تحليل المسار في نموذج الانحدار اللوجستي مع تطبيق عملي*

د.افتخار عبد الحميد النقاش * هبة ابراهيم صالح**

المستخلص:

للوقوف على نمط تأثير العديد من العوامل على ظاهرة معينة يكون لبعضها تأثير مباشر واضح بينما البعض االاخر منها يتدرج بنفس الاهمية الا ان هذه التأثيرات لايمكن تحديدها على اساس علاقة مباشرة بينها وبين الظاهرة قيد الدراسة ولكن طبيعة هذا التأثير غير المباشر ينتج من خلال عامل وسيط او اكثر من ضمن العوامل التوضيحية ذات التأثير المباشر , اذ ان هذا المفهوم هو مايمثله اسلوب تحليل المسار (Path analysis) الذي استخدم في بحثنا هذا العرض اسلوب جديد لتحليل البيانات المصنفة (المتعددة الاستجابة) وهي البيانات التي تعتمد العد بدلا" من القياس لدراسة التأثيرات المباشرة وغير المباشرة لمجموعة من العوامل المؤثرة (الجنس , العمر , منطقة السكن) على اصابة الاشخاص دون سن الثامنة عشر بمرض فقر الدم (سوء التغذية , عوز الحديد , البحر الابيض المتوسط) من خلال دراسة نموذج اللوجست متعدد الحدود الشرطي وتحديد العلاقات السببية بين المتغيرات قيد الدراسة للنظام السببي التعاقبي التام وقد تبين بان الاصابة بمرض فقر الدم له علاقة قوية بعامل العمر يليه عامل الجنس حيث تتركز اكثر اصابات الذكور في الفئة العمرية (10-13)سنة بينما تتركز اكثر اصابات الانثى في الفئة العمرية اكبر من (13)سنة .

* استاذ مساعد/ الجامعة المستنصرية/ كلية الادارة والاقتصاد / قسم الاحصاء

* * مدرس مساعد/ الجامعة المستنصرية/ كلية الادارة والاقتصاد / قسم الاحصاء

مقبول للنشر بتاريخ 2006/4/11

*** بحث مستل من رسالة ماجستير بالعنوان ذاته .**

((ABSTRACT))

" Path analysis of logistic regression model , with an applicative work"

In order to know the effect manner of various factors on a certain phenomenon , there will be a clear and direct effect for some of them, while the others are listed with the same importance , but these effects can not be determined according to the direct relation between them and the phenomenon under study. But the nature of this indirect effect that is resulted by one mediator factor or more of the explanatory factors with direct effect.

This concept is represented by path analysis method which has been used in this research for a new method exposition of categorical data analysis (multi- response). These data depend on the accounting rather than gauging for the study of the direct an indirect effects of the affected factors (sex, age , habitation) on the anemia infection that infect people who are under eighteen years old (ill- nutritional, iron deficiency, and Thalassemia). This could be done by study of the conditional multi- limits logic model and determining the causal relations among the variables under study of the fully recurrent causal system.

It has been elucidated that anemia infection has a strong relation with the age factor and then sex factor where the most male infections are concentrated on the age group (10 -13) years old, while for the female are concentrated on the age group that is more than (13) years old.

1.المقدمة

يعد اسلوب تحليل المسار من الاساليب الاحصائية الكفؤة في تحليل البيانات اذ تمكن الباحث من التعيين وبوضوح العلاقات السببية المحتملة لمجموعة من العوامل وبيان تأثيراتها المباشرة وغير المباشرة والكلية على الظاهرة قيد الدراسة وبهذا تساعد الباحث على استنتاج التفسيرات المنطقية للظاهرة وبصورة اكفأ في تحليل البيانات . وتبرز اهمية اسلوب تحليل المسار من خلال القدرة على دراسة تاثيرات عدة عوامل على ظاهرة معينة بشكل غير مباشر عبر عدة عوامل توضيحية، خلافاُ لاسلوب تحليل الانحدار الذي قد يهمل ذلك الجانب اذ يفترض اتجاها معينا للسببية من خلال اهتمامه بتحليل الاثار المباشرة للمتغيرات التوضيحية على متغير الاستجابة.

ويتم ضمن هذا التحليل جعل تقييم الاهمية النسبية للمتغيرات الخارجية اكثر وضوحا في تحديد الاختلافات الكلية للمتغير الداخلي من خلال تحليل العلاقات بين المتغيرات في نماذج سببية(causal models) مبنية على اساس نظرية ما لاختبار الفرضيات المحددة للعلاقات السببية ولمجموعة من المتغيرات.

2. هدف البحث

يتمثل هدف البحث في استخدام اسلوب جديد لتحليل البيانات المصنفة والمتمثل باسلوب تحليل المسار من خلال بناء النماذج السببية المقترحة وتحديد الاهمية السبية بين المتغيرات المحددة في النماذج ودراسة تأثيراتها المباشرة وغير المباشرة على الظاهرة المدروسة للنظام السببي التعاقبي التام في حالة كون متغيرات الاستجابة وصفية (متعددة الاستجابة) باستخدام النموذج اللوجستي ذو الاحتمال الشرطي لوصف النظام السببي للمتغيرات المصنفة وتفسير تلك التاثيرات باستخدام قاعدة لوغاريتم نسبة الارجحية (Log ODD Ratio)

3. الجانب النظري

3-1 النماذج السببية causal models

وهي النماذج التي تستخدم لتفسير وتحديد العلاقة بين المتغيرات ومن ثم ترجمة تلك العلاقات رياضيا حيث تعبر عن المكونات الاساسية للظاهرة قيد الدراسة، اذ ان صياغة المعادلات السببية

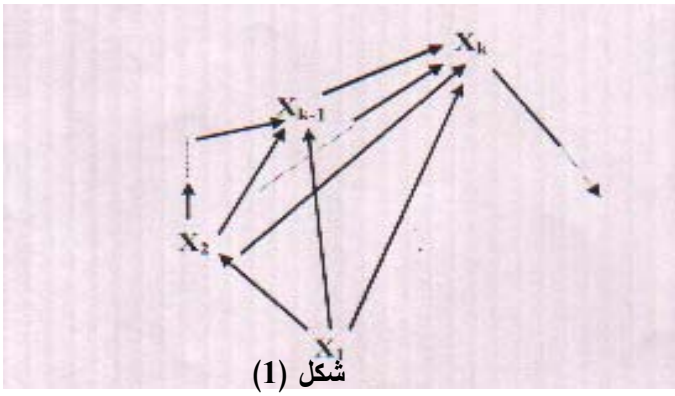
وتحديد تلك المعادلات في النماذج تتم من قبل الباحث وفقا للفرضية التي يحدد من خلالها العلاقة بين المتغيرات السببية(Causes) والمتغيرات المتأثرة بها (Effects) وتنقسم الانماذج السببية الى:

1. النموذج السببي غير التعاقبي (أنموذج التغذية العكسية)
2. النموذج السببي التعاقبي (العلاقة السببية احادية الاتجاه)

اما بالنسبة للأنموذج السببي قيد الدراسة والمتمثل بالأنموذج السببي التعاقبي فان تقنيات هذا الأنموذج تحتم على الباحث بذكر او تحديد اتجاه العلاقة السببية ان وجدت بين متغيرين، وقد وضح[1973] Coodman وجود الترتيب الزمني أو الاولوية السببية بين المتغيرات في الأنموذج التعاقبي فلو فرضنا ان هناك ثلاثة متغيرات مرتبة على التوالي حسب الاولوية السببية كالاتي

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

فمثلا المتغير الخارجي X_1 (Cause) يؤثر على المتغير الداخلي X_2(Effect) والذي يعد بدوره متغير خاجي في تاثيره على المتغير الداخلي X_3 ، وبعبارة اخرى فان المتغير الداخلي X_2 تتحدد اختلافاته بأسباب او متغيرات سابقة له مثل X_1 ولا تتحدد بالمتغيرات السببية اللاحقة مثل X_3 ومن هنا جاءت تسميته بالأنموذج السببي التعاقبي . ويطلق على هذا الأنموذج بالتعاقبي التام(Fully Recursive model) عندما تتحدد اختلافات المتغير الداخلي بكل او جميع المتغيرات السابقة له فمثلا المتغير الداخلي X_{K-1} تتحدد اختلافاته بكل المتغيرات السببية السابقة له مثل $X_1, X_2,$ .....X_{K-1}) ولا تتحدد اختلافاته بالمتغير اللاحق X_K وهكذا.



شكل (1)

العلاقات السببة للنظام السببي التعاقبي التام

2-3 مراحل بناء النماذج السببية

عند اقامة او بناء النماذج السببية هناك بعض النقاط التي يجب اخذها بنظر الاعتبار وهي كالاتي :

1- تحديد العلاقة بين المتغيرات بالاعتماد على الأسس النمطقية او النظريات العلمية ومراعاة التسلسل الزمني الذي يجب اخذه بنظر الاعتبار عند ترتيب المتغيرات وملائمة البيانات مع الأنموذج المفترض تعد من الاساسيات الواجب اتباعها عند بناء الأنماذج السببية.

2- تحديد الشكل الرياضي للانموذج وتدعى هذه الخطوة بالتخصيص (Specification) ويقصد بها تحويل الفروض النظرية الى مجموعة معادلات لغرض تشكيل الأنموذج السببي .

3- تشخيص كل معادلة في الأنموذج (Identification) .

4- ايجاد التقديرات الاحصائية للمعلمات في الأنموذج المفترض .

5- تقييم اداء الأنموذج السببي بأجراء الاختبارات المناسبة .

6- تحليل الأنموذج وتفسير النتائج ووضع التوصيات الملائمة .

3-3 الأساليب المستخدمة في تحليل المسار

اولاً: التمثيل البياني Graphical representation

يمكن اعتبار مخطط المسار (path diagram) احد الأساليب المستخدمة بتحليل المسار اذ يعتبر وسيلة لعرض العلاقات الفرضية بين المتغيرات الخارجية (causes) والمتغيرات الداخلية (effects) ، لذا فمن خلال مجموعة من الأسهم والمسارات يتم تحديد الأهمية النسبية للمتغيرات المحددة في الانماذج ودراسة تأثيراتها المباشرة وغير المباشرة .

ويمكن تلخيص قواعد مخططات المسار كما يلي :

1- رسم أسهم ذات اتجاه واحد متجهة من السبب (cause) (المتمثل بالمتغيرات الخارجية) الى ناحية التأثير (Influence) وصولاً الى الأثر (Effect) (المتمثل بالمتغيرات الداخلية) ان هذه الأسهم هي ما تدعى بالمسارات (paths) .

2- ان عدم وجود أسهم بين متغيرات معينة تشير الى عدم وجود علاقة سببية بينهما .

3- يتم رسم سهم ذي اتجاهين بين متغيرين ليشير الى الطبيعة المتبادلة في التأثير بينهما .

وفيما يلي بعض الملاحظات الخاصة بمخطط المسار للمتغيرات المصنفة المتعددة الاستجابة ، مع الملاحظة بان الملاحظات اللاحقة لاتنطبق على منظومة المتغيرات ذات الاستجابة الثنائية .

1- لايوجد حساب لقيمة عددية واحدة لمعامل المسار (path coefficient) في حالة المتغيرات المصنفة خلافاً لأسلوب تحليل المتغيرات المستمرة اذ ان كل مسار يملك قيمة واتجاه ليعبر عن الأثر المتوقع للمتغير الخارجي على المتغير الداخلي .
2- التصنيفات المتعددة للمتغير تقودنا الى معاملات مسار متعددة والمرتبطة بالمسار المعطى في المخطط .
3- بسبب عدم وجود حساب لقيمة عددية واحدة لمعامل المسار لذا لاتوجد طريقة محددة لمعرفة اية قيمة معلمية مناسبة لتؤشر عل المسار المراد بيان تأثيره .

ثانياً: تحليل النماذج التركيبية (نماذج متعددة المعادلات)

Analysis of structural equation models (multi equation models)

ان النماذج المقدرة وخاصة نماذج الانحدار تحتوي على معادلة مفردة فقط نتمكن من خلالها تفسير كل متغير توضحي او خارجي (Explatory or Exgenous variable) وتاثيره المباشرعلى المتغير المعتمد او الداخلي (Dependent or Endogenous variable) والخروج باستنتاجات منطقية للظاهرة قيد الدراسة .

ومع ذلك فان العديد من النظريات التجريبية تميل بان تكون اكثر تعقيداً اذ يتطلب تفسيرها وجود علاقات او معادلات لتعبر عن الظاهرة قيد الدراسة . ففي هذه الحالة فان مجموع التغيرات المتوقعة لاتقتصر على التأثير المباشر للمتغير الخارجي ولكن ايضاً تتضمن التغيرات الناتجة في المتغيرات التوضيحية الأخرى وهذه التغيرات تدعى بالتأثيرات غير المباشرة . وبما ان عملية تحليل المتغيرات المصنفة تتطلب استخدام نماذج سببية اكثر تعقيداً بالمقارنة مع النماذج السببية للمتغيرات المستمرة والتي تعتمد معادلة الانحدار الخطية (Linear Regression equation) لوصف النظام السببي لتلك المتغيرات ، لذا تم اعتماد أنموذج انحدار اللوجستك متعدد الحدود للتعامل مع النظام السببي في حالة تصنيف المتغيرات .

ولبيان تأثيرات المتغيرات التوضيحية على أنماذج انحدار اللوجستك التركيبية (structural logistic regression model) فقد تم الاعتماد على الاحتمالات الشرطية للأنموذج اللوجستي لتوضيح تلك التاثيرات لتكن $\{x_i, i = 1,2,...k\}$ متغيرات مصنفة ذات التصنيفات الممكنة $\{1,2,.....,I_j\}$ لذا نعبر عن العلاقة التركيبية بين موجه الاستجابة X_k وموجه المتغيرات التوضيحية

$X_{pa(k)} = (X_{1,} X_{2,} ... X_{k-1})'$ بالاحتمال الشرطي لموجه الاستجابة $X_k = x_k$ باعطاء $X_{pa(k)} = x_{pa(k)} = (x_{1,} x_{2,} ...., x_{k-1})'$ كالاتي :

$$P\left(\frac{X_k}{X_{pa(k)}}\right) = \frac{\exp\left(\alpha_{xk} + \sum_{i-1}^{k-1} \beta_i x_i x_k\right)}{\sum_{X_{k-1}}^{I_k} \exp\left(\alpha_{xk} + \sum_{i-1}^{k-1} \beta_i x_i x_k\right)} \quad \text{.....(1)}$$

حيث

α_{xk} : تمثل معلمة التقاطع Intercept parameter

$\beta_{ix_i x_k}$: يمثل تاثير معلمة المتغير التوضيحي X_i على متغير الاستجابة X_k .

للأنموذج السابق نعرف المتغيرات الوهمية (Dummy variable) الاتية .

$$X_{ij} = \begin{cases} 1 & (X_{i=j}) \\ 0 & (X_{i \neq j}) \end{cases} \qquad (j = 1,2,\ldots\ldots, I_i \; ; i = 1,2,\ldots.., k)$$

ولغرض الحصول على مقدرات لمعلمات الأنموذج السببي المدروس من خلال تشخيص الأنموذج اعلاه بافتراض القيود الاتية على المعلمات الاتية :

$\alpha_1 = 0$

$\beta_{i1j} = 0 \quad (j = 1,2,\ldots..,I_k \;\; i = 1,2,\ldots..,K-1)$

$\beta_{ij1} = 0 \quad (j = 1,2,\ldots....,I_i \; i = 1,2,\ldots....,K-1)$

لموجه المتغيرات الوهمية العشوائية $X_i = (X_{i1}, X_{i2},...., X_{iIi})' (i = 1,2,...., K)$ يعاد صياغة الأنموذج (1) كالاتي :

$$P(x_k / x_{pa(k)}) = \frac{\exp(x_k' \alpha + \sum_{i=1}^{k-1} x_k' \beta_i x_i)}{\sum_{x_K} \exp(x_k' \alpha + \sum_{i=1}^{k-1} x_k' \beta_i x_i)} \qquad \ldots\ldots\ldots\ldots (2)$$

اذ ان

$$P(x_k / x_{pa(k)}) = P(X_K = x_k / X_{pa(k)} = x_{pa(k)}); X_{pa(k)} = (X_1^{/}, X_2^{/}, ..., X_{K-1}^{/})^{/}$$

عندما $\alpha = (\alpha_{1,}\alpha_2, ..., \alpha_{1_K})^{/}$

وان

$$\beta_i = \begin{bmatrix} \beta_{i11} & \beta_{i12} & . & . & . & \beta_{i1I_i} \\ \beta_{i21} & \beta_{i22} & . & . & . & \beta_{i2I_I} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ \beta_{iI_k1} & \beta_{iI_k2} & . & . & . & \beta_{iI_kI_i} \end{bmatrix}$$

3-4 حساب معاملات المسار باستخدام معلمات الأنموذج اللوجستي

Calculation of path coefficient by using logistic parameters models

يلاحظ عادة بان تأثير المتغيرات التوضيحية على متغير الاستجابة يتم تحديده من خلال تقدير معلمات الأنموذج المفترض حيث يمكن عن طريقها تحليل هذه النماذج ولكنها لا تزودنا بشرح واضح عن التاثيرات المباشرة وغير المباشرة والكلية لذا فمن خلال هذا البحث تم عرض طريقة جديدة لتفسير تلك التأثيرات وبالاعتماد على معلمات الأنموذج اللوجستي.

حيث يمكن صياغة الاحتمالات الشرطية لنموذج اللوجستي بشكل اخر حسب قاعدة لوغاريتم نسبة الارجحية، اذ ان لوغاريتم نسبة الارجحية لموجه الاستجابة $X_k = x_k$ **على موجه الاساس** $X_k = x_k^*$ **يعرف كالاتي:**

$$Log\frac{P(x_k / x_{pa(k)})}{P(x_k^* / x_{pa(k)})} - Log\frac{P(x_k / x_{pa(k)}^*)}{P(x_k^* / x_{pa(k)}^*)} = \sum_{i=1}^{k-1} tr\beta_i (x_i - x_i^*)(x_k - x_k^*)^{/} \quad .....(3)$$

ان صيغة اللوجستك المذكورة انفا تتضمن زيادة في ارجحية موجه المتغيرات التوضيحية $X_{pa(k)} = x_{pa(k)}$ على موجه المتغيرات الاساس $X_{pa(k)} = x^{*}{}_{pa(k)}$ كما يمكن اعادة صياغة المقدار السابق كالاتي :

$$\text{Log OR}\,(\chi_k, \chi_k^* / \chi_{pa(k)}, \chi^*_{pa(k)}) = \sum_{i=1}^{k\,1} \text{tr}\,\beta_i\,(\chi_i - \chi_i^*)\,(\chi_k - \chi_k^*)' \qquad \ldots\ldots (4)$$

وتعرف الصيغة اعلاه للوغاريتم نسبة الارجحية كحاصل الضرب الداخلي لموجه المتغيرات التوضيحية $\sum_{i=1}^{k-1}\beta_i\,(\chi_i - \chi_i^*)$ وموجه الاستجابة $(\chi_k - \chi_k^*)'$.

عندما:

$$\text{OR}\,(\chi_k, \chi_k^* / \chi_{pa(k)}, \chi^*_{pa(k)}) = \frac{P\,(\chi_k / \chi_{pa(k)})\;P\,(\chi_k^* / \chi^*_{pa(k)})}{P(\chi_k^* / \chi_{pa(k)})\,P(\chi_k / \chi^*_{pa(k)})}$$

حيث يتم حساب التأثير الكلي لموجه المتغير التوضيحي X_i على موجه الاستجابة X_k حسب قاعدة لوغاريتم نسبة الارجحية كالآتي:

$$e_T\,(\chi_i \to \chi_k)$$
$$= Log\,OR\{\chi_k, \mu_k; (\chi_i, \mu_{i+1}(\chi_i), \ldots, \mu_{k-1}(\chi_i)), (\mu_i, \mu_{i+1}, \ldots, \mu_{k-1} / \mu_{pa(i)})\}$$

$$= \text{tr}\;\beta_i\,(\chi_i - \mu_i)(\chi_k - \mu_k)' + \sum_{j=i+1}^{k-1} \text{tr}\,\beta_j(\mu_j(\chi_i) - \mu_j)(\chi_k - \mu_k)' \qquad \ldots\ldots (5)$$
$$(i = 1,2,\ldots, k-2)$$

إذ ان التأثير الكلي عبارة عن حاصل جمع المكونين الاول والثاني فالمكون الاول يمثل التأثير المباشر لموجه المتغير التوضيحي X_i على موجه الاستجابة X_k ، إذ ان التأثير المباشر للموجه التوضيحي $X_i = \chi_i$ على موجه الاستجابة $X_k = \chi_k$ بوجود المتغيرات التوضيحية الاخرى $X_j = \chi_j (j = i+1, i+2, \ldots, k-1)$ يعرف كالآتي:-

$$e_d(\chi_i \to \chi_k / \chi_{i+1},..., \chi_{k-1}) = \text{Log OR}(\chi_k, \mu_k; \chi_i, \mu_i / \mu_{pa(i)}, \chi_{i+1},..., \chi_{k-1}) \quad \ldots 6$$
$$= \text{tr}\,\beta_i(\chi_i - \mu_i)(\chi_k - \mu_k)'$$

ويلاحظ ان التأثير المباشر لموجه المتغير التوضيحي $X_{k-1} = \chi_{k-1}$ على موجه الاستجابة $X_k = \chi_k$ يكون مساويا للتأثير الكلي لهما، إذ أن $e_d(\chi_{k-1} \to \chi_k) = e_T(\chi_{k-1} \to \chi_k)$.

اما المكون الثاني فيمثل التأثير غير المباشر وكالآتي:

$$e_{ind}(\chi_i \to \chi_k) = \sum_{j=i+1}^{k-1} tr\,\beta_j(\mu_j(\chi_i) - \mu_j)(\chi_k - \mu_k)' \quad \ldots\ldots (7)$$

$$= Log\,OR\{\chi_k, \mu_k; \mu_{i+1}(\chi_i),..., \mu_{k-1}(\chi_i))\ ,\ (\chi_i, \mu_{i+1},..., \mu_{k-1} / \mu_{pa(i)})\}$$

ويتم حساب معدل التأثير المباشر وغير المباشر والكلي لموجه المتغير التوضيحي X_i على موجه الاستجابة X_k بأخذ التوقع للصيغة (5) وكالآتي:-

$$E\{e_T(X_i \to X_k)\} = E\{e_d(X_i \to X_k)\} + E\{e_{ind}(X_i \to X_k)\}$$
$$= tr\,\beta_i \,\text{cov}(X_i, X_k) + \sum_{j=i+1}^{k-1} tr\,\beta_j \,\text{cov}(\mu_j(X_i), \mu_k(X_i)) \quad \ldots 8$$

عندما :-

$$Cov(X_i, X_k) = E\{(X_i - \mu_i)(X_k - \mu_k)'\}$$

وإن:-

$$Cov(\mu_j(X_i), \mu_k(X_i)) = E\{(\mu_j(X_i) - \mu_j)(\mu_k(X_i) - \mu_k)'\}$$

4. الجانب التطبيقي

لقد خصص موضوع الدراسة لبحث استخدام الاساليب الاحصائية في الجوانب الصحية اسهاما في زيادة المعرفة بالطرائق الاحصائية وتطبيقاتها للوقوف على نمط تأثير العديد من العوامل على الاصابة بالأوبئة والامراض وتركز الاهتمام بمرض فقر الدم باعتباره احدى المشاكل الصحية الاولى في العالم بالإعتماد على البيانات المتوفرة في المراكز الصحية لمدينة بغداد.

شمل التطبيق دراسة المصابين بفقر الدم من دون سن الثامنة عشر والمصنفين على اساس انواع فقر الدم الى ثلاثة انواع مرتبة حسب اهميتها في الدراسة، ويهدف التطبيق الى بناء انموذج سببي لدراسة وتقدير العلاقات الفرضية بين أهم العوامل المؤثرة على الاصابة بالمرض، كما يهدف الى تحديد العوامل الأكثر تمييزا من مجموعة العوامل المدروسة (المتغيرات التوضيحية في الأنموذج) ودراسة تأثيراتها المباشرة وغير المباشرة على الاصابة بالأنواع المختلفة لمرض فقر الدم.

ولدراسة كافة العلاقات السببية التي تربط المتغيرات المصنفة وفق النظام السببي التعاقبي التام يتم:-

اولا- دراسة التأثيرات المباشرة وغير المباشرة لمتغيري العمر والجنس على الاصابة بالمرض :

بعد جمع البيانات عن المصابين بفقر الدم المشمولين بالدراسة البالغ عددهم (505) مصاب تم خزنها في الحاسب الألكتروني ولما كانت المتغيرات مصنفة فقد تم وضع رمز لكل مستوى من مستويات المتغيرات كالآتي:

X_1 متغير عمر المصاب مصنف الى اربع فئات مستوياتها:

(1) الفئة العمرية أقل من 6 سنوات (2) من 6-10 سنوات

(3) من 10-13 سنة (4) الفئة العمرية من 13-18 سنة

X_2 متغير جنس المصاب مستوياته

(1) ذكر (2) أنثى

X_3 متغير الاستجابة يمثل نوع الاصابة بفقر الدم مستوياتها:

1. فقر دم سوء التغذية
2. فقر دم عوز حديد.
3. فقر دم البحر الابيض المتوسط.

وباستخدام البرنامج الاحصائي الجاهز SPSS (V11)-Descriptive statistics-crosstables تم الحصول على التكرارات الحقيقية لتصنيف المتغيرات المدروسة والموضحة في الجدول (1):

جدول (1)

توزيع المرضى مصنفين حسب العمر والجنس ونوع المرض

العمر بالسنوات Address	الجنس Sex	فقر الدم Anemia		
		فقر الدم سوء التغذية Nutr.	فقر الدم عوز حديد IDA.	فقر الدم البحر الابيض المتوسط Thala
6-	ذكر	31	20	10
	انثى	23	21	7
10-6	ذكر	53	13	5
	انثى	42	14	4
13-10	ذكر	37	18	8
	انثى	26	7	9
18-13	ذكر	14	15	11
	انثى	34	54	29

1. بناء الأنموذج السببي الملائم:

بالاعتماد على العلاقات المنطقية فضلاً عن ما حددته النظرية الطبية حول مرض فقر الدم ولغرض تحديد العلاقات الدالية بين المتغيرات تم توفيق الأنموذج الآتي لبيان تأثير العمر X_1 الجنس X_2 على الاصابة بمرض فقر الدم X_3 .

$$P(x_3 / x_{p\alpha(3)}) = \frac{\exp(x_3'\alpha + \sum_{i=1}^{2} x_3'\beta_i x_i)}{\sum_{x_3=1}^{3} \exp(x_3'\alpha + \sum_{i=1}^{2} x_3'\beta_i x_i)}$$

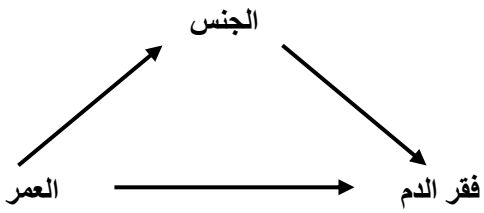
إذ إن:

$\chi_{p\alpha} = (x_1', x_2')'$: موجه يمثل المتغيرات التوضيحية في الأنوذج

$X_1' = (X_{11}, X_{12}, X_{13}, X_{14})$: موجه وهمي يمثل مجاميع تصنيف المتغير التوضيحي(العمر)

$X_2' = (X_{21}, X_{22})$:موجه وهمي يمثل مجاميع تصنيف المتغير التوضيحي(الجنس)

$X_3' = (X_{31}, X_{32}, X_{33})$: موجه وهمي يمثل مجاميع تصنيف متغير الاستجابة (نوع المرض)



شكل (2)

شكل مخطط المسار للمتغيرات أعلاه للنظام السببي التام

وبتطبيق الاساليب التكرارية لطريقة الامكان الاعظم في تقدير معلمات الأنموذج السببي وبأستخدام البرنامج الاحصائي الجاهز SPSS (V11) -Multinomial logistic model تم الحصول على القيم التقديرية النهائية لمعلمات الأنموذج المفترض والموضحة في الجدول (2).

دول (2)

القيم التقديرية لمعلمات الأنموذج،الأخطاء المعيارية وقيم إحصاءة Wald ومستويات معنوياتها لبيان تأثير متغيري (العمر،الجنس)

				S.E	β	اختبار Wald	
						W	P.value
فقر دم سوء التغذية	معلمة التقطيع			0.226	0.153	0.457	0.499
	العمر بالسنوات	6-		0.360	0.941	6.852	0.009
		10-6		0.416	2.144	26.567	0
		13-10		0.359	1.090	9.208	0.002
		18-13		0	0	0	0
	الجنس	ذكر		0.276	0.112	0.165	0.685
		انثى		0	0	0	0
فقر دم عوز الحديد	معلمة التقطيع			0.211	0.562	7.064	0.008
	العمر بالسنوات	6-		0.360	0.354	0.969	0.325
		10-6		0.440	0.571	1.685	0.194
		13-10		0.384	-0.137	0.128	0.721
		18-13		0	0	0	0
	الجنس	ذكر		0.289	-0.067	0.054	0.816
		انثى		0	0	0	0

- القيم الصفرية تمثل القيم الثابتة لتقدير المعلمات(تشخيص الأنموذج)
- لبيان معنوية كل معلمة تقارن قيم احصاءة Wald مع مربع كاي بدرجة حرية تساوي واحد.
- فقر دم البحر الابيض المتوسط يمثل التصنيف الاساس base category.

2. تقدير معاملات المسار

باستخدام اسلوب تحليل المسار لتحديد الاهمية النسبية للمتغيرات الخارجية العمر X_1 والجنس X_2 على المتغير الداخلي نوع مرض فقر الدم X_3 وبالاعتماد على معلمات الأنموذج المقدر الذكور في جدول (2) وباستخدام الامكانات المتوفرة في برنامج Excel تم حساب التأثيرات المباشرة وغير المباشرة للمتغيرات الخارجية وفق الصيغ (6) و (7) وكذلك تم حساب معدل التاثيرات للمتغيرات المدروسة وفق الصيغة (8) والنتائج التي التي حصلنا عليها موضحة في الجدول(3).

جدول(3)

التأثير المباشر وغير المباشر لكل من العمر والجنس على الاصابة بمرض فقر الدم

		العمر بالسنوات Age				الجنسSex	
		-6	6-10	10-13	13-18	ذكر	انثى
التاثير المباشر	فقر الدم سوء تغذية	-0.074	0.440	0.156	-0.417	0.041	-0.035
	فقر الدم عوز حديد	0.132	-0.341	-0.278	0.376	-0.055	0.048
	فقر الدم البحري	-0.024	-0.714	0.057	0.574	-0.019	0.017
التاثير غير المباشر	فقر الدم سوء تغذية	0.006	0.006	0.005	-0.016		
	فقر الدم عوز حديد	-0.008	-0.008	-0.007	0.022		
	فقر الدم عوز حديد	-0.003	-0.003	-0.003	0.008		
معدل التاثير	المباشر وغير المباشر الكلي	0.119 -0.017 0.102				0.004 - 0.004	

من خلال تحليل النتائج الواردة في تحليل الأنموذج المقترح لاحظنا بان الجزء الاكبر من التاثير يتاتى من عامل العمر . وبالنسبة لعامل العمر X_1 يلاحظ بان التأثير غير المباشر اصغر بكثير من التأثير المباشر، اذ نلاحظ بان الفرد المصاب للفئة العمرية اقل من ست سنوات اكثر عرضة للاصابة بمرض فقر الدم الناتج عن عوز حديد، اذ ان ارجحية الاصابة بهذا المرض تساوي 1.141 = exp(0.132) مرة اعلى عن متوسط عامل العمر.

اما بالنسبة للفئة العمرية 6-10 سنوات فان ترجيح اصابتها بمرض فقر الدم الناتج عن سوء تغذية يساوي 1.533= exp (0.440) مرة اعلى عن متوسط عامل العمر.وبالنسبة للفئة العمرية 10-13 سنة فهي اكثر عرضة للاصابة بامراض فقر الدم الناتجة عن عوز التغذية ومن ثم فقر الدم البحري ، اذ ان ترجيح اصابتها بمرض فقر الدم سوء تغذية وفقر الدم البحري يساوي على التوالي 1.169 و 1.059 مرة اعلى عن متوسط عامل العمر ، وأخيرا فأن الفئة العمرية من 13 الى 18سنة أكثر عرضة للأصابة بمرض فقر الدم البحري ومن ثم فقر دم عوز حديد، إذ إن ترجيح

الاصابة بمرض فقر الدم البحري يساوي 1.775 مرة أعلى عن متوسط عامل العمر وإن ترجيح الاصابة بفقر الدم عوز حديد تساوي 1.456 مرة اعلى عن متوسط عامل العمر.

أما بالنسبة للتأثيرات غير المباشرة للعمر من خلال عامل الجنس. نلاحظ بأن الاصابة بمرض فقر الدم وللفئات العمرية (اقل من ست سنوات، 6-10 سنوات، 10-13 سنة) متساوي وأكثر عرضة للأصابة بفقر الدم الناتج عن سوء تغذية، إذ ان ارجحية الاصابة بالمرض تساوي 1.006 مرة أعلى عن متوسط عامل العمر وللفئات العمرية الثلاث الاولى. اما بالنسبة للفئة العمرية من 13 الى 18 سنة فتميل للاصابة بمرض فقر الدم عوز حديد ومن ثم فقر الدم البحري، اذ ان ارجحية الاصابة على التوالي تساوي 1.022 و 1.008 مرة اعلى عن متوسط عامل العمر.

اما التأثير المباشر لمتغير الجنس فيلاحظ بان الذكور تميل الى الاصابة بمرض فقر الدم سوء تغذية، حيث ان ترجيح الاصابة بالمرض تساوي exp (0.041)=1.042 مرة أعلى عن متوسط عامل الجنس.

وبالنسبة للإناث فان ترجيح الاصابة بمرض فقر الدم عوز حديد تساوي exp(0.048)=1.048 مرة اعلى عن متوسط عامل العمر وإن ترجيح الاصابة بفقر الدم البحري تساوي exp (0.017)=1.017 مرة اعلى عن متوسط عامل الجنس.

ثانياً: دراسة التأثير المباشر لمتغير العمر على جنس الفرد المصاب.

1. بناء الأنموذج السببي الملائم.

لغرض تحديد العلاقة الدالية بين المتغيرات ولبيان تأثير متغير العمر X_1 على جنس الفرد المصاب X_2 تم توفيق الأنوذج الآتي:

$$P(\chi_2 / \chi_{pa(2)}) = \frac{\exp(\chi_2' \alpha + \chi_2' \beta_1 \chi_1)}{\sum_{\chi_2=1}^{2} \exp(\chi_2' \alpha + \chi_2' \beta_1 \chi_1)}$$

إذ أن:

$\chi_{pa(2)} = X_1' = (X_{11}, X_{12}, X_{13}, X_{14})$: موجه يمثل مجاميع تصنيف المتغير التوضيحي (العمر).

$X'_2 = (X_{21}, X_{22})$: موجه وهمي يمثل مجاميع تصنيف متغير الاستجابة (جنس المصاب)

جدول (4)

القيم التقديرية لمعلمات الأنموذج، الاخطاء المعيارية وقيم احصاءة Wald ومستويات المعنوية لبيان تأثير متغير العمر

	β	S.E	اختبار Wald	
			W	P.value
معلمة التقاطع	-1.073	0.183	34.339	0.000
6-	1.252	0.264	22.551	0.000
10-6	1.242	0.254	23.976	0.000
13-10	1.479	0.271	29.861	0.000
18-13	0	0	0	0

- القيم الصفرية تمثل القيم الثابتة لتقدير المعلمات (تشخيص الأنموذج)
- لبيان معنوية كل معلمة تقارن قيم احصاءة Wald مع مربع كاي بدرجة حرية تساوي واحد.
- جنس الفرد المصاب (انثى) يمثل التصنيف الاساس base categor .

2. تقدير معاملات المسار

باستخدام اسلوب تحليل المسار لتحديد الاهمية النسبية للمتغير الخارجي العمر X_1 على المتغير الداخلي جنس المصاب X_2 وبالاعتماد على الامكانات المتوفرة في برنامج Excel تم حساب التأثير المباشر من خلال الصيغة (6) وكذلك حساب معدل التاثير المباشر حسب الصيغة (8).والنتائج موضحة في الجدول (5)

جدول (5)

التأثير المباشر الكلي لمتغير العمر على جنس المصاب

الجنس Sex		العمر بالسنوات Age			
		−6	10-6	13-10	18-13
ذكر male		0.184	0.179	0.306	-0.485
انثى female		-0.160	-0.156	-0.266	0.422
معدل التأثير	المباشر	0.088			
	غير المباشر	-			
	الكلي	0.088			

من خلال تحليل نتائج الأنموذج المقترح وجد بان اكثر اصابات الذكور تتركز في الفئات العمرية (اقل من ست سنوات ، 6-10 سنوات، 10-13 سنوات) منه للفئة العمرية اكبر من 13 سنة، اذ ان ارجحية اصابة الذكور وللفئة العمرية اقل من ست سنوات تساوي exp(0.184)=1.202 مرة اعلى عن متوسط عامل العمر، وللفئة العمرية (6−10) تساوي 1.196 = exp(0.179) مرة اعلى عن متوسط عامل العمر ، اما بالنسبة للفئة العمرية (10−13) سنوات تساوي 1.358 = exp(0.358) مرة اعلى عن متوسط عامل العمر .

اما بالنسبة للاناث فتتركز الاصابة في الفئة العمرية اكثر من 13 سنة ، اذ ان ارجحية اصابة الاناث وللفئة العمرية اكثر من 13 سنة تساوي 1.525 = exp(0.422) مرة اعلى عن متوسط عامل العمر ، وهذا يوافق النظرية الطبية التي تشير الى ان نسبة الاصابة بفقر الدم متقاربة للاعمار الصغيرة ولكلا الجنسين وترجح اصابة الاناث للاعمار الكبيرة نسبياً نتيجة تعرضهن الى الكثير من التغيرات الفسلجية التي تؤدي بهن للاصابة بالمرض .

5. الاستنتاجات والتوصيات

5-1 الاستنتاجات

1.تبين من خلال التحليل بان الاصابة بمرض فقر الدم له علاقة قوية بعامل العمر ، اذ يعد هذا العامل المؤثر الاول للاصابة بالمرض حيث كان معدل التاثير الكلي لهذا العامل مساوي الى

المقدار (0.102) نتيجة اختلاف حالات الاصابة للفئات العمرية ضمن العينة ، يليه عامل الجنس حيث كان معدل التاثير الكلي لهذا العامل مساوي الى المقدار (0.004) يرجع الى ان نسبة الاصابة بالمرض لكلا الجنسين بمستوى متقارب .

2. تبين من خلال هذا التحليل بان الاصابة بمرض فقر الدم تظهر في كلا الجنسين (الذكور ، والاناث) وتتركز اكثر اصابات الذكور في الفئة العمرية (10–13) سنة بينما تتركز اكثر اصابات الاناث بالمرض في الفئات العمرية الكبيرة نسبياً (من 13 الى 18سنة) .

3. كما ظهر لنا بان الاصابة بمرض فقر الدم الناتج عن سوء تغذية له علاقة بعامل العمر ، حيث يلاحظ بان الاصابة بهذا النوع من المرض تتركز في الفئات العمرية الصغيرة نسبياً بالمقارنة مع الفئة العمرية (من 13 الى 18سنة) اذ تمثل هذه المرحلة حالة الحركة والنشاط الجسماني لدى هذه الفئات وللذكور اكثر منه بالنسبة للاناث.

4. أما بالنسبة للاصابة بمرض فقر الدم الناتج عن عوز حديد وفقر الدم البحري فتتركز أكثر حالات الاصابة بالمرض في الفئة العمرية (أكبر من 13 سنة) وللاناث اكثر منه بالنسبة للذكور نتيجة التغيرات الفسلجية لدى الاناث لهذه الفئة العمرية.

2-5 التوصيات.

1. اهمية استخدام اسلوب تحليل المسار في تحليل البيانات المصنفة من خلال افتراض نماذج سببية تساهم في التنبؤ بقيم متغير ما او مجموعة من المتغيرات ودراسة التأثيرات المباشرة وغير المباشرة ما بين المتغيرات، فضلاً عن تحديد مدى تأثير كل متغير.

2. دراسة تأثير المتغيرات التوضيحية على نماذج انحدار اللوجستك التركيبية في حالة وجود تفاعل بين المتغيرات ودراسة تأثيراتها المباشرة وغير المباشرة على الظاهرة قيد الدراسة.

3. ضرورة تشجيع وتوسع البحوث الصحية ذات العلاقة بمرض فقر الدم من خلال جمع البيانات حول انتشار المرض ونسبة الاصابة والعوامل المرافقة لظهوره واهمية اعتماد المستشفيات الحكومية على اسلوب التشخيص الدقيق لمرض فقر الدم نظراً للصعوبات التي واجهها الباحث في الحصول على تشخيص دقيق لنوع مرض فقر الدم.

4.نوصي كذلك بضرورة رفع المستوى الصحي للمواطنين بصورة عامة وللفئات العمرية من دون سن الثامنة عشر بصورة خاصة الأكثر عرضة للاصابة من خلال التشجيع على الرضاعة الطبيعية

للاطفال والاطعمة التكميلية الملائمة والمحافظة على الاغذية السليمة للتناول والمحافظة على خصائصها التغذوية وتحسين تدبير مصادر المياه في السيطرة على البلهارسيات والملاريا في اماكن انتشارها تعتبر من الاجراءات المهمة في الوقاية من مرض فقر الدم .

6. المصادر

1- باجلان . بتول عبد الكريم (1983),"استخدام تحليل المسالك path analysis في دراسة العوامل الاجتماعية والاقتصادية المؤثرة على زيادة تركيز السكر في الدم" ,رسالة ماجستير, جامعة بغداد ,كلية الادارة والاقتصاد.

2- الطائي , عبد الحسين حسن(2000) , تقدير وتحليل معادلة الانحدار المتعدد في حالة كون المتغيرات المعتمدة وصفية ومحددة " , رسالة دكتوراه , جامعة بغداد , كلية الادارة والاقتصاد.

3- Agresit , A.(1990)"Categorical Data Analysis", John Wiley Sons , Inc , New York .

4- Anderson , J . A . (1984) ," Regression and Order Categorical Variables ",Journal of Royal Statistical Society , B, 46 , 1-30.

5- Asher , H.B.(1983), "Causal Modeling " , Chapman and Hall , London.

6. Imbeds . (2004) ,"Discrete Response Model ",Lecture Notes , Us Berkeley Department of Agricultural and Resource Economics . are . Berkeley . end / courses / ARE213 / fa112004/prid4-04sop.pdf.

7- pryanishnikov , I.,Zigova , K. (2003)"Multinomial logit Model for the Austrian Labor Market ",Austrian Journal , No.32,No. 4,267-282. www.statistik.tuwien.ac.at/oezstat/ausgo34/papers/pryan .pdf

8- Train , Kenneth .(2003)"Discrete Choice Model with Simulation "Cambridge university press . http ://elsa Berkeley.

9- weesis . (2002) ,"on pearson X^2 for categorical Response variable", German Stat User's Meeting http:// www.stat com/meeting/2dutch/pearson X^2 .pdf .

.....

.....

.....