاص بالانحدار الذاتي AR والمعروف بالرمز (P) ودرجة القسم الخاص بالمتوسطات المتحركة MA والمعروفة بالرمز (q)، وقد تم تطبيق معايير التعرف التي تعتمد على شكل منحنى دالة الارتباط الذاتي (ACF)، وشكل منحنى دالة الارتباط الذاتي الجزئي (PACF)، وعند مطابقة قيم معاملات الارتباط الذاتي والجزئي للسلسلة الزمنية بعد أخذ الفرق الثاني للبيانات اللوغاريتمية مع السلوك النظري لها كما في الشكل (6.8) و (7.8) لوحظ منحنى دالة الارتباط الذاتي (ACF) تتناقص تدريجياً مع زيادة فترات الإزاحة k سالكة سلوك دالة الجيب المتضائلة تدريجياً، في حين

#### التنبؤ بأعداد مرضى السرطان باستخدام منهجية بوكس وجينكنز في محافظات غزة

لوحظ قطع بعد الإزاحة (الفترة) الثالثة (cut-off) لدالة الارتباط الذاتي الجزئي (PACF)، من خلال ذلك نستنتج أن النموذج الملائم هو نموذج انحدار ذاتي من الدرجة الثالثة (3,2,0) ARIMA. ولتحديد درجة الانحدار الذاتي بشكل أدق تمت مقارنة النموذج المقترح (3,2,0) ARIMA مع النماذج (2,2,0) ARIMA ، (0,2,0) ARIMA ، (1,2,0) ARIMA ، (4,2,0) ARIMA. وسيتم تقدير معالم هذه النماذج من خلال دراستها للوصول إلى أفضل نموذج لتمثيل البيانات بحيث يكون الخطأ العشوائي لهذه النماذج يتبع التوزيع الطبيعي، ومن ثم سيتم المفاضلة بين هذه النماذج حسب أقل قيمة لمعيار المعلومات البيزي (BIC)، والجدول التالي يعرض النماذج وقيم المعالم الخاصة بكل نموذج وقيمة اختبار (BIC) وهي مبينة في جدول (5.8).

جدول (5.8): نتائج اختبار BIC لنماذج ARIMA المقدرة

النموذج	Ø1	Ø2	Ø3	Ø4	Ø1	Ø2	BIC
ARIMA (3,2,0)	-1.158	-0.786	-0.395	-	-	-	5.693
ARIMA (2,2,0)	-0.998	-0.378	-	-	-	-	5.698
ARIMA (1,2,0)	-0.719	-	-	-	-	-	5.841
ARIMA (4,2,0)	-1.221	-0.912	-0.576	-0.151	-	-	5.758

مما تقدم نستنتج أن أقل قيمة لمعيار المفاضلة BIC المبينة أعلاه يحملها النموذج ذو الدرجة الثالثة، وبالتالي فإن نموذج الانحدار الذاتي المتكامل (3,2,0) ARIMA هو النموذج الملائم للبيانات وصيغته الرياضية:

$$Y_t = 2y_{t-1} - y_{t-2} - 1.158(y_{t-1}) - 0.786(y_{t-2}) - 0.395(y_{t-3}) + \varepsilon_t$$

#### 4.8. مرحلة تقدير النموذج

بعد التحقق من ملائمة النموذج، تأتي الخطوة التالية من مراحل بناء نموذج السلاسل الزمنية هي تقدير معالم النموذج (3,2,0) ARIMA الذي تم اختياره كنموذج مقترح لسلسلة أعداد المصابين بأمراض السرطان في محافظات غزة، وبتطبيق طريقة المربعات الصغرى الاعتيادية (Least squares ordinary) على بيانات السلسلة وباعتماد على البرنامج الإحصائي (SPSS)، تم الحصول على النموذج المقدر الذي يحتوي على ثلاثة معالم Ø1 و Ø2 و Ø3 وتم حساب تقدير هذه المعلمة كما في الجدول (6.8).

جدول (6.8): نتائج تقدير معالم النموذج

المعلمة المقدرة	التقدير	الخطأ المعياري	T-قيمة	P-value
$\phi_1$	-1.158	.105	-11.049	0.000
$\phi_2$	-.786	.144	-5.466	0.000
$\phi_3$	-.395	.105	-3.777	0.000

ويتبين من خلال الجدول أن قيمة P-value التي تساوي 0.000 أقل من 0.05 عند مستوى دلالة 5% مما يدل على أن تقدير  $\phi_1$  و  $\phi_2$  و  $\phi_3$  يختلف معنوياً عن الصفر.

#### 12-5 : مرحلة اختبار دقة النموذج

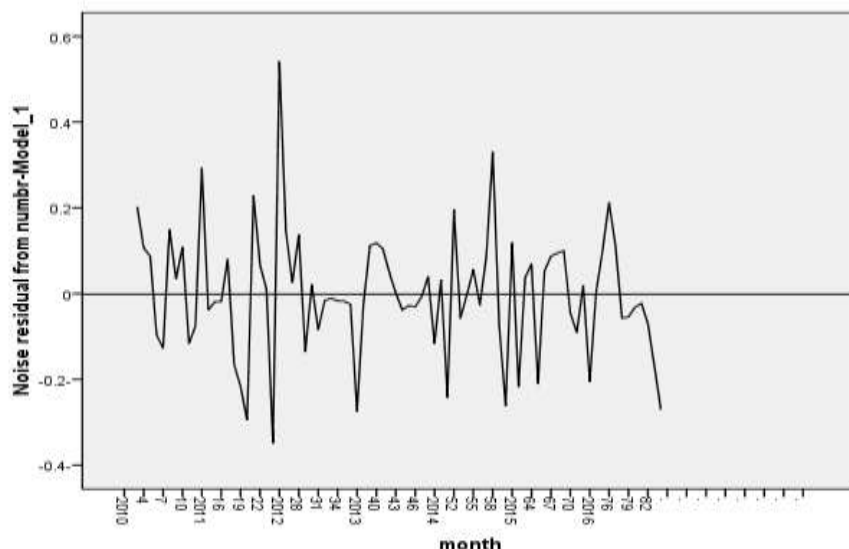
بعد أن تم تحديد النموذج الملائم لبيانات السلسلة وتقدير معالمه، سنقوم باختبار ملائمة فروض هذا النموذج النظرية لبيانات السلسلة الزمنية بهدف تحسينه أو الإبقاء عليه، وتعتبر هذه المرحلة من أهم مراحل التحليل الحديث. وقد أجريت بعض الاختبارات لتقويم النموذج منها تحليل الاستقرار والانعكاس وتحليل البواقي ومن ثم توفيق النموذج الأعلى والأدنى مباشرة.

#### 12-5-1 : تحليل الاستقرار والانعكاس

بالنظر للنموذج ARIMA (3,2,0) المقترح، حيث نجد أنه يحتوي على ثلاثة معالم انحدار ذاتي أقل من الواحد الصحيح  $\phi_1 = -1.158$  و  $\phi_2 = -.786$  و  $\phi_3 = -.395$ ، حيث أن النموذج المقدر يحقق شروط الاستقرار حيث تقع جذور المعادلة  $\phi(p)(L) = 0$  خارج حدود دائرة الوحدة أي أن تنحصر معاملات النموذج ما بين  $-1 < \phi_p < 1$ ، كما أن جميع سلاسل الانحدار الذاتي تكون دائماً منعكسة وهذا يعني أن النموذج المقترح يحقق شرطي الاستقرار والانعكاس.

#### 12-5-2 : تحليل البواقي

يعتبر تحليل البواقي  $e_t$  جزئية أساسية لمعرفة صلاحية النموذج المستخدم للتنبؤ، ويجب أن تحقق الشروط النظرية المتعلقة بالأخطاء الحقيقية  $\epsilon_t$  وهي تتأرجح بتباين ثابت حول الصفر وتكون طبيعة التوزيع، وأن لا يوجد ارتباط ذاتي بينها. أولاً: تجانس التباين حول الصفر: وتعتبر أول خطوة من خطوات تحليل البواقي، وعن طريق الرسم البياني لسلسلة البواقي (e) كما في شكل (8.8) نلاحظ أن البواقي تتأرجح بصورة ثابتة حول الوسط صفر.



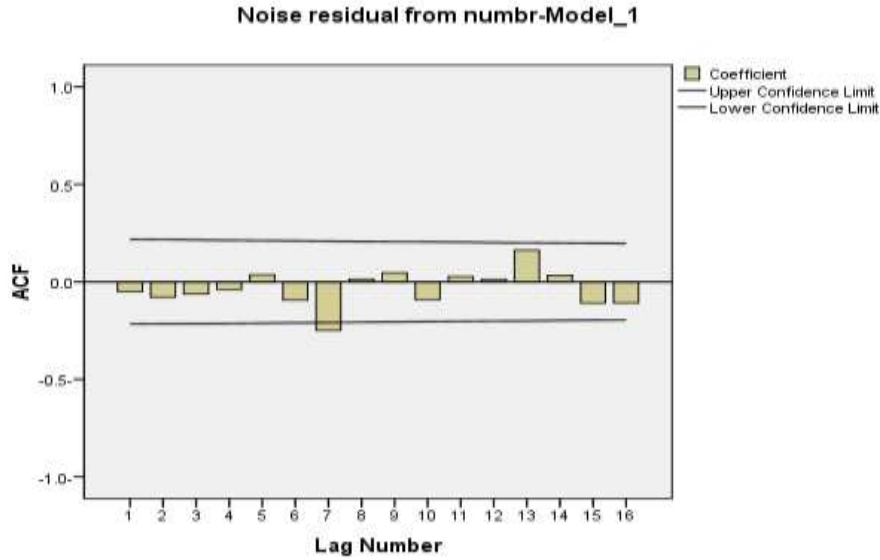
شكل (8.8): رسم البواقي لنموذج ARIMA (3,2,0)

ثانياً: استقلال البواقي

لاختبار عشوائية سلسلة البواقي تم استخراج معاملات الارتباط الذاتي للبواقي المقدر، وكما هو موضح في الجدول (7.8) والشكل (9.8) أن جميع معاملات الارتباط الذاتي لسلسلة البواقي  $r_k(e)$  تقع ضمن حدود الثقة:  $(e) \leq 0.20$ ، ولغرض التأكد من ملائمة النموذج تم تطبيق إحصاءة الاختبار Ljung & Box ( $Q.stat$ ) ومن الملاحظ أن القيمة المحسوبة  $14.058 = 16$  وهي أقل من القيمة الجدولية  $26.296 = \chi^2_{(16,0.05)}$ ، ونجد أن  $p\text{-value} > 0.05$  عند جميع الفجوات (1, 4, 8, 12, 16)، وهذا يعني قبول فرضية العدم التي تقوم بأن البواقي عشوائية مستقلة (White Noise)، وبالتالي فإن النموذج ARIMA (3,2,0) هو النموذج المناسب للبيانات.

جدول (7.8): نتائج اختبار Ljung – Box

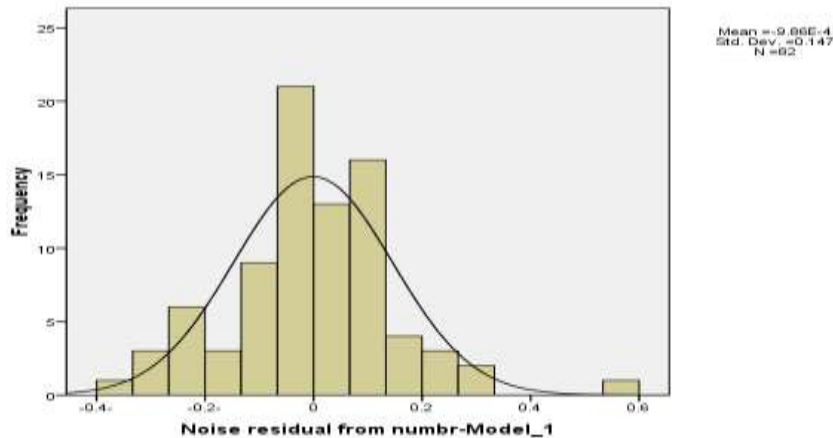
lag	value (x-squared)	DF	p – value
1	0.210	1	0.647
4	1.244	4	0.871
8	7.853	8	0.448
12	8.920	12	0.710
16	14.058	16	0.594



شكل (9.8): رسم حدود الثقة للبواقي المقدرة لنموذج  $ARIMA(3,2,0)$

ومن الشكل (9.8) نلاحظ أن جميع معاملات الارتباط الذاتي تقع داخل حدود الثقة باستثناء قيمة واحدة ويمكن إهمال ذلك، حيث من المتوقع عملياً وجود أحد المعاملات خارج حدود الثقة. ثالثاً: طبيعة التوزيع للبواقي

بهدف التأكد من معنوية معلمة النموذج واحتساب حدي الثقة للتنبؤ، تم التأكد أن الأخطاء (e) تتوزع طبيعياً، وذلك من خلال تمثيل سلسلة بواقي النموذج المقدر بيانياً على شكل المنحنى المدرج التكراري كما في الشكل (10.8) والذي يوضح فيه بأن توزيع البواقي طبيعي.



شكل (10.8): التوزيع الطبيعي لبواقي النموذج المقدر



### 12-5-3 : توفيق النماذج الأعلى والأدنى مباشرة

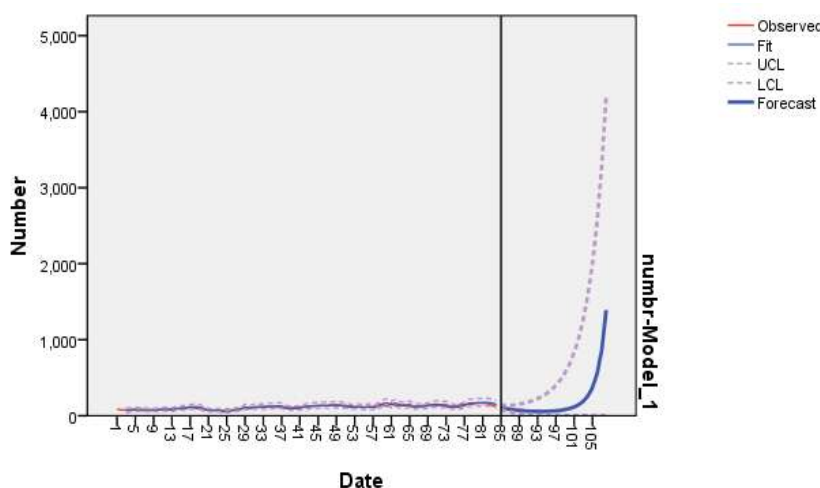
بمقارنة النموذج  $ARIMA(3,2,0)$  بالنماذج الأعلى مباشرة والأدنى مباشرة حسب المعيار الإحصائي (BIC) نجد أن نموذج  $ARIMA(3,2,0)$  هو أفضل هذه النماذج لأن لديه أقل قيمة للمعيار الإحصائي (BIC) كما في الجدول (8.8)، وهذا يعني أن إضافة أو طرح معلمة من النموذج لم ترفع كفاءة النموذج بل بالعكس عملت على الحد من كفاءة النموذج.

جدول (8.8) : المعيار الإحصائي للنماذج الأعلى والأدنى مباشرة للنموذج المقترح

النموذج	المعيار الإحصائي	BIC
النموذج المقترح	AMIRA (3,2,0)	5.693
الأدنى مباشرة	ARIMA (2,2,0)	5.698
	ARIMA (1,2,0)	5.841
الأعلى مباشرة	ARIMA (4,2,0)	5.758

### 12-6 : مرحلة التنبؤ

تم في هذه المرحلة استخدام نموذج الانحدار الذاتي المتكامل  $ARIMA(3,2,0)$  الملائم للبيانات بصيغته الرياضية المبينة في مرحلة تحديد النموذج بند (3.8)، وذلك للتنبؤ بأعداد مرضى السرطان في محافظات غزة للفترة (2017-2018) والتي عرضت نتائجه في جدول (9.8)، كما وتم تمثيل السلسلة الزمنية لهذه التنبؤات في شكل (11.8) والتي أظهرت ازدياد مضطرب بأعداد مرضى السرطان في محافظات غزة نهاية 2017 وسنة 2018.



شكل (11.8): منحني القيم الحقيقية مقابل القيم التنبؤية والمقدرة لنموذج  $ARIM(3,2,0)$  لسلسلة أعداد مرضى السرطان للفترة (2017-2018)

شريف شحدة مصلى

جدول رقم (9.8): أعدد مرضى السرطان المتنأ بها بمحافظات غزة (2017-2018)

Month/ Year	2017	2018
January	109	64
February	94	71
March	84	81
April	75	95
May	71	115
June	64	145
July	61	190
August	58	259
September	57	368
October	56	547
November	58	852
December	60	1391

12-6-1 : قياس دقة التنبؤ

تعتبر هذه المرحلة ذات أهمية في تقييم النموذج للأغراض التنبؤية حيث يعتبر جذر متوسط مربعات البواقي RMSE من أهم المعايير المستخدمة في المفاضلة بين مجموعة النماذج، حيث أظهرت بأن النموذج  $ARIMA(3,2,0)$  الذي لديه أصغر معيار RMSE وقيمته (15.891) هو أفضل نموذج للأغراض التنبؤية كما موضح في جدول (10.8).

جدول (10.8): نتائج معيار RMSE لتقييم النماذج للأغراض التنبؤية

النماذج	المعيار الإحصائي RMSE
$ARIMA(3,2,0)$	15.891
$ARIMA(2,2,0)$	16.368
$ARIMA(1,2,0)$	18.056
$ARIMA(4,2,0)$	15.985

## الاستنتاجات والتوصيات

### أولاً: الاستنتاجات

خرجت الدراسة بعدد من الاستنتاجات لعل أهمها:

- (1) نلاحظ من خلال دراسة سلسلة أعداد مرضى السرطان في محافظات غزة إنها غير مستقرة في التباين والمتوسط وأن هناك اتجاه عام واضح في السلسلة يتضح كثيراً بعد العدوان الصهيوني على محافظات غزة عام 2008 و 2012 و 2014، حيث إن ظهور تأثير الأسلحة المحرمة دولياً زاد عدد مرضى السرطان إضافة إلى نقص الأدوية والأجهزة الحديثة للكشف المبكر عن أمراض السرطان والمستلزمات الطبية.
- (2) تم تحقيق استقرارية السلسلة الزمنية بعد ميل التباين للثبات بأخذ اللوغاريتم، ومن ثم أخذ الفرق الثاني للبيانات اللوغاريتمية، وبعد مطابقة معاملات الارتباط الذاتي والجزئي للسلسلة الزمنية مع السلوك النظري لدالتي الارتباط الذاتي والجزئي، فقد اتضح أن دالة الارتباط الذاتي (ACF) تتناقص تدريجياً مع زيادة فترات الإزاحة  $k$  سالكة سلوك دالة الجيب المتضائلة تدريجياً، في حين لوحظ قطع في دالة الارتباط الذاتي الجزئي (PACF) للعينة بعد الإزاحة الثالثة.
- (3) باستخدام معيار المفاضلة بين عدة نماذج، وهو أقل قيمة لمعيار المعلومات البيزي (BIC) وجد أن النموذج الأمثل للبيانات هو نموذج الانحدار من الدرجة الثالثة، وقد تم التأكد من صحة التعرف على النموذج من خلال الاختبارات الإحصائية (معنوية المعلمات المقدرة، تحليل الاستقرار والانعكاس، تحليل البواقي وتوفيق النماذج الأعلى والادنى مباشرة للنموذج المقترح).
- (4) وجدنا أن النموذج الأمثل لبيانات السلسلة الزمنية هو نموذج الانحدار الذاتي المتكامل  $ARIMA(3,2,0)$ .
- (5) باستخدام النموذج الأمثل  $ARIMA(3,2,0)$  للتنبؤ بأعداد مرضى السرطان في محافظات غزة للفترة (2017-2018) أظهرت القيم التنبؤية بأنها قريبة من القيم الأصلية للسلسلة وأن هناك تزايداً مضطرباً في أعداد مرضى السرطان بمرور الزمن.
- (6) لقد استخدمنا معيار المفاضلة بين مجموعة النماذج لقياس دقة التنبؤ وهو أقل قيمة لمعيار جذر متوسط مربعات البواقي RMSE، فوجدنا أن النموذج  $ARIMA(3,2,0)$  الذي لديه أقل قيمة RMSE هو أفضل نموذج للأغراض التنبؤية.

### ثانياً: التوصيات

من خلال النتائج التي تم التوصل إليها نوصي بما يأتي:

- (1) استخدام نموذج بوكس وجينكنز الأمتل (3,2,0) ARIMA في التنبؤ بأعداد مرضى السرطان في محافظات غزة من قبل الجهات ذات العلاقة.
- (2) استخدام أساليب تنبؤية أخرى كأسلوب نماذج الشبكات العصبية الاصطناعية خاصة بوجود السلاسل الزمنية الطويلة والتي لا تتسم بالارتباط الذاتي والموسمية، واستخدام أسلوب الهجين بين الشبكات العصبية الاصطناعية ونماذج بوكس وجينكنز ومقارنة نتائج هذه الأساليب مع نتائج نموذج بوكس وجينكنز الأمتل.
- (3) مطالبة وزارة الصحة الفلسطينية بضرورة الأخذ بنتائج هذا البحث واتخاذ التدابير اللازمة للحد من هذه الظاهرة، بإخضاع سكان محافظات غزة لفحص طبي شامل سنوياً للوقاية من هذه الأمراض مع توفير الإمكانيات المادية والبشرية لذلك.

**المراجع:**

**أولاً: المراجع والمصادر العربية**

- حشمان، مولود (2010). "السلاسل الزمنية وتقنيات التنبؤ القصير المدى". د.م.ج، الجزائر.
- حشمان، مولود (2002). "نماذج وتقنيات التنبؤ القصير المدى". د.م.ج، الجزائر.
- عكاشة، محمود خالد (2002). "استخدام نظام SPSS في تحليل البيانات الإحصائية". غزة: جامعة الأزهر.
- شعراوي، سمير (2005). "مقدمة في التحليل الحديث للسلاسل الزمنية". الطبعة الأولى. جامعة الملك عبد العزيز: مركز النشر العلمي.
- فاندلر، ولتر (1992). "السلاسل الزمنية من الوجهة التطبيقية ونماذج بوكس وجينكنز". ترجمة عبد الرحمن حامد، الرياض: دار المريخ للنشر.
- كاناموس، زدون ميلر (2004). "الإحصاء للتجارين مدخل حديث". الطبعة الإنجليزية، ترجمة سلطان محمد عبد الحميد، الرياض: دار المريخ للنشر.
- تومي، صالح (1999). "مدخل لنظرية القياس الاقتصادي". الجزء الثاني. الجزائر: ديوان المطبوعات.
- الصراف، نزار مصطفى (1981). "تحليل السلاسل الزمنية باستخدام التقنية الإحصائية للتنبؤات الاقتصادية في العراق". رسالة ماجستير في الإحصاء. كلية الإدارة والاقتصاد، جامعة بغداد.
- الجبوري، وليد دهان (2010). "التنبؤ بمستوى التضخم في أسعار المستهلك الشهرية في العراق باستخدام السلاسل الزمنية ثنائية المتغيرات". رسالة ماجستير في الإحصاء. الإدارة والاقتصاد، الجامعة المستنصرية.
- طعمة، سعدية عبد الكريم (2012). "استخدام تحليل السلاسل الزمنية للتنبؤ بأعداد المصابين بالأورام الخبيثة في محافظة الأنبار". مجلة جامعة الأنبار للعلوم الاقتصادية والإدارية، المجلد 4، العدد 8.
- الطائي، فاضل عباس، والشرابي، نجلاء سعد (2009). "تمودج الانحدار المضطرب مع التطبيق". المجلة العراقية للعلوم الإحصائية، العدد 5.
- الخضير، محمد قدوري (1996). "دراسة مقارنة لطرائق التقدير والتنبؤ لبعض نماذج بوكس وجينكنز الموسمية". رسالة ماجستير في الإحصاء. كلية الاقتصاد والإدارة، جامعة بغداد.
- وزارة الصحة الفلسطينية، وحدة نظم المعلومات الصحية (2018). "سلسلة أعداد مرضى السرطان حسب الأشهر في الفترة 2010-2016". غزة: مركز المعلومات الصحية.

منظمة الصحة العالمية (2018). "مواضيع صحية: السرطان"، موقع منظمة الصحة العالمية الإلكتروني.

ثانياً: المراجع والمصادر الأجنبية

- Box, G.E.P and Jenkins, G.M. (1979). "Time series Analysis, Forecasting and control". San Francisco.
- Matroushi, saeed (2011). "Hybrid computational intelligence systems basedon statistical and neural networks methods for time series forecasting : the case of gold price". Lincoln university, United Kingdom.
- Kaiser, R. and Maravall, A. (2001). "Notes on Time series Analysis ARIMA Models and signal Extraction". Benco de Espona servicio Estudios.
- Box, G.E. and Jenkins, G.M. (1976). "Time series Analysis: Forecasting and control." Holden day, London.
- Ljung, G.M. and Box, G.E.P. (1978). "On ameasure of the lack of fit in time series models". Biometrika, n66.
- Douglas, C.M. and contreas, J.G. (1976). "Note on Forecasting with Adaptive Filtering". O.P.Q, Vol. 24, No.4.
- Makridakis, S. and wheelwright, S. and Mcgee, E. (1998). "Forecasting: Methods and Applications". 2nd ed. Jhon Wiley & sons New York USA.
- Khan, H. A. (2013) , " Forecasting of gold prices (Box Jenkins Approach) ", In ternational journal of Emerging technology and advanced Engineering, vol 3, Issue3.