تحليل الاقتران أللامعلمي في ظل زمن الفشل لبيانات المخاطر المنافسة متعددة المتغيرات

أ.م.د. لقاء علي محمد م.م. امال صادق الموسوي (*)

كلية الادارة والاقتصاد / جامعة بغداد

المستخلص :

ان التحليل اللامعلمي لبيانات المخاطر المنافسة الثنائية ذو اهمية كبيرة في الدراسات الاجتماعية والطبية لاسيما علم الامراض الجينية اذ ان اسباب الفشل المتعددة تتفاعل مع بعضها مما يبطل التحليل اللامعلمي بسبب انتهاك فرضية استقلالية بيانات المخاطر المنافسة وهو شائع في المجتمعات التي تعتمد على بداية مؤكدة للفشل كالامراض المزمنة والاضطرابات النفسية .....الخ اذ تكون المراقبة مرتبطة مع الموت .في هذا البحث قمنا بتطوير مقدرات لامعلمية لدالة خطورة سبب- معين الثنائية ودالة الحادثة التجميعية الثنائية دون الحاجة الى وجود فرضيات حول استقلالية بيانات المخاطر المنافسة ان هذه الطريقة لم تطرح سابقا لاستعمالها بيانات المخاطرة المنافسة بدلا من بيانات اوقات الفشل الاعتيادية وان مقدرات هذه الطريقة تتصف بكونها متسقة ،موحدة وتقترب بضعف الى عمليات Gaussain

Abstract:

The non parametric analysis of bivariate competing risks data is too importance in social studies and medical studies ,especially pathologist genetic where the multi causes of failure interaction ,with the others ,violating independence hypothesis of censoring data which is common in communities that depend on the beginning of the proven failure as in chronic disease estimators where the censoring are associated with death developed non parametric estimators for the bivariate cause-specific hazard function and the bivariate cumulative incidence function without make any assumptions about the independence of the risks data. This method has not been touched upon previously used data competition rather than risk the usual times of failure data .

(*) جزء مستل من اطروحة دكتوراه للباحثة الثانية .

And estimators of this method are characterized by being consistent ,uniform and approaching weakly to Gaussain process.

الهدف من البحث The Aim of Re search :

ان الهدف الرئيسي من هذا البحث هو تقديم صياغة لبيانات ثنائية وتوسيعها الى بيانات متعددة المتغيرات في ظل تحليل لامعلمي للاقتران باستخدام طرق تجريبية في تقدير الدوال الامعلمية الاساسية في صياغة المخاطر المنافسة دون الحاجة الى وجود افتراضات لاجراء الاختبار .

مشكلة البحث The problem of Re search :

ان تقنيات تحليل البقاء الكلاسيكية هي غير ملائمة الاستخدام في بيانات المخاطر المنافسة compent risk data)) لما لهذه البيانات من خاصية انما يختار حدثا واحدا فقط لتحليل الاسباب المنافسة للفشل ،اما بقية الاسباب فهي تهمل وتعتبر مشاهدات مراقبة يمنى مما يؤدي الى تحيز في اهم طرق التقدير الا وهي طريقة كابلان – مير (kaplan-meier)) وطريقة نيلسن – الن Nelson-Alenبسبب انتهاك احدى الفرضيات الاساسية للمخاطر المنافسة الا وهي فرضية استقلال اوقات الفشل والاستقلالية بين الاسباب المختلفة للفشل مما يجعل استخدام هذه الطرق مستحيلا .

الفكرة المقترحة proposed idea :-

الحصول على مقياس استدلالي كمي قادر على تمثيل احتمالات الفشل للإصابة بمرض السكري لبيانات المخاطر المنافسة وفي نفس الوقت يوصف الاقترانات الثنائية في حالة الاصابة او الموت عند اعمار معينة للاباء والابناء ضمن العائلة الواحدة التي لها استعداد جيني للاصابة باي مرض وراثي (مرض السكري) من خلال تطوير مقدرين لامعلميين لبيانات المخاطر المنافسة احدهما هو مقدر دالة الخطورة الثنائية التجميعية بسبب الاصابة بالمرض والثاني هو مقدر دالة الحادثة (المرض)الثنائية التراكمية من دون ان يكون هناك فرضيات حول استقلالية بيانات المخاطر المنافسة .

عينة البحث sample of re search :-

لقد تم استخدام عينة تتالف من 500 فردا بشكل ازواج (125) زوجا ذكورا واناثا ولمختلف الاعمار ممن لهم سابقة ظهور مرض السكري الوراثي في العائله.وقد وجدت ان هذه البيانات تتبع توزيع كاما ذو معلمين حسب اختبار Goodness of fit و استخدمت بيانات مراقبة مستقلة من النوع الثاني يمين censoring type tow right data.

ان عينة البحث قد شملت جميع الاعمار ذكوراً واناثاً وحسب الجدول الاتي :-

العمر / الحالة		0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-	المجموع	
مصاب	ذ	28	16	33	45	41	23	20	6	216	320
	أ	16	15	15	21	24	10	7	-	108	
غير مصاب	ذ	14	19	8	8	7	6	4	1	67	180
	أ	11	27	35	13	10	8	7	2	113	
المجموع		77	91	87	87	52	38	38	9	500	

مقدمة البحث :

تظهر بيانات المخاطر المتنافسة competing risks عادة في الدراسات التي يكون فيها فشل الفرد يصنف الى واحد منk حيث ($k \geq 1$) من الاسباب المختلفة للفشل ، ولعل اهم مشكلة تواجه علم الامراض الوراثية هي في تحديد العوامل الوراثية التي تتفاوت من مريض لأخر فيحاول البحث دوما عن الاقتران بين الجينات والامراض والصفات الوراثية من جيل الى اخر ولعل اهمها تلك الامراض الوراثية المرتبطة بالموت .

وفي هذا البحث ندرس مدى اقتران مثل تلك الامراض بين افراد العائلة الواحدة ضمن جيلين هما (الاباء و الابناء) ولماذا يوجد في اشخاص وينعدم في اشخاص اخرين لهم نفس الظروف ، وما هي العوامل المسببة في تعجيل زمن الفشل(ظهور المرض) عند اشخاص في اعمار مبكرة او مبكرة جدا وباعمار متاخرة عند اشخاص اخرين ضمن نفس نطاق العائلة واجراء اختبار حول اعتمادية هذا الزمن .

تحت فرضية الاستقلال يكون توزيع متعدد المتغيرات قابل للتشخيص وبانعدام وجود هذه الفرضية تصبح بيانات مخاطر المنافسة متعدد المتغيرات غير قابل للتشخيص وغير قابلة ايضاً للتقدير لامعلميا بسبب وجود عدة اسباب مستترة أهمها عدم وجود الاستقلالية بين اوقات الفشل عندها يصبح التطبيق غير صحيح .

لذلك كان علينا صياغة بيانات ثنائية تتوسع الى بيانات متعددة المتغيرات باستخدام طرق وشكل دوال لامعلمية اساسية في صياغة بيانات المخاطر المنافسة وقادرة على تحليل الاقتران العائلي اللامعلمي لاي مرض جيني .

- **وصـف البيانـات** Data description : البيانـات تتكـون مـن n مـن الازواج أبـاء و أبنـاء في i^{th} عنقود لعينة تتوزع مشاهدتها (iid) بالشكل الاتي :

(Yi1 , ……, Yin , ηi1,……, ηin) , i =1 , …., n

مع وجود بيانات مراقبة مستقلة هي :

(C1,….,Cn) , (T1,…., Tn, ∈1 ,…. , ∈n)

اذا كان الفرد مريضا $T_1^{(j)}$ $Yj = Tj^{(1)} , \eta\, j = 1$

اذا مات الفرد دون مرض $T_2^{(j)}$ $Yj = Tj^{(2)} , \eta\, j = 2$

اذا كان الفرد على قيد الحياة وخال من المرض $Yj = cj , \eta\, j = 0$

علما ان : $Tj^{(1)}$ يمثل الاب او الام و $Tj^{(2)}$ يمثل الابن او البنت

هؤلاء الافراد هم ضمن عنقود خاضعين لنوعين من انواع الفشل j حيث ان $J = 1,2$ هما :

وقت الفشل المستتر المهم (وهو يمثل الاصابة بالمرض) $T1^{(J)}$

وقت الفشل المستتر للخطورة المنافسة (ويمثل الموت) $T2^{(J)}$

أن صيغة المشاهدة الواحدة تكون كالاتي : $TJ = \min(T1^{(J)} , T2^{(J)}) = T1^{(J)} \wedge T2^{(J)}$

$$\in j = I (Tj = T1^{(j)}) + 2\, I (Tj = T2^{(j)}) \quad \text{for } J = 1 , 2$$

حيث ان I هو مؤشر الدالة .

وبوجود بيانات مراقبة $C^{(2)}$ و $C^{(1)}$ صيغة البيانات المشاهدة ستكون بحسب الأتي :

$$sJ = I (Tj \le C^{(j)})$$

$$Yj = T1^{(j)} \wedge T2^{(j)} \wedge C^{(j)}$$

الشكل النهائي للمشاهدة في العنقود الواحد ستكون :

$$\eta\, j = \in j\ sj = I (Tj = T1^{(j)}) + 2\, I (Tj = T2^{(j)})\ \ I (Tj \le C^{(j)})$$

وكمـا ذكرنـا سـابقا بسـبب وجـود ارتبـاط اوقـات الفشـل للأفـراد سـتكون غيـر قابلـة للاختبـار كمـا ان توزيعاتهـا الحديـة هـي غيـر مشخصـة لا معلميـا ، علـى هـذا الاسـاس انصـب التركيـز حـول أي مـن الكميـات يمكـن ان تكـون قابلـة للتشـخيص لا معلميـا وايهمـا يمكـن الاسـتفادة منهـا و تكييفها والاستفادة منها بشكل كبير في تطبيق الاختبار .

يعتمد تكوين بيانات مخاطر المنافسة على دالتين اساسيتين هما :

١- دالة خطورة سبب – معين التجميعية

٢- دالة الحادثة التجميعية

والسبب في ذلك يرجع الى ان هاتان الدالتان تصفان الزمن الاول للفشل T والحالة المسببة للفشل وان صيغة دالة خطورة سبب – معين التجميعية هي :

$$\Lambda^{(j)}{}_{k}(t) = \int_0^t \lambda_k^{(j)}(s)\, d s \qquad k , j = 1 , 2$$

اما دالة الحادثة التجميعية F فتستخرج بالاعتماد على دالة الباقين على قيد الحياة اللامعلمية وهي دالة بسيدو للبقاء :

$$\text{Exp}\{-\Lambda^{(j)} k(t)\} \quad , \qquad k , j = 1 , 2$$

عندئذ يمكن استخراج $Fk^{(j)}(s)$ وكالاتي :

$$Fk^{(j)}(s) = p(Tj \le s , Ej = k) \qquad , \ k , j = 1 , 2$$

وبتوسيع الدوال السابقة الاحاديتان الى دوال ثنائية فان دالة خطورة سبب – معين التجميعية الثنائية ستكون كالاتي :

$$\Lambda kl(s , t) = \int_0^s \int_0^t \lambda kl\ (u ,v)\, du\, dv \qquad k , j = 1 , 2$$

و دالة الحادثة الثنائية التجميعية فهي :

$$Fkl(s ,t) = p(T1 \le s , E1 = k , T2 \le t , E2 = L)$$

$$= \int_0^s \int_0^t \lambda kl\ (u ,v)\, S(\bar{U},\tilde{V})\, du\, d$$

اذ ان

$$S(u ,v) = p(T1 \ge u , T2 \ge v)$$

تمثل دالة الباقين على قيد الحياة الثنائية الكلية .

1- الا ان الذي يهمنا هنا هو $\lambda 11$ و $\Lambda 11$ و $F11$ لأنها تمثل الكميات المشخصة لاقترانات سبب ضمن العناقيد وان دالة الخطورة $\lambda 11$ ودالة الخطورة التجميعية $\Lambda 11$ هي كمية العنقدة في خطورات سبب – 1الآنية عند النقطة (s , t) حيث s يمثل عمرعند الاصابة للاباء و t عمر الاصابة للابناء بشرط ان يكون الشخص على قيد الحياة عند (s , t) ، لذلك نجد إن اغلب اختبارات الاقتران يكون عن طريق F أكثر مما هي عن طريق Λ بسبب كونها الأكثراظهاراً وتشخيصا لأسباب ذلك المرض وبالوقت نفسه هي قابلة للمقارنة.الان هذه الكميات نفسها مع البيانات الثنائية تكون غير قابلة للتبديل اذ لايمكن افتراض ان

$$\Lambda_1^{(1)} = \lambda_1^{(2)} \quad \text{or} \quad F_1^{(1)} = F_1^{(2)}$$

فالتحليلات اللامعلمية هي معقدة حتى مع السبب الفردي للفشل ولحسم هذه التحديات اوجدنا اسلوبا اخر في اجراء اختبار الاقتران أولا نسبة الخطورة البديلة وهي :

$$\Phi(s , t) = \Lambda 11(s , t)\ \{ \Lambda 1^{(1)}(s) \cdot \Lambda 1^{(2)}(t) \}^{-1}$$

وهـي بالحقيقـة تمثـل المسـاهمة النسـبية للمخـاطر التجميعيـة لسـبب ـــ 1 للآبـاء والابنـاء مقسـوما على المخاطر التجميعية للسبب ـ 1 الاحادية للآباء والابناء

وهـي معرفـة جيـدا عنـد النقطـة (s,t)عمـر الاصـابة بـالمرض للآبـاء والابنـاء بسـبب إن الكميـات في Φ هي عبارة عن مجموع للخطورات ألاحادية والثنائية .

الاجراء الثاني : اعتمدنا على نسبة دالة الحادثة الثنائية التجميعية

$$\Psi(s,t) = F11(s,t)\{F1^{(1)}(s) \cdot F1^{(2)}(t)\}^{-1}$$

وهي معرفة ايضا عند نقاط الزمن لـ Φ نفسها .

اصـبح الان اختبـار اسـتقلالية بيانـات المخـاطر المنافسـة ممكنـا وبـالأخص اسـتقلالية سـبب – 1(مرض السكري الوراثي) اذ يمكننا التعرف على وجوده عن طريق λ11 عندما تكون :

$$\lambda 11(s,t) = \lambda 1^{(1)}(s) \cdot \lambda 1^{(2)}(s) \quad \text{for all } s, t$$

$$\Phi(s,t) = 1 \quad \text{for all } s, t$$

وكذلك نفس الشي فيما يخص F11 هو :

$$F11(s,t) = F1^{(1)}(s).\ F1^{(2)}(t) \quad \text{for all } s, t$$

$$\Psi(s,t) = 1 \quad \text{for all } s, t$$

إن كـل مـن Φ(s,t) و Ψ(s,t) مؤشـران علـى نسـبة الزيـادة والنقصـان للاحتمـالات المشـتركة للإصـابة بـالمرض للآبـاء والابنـاء عنـد زمـن الاصـابة بـالمرض (s,t) فعنـدما يكـون كـل مـن Φ و Ψ مسـاويان إلـى 1 الصـحيح فهـذا يعنـي ان كـل مـن (∈1 , t1) و (∈2 , t2) همـا مستقلين في النقطة (s,t) .

ولكـن و بصـورة عامـة إن تكـون Φ (s,t)=1 هـذا لا يعنـي بالضـرورة إن Ψ(s,t) =1والعكس صحيح ويرجع هذا الاختلاف بسبب الاقتران في F11 أو بسبب تأثير بقية الاقترانات القوية في λ12 ، λ22 ، λ22 وهي عكس التي في λ11 .

بعـد ذلـك وللحصـول علـى مقـدرات ل Ψ و Φ نسـتخدم طريقـة التمهيـد smoothing مـن خـلال الممهـد الخطـيLLS الـدي يعتبـر افضـل ممهـد خطـي لأنـه يتكيـف مـع كـل التصـاميم الثابتـة والعشـوائية وعـن طريـق المـلء (plug in) لــ Φ و Ψ مـن خـلال برنـامج bootstrap نحصل على :

$$\Phi^{\wedge} = \Lambda^{\wedge}11,\ \Lambda^{\wedge}1^{(1)}\ \Lambda 1^{(2)}\}^{-1} \quad \& \quad \Psi^{\wedge} = F^{\wedge}11\{F^{\wedge}1^{(1)}\ F1^{(2)}\}^{-1}$$

ان المقـدرين اعـلاه همـا معلمتـي التمهيـد band width وهمـا يـؤثران تـأثيرا كبيـرا علـى تمهيـد المنحنى المقدر وأقترابه الى المنحني الحقيقي وأن اي زيادة لهاتين المعلمتين يؤديان الى

تعظـيم التحيـز وتصـغير التبـاين والعكـس صـحيح وهنـا معلمتـي التمهيـد $\Psi^{\wedge}$ و $\Phi^{\wedge}$ تمـثلان معلمتي التباين.

ولتوحيد المعلومات عبر الزمن لتكون اكثر اختصارا نكامل المعدلات الموزونة ا لى كل من $\Psi^{\wedge}$ و $\Phi^{\wedge}$:

$$\Phi^{\wedge}* = \frac{\int (t1,t2) w^{\wedge}(s,t) \Phi^{\wedge}(s,t) dsdt}{\int [t1,t2] w^{\wedge}(s,t) dsdt} = \int (t1,t2)\ w^{\wedge}(s,t)\ \Phi^{\wedge}(s,t)\ dsdt$$

$$\Psi^{\wedge}* = \frac{\int (T1,T2) w^{\wedge}(s,t) \Psi^{\wedge}(s,t) dsdt}{\int [T1,T2] w^{\wedge}(s,t) dsdt} = \int (t1,t2)\ w^{\wedge}(s,t)\ \Psi^{\wedge}(s,t)\ dsdt$$

حيث ان

$$\tilde{w}(s,t) = \frac{w^{\wedge}(s,t)}{\int [T1,T2] w^{\wedge}(s,t) dsdt} \qquad \text{as } n \to \infty$$

وان $w^{\wedge}(s,t)$ تمثل دالة المعدل الموزون للاستجابات اللذان يساويان الواحد الصحيح تحت الاستقلالية لكل من النقطتين s , t ويمكن استخراج حدود الثقة للمقدر اللامعلمي $F^{\wedge}11$ (وهو الاكثر قوة) عند مستوى دلالة $\propto = 0.05$

من خلال حساب ولكل مجموعة بيانات$F^{\wedge}11\ (s,t)$ عند فترة الثقة 0. 95 حسب الاتي :

$$\Phi^{\wedge}(s,t) = \text{logit} \{ F11\ (s,t) \}$$

$$\Phi^{\wedge}(s,t) = \log \{ \frac{F11(s,t)}{1 - F11(s,t)} \}$$

حدود الثقة لكل مجموعة ستكون

$$\Phi^{\wedge}(s,t) \pm 1.96\ Se \{ \Phi^{\wedge}(s,t) \}$$

اذ ان { $\Phi^{\wedge}(s,t)$ } Se هو الخطا المعياري المستخرج من bootstrap لعينة تتكون من n من الازواج مع ارجاع وان مقدرات طريقة البوتستراب تتصف بانها متسقة ، موحدة وتقترب من عمليات (Gaussian) .

المحاكاة simulation :

- توليد ارقام عشوائية عن طريق RND (I) بين الصفر والواحد تتوزع توزيعا منتظما U (0,1) .
- توليد بيانات الاباء $T_2^{(1)}$, $T_1^{(1)}$ والابناء $T_2^{(2)}$, $T_1^{(2)}$ تتوزع توزيع كاما الضعيف بالمعلمتين $\propto = \beta = \frac{1}{\sigma 2}$.

- ترتيب البيانات بشكل عناقيد وحسب طريقة الربط المفرد .
- توليد بيانات مراقبة مستلقة $C^{(1)}$ و $C^{(2)}$ من النوع الثاني يمين censor type two data تتوزع توزيع منتظم U (0,2) بعدد n = 250 .
- استخراج قيم الدوال الاحادية والثنائية لكل من λ و Λ و F .
- تقدير الدوال الثنائية لـ S و Λ_{11} و F_{11} عن طريق برنامج بوتستراب للحصول على المقدرات $S^{\Lambda^{\#}}$ و $\Lambda^{\Lambda^{\#}}_{11}$ و $F^{\Lambda^{\#}}_{11}$.
- استخراج مقدر التباين للـ $\hat{\Phi}$ و $\hat{\Psi}$ باستخدام دالة انفلونس التجريبية $I_i(s , t)$.

استخراج الخطا القياسي لبوتستراب وحسب الصيغة :-

$$BMSE = \frac{\sum_{I=1}^{N} ei^2}{N-p}$$

- مقارنة الاخطاء القياسية لبوتستراب مع الاخطاء التجريبية للمقدر المقترح .

التوصيات :-

يفضل ان يكون اختبار استقلالية الاصابة بمرض السكري الوراثي عن طريق استخدام $\Psi (s , t)$ بدلا من $\Phi (s , t)$ لأنها تزيد من قوة الاختبار عند نقاط زمن الاصابة بالمرض عند اعمار معينة اي (s , t) .

المصادر :

١- علوان ، اسراء سعدون " مقارنة بين طريقتي simex و Lls لتقدير دالة الانحدار اللامعلمية باستخدام المحاكاة " أطروحة ماجستير جامعة بغداد / كلية الادارة والاقتصاد ٢٠٠٣ م .

٢- النزال ، رافد اسماعيل " طرق دراسة الـ jack knifE وال bootstrap) اطروحة ماجستير جامعة بغداد / كلية الادارة والاقتصاد ١٩٨٩ م .

٣- *Bandeen – roche , k . and liang , k (2002) modeling multivariate failure time association in the presence of a competing risk . biometrika , 89, 299-314 .*

٤- *Dabrowska , D.M (1989) Kapalan Meier estimate on the plane : work convergence ,LIL and the bootstrap , J. multiv. Anal,29,308-25.*

٥- *Kosorok , M.R.(2006,to appear) introduction to empirical processes and semi parametric inference New York ;Springer – verlag .*