التنبوء بعدد المصابين بسرطان الدم باستخدام نماذج السلاسل الزمنية من (2015-2017) في محافظة اربيل

(PP 308 - 322)

ID No. 2062

https://doi.org/10.21271/zjhs.22.4.19

برى خان عبدالله عمر

جامعة صلاح الدين – اربيل / كلية الادارة والاقتصاد / قسم الاحصاء

Pareekhan.omer@su.edu.krd

الاستلام:2017/02/20

القبول : 2018/04/24

النشر: 2018/09/10

ملخص

يهدف هذا البحث الى استخدام احدى الاساليب الاحصائية الكفوءة للتنبوء بمرض السرطان. والتي هو عبارة عن تكاثر الخلايا بشكل غير طبيعي ومنظم في الجسم لتصبح شاذة وتتكاثر دون تحكم وتخرج عن خط النمو الطبيعي، تعتبرانجاز بحث في هذا المجال لابد ان تستند على التحليل الاحصائي العلمي الذي يعتمد على الاساليب والطرق الاحصائية الكفوءة بمعايير ومقاييس كمية وعلمية عالية المعنوية، لذا تم استخدام تحليل السلاسل الزمنية للتنبؤ بعدد المصابين بسرطان الدم خلال السنة القادمة او فيما سنواجهه في المستقبل القريب و ذلك اعتمادا على البيانات الحقيقية التي اخذت من دائرة الصحة ومستشفى نانةكةلى الخاصة بسرطان الدم من (2015-2017) والهدف من استخدام تحليل الإحصائي للسلسلة الزمنية للبيانات هو فهم خصائصها الأساسية (تغيراتها ودوراتها و تذبذباتها غير المنتظمة) ومن ثم التقديروالتنبؤ بسلوك السلسلة الزمنية في المستقبل لكلا الجنسين وتم إيجاد أفضل نموذج بالاعـتماد على قيمة متوسط مربعات الخطأ (MSE) وتبين من نتائج تحليل السلسلة الزمنية لكلا الجنسين ان هناك زيادة في عدد المصابين بسرطان الدم اي تتكاثر مع مرورالزمن اعتمادا على اقل قيمة من المتوسط مربعات الخطا من نموذج المقترح $ARIMA(0,1,0)$ الاناث ونموذج المقترح $ARIMA(1,0,0)$ للذكور.

1. مقدمة

عد أسلوب تحليل السلاسل الزمنية Time Series Analysis من الأساليب الإحصائية الجديرة بالاهتمام والتي تطورت كثيراً في تحليل الكثير من الظواهر، وأصبح بالإمكان استخدامها لغرض التوقع للمستقبل فى اي مجالات الحياة. ويعتمد هذا الاسلوب على تتبع الظاهرة (أوالمتغير) على مدى زمني معين (عدة سنوات مثلاً)، عادتة تكون السلسلة الزمنية عبارة عن مجموعة من المشاهدات حول الظاهرة لفترة زمنية معينة وتتلخص أهم أهداف تحليل السلسلة الزمنية في الحصول على وصف دقيق للسلسلة الزمنية وبناء نموذج مناسب لتفسير سلوك السلسلة الزمنية وإستخدام النتائج للتنبؤ بسلوك السلسلة الزمنية في المستقبل.

2. السلاسل الزمنية Time Series

أن معظم البيانات في الاقتصاد والعلوم الطبيعية وغيرها تأخذ شكل السلاسل الزمنية وهي تلك النماذج التي تتضمن مجموعة من المشاهدات يتم جمعها لظاهرة معينة لمدة معينة في وحدة زمنية متساوية او متقطعة حسب طبيعة الدراسة و تتكون من متغيرين أحدهما توضيحي وهو متغير الزمن والأخر متغير الاستجابة وهو قيمة الظاهرة المدروسة (احمد، 2004) ويمكن التعبير رياضيا عن هذه العلاقة:

$$Y = f(t) \quad .....(1)$$

إما إذا كان هناك متغيرات أخرى إلى جانب الزمن تؤ ثر على Y يستخدم العلاقة الرياضية الآتية:

$$Y = f(t, x_1, x_2, x_3, ... x_n) \quad .....(2)$$

ويمكن تصنيف السلاسل الزمنية إلى:

- السلاسل الزمنية المستمرة Continuous Time Series

وهي السلاسل التي يتم تسجيل مشاهدتها بشكل مستمر من الزمن , ومن الممكن مشاهدة قيم السلاسل في كل لحظة من الزمن , مثل : درجة حرارة , الأسعار ,... الخ.

- السلاسل الزمنية المتقطعة Discrete Time Series

وهي تلك السلاسل الزمنية التي تأخذ المشاهدات الظاهرة ما عند نقاط زمنية سابقة لفترات متساوية , وقد تكون ساعة أو شهر أو سنة...الخ.

1.2 تصنيف السلاسل الزمنية حسب ثباتها

1.1.2 الاستقراريه Stationary

السلسلة الزمنية تكون مستقرة في حالة عدم وجود تغير في وسطها الحسابي و تباينها وكذلك اذا كان التغيرات الدورية غير موجودة و هناك نوعان من الاستقرارية:

- السلسلة الزمنية تامة الاستقرارية Strictly Stationary Time series

تتميز السلسلة الزمنية بالاستقرارية التامة إذ كان التوزيع الاحتمالي المشترك للمشاهدات $(Z_{t_1}, Z_{t_2}, \ldots\ldots Z_{t_n})$ هو نفس التوزيع الاحتمالي المشترك للمشاهدات $(Z_{t_1+k}, Z_{t_2+k}, \ldots\ldots Z_{t_n+k})$ بمعنى أن التوزيع المشترك لأي مجموعة من المشاهدات لا يتأثر بتغير الزمن:

$$fz_{(t_1)}\ldots z_{(t_n)}(Z_1,\ldots.Z_n) = fz_{(t_1+k)}\ldots z_{(t_n+k)}(Z_1,\ldots.Z_n) \quad \ldots\ldots\ldots(3)$$

اذا تحققت فيها الشروط الاتية:

1. القيمة المتوقعة (الوسط الحسابي) لـ Z_t ثابت مهما اختلف الزمن لجميع قيم t: $E(Z_t) = \mu \quad \ldots\ldots\ldots(4)$
2. تباين السلسة الزمنية Z_t ثابت مهما اختلف الزمن أي: $E\left[(Z_t - \mu)^2\right] = \sigma_z^2 \quad \ldots\ldots\ldots(5)$
3. دالة التغاير الذاتي لا تعتمد على الزمن بل تعتمد على الإزاحة (k):

$$Cov(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}) = \gamma_k \quad \ldots\ldots\ldots(6)$$

- سلسلة زمنية ضعيفة الاستقرارية Weakly Stationary Time Series

تكون السلسلة الزمنية ضعيفة الاستقرارية أوالاستقراريه من الدرجة الثانية (Second Order Stationary) إذا تحقق فيما الشرطان (4) & (6).

2.1.2 عدم الاستقراريه Non-Stationary

عندما تتغير خصائص السلسلة الزمنية مع مرور الزمن تكون السلسلة غير مستقرة، وعدم وجود الاستقرارية في السلاسل الزمنية يعني حدوث تغيرات النظامية حول الوسط و التباين مع تغيير الزمن، لذا يجب ازالة عدم الاستقرارية قبل تحليل السلسلة الزمنية لان وجوده سيعطي نتائج مضللة في التنبؤ((Chan, N. H. (2002)، يمكن تحويل السلسلة الزمنية غيرالمستقرة حول التباين إلى سلسلة مستقرة باستخدام تحويلات القوى Power Transformation وإذا كانت السلسلة الزمنية غير مستقرة حول الوسط الحسابي، فيمكن تحويلها إلى سلسلة زمنية مستقرة وذلك بأخذ الفروق Differences المناسبة للسلسلة الزمنية الأصلية (فاندل، 1992).

نفترض أن : Z_t = السلسلة الزمنية الأصلية

∇ = عامل الفرق (Difference Operator)

W_t = السلسلة الزمنية الجديدة بعد اخذ الفرق

فإن :

$$W_t = \nabla Z_t = Z_t - Z_{t-1} = (1-\beta)Z_t$$

$$\nabla = 1 - \beta$$

β : تمثل عامل الارتداد الخلفي . وإذا لم تحقق الإ ستقرارية حول الوسط بعد اخذ الفرق الأول للسلسلة الزمنية الأصلية ، نقوم بأخذ الفرق الثاني لها و كالآتي :

$$W_t = \nabla^2 Z_t = \nabla\nabla Z_t = \nabla(\nabla Z_t) = \nabla(Z_t - Z_{t-1})$$
$$= \nabla Z_t - \nabla Z_{t-1}$$
$$= Z_t - Z_{t-1} - Z_{t-1} + Z_{t-2}$$
$$= Z_t - 2Z_{t-1} + Z_{t-2}$$
$$= (1 - 2\beta + \beta^2) Z_t \qquad = (1-\beta)^2 Z_t$$

وبصورة عامة فان: $W_t = \nabla^d Z_t = (1-\beta)^d Z_t$ .....(7)

d : الحد الأدنى لعدد الفروق اللازمة لتحقيق الاستقررية حول الوسط حيث $d = 0,1,2,\ldots$.

ويمكن اختبار الاستقرارية لسلسلة زمنية معينة عن طريق استخراج قيم معاملات الارتباط الذاتي(Auto Correlation Coefficients) لتلك السلسلة، فإذا كانت السلسلة مستقرة فان قيم معاملات الارتباط الذاتي تقترب من الصفر بعد فترة الإزاحة الثانية أوالثالثة، والسلسلة غيرالمستقرة تقترب من صفر بعد عدد كبير من الفترات الزمنية وقد تصل إلى السابعة أو الثامنة.

3.1.2 دالة التغاير الذاتى Auto Covariance function

وهو مقياس للتغاير بين مشاهدتين Z_t و Z_{t+k} للسلسلة نفسها في الفترات الزمنية المزاحة (k) ، ويكتب بالشكل الآتي:

$$\gamma_k = \mathrm{E}[(Z_t - \mu)(Z_{t+k} - \mu)] = Cov(Z_t, Z_{t+k}) \quad .....(8)$$

4.1.2 دالة الارتباط الذاتي Auto Correlation function

وهو مقياس لتحديد العلاقة بين مشاهدات السلسلة نفسها في فترات زمنية مختلفة، إن الصيغة الرياضية لمعاملات الارتباط الذاتي للسلسلة الزمنية المستقرة هي (فاندل، 1992):

$$\rho_k = \frac{\mathrm{E}[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sqrt{\mathrm{E}(Z_t - \mu)^2 \, \mathrm{E}(Z_{t+k} - \mu)^2}} \quad .....(9)$$

k : الإزاحة (lag)

ρ_k : معامل الارتباط الذاتي عند الإزاحة k.

وبما أن التباين للسلسة الزمنية المستقرة ثابت ومتساوي لكل الفترات الزمنية: $\sigma_{Z_t} = \sigma_{Z_{t+k}} = E(Z_t - \mu)^2$

وان دالة : $\mathrm{E}[(Z_t - \mu)(Z_{t+k} - \mu)] = Cov(Z_t, Z_{t+k}) = \gamma_k$ ولذالك يمكن كتابة معامل الارتباط الذاتي بالصيغة الآتية:-

$$\rho_k = \frac{\mathrm{E}[(Z_t - \mu)(Z_{t+k} - \mu)]}{\mathrm{E}(Z_t - \mu)^2} = \frac{\gamma_k}{\gamma_0} \quad .....(10)$$

وان دالة الارتباط الذاتي تتصف بخصائص عدة هي :-

- $\rho_0 = 1$
- هذه الدالة هي دالة زوجية أي $\rho_k = \rho_{-k}$
- $|\rho_k| \le 1$ لكل قيم k .
- مصفوفة الارتباط الذاتي هي مصفوفة الشبة القطعي الموجب (positive semi definite) وهذا يعني أن قيم محدداتها إما موجبة او انها تساوي صفر.

5.1.2 دالة الارتباط الذاتي الجزئي Partial Auto Correlation function

يقيس درجة العلاقة بين Z_t, Z_{t-k} للسلسلة نفسها عندما يكون تأثير الفترات الزمنية الاخرى ثابتا، ويساعد في عملية تحديد رتبة الأنموذج ونوعه (Enders W. 2004)، ويتم تقدير معاملات الإرتباط الجزئي بإعتماد معاملات الإرتباط الذاتي. ويشار إلى إرتباط الذاتي الجزئي بالرموز $[\phi_{kk}]$ التي تمثل مجموعة الإرتباط الذاتية الجزئية عند فترات فاصلة فختلفة وهذه تعرف كما يلي:-

$$\phi_{kk}=\frac{|R_k^*|}{|R_k|} \quad .....(11)$$

حيث (R_k) تمثل مصفوفة الإرتباطات الذاتية ذات رتبة $(k\times k)$:

$$R_k=\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{pmatrix}$$

وان (R_k^*) هى مصفوفة (R_k) بعد استبدال المتجه الأخير بالمتجه: $\begin{pmatrix} \rho_1 \\ \rho_2 \\ : \\ \rho_k \end{pmatrix}$

وهكذا فان: $\phi_{11}=\rho_1$

$$\phi_{22}=\frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}=\frac{\rho_2-\rho_1^2}{1-\rho_1^2} \quad .....(12)$$

6.1.2 الارتباط المتزامن Cross Correlation

يقيس درجة العلاقة بين قيم سلسلة زمنية معينة مع القيم الماضية والحالية والمستقبلية لسلسلة زمنية اخرى ويستخدم في تحليل السلاسل الزمنية متعددة المتغيرات ويهدف الى ما يهدف اليه معامل الارتباط الذاتي في تحليل السلاسل الزمنية ذات المتغير الواحد، وصيغة كالاتي:

$$\rho_{xy}=\frac{\gamma_{xy}(k)}{\sqrt{\gamma_{xx}(0).\gamma_{yy}(0)}} \quad .....(13)$$

حيث: $\gamma_{xy}(k)$: تغاير الذاتي بين السلسلة X والسلسلة Y عند الازاحة k.

$\gamma_{xx}(0)$: تباين السلسلة X.

$\gamma_{yy}(0)$: تباين السلسلة Y.

فهو مقياس يقيس الاتباط الموجود بين X_t و Y_{t+k} ويتصف بهذه الخواص :

- $\rho_{xy}(k)=\rho_{yx}(-k)$
- $|\rho_{xy}(k)|\le 1$
- $\rho_{yy}(0)=1$ و $\rho_{xx}(0)=1$

3. النماذج الخطية للسلاسل الزمنية المستقرة

1.3 نموذج الإنحدار الذاتي Auto Regressive Model (AR)

الصيغة العامة للأنموذج هي : -

$$Z_t=\phi_1 Z_{t-1}+\phi_2 Z_{t-2}+\cdots+\phi_p Z_{t-p}+\varepsilon_t \quad .....(14) \quad , \quad \varepsilon_t \sim N(0,\sigma_\varepsilon)$$

حيث :

ϕ_i : تمثل معلمات الانموذج حيث ان $(i=1,2,...,p)$

وتسمى المعادلة (14) بأنموذج الانحدار الذاتي وترمز لها بارمز $AR(p)$. ويمكن كتابة الأنموذج باستخدام عامل الارتداد الخلفي Back Shit Operator :

$$\phi(B)Z_t=\varepsilon_t$$

$$\phi(B)=1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p \quad .....(15)$$

حيث : B : عامل الارتداد الخلفي .

ولغرض التحقق من شرط استقرارية الأنموذج يجب ان تكون جذور المعادلة المميزة $(\phi(B)=0)$ بقيمتها المطلقة أكبر من واحد. ويمكن أن يكون أنموذج الانحدار الذاتي ذا صيغة موسمية والسلسلة الزمنية هي التي تعيد نفسها في فترات زمنية ثابتة (العمري، 2006).

ففي حالة وجود الموسمية يسمى الأنموذج بأنموذج الانحدار الذاتي الموسمي Seasonal Auto Regressive model ، ويرمز له ب $AR(P)^s$ وصيغة كالآتي :

$$Z_t=\Phi_1 Z_{t-s}+\Phi_2 Z_{t-2s}+...+\Phi_p Z_{t-ps}+\varepsilon_t \quad .....(16)$$

حيث : P: رتبة أنمزذج الانحدار الذاتي الموسمي.

Φ_i : ترجيحات أنموذج الانحدار الذاتي الموسمي.

S: طول فترة الموسم.

ويمكن كتابة الأنموذج بواسطة عامل الارتداد الخلفي: $\Phi(B)^s Z_t=\varepsilon_t$

والارتباط الذاتي للأنموذج هو: $\rho_{ks}=\Phi^k \quad k=1,2,3,... \quad .....(17)$

2.3 أنموذج الأوساط المتحركة Moving-Average Model (MA)

الصيغة العامة للأنموذج هي : $Z_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-...-\theta_q\varepsilon_{t-q} \quad .....(18)$

وتسمى المعادلة (18) بأنموذج الأوساط المتحركة وترمز لها بالرمز MA(q). ويمكن كتابة الأنموذج باستخدام عامل الارتداد الخلفي :

$$Z_t=\theta(B)\varepsilon_t$$

حيث ان:

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q$$

إن تباين الأنموذج هو:

$$\gamma_0=Var(Z_t)=(1-\theta_1^2+\theta_2^2+...+\theta_q^2)\sigma_\varepsilon^2 \quad .....(19)$$

ودالة الارتباط الذاتي لهذا الأنموذج هي :

$$\rho_k=\begin{cases}\dfrac{-\theta_k\theta_1\theta_{k+1}+\theta_2\theta_{k+2}+...+\theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+...+\theta_q^2} & \forall k=1,2,...,q \\ 0 & \forall k\geq q+1\end{cases} \quad .....(20)$$

حيث نلاحظ ان دالة الارتباط الذاتي مساوية للصفر بعد الرتبة (q)، أي انها تنقطع بعد الفترة (q). في حين ان دالة الارتباط الجزئي تتناقص اسياً و حينما يكون الأنموذج موسمياُ ((Chan, N. H. (2002)، فانه يسمى بأنموذج الأوساط المتحركة الموسمية Seasonal Moving Average Model ويرمز له ب $MA(Q)^s$ و صيغتها :

$$Z_t=\varepsilon_t-\Theta_1\varepsilon_{t-s}\Theta_2\varepsilon_{t-2s}-\ldots-\Theta_Q\varepsilon_{t-Qs} \quad \ldots\ldots(21)$$

حيث S : طول فترة الموسم .

Q : رتبة أنموذج الأوساط المتحركة الموسمية .

Θ_i : ترجيحات أنموذج الأوساط المتحركة الموسمية.

ويمكن كتابة الأنموذج بواسطة عامل الارتداد الخلفي : $Z_t=\Theta(B)^S\varepsilon_t$

حيث :

$$\Theta(B)=1-\Theta_1B^S-\Theta_2B^{2S}-\ldots-\Theta_QB^{QS} \quad \ldots\ldots(22)$$

3.3 الأنموذج المختلط (الانحدارالذاتي-الأوساط المتحركة) $ARMA$

Auto-Regressive Moving-Average Model

إن الكثير من السلاسل الزمنية المستقرة لايمكن تمثيلها كأنموذج الأوساط المتحركة أو الانحدار الذاتي فقط، لان هذه السلاسل غالباُ ما تحتوى على خواص كلا النموذجين اي ان نموذج واحد يمكن يعطي مواصفات كلا النموذجين وهو ما يسمى بالأنموذج المختلط (الانحدار الذاتي- الأوساط المتحركة)، ويختصر إلى $ARMA(p,q)$ حيث يمثل (p) رتبة الانحدار الذاتي و يمثل (q) رتبة الأوساط المتحركة وبذلك الزوج المرتب (p,q) تمثل رتبة الأنموذج المختلط (احمد، 2004) و(Anderson, O. D. and Perryman (1982) والصيغة العامة له:

$$Z_t=\phi_1Z_{t-1}+\phi_2Z_{t-2}+\ldots+\phi_pZ_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q} \quad \ldots\ldots(23)$$

ويختصر بواسطة عامل الارتداد الخلفي إلى: $\phi(B)Z_t=\theta(B)\varepsilon_t$

ودالة الارتباط الذاتي لهذا الأنموذج هي :

$$\rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2}+\ldots+\phi_p\rho_{k-p} \quad \ldots\ldots(24)$$

وهي تتناقص اسياً. كذلك فان دالة الارتباط الذاتي الجزئي تسلك نفسه.

4.3 الأنموذج الموسمي المضاعف Multiplicative Seasonal Model

في حالات نادرة يتم دمج كل من الأنموذج $(ARMA)$ الموسمي مع الأنموذج غير الموسمي الذي يدعى بالأنموذج المضاعف ويرمز لـه بـ $ARMA(p,q)(P,Q)^s$ ، وصيغة كالآتي بواسطة عامل الارتدادج الخلفي :

$$\phi(B)\Phi(B^S)Z_t=\theta(B)\Theta(B^S)\varepsilon_t$$

ان تباين الأنموذج هو :

$$\gamma_0=(1+\theta^2)(1+\Theta^2)\sigma_\eta^2 \quad \ldots\ldots(25)$$

4. نماذج السلاسل الزمنية الخطية غير المستقرة

ان كثيراً من السلاسل الزمنية المستخدمة في المجالات التطبيقية هي غير مستقرة حول الوسط أو التباين لذا يجب إزالة عدم الاستقرارية قبل تحليلها لان وجوده سيعطي نتائج مضللة في التنبؤ، وهناك حالتان تتطلبان إزالة عدم الاستقرارية:

1. إذا كانت السلسلة الزمنية المستقرة حول الوسط، ويمكن إزالة عدم الاستقرارية بواسطة أخذ عدد معين من الفروقات اي اضافة (Integrated) الى اسم الانموذج السلسلة الأصلية ((Anderson, O. D. and Perryman (1982):

$$W_t=\nabla^dZ_t \quad \ldots\ldots(26)$$

2. إذا كانت السلسلة الزمنية غير المستقرة حول التباين، يمكن إزالة عدم الاستقرارية في هذه الحالة بواسطة التحويلات . ومن أبرزها :

i. التحويلات اللوغاريتمية Logarithmic-Transformation

ii . تحويلات القوى Power- Transformation

iii. تحويلات الصيغة القياسية Standardization-Transformation

5. مراحل بناء الأنموذج Stages for building model

ان الهدف الرئيسي لبناء الأنموذج في السلاسل الزمنية هو إمكانية استخدام ذلك الأنموذج للتنبؤ بالقيم المستقبلية للسلسلة:

1.5 التشخيص Identification

في هذه المرحلة يتم تشخيص واختيار الأنموذج افضل الذي يمثل السلسلة الزمنية وفي هذه المرحلة يتم رسم السلسلة الأصلية للتعرف على خصائصها (الاتجاه، التغيرات الدورية، التغيرات الموسمية، ...) ثم اختبار السلسلة من حيث الاستقرارية حول الوسط و التباين ليتم بعد ذلك احتساب معاملات الارتباط الذاتي والارتباط الذاتي الجزئي (جميل، 2007) . فاذا كانت دالة الارتباط الذاتي تتناقص اسياً ومعاملات الارتباط الذاتي الجزئي تنقطع بعد الفترة (p) فان الأنموذج هو $AR(p)$. أما اذا كانت دالة الارتباط الذاتي الجزئي تناقص اسياً ومعاملات الارتباط الذاتي تنقطع بعد الفترة (q) فان الأنموذج هو $MA(q)$. ولكن اذا كان الارتباط الذاتي والارتباط الذاتي الجزئي ينحدران اسياً فهذا يعني وجود الأنموذج $ARMA(p,q)$.

2.5 التقدير Estimation

في هذه المرحلة يتم تقدير معالم الأنموذج المشخص وهناك عدة طرق ومنها ((Box G.E. & Jenkins G.M.T. (1994) :-

1. طريقة المربعات الصغرى Least-Square Method
2. طريقة الإمكان الأعظم Maximum Likelihood Method
3. طريقة يول-ولكر Yule-Walker Method

3.5 التحقق من الأنموذج Model Diagnostic Checking

بعد ايجاد تقديرات معاملات الأنموذج نقوم باعادة احتساب مشاهدات السلسلة الزمنية باستخدام الأنموذج المشخص ومن ثم احتساب البواقي واختبار هذه البواقي للتأكيد من صحة وملاْءمة الأنموذج ويتم ذلك بطرق مختلفة منها.

1.3.5 اختبار الارتباط الذاتي للبواقي:

عندما لاتوجد اختلافات كبيرة بين دالة الارتباط الذاتي للسلسلة المقدرة ودالة الارتباط الذاتي للسلسلة الأصلية فان هذا يدل على وملاْءمة الأنموذج, اما اذا دالتا الارتباط الذاتي لا تختلفان بشكل مميز, فاننا نقوم باحتساب الارتباط الذاتي للبواقي r_ε فاذا كانت معاملات الارتباط الذاتي للبواقي واقعة ضمن حدود الثقة فان هذا يعني ان البواقي غير نظامية أي عشوائية وبالتالي يكون الأنموذج المشخص ملائم (احمد، 2004) وحدود الثقة هي:

$$-1.96S(r_\varepsilon) \le \rho(\varepsilon) \le 1.96S(r_\varepsilon)$$

$S(r_\varepsilon)$: الانحراف المعياري للارتباط الذاتي للبواقي:

$$r_\varepsilon = \frac{\sum_{t=1}^{n-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2} \qquad .....(26)$$

$$S(r_\varepsilon) = N^{-\frac{1}{2}} \left[1 + 2\sum_{k=1}^{k-1} r_\varepsilon \right] \qquad .....(27)$$

2.3.5 اختبار حسن المطابقة Goodness of Fit Test

قام كل من (Box & Pieres 1970) بدراسة الأنموذج المختلط $ARIMA(p,d,q)$ ويفترض ان :

$$Q = n\sum_{k=1}^{m} r_k^2(\hat{\varepsilon}) \sim \chi^2_{(m-p-q)} \qquad .....(28)$$

ويمكن معرفة مدى ملاءمة الأنموذج هذا الاختبار الذي يتم وفق الفرضية التالية :

فرضية العدم H_0 : الانموذج ملاءم

الفرضية البديلة H_1 : الانموذج غير ملاءم

فاذا كان $P-value < 0.05$ لهذا الاختبار إذن نرفض H_0 أي ان الأنموذج غير ملائم ,أما إذا كان $P-value \geq 0.05$ فإننا لانرفض H_0 أي ان الانموذج ملائم.

6. التنبؤ Forecasting

يعرف التنبؤ بأنه معرفة السلوكية المستقبلية لظاهرة معينة بأقل خطأ ممكن عند مقارنتها بالواقع، وبعد إنجاز المراحل الثلاث الأولى وتحديد الأنموذج الملائم و بعد التحقق من ملاءمة الأنموذج المشخص, نقوم باستخدامه في التنبؤ بحيث يجعل متوسط مربعات الخطأ (MSE) أقل ما يمكن, أي بمعنى يجب ان يختار $\hat{Z}_t(l)$ ليصبح (MSE) أقل ما يمكن (احمد، 2004).

$$E\left[\varepsilon_t^2(l)\right] = \left[E\left(Z_{t+l} - \hat{Z}_t(l)\right)\right]^2 \qquad .....(29)$$

حيث : $\hat{Z}_t(I)$: القيمة التنبؤية .

t : الفترة الزمنية الأصلية .

وبمساواة معادلة (29) للصفر نحصل على : $\hat{Z}_t(l) = E(Z_{t+l})$ .....(30)

وبتعويض معادلة (20) في (30) فان متوسط مربعات الخطأ ل $\hat{Z}_t(I)$ يكون :

$$MSE = E\left[Z_{t+l} - E(Z_{t+l})\right]^2 = Var(Z_{t+l})$$

وبذالك نحصل على التنبؤ الأمثل عند التوقع الشرطي إلى (Z_{t+l}) تساوي $\hat{Z}_t(l)$ أي:

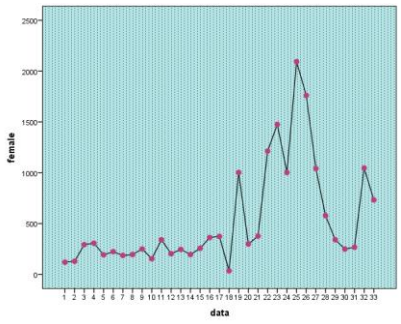
$$\hat{Z}_t(l) = E\left(Z_{t+l} \middle| Z_t, Z_{t-1} \ldots, Z_1\right)$$

7. وصف البيانات

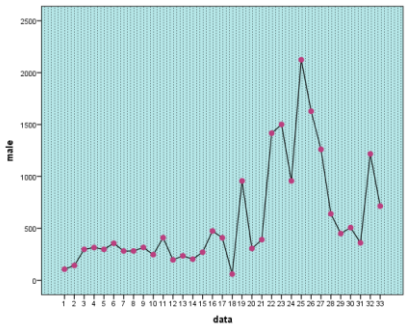
يتناول هذا المبحث الجانب التطبيقي للبحث للبيانات التي أخذت من دائرة الصحة / أربيل و مستشفى (نانه كلي) الخاصة (بالسرطان الدم) ولعدم امكانية الحصول على البيانات اكثر فقط تمت الدراسة اعتماداً على عينة بحجم (33) مشاهدة والتي تعتبر عدد المصابين لـ (33) شهراً و لكلا الجنسين ولفترة (2015 – 2017) ، لتأكد من العدد المصابين في السنوات القادمة بهذا المرض الذي تعتبر من احد اخطر أنواع الأمراض التي لا نستطيع إشتثاث هذا المرض من أصلها أوعلاجها نهائياً بسهولة اومعرفة فيما نواجه في المستقبل بهذا المرض شيء مهم جداً لذا تم إستخدام البيانات للتنبؤ بعدد المصابين بالسرطان الدم لـ(12) شهورالقادمة ، ويتم مناقشتها في ضوء الجانب النظري و تحليل البيانات كانت باعتماد على البرمجيات الجاهزة Statgraphics and SPSS version 22 .

1.7 رسم السلسلة الزمنية:

تعتبر خطوة الاساسية من تحليل السلاسل الزمنية هي رسم السلسلة الزمنية للظاهرة ومنها التعرف على بعض خصائصها التي موجودة في السلسلة.



الشكل (1) السلسلة الزمنية للاناث المصابات بسرطان الدم



الشكل (2) السلسلة الزمنية للذكور المصابون بسرطان الدم

نلاحظ من خلال الشكلين للسلسلة الزمنية لكل من الجنسين على التوالي، وجود الاختلافات في شكل تذبذبات السلاسل الزمنية وهذا يعني عدم استقراريتها حول الوسط والتباين.

2.7 اختبار العشوائية

لتاكد من معرفة مدى عشوائية السلسلة الزمنية يجب ان توضع الفرضيتين كما يتضح في الجدول (1) المعتمدة علي قيمة P-Value للاختبار Box & Pierce. واذا كانت $(P-value>\alpha=0.05)$ فان السلسلة الزمنية عشوائية مما يعني لايمكن ان نرفض فرضية العدم القائلة بان السلسلة عشوائية، ونرفضه اذا كان عكس ذلك.

الجدول (1) الفرضيات و P-Value للاختبار لكل السلسلة الزمنية

الجنس	الفرضيات	P-Value Box & Pierce
الاناث	H_0: السلسلة عشوائية	0.00175029
	H_1: السلسلة غير عشوائية	
الذكور	H_0: السلسلة عشوائية	0.00130247
	H_1: السلسلة غير عشوائية	

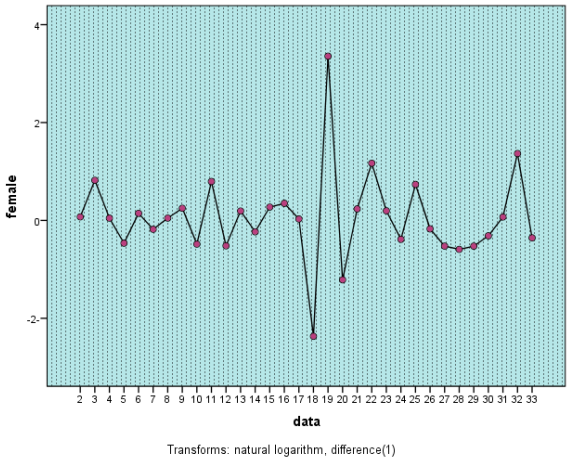
يتبين من الجدول اعلاه بان قيمة $(P-Value)$ للاختبار (Box & Pierce) لكل من السلسلة الزمنية (الاناث، الذكور) غيرعشوائية لان قيمة $(P-Value<\alpha)$ اي نرفض فرضية العدم. وبعد معرفة ان السلاسل الزمنية غيرعشوائية تم استخدام الفرق الاول لكل السلسلة لازالة اتجاه وحصول على استقرارية حول الوسط واستخدام تحويل (Log) لمحاولة الحصول على الاستقرارية حول التباين.

الجدول (2) الفرضيات وP-Value لكل السلسلة الزمنية المحولة

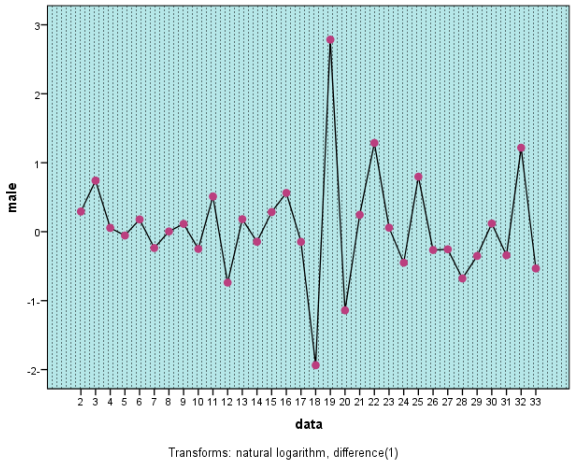
الجنس	الفرضية	P-Value Box & Pierce
اناث	H_0: السلسلة عشوائية	0.190245
	H_1: السلسلة غير عشوائية	
ذكور	H_0: السلسلة عشوائية	0.133073
	H_1: السلسلة غير عشوائية	

ويتبين من الجدول اعلاه بان استخدام الفرق الاول وتحويل (Log) محاولة جيدة للحصول على تقريب للاستقرارية وليس الاستقرارية تامة وذلك اعتمادا على قيمة $(P-Value>\alpha)$ اي بمعنى لايمكننا ان نرفض فرضية العدم.

3.7 رسم السلسلة الزمنية بعد تحويلات



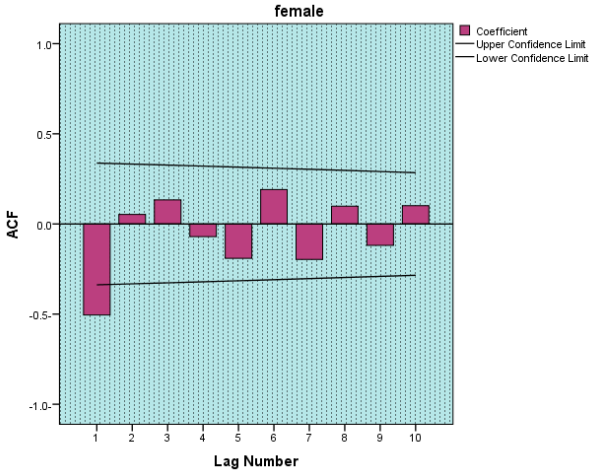
الشكل (3) السلسلة الزمنية المحولة للاناث المصابات بسرطان الدم



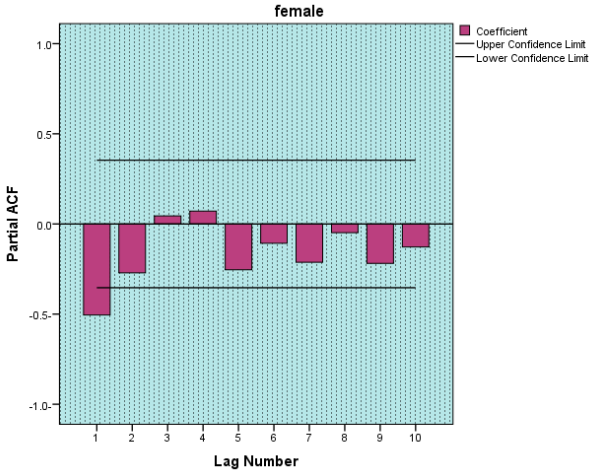
الشكل (4) السلسلة الزمنية المحولة للذكورالمصابون بسرطان الدم

1.3.7 تحقيق الاستقرارية حول الوسط

ليس النظر الى رسم السلسلة الزمنية بعد تحويلات كافيء لمعرفة مدى الاسقرارية السلسلة الزمنية للاناث، لذلك نلجا الي حصول على قيم ورسوم معاملات الارتباط الذاتي والارتباط الذاتي الجزئي، وقد تم احتساب الارتباط الذاتي للبواقييات كما ياتي:



الشكل (5) الارتباط الذاتي للسلسلة المحولة للاناث المصابات بسرطان الدم



الشكل(6) الارتباط الذاتي الجزئي للسلسلة المحولة للاناث المصابات بسرطان الدم

الجدول (3) معاملات الارتباط الذاتي وحدود الثقة للبواقيات للسلسلة الزمنية للاناث

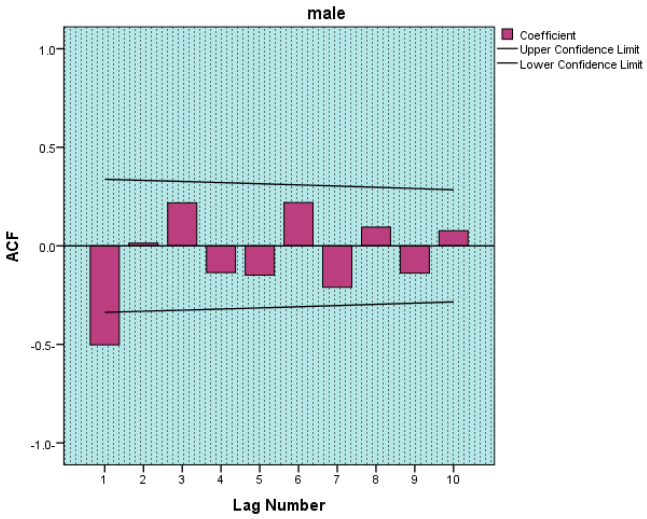
Lag	Autocorrelation	Stnd. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.504799	0.176777	-0.346477	0.346477
2	0.0527416	0.217201	-0.425707	0.425707
3	0.13377	0.217601	-0.426491	0.426491
4	-0.070546	0.220156	-0.431499	0.431499
5	-0.190722	0.220861	-0.432881	0.432881

6	0.190741	0.225949	-0.442853	0.442853
7	-0.196873	0.230926	-0.452608	0.452608
8	0.0988811	0.236113	-0.462774	0.462774
9	-0.118859	0.237404	-0.465304	0.465304
10	0.101159	0.239256	-0.468934	0.468934

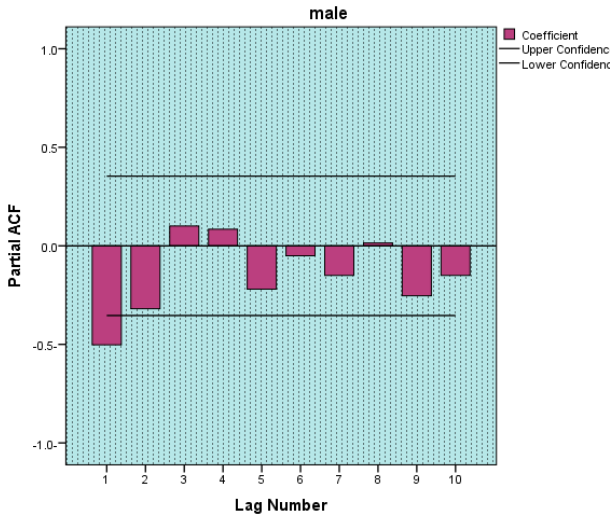
الجدول (4) معاملات الارتباط الذاتي الجزئي وحدود الثقة للبواقيات للسلسلة الزمنية للاناث

Lag	Partial Autocorrelation	Stnd. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.504799	0.176777	-0.346477	0.346477
2	-0.271184	0.176777	-0.346477	0.346477
3	0.0444979	0.176777	-0.346477	0.346477
4	0.0710269	0.176777	-0.346477	0.346477
5	-0.253908	0.176777	-0.346477	0.346477
6	-0.106859	0.176777	-0.346477	0.346477
7	-0.212349	0.176777	-0.346477	0.346477
8	-0.0481869	0.176777	-0.346477	0.346477
9	-0.21859	0.176777	-0.346477	0.346477
10	-0.127587	0.176777	-0.346477	0.346477

يتبين من الاشكال رقم (5) و(6) والجدول (3) و(4) بان جميع قيم معاملات الارتباط الذاتي و معاملات الارتباط الجزئي للبواقيات ضمن حدود الثقة ($-1.96S(r_t) \le \rho(\varepsilon) \le 1.96S(r_t)$) ماعدا فترة الازاحة الاولى.



الشكل (7) الارتباط الذاتي للسلسلة المحولة للذكور المصابون بسرطان



الشكل (8) الارتباط الذاتي الجزئي للسلسلة المحولة للذكورالمصابون بسرطان

الجدول (5) معاملات الارتباط الذاتي وحدود الثقة للسلسلة الزمنية للذكور

Lag	Autocorrelation	Stnd. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.502457	0.176777	-0.346477	0.346477
2	0.0141166	0.216862	-0.425042	0.425042
3	0.218404	0.21689	-0.425098	0.425098
4	-0.135818	0.223657	-0.438361	0.438361
5	-0.14906	0.22622	-0.443384	0.443384
6	0.219776	0.229269	-0.44936	0.44936
7	-0.210643	0.235761	-0.462083	0.462083
8	0.0954881	0.24157	-0.47347	0.47347
9	-0.138587	0.242747	-0.475776	0.475776
10	0.0765625	0.245207	-0.480598	0.480598

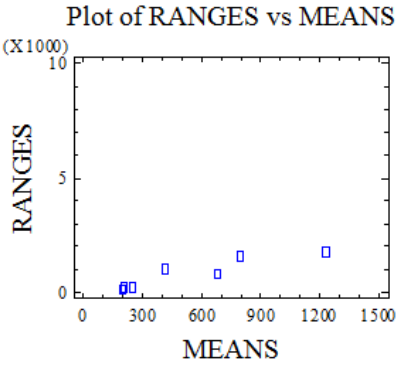
الجدول (6) معاملات الارتباط الذاتي الجزئي وحدود الثقة للسلسلة الزمنية للذكور

Lag	Partial Autocorrelation	Stnd. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.502457	0.176777	-0.346477	0.346477
2	-0.318842	0.176777	-0.346477	0.346477
3	0.100596	0.176777	-0.346477	0.346477
4	0.0842563	0.176777	-0.346477	0.346477
5	-0.220074	0.176777	-0.346477	0.346477
6	-0.0505451	0.176777	-0.346477	0.346477
7	-0.150532	0.176777	-0.346477	0.346477
8	0.0144222	0.176777	-0.346477	0.346477
9	-0.25378	0.176777	-0.346477	0.346477
10	-0.150309	0.176777	-0.346477	0.346477

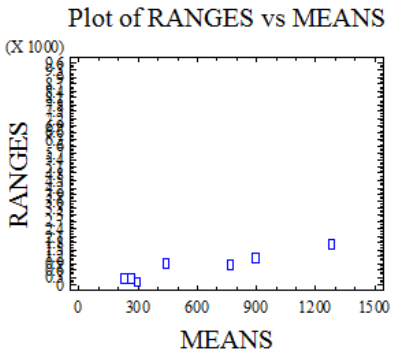
يتبين من الاشكال رقم (7) و(8) والجدول (5) و(6) بان جميع قيم معاملات الارتباط الذاتي و معاملات الارتباط الجزئي ضمن حدود الثقة ($-1.96S(r_t) \le \rho(\varepsilon) \le 1.96S(r_t)$) ماعدا فترة الازاحة الاولى.

2.3.7 تحقيق الاستقرارية حول التباين

يتبين من اشكال ادناه بان حصلنا على الاستقرارية حول التباين عند استخدام تحويل (Log) ولمعرفة مدى استقرارية السلسلة حول التباين وملاءمة نلجاء الى رسم (المدى والمتوسط) حيث نقسم السلسلة الى (6) مجموعات جزئية على اساس ان طول كل مجموعة جزئية (6) مشاهدات ويتم احتساب المدى والمتوسط لكل مجموعة وبعدها رسم (المتوسط والمدى) عندها نجد ان البيانات منتشرة تقريباً عشوائياً حول المحور الافقي، مما يدل على استقرارية السلسلة الزمنية حول التباين.



الشكل (9) تشتت السلسلة الزمنية المحولة للاناث المصابين بسرطان الدم



الشكل (10) تشتت السلسلة الزمنية المحولة للذكور المصابين بسرطان الدم

4.7 اختيار الأنموذج الملائم

بعد تاكد من استقرارية السلاسل الزمنية حول المتوسط والتباين يجب تحديد الانموذج الملائم وذلك بدراسة سلوك دالتي الارتباط الذاتي والارتباط الذاتي الجزئي، لذلك تم اقتراح خمسة(5) نماذج للسلسلة الزمنية كما يتضح في الجدول الادناه ومن ثم يتم اختيار أنموذج ملائم وذلك اعتمادا على قيمة أقل متوسط مربعات الخطأ (MSE).

الجدول (7) النماذج المقترحة وقيمة MSE للسلسلة الزمنية للاناث

النماذج المقترحة	MSE
$ARIMA(0,1,0)$	276.031
$ARIMA(1,1,0)$	282.776
$ARIMA(1,0,0)$	286.697
$ARIMA(0,1,1)$	283.919
$ARIMA(1,1,1)$	284.587

يبين من الجدول (7) بان الانموذج الملائم هو نموذج الاول أي $ARIMA(0,1,0)$ حيث لها اقل متوسط مربعات الخطأ (MSE).

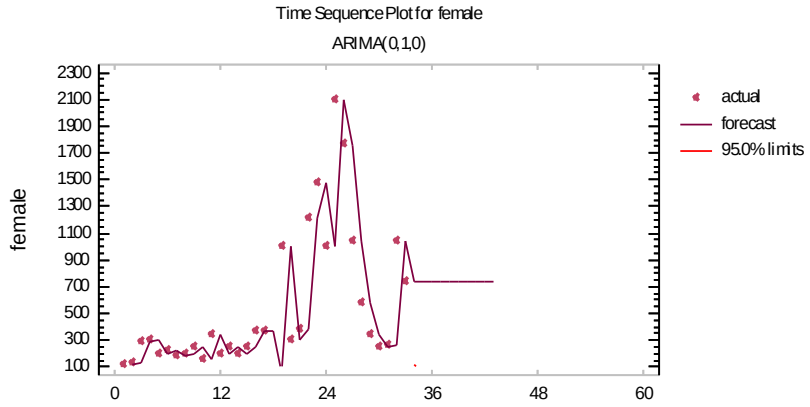
الجدول (8) النماذج المقترحة وقيمة MSE للسلسلة الزمنية للذكور

النماذج المقترحة	MSE
$ARIMA(0,1,0)$	290.375
$ARIMA(1,1,0)$	277.888
$ARIMA(0,1,1)$	268.39
$ARIMA(1,1,1)$	274.049
$ARIMA(1,0,0)$	254.416

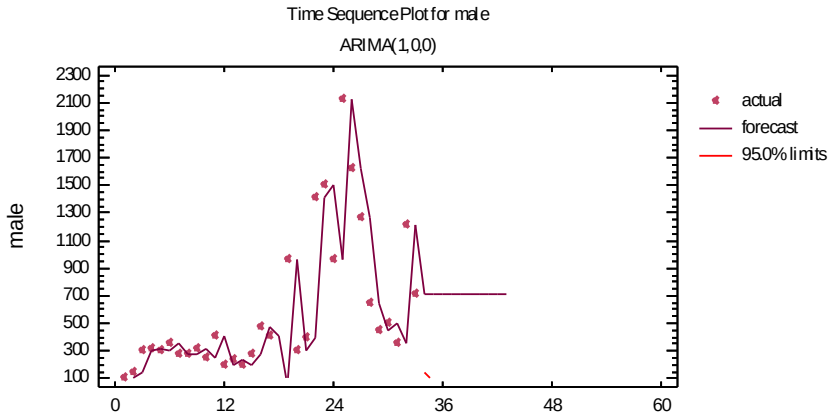
يبين من الجدول (8) بان الانموذج الملائم هو نموذج الاخير أي $ARIMA(1,0,0)$ حيث لها اقل متوسط مربعات الخطأ (MSE).

5.7 التنبؤ Forecasting

تعتبر التنبؤ المرحلة الاخيرة من المراحل التي تم اجرائها لكل من تحويل السلسلة الزمنية للذكور وللاناث المصابين بسرطان الدم وبعد ذلك تمت تحديد الانموذج الملائم من خلال اختيار النماذج المقترحة وحساب قيمة (MSE) وإعتماداً على اقل القيمة لهذا المقياس تعتبر نموذج ملائم، واستخدمت تلك النماذج المقترحة للتنبؤ بالقيم المستقبلية لكل السلسلة الزمنية حيث تم استخراج رسم السلسلة التنبؤية لـ(12) شهورالقادمة كما مبين ادناه.



الشكل(11) السلسلة الزمنية للقيم التنبؤية للاناث (12) شهور القادمة حسب النموذج الملائم المقترح



الشكل(12) السلسلة الزمنية للقيم التنبؤية للذكور(12) شهور القادمة حسب النموذج الملائم المقترح

8. الاستنتاجات:

1. بما ان السلسلة الزمنية الاصلية غير مستقرة حول الوسط و التباين حسب قيمة (p-value) للاختبار (Box & Pierce) لقد حولنا الى الاستقرارية من خلال استخدام (الفرق الاول، log) وهذا لايؤدي الى الاستقرارية التامة وانما قد يتقارب من حالة الاستقرارية ، ومن خلال رسم وحساب قيمة معاملات الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي للبواقيات (PACF) تاكد استقرارية حول الوسط، ورسم المدى والمتوسط (Range-mean) تاكد المشاهدات غير النظامية يعني البيانات منتشرة تقريباً عشوائياً حول المحور الافقي، مما يدل على استقرارية السلسلة الزمنية حول التباين .

2. تم استخدام (5) النماذج المقترحة لكل من الاناث والذكور وتبين ان نموذج $ARIMA(0,1,0)$ افضل الانموذج المقترح للاناث لانه لها اقل القيمة لمتوسط المربعات الخطا ($MSE = 276.031$)، وكذلك نموذج المقترح $ARMA(1,0,0)$ ايظاً تعتبر افضل النموذج للذكورلانه لها اقل قيمة لمتوسط المربعات الخطا $MSE = 254.416$.

3. استخدمت النماذج المقترحة باقل قيمة لـ(MSE) للتنبؤ بالقيم المستقبلية للسلسلتين الزمنية للاناث والذكور وتبين من خلال الشكلين (11) و(12) بان هناك تكاثر في عدد المصابين (12) شهورالقادمة.

المصادر:

1. احمد، كامران حسن (2004): استخدام نماذج السلاسل الزمنية للتنبؤ ببعض محاصيل الانتاج النباتي و الثروة الحيوانية في محافظة اربيل، رسالة الماجستير علوم في الاحصاء، قسم الاحصاء – كلية الادارة و الاقتصاد، جامع صلاح الدين – اربيل.

2. العمري، هيء انس عبد المجيد (2006): بعض طرائق الكشف عن التغذية العكسية في النماذج الحركية مع التطبيق على الانواء الجوية في محافظة نينوى، رسالة ماجستير في الاحصاء، كلية علوم الحاسبات و الرياضيات، جامعة الموصل، العراق.

3. جميل، رقية عبد القادر (2007): التنبؤ بانتاج واستهلاك الطاقة الكهربائية في مدينة السليمانية باستخدام نماذج بوكس- جنكنز، رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة السليمانية، العراق.

4. فاندل، والتر (1992): السلاسل الزمنية من الوجهة التطبيقية ونماذج بوكس- جنكنز، تعريب د. عبد المرضي حامد عزام ومراجعة د. احمد حسين هارون ، دار المريخ، الرياض، المملكة العربية السعودية.

5. Anderson, T. W (1971): The Statistical Analysis of Time Series, John Wiley& Sons, Inc, New York, USA.

6. Anderson, O. D. and Perryman (1982): Applied Time Series Analysis, North Holland – Publishing Company, New York, USA.

7. Box G. E. & Jenkins G. M. T. (1994): Time Series Analysis: Forecasting and Control, 3ed edition, Printice-Hall, Inc, Upper Saddle River, N.J. , USA.

8. Chan, N. H. (2002): Time series Applications to Finance, John Wiley & sons, Inc. New York, USA.

9. Enders W. (2004): Applied Econometric Time Series Analysis, John Wiley & Sons, Inc, Iowa state University, USA.

پوختە

شیرپەنجە بریتیە لە زیادبوونی خانەکان بەشیوەیەکی ناسروشتی وناریکی لە جەستەدا کەوایلیدەکات ناوازە بیت وەبەردەوام لەزیادبوون دەبیت کەناتوانریت کونترول بکریت وەدەردەچیت لە هیلی گەشەی سروشتی. وەیەك لە بیردۆزە نوییەکان دەلی بەهۆی بوونی ناریکیەکی جینی سادە کە ناتوانریت ئامیری بەرگری تیبینی بکات وە ئەو ناریکیە لەگەل تیپەربوونی کاتدا دەبیتە هۆی دەرچوونی ئەو خانەیە لەکونترولکردن بەم شیوەیە شیرپەنجە پەیدا دەبیت. وەئەنجامدانی توێژینەوە بەم جورە سیفات وتایبەتمەندیە پیویستی بە شیکردنەوەی زانستی هەیە کە پشت ببەستیت بە شیواز وریگاکانی ئاماری لیهاتوو بەپیوانەکان وپیوەرەکانی ژمیرەیی وزانستی بەرزی بەرجەستەیی. بویە زنجیرەکانی کاتی بەکارهینرا بوپیشبینی کردن بە ژمارەی توشبوەکان بە شیرپەنجەی خوین لە سالی داهاتوو یان ئەوەی روبەروومان دەبیتەوە لە ئایندەدا بەپشت بەستن بەداتای راستی کە وەرگیراوە لە (2015 -2017) لە بەریوەبەرایەتی تەندروستی ونەخوشخانەی نانەکەلی کە تایبەتە بە شیرپەنجەی خوین وەئامانج لەم شیکردنەوە ئامارییە بو زنجیرەی کاتی تیگەیشتنە لە تایبەتمەندیە سەرکیەکانی (گورانکاریەکان و سورانەوەکان وەلەرەلەرە ناریکیەکانی) دوای ئەوە خەملاندن وپیش بینیکردن بەهەلسوکەوتی زنجیرەکاتیەکە لە ئایندەدا بو هەردوو رەگەز وە دوزینەوەی باشترین مودیل بە پشتبەستن بەنرخی ناوەندی دووجای هەلە(MSE)، وەدەرکەوتووە لەئەنجامی شیکردنەوەی زنجیرەکاتیەکەبوهەردوو رەگەز کەزیادبوون هەیە لە ژمارەی توشبوەکان بەشیرپەنجەی خوین لەگەل تیپەربوونی کات.

Abstract

Cancer is an abnormal reproduction of cells and irregular in the body, it becomes abnormal and multiplies without controlling and out of the natural growth line, one of the modern theories says that the cause is a simple genetic defect that cannot be observed by the immune system and this defect with time causes the cell to come out of the control Then the cancer will appear. The completion of researches such as this type of specifications and characteristics must be based on the scientific analysis which depends on the efficient statistical methods with the quantitative measures and scientific high significant. Therefore, the time series have been used to predict the number of people infected the cancer over the next year or what we face it in the sooner future depending on the real data from (2015-2017) that taken from the department of the health and Nanakaly Hospital. The aim of using the statistical analysis of the time series analysis is to understand its basic characteristics (changes, cycles and irregular fluctuations), estimate and predict the behavior of the time series in the future for both of (Female and Male) and finding the best model by relying on the value of mean squares error (MSE), the results of time series analysis for both (Female and Male) shows that there is increased in the number of infected or multiplies over the time.