

أ.م.د. سوزان صابر حيدر*

م. فرهاد علي احمد**

١. الملخص:

تعد المياه من اهم الموارد الطبيعية على الاطلاق حيث يعد عاملا اساسيا تركز عليه حياة الانسان وكافة انشطته الاجتماعية والاقتصادية في مختلف المجالات وتتميز المياه من غيرها من الموارد الطبيعية بكون كميته ثابتة في الكرة الارضية ويتجدد خلال فترة محددة من الزمن بفضل الدورة الهيدرولوجية. ولكن أصبحت المياه في خطر كبير بسبب التلوث التي تتعرض له سواء بسبب ظروف طبيعية أو بسبب نشاط الإنسان. وتلوث الماء لاسباب كيميائية او فيزيائية او بايولوجية تعد من المشكلات الرئيسية لموارد المياه، وصيانتها. وان احدى اهم مواضيع مكافحة موارد المياه من التلوث هو تشخيص وايجاد العوامل (الفيزيائية او الكيميائية) المؤثرة عليها، وايجاد العلاقة (Correlation) بينهما. وان المعادن الثقيلة (Heavy Metals) موجودة بصورة طبيعية في النظام البيئي، مع اختلافات كبيرة في التركيز. لكن ازدياد نسبها مؤخراً يرجع سببها إلى المصادر الصناعية والنفايات الصناعية السائلة، وان تراكم هذه المواد وعدم تفسيخها تعتبر من العوامل الرئيسية المسببة لتلوث المياه. لذ يتناول هذا البحث دراسة استخدام التحليل العاملي (Factor Analysis) لتحديد خصائص الرئيسية للماء واستخدام (Canonical Correlation) في دراسة قوة العلاقات بين المتغيرات ومن ثم دراسة تأثير مجموعة من المتغيرات المستقلة على تلوث المياه(من خلال ايجاد الارتباطات القوية بين المجموعتين من المتغيرات وتحديد مدى معنويتها ونوع الارتباط).

الكلمات الدالة:

نوعية المياه، تحليل الارتباط القويم، التحليل العاملي، تحليل متعدد المتغيرات.

Use a straight correlation to determine the correlation between qualitative variables and heavy metals in the water of Sulaymaniyah province

Abstract:

Water is considered one of the most important natural resources at all. It is considered a fundamental factor that underlies human life and all its social and economic activities in various fields. Water from other natural resources is characterized by the fact that its quantity is constant in the globe and is renewed within a specified period of time thanks to the hydrological cycle. However, water is at high risk due to pollution, both natural and human activity. Water pollution for chemical, physical or biological reasons is one of the major problems of water resources and its maintenance. One of the most important issues in combating water resources from pollution is the diagnosis and finding of the factors (physical or chemical) affecting them, and finding the relationship (Correlation) between them.

Heavy metals are naturally present in the ecosystem, with significant differences in concentration. However, the recent increase in their rates is due to industrial sources and industrial effluents, and the accumulation of these materials and non-decomposition are the main factors causing water pollution. Therefore, this study examines the use of factor analysis (Factor Analysis) to determine the main characteristics of water and use (Canonical Correlation) in the study of the strength of relationships between variables and then study the impact of a set of independent variables on water pollution (by finding

* جامعة السليمانية – كلية الادارة والاقتصاد

** جامعة السليمانية – كلية الادارة والاقتصاد

مقبول للنشر بتاريخ 2018/4/23

strong correlations between the two sets of variables and determine the extent Morale and type of link).

Key words:

Water quality, correlation analysis, factor analysis, multivariate analysis.

. المقدمة:

من الجانب الاقتصاد المعاصر، التطور الاقتصادي والاجتماعي في اي مجتمع يعتمد على كمية المياه المتوفرة لديها، وان زيادة نسبة السكان والاحتياجات الاجتماعية والتطور الاصطناعي وخاصة الزراعة المهنية ادى الى زيادة نسبة المعادن الثقيلة (Heavy Metals) والمواد السامة الاخرى في الماء والتربة. وفي الدول التي تكثر فيها اليابسة مثل (العراق – كوردستان) تعد المياه الجوفية مصدرا ضروريا لتغطية الاحتياجات المائية، بصفة عامة تعتبر المياه الجوفية نقية وخالية من التلوث والبكتيريا الضارة ولكنها قد تتعرض للتلوث نتيجة بعض العوامل الخارجية.

و تتعرض الطبقات السطحية الحاملة للمياه للتلوث بدرجة كبيرة وكلما كان مستوى الماء في تلك الطبقات قريب من سطح الأرض كلما إزدادت قابليتها للتلوث. وقد تنتقل البكتيريا إلي طبقات أعمق خاصة إذا كانت المواد الصخرية المكونة لتلك الطبقات عالية المسامية والنفاذية. لذا المياه الجوفية دائما تتأثر بالعوامل الطبيعية والخارجية، ولحماية آبار المياه فإنه ينبغي أن تحدد مواقعها بعيداً عن مصادر التلوث ويراعي عند تصميمها وإنشائها الحماية الصحية اللازمة. وعلية قبل استخدامنا للمياه الجوفية يجب ان تكون لدينا معلومات دقيقة جدا عن المركبات الكيميائية لمثل هذا النوع من المياه.

وان وجود المعادن الثقيلة (Heavy Metals) في المياه تسبب مشكلة كبيرة في جميع انحاء العالم، وتزدادت بشكل كبير بسبب فعالية الانسان. وان مصدر هذا التلوث هو استخدام النفط ومشتقاته وأستخراج المعادن وبقايا نفايات الاعمال الصناعية والنشاطات الزراعية. وهذه المعادن تدخل جسم الانسان عن طريق الاكل والشرب ووجوده خطر على صحة الانسان.

ولتوضيح (نوعية المياه) بشكل ادق، يجب فحص وتحليل الخصائص الكيميائية والفيزيائية للماء مثل درجة الحرارة – التوصيل الكهربائي – الداله الحامضيه PH – الامونيا - فلوريد – نترات – فوسفات PO4 الخ ولتحليل هذه الكمية الكبيرة من المتغيرات يحتاج الى وقت زمني طويل ولكون وجود علاقة بين اغلب هذه الخصائص، نستطيع ان نأخذ بنظر الاعتبار تركيبة من هذه الخصائص كعوامل اساسي. واحدة من اهم مواضيع مراقبة نوعية الماء: تشخيص العوامل الاساسية وتفكيك العوامل الكيميائية والفيزيائية للماء والعلاقات بينهم. لذا العوامل الفيزيائية بشكل عام تنشأ من خصائص المنطقة المتواجد فيها الماء وطبيعتها، والتي يمكن تشخيصها والحصول عليها استنادا على اللمس – الطعم – الرائحة مثل المياه الغير صافية والمواد العالقة الصلبة، بينما العوامل الكيميائية تنشأ وتتأثر بالمصادر التي يصنعها الانسان مثل المجاري، المجاري الصناعية، سماد كيميائي، لذا الخصائص الكيميائية للماء فله علاقة مباشرة مع خاصية الانحلال للماء مثلاً الحموضة والملوحة والفلوريد والمعادن الثقيلة (Heavy Metals). لذا اذا كان هنالك علاقة (Correlation) قوية بين هذين مجموعتين من المتغيرات (الفيزيائية والكيميائية) يدل على ان كلا المجموعتين نشأ من نفس المصدر. وان التحليل عن طريق المختبرات يحتاج الى وقت طويل وكلفة عالية وان افضل طريقة للتحليل هو استخدام (Multivariate Analysis).

وقد توسع استخدام (Multivariate Analysis) لتجزئة وتحليل البيانات البيئية في السنوات الاخيرة، على سبيل المثال Principle Component Analysis و Canonical Correlation Analysis لتفسير البيانات المعقدة والتعرف على النوعية الجيدة للماء، وعن طريق هذه التحليلات كان بالامكان تحديد وقت ومكان العوامل المؤثرة

والمسببة لتلوث البيئة. وأن استخدام كل من طرق تحليل متعدد المتغيرات Canonical Correlation Analysis, Cluster Analysis, Factor Analysis يعتبر عامل مساعد جيد لتفسير بيانات نوعية الماء.

3. **التحليل العاملي** : هو أسلوب احصائي يستخدم في دراسة الظواهر بهدف ارجاعها الى العوامل المؤثرة فيها، وهو عملية رياضية تستهدف تفسير معاملات الارتباط الموجبة- التي لها دلالة احصائية- بين مختلف المتغيرات. ومن اكثر الطرق التحليل العاملي دقة وشيوعا واستخداماً هي طريقة المكونات الاساسية.

1.3 تحليل المكونات الاصلية Principle Component Analysis

لتشخيص وتصنيف الخصائص المختلفة للماء يتم استخدام تحليل المكونات الاصلية (Principle Component Analysis)، وان عملية تحليل وتجزئة المكونات الاصلية مهمة جدا لتقليل ابعاد البيانات واستخراج عدد قليل من العوامل المستترة بهدف تحليل العلاقة بين المتغيرات. وان عملية تقليل البيانات تحدث عن طريق تبديل البيانات الى مجموعة جديدة من المتغيرات والتي تسمى بالمكونات الاساسية. وان هذه المجموعة الجديدة من المتغيرات لها علاقة قوية مع بعضها البعض وقريبة من بعضها، وفي النهاية فإن جميع المتغيرات تتلخص بعدد من المجاميع وكل مجموعة تسمى بالمكون الاساسي. واكثر ما يميز هذه الطريقة هو ان كل عامل يستخلص اقصى تباين ممكن ويؤدي الى اقل قدر من البواقي كما ان المصفوفة الارتباطية تختزل الى اقل عدد من العوامل المتعامدة.

$$PC_{ij} = \sum_{k=1}^{\infty} w_{ki} x_{kj}$$

إذ ان:

PC_{ij} : المكون (i) من المجموع (j)

w_{ki} : وزن المتغير (k) من المكون (i)

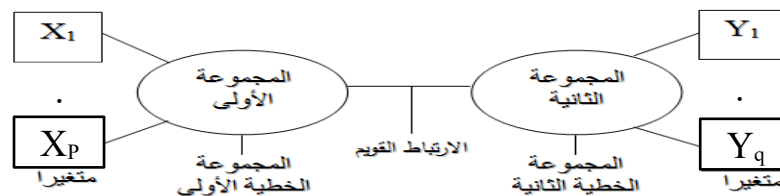
x_{kj} : القيمة الخاصة

وفي النتيجة فإن القرار النهائي للمتغيرات المستخدمة في هذا الجزء (Analysis Principle Component) ومدى تأثيرها على العوامل يعتمد على المجموع الكلي للمتغيرات والتي نحصل عليها من التباين الكلي. وان القيمة الخاصة (Eigen value) لكل عامل هي نسبة من مجموع التباين الكلي للمتغيرات، وأن هذه القيمة مهمة جدا لاكتشاف العلاقة بين العوامل والمتغيرات.

4. الارتباط القوي Canonical Correlation

قبل أكثر من ستة عقود ماضية تم تطوير أسلوب الارتباط القوي من قبل الباحث Hotelling امتداداً وتطويراً للارتباط البسيط Simple Correlation الذي يقيس قوة العلاقة الخطية بين متغيرين (X, Y)، وأيضاً الارتباط المتعدد Multiple Correlation الذي يقيس قوة العلاقة الخطية بين مجموعة من المتغيرات X'S إذ إن ارتباطها بالمتغير (Y) يكون أعظم ما يمكن، وفي هذه الحالة فإن تحليل الارتباط القوي هو حالة عامة للانحدار المتعدد إذ يقوم بإيجاد العلاقة بين مجموعة من متغيرات (Y) مع مجموعة من متغيرات (X). أي ان الارتباط القوي و (Canonical Correlation) يشمل الارتباط بين مجموعتين من المتغيرات الأولى X'S والتي تضم (X1, X2... Xp) والثانية Y'S والتي هي (Y1, Y2... Yq).

أصبح استخدام الارتباط القوي واسعاً في جميع المجالات والعلوم التي تتطلب الحالة فيها دراسة العلاقة بين مجموعة من المتغيرات يمكن فصلها إلى مجموعتين من المتغيرات وبفضل التطور السريع في الحاسبة واستخدام البرامج الجاهزة مكن ذلك الباحثين من استخدامه.



الشكل (1) يوضح الارتباط القوي بين المتغيرات

5. نموذج الارتباط القوي Model of Canonical correlation

هي إحدى الطرق الشائعة لتحليل المتعدد للمتغيرات والغاية من هذا التحليل هو تحديد الارتباط الخطي بين المتغيرات (متغيرات متعددة الارتباط) وفي اغلب البحوث الاحصائية تنقسم المتغيرات الى مجموعتين وهما المتغيرات المستقلة والمتغيرات المعتمدة. وان تحليل Canonical Correlation Analysis هي طريقة لدراسة العلاقة Correlation بين المجموعتين من المتغيرات والمتمثلة بمصفوفة Y, X.

لذا Canonical Correlation Analysis تقوم بتكوين مجموعة من المتغيرات القانونية (p) للمتغيرات X وكذلك تقوم بتكوين مجموعة من المتغيرات القانونية (q) للمتغيرات Y، وهذا يطابق العوامل المكونة عن طريق Factor Analysis. والهدف الرئيسي من استخدام Canonical Correlation Analysis هو تكوين مجموعتين جديتين من المكونات والمتمثلة بـ $U = px$ و $V = qy$ هذه تراكيب خطية من المتغيرات الأولية. وفي هذه الحالة فإن قيمة كل من p, q يتم تحديدها بطريق تجعل الارتباط بين U, V اكبر مايمكن.

وفي هذا البحث تم استخدام المتغيرات المعتمدة والمتمثلة بالعوامل الفيزيائية والمتغيرات المستقلة هي العوامل الكيميائية. وان طريقة Canonical Correlation Analysis تستعمل لتقليل حجم البيانات المستخدمة. ويحاول تحديد العلاقات الخطية المتعدد بين المجموعتين وبالتالي اختزال هذه العلاقات إلى أقل عدد من المتغيرات (العوامل) القوية، إذ إن الارتباط بين كل زوج من المتغيرات سوف يكون الارتباط بينهما ارتباطاً بسيطاً.

فإذا كانت لدينا (N) من المشاهدات و (P) تمثل المتغيرات من المجموعة الأولى و (q) تمثل المتغيرات من المجموعة الثانية بحيث $N \geq p + q$ وإن $p \leq q$ أو $p \geq q$ أي إن:

$$X = [X_{ij}] \quad i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, P$$

$$Y = [Y_{ij}] \quad i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, q$$

لذا فإن التراكيب الخطية $c'X$ و $d'Y$ هي:

$$Z_{x_2} = U_i = c_2'X, \quad Z_{x_r} = U_i = c_r'XZ_{x_1} = U_i = c_1'X$$

$$Z_{y_2} = V_j = d_2'Y, \quad Z_{y_r} = V_j = d_r'YZ_{y_1} = V_j = d_1'Y$$

$$r = \min(p, q) \quad i, j = 1, 2, \dots, n$$

وان r يمثل عدد الأزواج التراكيب الخطية، (c') و (d') يمثلان متجه الأوزان، U_i, V_j تراكيب خطية لـ p من متغيرات X'S و q من متغيرات Y'S.

وكما قلنا ان الارتباط بين أزواج المتغيرات القوية يدعى بالارتباط القوي Canonical correlation Analysis (C.C.A) وان كل زوج مرتبط من هذه المتغيرات يكون غير مرتبط مع المتغيرات القوية التي ارتبطت فيما بينها أي يكون الارتباط صفراً. ويجب ملاحظه ان الأوزان تختار بحيث يكون كل متغير قويم قياسياً بوسط حسابي صفر وتباين واحد.

تكم اهمية الارتباط القوي في تحليله والذي يقوم على اساس اختيار الأوزان c و d بحيث تظهر اهمية الارتباط القوي والتي تكون فيها قيمته بين Z_x, Z_y للمجموعتين اعظم ما يمكن، إذ ان الزوج الاول من الارتباط القوي من التراكيب الخطية ليس بالضرورة ان يكون هو الوحيد الذي يعطي اعظم ارتباط ممكن الحصول عليه بين المجموعتين، إذ يكون هناك على الاقل زوجاً ثانياً من التراكيب للمجموعتين أي ان العدد الاكبر للارتباطات القوية التي يمكن ان تعرف للمشكلة معنوية المعطاة عددها مساو لعدد المتغيرات في المجموعة الصغرى ولكن ليس بالضرورة أن تكون جميع الارتباطات القوية التي تتضمنها المشكلة معنوية.

6. تحديد الأوزان Determine the weights

ان خطوة تحديد الأوزان الخاصة بالزوج الاول ليست الوحيدة اذا قد يكون هناك ارتباط قويم اخر يتطلب ايضاً تحديد الأوزان الخاصة وهكذا. فتحديد الأوزان الخاصة بالزوج الثاني والذي يمثل اعظم ارتباط قويم يمكن الحصول عليه بالنسبة لبقية الأزواج الاخرى باستثناء الزوج الاول الذي حدد في الخطوة الاولى، وقد نستمر في العملية بحيث نحصل على عدد غير محدد من الأزواج الا في حالة كون $P \leq q$ فان اوزان الأزواج من التراكيب الخطية سوف يكون لـ P من الأزواج أو العكس، فكل زوج من هذه التراكيب لها اوزان مختلفة من الاخرى.

ان ايجاد المتجهين (c') و (d') إذ ان الارتباط بين المتغيرات القوية (U_x) و (V_y) يكون اعظم ما يمكن بالمتغيرات القوية Canonical variables والمعاملات (c) و (d) تعرف بالمعاملات القوية Canonical coefficient. وكل زوج من المتغيرات القوية يشكل نسبة التباين المشترك لوصف العلاقة القائمة بين المجموعتين، فالزوج الاول يشكل النسبة الاكبر، يشكل الزوج الثاني نسبة اقل من الزوج الاول وهكذا لبقية الأزواج (اي ان الزوج الاول هو صاحب اعلى نسبة يليه الثاني والثالث وهكذا).

لغرض اشتقاق وتحديد الأوزان المناسبة للمتغيرات القوية بين المتغيرات في كل زوج والتي تجعل الارتباط في أقصى حد، علينا حساب معامل الارتباط واشتقاق معادلته:-

نفرض ان لدينا $V = d'y, U = c'x$ تركيب خطية الاولى لها P من المتغيرات والثانية q من المتغيرات وتخضع

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N_{(p+q)}(M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix})$$

للتوزيع الطبيعي متعدد المتغيرات:

ان كل تركيبة خطية يمكن كتابتها:

$$Z_{Xr} = U_i = C_{1i}X'x = U_1X_1 + \dots + U_pX_p \quad i = 1, 2, \dots, n \quad r = 1, 2, \dots, p$$

$$Z_{Yr} = V_j = d_{1j}Y_1 + \dots + d_{qj}Y_j \quad j = 1, 2, \dots, n \quad r = 1, 2, \dots, q$$

U_i تمثل التركيبة الخطية الاولى (المتغير القويم للمجموعة X'S)

V_j تمثل التركيبة الخطية الثانية (المتغير القويم للمجموعة Y'S)

ويمكن كتابة التراكيب الخطية بالصيغة الاتية:

$$Z_{Xr} = U'_x = U_1X_1 + \dots + U_pX_p$$

$$Z_{rn} = U_{yn} = V'_y = V_1Y_1 + \dots + V_qY_q$$

ويمكن اثبات:

$$\bar{Z}_x = U_1\bar{X}_1 + \dots + U_p\bar{X}_p$$

$$S_{xxv} = Z_x - \bar{Z}_x = U_1(X_1 - \bar{X}_1) + \dots + U_p(X_p - \bar{X}_p)$$

وبتربيع الطرفين واخذ المجموع لهما الى (n) من المشاهدات نحصل على :

$$S_{xxu} = \sum_{i=1}^n ZX^2 = U_1^2 \sum_{i=1}^n (X_1 - \bar{X}_1)^2 + \dots + U_p^2 \sum_{i=1}^n (X_p - \bar{X}_p)^2 + 2U_1U_2 \sum_{i=1}^n (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) + \dots + 2(U_{p-1}U_p) \sum_{i=1}^n (X_{p-1} - \bar{X}_{p-1})(X_p - \bar{X}_p)$$

بافتراض ان جميع المتغيرات لها متوسط (M) وتباين (Σ) , اي ان: $x \sim N(M, \Sigma)$ ان Σ تمثل مصفوفة التباين والتباين المشترك للمجتمع و S تمثل تقدير للمصفوفة .

$$U'S_{xx}U = U_1^2S_{11} + \dots + U_p^2S_{pp} + 2U_1U_2S_{12} + 2U_{p-1}U_p - S_{p-1}S_p$$

$$S_{jk} = \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \quad j, k = 1, 2, \dots, p$$

$$\therefore U = \begin{bmatrix} U_1 \\ \vdots \\ U_p \end{bmatrix} \quad S_{xx} = \begin{bmatrix} S_{11} & \dots & S_{1p} \\ \vdots & \ddots & \vdots \\ S_{p1} & \dots & S_{pp} \end{bmatrix}$$

$$\sum_{i=1}^n ZX^2 = U'S_{xx}U$$

وبنفس الطريقة نبرهن:

$$\sum_{i=1}^n Z_x Z_y = U'S_{xy}V$$

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_q \end{bmatrix} \quad S_{yy} = \begin{bmatrix} S_{11} & \dots & S_{1q} \\ \vdots & \ddots & \vdots \\ S_{q1} & \dots & S_{qq} \end{bmatrix} \quad S_{xy} = \begin{bmatrix} S_{x1y1} & \dots & S_{xpy1} \\ \vdots & \ddots & \vdots \\ S_{xpy1} & \dots & S_{xpyq} \end{bmatrix}$$

$$S_{yy}S_{jk} = \sum_{i=1}^n (Y_{ij} - \bar{Y}_j)(Y_{ik} - \bar{Y}_k)$$

$$S_{xy}Y_k = \sum_{i=1}^n (X_{ij} - \bar{X}_j)(Y_{ik} - \bar{Y}_k) \quad j = 1, 2, \dots, p, \quad k = 1, 2, \dots, q$$

ومن المعادلات اعلاه فان معامل الارتباط:

$$P_{x,y} = \frac{U'S_{xy}V}{\sqrt{(U'S_{xx}U)(V'S_{yy}V)}}$$

S_{xx} : مصفوفة التباين للمستويات في المجموعة X

S_{yy} : مصفوفة التباين للمستويات في المجموعة Y

S_{xy} : مصفوفة التباين للمستويات في المجموعة X, Y

ويمكن حساب الارتباط القويم عن طريق مصفوفة الارتباط اذا كان لدينا P من المستويات في المصفوفة X و q من المستويات في المصفوفة y فأن:

$$D_x = \begin{bmatrix} \sqrt{S_{x1x1}} & \dots & 0 \\ \vdots & & \\ 0 & \dots & \sqrt{S_{xpxp}} \end{bmatrix}, D_y = \begin{bmatrix} \sqrt{S_{y1y1}} & \dots & 0 \\ \vdots & & \\ 0 & \dots & \sqrt{S_{yqyq}} \end{bmatrix}$$

$$R_{xx} = \begin{bmatrix} 1 & r_{x1x2} & \dots & r_{x1xp} \\ r_{x2x1} & 1 & \dots & r_{x2xp} \\ \vdots & \vdots & \ddots & \vdots \\ r_{xpx1} & r_{xpx2} & \dots & 1 \end{bmatrix}, R_{yy} = \begin{bmatrix} 1 & r_{y1y2} & \dots & r_{y1yq} \\ r_{y2y1} & 1 & \dots & r_{y2yq} \\ \vdots & \vdots & \ddots & \vdots \\ r_{yqy1} & r_{yqy2} & \dots & 1 \end{bmatrix}, R_{xy} = \begin{bmatrix} r_{x1y2} & r_{x1y2} & \dots & r_{x1yq} \\ r_{x2y1} & r_{x2y2} & \dots & r_{x2yq} \\ \vdots & \vdots & \ddots & \vdots \\ r_{xpy1} & r_{xpy2} & \dots & r_{xpyq} \end{bmatrix}$$

ويمكن اعادة صياغة المعادلات بالشكل الاتي:

$$\sum_{i=1}^n Z_x^2 = U' d_x R_{xx} d_x U$$

$$\sum_{i=1}^n Z_y^2 = V' d_y R_{yy} d_y V$$

$$\sum_{i=1}^n Z_x Z_y = U' d_x R_{xy} d_y V$$

ولتبسيط المعادلات اعلاه نفرض:

$$U' d_x = c', \quad V' d_y = d', \quad d_x U = c, \quad d_y V = d$$

فتكون لدينا المعادلات الاتية:

$$\sum_{i=1}^n Z_x^2 = c' R_{xx} c$$

$$\sum_{i=1}^n Z_y^2 = d' R_{yy} d$$

$$\sum_{i=1}^n Z_x Z_y = c' R_{xy} d$$

$$r_{zxzy} = \frac{c' R_{xx} d}{\sqrt{(c' R_{xx} c)(d' R_{yy} d)}}$$

ولايجاد متجهات الاوزان (c') ، (d') بحيث يكون كل متغير قويم قياسياً بوسط حسابي صفر وتباين 1 وبما ان حاصل ضرب كل من Z_x ، Z_y بوسط ثابت اعتباطياً لا يغير قيمة معامل الارتباط بينهم فانه من الضروري تحديد c ، d بالنسبة للثوابت في اختيار c ، d في هذه الحالة تحقق الشرط الاتي :-

$$c' R_{xx} c = d' R_{yy} d = 1$$

ويتم استخراج الاوزان وحساب معامل الارتباط القويم باستعمال دالة التعظيم الاتية:

$$f = c' R_{xy} d - \frac{\sqrt{\lambda_1}}{2} c' R_{xx} c - \frac{\sqrt{\lambda_2}}{2} d' R_{yy} d$$

إذ ان $\frac{\sqrt{\lambda_1}}{2}$ ، $\frac{\sqrt{\lambda_2}}{2}$ مضاعفات لا كرانج. وبأخذ المشتقة لـ F بالنسبة لـ c ، d ومساواتها للصفر نحصل :

$$\frac{\partial f}{\partial c} = R_{xy} d - \sqrt{\lambda_1} R_{xx} c = 0$$

$$\frac{\partial f}{\partial d} = c' R_{xy} - \sqrt{\lambda_2} d' R_{yy} = 0$$

نضرب المعادلتين اعلاه في c' و d يكون لدينا:

$$c'R_{xy}d - \sqrt{\lambda_1} c'R_{xx}c = 0$$

$$c'R_{xy}d - \sqrt{\lambda_2} d'R_{yx}d = 0$$

وبحل المعادلتين اعلاه وبالتعويض عن قيمة c نحصل على :

$$R^{-1}_{yy}R_{yx}R^{-1}_{xx}R_{xy} - \lambda I = 0$$

يطلق على المعادلة اعلاه بالمعادلة المميزة (Eigen Equation) للمصفوفة $R^{-1}_{yy}R_{yx}R^{-1}_{xx}R_{xy}$ وان عدد الجذور التي لا تساوي صفر المتحقق من هذه المعادلة تساوي q وتدعى بالقيم المميزة (Eigen values).

7. الجانب التطبيقي :

تم جمع البيانات من (مركز البحوث العلمية في محافظة السليمانية) لـ (30) عينه تم اختيارهم عشوائيا للمياه الجوفية في المحافظة واطرافها، فقد تم الاعتماد على نسبة المواد الاولية والمعادن الثقيلة في الماء، في قيد هذه الدراسة تم استخدام 15 مكون من مكونات الماء لتحديد المكونات الاساسية منها. ولايجاد هذه المكونات الاساسية تم الاستفادة من Principle Component، فقد تم تقسيم المتغيرات الـ (15) في هذه الدراسة الى مجموعتين، حيث ان متغيرات المجموعة الاولى Y'S وتضم (Y1, Y2...Y9)، وتمثل (9) متغيرات نوعية، اما متغيرات المجموعة الثانية X'S وتضم (X1, X2...X6)، وتمثل (6) متغيرات المعادن الثقيلة. وبعد التحقق من ان بيانات المتغيرات المعتمدة تتوزع طبيعيا، ادخلت مجموعتي Y'S و X'S في برنامج (SPSS Ver (24)) لحساب الارتباط القويم بين المجموعتين.

الجدول رقم (1) يوضح مصفوفة الارتباط بين المواد الاولية والمعادن الثقيلة في الماء

	الفلوريد	الكالسيوم	السلفات	بيكاربونات	النترات	الفوسفات	قلوية الماء	الصوديوم	البوتاسيوم	الزنك	الكاديوم	المنيوم	البروم	المغنيسيوم	المنغنيز
	F	Cl	SO4	HCO3	NO3	PO4	Alkalinity	Na	K	ZN	Cd	Al	Br	Mg	Mn
الفلوريد F	1														
الكالسيوم Cl	0.474	1													
السلفات SO4	0.466	0.553	1												
بيكاربونات HCO3	-0.2	0.109	-0.266	1											
النترات NO3	-0.417	0.06	-0.156	0.185	1										
الفوسفات PO4	-0.106	0.013	-0.095	0.05	0.24	1									
قلوية الماء	0.373	0.578	0.819	-0.096	-0.178	-0.125	1								
الصوديوم Na	0.233	0.564	0.291	0.205	-0.197	-0.088	0.568	1							
البوتاسيوم K	0.133	0.52	0.219	0.333	0.534	0.425	0.088	0.115	1						
الزنك ZN	-0.388	-0.175	-0.286	0.096	-0.061	0.066	-0.382	-0.293	-0.059	1					
الكاديوم Cd	-0.092	-0.101	-0.098	0.185	0.001	-0.032	-0.14	-0.159	-0.107	0.199	1				
المنيوم Al	0.008	-0.191	0.168	-0.477	0.047	-0.089	0.091	-0.097	-0.293	-0.15	0.113	1			
البروم Br	0.266	0.269	0.057	0.096	0.207	0.612	-0.04	0.602	0.027	-0.062	-0.17	1			
المغنيسيوم Mg	0.467	0.814	0.706	0.187	-0.183	0.042	0.785	0.592	0.335	-0.23	-0.029	-0.32	0.165	1	
المنغنيز Mn	0.399	0.569	0.348	0.168	-0.247	-0.041	0.44	0.528	0.284	-0.16	-0.031	-0.07	0.203	0.604	1

تحديد المكونات الاساسية:

ولايجاد المكونات الاصلية من مجموعة مكونة من (P) من المتغيرات، نستخدم طريقة الدوران في تحليل العوامل (Factor Analysis)، فالمكون الاساسي الاول هو المكون الذي له اكبر قيمة تباين. وهكذا بالنسبة للمكونات الاخرى بالتوالي. وفي هذه الدراسة لتسهيل المصفوفة المستخدمة تم الاستفادة من طريقة (Normal Varimax) في (Factor Analysis) اعتمادا على قيم Eigen value. وفي دراستنا تم اختيار 3 مكونات التي تكون فيها قيم

(Eigen value) أكبر ما يمكن، ويتبين أيضا أن قيم التباين تكون مطابقة للمكونات الجديدة بشكل متوالي، كما موضح في الجدول رقم 2، وهذه النتائج تعتبر نسبة مئوية من التباين الكلي .
الجدول رقم (2) قيم Eigen value بالنسبة للمكونات الجديدة

Component					
PC5	PC4	PC3	PC2	PC1	
-0.075	-0.463	0.201	-0.221	0.625	الفلوريد
0.088	0.081	-0.043	0.190	0.854	الكلوريد
0.233	0.090	0.292	-0.277	0.752	السلفات
0.011	0.206	-0.662	0.533	0.069	بيكاربونات
0.099	0.658	0.311	0.589	-0.175	النترات
0.040	-0.260	0.392	0.645	0.005	الفوسفات
0.096	0.233	0.067	-0.312	0.815	قلوية الماء
-0.171	0.166	-0.301	-0.075	0.682	الصوديوم
0.037	0.137	0.196	0.782	0.418	البوتاسيوم
0.320	-0.332	-0.320	0.218	-0.398	الزئبق
0.851	-0.069	-0.263	0.015	-0.174	الكالسيوم
0.366	0.184	0.558	-0.472	-0.146	المنيوم
0.019	-0.410	0.365	0.678	0.268	البروم
0.111	0.003	-0.193	0.046	0.920	المغنيسيوم
0.082	-0.163	-0.218	0.030	0.706	المنغنيز
1.099	1.217	1.652	2.657	4.652	Eigen value
7.326	8.112	11.010	17.716	31.013	التباين الكلي%
75.177	67.852	59.740	48.730	31.013	التباين التراكمي%

من النتائج التي تم الحصول عليها من تحليل وتجزئة المكونات الأساسية يتضح أن هنالك خمسة مكونات أساسية التي تكون بها قيم Eigen value أكبر من الواحد وتعد أهم خمسة مكونات أساسية للماء التي لها أكبر تأثير ممكن على نوعية الماء وبأخذ بنظر الاعتبار تفسير وإيرانس التي تكون فيها التباين بنسبة 75.177% من التباين الكلي .
يعتبر خاصية كل من الكلور Cl، الكبريتات أو السلفات SO_4 ، المغنيسيوم Mg من أهم الخواص في المكون الأساسي الأول بنسبة تباين 31.013%، و خاصية الفوسفات PO_4 ، Al من أهم الخواص في المكون الأساسي الثاني بنسبة تباين 17.716%، و خاصية البروم Br من أهم الخواص في المكون الأساسي الثالث بنسبة تباين 11.01%، و خاصية النترات NO_3 من أهم الخواص في المكون الأساس الرابع بنسبة تباين 8.112% وأخيرا Cd من أهم الخواص في المكون الأساسي الخامس بنسبة تباين 7.326% .

العلاقة بين مكونات نوعية الماء والمعادن الثقيلة (Heavy Metals):

في هذه الدراسة تم الاستفادة من Canonical Correlation Analysis لتشخيص العلاقة بين مجموعة المتغيرات المستقلة (predict variables) والمتمثلة بـ $ALK, \dots, PO_4, NO_3, HCO_3, SO_4, CL, F$ ومجموعة المتغيرات المعتمدة (response variables) والمتمثلة بـ Mn, Mg, Cd, Zn . ومن النتائج التي تم الحصول عليها من هذا التحليل تمثل قيم Correlation بين متغيرات القانونية (R)، مستوى المعنوي significance level، درجة الحرية degrees of freedom و نتائج F test، وعلى أساس هذه النتائج التي تم الحصول عليها نستطيع أن نقرر فرضية البحث من وجود أو عدم وجود علاقة بين المجموعتين من المتغيرات. ويجب أن يكون عدد المجاميع القانونية متساوية مع عدد المتغيرات المعتمدة. ولدراسة العلاقة بين المتغيرات المسقلة والمتغيرات المعتمدة نعتمد على قيمة Correlation Coefficient كما موضح في الجدول رقم (3)، ويتبين من خلال الجدول رقم (4) أن فقط المجموعة القانونية 1, 2, 3 على التوالي لها قيمة R بحدود (0.77, 0.788, 0.975) ويمكن الاستفادة منها للدراسة. الجدول رقم (3) يوضح قيم الارتباطات القوية للمجموعة الأولى والثانية

Y-set canonical variates							x-set canonical variates						
6	5	4	3	2	1		6	5	4	3	2	1	
0.12	-0.58	-0.04	0.46	-0.22	-0.76	زنك	0.47	0.72	0.13	-0.71	0.51	0.64	فلوريد
0.6	0.82	-0.09	0.09	0.02	-0.13	الكالسيوم	0.15	-0.09	0.43	0.01	-0.28	-0.81	كلوريد
0.45	-0.5	0.34	-0.7	-0.32	-0.33	المنيوم	1.36	0.34	0.41	-0.54	-0.39	-0.92	سلفات
0.12	-0.05	0.11	-0.31	0.91	-0.34	بروم	0.22	1.16	0.18	0.3	-0.04	-0.22	بيكاربونات
-0.31	0.12	-0.67	-0.13	-0.39	1.09	مغنيسيوم	-0.02	0.77	0.06	-1.12	-0.03	0.09	نترات
0.69	-0.41	-0.12	0.61	0.17	0.78	منغنيز	0.18	0.19	0.26	-0.35	0.42	-0.66	فوسفات
							-1.87	-0.02	0.29	0.12	-0.09	0.68	قلوية الماء
							0.61	-0.31	0.07	-0.34	0.15	0.62	صوديوم
							-0.76	-1	-0.25	0.66	0.74	0.62	بوتاسيوم

الجدول رقم (4) يوضح قيم الجذور المميزة واختبار F للمجموعتين

Sig.	D.F	F	Eigenvalue	Rc	Root
000	54	3.234	19.321	0.975	1
0.055	40	1.525	1.637	0.788	2
0.218	28	1.258	1.459	0.77	3
0.787	18	0.71	0.432	0.549	4
0.9	10	0.471	0.213	0.419	5
0.978	4	0.108	0.02	0.139	6

تبين من هذه الدراسة بأن في المكون القانوني الاول ان كل من كلوريد و سولفات في المتغيرات النوعية و زنك و مغنيسيوم و منغنيز في المتغيرات المواد الثقيلة لها اكبر تأثير ممكن على هذا المكون. اما بالنسبة للمكون القانوني الثاني ان كل من بوتاسيوم في المتغيرات النوعية و بروم في المتغيرات المواد الثقيلة لها اكبر تأثير ممكن على هذا المكون. وبالنسبة للمكون القانوني الثالث ان كل من فلوريد في المتغيرات النوعية و المنيوم في المتغيرات المواد الثقيلة لها اكبر تأثير ممكن على هذا المكون. بينما المتغيرات الاخرى التي لم تذكر لها تأثير قليل على هذه المكونات ويمكن الاستغناء عنها، كما موضح في الجدول رقم (3).

ومن الجدول رقم (4) ويمكننا القول بأن هنالك علاقة كبيرة بين مكونات نوعية للماء والمواد الثقيلة وبمعنى اخر انه توجد هنالك ارتباط قوي بين هاتين المجموعتين، لذا نستطيع ان نقرر بالاستناد على هذه المجاميع القانونية الثلاثة أن سبب هذه العلاقة القوية هي ان هاتين المجموعتين من المتغيرات تنشأ من نفس المصدر، بينما المجاميع الاخرى من المتغيرات القانونية لها ارتباط ضعيف ويمكن الاستغناء عنها. ومن نفس الجدول رقم (4) نلاحظ ان جميع القيم للاختبار F غير معنوية لان القيمة الاحتمالية للجذور المميزة جميعها اكبر من المستوى المعنوية ($\alpha = 0.05$) ماعدا الجذر المميزة الاول، حيث ان القيمة الاحتمالية لها يساوى 0.000 و عند مقارنة هذه القيمة مع مستوى المعنوية (α) نلاحظ بأن القيمة اقل من 0.05 اي ان الجذر المميزة الاول معنوي من الناحية الاحصائية. والجدول ادناه رقم (4) يبين الاوزان القوية للمجموعتين من المتغيرات القوية والتي تجعل الارتباط في أقصى حد.

الجدول رقم (5) يبين الاوزان القوية للمجموعة الاولى و الثانية من المتغيرات

Y-set canonical vareate							X-set canonical vareate						
6	5	4	3	2	1		6	5	4	3	2	1	
0.13	-0.3	-0.71	0.51	-0.08	-0.34	زنك	0.46	0.05	0.38	-0.12	0.27	0.53	فلوريد
0.66	0.66	-0.3	0.1	-0.12	-0.1	الكالسيوم	0.02	-0.09	0.02	0.26	0.16	0.82	كلوريد
0.53	-0.32	-0.01	-0.71	-0.33	-0.02	المنيوم	0.07	-0.15	-0.09	-0.2	-0.2	0.83	سولفات
0.1	-0.1	-0.25	-0.08	0.93	0.21	بروم	-0.16	0.67	0.06	0.58	0.23	0.01	بيكربونات
-0.06	0.13	0.02	0.3	0.02	0.94	مغنيسيوم	-0.52	0.2	-0.19	-0.33	0.27	-0.13	نترات
0.46	-0.25	0.45	0.44	0.18	0.55	منغنيز	-0.14	-0.07	-0.52	-0.17	0.7	0.09	فوسفات
0.96	0.72	0.86	1.07	1.03	1.36	مجموع المربعات	-0.26	-0.06	0.19	-0.09	-0.27	0.88	قوة الماء
0.16	0.12	0.14	0.18	0.17	0.23	نسبة التباين	0.08	-0.15	0.5	0.2	-0.01	0.6	صوديوم
							-0.27	-0.04	-0.09	0.15	0.73	0.32	بوتاسيوم
							0.69	0.56	0.76	0.66	1.36	2.91	مجموع المربعات
							0.08	0.06	0.08	0.07	0.15	0.32	نسبة التباين

الاستنتاجات:

تم التوصل إلى مجموعة من النتائج عن طريق الدراسة النظرية والتطبيقية ، أهمها.
ان اسلوب التحليل العاملي هو مجموعة من الاساليب الاحصائية التي تهدف الى الكشف عن المتغيرات المشتركة لظاهرة معينة وتلخيصها في عدد قليل من العوامل ليفسر اكبر نسبة من التباين في مجموعة من المتغيرات.
تعد طريقة المكونات الاساسية اكثر الطرق التحليل العاملي استعمالا في استخلاص العوامل، والتي تهدف الى تلخيص البيانات مع المحافظة على اكبر قدر من المعلومات.

توجد هنالك علاقة قوية بين مكونات النوعية للماء والمواد الثقيلة وهذا يدل على انه توجد هنالك ارتباط قوي بين هاتين المجموعتين ونستنتج من ذلك بأن سبب هذه العلاقة القوية هي ان هاتين المجموعتين من المتغيرات تنشأ من نفس مصدر التلوث.

ان طريقة Canonical Correlation Analysis طريقة جيدة لتقليل حجم البيانات المستعملة . وتحديد قوة العلاقات الخطية المتعددة بين المجموعتين وبالتالي اختزال هذه العلاقات إلى أقل عدد من المتغيرات (العوامل). النتائج التي حصلنا عليها من تحليل الارتباط القويم بين المتغيرات النوعية والمتغيرات المعادن الثقيلة يدل على وجود ارتباط قوي بينهما عند المستوى المعنوي 0.05 . وان الملوثات مثل الفوسفات، الامونيا، والنترات تنشأ عادة بكثرة من فعاليات الانسان مثل المجاري العامة والزراعة والصناعة. وكذلك ان منبع المواد الثقيلة تنشأ لاسباب جيولوجية او عن طريق الانسان. واستنادا على النتائج التي تم الحصول عليها من تحليل القويم تظهر ان كلا من المجموعتين من المتغيرات تنشأ من منبع واحد هو صنع البشر.

ان المتغيرات المستقلة نجحت في تفسير نسبة (75.177) من التباين الكلي الحاصل في المتغيرات المعتمدة.

التوصيات:

استنادا على المعلومات النتائج التي حصلنا عليها والبحوث والدراسات السابقة نستطيع اقتراح بان الطرق الاحصائية مثل تحليل المكونات الاساسية والارتباط القويم طرق حل مناسبة لادارة جودة المياه وايجاد منابع تلوث المياه.

المصادر

1. الراوي, اسماء غالب 1996 استخدام الارتباط القويم معياراً لدمج الجداول التوافقية مع تطبيق عملي (رسالة ماجستير), الجامعة المستنصرية, كلية الادارة والاقتصاد.
2. الصوفي , عبد المجيد رشيد 1985 , اختبار χ^2 واستخدامه في التحليل الاحصائي , بيروت: الطبعة الاولى .
3. الكبيسي , مائل كامل 1998 , استخدام الارتباط القويم في دراسة العلاقة بين درجات مواد المفاضلة في القبول ودرجات المواد العلمية للسنة الاولى في كليات المجموعة الطبية. رسالة ماجستير, الجامعة المستنصرية , كلية الادارة والاقتصاد
4. الالوسي, احمد صالح 1987 مقدمة تحليل متعدد المتغيرات الجزء الثاني-جامعة بغداد الادارة والاقتصاد.
5. جاسم, سليمة حمادي " (2006) الارتباط القويم والشبكات العصبية الاصطناعية "رسالة ماجستير كلية الإدارة والاقتصاد ، بغداد.
6. يوسف , هيام تحسين (1990) استخدام الارتباط القويم لمتعدد المتغيرات في تحليل العلاقة بين اداء الطالب في السنة الاخيرة للجامعة والسنوات السابقة لها , رسالة ماجستير , الجامعة المستنصرية , كلية الادارة والاقتصاد.
- 7- Levine. S, 1989 " Canonical analysis and Factor comparison " Sage university papers ,Beverly Hills Landon.
- 8- Thompson, B, 1985 "Canonical Correlation analysis "uses and interpretation sage university papers, London.
- 9- Thorndike, R.M., 2000, "Handbook of applied multivariate statistics and mathematical modeling ".
- 10- Leurgans, S.E et al, 1993, canonical correlation analysis when the data are curves, J.R. statist Soc (55) No. (39).