استخدام التحليل العاملي لدراسة العوامل المؤثرة على الإصابة بمرض فقر الدم عند الأطفال

م.م أسيل عبد الرزاق/ قسم الإحصاء / كلية الإدارة والاقتصاد / الجامعة المستنصرية

المستخلص

يعد التحليل ألعاملي من الطرق الإحصائية المهمة التي تهدف إلى دراسة الظواهر المختلفة لأجل استخلاص أهم العوامل التي أثرت فيها من خلال تحليل معاملات الارتباط بين متغيرات الظاهرة المدروسة، كما ويتميز التحليل ألعاملي بقدرته على اختصار المتغيرات الكثيرة وترتيبها في عدد ضئيل من المركبات الخطية.

سيتم في هذا البحث دراسة المتغيرات الأساسية المؤثرة على مرض فقر الدم بناءا على مجموعة من القياسات الخطية الخاصة لكل مريض والتي يعتقد بانها تساهم في تشخيص المرض وتحديد اسلوب علاجه

1:- المقدمة:

على الرغم من اختلاف الباحثين في تحديد مفهوم واحد للتحليل ألعاملي إلا أنهم يؤكدون في الغالب على اعتبار طريقة إحصائية تهدف إلى دراسة الظواهر المعقدة لاستخلاص العوامل التي أثرت فيها من خلال تحليل معاملات الارتباط بين المتغيرات كما يهدف التحليل ألعاملي إلى إيجاد مجموعة من العوامل Factor التي تكون مسؤولة عن توليد الاختلافات Variations في مجموعة مكونة من عدد كبير من متغيرات الاستجابة Response variables حيث يمكن التعبير عن المتغيرات المشاهدة كدالة في عدد من العوامل المستقرة Factors وغالبا ما يعبر عن متغيرات الاستجابة كتركيب خطي Linear compounds من العوامل المستقرة حيث تكون العلاقات بين المتغيرات داخل العامل الواحد أقوى من العلاقة مع المتغيرات في عوامل أخرى. أن التحليل ألعاملي يساعد على فهم تركيب مصفوفة الارتباط أو التباين المشترك من خلال عدد قليل من العوامل.

2:- هدف البحث:

يهدف هذا البحث إلى دراسة مجموعة من المتغيرات ذات العلاقة بمرض فقر الدم عند الأطفال دون الخمس سنوات وذلك استنادا إلى بيانات واقعية لمرضى مسجلين في احدى مؤسسات

وزارة الصحة لخمس مناطق في قطاع الرصافة وهي الاعظمية ، بغداد الجديدة ، الشعب ، مدينة الصدر ، البلديات لعينة حجمها 245 طفل مصاب في عام 2007.

3:- الجانب النظري:

سوف يتم وصف الاساسيات الاولية للتحليل العاملي.

3-1: النموذج ألعاملي Factor model

يتكون النموذج الرياضي للتحليل ألعاملي من مجموعة من المتغيرات المشاهدة (Observed Variables) وعددها (P) ، أخذت من عينة حجمها (n) على أساس دالة خطية ل(q) من العوامل المشتركة (common factors) حيث (p>q) والى (p) من العوامل الوحيدة (Unique factors) لكل متغير أي أن :

حيث أن :

$$\underset{p\times 1}{\underline{X}} = \underset{p\times q}{\underline{A}}\ \underset{q\times 1}{\underline{F}} + \underset{p\times 1}{\underline{U}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$\underline{X}$: الموجه العشوائي للمتغيرات المشاهدة

A: مصفوفة تحميلات العوامل Factor Loading))

$\underline{F}$: الموجه العشوائي للعوامل المشتركة (Common Factors)

$\underline{U}$: الموجه العشوائي للعوامل الوحيدة (Unique Factors)

ان مصفوفة التغاير لـــ($\underline{X}$) توصف بالصيغة:-

$$\underset{p\times p}{\sum} = E[\underline{X}\,\underline{X}'] - (m)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

وان $\sum$ مصفوفة موجبة متماثلة ومن الرتبة P

3-2:ألفرضيات الأساسية للتحليل ألعاملي

Basic Assumption of factor analysis

ألفرضية الأولى:-

تعتمد هذه الفرضية على أساس وجود ارتباطات بين مجموعة من المتغيرات وأن هذه الارتباطات ناتجة عن وجود عوامل مشتركة فيما بينها ، والتحليل ألعاملي يهدف إلى تفسير هذه الارتباطات بعوامل تكون اقل من المتغيرات المستخدمة وان هذه الفرضية تأخذ القيمة المعيارية

للمتغيرات وذلك للحصول على متغيرات تتوزع توزيعا طبيعيا بوسط قدره صفر وتباين قدره واحد لتسهيل العمليات الحسابية وكذلك للتخلص من اختلاف وحدات القياس للمتغيرات أن وجد وتحت هذه الفرضية يقسم التباين الكلي للمتغيرات إلى ثلاثة أنواع هي :-

1:- التباين المشترك العامCommon Variance

ويسمى كذلك بالتباين المشاع أو كميات الشيوع وهو ذلك الجزء من التباين الذي يرتبط مع بقية المتغيرات الأخرى من خلال العوامل المشتركة ويحسب من معاملات العوامل العامة .

2:- التباين الخاص Specific Variance

هو ذلك الجزء من التباين الكلي الذي لا يرتبط مع أي متغير وإنما مع المتغير نفسه فقط.

3:- تباين الخطأ Error Variance

هو ذلك الجزء من التباين الكلي الناتج من العوامل غير المفسرة في النموذج.

ألفرضية الثانية

تقوم الفرضية الثانية في التحليل ألعاملي بافتراض وجود ارتباط بين المتغيرين (i, j) ويمكن حسابه على أساس طبيعة وتأثير تحميلات (تشبعات) العوامل المشتركة . ويمكن تمثيل هذا الفرض بالنسبة للعوامل المتعامدة (Orthogonal) بالمعادلة التالية:

$$r_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + a_{i3}a_{j3} + \ldots\ldots\ldots\ldots + a_{iq}a_{jq} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

أي أن معامل الارتباط بين متغيرين يساوي مجموع حاصل ضرب تحميلات المتغيرات بالعوامل المشتركة بينهما ويمكن إعادة كتابة المعادلة السابقة بالشكل الأتي :

$$R = AA'$$

حيث أن

R: تمثل مصفوفة الارتباط

A: تمثل مصفوفة تحميلات العوامل

قيم الشيوع The communalities

أن كمية الشيوع للمتغير zj هي مجموعة من مربعات تحميلات (تشبعات) ذلك المتغير وتمثل نسبة التباين الذي تفسره العوامل المشتركة الناتجة من تحليل مصفوفة الارتباط R أي أنها تعطي مدى التداخل بين المتغيرات والعوامل المشتركة ويرمز لها بالرمز h_j^2 والمتمثلة بالعلاقة

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \ldots\ldots\ldots + a_{jq}^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$
$$j = 1,2,\ldots,p$$

حيث أن a_{jq} يمثل وزن العامل q بالنسبة للمتغير j وهي معاملات مصفوفة العوامل F وتعرف بتحميلات العوامل أو تشبعات العوامل ومن خصائص كميات الشيوع h_j^2 أنها موجبة وتقع بين الصفر والواحد أي أنها $o \le h_j^2 \le 1$

3-3: طريقة المكونات الأساسية

Principal components method

تعمل هذه الطريقة على تحليل مجموعة من المتغيرات المرتبطة ($X_1,X_2\ldots\ldots X_P$) إلى مجموعة جديدة من المركبات الأساسية ولتكن ($F_1,F_2,\ldots\ldots F_K$,) وتكون هذه المركبات على هيئة تركيب خطي مشتق من المتغيرات ($X_1,X_2\ldots\ldots X_P$) لتحل محلها بحيث تكون هذه المركبات مؤهلة لتفسير معظم التباين الكلي للقيم الأصلية علما بأن هذه المركبات الجديدة ستكون متعامدة أي أنها غير مرتبطة مع بعضها .

من خلال مصفوفة التباين –التباين المشترك ($\sum$) يتم إيجاد المعادلة المميزة للمصفوفة بأخذ المحدد

$$|\Sigma - lI| = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

والشكل العام لها هو :

$$l^P + C_{P-1} l^{P-1} = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

حيث أن (C_O ..... C_{P-2} C_{P-1}) هي معادلة كثيرة الحدود polynomial في l من الدرجة p وعند حل هذه المعادلة سوف يتم الحصول على p من الجذور المميزة $(l_1, l_2, \ldots\ldots\ldots\ldots, l_p)$

بحيث أن

$l_1 \ge l_2 \ge \ldots\ldots\ldots\ldots \ge l_p$ 0 <

وأن لكل جذر مميز (l_j) يكون هناك متجه مميز (a_j) ويتم إيجاده بعد تعويض قيمة الجذر المميز (l_j) حسب العلاقة التالية

$$(\Sigma - lI)\underline{a} = 0 \text{.....(7)}$$

والمركب الرئيسي الأول هو الذي يقابل أكبر جذر مميز لمصفوفة التباين المشترك (Σ) والذي تكون معاملاته تمثل قيم المتجه المميز (a_j) والمقابل للجذر المميز (l_j) فالمركب الرئيسي (j) يكتب بالصيغة التالية :-

$$y_j = a_j' x \text{ .....(8)}$$

وأن قيم المتجه المميز هي معاملات a_{ij} المركب الرئيسي بحيث تحقق خاصية normalize

أي أن $a_j' a_j = 1$

وأن جميع الجذور المميزة لمصفوفة الثابت المشترك (Σ) تكون موجبة $(l_j \geq 0)$ وان مجموعة الجذور

$$\sum_{j=1}^{p} l_j = tr(\Sigma) \text{--------(9)}$$

هو عناصر القطر الرئيسي لمصفوفة (Σ) وان

$$|\Sigma| = \prod_{j=1}^{p} l_j \text{---------(10)}$$

وان مقياس أهمية المركب الأساسي (y_j) في حساب التباين الكلي تحسب بموجب العلاقة الآتية

$$\frac{l_j}{\sum_{j=1}^{p} l_j} = \frac{l_j}{tr(\Sigma)} \text{----------(11)}$$

وكل مركب رئيسي له تباين مساوي إلى قيمة الجذر المميز أي أن

$$\text{cov}(y_i y_j) = \begin{cases} l_j a_j' a_j & \\ 0 & if \quad i \neq j \\ l_j & if \quad i = j \end{cases} \text{----------(12)}$$

3-4طريقة تعظيم التباين لـــ Kaiser:

The Kaiser varimax method

تعتبر هذه الطريقة أكثر قربا للتركيب البسيط ، إذ أنها من أكثر طرق التدوير المتعامد شيوعا وتعتمد تبسيط العوامل من خلال تباين مربعات تحميلات (s_p^2)

$$S_p^2 = \frac{1}{n}\sum_{j=1}^{n}\left(a_{jp}^2\right)^2 - \frac{1}{n^2}\left[\sum_{j=1}^{n} a_{jp}^2\right]^2 \text{ ----------(13)}$$

حيث أن

a_{jp} هو العنصر الواقع في الصف j للعمود p في المصفوفة المدورة وعندما يكون التباين أعظم ما يمكن .

يكون للعامل قابلية للتغيير والتبسيط على أساس أن تحميلا ته تتجه حول الصفر والواحد . فمقياس أعلى تبسيط لمصفوفة العامل الكامل يعرف بأنه تعظيم لمجموع التبسيطات للعوامل المفردة فإذا ما جمعت المعادلة (13) لكل العوامل فأن

$$S^2 = \sum_{P=1}^{m} S_P^2 = \frac{1}{n}\sum_{p=1}^{m}\sum_{j=1}^{n} a_{jp}^4 - \frac{1}{n^2}\sum_{p=1}^{m}\left[\sum_{j=1}^{n} a_{jp}^2\right]^2 \text{ ----------(14)}$$

حيث أن

a_{jp} :- تمثل قيمة تشبع j بالعامل p

n:- هي عدد المتغيرات

أن تعظيم المعادلة 14 يسما بمقياس تعظيم التباين العام وبترجيح المعادلة أعلاه بدلالة قيم الشيوع h_j^2 فأن معيار الاقتراب لتعظيم التباين هو جعل قيمة v)) أعظم ما يكون للحصول على أفضل تحميلات للعوامل وكما يلي :

$$v = n\sum_{p=1}^{m}\sum_{j=1}^{n}\left(\frac{a_{jp}}{h_j}\right)^4 - \sum_{p=1}^{m}\left[\sum_{j=1}^{n}\frac{a_{jp}^2}{h_j^2}\right]^2 \text{ ----------(15)}$$

4:- الجانب التطبيقي

يعتبر مرض فقر الدم (الأنيميا) من أكثر ألأمراض شيوعا عند ألأطفال ألناتج عن نقص الحديد في جسم الطفل والفئة العمرية المعرضة للاصابة بالمرض تتراوح بين 24-9 شهر لأنه خلال فترة الشهور الستة الأولى من عمر الطفل يكون لدى الجسم كمية حديد مختزنة تكفيه خلال تلك الفترة ولكن بعدها مباشرة يبدأ الجسم بالبحث عن الحديد في طعام الطفل اليومي لذا فمن الممكن حدوث فقر الدم لديه لعدم حصوله على التغذية الصحيحة لهذه المرحلة من عمره

وكذلك يحدث فقر الدم أيضا بسبب إصابة الطفل بأمراض الجهاز الهضمي وعدم إعطاء الحديد كعلاج وقائي .

تم الحصول على البيانات الصحية المسجلة لعينة حجمها 245 من مركز بحوث التغذية في الرصافة لعام 2007 للمناطق (الاعظمية ، بغداد الجديدة ، الشعب ، مدينة الصدر ، البلديات) وقد تم اختيار متغيرات الدراسة بعد التداول مع الاطباء الاستشاريين وهي على التوالي:

X_1: العمر	X_7: مشاكل تنفسية
X_2: العشو الليلي	X_8: فيتامين A
X_3: الوزن	X_9: نسبة الهيموكلبين في الدم (Hb)
X_4: الطول	X_{10}: ديدان
X_5: الجنس	X_{11}: إسهال
X_6: علاج ديدان في الشهر السابق	X_{12}: منطقة السكن

وقد تم استخدام نظام (SPSS) لبناء نموذج التحليل ألعاملي من خلال طريقة المكونات الرئيسية الاعتيادية.

4-1:- تفسير نتائج التحليل ألعاملي

بعد تحليل مصفوفة الارتباط الخاصة بالمتغيرات التي تخص مرضى فقر الدم بواسطة طريقة المكونات الرئيسية فقد وجدت (ستة) عوامل رئيسية تؤثر في الإصابة بفقر الدم يمثلها عدد القيم القاعدية التي هي اكبر من واحد وكما هو مبين في جدول (4-1) إن اهمية العامل من مقدار تفسيره للتباين وكذلك مقدار الزيادة التي يضيفها عامل بوجود عوامل وتشكل هذه العوامل 911.62 من التباين الكلي للبيانات.

جدول (4-1)

يبين القيم الذاتية ونسبة تباين العامل من التباين الكلي والتباين المتجمع للعامل

العامل	القيم الذاتية Initial Eigen values			Extraction Squared Loadings		
Component	المجموع Total	نسبة تباين العامل %of Variance	التباين المتجمع Cumulative %	المجموع Total	نسبة تباين العامل %of Variance	التباين المتجمع Cumulative %
1	544.1	863 .12	12.863	544.1	863 .12	12.863
2	354.1	11.287	24.150	354.1	11.287	24.150
3	224.1	203 .10	34.353	224.1	203 .10	34.353
4	187.1	9.895	44.248	187.1	9.895	44.248
5	146.1	552 .9	53.800	146.1	552 .9	53.800
6	093.1	111 .9	62.911	093.1	111 .9	62.911
7	0.998	316 .8	71.227			

8	0.856	136 .7	78.364			
9	819.0	829 .6	85.192			
10	0.717	978 .5	91.171			
11	564.0	699 .4	95.870			
12	496.0	130 .4	100.000			

ومن خلال هذه النتائج ودرجة تحميلاتها في كل عامل عن طريق استخدام مصفوفة تحميلات العوامل المدورة المبينة في جدول (4-2)

جدول (4-2)

مصفوفة تحميلات العوامل المدورة

المتغيرات Variables		العوامل Factors					
		1	2	3	4	5	6
X_1	العمر	0.026	0.741	0.072	0.295	0.038	0.119
X_2	عشو ليلي	0.021	-0.092	0.029	-0.352	0.687	0.208
X_3	الوزن	0.166	-0.013	0.150	0.790	-0.053	0.053
X_4	الطول	0.027	0.803	-0.009	-0.358	0.024	-0.080
X_5	الجنس	0.033	-0.134	0.309	-0.147	-0.063	-0.728
X_6	علاج ديدان	0.302	0.28	0.469	-0.367	-0.137	0.101
X_7	مشاكل تنفسية	0.046	-0.017	0.683	0.072	-0.070	-0.068
X_8	فيتامين A	0.068	-0.067	0.279	-0.087	-0.077	0.686

X_9	Hb	0.781	-0.173	0.127	0.037	0.150	-0.018
X_{10}	ديدان	0.077	0.123	-0.032	0.179	0.780	-0.186
X_{11}	إسهال	-0.214	0.110	0.592	0.221	0.260	0.097
X_{12}	المنطقة	-0.752	-0.268	0.132	-0.118	0.053	-0.053

العامل الأول

حيث كون هذا العامل 863.12 من التباين الكلي وتشير هذه النسبة إلى أن هذا العامل على درجة كبيرة من الأهمية ويشمل كل من المتغيرين :-

1:- hb ويمثل نسبة الهيموكلوبين بالدم وبمقدار تشبع 781.0

2:- المنطقة وبمقدار تشبع -752.0

أوضح البحث أن نسبة الهيموكلوبين بالدم تؤثر تأثيرا كبيرا على مرض فقر الدم حيث يتم من خلالها تحديد نسبة فقر الدم إذا كانت قوية أو متوسطة أو ضعيفة إضافة إلى ذلك هناك تناقصا في تحديد مرض فقر الدم بالنسبة لعامل منطقة سكن المريض وقدرته على مراجعة المراكز الصحية

العامل الثاني

ويأتي هذا العامل بالمرتبة الثانية حيث كون نسبة 11. 287 من التباين الكلي وبتراكم مقداره 24. 150 ويشمل على المتغيرين

١:- العمر وبمقدار تشبع 0. 741

٢:- الطول وبمقدار تشبع 0. 803

أن متغير العمر له دور أساس في تحديد نسبة فقر الدم وذلك لكون الأعراض المرضية عند الأطفال تقتضي مراجعة المراكز الصحية بأقرب فرصة ممكنة اما بالنسبة لمتغير الطول فان فقر الدم يؤدي الى نقص في المواد الغذائية وهرمونات النمو وبالتالي يؤدي الى نقص في نمو المريض مما يؤثر على الطول.

العامل الثالث

وتأتي أهمية هذا العامل بالمرتبة الثالثة حيث تكون نسبة 203.10 من التباين الكلي وبنسبة تراكمية 353.34 ويشمل على متغيرين:-

١:- مشاكل تنفسية وبمقدار تشبع 683.0

٢:- الإسهال وبمقدار تشبع 592.0

بالنسبة لمتغير المشاكل التنفسية فان الاوكسجين ينتقل الى خلايا الدم من خلال كريات الدم الحمراء وعلية فان فقر الدم يؤثر في كمية الاوكسجين اما بالنسبة لمتغير الاسهال فان الاطفال المصابين به تكون لديهم مناعة اقل وضعف بدني.

العامل الرابع

بلغت نسبة هذا العامل 895.9 من التباين الكلي وبتراكم مقداره 44. 248 ويشمل متغير واحد فقط هو الوزن وبمقدار تشبع 790.0

ان متغير الوزن يؤثر في اصابة الطفل حيث ان نقص الوزن يدل على عدم توفر الغذاء الكافي في خلايا الجسم.

العامل الخامس

يفسر هذا العامل 9. 553 من التباين الكلي وبنسبة تراكمية قدرها 800.53 ويشمل على متغيرين وهما

١:- العشو الليلي وبمقدار تشبع 0. 687

٢:- الديدان وبمقدار تشبع 780.0

يمكن تفسير المتغيرين بأن إصابة الأطفال بالعشو الليلي يضعف قدرة المريض على التركيز وذلك يؤثر في صحته من حيث تقبله للغذاء اما بالنسبة لمتغير الديدان فانها طفيليات تدخل في جسم الطفل تمتص الغذاء الذي يدخل الى الأمعاء مما يؤدي الى سوء التغذية والاصابة بفقر الدم.

العامل السادس

بلغت نسبة هذا العامل 111.9 من التباين الكلي وبنسبة تراكمية قدرها 911.62 ويشمل على متغيرين وهما

١:- الجنس وبمقدار تشبع -728.0

٢:- فيتامين A وبمقدار تشبع 0 . 686

ان متغير الجنس ليس بالاهمية وذلك واضح حيث ان التباين في جنس الاطفال دون سن الخامسة يكون قليل مقارنتاً بالاعمار فوق سن الخامسة اما بالنسبة لمتغير فيتامين A فان نقصه يؤثر في صحة الطفل بالاضافة الى ذلك ان نقصه يؤدي الى الاصابه بمرض العشو الليلي.

5- الاستنتاجات والتوصيات:

1:- لقد اظهرت النتائج ان المتغير علاج ديدان في الشهر السابق لم يظهر تأثيره في اي من العوامل الستة مما يعني ان اخذ المريض لجرع علاج الديدان لا تؤثر في مرض فقر الدم.

2:- ان هذه العوامل مجتمعة تكون متوافقة مع ما يلجأ اليه الاطباء عند تشخيصهم للمرض وبذلك تدل النتائج الى مطابقتها لواقع حال العلاج الطبي.

3:- نوصي بان تكون المؤسسات الصحية اكثر دقة في جمع البيانات واكثر تعاونا مع الباحثين واستشارتهم في اساليب جمع البيانات الحديثة.

4:- نوصي بدراسة العوامل المؤثرة لمرض فقر الدم للأطفال فوق سن الخامسة ومقارنة النتائج بدراسة الاطفال تحت سن الخامسة.

6- المصادر:

1:- الياس ، جوان غالب : (1993) ،" استخدام التحليل ألعاملي لتحديد أهم العوامل المؤثرة في الإصابة بمرض سرطان الثدي " ، رسالة ماجستير مقدمة إلى قسم الإحصاء – كلية الإدارة والاقتصاد – الجامعة المستنصرية .

2:- الهاشمي ، ندى محمد جواد : (1999) ،" استخدام التحليل ألعاملي لتحديد العوامل ذات العلاقة بظهور حالات الإصابة بأورام الدماغ " ، رسالة ماجستير مقدمة إلى قسم الإحصاء – كلية الإدارة والاقتصاد – الجامعة المستنصرية .

3 -:Anderson , T.W).1984":(An introduction to multivariate statistical analysis ", 2nd edition , john Wiley & sons ,new York.

4 -:Timm , N.H).1975": (Multivariate Analysis with Application in Education and psychology ",Wadsworth publishing Company , California .