

## إستخدام التحليل التمييزي لتصنيف الأطفال دون سن الخامسة الى مصابين وغير

### مصابين بمرض الإسهال في محافظة واسط

أ.م.د. سعد صبر محمد

م.م. إخلاص علي حمودي

كلية الادارة والاقتصاد / جامعة واسط

### المستخلص :-

في هذا البحث تم استخدام التحليل التمييزي لتصنيف الأطفال الى مصابين وغير مصابين بمرض الإسهال اعتماداً على متغيرات ذات صفات تمييزية من خلال بيانات الأطفال لدون الخمس سنوات في العراق لعام ٢٠١١ من خلال المسح المتعدد المؤشرات ( MICS4 ) والذي تم تنفيذه من قبل الجهاز المركزي للإحصاء وتكنولوجيا المعلومات ، وباستخدام الدالة المميزة الخطية تم تصنيف الأطفال الى مصابين وغير مصابين بمرض الإسهال على اساس مجموعة من المتغيرات المصاحبة للمرض .

### ABSTRACT:-

*In this research the discriminant analysis has been used to classify the children to those infected and non-infected diarrhea disease depending on variables with recipes discriminatory through the children data under the age of five years in Iraq for 2011 through multiple indicator survey ( MICS4 ) which was implemented by the central bureau of statistics and information technology, by using the linear discriminant function . children were classified to those infected and non-infected diarrhea disease on the basis of a set of variables associated with the disease .*

### Introduction

### 1-1 المقدمة [5][8] :

يعد التحليل التمييزي ( Discriminant Analysis ) ويرمز له اختصاراً (DA) احد الاساليب الاحصائية لتحليل المتغيرات المتعددة ( Multivariate Analysis ) ولقد قدم العالم فيشر ( R. A. Fisher ) ذكر هذا المصطلح عند اول معالجة حديثة لمشاكلات الفصل (العزل) ( Separation ) والتمييز ( Discriminant ) . اذ اوجد دالة عرفت بأسمه دالة فيشر الخطية للتمييز ( Fisher's Linear Discriminant Function ) ، ويستخدم التحليل التمييزي ( DA ) في

دراسة تصنيف الاشخاص او العلامات الى مجموعات كأن تكون مجموعتين او اكثر من ذلك بناء على الأزواج او النسب او الدرجات التي يحصلون عليها في توليف خطي للمتغيرات المستقلة اذ يتم فيها تصنيف عضوية الحالات في مجموعتين معرفتين او اكثر على ان تكونا محددين مسبقا على خلاف

التحليل العنقودي ( Clustter Analysis ) الذي لا يكون فيه عدد المجموعات معرّفا ولا عضوية الحالات فيه معرّفا . كأن يتم تصنيف الزبائن الى زبائن يتوقع منهم ان يكونوا راضين او غير راضين عن منتج معين او تصنيف الشركات الى شركات يتوقع تعثرها او شركات لا يتوقع تعثرها . ويظهر من ذلك بان التحليل التمييزي ( DA ) يدرس العلاقات السببية بين المتغيرات ويستخدم كوسيلة استكشافية في فهم العلاقات السببية بدرجة كافية . وهو يشبه الى حد كبير اسلوب تحليل الانحدار (Regression Analysis) . ففي اسلوب الانحدار نجد بان المتغير المعتمد (Y) Dependent Variable هو متغير مستمر ( Continuous Variable ) ولكن متى ما وجد بان هذا المتغير هو متغير اسمي (Nominal Variable) فان الباحثون يلجأون الى احد خيارين هما :

**أولاً :** التحليل التمييزي (DA) ( Discriminant Analysis ) ويستعمل التحليل التمييزي عندما تتوزع بيانات المتغير المعتمد (Y) توزيعاً طبيعياً ( Normally ) . ويدخل هذا ضمن الاسلوب المعلمي للاحصاء ( Parametric Technique ) .

**ثانياً :** الانحدار اللوجستي ( Logistic Regression ) ويستعمل عندما لا تتوزع البيانات للمتغير المعتمد ( Y ) توزيعاً طبيعياً . ويدخل هذا ضمن الاسلوب اللامعلمي للاحصاء (Nonparametric Technique) .

وعليه فانه في اغلب الاحيان يستعمل تحليل الانحدار عندما يراد تأسيس او اقامة العلاقة بين المتغيرات بحيث يكون فيها المتغير المعتمد متغيراً مستمراً ، اما في حالة انتهاك مثل هذه الفرضية فان استعمال تحليل الانحدار لا يكون مناسباً دائماً ، فعلى سبيل المثال لو كان استعمال دخول موقع للانترنت فان المتغير المعتمد هو متغير اسمي ويقسم الى مستويين مستعمل (user=1) وغير مستعمل (nonuser=2) . وفي مثل هذه الحالة فان تحليل الانحدار لا يكون على طول الزمن مناسباً وان اختيارنا هو التحليل التمييزي عندما تكون العينة مختارة من مجتمع يتبع التوزيع الطبيعي اذ في حالة انتهاك مثل هذه الفرضية نستخدم الانحدار اللوجستي الذي يمثل أحد اختبارات التوزيع الحر وهو احد الاختبارات التي لانتاج فيها الى متطلبات التوزيع الطبيعي .

## **2-1 مشكلة البحث :**

ان التطورات العلمية تؤدي بطبيعة الحال الى زيادة المعرفة والدقة في تقليل الاخطاء وتسمح بسرعة ايجاد الحلول المناسبة . ومن الواضح انه اذا كانت قيم المتغيرات المستقلة X's تختلف اختلافاً واضحاً بالنسبة للمفردات في المجموعتين فهذا يعني وجود مشكلة لأمكانية التمييز بين المجموعتين وانه من الممكن تعيين المفردات الجديدة لأي من المجموعتين باستخدام التمييز بكفاءة

عالية اذ ان العوامل المؤدية للإصابة بالإسهال عند الاطفال حديثي الولادة تختلف قيمها بين المصابين بالإسهال وغير المصابين بالإسهال .

## **3-1 أهمية البحث :**

تكمّن أهمية البحث في استخدام اسلوب التحليل التمييزي كأداة او وسيلة مهمة تساعد المخططين في تصنيف الأطفال المعرضين للإصابة بالإسهال الى فئتين مصاب بالإسهال وغير مصاب بالإسهال مع تحديد العوامل المهمة المؤدية للإصابة بذلك المرض .

## **4-1 اهداف البحث :**

### **Research Aims**

يمكن تلخيص اهداف البحث بالاتي :

1. تحديد العوامل والمتغيرات المؤثرة في حالة الاسهال عند الطفل في الاسبوعين الاخيرين للإصابة وهما ( عمر الطفل ، الرضاعة الطبيعية ، شرب السوائل قبل يوم من الإصابة بالاسهال ، الجنس ، تحصيل الام الثقافي ) . كمتغيرات مستقلة مؤثرة على الإصابة بالاسهال كمتغير نوعي مصاب بالاسهال (yes=1) وغير مصاب بالاسهال ( No=2 ) .
2. اختبار وجود فروق معنوية ذات دلالة احصائية بين المجموعتين مصاب بالاسهال (yes=1) وغير مصاب بالاسهال (No=2) بالنسبة للمتغيرات المستقلة .
3. تحديد المتغيرات المستقلة التي تساهم باكبر قدر من الاختلاف بين المجموعتين للمتغير التابع .
4. تقسيم الحالات بين فئات المتغير التابع بناءً على قيم المتغيرات المستقلة .
5. تقويم دقة التقسيم .

### **5-1 فروض البحث<sup>[10]</sup> :**

يعتمد البحث على الفروض الاحصائية الاتية :

1. ان العينة المختارة من مجتمع يتبع التوزيع الطبيعي ، بمعنى ان بيانات المتغيرات تتبع التوزيع الطبيعي .
2. تجانس تباين المجموعتين .
3. يجب وجود مجموعتين قابلتين للتمييز على الاقل .
4. كل مجموعة من المجموعات المختارة يجب ان تكون معرفة ومحددة مسبقا وتختلف كل مجموعة عن المجموعات الاخرى .
5. ان تكون قيم المتغيرات مختلفة اختلافا واضحا بالنسبة للمفردات في المجموعات .
6. عدم وجود ارتباط عال بين المتغيرات المستقلة فان كان هنالك ارتباطا عالياً كان لزاماً ازالة بعض هذه المتغيرات من التحليل .
7. عدم وجود قيم شاذة في بيانات العينة المختارة .

### **6-1 عينة البحث<sup>[3]</sup> :**

لقد تم الحصول على البيانات من المسح المتعدد المؤشرات MICS4 الذي تم تنفيذه من قبل الجهاز المركزي للإحصاء وتكنولوجيا المعلومات في العام ٢٠١١ اذ نفذ الجهاز المركزي للإحصاء وهيئة اقليم كردستان هذا المسح بالتعاون مع وزارة الصحة وقدمت منظمة الامم المتحدة للطفولة ( اليونيسيف ) الدعم المالي والفني لتنفيذ هذا المسح . والمسح العنقودي متعدد المؤشرات برنامج دولي لاجراء دراسات مسحية على الاسر اعدته وطورته منظمة اليونيسيف وقد اجري المسح العنقودي متعدد المؤشرات في العراق كجزء من الجولة العالمية الرابعة للمسوح العنقودية متعددة المؤشرات ( MICS4 ) . ويوفر هذا المسح معلومات حديثة عن وضع الاطفال والنساء وقياس المؤشرات الاساسية التي تسمح للدول رصد ماتحرزه من تقدم تجاه تحقيق الاهداف الانمائية للاللفية ( MDGs ) ، الى جانب قياس مستوى الوفاء بالالتزامات الاخرى المتفق عليها دوليا . ولبيان حالة الإصابة بمرض الاسهال عند الاطفال تم اخذ ( 1997 ) حالة كانت تخص محافظة واسط ، اذ ان هذا العدد يمثل حجم العينة في هذا البحث .

## المبحث الثاني الجانب النظري

### 1-2 المعادلة الخطية للتحليل التمييزي [2]:

#### Discriminant Analysis Linear Equation

يعرف التحليل التمييزي أو تحليل الدالة التمييزية بأنه أسلوب احصائي لتحليل المتغيرات المتعددة في تحديد المتغيرات الكمية (أو المتنبأ بها) (Predictors) أو المتغيرات المستقلة بأفضل تمييز (Discriminant) بين مجموعتين معرفتين أو أكثر. وإن هذا التحليل يجعل الدالة التمييزية التي تعد توضيحاً خطياً للأوزان (Weightings) والعلامات أو الحالات (Scores) لهذه المتغيرات وإن معادلة التحليل التمييزي (DA) تشبه صيغة معادلة الانحدار وتكتب بالصيغة الآتية:

$$Z = A + B_1 x_1 + B_2 x_2 + \dots + B_p x_p$$

A : ثابت ويمثل المقطع الطولي لمحور الصادات

B : المعاملات التمييزية أو الوزن للمتغير

x : علامات الاستجابة للمتغير

P : عدد المتغيرات المستقلة

Z : الدالة التمييزية

وإن المعاملات B's هي معاملات تمييزية غير معيارية (Unstandard) وهي مشابهة (أو منازرة) إلى معاملات الانحدار  $\hat{\beta}, s$ . وهذه المعاملات تعظم المسافة بين الأوساط لمعيار المتغير المعتمد الاسمي (Y) أما معاملات التمييز المعيارية (Standard) فيمكن كذلك استعمالها وهي تشبه الأوزان (b's) في الانحدار.

وإن عدد الدوال التمييزية يحدد بحسب القاعدة الآتية:

العدد الأقل لـ (عدد المجموعات - 1) أو (عدد المتغيرات الكلية) يفرض لدينا ثلاث مجموعات وخمس متغيرات كمية فإن عدد الدوال التمييزية يساوي

$$\text{عدد المجموعات} - 1 = 5 - 1 = 4$$

$$\text{عدد المتغيرات الكلية} = 5$$

وعليه فإن عدد الدوال التمييزية يساوي (4) لأن هذا الرقم هو الأقل والشيء نفسه في بحثنا فإن عدد المجموعات يساوي اثنين وعدد المتغيرات يساوي خمس فإن عدد الدوال التمييزية يساوي

$$\text{عدد المجموعات} - 1 = 2 - 1 = 1$$

$$\text{عدد المتغيرات الكلية} = 5$$

وعليه فإن عدد الدوال التمييزية هو دالة واحدة.

### 2-2 الدالة المميزة الخطية في حالة مجموعتين [6]

#### The Linear Discriminant Function-Two Groups

ان دالة التمييز هي نموذج يمكن صياغته اعتماداً على مؤشرات العينة التي اختيرت مفرداتها بشكل عشوائي ووضعت في مجموعتين مختلفتين ، بواسطة هذه الدالة نستطيع ان نختبر المفردة ونحدد عانديتها الى اي مجموعة . فلو فرضنا ان مجال العينة هو W فسوف يقسم الى قسمين :

( R ) يعود الى المجموعة الاولى ، و ( W-R ) يعود الى المجموعة الثانية ، اما الحد الفاصل بين المجموعتين فيمكن ان يعود الى اية مجموعة من هاتين المجموعتين وفي هذه الحالة فإن :

$$f_1(\underline{X}) = f_2(\underline{X})$$

$$\text{Log } f_1 = \text{Log } f_2$$

وبافتراض ان مفردات المجموعتين تتوزع توزيعاً طبيعياً متعدد المتغيرات بمتوسطي  $\mu_1, \mu_2$

ومصفوفة تباين وتباين مشترك  $\Sigma$  .

$$(\underline{X} - \underline{\mu}_1)' \Sigma^{-1} (\underline{X} - \underline{\mu}_1) = (\underline{X} - \underline{\mu}_2)' \Sigma^{-1} (\underline{X} - \underline{\mu}_2)$$

$$(\underline{X} - \underline{\mu}_1)' \Sigma^{-1} (\underline{X} - \underline{\mu}_1) - (\underline{X} - \underline{\mu}_2)' \Sigma^{-1} (\underline{X} - \underline{\mu}_2) = 0$$

$$\underline{X}' \Sigma^{-1} (\underline{X} - \underline{\mu}_1) - \underline{\mu}_1' \Sigma^{-1} (\underline{X} - \underline{\mu}_1) - \underline{X}' \Sigma^{-1} (\underline{X} - \underline{\mu}_2) + \underline{\mu}_2' \Sigma^{-1} (\underline{X} - \underline{\mu}_2) = 0$$

$$\underline{X}' \Sigma^{-1} \underline{X} - \underline{X}' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{X} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{X}' \Sigma^{-1} \underline{X} + \underline{X}' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma^{-1} \underline{X} + \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2 = 0$$

$$\underline{X}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X} + (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = 0$$

$$\underline{X}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X}'$$

$$2 \underline{X}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) + (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = 0$$

$$\underline{X}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = 0$$

وهي تمثل قيم دالة التمييز عند الحد الفاصل بين المجموعتين عندما تكون معالم المجتمع معلومة اما

عندما تكون المعالم مجهولة فتقدر اعتماداً على قيمة العينتين  $n_1, n_2$  وبأوساط حسابية  $\bar{x}_1, \bar{x}_2$  .

وبتقدير مصفوفة التباين والتباين المشترك فإن دالة التمييز تكون :

$$\underline{X}' S^{-1} (\bar{X}_1 - \bar{X}_2)' - \frac{1}{2} (\bar{X}_1 + \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2) = 0 \quad \dots\dots\dots (1)$$

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

$$S = \frac{1}{n_1 + n_2 - 2} (\underline{X}_1' \underline{X}_2 - \underline{X}_2' \underline{X}_2)$$

$S_1, S_2$  تمثل تقدير مصفوفة التباين والتباين المشترك اعتماداً على العينة الاولى والثانية على التوالي . والمعادلة رقم (1) تمثل الدالة المميزة الخطية المقدرة عند الحد الفاصل بين المجموعتين ، نعتمد على هذه المعادلة في تصنيف اية مفردة جديدة ، فالمفردة اما تعود الى المجموعة الاولى اذا كان الجانب الايسر اكبر من صفر ، او تعود الى المجموعة الثانية اذا كان اقل من صفر ، وتصنف عشوائيا فيما عدا ذلك .

فاذا كان لدينا المتغيرات  $(X_1, X_2, \dots, X_p)$  فالدالة الخطية تأخذ الصيغة التالية :

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_p X_p$$

حيث  $Y$  تمثل تركيبة خطية للمتغيرات التوضيحية تدعى بدالة التمييز وتكتب بالصيغة التالية وعلى شكل مصفوفة :

$$Y = \underline{b}'X$$

حيث ان المعاملات  $(b_1, b_2, \dots, b_p)$  تقدر بحيث تجعل الدالة تعطي افضل تمييز بين المجموعتين وذلك بجعل مربع الفرق بين متوسطي المجموعتين الى التباين ( مجموع المربعات داخل المجموعات ) اكبر مايمكن .

$$Q = \frac{\text{Between groups}}{\text{Within groups}}$$

$$Q = \frac{[\bar{Y}_1 - \bar{Y}_2]^2}{\sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2} = \frac{[\underline{b}'\mu_1 - \underline{b}'\mu_2]^2}{\sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}$$

حيث ان

$$\bar{Y}_1 = b_1 \bar{X}_{i1} + b_2 \bar{X}_{i2} + \dots + b_p \bar{X}_{ip}$$

وعندما تكون المعلمات غير معلومة نستخدم تقديرات العينة  $\bar{X}_1$  بدلاً من  $\mu_1$  ،  $S$  بدلاً من  $\sum$  حيث ان

:

$$\bar{X}_1 = (\bar{X}_{i1}, \bar{X}_{i2}, \dots, \bar{X}_{ip}) \quad i = 1, 2$$

$$S = \frac{1}{n_1 + n_2 - 2} (X_1' X_1 - X_2' X_2)$$

$$\therefore Q = \frac{[\underline{b}'(\bar{X}_1 - \bar{X}_2)]^2}{\sum_{i=2}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_1)^2}$$

نفرض ان  $\bar{X}_1 - \bar{X}_2 = \underline{\alpha}$

$$\begin{aligned}
d &= \bar{y}_1 - \bar{y}_2 \\
&= \underline{b}'(\bar{X}_1 - \bar{X}_2) \\
D &= \bar{Y}_1 - \bar{Y}_2 = \underline{b}'\alpha \\
d^2 &= (\underline{b}'\alpha)^2
\end{aligned}$$

ومجموع مربعات الفرق داخل المجاميع يحسب كما يلي :

$$(Y_{ij} - \bar{y}_i) = b_1(X_{1ij} - \bar{X}_{1i}) + b_2(X_{2ij} - \bar{X}_{2i}) + \dots + b_p(X_{pij} - \bar{X}_{pi})$$

وبأخذ المجموع وتربيع طرفي المعادلة نحصل على :

$$\sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \underline{b}'S\underline{b}$$

$$Q = \frac{(\underline{b}'\alpha)}{\underline{b}'S\underline{b}}$$

$$\frac{\partial Q}{\partial \underline{b}} = \frac{\underline{b}'S\underline{b} * 2(\underline{b}'\alpha)\alpha - (\underline{b}'\alpha)^2 * 2S\underline{b}'}{(\underline{b}'S\underline{b})^2} = 0$$

$$\underline{b}'S\underline{b} * \underline{b}'\alpha\alpha - (\underline{b}'\alpha)^2 * S\underline{b} = 0$$

وبقسمة الطرفين على  $(\underline{b}'\alpha)^2$

$$\frac{\underline{b}'S\underline{b}}{\underline{b}'\alpha}\alpha - S\underline{b} = 0$$

$$\frac{\underline{b}'S\underline{b}}{\underline{b}'\alpha}\alpha = S\underline{b}$$

$$\underline{b}' = S^{-1} \frac{\underline{b}'S\underline{b}}{\underline{b}'\alpha}\alpha$$

ويمكن اعتبار المقدار  $\frac{\underline{b}'S\underline{b}}{\underline{b}'\alpha}$  ثابتا ومساويا الى الواحد فإن الحل سيكون كالآتي :

$$\begin{aligned}
\underline{b}' &= S^{-1}\alpha \\
&= S^{-1}(\bar{X}_1 - \bar{X}_2)'
\end{aligned}$$

وهي توليفة خطية ناتجة عن تعظيم نسبة التباين بين المجاميع الى داخل المجاميع .  
 ففي حالة وجود مجموعتين يكون لدينا دالة تمييز واحدة فقط ، وفي حالة وجود ثلاث مجاميع يكون لدينا دالتين تمييزيتين . وبعد استخراج المعاملات  $\underline{b'}$  ، تصنف المشاهدات الى احدى المجموعتين بالاعتماد على نقطة وسط المجموعتين (L) التي تجعل احتمال التصنيف الخاطيء اقل مايمكن .

$$L = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

تصنف المشاهدات الى المجموعة الاولى اذا كانت  $\hat{Y} > L$  .

تصنف المشاهدات الى المجموعة الثانية اذا كانت  $\hat{Y} < L$  .

تصنف المشاهدات عشوائيا الى المجموعة الاولى او الثانية اذا كانت  $\hat{Y} = L$  .  
 حيث ان :

$$\hat{Y} = (\bar{X}_1 - \bar{X}_2)' S^{-1} X = \hat{\alpha} X$$

$$Y_1 = (\bar{X}_1 - \bar{X}_2)' S^{-1} X_1$$

$$\bar{Y}_1 = \hat{\alpha}' \bar{X}_1$$

$$\bar{Y}_2 = \hat{\alpha}' \bar{X}_2$$

حيث ان :  $\hat{\alpha}' = (\bar{X}_1 - \bar{X}_2) S^{-1}$  و  $\bar{Y}_1, \bar{Y}_2$  تمثل متوسطي المجموعتين الاولى والثانية .

### 3-2 اختبار معنوية الدالة المميزة الخطية :

#### 1-3-2 اختبار $T^2$ -Hotelling :

عندما يراد التمييز بين مجموعتين ، فانه يمكننا ان نختبر الفرضية التي تنص على تساوي متوسطات المجموعات :

$$H_0 : M_1 = M_2$$

vs

$$H_1 : M_1 \neq M_2$$

وان احصاء الاختبار المستخدمة في حالة التمييز بين مجموعتين هي  $T^2$  Hotelling وصيغته الرياضية بالشكل التالي :

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} * D^2 \quad \text{حيث ان :}$$

$$D^2 = (\bar{X}_1 - \bar{X}_2)' S^{-1}_{pooled} (\bar{X}_1 - \bar{X}_2)$$

وباستخدام اختبار F والتي تكون صيغته بالشكل التالي :

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} * T^2$$

وبدرجات حرية  $(p, n_1 + n_2 - p - 1)$  . وترفض  $H_0$  بمستوى معنوية  $F_\alpha$  اذا كانت

$$F_{cal} > F_{\alpha}(p, n_1 + n_2 - p - 1)$$



ونقبل  $H_1$  وهذا يدل على ان المتوسطات المجموعتين غير متساوية ، اي يوجد فرق معنوي بين المجموعتين ، ومهناه ان الدالة المميزة الخطية قابلة للتمييز بدرجة عالية .

## 2-3-2 اختبار ولكس Wilks :

الاحصاءة الاخرى للاختبار هي مقياس ولكس Wilks-Criteria او ولكس لمدا Wilks Lamda وهو مقياس يستخدم في اختبار الفرضيات في التحليل متعدد المتغيرات لايجاد هل ان هناك اختلاف بين متوسطات المجموعات المكونة من خليط من المتغيرات في حالة التحليل التمييزي ، وصيغته بالشكل التالي [9] :

$$\Lambda = \frac{|w|}{|T|} = \frac{|w|}{|w + B|}$$

W : مصفوفة التباين والتباين المشترك داخل المجموعات .

T : مصفوفة التباين والتباين المشترك الكلي .

B : مصفوفة التباين والتباين المشترك بين المجموعات .

تتراوح قيمة (  $\Lambda$  ) بين الصفر والواحد ، فإذا كانت مساوية للواحد فإن ذلك يشير الى ان متوسطات المجموعات متساوية وبذلك فلا يوجد تمييز بين المجموعات ، وهذا يعني ان دالة التمييز فاشلة . اما اذا كانت قيمتها اقل من الواحد او قريبة من الصفر فإن ذلك يدل على قوة التمييز .

## 3-3-2 اختبار بارتليت Barttlete:

يعد هذا المقياس اكثر دقة من مقياس (  $\Lambda$  ) وتكون صيغته الرياضية كالآتي :

$$x^2 = -N \log (\Lambda)$$

ويكون توزيعه مقارباً الى  $x^2$  بدرجة حرية (P(K-1) ، وقد طورت صيغته من قبل Barttlete الى الشكل التالي :

$$x^2 = - \left[ N - 1 - \frac{(P + K)}{2} \right] \log (\Lambda)$$

بدرجة حرية (P(K-1)

## 4-3-2 مقياس F :

وبالنظر لصعوبة الحصول على القيمة الجدولية لعدم توفر الجداول الخاصة بمقياس ولكس ، وضع معياراً بديلاً اخر من قبل Rao ، وهذا المعيار هو مقياس F وصيغته الرياضية هي :

$$F = \frac{1 - \Lambda^{\frac{1}{s}}}{\Lambda^{\frac{1}{s}}} * \frac{ms - 2\lambda}{P(K - 1)}$$

بدرجات حرية  $df_1 = P(K - 1)$

$df_2 = ms - 2\lambda$

$$m = N - \frac{1}{2} [P + K]$$

$$S = \left[ \frac{P^2 (1 - K)^2 - 4}{(1 - K)^2 + P^2 - 5} \right]^{1/2}$$

$$\lambda = \frac{P(K - 1) - 2}{4}$$

## 4-2 احتمال خطأ التصنيف :

هناك نوعان من احتمال خطأ التصنيف هما :

1. احتمال خطأ التصنيف  $P_{12}$  وهو احتمال تصنيف المفردة الى المجموعة الثانية وهي اصلا

تعود الى المجموعة الاولى  $P_{12} = \phi \left( -\frac{\delta}{2} \right)$  حيث ان  $\phi$  تمثل دالة التوزيع القياسي

2. احتمال خطأ التصنيف  $P_{21}$  وهو احتمال تصنيف المفردة الى المجموعة الاولى وهي اصلا

تعود الى المجموعة الثانية  $P_{21} = \phi \left( -\frac{\delta}{2} \right)$

$\delta^2$  يمثل مربع مسافة وكذلك  $D^2$  يمثل مربع مسافة مقياس مهالونوبيس اذن :

$$\delta^2 = D^2 = (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)$$

وبذلك سوف يكون تقدير احتمال التصنيف :

$$P_{21} = P_{12} = \phi \left( -\frac{D}{2} \right)$$

حيث ان D هو جذر مقياس مهالونوبيس .

## المبحث الثالث

### الجانب التطبيقي

## 1-3 المقدمة<sup>[7]</sup> :

يمثل مرض الإسهال ثاني أهم أسباب وفيات الأطفال دون سن الخامسة إذ أنه يؤدي بحياة ١.٥ مليون طفل كل عام. ويمكن أن يدوم هذا المرض عدة أيام ويترك الجسم دون كمية الماء والأملاح الضرورية للبقاء على قيد الحياة. ومعظم الذين يقضون نحبهم بسبب الإسهال يتوفون في واقع الأمر جرّاء الإصابة بالجفاف الوخيم ، ونتيجة ضياع سوائل أجسامهم. ويواجه الأطفال الذين يعانون من سوء التغذية أو خلل في المناعة أكبر مخاطر الوفاة بسبب الإسهال. والإسهال هو إخراج براز رخو أو سائل ثلاث مرّات أو أكثر في اليوم (أو بوتيرة تفوق البوتيرة المعتادة). ولا يُعد إخراج البراز المتشكّل إسهالاً، شأنه شأن البراز الرخو أو العجيني الذي يخرج الرضيع. ذلك أنّ الإسهال يمثل، عادة أحد أعراض الإصابة بنوع من أنواع العدوى التي تسببها طائفة متنوعة من الجراثيم أو الفيروسات أو الطفيليات. وتنتشر العدوى، عادة، عن طريق الأغذية أو مياه الشرب الملوثة أو من شخص إلى آخر بسبب تدني مستوى النظافة الشخصية. ويمكن علاج الإسهال بمحلول من الماء النقي والسكر والملح وبأقراص الزنك.

وقد تم دراسة المتغيرات التالية لكل مجموعة من المصابين وغير المصابين بالإسهال عند الأطفال

١- عمر الطفل      ٢- الرضاعة الطبيعية      ٣- تناول الحليب الصناعي

٤- جنس الطفل  
٥- مستوى تعليم الأم  
وكانت توزيع مفردات العينة حسب المجموعات ونسبة كل مجموعة على النحو التالي :

الجدول رقم (١)

المجموعات	اسماء المجموعات	عدد الحالات	الاحتمال
1	مصاب	296	0.15
2	غير مصاب	1683	0.85
	المجموع	1979	1.00

### 2-3 المتغيرات المستخدمة في الدراسة:

أ- المتغير المعتمد الذي يمثل الإصابة بمرض الإسهال :  
(١) مصاب (٢) غير مصاب (متغير نوعي اسمي)  
ب- المتغيرات المستقلة :

- ١- عمر الطفل ( $X_1$ ) متغير كمي
- ٢- الرضاعة الطبيعية ( $X_2$ ) نعم (1) ، لا (2) ، متغير نوعي اسمي
- ٣- تناول الحليب الصناعي ( $X_3$ ) نعم (1) ، لا (2) ، متغير نوعي اسمي
- ٤- جنس الطفل ( $X_4$ ) ذكر (1) ، انثى (2) ، متغير نوعي اسمي

٥- مستوى تعليم الأم ( $X_5$ ) غير متعلمة (1) ، ابتدائي (2) ، ثانوي (3) ، تعليم غير نظامي (4)  
الجدول رقم (٢)

يبين الاوساط الحسابية والانحرافات المعيارية للمتغيرات المستقلة وذلك لكل مجموعة من المجموعتين  
ولاجمالي المجموعتين : المصابين وغير المصابين

Variables	Mean	Std.Deviation
Yes $X_1$	1.40	1.272
$X_2$	1.12	0.323
$X_3$	1.75	0.432
$X_4$	1.50	0.501
$X_5$	2.01	0.722
No $X_1$	2.05	1.408
$X_2$	1.10	0.305
$X_3$	1.88	0.323
$X_4$	1.51	0.500
$X_5$	2.01	0.724
Total $X_1$	1.96	1.408
$X_2$	1.11	0.307
$X_3$	1.86	0.344
$X_4$	1.51	0.500
$X_5$	2.01	0.724

### 2-3 اختبارات الفروض الأساسية :

عندما يراد التمييز بين مجموعتين فإنه يمكننا ان نختبر الفروض التالية<sup>[4]</sup> :  
أولاً : ان العينة المختارة من مجتمع يتبع التوزيع الطبيعي ، بمعنى ان بيانات المتغيرات تتبع التوزيع الطبيعي .

الجدول رقم (٣)

يبين ان كل المتغيرات مستوى الدلالة فيها اكبر من 0.05 اي يتبع التوزيع الطبيعي

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
N	1979	1979	1979	1979	1979
Normal Parameters Mean	1.96	1.11	1.86	1.51	2.01
Kolmogrov-Smirnov z	7.394	23.524	23.028	15.455	11.016
Asymp.Sig.(2-tailed)	0.15	0.08	1.37	0.17	0.13

ثانياً : تجانس تباين المجموعتين .

الجدول رقم (٤)

يبين قيم Log Determinant تقريباً متساوية للمجموعتين ويذكر البياتي<sup>[1]</sup> ان تساوي قيم Log Determinant نسبياً لجميع المجموعات يفترض تجانس المصفوفات لتباينات المشتركة

	Rank	Log Determinant
Yes	5	-5.760
No	5	-6.213
Pooled within-groups	5	-6.110

الجدول رقم (٥)

يشير الى مستوى المعنوية 0.000 باستخدام اختبار Box's M

Box's M	70.397
F	4.664
Approx.	15
Df1	1073425.138
Df2	0.000
Sig.	

ثالثاً : عدم وجود ارتباط بين المتغيرات المستقلة فان كان هنالك ارتباط كان لازماً ازالة بعض هذه المتغيرات من التحليل

الجدول رقم (٦)

يبين عدم وجود ارتباط ذاتي بين المتغيرات لأن قيم (VIF) اقل من 5.00<sup>[4]</sup>

	Tolerance	VIF
X <sub>1</sub>	0.857	1.167
X <sub>2</sub>	0.929	1.076
X <sub>3</sub>	0.792	1.263
X <sub>4</sub>	0.998	1.002
X <sub>5</sub>	0.981	1.020

رابعاً : عدم وجود قيم شاذة في بيانات العينة المختارة

الجدول رقم (٧) يعرض فقط اول 20 حالة ويبين عدم وجود قيم شاذة وذلك بأجراء اختبار Mahalanobis علما انه تم التأكد لكل الحالات لعدم وجود قيم شاذة لأن قيم MA11.1 اقل من القيمة الجدولية لمربع كاي عند درجة حرية ٤ (عدد المجموعات-١) ومستوى دلالة 0.001 وتساوي 18.47 نقبل فرضية العدم لعدم وجود قيم شاذة<sup>[4]</sup>

جدول رقم (٧)  
يبين عدم وجود قيم شاذة في بيانات العينة المختارة

	MA11.1
1	5.194469
2	3.834207
3	3.589115
4	3.333723
5	1.255548
6	3.483634
7	3.685411
8	5.148921
9	3.699579
10	3.106511
11	11.38685
12	3.685411
13	9.517074
14	8.609129
15	3.483634
16	3.139846
17	2.025003
18	3.38669
19	1.255548
20	3.38669

خامساً : وجود مجموعتين قابلتين للتمييز على الأقل  
الجدول رقم (٨) يبين المقاييس المستخدمة في التحليل الإحصائي : يبين اختبار مربع كاي اكبر من قيمها الجدولية اذ يمكن الاستنتاج ان هناك فروقاً معنوية بين المجموعتين تعود الى المتغيرات الخمسة اي ان هناك تمييز بين المجموعتين ومعيار wilks قيمته اقل من واحد وهذا يدل على قوة التمييز للدالة

الجدول رقم (٨)  
يبين اختبار wilks لقوة تمييز الدالة

Test	Wilks' Lambda	Chi-square	df	Sig.
1	0.900	67.655	5	0.000

### 3-3 اختبار معنوية المتغيرات في الدالة التمييزية:-

يتم اختبار معنوية جميع المتغيرات وذلك لمعرفة أهمية كل متغير في الدالة ومدى تأثيره في تحليل النتائج باستخدام اختبار F .

الجدول رقم (٩)  
يبين اختبار F لكل متغير في الدالة المميزة الخطية

variables	F	Df1	Df2	Sig.
X <sub>1</sub>	55.978	1	1977	0.000
X <sub>2</sub>	0.588	1	1977	0.443
X <sub>3</sub>	35.590	1	1977	0.000
X <sub>4</sub>	0.100	1	1977	0.751
X <sub>5</sub>	0.006	1	1977	0.940

نلاحظ من خلال الجدول ان المتغير (X<sub>1</sub>) العمر يمتاز بمعنوية عالية وله تأثير كبير في عملية التصنيف يليه المتغير (X<sub>3</sub>) تناول الحليب الصناعي له تأثير معنوي ولكن بشكل اقل ثم تليه المتغيرات (X<sub>2</sub>) الرضاة الطبيعية و (X<sub>4</sub>) الجنس و (X<sub>5</sub>) مستوى تعليم الأم حيث كان تأثيرها ضعيف .  
تحتسب الدلة المميزة للمجموعتين باستخدام نموذج Fisher للدالة المميزة وفق الصيغة التالية :

$$Z = A + B_1 x_1 + B_2 x_2 + \dots + B_p x_p$$

ويتم تقدير معلمات الدالة المميزة B وفق الصيغة التالية :

$$B = S^{-1} (\bar{x}_{pi} - \bar{x}_{qi}) \quad p = q$$

الجدول رقم (١٠)  
يبين تقدير معاملات الدالة المميزة

variables	Function
X <sub>1</sub>	0.528
X <sub>2</sub>	0.142
X <sub>3</sub>	1.397
X <sub>4</sub>	0.045
X <sub>5</sub>	0.063
constant	-3.987

نلاحظ ان المتغير (X<sub>3</sub>) تناول الحليب الصناعي ذو تأثير عالي قياسا بالمتغيرات الاخرى يليه المتغير (X<sub>1</sub>) عمر الطفل يليه المتغيرات (X<sub>2</sub>) الرضاة الطبيعية و (X<sub>5</sub>) مستوى تعليم الأم و (X<sub>4</sub>) الجنس .

### 4-3 حساب احتمال التصنيف الصحيح :-

الجدول رقم (١١)

### يبين نتائج التصنيف التنبئي

	Predicted Group Membership		Total
	Yes	No	
Original Count Yes	193	103	296
No	666	1017	1683
% Yes	65.2	34.8	100.0
No	39.6	60.4	100.0

نسبة التصنيف الصحيح في دالة التمييز الخطية كانت (61%) اذ ان (1207) مفردة من اصل (1979) مفردة صنفت بشكل صحيح الى المجاميع التي تنتمي اليها ، بينما (772) مفردة صنفت بشكل خاطيء ، وسوء التصنيف هذا يحدث بسبب التشابه الكبير بين المتغيرات الموجودة في الدالة .

### References

### المصادر :

#### أولاً : المصادر العربية

1. البياتي ، محمود مهدي ، تحليل البيانات الاحصائية باستخدام البرنامج الاحصائي spss ، عمان ، دار الحامد ، (2005) .
2. الجاعوني ، احمد فريد ، د. عدنان غانم " التحليل الاحصائي متعدد المتغيرات ( التحليل التمييزي ) في توصيف وتوزيع الاسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع " ، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية ، المجلد 23 العدد الثاني (2007) .
3. المسح العنقودي متعدد المؤشرات ، 2011 ، [www.childinf.org](http://www.childinf.org) ،
4. د. جودة، محفوظ " التحليل الاحصائي المتقدم باستخدام Spss " ، دار وائل للنشر ، الطبعة الاولى ، عمان ( 2008 ) .
5. جونسن، ريتشارد ، دين وشرن ترجمة د. عبد الرحمن حامد عزام " التحليل الاحصائي للمتغيرات المتعددة من الجهة التطبيقية " ، دار المريخ للنشر ، الرياض ، المملكة العربية السعودية ص . ص ( 735-799 ) ( 1997 ) .
6. صالح ، عائدة هادي " استخدام التحليل المميز لتشخيص بعض امراض العيون " ، مجلة الادارة والاقتصاد ، العدد السابع والستون ( 2008 ) .
7. منظمة الصحة العالمية ، مركز وسائل الاعلام ، صحيفة وقائع رقم 331 . تشرين الثاني 2012 ، [www.Who.int](http://www.Who.int) .

#### ثانياً : المصادر الاجنبية

8. Donald F. Morrisn " Multivariate Statistical Methods , 2<sup>nd</sup> Edition, Graw-hill boole Company, Newyork .PP ( 230-245 ) ( 1976 ) .
9. [www.en.wikipedia.org/wiki/wilks-lambda-distribution](http://www.en.wikipedia.org/wiki/wilks-lambda-distribution) .
10. [www.uk.sagepub.com/.../chapter%20-Dis](http://www.uk.sagepub.com/.../chapter%20-Dis)