دراسة السلسلة الزمنية لعدد حالات الإصابة بمرض سرطان الرئة في مدينة الموصل (١٩٨٠-١٩٩٠)

عبد الغفور جاسم سالم

قسم الرياضيات، كلية التربية للبنات، جامعة تكريت، تكريت، جمهورية العراق

الملخص:

تم في هذا البحث دراسة السلسلة الزمنية لعدد الإصابات بمرض سرطان الرئة كمتسلسلة ماركوف ، بوضع افتراضات على عدد الإصابات لصياغة المسألة وفق أنموذج متسلسلة ماركوف . تم إيجاد مصفوفة الانتقال من الدرجة (5X5) بالاعتماد على عدد الحالات . وبعد إيجاد صفات هذه السلسلة تبين أنها ثبوتية (ergodic) وتم إيجاد التوزيع المستقر (Stationary Distribution) لكل من حالتي الشفاء والإصابة وتبين أن احتمالية الشفاء ضعيفة مقارنة باحتمالية الإصابة التي تمثل أعداد المصابين بالمرض . كما تم إيجاد بعض المتغيرات ذات العلاقة لكلا الحالتين آنفتي الذكر (فترة المكوث Duration of Excursion التقاطع العلوي Upcrossing فترة الانتظار Waiting time).

المقدمة :

إن متسلسلة ماركوف(Markov-chain) هي حالة خاصة للعملية التصادفية (Stochastic Process) التي تشمل عدد من الحالات (States) بمعلمة (Parameter) زمنية . يعد العالم Markov (1907) من الرواد في وضع المفاهيم الأساسية لمتسلسلة ماركوف وطورت هذه المفاهيم من قبل العديد من الرياضيين. واستخدمت متسلسلات ماركوف في مجالات عديدة منها المجال الاقتصادي ، المجال الزراعي ، وعلم الاجتماع... حاول (1981) Battagla دراسة منسوب النهر باعتباره متسلسلة ماركوف بحالتين (حالة الجفاف و حالة الفيضان) ، وهناك دراسة للسلسلة الزمنية للأمطار كمتسلسلة ماركوف بحالتين(حالة الجفاف و حالة غزير المطر) [2] حيث هنالك دراسة منسوب النهر بثلاث حالات (الفيضان – الجفاف – الوضع الطبيعي) [5] وكذلك سلسلة الأمطار بثلاث حالات [7] ، ودراسات عديدة أخرى في شتى المجالات .

تركز اهتمامنا في هذا البحث على صياغة عدد الإصابات بمرض سرطان الرئة في مدينة الموصل كمتسلسلة ماركوف وتم إيجاد بعض الصفات الإحصائية لهذه السلسلة.

١- التحليل الإحصائي :

من خلال ملاحظة طبيعة البيانات والمؤشرات الأساسية المراد الوقوف عليها ، نجد أن البيانات ذات طبيعة غير خطية وان عدد الإصابات بهذا المرض يبلغ ذروته في عمر (49-40) سنة. درست هذه السلسلة من قبل الصفاوي 2005 باستخدام أحد النماذج الخطية (ARMA) . وكمحاولة أخرى سيتم دراسة هذه السلسلة باستخدام متسلسلة ماركوف Markov-Chain.

الخطوة الأولى للمسألة هي تعريف الخطوة (step) والحالة (state) لعدد المصابين لتستخدم في متسلسلة ماركوف . ومن ثم وضع عدة افتراضات لتشمل الحالات (States) وكيفية حركة هذه الحالات التي تمثل أعداد المرضى لتكوين مصفوفة الانتقال (transition matrix) .

في هذه المسألة عرفنا الحالة بعدد المرضى لفترة عشرة سنوات حسب البيانات المتوفرة أما الخطوة فهي زيادة عدد المرضى من فترة زمنية (عشرة سنوات) إلى أخرى.

إن حركة المرضى خلال الفئات التي تمثل كل منها عشرة سنوات باعتبارها عملية عشوائية فقد وضع عمر المريض في واحدة من هذه الفئات التالية :

جدول رقم -١- فئات الأعمار والحالات

State	Class
S_1	10-19
S_2	20-29
S_3	30-49
S_4	50 and more

وبالاعتماد على هذه الافتراضات تم الحصول على الجدول الآتي الذي يمثل عدد المرضى(ذكور وإناث) المصابين بسرطان الرئة بعد تبويبها .

جدول رقم -٢-أعداد المرضى ضمن الحالات للسنوات(١٩٨٠-١٩٩٠)

	S_1	S_2	S_3	S_4
1980	0	42	24	0
1981	1	29	21	0
1982	0	18	29	0
1983	7	27	46	0
1984	9	42	34	14
1985	0	53	55	7
1986	0	46	75	0
1987	0	53	97	0
1988	0	88	46	4
1989	0	66	125	0
1990	0	103	94	0

حيث S_i $i = 1,2,3,4$ تمثل الحالات المختلفة لمتسلسلة ماركوف (فئات العمر) .

لتكوين المصفوفة الاحتمالية الانتقالية Transition probability P) Matrix) التي هي مصفوفة الانتقال من أية حالة إلى أخرى في وحدة زمنية واحدة ، لذا نحتاج إلى بيانات تصف حركة المرضى بصورة منفردة عبر الزمن وهذه البيانات غير متوفرة لعدة أسباب منها :

١- عدم مراجعة المريض في الأدوار الأولى للمرض. ومراجعة المريض لأكثر من طبيب .

٢- تخوف المريض وإحجامه عن المراجعة .

٣- شحة الأدوية والمستلزمات الطبية .

أما المتوفر من البيانات فيعطي معلومات عن العدد الكلي للمرضى بفئات العمر المختلفة واتخذت الفترة الزمنية (سنة واحدة) كفترة أساس ملائمة لمصفوفة التحول التي يمكن تكوينها بوضع افتراضات ملائمة لوصف حركة المرضى بين فئات العمر المختلفة وهي :

١- أي مريض يصل إلى الحالة S_4 يبقى ضمنها .

٢- أي مريض يزداد عمره سينتقل إلى المستوى الذي يكون أعلى منه ، فالزيادة في عدد المرضى في أية حالة S_i تأتي من الحالة السابقة لها مباشرة S_{i-1} .

٣- التناقص في عدد المرضى ناتج عن انتقالهم إلى الحالة S_0 المتمثلة بحالة الشفاء أو الوفاة .

باستخدام هذه الافتراضات والبيانات المتوفرة لفترة سنة واحدة (فترة الأساس) جدول رقم –٣–أدناه

جدول رقم –٣– أعداد المرضى ضمن الحالات للسنوات (١٩٨٠–١٩٨١)

States	S_1	S_2	S_3	S_4
1980	0	42	24	0
1981	1	29	21	0

لذا يمكننا تبيان الحركة التقديرية للمرضى من حالة إلى أخرى خلال(١٩٨٠–١٩٨١) كما موضحة في الجدول رقم –٤–.

جدول رقم –٤ –الحركة التقديرية للمرضى للفترة (١٩٨٠–١٩٨١)

1980/1981	S_0	S_1	S_2	S_3	S_4	مجموع الصفوف
S_0	0	0	0	0	0	0
S_1	16	1	25	0	0	42
S_2	0	0	4	20	0	24
S_3	0	0	0	1	0	1
S_4	0	0	0	0	0	0
مجموع الأعمدة	16	1	29	21	0	

وبنفس الطريقة يتم تكوين الجداول الأخرى الخاصة بحركة المرضى من حالة إلى أخرى خلال (٨٢–٨٣).....(٨٩–٩٠). وعندما نجمع هذه الجداول وفقاً لقاعدة جمع المصفوفات نحصل على الجدول التالي (جدول –٥) الذي يوضح التحولات (الحركات) التقديرية للمرضى من عام (١٩٨٠) لغاية (١٩٩٠).

جدول رقم –٥–الحركة التقديرية للمرضى للفترة (١٩٨٠–١٩٩٠)

S_i / S_{i-1}	S_0	S_1	S_2	S_3	S_4	مجموع الصفوف
S_0	0	16	146	0	0	162
S_1	16	1	35	7	0	59
S_2	16	0	264	166	0	446
S_3	0	0	80	424	18	522
S_4	0	0	0	18	7	25
مجموع الأعمدة	32	17	525	615	25	

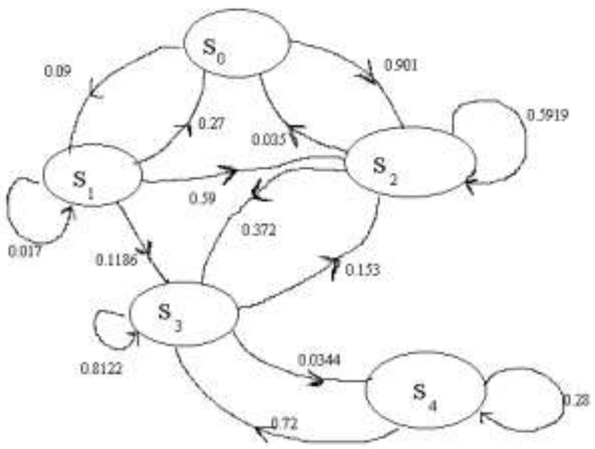
من الجدول الذي في أعلاه نقسم عناصر كل صف على المجموع الكلي للصف الذي تقع فيه ، فينتج عن ذلك مصفوفة عشوائية تستخدم كمصفوفة الانتقالات الاحتمالية للفترات (١٩٨٠–١٩٩٠) . وهي تعكس لنا الافتراضات الأولية عن حركة المرضى حيث S_0 و S_4 حالتان منتهيتان وحصلنا على الجدول الآتي (جدول –٦–) .

جدول رقم –٦– مصفوفة الانتقال

P=

	S_0	S_1	S_2	S_3	S_4
S_0	0	0.0987654	0.9012346	0	0
S_1	0.271186	0.016949	0.59322	0.11864	0
S_2	0.03587	0	0.591928	0.372197	0
S_3	0	0	0.153257	0.81226	0.034483
S_4	0	0	0	0.72	0.28

ويمكن تمثيل مصفوفة الانتقال P بمخطط شجيري (Digraph) الذي يوضح صفات متسلسلة ماركوف الممثلة بالمصفوفة P



الشكل (١) :المخطط الشجيري لمصفوفة الانتقالP

نلاحظ من المخطط الشجيري ومن تعريف الحالة المغلقة نجد أن الحالة المغلقة الوحيدة هي $\{S_0, S_1, S_2, S_3, S_4\}$ أي أن P غير قابلة للتجزئة (تكون السلسلة غير قابلة للتجزئة إذا كانت مغلقة) .

لبرهان أن P ثبوتية (Ergodic) بقي آن نبرهن إن P بدائية (Primitive) . أي علينا إيجاد القيم الذاتية $\lambda_i , i = 1,2,3,4,5$ واثبات أن إحدى هذه القيم تساوي واحد وبقية القيم اقل منها بالقيمة المطلقة ، أي إن $|\lambda_i| < \lambda_1$ (على اعتبار أن $\lambda_1 = 1$). i≠1

لإيجاد القيم الذاتية نتبع حل النظام $|P - \lambda I| = 0$ ، (لاحظ[3]).

وبحل النظام حصلنا على المعادلة الآتية:

$$\lambda^5 - 1.7011\lambda^4 + 0.6595878\lambda^3 - 0.058043\lambda^3 + 0.0362726\lambda + 0.0632823 = 0$$

وبحل هذه المعادلة حصلنا على الجذور (القيم الذاتية λ_i ، $i = 1,2,.....5$)

$\lambda_1 = 1$

$\lambda_2 = 0.936$

$\lambda_3 = 0.045 - 0.454i$

$\lambda_4 = 0.045 + 0.454i$

$\lambda_5 = -0.325$

ومـن هـذه القـيم نجـد أن $|\lambda_i| < \lambda_1$ i=2,3,4,5 . أي أن P هـي مصفوفة بدائية (Primitive) .

وبمـا أن Pغيـر قابلـة للتجزئـة (Irreducible) وبدائيـة (Primitive) نستنتج أن Pهي مصفوفة ثبوتية (Ergodic).

التوزيع المستقر Stationary distribution :

بما أن P هي مصفوفة ثبوتية (Ergodic) فانه يوجد توزيع مستقر ووحيد (لاحظ [3]) .

ولإيجاد هذا التوزيع نتبع حل النظام الآتي:

$$\pi P = \pi$$

$$\sum_{i=0}^{4} \pi_i = 1 \qquad \qquad ....(1)$$

حيث $\pi_i = \Pr(\text{State}\, i)$

وبحل النظام (١) نحصل على

$$\pi_0 = 0.010277$$
$$\pi_1 = 0.0010325$$
$$\pi_2 = 0.2786665$$
$$\pi_3 = 0.6775646$$
$$\pi_4 = 0.0324503$$

بما أن π_i تمثل احتمالية كل حالة ($(\pi_i = \Pr(\text{State}\, i)$حيث لدينا خمسة حالات هي $\{S_0, S_1, S_2, S_3, S_4\}$ ، وأن S_0تمثل حالة الشفاء و S_1, S_2, S_3, S_4تمثـل حـالات الإصـابة والممثلـة بأعـداد المصـابين بمختلف الفئات العمرية. وسنرمز لها بـ S_M أي أن S_M تمثل حالة الإصابة و S_0 تمثل حالة الشفاء . سيتم إيجاد بعض المتغيرات مثل (فترة المكوث Duration of Excursion وزمن العبور Upcrossing وزمن الانتظار Waiting time) لكل من حالتي الإصابة والشفاء .

بما أن $\Pr(\text{State } S_0) = 0.0103 = \alpha$

نحصل على أن $\Pr(\text{State}\, S_M) = 00.9897 = \beta$

١- فترة البقاء : Duration of Excursion

هي فترة مكوث المتسلسلة في كل من حالتي الشفاء S_0 والإصابة S_M ويرمز لها بـالرمز D_i حيث $i = S_0, S_M$ التوزيع الاحتمالي لفترة البقاء(المكوث) (لاحظ [7]) حيث يعرف بما يلي :

$$\Pr(\mathrm{D}_i = k) = \Pr(\mathrm{J}_j = i, 1 \le j \le k, \mathrm{J}_{k+1} \ne i / \mathrm{J}_0 \ne i, \mathrm{J}_1 = i)$$

حيث $i = S_0, S_M$ (حالتي الشفاء S_0 والإصابة S_M)

وباستخدام التعريف أعلاه للحالتين S_0 و S_M تم الحصول على النتائج التالية :

$$\left.\begin{array}{l}\Pr(\mathrm{D}_{S_0} = k) = \alpha(1-\alpha)^{k-1} \\ \Pr(\mathrm{D}_{S_M} = k) = \beta(1-\beta)^{k-1}\end{array}\right\} \qquad k = 1,2,3,.....$$

$$\mathrm{E}(\mathrm{D}_{S_0}) = \frac{1}{\alpha} \quad ; \quad \mathrm{E}(\mathrm{D}_{S_M}) = \frac{1}{\beta} \quad ; \quad \mathrm{V}(\mathrm{D}_{S_0}) = \frac{1-\alpha}{\alpha^2} \quad ; \quad \mathrm{V}(\mathrm{D}_{S_M}) = \frac{1-\beta}{\beta^2}$$

٢- الـزمن بـين عبـورين متتـاليين : Time between two consecutive crossing

لإيجـاد الفتـرة الزمنيـة بـين حـدوث حـالتين والتـي تمثـل الفتـرة الزمنيـة لتكـرار حالـة الشـفاء S_0 أو حالـة الإصـابة S_M و يرمـز لهـا بـالرمز T_i حيث $i = S_0, S_M$. التوزيـع الاحتمـالي لبقـاء الحالـة بـين عبـورين متتـاليين (انتقالين إلى حالة الشفاء S_0) يعرف كما يلي : (لاحظ [3])

$$\Pr(\mathrm{T}_{S_0} = k) = \sum_{j=2}^{k} \Pr(X_j = S_0 \quad for \quad i < j, X_M = S_M \quad for \quad j \le M \le t, X_{t+1} = S_0 / X_0 = S_M, X_1 = S_0)$$

وباستخدام التعريف حصلنا على :

$$\Pr(\mathrm{T}_{S_0} = k) = \alpha\beta\left[\frac{(1-\alpha)^{k-1} - (1-\beta)^{k-1}}{\beta - \alpha}\right], \qquad k = 1,2,3,........$$

$$\mathrm{E}(\mathrm{T}_{S_0}) = \frac{\alpha+\beta}{\alpha\beta} \qquad ; \mathrm{Var}(\mathrm{T}_{S_0}) = \frac{1-\beta}{\beta^2} + \frac{1-\alpha}{\alpha^2}$$

حيث α و β معرفان سابقا .

٣- زمن الانتظار : Waiting time

هو زمن انتظـار حالـة معينـة بعد زوالهـا ويرمـز لهـا بـالرمز W_i حيث $i = S_0, S_M$ وأن التوزيـع الاحتمـالي لـزمن الانتظـار W_i يعـرف كمـا يلي: (لاحظ [3])

$$\Pr(\mathrm{W}_i = k) = \Pr(\mathrm{J}_2 \ne i, ..., \mathrm{J}_k \ne i / \mathrm{J}_0 = i, \mathrm{J}_1 \ne i)$$

حيث $i = S_0, S_M$

وباستخدام التعريف الذي في أعلاه حصلنا على :

$$\Pr(\mathrm{W}_{S_0} = k) = \frac{\beta\pi_M}{1-\pi_0}\left[\frac{(1-\beta)\pi_M}{1-\pi_0}\right]^{k-1}$$

$$\Pr(\mathrm{W}_{S_M} = k) = \frac{\alpha\pi_M}{1-\pi_M}\left[\frac{(1-\alpha)\pi_0}{1-\pi_M}\right]^{k-1} \qquad k = 1,2,....$$

حيث $\pi_M = \beta \quad , \pi_0 = \alpha$

الاستنتاجات :

من خلال دراستنا لعدد الإصابات بمرض سرطان الرئة وتحويل المتسلسلة إلـى متسلسـلة مـاركوف تبـين أن المتسلسـلة ثبوتيـة(. وبمـا أنهـا ثبوتيـة تـم إيجاد التوزيع المستقر لهذه المتسلسلة حيث وجدنا أن احتمالية الشفاء S_0 ضـعيفة جدا مقارنـة باحتماليـة الإصـابة S_M وأن الفتـرة الزمنيـة لكـي تعـود الحالـة (Upcrossing) وخاصـة لفتـرة الشـفاء هـي ضـعيفة جـدا . بالإضـافة الـى أن احتماليـة زمـن الانتظـار لحالـة الشـفاء كبيـرة حيـث عنـدما $k = 1$ فإنها تكون 0.9897 .

المصادر :

1- Battagla, F. (1981) “Up crossing of discrete hydrologic processes and Markov chains”, J. of Hydrology, 53, 1-16.
2- Cox, D. R. and Miller, H. D. (1965) “ The Theory of stochastic Processes”, Methuen, London.
3- Dean, L. Isaacson (1976) “Markov chains Theory and application“, John Wiley and Sons, Inc.

4- Iosifescu, M. (1980) “ Finite Markov processes and their applications“ John Wiley and Sons, Inc.
5- Salim A.G. and Thanoon, B.Y. (1997) , “A Markov-chain Model for river flow “, J. science ,Kattar Univ.-6-.

٦- الصفاوي . صفاء يونس (٢٠٠٤)، " النمذجة الحصينة للسلاسل الزمنية المتعددة " . أطروحة دكتوراه . جامعة الموصل .
٧- العبيدي . د. عبد الغفور جاسم و وفاء محي الدين (٢٠٠٣) ، " دراسة السلسلة الزمنية للأمطار كمتسلسلة ماركوف " . مجلة تكريت للعلوم الصرفة، مجلد (٩). العدد (٢) .

Studying a time series of number lung’s cancer cases in Mosul City (1980-1990)

Abd AL-ghafoor Jasim Salim

Department of Mathematics, College of Education for Women, University of Tikrit, Tikrit, Iraq

Abstract:

This paper studies a time series for the number of infection of Tuberculosis of the lungs as a Markov – chain. We have made hypotheses concerning the number of infections in order to set the problem according to Markov-chain. Then I’ll try to find the characteristic of this chain that proved to be ergodic. The stationary distribution as well as the probability distribution of the some related variables are derived.