طريقة لتقدير معاملات الارتباط القويم في توزيع متعدد المتغيرات

م.م انتصار عبيد حسون

الخلاصة :

يتناول هذا البحث تقدير معاملات الارتباط القويم وكذلك تقدير مصفوفة التباين والتباين المشترك بين مجموعة المتغيرات، وتم التوصل ايضا الى ان قيم الارتباط القويم للمتغيرات الاصلية والتراكيب الخطيةهي نفسها ، وثم تم تقدير واختبار معنوية معامل الارتباط القويم.

كلمات مفتاحية :

الارتباط القويم ، تحليل الارتباط القويم ، المتغيرات القويمة ، المتغيرات القانونية ، مصفوفة التباين والتباين المشترك .

المقدمة :

يسمى المقياس الإحصائي الذي يقيس العلاقة بين المتغير x والمتغير y بمعامل الارتباط البسيط وهناك مقاييس أخرى منها الارتباط الجزئي والارتباط المتعدد ، أما المقياس الإحصائي الذي يستخدم لقياس العلاقة بين مجموعة المتغيرات

$(x_1, x_2, \ldots\ldots. x_p)$

ومجموعة ثانية من المتغيرات هي :

$(x_{p+1}, x_{p+2}, \ldots\ldots. x_q)$

فيسمى بالارتباط القويم (Canonical correlation) وقد أشار إليه كثير من الباحثين نذكر منهم الباحث **Cooley** عام ١٩٧١ والباحث (**Rancher,** ١٩٩٦) ، **الباحث Tabachick عام ١٩٩٦ .**

وتركزت معظم البحوث على تحديد قوة العلاقة بين المجموعة X_p ، والمجموعة X_q من **خلال إجراء تحليل كامل متعدد المتغيرات يقوم على إيجاد مصفوفة التباين والتباين المشترك** بين المتغير المعتمدY والمتغيرات $(x_1, x_2, \ldots\ldots. x_p)$ ، وكذلك حساب الاعتمادية بين **مجموعة المتغيرات p ومجموعة المتغيرات q من خلال وضع تراكيب خطية لكل متغير x بدلالة مجموعة المتغيرات الأخرى$(x_{p+1}, x_{p+2}, \ldots\ldots. x_{p+q})$.**

وهذا مهم جداً في البحوث التطبيقية عندما يرغب الباحث في تحليل العلاقات بين مجموعة كبيرة من المتغيرات خاصة في التجارب الفيزيائية ، والطبية ، والكيميائية ، وفي بحثنا هذا نحاول عرض موضوع الارتباط القويم بشكل مبسط وتوضيح المصفوفات التي نحتاج إليها والتراكيب الخطية المهمة بن المتغيرات ، وكذلك مقدر الارتباط القويم وإيجاد إحصاءه أختبار لاختبار فرضية الارتباط القويم .

هدف البحث :

يهدف البحث إلى تقدير معاملات الارتباط القويم للمجتمع اعتمادا على مجموعات المشاهدات المستقلة المأخوذة من العينات، وكذلك نبرهن ان قيم الارتباط القويم للمتغيرات الأصلية $(X_1, X_2, \ldots X_{p+9})$ والتراكيب الخطية $(Y_1, Y_2, \ldots Y_{p+9})$ هي نفسها ، كذلك نهدف إلى تقدير مصفوفة التباين والتباين المشترك بين مجموعة المتغيرات وكذلك اشتقاق إحصاءه **لاختبار فرضية الارتباط القويم ،أي اختبار فرضية الاستقلالية بين مجموعتين من المتغيرات .**

Canonical correlation analysis: [1,2,3,4,5,6,7]

تعريف الارتباط القويم :-

يسمى المقياس الإحصائي الذي يستخدم لقياس قوة العلاقة بين المتغيرين x,y بمعامل الارتباط البسيط ، وكذلك يستخدم معامل الارتباط المتعدد لقياس العلاقة أو الارتباط بين مجموعة المتغيرات $(x_1,x_2,.......x_p)$

وكذلك يستخدم الارتباط الجزئي لقياس العلاقة بين مجموعة متغيرات بثبوت مجموعة أخرى .

وإذا كان $\hat{y}$ يمثل تركيب خطي من المتغيرات $(x_1,x_2,.......x_p)$ وهو

$$\hat{y} = \hat{B}_1 X_1 + \hat{B}_2 X_2 + ... + \hat{B}_p X_p$$

ويتم تحديد القيم التقديرية للمعالم

$\hat{B}_1,........,\hat{B}_p$ بطريقة تجعل مجموع مربعات الخطأ $E(y-\hat{y})^2$ أقل ما يمكن .

ويسمى الارتباط بين القيم الحقيقية Y والقيم التقديرية $\hat{y}$ بمعامل الارتباط المتعدد .

ولنفترض أن $\sum$ هي مصفوفة التباين والتباين المشترك بين المتغيرات $(x_1,x_2,.......x_p)$ والمتغير المعتمد y .

$$\sum = \begin{pmatrix} \sigma_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$$

علماً بأن

$$v(y) = \sigma_{11}$$

مصفوفة التباين

$$\sum_{22} = \text{cov matrix of } (x_1,x_2,.......x_p)$$

والتباين المشترك بين $(x_1,x_2,.......x_p)$

متجه صفري يعبر عن التباين المشترك بين Y والمتغيرات $(x_1,x_2,.......x_p)$

فأن معامل الارتباط بين y والمتغيرات $(x_1,x_2,.......x_p)$ هو

$$\rho_{y(x_1,x_2,......xp)} = \sqrt{\frac{\sum_{12}\sum_{22}^{-1}\sum_{21}}{\sigma_{11}}} \quad ....(1)$$

$$= \sqrt{1-\frac{\left|\sum\right|}{\sigma_{11}\left|\sum_{22}\right|}} \quad ....(2)$$

وسوف نعتمد على مفهوم الارتباط القويم **canonical correlation** لتعميم مفهوم معامل الارتباط المتعدد وكما يلي :

لنفرض لدينا مجموعة المتغيرات العشوائية وسوف تسمى (p-group) $(x_1,x_2,.......x_p)$ ومجموعة أخرى $(x_{p+1},x_{p+2},.......x_{p+q})$ وتسمى هذه (q-group) ، إذ $q \geq p^{-1}$ ونحن بصدد الحصول على مقياس لقوة الاعتمادية بين مجموعة المتغيرات p ، ومجموعة المتغيرات q ، ولتحقيق ذلك نفرض التراكيب الخطية من المتغيرات $x_1,x_2......x_p,x_{p+1}\ .....x_{p+q}$ وكالآتي :

$$x_1 = c_{11}x_1 + c_{12}x_2 + ... + c_{1p}x_p$$

$$x_2 = c_{21}x_1 + c_{22}x_2 + ... + c_{2p}x_p \quad ....(3)$$

.

$$
\begin{aligned}
x_p &= c_{p1}x_1 + c_{p2}x_2 + \dots\dots + c_{pp}x_p \\
x_{p+1} &= d_{p+1,1}x_{p+1} + d_{p+1,2}x_{p+2} + \dots\dots + d_{p+1,q}x_{p+q} \\
x_{p+2} &= d_{p+2,1}x_{p+1} + d_{p+2,2}x_{p+2} + \dots\dots + d_{p+2,q}x_{p+q} \\
x_{p+q} &= d_{p+q,1}x_{p+1} + d_{p+q,2}x_{p+2} + \dots\dots + d_{p+q,q}x_{p+q}
\end{aligned}
$$

وللسهولة سنفرض إن جميع المتغيرات ($x_1, x_2 \dots\dots x_{p+q}$) لها متوسط صفر وتباين ١ ، وان أي متغير x في المجموعة p هو مستقل (غير مرتبط) مع أي متغير آخر في نفس المجموعة ، وكذلك أي متغير x في المجموعة q فهو مستقل عن أي متغير في المجموعة q .

ولنرمز لمعامل الأرتباط بين المتغير x_i , x_{p+i} بالرمز ρ_i

ρ_i **=correlation (x_i, x_{p+i})** $\forall_i$ **=١،٢,....p**

أي أن:

ρ_1 **=corr(x_1, x_{p+1})**

ρ_2 **=corr(x_2, x_{p+2})** **corr(x_1, x_{p+3})=٠**

:

:

.

ρ_p **=corr(x_p, x_{2p})**

ويمكن التعبير عن ذلك بالمصفوفة التالية

$$
\begin{array}{cc}
 & \begin{array}{cc} \mathbf{x_1, x_2 \dots x_p} & \mathbf{x_{p+1} \dots x_{p+q}} \end{array} \\
\Lambda = & \left[\begin{array}{ccc|ccc}
1 & & 0 & \rho_1 & & 0 \\
 & 1 & & & & \\
\hdashline
0 & & 1 & 0 & & \rho_p \\
\rho_1 & & 0 & 1 & & 0 \\
 & \rho_2 & & & 1 & \\
0 & & \rho_p & 0 & & 1
\end{array}\right]
\end{array}
$$

و يمكن أن نبرهن أن قيمة محدد Λ هي

$$|\Lambda| = \pi_{i=1}^{\rho}\left(1 - p_i^2\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5)$$

البرهان

$$|\Lambda| = \left\|\begin{pmatrix} I & \rho & 0 \\ \rho & I & 0 \\ 0 & 0 & I \end{pmatrix}\right\|$$

$$=\left|\begin{array}{cc:c} I_{p\times p} & diag(pi) & 0_{(p+q),p} \\ diag(pi) & I_{p\times p} & 0_{(p+q),p} \\ \hdashline 0_{p,(q-p)} & 0_{p,(q-p)} & I_{(p+q),(q-p)} \end{array}\right|$$

وتسمى الارتباطات ($\rho_1, \rho_2, \ldots\ldots\ldots\ldots\ldots \rho_p$) بالارتباط القويم ، أما المتغيرات ($X_1, X_2, \ldots\ldots\ldots X_p, X_{p+1}, \ldots\ldots X_{p+q}$)تسمى المتغيرات القويمة Canonical variables

والآن دعنا نعرف التراكيب الخطية

$$X = C_1X_1 + C_2X_2 + \ldots\ldots\ldots\ldots + C_pX_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

$$Y = d_{p+1}X_{p+1} + d_{p+2}X_{p+2} + \ldots\ldots\ldots + d_{p+q}X_{p+q} \ldots\ldots\ldots\ldots\ldots (7)$$

وحيث أننا نهتم بمقياس الارتباط ، وإذا افترضنا ان متوسطات جميع المتغيرات تساوي صفر ،فأن

$$V(x) = c'\sum_{11} c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

$$V(y) = d'\sum_{22} d \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9), \quad \sum = \left(\begin{array}{c:c} \sum 11 & \sum 12 \\ \hdashline \sum 21 & \sum 22 \end{array}\right)^{p}_{q \ (p+q)}$$

$$\text{Cov}(x,y) = c'\sum_{12} d \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

مصفوفة من الرتبة (p+q)

وتعبر $\sum$ عن مصفوفة التباين والتباين المشترك بين المتغيرات ($X_1, X_2, \ldots\ldots\ldots X_p$) والمتغيرات ($X_p, X_{p+1}, \ldots\ldots X_{p+q}$)، وإذانمقياس الارتباط خالي من الوحدات ، وعليه يمكن تحويل التراكيب الخطية الى صيغة قياسية، إذ :-

$$V(x) = c'\sum_{11} c. = 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

$$V(y) = d'\sum_{22} d = 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$\text{Cov}(x,y) = \rho = c'\sum_{12} d \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

والهدف هو تعظيم $c'\sum_{12} d$ = Corr(x,y) ρ = نسبة إلى المتجه d,c وطبقاً للشروط(١٢) ، (١١) ، ويمكننا التعبير عن هذا الهدف بالدالة :

$$\textbf{Max} \ \Phi = c'\sum_{12} d - \frac{\lambda}{2}\left(c'\sum_{11} c - 1\right) - \frac{\mu}{2}\left(d'\sum_{22} d - 1\right) \ldots\ldots\ldots (14)$$

$$\frac{\partial \Phi}{\partial c} = \sum_{12} d - \lambda \sum_{11} c = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

$$\frac{\partial \Phi}{\partial d} = c'\sum_{12} - \mu \sum_{22} d = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

بضرب المعادلة (١٥) بـ c' والمعادلة (١٦) بـ d' من اليسار نحصل على

$$c'\sum_{12} d = \lambda c'\sum_{11} c$$
$$d'\sum_{21} c = \mu d'\sum_{22} d$$
$$\therefore \rho = \lambda$$
$$\rho = \mu \Rightarrow \rho = \mu = \lambda$$
$$\therefore \sum_{12} d - \rho\sum_{11} c = 0$$
$$\sum_{21} c - \rho\sum_{22} d = 0$$

والتي يمكن التعبير عنها بالمصفوفات كالاتي :

$$\begin{pmatrix} -\rho\sum_{11} & \sum_{12} \\ \sum_{21} & -\rho\sum_{22} \end{pmatrix}\left(\frac{c}{d}\right)_q^p = 0$$

او

$$\left| \begin{pmatrix} -\rho\sum_{11} & \sum_{12} \\ \sum_{21} & -\rho\sum_{22} \end{pmatrix} \right| = 0$$

$$\left| \rho^2\sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21} \right| = 0 \quad \text{.........(17)}$$

لنفرض أن جذور المعادلة (١٧) رتبت بحيث أن :-

$$\rho_1^2 \geq \rho_2^2 \geq \text{.........} \geq \rho_p^2$$

وحيث أن هدفنا تعظيم الارتباط ، وإن هذا لا يسمح بقمة سالبة للارتباط $0 \Leftrightarrow \rho$ لذلك نأخذ الجذور الموجبة فقط

$$\rho_1 \geq \rho_2 \geq \text{.........} \geq \rho_p$$

لإيجاد المتجهات d ، c فان هذا ممكن وبصورة منفصلة ، من المعادلة (١٦) وكالآتي تسمى المعادلة $|A-\lambda B| = 0$ بالمعادلة المميزة للجذور v للمصفوفة A في B ،ان مجموعة المعادلات الخطية $(A-\lambda iB)c_i = 0$ لها حل غير صفري هو ci

إذا كان $|A-\lambda iB| = 0$ **i = ١،٢,......**

ويمكن إيجاد c ،d من النظرية التالية :

<u>نظرية:</u> لتكنB,A مصفوفات متماثلة من الدرجة P وان B محددة موجبة ، وعليه توجد مصفوفة مربعة مثل c هي :

$$C_{P\times P}=\begin{pmatrix} C_{11}C_{12}\ldots\ldots\ldots\ldots\ldots C_{1P} \\ C_{21}C_{22}\ldots\ldots\ldots\ldots\ldots C_{2P} \\ \\ C_{P1}C_{P2}\ldots\ldots\ldots\ldots\ldots C_{PP} \end{pmatrix}$$

بحيث أن :-

$$C'AC=\Lambda=\begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_P \end{pmatrix}, \qquad C'BC=I$$

وبعبارة أخرى يمكن القول أن التحويل الخطي Y=CX يحول الصيغتان التربيعيتان من

$$X'BX\Rightarrow\sum_{i=1}^{P}yi^2 \qquad ,x'Ax\Rightarrow\sum_{i=1}^{P}\lambda iyi^2$$

ومن ثم من المعادلة

$$|A-\lambda B|=0$$

نجد أن

$$\left|\sum{}_{12}\sum{}_{22}\overset{-1}{\sum}{}_{21}-\rho^2\sum{}_{11}=0\right|$$

وبجعل $\lambda=\rho^2$ **فان**

$$A=\sum{}_{12}\sum{}_{22}\sum{}_{21} \qquad B=\sum{}_{11}$$

إذن توجد المصفوفة

$$c=(c_1,c_2,\ldots\ldots\ldots\ldots\ldots\ldots\ldots c_p)$$

بحيث ان :

$$c'Bc=c'\sum{}_{11}c=I, \quad c'\sum{}_{12}\overset{-1}{\sum}{}_{22}\sum{}_{21}c=\begin{pmatrix} \rho_1^2 & & 0 \\ & \rho_2^2 & \\ 0 & & \rho_p^2 \end{pmatrix}$$

بعد إيجاد المتجهات ($C_1,C_2,\ldots\ldots C_p$) نحصل على المتغيرات القانونية (p) وهي

$$X_1 = C_1' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = C_{11}X_1 + C_{12}X_2 + ... + C_{p1}X_p$$

$$X_2 = C_2' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = C_{12}X_1 + C_{22}X_2 + ... + C_{p2}X_p$$

$$X_p = C_p' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = C_{1p}X_1 + C_{2p}X_2 + ... + C_{pp}X_p$$

ومن هنا نلاحظ ان المتغيرات($x_1, x_2, ...x_p$) تملك تباين واحد وأنها غير مرتبطة مع نفسها ونلاحظ أيضا

$$corr(x_i x_j) = c' \sum_i {}_{11} c_j = \begin{cases} 1 & if \quad i = j \\ 0 & \quad i \neq j \end{cases}$$

وأن **(I , j)th elements in** $c' \sum_{11} c = I$

وللحصول على مجموعة المتغيرات القويمة ($X_{p+1}, ..., X_{p+q}$)

من المعلوم انه توجد صيغة أخرى مناظرة للمعادلة

$$\left| \rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21} \right| = 0$$

ناتجة من التماثل وهي

$$\left| \rho^2 \sum_{12} - \sum_{21} \sum_{11}^{-1} \sum_{12} \right| = 0$$

ولها نفس الجذور

$$\rho_1^2, \rho_2^2, ... \rho_p^2$$

وباستخدام نفس الرموز السابقة

$$\lambda = \rho^2 \quad , \quad \Lambda = \sum_{21} \sum_{11}^{-1} \sum_{12} \quad , B = \sum_{12}$$

إذن توجد مصفوفة مثل D بحيث أن : **d_1 d_2 ....d_p**

$$D = \begin{pmatrix} d_{11} & d_{12} & .... & d_{1q} \\ d_{21} & d_{22} & .... & d_{2q} \\ \vdots \\ \vdots \\ d_{q1} & d_{q2} & .... & d_{qq} \end{pmatrix}_{q\times q}$$

وان المتغيرات القويمة (q) الباقية هي :

$$X_{p+1} = d_1' \begin{pmatrix} x_{p+1} \\ \vdots \\ \vdots \\ x_q \end{pmatrix} = d_{11}X_{p+1} + d_{21}X_{p+2} + ... + d_{q1}X_{p+2}$$

$$X_{p+q} = d_q' \begin{pmatrix} x_{p+1} \\ \vdots \\ \vdots \\ x_{q+q} \end{pmatrix} = d_{1q}X_{p+1} + d_{2q}X_{p+2} + ... + d_{qq}X_{p+q}$$

حيث ان المتغيرات X_{p+q}....... ,X_{p+1} لها تباين واحد وهي غير مرتبطة بينها ، ولكن

Corr(X_1,X_{p+1})= ρ_1

Corr(X_2,X_{p+2})= ρ_2

Corr(X_p,X_{2p})= ρ_q

$\rho_1 \geq \rho_2 \geq .... \geq \rho_p$

وحيث $corr(x_1, x_{p+1}) \geq corr(x_2, x_{p+2}) \geq ....$

وهي الصفة مهمة جداً في البحوث التطبيقية وخاصة في حالة وجود مجموعتين من المتغيرات الكثيرة والباحث مهتم بدراسة العلاقات الداخلية بين المجموعتين ، فهو يستطيع ان يدرس تركيب خطي من المتغيرات في كل مجموعة وهذا له اكبر ارتباط ، وهذه الارتباطات تمثل أعلى ارتباطات بين متغيرات التركيب ، فمثلاً المجموعة الأولى من المتغيرات تمثل القياسات الخاصة بالخصائص الفيزيائية الطول ، العمل ، والمجموعة الثانية تمثل خصائص الصلادة ، والأوزان والاختبارات الخاصة بها .

أن الارتباطات بين التراكيب الخطية من مجموعتي المتغيرات توضح على الأغلب العلاقات الداخلية ، أي الارتباط الداخلي بين المجموعتين نفسهما .

إذا كانت لدينا مجموعتي المتغيرات

$$\underline{x}_1 = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_p \end{pmatrix} , \quad \underline{x}_2 = \begin{pmatrix} x_{p+1} \\ x_{p+2} \\ . \\ . \\ x_{p+q} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{q}^{p}$$

فان معاملات الارتباط القويم هي جذور المعادلة

$$\left|\rho^2 \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}\right| = 0 \ldots\ldots\ldots\ldots * $$

وعليه فان

$$\left|\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1}\right| = 0$$

وعند تحويل المتجهات X_1, X_2 الى متجهات جديدة بواسطة التحويل

$$y_1 = cx_1$$
$$y_2 = Dx_2$$

فان مصفوفة التباين والتباين المشترك لـy_1 هي : $\Lambda_{11} = c\Sigma_{11}c'$

وكذلك مصفوفة التباين والتباين المشترك لـ y_2 هي

$$\Lambda_{11} = D\Sigma_{22}D'$$

$$\text{cov}-matrix \;\; of \;\; y_1 \;\&\; y_2 = \Lambda_{12} = c\Sigma_{12}D'$$

عندئذ يكون الارتباط القويم بين المجموعة y_1, y_2 هو جذور المعادلة

$$\left|\rho^2 \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right| = 0 \ldots\ldots\ldots\ldots **$$

ويمكننا برهنة ان المعادلة* ، ** لهما نفس الجذور وكما يلي :

$$\left|\rho^2 c\Sigma_{11}c' - c\Sigma_{12}D^1 D^{-1}\Sigma_{22}^{-1} D^{-1} D\Sigma_{21}c'\right|$$
$$\left|c\left(\rho^2 \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)c'\right| = 0$$
$$|c|\left|\rho^2 \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right||c'| = 0$$

وهي المعادلة *

نأتي الآن الى تقدير واختبار معنوية معامل الارتباط القويم

Estimation and tests of significance of canonical correlation

لنكن لدينا المركبة (p+q) من المتجهات

$$\begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ x_2 \end{pmatrix}_q^p - N_{p+q}\left(D, \sum\right)$$

ولتقدير الارتباطات القويمة v المعرفة كجذور للمعادلة

$$\left|\rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}\right| = 0$$

لنفرض لدينا مجموعة من المتجهات ذات N من المشاهدات والمستقلة عن بعضها البعض وان S هي مصفوفة مربعة من الرتبة (p+q) (p+q) تمثل تقدير مصفوفة التباين والتباين المشترك

$$s = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}_q^p$$

وحيث أن مقدرات الإمكان الأعظم لـ $\sum_{11}, \sum_{12}, \sum_{22}$ هي :

$$\hat{\sum}_{11} = \frac{s}{N} \;,\; \hat{\sum}_{12} = \frac{s_{12}}{N} \;,\; \hat{\sum}_{22} = \frac{s_{22}}{N}$$

وعند تعويض هذه المقدرات في المعادلة أعلاه تحصل على مقدرات الإمكان الأعظم MLE'S لمعاملات الارتباط القويم وهي

$$\hat{\rho}_1^2 = r_1^2, \hat{\rho}_2^2 = r_2^2, \dots\dots\dots \hat{\rho}_p^2 = r_p^2$$

وهي جذور للمحددة

$$\left|r^2 s_{11} - s_{12} s_{22}^{-1} s_{21}\right| = 0$$

وتعرف $r_1^2, r_2^2, \dots.. r_p^2$ معاملات الارتباط القويم للعينة ، وعليه فان

$$z = \prod_{i=1}^{r}\left(1 - \rho_i^2\right) = \prod_{i=1}^{r}\left(1 - r_i^2\right)$$

بعد ان تم بحث موضوع الارتباط القويم وخصائصه.
سنأتي الى كيفية اختبار الفرضية المتعلقة بالارتباط القويم .[٨]
ولنفرض لدينا مجموعة المتغيرات (p+q)

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_q^p \quad - \quad N_{p+q}\left(D, \sum\right)$$

ولتقدير الارتباطات القويمة v الناتجة من المعادلة

$$\left| \rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21} \right| = 0$$

نفرض ان S مصفوفة مربعة ذات بعد (p+q) × (p+q) تمثل مصفوفة التباين والتباين المشترك

$$s = \begin{pmatrix} s_{11} & \vdots & s_{12} \\ \cdots & & \cdots \\ s_{21} & \vdots & s_{22} \end{pmatrix}_q^p$$

وان مقدرات الإمكان الأعظم لكل من

$$\sum_{11}, \quad \sum_{12}, \quad \sum_{22}$$

هي

$$\hat{\sum}_{11} = \frac{s}{N}, \hat{\sum}_{12} = \frac{s_{12}}{N}, \hat{\sum}_{22} = \frac{s_{22}}{N}$$

حيث N عدد المشاهدات المستقلة لكل قيمة من التوزيع الطبيعي

وعند تعويض هذه المقدرات في المعادلة (١)نجد ان مقدر الإمكان الأعظم للارتباطات القويمة هو

$$\hat{\rho}_1^2 = r_1^2, \hat{\rho}_2^2 = r_2^2, \ldots\ldots\ldots, \hat{\rho}_p^2 = r_p^2$$

وهى جذور للمعادلة

$$\left| r^2 s_{11} - s_{12} s_{22}^{-1} s_{21} \right| = 0$$

حيث أن $r_1^2, r_2^2, \ldots r_p^2$ هي الارتباطات القويمة للعينة وحيث ان

$$z = \pi_{i=1}^{r} \left(1 - \rho_i^2\right)$$

$$z = \pi_{i=1}^{p} \left(1 - r_i^2\right)$$

ولاختبار الفرضية X_1 ، X_2 هي مستقلة أي أن $\sum_{12} = 0$

$H_0 : x_1, x_2$ **are independent**

نرفض الفرضية $H_\circ$ إذا كان (ثابت=Z) ، ثابت $\log Z \leq$
وعليه ترفض $H_\circ$ إذا كان

$$-\left[(N-1)-\frac{1}{2}(p+q+1)\right]\log Z \geq \chi^2_{pq}(\alpha)$$

$$\because Z = \underset{i=1}{\overset{p}{\pi}}\left(1-\rho_i^2\right) = \frac{\left|\sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21}\right|}{\left|\sum_{11}\right|}$$

ومنها نجد أن

$$Z = \underset{i=1}{\overset{p}{\pi}}\left(1-r_i^2\right) = \frac{\left|s_{11} - s_{12}s_{22}^{-1}s_{21}\right|}{\left|s_{11}\right|}$$

وهكذا رأينا كيف تم التوصل إلى مقياس لرفض فرضية الاستقلالية باعتماد الارتباطات القويمة .

بعض مقايس الاعتمادية بين مجموعتين من المتغيرات :-

١. معامل الارتباط العام وهو

$$Q = \underset{i=1}{\overset{p}{\pi}}\rho_i^p$$

وبما أن

$$\left|\rho^2\sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21}\right| = 0 \quad \ldots\ldots 1$$

فأن

$$\left|\sum_{12}\sum_{22}^{-1}\sum_{21}\sum_{11}^{-1}\right| = 0$$

$$\therefore Q = \underset{i=1}{\overset{p}{\pi}}\rho_i^{\rho} = \frac{\left|\sum_{12}\sum_{22}^{-1}\sum_{21}\right|}{\left|\sum_{11}\right|}$$

وان Q=٠ عندما $\sum_{12} = 0$ أي أن

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \text{ غير مرتبط مع } \begin{pmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_{p+q} \end{pmatrix}$$

٢. المعامل العام المرادف للارتباط $z = \underset{i=1}{\overset{p}{\pi}}\left(1-\rho_i^2\right)$

وحيث ان

$$\left|(1-\rho^2)\sum_{11}-\left(\sum_{11}-\sum_{12}\sum_{22}^{-1}\sum_{21}\right)\right|=0$$

$$\because Z=\pi_{i=1}^{p}(1-\rho_i^2)=\frac{\left|\sum_{11}-\sum_{12}\sum_{22}^{-1}\sum_{21}\right|}{\left|\sum_{11}\right|}$$

وعندما Z=١ فهذا يعني ان المجموعة

$$x_1 \sim N_p(\mu_1,\sum_{11})$$

$$x_2 \sim N_p(\mu_2,\sum_{22})$$

والمجموعة

ومنها يكون التوزيع الشرطي لـ x١ بوجود x٢

$$x_1|x_2 \sim N_p\left[\mu_1-\sum_{12}\sum_{11}^{-1}(x_2-\mu_2),\sum_{11}-\sum_{12}\sum_{22}^{-1}\sum_{21}\right]$$

إضافة لما تقدم يمكننا أن نبين بأن الارتباط القويم بين مجموعة المتغيرات X١ والمجموعة X٢

$x_1=\begin{pmatrix}x_1\\x_2\\ \cdot\\x_p\end{pmatrix}$ ، $X_2=\begin{pmatrix}x_{p+1}\\x_{p+2}\\ \cdot\\x_{p+q}\end{pmatrix}$ **لا يتغير تحت أي تغير خطي يكون (non singular) وكالآتي**

لتكن لدينا مجموعتان من المتغيرات

$$\underline{X}_1=\begin{pmatrix}x_1\\x_2\\ \vdots\\x_p\end{pmatrix},\quad \underline{X}_2=\begin{pmatrix}x_{p+1}\\x_{p+2}\\ \vdots\\x_{p+q}\end{pmatrix}$$

$$\sum=\begin{pmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{pmatrix}_{q}^{p}$$

فان معاملات الارتباط القويم هي جذور المعادلة

$$\left|\rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}\right| = 0$$

وعند تحويل المتجهات X_1, X_2 إلى متجهات جديدة بواسطة

$y_1 = cx_1$
$y_2 = cx_2$

وان مصفوفة التباين والتباين المشترك لـ y_1 هي

$$\Lambda_{11} = c \sum_{11} c'$$

ومصفوفة التباين والتباين المشترك لـ y_2

$$y_2 = \Lambda_{22} = D \sum_{22} D'$$
$$\operatorname{cov}(y_1, y_2) = \Lambda_{12} = c \sum_{12} D'$$

فان الارتباط القويم بين المجموعتين y_1 ، y_2 هو جذر المعادلة

$$\left|\rho^2 \Lambda_{11} - \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21}\right| = 0 \ldots\ldots\ldots\ldots\ldots 2$$

ويمكننا أن نبرهن أن المعادلتين ١ ، ٢ لهما نفس الجذور

$$\left|\rho^2 C \sum_{11} C' - C \sum_{12} D^1 D^{-1} \sum_{22}^{-1} D^{-1} D \sum_{21} C'\right|$$
$$\left|C\left(\rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}\right) C'\right| = 0$$
$$|C| \left|\rho^2 \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}\right| |C'| = 0$$

وهو نفس المقدار في المعادلة (١)

الاستنتاجات:

في ظل ما تقدم نستنتج ما يلي :

1. يتم تقدير معاملات الارتباط القويم $\rho_1, \rho_2, ...\rho_p$ بواسطة معاملات ارتباط العينة .
2. مقياس فرضية الاستقلالية للفرضية .H التي تؤكد على ان $\sum_{12=0}$ أي انه لا توجد هناك علاقة بين مجموعة المتغيرات $X_1, X_2, ...X_p$ ومجموعة المتغيرات $X_{p+1}, ...X_{p+q}$ هو مقياس بسيط يعتمد على عدد مشاهدات العينة لكل متغير (وهي متساوية طبقاً لجميع المتغيرات) ويعتمد على logz ، إذ ان $z = \pi_{i=1}^{p}\left(1-\rho_i^2\right)$.
3. تمكنا من برهان ان مصفوفة التباين والتباين المشترك لمجموعة المتغيرات $\underline{X}_1, \underline{X}_2$ هي نفسها للمتغيرات Y_1, Y_2 التي تمثل تراكيب خطية من المتغيرات الأصلية أي ان مصفوفة التباين والتباين المشترك لا تتغير تحت أي تغيير خطي يكون nonsingular .
4. معاملات الارتباط القويم مهمة جداً في البحوث التطبيقية خاصة عندما يرغب الباحث في تحليل العلاقات بين مجموعة كبيرة من المتغيرات خاصة في التجارب الفيزيائية والطبية والكيميائية .
5.

التوصيات :

1. نوصي باستخدام مقياس الارتباط القويم في بحث العلاقة بين مجموعتين من المتغيرات .
2. نوصي باستخدام الإحصاءه التي يتم اشتقاقها في اختبار فرضية الاستقلالية بين مجموعتين من المتغيرات لأنها إحصاءه بسيطة وسهلة جداً .

References / المصادر :

١. Bryan, F. J. Manly. ٢٠٠٥ . Multivariate statistical methods: a primer. US ,pp ٢١٤.

٢. Mia, H. ٢٠٠٤. Theory and applications of recent robust methods . Belgium , pp ٤٠٠.

٣. Barbara , G. T. ٢٠٠١& Linda S. F.. Using multivariate statistics. University of Michigan US pp ٩٦٦.

٤. **Alan ,J.٢٠٠٨.Modern multivariate statistical techniques : regression , classification and ... manifold learning . New York US , pp ٧٣٢.**

٥. **Takakazu , S., Toru , O. , Fumitake , S. & Tomoya , Y. ٢٠٠٧. Distributions and the bootstrap methods of some statistics in principle canonical correlation analysis . J-Japan Satist. Soc. Vol.٣٧ No.٢ ٢٠٠٧ ,pp٢٣٩-٢٥١.**

٦. **Bennett Mangasraian .١٩٩٤. serial and parallel Multi-category Discrimination .technometrics Vol .٨ ١٩٩٤ ,**

٧. **Fisher , L. and J.W. Van Ness .١٩٧٣ . "Admissible Discriminate Analysis " JASA ٦٨ pp ٦٠٣-٦٠٧ .**

٨. **Richard ,L.G. and Robert , J. Power of tests for quality of covariance matrices. technometrics Vol.١٦ No. ١ feb ١٩٧٤.**

Abstract :

In this research , the interdependence between two variables x , y is measured by the coefficient of correlation . While the independence between two sets of variable is measured by canonical correlation, the method of estimating canonical correlation is explained, and all the notation and assumption necessary for it is represented. Also the statistic for testing the hypothesis of independence is derived, and we prove that the canonical correlation is not changed under any non singular transformation.

Also a procedure for obtaining the set of canonical variables is considered in this research.