مقارنة مقياسي الارتباط الانتروبي مع بعض المقاييس الشائعة لقياس الارتباط في جداول التوافق من درجة

(KxK)

م.د. باسم شليبه مسلم

كلية الادارة والاقتصاد / جامعة واسط

المستخلص :-

في هذا البحث تم مقارنة مقياسي الارتباط الانتروبي (Entropy correlation) مع بعض المقاييس الشائعة لاختبار الاستقلالية من عدمها لمتغيرات جداول التوافق ومنها (اختبار مربع كاي – اختبار يتس – اختبار نسبة الإمكان الأعظم – اختبار فشر) ومقارنتهما لقياس الارتباط مع بعض المقاييس الشائعة لقياس الارتباط ومنها (معامل فاي – معامل التوافق –مقياس يول) واغلبها تعتمد في احتسابها على قيمة اختبار مربع كاي لاختبار الاستقلالية بين المتغيرين في جدول التوافق من درجة (kxk) أو(2xk) .وقد اعتمد البحث في الجانب التجريبي والتطبيقي على مجموعة من جداول التوافق من درجة (2x2) اذ تم اختيارها من البحث [14] فضلا عن جدولين توافق لبيانات ظاهرة الطلاق لمرة واحدة ولمرتين بحسب نوع جنس المطلق (ذكور وإناث)والبيئة المكانية لحدوث الطلاق (ريف ومدينة) في محافظة واسط للمدة الزمنية (2004-2010). ومن أهم ما خلص إليه البحث امتلاك مقياسي الارتباط الانتروبي صفة التفسير المباشر والقوي للكشف عن وجود العلاقة وقوتها في آن واحد على الرغم من ضعف احدهما مقارنة بالمقاييس الأخرى بخلاف المقاييس الشائعة الأخرى التي تعتمد على قيمة مربع كاي لاختبار الاستقلالية وهي قيمة عددية لا تمتلك دلالة تفسيرية لمفهوم علاقة الارتباط وكذلك أن مقياس الارتباط الانتروبي لا يتأثر بالعدد الموجود في إي خلية من خلايا الجدول كونه يعتمد على الثقل الاحتمالي للخلية بخلاف اختبار مربع كاي الذي يتأثر بوجود عدد اقل من 5 مشاهدات (تكرارات) كون عملية اشتقاقه اعتمدت على العينات الكبيرة في جدول التوافق فضلا عن استنتاجات أخرى يمكن مشاهدتها في البحث .

Abstract:

In this research we compared measurements of correlation entropy (Entropy correlation) with some measures common to test independence of whether or not to variables contingency tables including (chi square test - Yates test - ML ratio test - Fisher test) and comparable to measure the correlation with some measures common to measure the correlation including (Phi coefficient - contingency coefficient - yale measure) and mostly rely on calculated on the value of the chi square test to test the independence between the two variables in contingency Table of degree (kxk) or 2xk)). has adopted Search side experimental and applied to a set of contingency tables of degree (2x2) was selected from research [14] as well as two tables agree to data phenomenon of divorce once and twice, according to the sex of the absolute (male and female) and spatial environment for a divorce (Rural and city) in Wasit province for the time (2004-2010). One of the main findings of the research owning A standard correlation entropy interpretation direct and strong to detect the existence of the relationship and strength at the same time despite the weak one compared standards other than metrics other common that depend on the value of chi square test of independence is a numerical value has no significant explanatory concept of relationship correlation, as well as to measure the correlation entropy is not affected by number located in any cell of the table because it depends on the gravity of probability cell unlike chi square test that is affected by a number less than 5 hits (frequency) that the process of derivation based on large samples in contingency Table as well as the conclusions Others can be seen in the search

الكلمات المفتاحية :- **الانتروبي , الانتروبي الحدي , الانتروبي المشترك جدول التوافق , معامل الارتباط البسيط . معامل الانحدار .**

1-المقدمة :-

يعرف الارتباط **(correlation)** بأنه العلاقة بين متغيرين أو أكثر بحيث أن أي تغيير في قيمة احدهما يؤثر في تغيير قيمة المتغير الأخر . لذا يعد موضوع الارتباط من المواضيع المهمة لدى الباحثين لدراسة

الظواهر وسلوكها تبعا لتغير متغيراتها التي تؤثر عليها سلبا أو إيجابا , إذ يقوم الباحثون في مختلف الاختصاصات العلمية باستخراج قيمة معامل الارتباط التي تدل على قوة أو ضعف العلاقة بين متغيرات الظاهرة تحت الدراسة على وفق صيغ تختلف باختلاف نوع بيانات المتغيرات فهناك المتغيرات الكمية (**Quantitative variable**) التي يستخدم معامل الارتباط البسيط **simple correlation coefficient)** (معامل بيرسون) لقياس قوة ارتباطاتها والمتغيرات الوصفية (النوعية) **descriptive) variables)**التي تنقسم بدورها إلى نوعين الأول متغيرات وصفية قابلة للترتيب (يمكن ترتيبها تصاعديا أو تنازليا)مثل تقديرات الطالب الجامعي (جيد , مقبول,.الخ) يستخدم معامل ارتباط الرتب (معامل سبيرمان) **(spearman correlation coefficient)** لقياس ارتباطاتها , أما الثاني متغيرات وصفية غير قابلة للترتيب (متغيرات اسمية) ومثالها متغير نوع الجنس (ذكر,أنثى) أو الحالة العلمية (متعلم,غير متعلم) يستخدم معامل كرايمر **(Cramer's V coefficient)** أو معامل التوافق **(contingency coefficient)** الخ .

وقد تناولت الكثير من البحوث التطبيقية النوع الثاني من المتغيرات الوصفية متضمنة عملية تحليل بياناتها بعد وضعها في جدول يسمى بجدول التوافق **(contingency table)** لمتغيرين او (باتجاهين) **two way ()** (وهو موضع اهتمام البحث) أو بثلاث اتجاهات (ثلاث متغيرات) **(three way)** كل متغير(اتجاه)لديه مستويين أو أكثر إذ استخدم اختبارالاستقلالية **(independence test)** الذي يعد احد تطبيقات اختبار مربع كأي لبيان وجود العلاقة من عدمها بين مستويات المتغيرات ومن ثم يتم استخراج معامل الاقتران لقياس قوة الارتباط عندما يكون الجدول من الدرجة (2x2) ويستخدم معامل التوافق لقياس قوة الارتباط عندما يكون احد المتغيرات أوكلها لديه أكثر من مستويين ,ومن تلك البحوث على سبيل المثال مقرونة بأسماء الباحثين **Alan Agresti [1]** والباحثان **Fritse .Z, Jos M,F, Ten Berge [2]** والباحث **Jacob Cohen [4]** و **Kouji Yamamoto, Yohei Ban and Sadao Tomizawa [8]** والباحث **Narges Abbasi [9]** والباحث **Olusola Adeyemi [10]** والباحثان **Serpil Aktas and Meral Cetin [11]** **والباحث** **Yates [14]** .

إن الصيغ التي استخدمت لقياس قوة الارتباط (الاقتران) في تلك البحوث اعتمدت على قيمة اختبار مربع كأي المستخدم لاختبار الاستقلالية وهي تفتقر الى صفة التفسير القوي والمباشر لارتباط تلك المتغيرات الوصفية كونها تعتمد على قيم عددية لا تملك هذه الصفة إذ أن اختبار مربع كاي يعتمد باحتسابه على التكرارات المشاهدة والمتوقعة لخلايا جدول التوافق وقد اشترط في هذا الاختبار عدة شروط أهمها إن لا يقل حجم العينة عن 50 مفردة (مشاهدة) وكذلك لا يقل عدد المفردات في أية خلية من خلايا الجدول عن 5 مفردات وعند عدم تحقق هذا الشرط يتم دمج الصف (العمود)

الذي تقع فيه تلك الخلية إلى الصف (العمود) اللاحق أو السابق حتى يتحقق الشرط وهذا الدمج يؤثر بشكل مباشر على درجات الحرية للاختبار ويفقد مستويات قد تكون مهمة في الظاهرة المدروسة . ومن هنا جاء هدف البحث بإجراء مقارنة مقياسي الارتباط الانتروبي مع اختبار مربع كاي χ^2 للاستقلالية للكشف عن وجود العلاقة بين متغيرين في جدول التوافق من درجة (2xk) أو (kxk) لمتغيرين واختبارات أخرى شائعة وكذلك المقارنة مع بعض الاختبارات الشائعة التي تعتمد على قيمة مربع كاي χ^2 لقياس قوة الارتباط

(الاقتران) واختبارات أخرى أيضا للوقوف على أفضلها بحسب أحجام العينات الصغيرة والكبيرة و الكبيرة جدا وبحسب أسبقية المقياس في تحديد قوة العلاقة بين المتغيرين في جدول التوافق بعبارة أخرى تحديد المقياس الأفضل على التمييز الواضح والقوي لطبيعة الارتباط بين المتغيرين في جدول التوافق المربع (ذو المستويات المتساوية العدد) .

وعليه فقد تم تقسيم البحث على أربعة مباحث الأول منها تضمن المقدمة وهدف البحث والثاني الجانب النظري للبحث موضحا المفهوم الإحصائي للانتروبي وكذلك الانتروبي للمتغير المتقطع (discrete variable) واليات الحصول على مقياس الارتباط الانتروبي لجدول التوافق من درجة (2xk) والثالث الجانب التجريبي والتطبيقي إذ سيتم التحليل الإحصائي والمقارنة لجداول توافق تجريبية تم اختيارها من البحث [14] وجدولين توافق لبيانات حقيقية تتعلق بعدد مرات الطلاق في محافظة واسط للمدة الزمنية (2004-2010) مصنفة بحسب جنس المطلق والبيئة المكانية لحدوث حالة الطلاق تم الحصول عليها من محكمة أحوال الكوت والجداول جميعها من درجة 2x2 . أما المبحث الأخير فتضمن الاستنتاجات والتوصيات .

2- الجانب النظري :-

1-2 الانتروبي (entropy) :-

يعد الباحث Clausius أول من استخدم مصطلح الانتروبي لأول مرة في عام 1864عندما وصف دالة القياس المباشر لبعض الإحداثيات الكلية (macro coordinates) و لم يعتمد في دالته هذه على الأسس الاحتمالية إذ سمي هذا النوع (الانتروبي التقليدي) (classical entropy) [15] وقد استفاد منه بعض الباحثون في تخصص الفيزياء ضمن موضوع إحصاء الديناميكيا الحرارية إذ عرف الانتروبي (بأنه دالة بمواقع وسرعة الجزيئات التي يحتويها النظام الفيزيائي(physical system)) وبعد مدة من الزمن وتحديدا في عام 1948 قدم الباحث (Shannon) ملخص عن دالة الانتروبي بشكل مقياس لكمية المعلومات أو عدم اليقين (الغموض)(uncertainty) للتجارب الاحتمالية إي التجارب التي تخضع لشروط النظرية الاحتمالية إذ وضع نموذج رياضي لنظام الاتصالات (إعداد الرسائل وقنوات الاتصال) عليه الطابع الاحتمالي منطلقا من حقيقة مفادها إن كمية المعلومات (information) تكون مساوية إلى إزالة عدم اليقين (الغموض)وقد سمي هذا النوع (Shannon's entropy) ولذا تعتبر نظرية المعلومات (information theory) فرعا من فروع النظرية الاحتمالية, في هذا البحث سنهتم بالنوع الثاني لان النوعين مختلفين تماما إذ سيتم عرض الانتروبي لمتغير واحد (حدي) (marginal entropy) ومتغيرين (مشترك) (joint entropy)

والانتروبي الشرطي (conditional entropy) وكذلك المعلومات المتبادلة (mutual Information) وكالاتي :-

لقد عرف الباحث (Shannon) مستوى كمية المعلومات الواصلة لمشاهدة واحدة فقط من المتغير العشوائي x (Self Information) كالأتي[15],[8] :

$$I(x_i) = \log_2\left(\frac{1}{p_i}\right) = -\log_2(p_i) \qquad \ldots\ldots\ldots\ldots(1)$$

إذ إن (p_i) يشير الى الاحتمال للمشاهدة x_i وان (i=1,2,…n).

وقد ذكر الباحث إن الدالة الرياضية[1] أعلاه هي الوحيدة التي تحقق الشروط الآتية [12] :-

1- إن قيمة المعلومات لمتغيرين هي أعلى من قيمة المعلومات لمتغير واحد .
2- قيمة المعلومات لأي متغير هي غير سالبة.
3- قيمة المعلومات لمتغيرين مستقلين تساوي حاصل جمع معلوماتهما .

ويلاحظ إن كمية المعلومات المعرفة بالصيغة رقم (1) هي دالة بالاحتمال لحدوث المشاهدة x_i اذ إن كمية المعلومات التي تحتويها x_i هي نسبة من عدم اليقين (الغموض)(uncertainty) وهذا يعمم إلى كل المشاهدات الأخرى للمتغير العشوائي x .ولتوضيح (uncertainty) نفرض على سبيل المثال بأن لدينا فضائيين لتجربتين عشوائيتين كل فضاء يحتوي على مشاهدتين الفضاء الأول (a_1 , a_2) والاحتمالات المقابلة لها هي على التوالي (0.5,0.5) والفضاء الثاني هو(b_1,b_2) والاحتمالات المقابلة لها (0.95 ,0.04), يلاحظ إن عدم اليقين (الغموض)(uncertainty) موجود في التجربة الأولى بالنسبة للمشاهدة الأولى لكلا التجريبيتين أكثر من التجربة الثانية التي هي أكثر يقينا واقل غموضا لذا لا نستطيع التنبؤ بظهور الحالة (المشاهدة الأولى) لوجود الغموض فيها [12].

وبفرض توفر دالة الكتلة الاحتمالية للمتغير المتقطع X (probability mass function) الأتي :-

$$(X_i) \sim P(X_i) \qquad , \qquad (\, i = 1,2, \dots n \,)$$

$$P(X_i) = P[X_i = x_i \;] = p_i \qquad \dots\dots\dots\dots\dots\dots(.2)$$

إذ إن :-

$$p_i \geq 0 \qquad i = 1 \dots n \qquad\qquad \sum_{i=1}^{n} p_i = 1$$

يمكن كتابة دالة (Shannon's entropy) كالآتي [13] [6] [3] :-

$$H[X] = H_n(\mathrm{p}_1, p_2 \,, \dots\dots, p_n) = E[I(x_i)]$$

---

[1] اذ ان : nat فهي بدلالة $\log_2$ بدل دالة ln وفي حالة استخدام دالة bit الدالة تستخرج كمية المعلومات ب **1nat =ln2 bit.**

$$H[X] = E[-\log_2(p_i)] = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad \ldots\ldots\ldots.(3)$$

والصيغة (3) تمثل متوسط المعلومات (mean information) الذي يضم معلومات كل المشاهدات للمتغير العشوائي X وتكون قيمتها محصورة بين $(\log n, 0)$ إذ تكون قيمة الانتروبي أعلى ما يمكن في حالة التوزيع المنتظم (uniform dist.), في حين قيمة المقدار $(p_i \log_2 p_i)$ تكون مساوية الى الصفر عنما تكون قيمة p_i تساوي إلى الصفر, أي إن :-

$$\lim_{p_i \to 0} p_i \, log_2 \, p_i = 0$$

وبصورة مماثلة يمكن كتابة دالة الانتروبي لمتغير عشوائي أخر وليكن Y مثلا بقدر m من المشاهدات كالأتي :-

$$H[Y] = H_m(\mathrm{p}_1, p_2, \ldots\ldots, p_m) = E\left[I(y_j)\right] = -\sum_{j=1}^{m} p_j \, log_2 \, p_j \qquad \ldots\ldots\ldots\ldots\ldots.(4)$$

وتسمى الصيغ (3) و(4) الانتروبي الحدي (marginal entropy) للمتغيرين $(\boldsymbol{X}, \boldsymbol{Y})$ على التوالي .

إما في حالة توفر دالة الكتلة الاحتمالية المشتركة لمشاهدات المتغيرين X, Y (joint probability mass function) إي إن **:-**

$$(X_i, Y_j) \sim P(X_i, Y_j) \quad , \qquad (i = 1,2, \ldots n), (j = 1,2, \ldots m)$$

$$P(X_i, Y_j) = P[X_i = x_i, Y_j = y_j] = p_{ij} \qquad \ldots\ldots\ldots\ldots\ldots(5)$$

إذ إن :-

$$p_{ij} \geq 0 \qquad , \sum_{i=i}^{n} \sum_{j=1}^{m} p_{ij} = 1$$

فانه يمكن تعريف كمية المعلومات المشتركة للمشاهدتين X_i, Y_j كالأتي :-

$$I(X_i, Y_j) = \log_2 \frac{1}{P_{ij}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots.(5)$$

أما الانتروبي المشترك (joint entropy) للمتغيرين (X, Y) فيكون كالآتي [12] [3] :-

$$H(X, Y) = E\left[I(X_i, Y_j)\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

$$H(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 \frac{1}{p_{ij}} = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 p_{ij} \qquad \ldots\ldots\ldots\ldots\ldots\ldots.(7)$$

تمثل الصيغة (7) متوسط المعلومات المشتركة والذي يضم المشاهدات جميعها للمتغيرين (X, Y) .

ومن الشروط المهمة التي يجب تحققها للدالة (7) هو الأتي [6] :-

$$Max\big(H(X),H(Y)\big) \le H(X,Y) \le H(X)+H(Y) \quad \ldots\ldots(8)$$

وعندما يكون المتغيران (X,Y) مستقلين (independent variables) فان الطرف الأيمن والوسط من المتراجعة (8) يتساويان أي إن :-

$$H(X,Y) = H(X)+H(Y) \quad \ldots\ldots\ldots\ldots(9)$$

ومن الصيغة (9) يمكن الاستنتاج انه يمكن اعتماد أسس الاحتمالات في موضوع الانتروبي مع الأخذ بنظر الاعتبار أن كل عملية ضرب هي عملية جمع في الانتروبي وكل عملية قسمة مقدار (دالة)على أخر هو عملية طرح تلك المقادير, إذ أن الاستقلالية بين متغيرين تكون عندما يمكن إيجاد الاحتمال المشترك من حاصل ضرب الاحتمال الحدي للمتغيرين (X,Y) .

وعليه يمكن تعريف الانتروبي الشرطي (conditional entropy)الذي يمثل كمية المعلومات لأحد المتغيرات بعد ثبوت المتغير الآخر (تم حدوثه) كالأتي [3] [6] :-

$$H(Y|X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p\big(x_i\,,y_j\big)\ \log_2\ (y|x) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

وكذلك

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p\big(x_i\,,y_j\big)\ \log_2\ (x|y) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(11)$$

إذ إن :-

1- $H(Y|X) \ge 0$, $H(X|Y) \ge 0$
2- $0 \le H(Y|X) \le H(Y)$
3- $H(X)+H(Y|X) = H(Y)+H(X|Y)$
4- $H(Y|X) \ne H(X|Y)$

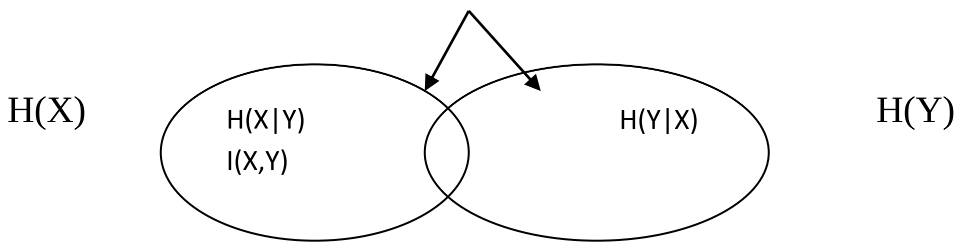
وكما هو الحال في موضوع الاحتمالات بوجود الحوادث المتبادلة (mutual events) توجد المعلومات المتبادلة بين المتغيرين (X,Y) التي يمكن تعريفها كالأتي :-

$$I(X,Y) = H(X)+H(X|Y) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

$$I(Y,X) = H(Y)+H(Y|X) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13)$$

ويمكن توضيح العلاقة بين الانتروبي والمعلومات المتبادلة بالشكل الأتي :-

H(X,Y)



الشكل رقم (1) يبين العلاقة بين الانتروبي والمعلومات المتبادلة.

كما وان :-

$I(X,Y) = I(Y,X)$

وتتميز المعلومات المتبادلة بالاتي[13] :-

$I(X,Y) \geq 0$

إن كمية المعلومات المتبادلة تكون مساوية إلى الصفر إذا وفقط إذا كان $H(Y|X) = H(Y)$ وهذا يعني ان المتغيران مستقلان [13].

ويمكن كتابة كمية المعلومات بصورة أخرى كالآتي :-

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(14)$$

إن الصيغة (14) غالبا ما تكون هي المهمة لدى اغلب الباحثين عند بحثهم عن صيغة قياس ارتباط باستخدام الانتروبي لجدول التوافق (2XK) , وهذا ما سيتم عرضه في المبحث القادم .

2-2 الارتباط الانتروبي(Entropy correlation) :-

قبل البدا بعملية توضيح فكرة الوصول إلى صيغة الارتباط الانتروبي يمكن توضيح شكل جدول التوافق باتجاهين (kxk) كالأتي :-

$X \Rightarrow$ / $Y \Downarrow$	X_1	X_2	X_3	....	X_K	المجموع
Y_1	O_{11}	O_{12}	O_{13}	....	O_{1K}	$O_{1.}$
Y_2	O_{21}	O_{22}	O_{23}	....	O_{2K}	$O_{2.}$
Y_3	O_{31}	O_{32}	O_{33}	....	O_{3K}	$O_{3.}$
. .	. .	. .	. .	....	. .	. .
Y_K	O_{k1}	O_{k2}	O_{k3}	....	O_{KK}	$O_{K.}$
المجموع	$O_{.1}$	$O_{.2}$	$O_{.3}$	....	$O_{.K}$	$O_{..}$

الشكل رقم (2) يبين جدول التوافق باتجاهين (kXk)

إذ أن :- الجدول أعلاه هو جدول مربع (Kxk).

وللوصول إلى صيغة لقياس الارتباط باستخدام الانتروبي لجدول التوافق المعرف بالشكل رقم (2) نبدأ من استخرج كمية المعلومات المتبادلة وهي النقطة المهمة لعملية الوصول إلى صيغة معامل الارتباط الانتروبي , وقد اعتاد الباحثون كتابتها كما في الصيغة (14) ومن ثم بدأ مقارنتها ومماثلتها مع صيغة معامل الارتباط البسيط (correlation coefficient) كما هو الحال مع الباحثتين (Jaakko Astola , Ilkka Virtanen) [5] اذ عرضتا المماثلة كالأتي :-

لدينا صيغة معامل الارتباط البسيط الأتي:-

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X\, \sigma_Y} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(15)$$

فعند الرجوع إلى الصيغة (14) ومقارنتها مع الصيغة (15) يمكن الاستنتاج أنها تماثلها بعد اجتراء التقديم و التأخير للمقادير إذ إن :-

$$I(X,Y) = -\big(\, H(X,Y) - H(X) - H(Y)\big) = H(X) + H(Y) - H(X,Y) \ldots (16)$$

ومنها يمكن القول أن $H(X,Y)$ يمثل التباين المشترك $cov(X,Y)$ وان$H(X)$ و $H(Y)$ يمثلان الانحراف المعياري للمتغيرين Y , X على التوالي مع الأخذ بنظر الاعتبار أن عملية الجمع والطرح استخدمت بسبب وجود log_2 في الانتروبي .

ولضمان بقاء قيمة معامل الارتباط الانتروبي محصورة بين الصفر والواحد يتم تقسيم الصيغة 16 على المقدار المعرف بالصيغة (26) كما سيتم عرضه حقا .

أما الطريق الأخر فقد عرضه الباحث (Suzuki) , إذ استخدم الصيغة (12) أو (13) للحصول على كمية المعلومات المتبادلة لجدول التوافق , ومن تطبيق الانتروبي على توزيع متعدد المتغيرات الطبيعي (Multivariate normal dist.) لمتغيرين استطاع إيجاد العلاقة الآتية [17]:-

$$I(X,Y) = -log \sqrt{1-\rho_{X,Y}^2} \quad \ldots\ldots\ldots(17)$$

ومنها إيجاد قيمة معامل الارتباط الانتروبي إي إن :

$$\rho_{X,Y} = \sqrt{1-e^{-2(I(X,Y))}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(18)$$

وقد ذكر الباحث انه يمكن كتابة كمية المعلومات المتبادلة بالصيغة الآتية :-

$$\boldsymbol{I(X,Y) = \frac{1}{O_{..}}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} O_{ij}\log_2 O_{ij} + O_{..}\log_2 O_{..} - \sum_{i=1}^{k} O_{i.}\log_2 O_{i.} - \sum_{j=1}^{k} O_{.j}\log_2 O_{.j}\right]} \ldots (19)$$

وقد تم التحقق من الصيغة *(19)*باشتقاقها من قبل الباحث (انظر الملحق).

واعتمادا على فكرة المقارنة والمماثلة في أسلوب الانتروبي الذي اعتمده الباحثان (Jaakko Astola , Ilkka Virtanen نستطيع عرض طريق أخر ومحاولة جديدة لإيجاد كمية المعلومات لجدول التوافق

بإيجاد الوسط الهندسي لمعاملي الانحدار$Y|X$ و $X|Y$ لكمية المعلومات مع الأخذ بنظر الاعتبار مسالة اللورغارتيم عند إيجاد الوسط الهندسي (Geometric mean) لمعاملي الانحدار المعلومات لإيجاد كمية المعلومات المتبادلة وكالاتي :-

إن صيغتي معاملي الانحدار في أنموذج الانحدار الخطي البسيط تعرف كالأتي:-

$$\boldsymbol{b_{Y|X}} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_X^2}} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_X \sigma_X}} \qquad \text{.........................}(20)$$

$$\boldsymbol{b_{X|Y}} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_Y^2}} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_Y \sigma_Y}} \qquad \text{.............................}(21)$$

وكذلك صيغة الوسط الهندسي لمعاملي الانحدار تكون كالأتي :-

$$G = \sqrt{b_{Y|X}\ .\ b_{X|Y}} = \rho(X,Y) \qquad \text{........................}(22)$$

وباستخدام أسلوب المقارنة والمماثلة يمكن كتابة الصيغ (17) و(18) باستخدام الانتروبي كالأتي :-

$$be_{Y|X} = H(X,Y) - 2H(X) \qquad \text{....................................}(23)$$

$$\mathrm{be}_{X|Y} = \mathrm{H}(X,Y) - 2H(Y) \qquad \text{...}(24)$$

إذ أن :-

$be_{Y|X}$: معامل انحدار كمية المعلومات للمتغير Y على كمية المعلومات المتغير X .

$\mathrm{be}_{X|Y}$:- معامل انحدار كمية المعلومات للمتغير X على كمية المعلومات المتغير Y.

ولأجل كتابة الوسط الهندسي المعرف بالصيغة (22) باستخدام الانتروبي لابد أن ننظر إليه على انه حاصل جمع معاملي الانحدار وليس ضرب بسبب استخدام $\log_2$ إي أن :-

$$G_e = \frac{1}{2}\left[be_{Y|X} + \mathrm{be}_{X|Y}\right] \qquad \text{.....................................}(25)$$

G_e : يمثل الوسط الهندسي الانتروبي لمعاملي انحدار كمية المعلومات للمتغيرين (X,Y) .

ويمكن البرهنة أن النتيجة النهائية للصيغة (22) بعد التعويض هي الحصول على الصيغة (16) مسبوقة بإشارة السالب إي أن :-

$$\begin{aligned} be_{Y|X} + \mathrm{be}_{X|Y} &= H(X,Y) - 2H(X) + \mathrm{H}(X,Y) - 2H(Y) \\ &= 2H(X,Y) - 2[H(X) + H(Y)] \\ &= 2H(X,Y) - 2[I(X,Y) + H(X,Y)] \end{aligned}$$

وبإجراء بعض العمليات الرياضية نحصل على الآتي :-

$$\left[be_{Y|X} + \mathrm{be}_{X|Y}\right] = -2\,I(X,Y) \qquad \dots\dots\dots\dots (26)$$

وبتعويض المعادلة (26) بالمعادلة (22) نحصل على الاتي :-

$$I(X,Y) = -G_e \qquad \text{................}(27)$$

إن الصيغة (27) تمثل مقدار المعلومات المتبادلة بدلالة قيمة الوسط الهندسي الانتروبي لمعاملي انحدار كمية المعلومات جدول التوافق من درجة (2xk) ولكون هذه الصيغة تكون قيمتها محصورة بين (0,∞) نقوم

بقسمة الصيغة على نصف مقدار الطرف الأيمن من الصيغة (8) لضمان بقاء قيمتها محصورة بين(0,1) بسبب أن [5] :-

$$I(X,Y) = -G_e \leq \frac{1}{2}\big(H(X)+H(Y)\big) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(28)$$

إذ أن :-

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) = H(Y,X) \ldots\ldots\ldots\ldots(29)$$

$$2\,H(X,Y) = H(X) + H(Y|X) + H(Y) + H(X|Y)$$

$$H(X,Y) = \frac{1}{2}\big(H(X) + H(Y|X) + H(Y) + H(X|Y)\big) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(30)$$

وبتعويض المعادلة (30) بالمعادلة (16) نحصل على الآتي :

$$I(X,Y) \leq \frac{1}{2}\big(H(X)+H(Y)\big) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(31)$$

أن حالة المساواة في الصيغة (31) تتحقق عندما يكون المتغيران غير مترابطين .

وعليه يمكن كتابة صيغة الارتباط لجداول التوافق من درجة (2xk) باستخدام الانتروبي كالأتي :-

$$\rho_{JG_e}(X,Y) = \frac{-G_e}{\frac{1}{2}(H(X)+H(Y))} = \frac{-2G_e}{(H(X)+H(Y))} = \frac{-[be_{Y|X}+be_{X|Y}]}{(H(X)+H(Y))} \ldots\ldots\ldots\ldots\ldots(32)$$

والصيغة (23) مكافئة للصيغة التي وردت في المصدر [5] .

كما ويمكن كتابة الارتباط الانتروبي كالأتي :-

$$\rho_{sG_e} = \sqrt{1-e^{2G_e}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(33)$$

والصيغة (33) مكافئة للصيغة التي وردت في المصدر [17] .

3-2 المقاييس الشائعة لاختبار الاستقلالية وقياس الارتباط :-

من المقاييس الشائعة لاختبار الاستقلالية بين المتغيرات في جدول التوافق التي تختبر الفرضية الآتية :-

H_0: $P_{ij}=p_i \times pj$

H_1: $P_{ij}\neq p_i \times pj$

والتي تم المقارنة بها في البحث هي :-

1- اختبار مربع كاي (chi square) وصيغته كالأتي :-

$$x^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij}-E_{ij})^2}{E_{ij}^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(34)$$

إذ أن :-

E_{ij} :- التكرار المتوقع للخلية بالصف i والعمود j .

2-اختبار يتس المصحح (Yates corrected test) وصيغته كالأتي :-

$$Y.C = \frac{o_{..}\;[|o_{11}\,o_{22} - o_{12}\,o_{21}| - 0.5\,o_{..}]^2}{o_{1.}o_{2.}o_{.1}o_{.2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(35)$$

وهذا الاختبار يستخدم عندما لجدول التوافق من درجة 2x2 .

3-اختبار نسبة الإمكان (Likelihood ratio test) وصيغته كالأتي :-

$$L.R = 2\sum_{i=1}^{m}\sum_{j=1}^{n} o_{ij} \ \log\frac{o_{ij}}{E_{ij}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(36)$$

4- اختبار فشر التام (Fisher exact test) وصيغته كالأتي :-

$$\mathrm{p.F.a} = \frac{\mathrm{o}_{1.}!\, o_{2.}\, !\, o_{.1}!\, o_{.2}!}{\mathrm{o}_{11}\, !\mathrm{o}_{12}!\mathrm{o}_{21}!\mathrm{o}_{22}!\mathrm{o}_{..}!} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(37)$$

أما المقاييس الشائعة لقياس الارتباط التي تم المقارنة بها فهي :-

1- مقياس فاي (Phi) وصيغته كالأتي :-

$$\emptyset = \sqrt{\frac{x^2}{o_{..}}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(38)$$

2- مقياس معامل التوافق (contingency coefficient) وصيغته كالأتي *:*-

$$C = \sqrt{\frac{x^2}{x^2+o_{..}}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(39)$$

3- مقياس يول Yule وصيغته كالأتي :-

$$Y = \frac{o_{11}o_{22}-o_{12}o_{21}}{o_{11}o_{22}+o_{12}o_{21}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(40)$$

4- معامل كرايمر وصيغته كالأتي :-

$$\mathrm{V} = \sqrt{\frac{x^2}{o_{..}(q-1)}} \qquad \ldots\ldots\ldots\ \ \ldots\ldots\ldots\ldots\ldots\ldots(41)$$

إذ أن q تمثل اقل عدد مستويات لكلا المتغيرين وفي حالة جدول التوافق من درجة *2x2* فان معامل كرايمر يتحول إلى معامل فاي .

3- الجانب التجريبي والتطبيقي :-

في هذا البحث سيتم تطبيق الجانب النظري على بيانات بعض التجارب التجريبية والتطبيقية التي تم ترتيبها بجدول توافق من درجة 2x2 وسيتم المقارنة أولا بين مقياسي الانتروبي للارتباط لوجود العلاقة بين المتغيرين في جدول التوافق وكذلك المقارنة بين قيمتي مقياسي الانتروبي للارتباط لتحديد أي من المقاييس

يتأثر بقيمة كمية المعلومات المتبادلة للمتغيرين هذا من ناحية وثانيا المقارنة بين المقياسين وبعض المقاييس الشائعة الأخرى للبحث عن وجود العلاقة في جدول التوافق وكذلك المقارنة في قيم الارتباط للمتغيرين في جدول التوافق لذا سيتم تحليل ثمان تجارب تم اعتماد الست الأولى من تجارب البحث [14] فضلا عن تجربتين تخص بيانات عدد مرات الطلاق (مرة واحدة –مرتان) بحسب جنس المطلق (ذكور –إناث) والبيئة المكانية للطلاق (مدينة – ريف) في محافظة واسط إذ تم الحصول عليها من محكمة أحوال الكوت في محافظة واسط وقد تم ترميزها بالتجربة 8 والتجربة 7 على التوالي وتجدر الإشارة أن التجارب اختيرت بعناية إذ كانت التجارب (1,4,7,8)تمثل العينات الكبيرة والتجارب (2,3) تمثل عينات كبيرة بوجود تكرار خلية واحد ا اقل من 5 مشاهدات أو تساويها بينما التجربة (6) تمثل جداول توافق بوجود أكثر من خلية

فيها اقل من 5 مشاهدات وان التجربة 8 تم ترميزها (طلاق مرة واحدة :1 ,طلاق مرتان :2) وجنس المطلق (ذكور :1 , إناث :2) و الشي نفسه مع التجربة 7 إذ أن البيئة المكانية للطلاق تم ترميزها (مدينة :1, ريف :2) وهي تندرج ضمن البيانات التطبيقية فضلا عن التجارب الأخرى التي تمثل

الجانب التجريبي للبحث والجدول رقم (1) يوضح البيانات وفق التجارب من (1-8) علما وكالاتي :

الجدول رقم (1) يبين بيانات لجدول توافق من درجة 2x2

التجربة	O_{11}	O_{12}	O_{21}	O_{22}
1	8123	83	424	1370
2	8330	2	622	1046
3	3954	3080	5	2961
4	2886	1363	1320	4431
5	2	3	1	4
6	4	30	2	40
7	637	107	157	44
8	509	235	135	66

وباستخدام البرنامج الجاهز spss وكذلك الإكسل لإجراء بعض الحسابات الرياضية تم الحصول على النتائج الآتية :-

جدول رقم (2) يبين قيم المعلومات المتبادلة ومقياسي الانتروبي للارتباط.

التجربة	$I(X,Y)=-G_e$	P_{SGe}	P_{JGe}
1	**0.389501**	**0.73562**	**0.610162**
2	**0.322396**	**0.689368**	**0.568419**
3	**0.267629**	**0.643802**	**0.290019**
4	**0.150142**	**0.509306**	**0.152786**
5	**0.034852**	**0.259479**	**0.037051**
6	**0.012048**	**0.154301**	**0.01733**
7	**0.004768**	**0.097422**	**0.006908**
8	**0.000087**	**0.013159**	**0.000105**

من الجدول رقم (2) يلاحظ الأتي :-

1- إن قيمة الانتروبي للارتباط باعتماد أسلوب الباحث(Suzuki) أعلى من القيمة نفسها باعتماد أسلوب الباحث (Jaakko Astola , Ilkka Virtanen) بشكل عام .
2- تأثر قيمة الانتروبي للارتباط باعتماد أسلوب الباحث(Suzuki) بشكل طبيعي عند تناقص قيمة المعلومات المتبادلة بعكس القيمة نفسها باعتماد أسلوب الباحث (Jaakko Astola , Ilkka Virtanen التي تتأثر بشكل كبير وتسوء القيمة كلما تناقصت قيمة المعلومات المتبادلة .

أما الجدول رقم (3) فيوضح قيم الاختبارات جميعها لاختبار الاستقلالية في جداول التوافق المؤشرة بالجدول رقم (1) وكالاتي :-

الجدول رقم (3) قيم اختبار مربع كاي ومقياس الارتباط الانتروبي بحسب الوسط الهندسي

التجربة	قيمة الاختبار								
	X^2	P	Y.C	P	R.L	P	p.F.a	P_{JGe}	P_{SGe}
1	**6.731E3**	**.000**	**6.725E3**	**.000**	**5.400E3**	**.000**	**.000**	**0.610162**	**0.73562**
2	**5.821E3**	**.000**	**5.821E3**	**.000**	**4.469E3**	**.000**	**.000**	**0.568419**	**0.689368**
3	**2.740E3**	**.000**	**2.738E3**	**.000**	**3.710E3**	**.000**	**.000**	**0.290019**	**0.643802**
4	**2.028E3**	**.000**	**2.026E3**	**.000**	**2.081E3**	**.000**	**.000**	**0.152786**	**0.509306**

5	**.467**	**.490**	**.000**	**1.000**	**.483**	**.487**	**1.000**	**0.037051**	**0.259479**
6	**1.267**	**.260**	**.487**	**.485**	**1.269**	**.260**	**.399**	**0.01733**	**0.154301**
7	**6.646**	**.010**	**6.098**	**.014**	**6.247**	**.012**	**.012**	**0.006908**	**0.097422**
8	**.114**	**.736**	**.064**	**.801**	**.113**	**.736**	**.734**	**0.000105**	**0.013159**

واعتمادا على الجدول رقم (3) يمكن تثبيت العلاقة للمتغيرين في جداول التوافق كالأتي :-

جدول رقم (4) يبين وجود العلاقة من عدمها بحسب مقياسي الانتروبي للارتباط والمقاييس الشائعة الأخرى.

التجربة	وجود العلاقة بين المتغيرين					
	X2	**Y.C**	**R.L**	**p.F.a**	**PJGe**	**P_{SGe}**
1	موجودة	موجودة	موجودة	موجودة	موجودة \متوسطة	موجودة \قوية
2	موجودة	موجودة	موجودة	موجودة	موجودة \متوسطة	موجودة \متوسطة
3	موجودة	موجودة	موجودة	موجودة	موجودة \ضعيفة	موجودة \متوسطة
4	موجودة	موجودة	موجودة	موجودة	موجودة \ضعيفة جدا	موجودة \متوسطة
5	غير موجودة	غير موجودة	غير موجودة	غير موجودة	تتلاشى	موجودة \ضعيفة جدا
6	غير موجودة	غير موجودة	غير موجودة	غير موجودة	تتلاشى	موجودة \ضعيفة جدا
7	موجودة	موجودة	موجودة	موجودة	تتلاشى	موجودة \ضعيفة جدا
8	غير موجودة	غير موجودة	غير موجودة	غير موجودة	تتلاشى	تتلاشى

من الجدولين رقم (3) و(4) نلاحظ الأتي :-

1- إن قيمة الاختبار ρ_{eG} لوجود العلاقة ابتعدت عن الحقيقة في تحديد وجود العلاقة بين المتغيرات في التجربة الرابعة بعكس الاختبارات الأخرى .
2- بشكل عام قدرة مقياسي الانتروبي للارتباط في تحديد قوة العلاقة كونهما يعتمدان على قيم الارتباط لتحديد وجود العلاقة بعكس المقاييس الشائعة الأخرى للبحث عن وجود العلاقة بين المتغيرات.

3- إن المقاييس الشائعة للبحث عن الاستقلالية بين المتغيرات اثبت وجود علاقة للتجربة السابعة وهي وجود العلاقة بين عدد مرات الطلاق والبيئة المكانية لحدوث الطلاق ولكن هذه العلاقة لم تحدد قوتها إلا بحساب قيمة ارتباط فاي ومعامل التوافق ومعامل يتس(Y.C) الذي اثبت ضعفها بعكس مقياس الانتروبي للارتباط ρ_{eS} الذي اثبت وجود العلاقة وحدد ضعفها مباشرة في حين أن ρ_{eG} لم يثبت العلاقة بين المتغيرات .

4- عدم وجود علاقة بين عدد مرات الطلاق وجنس المطلق (التجربة الثامنة)

أما قيم الارتباط للمقاييس الشائعة ومقياسي الانتروبي للارتباط . فيمكن توضيحها بالجدول رقم (5) كالاتي :-

الجدول رقم (5) يبين قيم الارتباط للمقاييس الشائعة ومقياسي الانتروبي للارتباط .

التجربة	قيمة الارتباط				
	∅	C	Y	P_{JGe}	P_{SGe}
1	.820	.634	.9940	0.610162	0.73562
2	.763	.607	.9997	0.568419	0.689368
3	.523	.464	.9970	0.290019	0.643802
4	.450	.411	.7530	0.152786	0.509306
5	.218	.213	.4545	0.037051	0.259479
6	.129	.128	.4545	0.01733	0.154301
7	.084	.084	.2505	0.006908	0.097422
8	.011	.011	.0286	0.000105	0.013159

من الجدول رقم (5) نلاحظ ملاياتي :-

1- بشكل عام يملك اختبار يتس اكبر قيمة ارتباط من المقاييس جميعها الشائعة والانتروبي في حالة وجود العلاقة من عدمها ولمختلف أحجام العينات .

2- بشكل عام يملك اختبار P_{JGe} اقل قيمة ارتباط من المقاييس جميعها الشائعة والانتروبي في حالة وجود العلاقة من عدمها ولمختلف أحجام العينات .

3- قيم الارتباط بالنسبة للمقاييس فاي ومعامل التوافق اكبر بنسبة ضئيلة بالنسبة لمقياس الانتروبي للارتباط P_{SGe} عند أحجام العينات الكبيرة في حين يكون العكس عند أحجام العينات المتوسطة والصغيرة .

4- رغم ان الاختبارات الشائعة لاستقلالية المتغيرات وعدم وجود العلاقة اثبت وجود العلاقة الا ان قيم الارتباط للمقاييس الشائعة اثبت ضعف العلاقة وهذا ما أثبته مقياس الانتروبي P_{SGe} للارتباط منذ البداية .

الاستنتاجات والتوصيات :-

أولا (استنتاجات الجانب النظري) :-

1- عدم تأثر طريقة قياس الارتباط (الاقتران)باستخدام الانتروبي لجدول توافق يحتوي على خلية واحدة أو عدة خلايا فيها اقل من 5 تكرارات إذ إن الطريقة تعتمد الثقل الاحتمالي للخلية من خلال احتساب الانتروبي الحدي والمشترك والشرطي بعكس اختبار مربع كأي الذي يتأثر بوجود خلية فيها عدد اقل من 5 .

2- يمكن الوصول إلى صيغة المعلومات المتبادلة بعدة أساليب إذا اعتمد مبدآ التماثل الذي اعتمد من قبل الباحثين كما هو الحال في الصيغة (27) .

ثانيا (استنتاجات الجانب التجريبي والتطبيقي) :-

3- تأثر قيمة مقياس الانتروبي P_{JGe} بشكل كبير عندما تصغر قيمة المعلومات المتبادلة مما يوثر في مصداقية المقياس للبحث عن وجود العلاقة بعكس مقياس الانتروبي P_{SGe} الذي يكون تأثره بشكل طبيعي وهو بذلك يتميز عليه .

4- إن آلية استخدام الانتروبي في جداول التوافق تؤدي للحصول على صيغة للبحث عن الاستقلالية بين المتغيرات وتحديد طبيعة العلاقة من خلال قياس الارتباط في الوقت نفسه .

5- بالنسبة للتجريبيتين السابعة والثامنة لا يمكن الجزم بوجود علاقة بين عدد مرات الطلاق ومتغير جنس المطلق أو البيئة المكانية لحالة الطلاق على الرغم إن التجربة السابعة اثبت المقاييس الشائعة ومقياس الانتروبي للارتباط بوجود علاقة إلا إنها ضعيفة .

6- تميز معامل فاي على مقياس الانتروبي في العينات الكبيرة جدا كما هو الحال في التجريبيتين الأولى والثانية في حين يتميز مقياس الانتروبي للارتباط في العينات الكبيرة والمتوسطة والصغيرة كما الحال في بقية التجارب .

وخلاصة ما تقدم يوصي الباحث إلى اعتماد أسلوب الانتروبي في قياس الارتباط في جدول التوافق من درجة 2xk كونه يستخدم كاختبار لاستقلالية المتغيرات وقياس قيمة الارتباط في الوقت نفسه بعكس الاختبارات الأخرى التي تختبر العلاقة ومن ثم تحسب الارتباط .وكذلك عندما تكون هناك خلية تخالف شروط إجراء اختبار مربع كأي (اقل من 5) الأمر الذي يستوجب دمج الصفوف أو الأعمدة وبهذا يفقد الجدول مستوى من مستويات المتغيرات نتيجة الدمج مما يوثر ذلك على النتائج التي من الممكن إن تكون غير دقيقة لان المقاييس الشائعة تعتمد اغلبها على قيمة مربع كاي الذي بني على العينات الكبيرة ($n \to \infty$) . وأخيرا يوصي الباحث اعتماد الانتروبي في قياس الارتباط في جدول التوافق من درجة 2xk في العينات الكبيرة والمتوسطة والصغيرة .

المصادر:-

1- المشهداني , كمال علوان خلف . عبودي , عماد حازم (2009) " **اختبار الفرضيات الإحصائية** " العراق , بغداد , مكتب الجزيرة للطباعة والنشر , الطبعة الأولى .

2- المشهداني , كمال علوان خلف . عبودي , عماد حازم . عبد الله , سهيل نجم (2012) " **الاختبارات الإحصائية تطبيقات محوسبة باستخدام برنامج spss** " شركة بابل للطباعة المحدودة

3- Alan Agresti (1992) " **A Survey of Exact Inference for Contingency Tables**" *Statistical Science*, Vol. 7, No. 1. pp. 131-153.

4-Fritse .Z, Jos M,F, ten BERGE (1985) " **Afamily of association coefficients for metricscales** " psychometrika – v.50 ,no. 1, 17-24 .

5- Jaakko Astola , Ilkka Virtanen (1982**)" Entropy Correlation Coefficient A measure of statistical dependence for categorized data "** paper presented at the 15 th European meeting of statisticians in Palermo .

6- Jacob cohen (1988) **" Set correlation and contingency tables "** Applied psychological measurement ,v. 12 ,no.4 , p.p 425-434.

7- James O.Berger (1985)**" statistical decision theory and Bayesian Analysis"** springer –verlag , second edition.

8-Jobst Heitzig and Jakob Runge (2011) **" Entropy train diagrams for information – based measures of statistical association "** Geophysical Research Abtracts ,v.13.

9- Joseph Lee Rodgers; W. Alan Nicewander (1988) **" Thirteen Ways to Look at the Correlation Coefficient"** *The American Statistician*, Vol. 42, No. 1. , pp. 59-66.

10- Kouji Yamamoto, Yohei Ban and Sadao Tomizawa(2008) **" Generalized Measure of Departure from No Three-Factor Interaction Model for 2 x 2 x K Contingency Tables"** *Entropy* 2008, *10*, 776-785

11- Narges Abbasi(2008) **"On Maximum Value of Correlation Coefficient "** International Mathematical Forum, no. 34, 1655 – 1658.

12- Nilgun. H , vujica. Y (1985) **" Transfer of Hydrologic Information Along Rivers Partially fed by Karstified Limestones"** karst water resources (proceedings of the Ankara – Antalya symposium) iahs publ .no 161.

13- Olusola Adeyemi (2010) **" Measures of Association for Research in Educational planning and Administration "** Journal of mathematics and statistics 3(3) p.p 82-90 .

14- Pang-ning tan,vipin kumar, jaideep srivastava(2004) **" selecting the right objective measure for assocaiation "** information systems o.29 p.p293-313.

15- Silviu Guisu (1977) **" Information Theory with Application "** McGraw – Hill.Inc.

16- Serpil Aktas and Meral C etin(2002) **" M-ESTIMATORS OF THE UNIFORM ASSOCIATION MODEL IN Rx C CONTINGENCY TABLES "** Journal of Mathematics and Statistics Volume 31 , 83-88.

17- Suzuki .E (1958) **" Weather forecast and Entropy in Information theory"** meteorological Research Institute ,vol. 9 , N.2 p.p 52-62.

18- Vincent Vignerron (2006) **" Entropy- based principle and Generalized Contingency Tables"** European symposium on Artificial neural networks .

19- Yates .f, (1984) **" Tests of significance for 2x2 contingency tables"** J.R.Statist.Soc.147,V.3, P.P 426-463 .

الملحق :-

$H(X) = -\sum_i \; p_i \; \log_2 p_i \qquad , \quad p_i = \frac{O_{i.}}{O_{..}}$

$$H(X) = -\sum_I \frac{O_{i.}}{O_{..}} \log_2 \frac{O_{i.}}{O_{..}}$$

$H(X) = -\frac{1}{O_{..}} [\sum_i O_{i.}(\log_2 O_{i.} - \log_2 O_{..})] \;\; H(X) = -\frac{1}{O_{..}} [\sum_i O_{i.} \log_2 O_{i.} - O_{i.} \log_2 O_{i.}] \qquad \ldots\ldots\ldots \;\; (1)$

وبالطريقة نفسها نستطيع كتابة الانتروبي Y الاتي :-

$H(Y) = -\frac{1}{O_{..}} [\sum_j O_{.j} \log_2 O_{.j} - O_{..} \log_2 O_{.j}] \qquad \ldots\ldots\ldots(2)$

وكذلك كتابة الانتروبي المشترك للمتغيرين (X,Y) الاتي :

$H(X,Y) = -\sum_i \sum_j P_{ij} \log_2 P_{ij} \qquad , p_{ij} = \frac{O_{ij}}{O_{..}}$

$$H(X,Y) = -\sum_i \sum_j \frac{O_{ij}}{O_{..}} \log_2 \frac{O_{ij}}{O_{..}}$$

$H(X,Y) = -[\sum_i \sum_j O_{ij} \log_2 O_{ij} - O_{..} \log_2 O_{..}] \qquad \ldots\ldots\ldots..(3)$

ان المعلومات المتبادلة يمكن تعريفها كالاتي :-

$\mathrm{I(X,Y) = -[H(X,Y) - H(X) - H(Y)]} \qquad \ldots\ldots\ldots\ldots(4)$

وبتعويض (1) و(2) و(3) في (4) نحصل الاتي :-

$$\boldsymbol{I(X,Y) = \frac{1}{O_{..}}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} O_{ij}\log_2 O_{ij} + O_{..}\log_2 O_{..} - \sum_{i=1}^{k} O_{i.}\log_2 O_{i.} - \sum_{j=1}^{k} O_{.j}\log_2 O_{.j}\right]}$$