المجلة العراقية للعلوم الاحصائية (17) 2010
عدد خاص بوقائع المؤتمر العلمي الثاني للرياضيات –الاحصاء والمعلوماتية
ص ص [227– 238]

معالجة البيانات غير التامة وتقديرها بطريقة المكونات الرئيسية

اسوان محمد طيب النعيمي*

الملخص

ان معالجة البيانات غير التامة (المفقودة) يتطلب معرفة سبب الفقدان ، وان اختلاف الآلية التي فقدت بها البيانات يؤدي الى اختلاف مواصفات العينة المتاحة ، والبيانات المعتمدة في الدراسة فقدت تماماً وبشكل عشوائي (MCAR) ، وتم تحويلها الى بيانات تامة . ثم ملاحظة وجود مشكلة التعدد الخطي وكشف عنها بمقياس معامل تضخيم التباين (VIF) ثم تقديرها بطريقة انحدار المركبات الرئيسة وطبق ذلك على متغيرات إنتاج النفط واستهلاك المنتجات المكررة في أقاليم معينة من العالم (ألف برميل يومياً) للفترة من (1970-2004) .

Mechanism of Missing Data and Estimating them by Principal Components

Abstract

The mechanism of missing data requires knowing the reason behind missing them because different mechanism easy in which data is missed give rise to different specifications of the available sample. The data on which this study depends has been missed completely at random (MCAR) and changed into complete one. It can also be possible to find the magnitude of unbias after changing the data into complete one. Then observing the problem of multi linear variance which can be detected by Variance Inflation Factor (VIF) gauge. The estimation of data by the principal component, method follows that. Finally the study applies this data on the variants of oil products and the consumption of the refined products in certain regions in the world (1000 barrels a day from 1970 to 2004).

* مدرس مساعد/ قسم نظم المعلومات الادارية/كلية الادارة والاقتصاد / جامعة الموصل
تاريخ التسلم :2009/7/1 _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ تاريخ القبول : 6 /12 /2009

مقدمة

ان فكرة انجاز بحث حول البيانات غير التامة جاءت من ظروف بلدنا العزيـز ومـا تعرض له من ويلات الحروب اذ أدى ذلك الى فقدان الكثير من البيانات المهمة وفـي جميـع نواحي الحياة الاقتصادية والطبيعية والصحية والعلمية الصرفة ... الخ وتحويل هذه البيانـات غير التامة الى بيانات تامة . كما ان أسباب الفقدان مختلفة ، منها ما يكون خارجا عـن ارادة المعنيين او تكون بإرادة المعنيين أي يكون مخطط لذلك بسبب الكلفة او المخاطرة او بـسبب عدم توفير الإمكانيات للمعاينة ، كما ان اسباب الفقدان تؤدي الى اختلاف الآلية التي فقدت بها البيانات وان اختلاف هذه الآلية يؤدي الى اختلاف مواصفات العينة المتاحة للدراسة .

وعند إيجاد المشاهدات المفقودة للمتغيرات التي تعاني من الفقدان فانه لابد من ظهور مشكلة تعدد العلاقة الخطية لخرق إحد فروض التحليل ، لذلك وجب معالجتها بطريقة انحـدار المركبات الرئيسة بعد الكشف عنها بمقياس تضخم تباين المعاملات (VIF) .

نبذة تاريخية

في حالة ظهور مشكلة فقدان قسم من البيانات للعينة غالباً ما يلجأ سابقاً الى حـذف المشاهدات التي تتضمن قيماً مفقودة من العينة المدروسة والاعتماد على العينة المتبقيـة فـي التحليل ، وقد استخدم هذا الأسلوب في عام (1932) من قبل Yates الـذي يعـد اول مـن استخدم هذا الأسلوب لمعالجة البيانات غير التامة.

وفي عام (1939) جاء Bartlett بأسلوب حسابي يكافئ الاسلوب الـذي جـاء بـه Yates معتمداً على حساب التباين والتباين المشترك واستخدام نتائجه لحساب معـالم نمـوذج الانحدار في حالة كون بعض قيم المتغير المستجيب $Y's$ مفقودة وقد تم تطبيق الاسلوب نفسه في عام (1952) من قبل Tacher ، في عام (1966) اسـتخدم Affif & Elashoff نفـس الأسلوب السابق ولكن في حالة كون المتغير المستجيب $Y's$ يتعرض لفقدان جزء من قيمـه. واستمر البحث في هذا الموضوع لعام (1967) وعام (1969) لنفس العالم أعـلاه واقتـصر على نماذج الانحدار الخطي البسيط .

وفي عام (1986) قام Glynn & Laird باستخدام أسلوب التحليل للبيانات التي يعتمد فقدانها على القيم المفقودة نفسها أي (Not MAR) وحصل على نتائج معقولة قياساً بطرائـق اخرى .

وقـدمت عـدة تقييمـات لهـذه الطريقـة أبرزهـا مـا جـاء بـه عـام (1987) (Litter & Rubin).

كما استخدم (الطائي ، 1998) البيانات غير التامة وتحليلهـا احـصائياً فـي نمـاذج الانحدار المتعدد.

آلية البيانات غير التامة (المفقودة)

وباختلاف البيانات غير التامة تختلف الطرائق الإحصائية لتحليل هذه البيانات وايـضاً تختلف الآلية التي تؤدي الى فقدان البيانات وان فهم هذه الآلية وتحديد طبيعتها يساعد كثيـراً في اختيار الطريقة المناسبة للتحليل بل يعد المدخل لتشخيص الطريقة التي تقترب نتائجها من الامثلية للبيانات المدروسة . وتكون علاقة فقدان البيانات لمتغير معين بقيم المتغير نفـسه او لقيم المتغيرات الأخرى وكما ياتي :

أ- ان فقدان X_j يكون مستقلاً عن قيم المتغيرات الأخرى وعن القيم المفقودة نفسها .
ب- ان فقدان قيم X_j يعتمد على القيمة نفسها .
ج- اعتماد القيم المفقودة X_j عن قيم المتغيرات الأخرى في العينة.

اذا كان سبب الفقدان مستقلاً عن القيمة المفقـودة وعـن قـيم المتغيـرات الأخـرى نفسها في العينة عندها يمكن القول ان البيانـات تفقـد تمامـاً بـشكل عـشوائي (MCAR) Miss Completely at Random . اما اذا كان سبب الفقدان له علاقـة بقـيم المتغيـرات الاخرى فقط ومستقلا عن القيمة المفقودة ففي مثل هذه يكون فقدان البيانات بـشكل عـشوائي Missing at Random (MAR) ، واحياناً يكون سبب الفقدان ناتجا عن القيمـة المفقـودة نفسها ومستقل عن قيم المتغيرات الأخرى . فالبيانات هنـا لا تفقـد بـشكل عـشوائي (not MAR) عند تحليل هذا النوع من البيانات ويجب اخذ آلية الفقدان بنظر الاعتبار . امـا فـي حالة (MCAR) و (MAR) يمكن ان يهمل التوزيع اليه . (الطائي، 1998).

ويمكن توضيح كل ما سبق بالمخطط الاتي :

آلية الفقدان

بشكل غير عشوائي	بشكل عشوائي	بشكل عشوائي تام
Not MAR	**MAR**	**MCAR**
ويكون سبب الفقدان ناتجا عن القيمة المفقودة نفسها ومستقل	**يكون سبب الفقدان له علاقة بقيم المتغير الأخر فقط ومستقل**	**اذا كان سبب الفقدان مستقلاً عن القيمة المفقودة وعن قيم**

عن قيم المتغيرات الأخرى	عن القيم المفقودة	المتغير الأخر
لا يهمل التوزيع اليه	يهمل التوزيع اليه	

طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة

Lest Square (LS) on Imputed Data
(Little & Rubin , 1987)

يوجد العديد من الطرائق في إيجاد البيانات غير التامة التي فقدت تماماً بشكل عشوائي (MCAR) او التي فقدت بشكل عشوائي (MAR) او بشكل غير عـشوائي (not MAR) ، لكن منها ما تبتعد عن استخدام المتغيرات التي تتضمن قيماً مفقودة ومنها ما تستخدم البيانـات المتاحة وفي بعض التطبيقات تتخلى عن جزءً منها كما في حساب التباينات المشتركة ومعامل الارتباط . اما طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة باسـلوبيها الأول التعويض بالمتوسط الشرطي Conditional mean Imputation والاخر اسلوب التعـويض بالمتوسط غير الشرطي Unconditional mean Imputation اذ تعتبـر هـذه الطريقـة بأسلوبيها بانها عامة ومرنة في معالجة مشكلة البيانات غير التامة حيث تعتمد هـذه الطريقـة على اشغال مواقع القيم المفقودة بقيم تقديرية تختلف بأسلوب تقديرها ولكن الهدف هو إكمـال العينة واستخدامها في التحليل النهائي ، اذ تعتبر هذه الطريقة ايجابية من جهـة لانهـا تـوفر بيانات بديلة عن البيانات المفقودة بشكل حسابي ولكنها تعتبر سلبية لانها تـستخدم الطرائـق التقليدية في تحليل بيانات العينة المكملة التي تضم بيانات حقيقية (مشاهدة) وبيانـات تقديريـة (محسوبة) ، والأخيرة قد تكون عرضة للتحيز عند تقديرها وهذا يرجع الى الأسلوب المعتمـد في التقدير ومدى امكانية الحصول على تقديرات كفوءة للبيانات المفقـودة وسـوف نـستخدم أسلوب التعويض غير الشرطي للبيانات غير التامة التي فقدت تماماً بشكل عشوائي في هـذا البحث.

التعويض بالمتوسط غير الشرطي Unconditional mean Imputation

يتم التعويض عن القيم المفقودة لاي متغير تحت الدراسة بمتوسط قيم المشاهدة له

$$\tilde{X}_j = \sum_{n_j} X_{obs} / n_j \quad \text{.....(1)}$$

n_j : يمثل عدد القيم المشاهدة فعلاً للمتغير X_j

وفي حالة إحلال $\tilde{X}_j$ محل القيم المفقودة X_{mis} لـ X_j فان المتوسط العـام للمتغيـر $\overline{X}_j$ يكون مساوياً لـ $\tilde{X}_j$ وتباين X_j للعينة المكملة هو حيث

$$(n_j - 1)S_{jj}/(n-1) \ldots\ldots\ldots\ldots (2)$$

S_{jj} : يمثل التباين المقدر من القيم المتاحة حسب الصيغة

$$\tilde{S}_{jj}^2 = \sum_{i=1}^{n_j} (X_{ij} - \tilde{X}_j)^2 / n_j \quad \ldots\ldots (3)$$

$$\tilde{S}_{kk}^2 = \sum_{i=1}^{n_k} (X_{ik} - \tilde{X}_k)^2 / n_k \quad \ldots\ldots (4)$$

$$\tilde{X}_j = \sum_{n_j} X_{ij} / n_j \quad \ldots\ldots\ldots\ldots (5)$$

$$\tilde{X}_k = \sum_{n_k} X_{ik} / n_k \quad \ldots\ldots\ldots\ldots (6)$$

وتحت فرض ان البيانات غير التامة بشكل عشوائي تماماً (MCAR) فـان S_{jj} هـو تقـدير للتباين الحقيقي كذلك فان تباين العينة متحيز بالمعامل $(n_j\text{-}1)(n\text{-}1)$ لان التعويض عن القـيم المفقودة تم باستخدام مركز التوزيع المتمثل بالوسط الحسابي وهـذا ينـسحب علـى التبـاين المشترك للمتغيرين X_j , X_k اللذين يتضمنان بيانات حقيقية مشاهدة وبيانات تقديرية محسوبة فيكون

$$\left[(n_{jk} - 1)/(n-1)\right]\tilde{S}_{jk} \quad \ldots\ldots\ldots (7)$$

حيث $\tilde{S}_{jk}$ محسوبة وفق الصيغة

$$\tilde{S}_{jk} = \sum_{i=1}^{n_{jk}} (X_{ij} - \tilde{X}_j)(X_{ik} - \tilde{X}_k)/(n_{jk} - 1) \quad \ldots\ldots (8)$$

وهو تقدير متسق للتباين المشترك المحسوب على وفق هذه الطريقة ومتحيزاً بالمعامل

$$(n_{jk} - 1)/(n-1) \quad \ldots\ldots (9)$$

وهـذا فـضلا عـن إمكانيـة ان تكـون مـصفوفة التبـاين والتبـاين المـشترك (Positive semi definite) .

انحدار المركبات الرئيسة Principal Component Regression

(الشكرجي ، 2005) ، (المشهداني ، 1994)

في حالة مصادفة معادلة الانحدار الخطي المتعدد مشكلة تعدد العلاقة الخطية لوجـود ارتباطات بين المتغيرات التنبؤية إذ أن التقديرات التي تنتج لدينا تكون متأثرة بالعلاقات بـين

المتغيرات وليس نتيجة العلاقة بين المتغير المستجيب والمتغيرات التوضيحية ، وذلك يعني خرق احد فروض التحليل التي تنص على أن أعمدة وصفوف المتغيرات التوضيحية X يجب ان تكون مستقلة خطياً مع بعضها البعض إذ توجد العديد من الطرائق المتحيزة للتخلص من مشكلة تعدد العلاقة الخطية بعد الكشف عنها بإحدى طرائق الكشف كمقياس تضخم تباينات المعاملات (VIF) التي ستذكر لاحقاً . ومن ثم استخدام طريقة المركبات الرئيسة المتحيزة في التحليل .

طريقة انحدار المركبات الرئيسة

تعتمد طريقة تحليل المركبات الرئيسة على اسلوب تحويل المتغيرات التوضيحية الأصلية الى متغيرات جديدة غير مرتبطة . هذه المتغيرات الجديدة تسمى بالمركبات الرئيسة، اذ ان كل مكون رئيسي هو عبارة عن تركيبة خطية في المتغيرات المستقلة الأصلية .

ويتم تحويل المتغيرات المستقلة الى المكونات الرئيسة بالشكل الآتي :

$$\overset{*}{Y} = \beta_0 \underline{I} + XVV'\underline{\beta} + \underline{U} \ldots\ldots\ldots (10)$$

حيث V عبارة عن مصفوفة المتجهات المميزة لمصفوفة الارتباط بين المتغيرات التوضيحية فاذا عوضنا عن XV بكمية ثابتة PC مصفوفة ذات بعد ($n*p$) أعمدتها عبارة عن معاملات انحدار النموذج المحور الذي يأخذ الشكل الآتي :

$$\overset{*}{Y} = \beta_0 \underline{I} + PC\overset{*}{\underline{\alpha}} + \underline{U} \ldots\ldots\ldots (11)$$

كما ويعرف مقدار التحيز للمكونات الرئيسة كالآتي :

$$b(\beta^{o}_{PC})E(\hat{\beta}_{PC}) - \beta = VV'\underline{\beta} \ldots\ldots\ldots (12)$$

مقياس تضخم تباينات المعاملات Variance Inflation Factors (VIF)

(سعيد ، 1996)

يعد هذا المقياس سبباً كافياً لإهمال المتغير X_j من التحليل او استخدام طريقة اخرى بديلا عن المربعات الصغرى في التقدير .

وقد اكد بعض الباحثين ان مقدار المقياس $[a_{JJ} = VIF \geq 10]$ فانه دليل على خرق فروض التحليل لوجود علاقة بين المتغيرات التوضيحية، ويعتمد هذا المقياس على فحص عناصر القطر الرئيسي للمصفوفة $(X'X)^{-1}$ حيث ان

$$a_{JJ} = (1 - R^2_j) \ldots\ldots\ldots (13)$$

وان R_j^2 عبارة عن معامل التحديد لانحدار X_j عن بقية المتغيرات التوضيحية ، توضح المعادلة (13) بانه لوجود تداخل خطي بين المتغيرات التوضيحية فان R_j^2 ستكون قريبة او مساوية للواحد وبذلك تكون a_{JJ} كبيرة الحجم . اما اذا كانت X_j مستقلة عن بقية المتغيرات التوضيحية الأخرى $R_j^2 = 0$ وبذلك تكون $a_{JJ}=1$.

جمع البيانات

أخذت البيانات من OPEC Annual Statistical Bulletin (2004) والمعتمدة في أطروحة دكتوراه* . والبيانات تمثل انتاج النفط واستهلاك المنتجات المكررة في اقاليم معينة من العالم (الف برميل يومياً) للفترة من (1970-2004) لـ 35 مشاهدة ولسبعة متغيرات هي :

X_1 انتاج أوربا وأمريكا من النفط

X_2 استهلاك أوربا وأمريكا من المنتجات المكررة

X_3 العجز الأوربي والأمريكي بين إنتاج النفط واستهلاك المنتجات المكررة

X_4 إنتاج آسيا والهادي من النفط

X_5 إنتاج آسيا والهادي من المنتجات المكررة

X_6 عجز آسيا والهادي بين إنتاج النفط واستهلاك المنتجات المكررة

Y تمثل عدد سنوات الدراسة للفترة من (1970-2004)

وآلية الفقدان للبيانات غير التامة بشكل عشوائي تماماً (MCAR).

التحليل الإحصائي

1- طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة

أخذت احد المتغيرات وليكن X_4 الذي يعاني من نقص في مشاهداته . إذ تم إيجاد المشاهدة المفقودة بالتعويض بالمتوسط غير الشرطي حسب الصيغة (1) والتباين حسب الصيغة (3) . وكانت قيمة المتوسط والتباين على التوالي عندما احد المشاهدات المفقودة للمتغير X_4 أي $n=34$ هي 5402.08 ، اما التباين 3224733 وعند التعويض عن القيمة المفقودة بالمتوسط 5402.08 فان المتوسط المقدر هو 5402.08 والتباين المقدر 3129888 أي عندما $n=35$.

* اطروحة دكتوراه مقدمة من عمار محمد سلو احمد ، كلية الادارة والاقتصاد ، جامعة الموصل (2006) الموسومة " السياسة الانتاجية والسعرية للملكة العربية السعودية ".

وللمقارنة مع المتوسط والتباين للمتغير X_4 قبل فقدان احدى قيمه أي عندما n=35 الحقيقية قبل الفقدان فان المتوسط يساوي 5341.96 والتباين يساوي 3256378 من ذلك يمكن ان نستنتج ان المتوسط والتباين المقدر قريب من المتوسط والتباين الحقيقي .

2- الكشف عن مشكلة تعدد العلاقة الخطية بمقياس معامل تضخم التباين (VIF)

مما تقدم وبعد ايجاد القيمة المفقودة لاي من المتغيرات لابد من ان يكون هنالك وجود مشكلة تعدد علاقة خطية بين المتغيرات وهذا يعني خرقا لإحد فروض التحليل ويمكن ملاحظة ذلك من خلال معامل تضخم التباينات (VIF) بين المتغيرات التوضيحية بالنسبة لبيانات انتاج النفط واستهلاك المنتجات المكررة في أقاليم معينة من العالم . ان هذا المقياس يعتمد اذ كان (VIF) اكبر من (10) فانه دليل على ان المتغير X_j يعاني من وجود مشكلة تعدد العلاقة الخطية ، اذ يلاحظ VIF للمتغيرات التوضيحية لبيانات إنتاج النفط كما في الجدول (1) .

الجدول (1)

يمثل مقياس تضخيم التباين VIF لمتغيرات إنتاج النفط

X_j	VIF
X_1	3.4
X_2	4.3
X_3	1.4
X_4	7.5
X_5	72.3*
X_6	49.6*

J=1,2,3,4,5,6

* تعني ان هذه المتغيرات تعاني من تضخم تباين معاملاتها

وبما ان قيم تضخيم تباين المعاملات VIF في الجدول (1) للمتغيرات التوضيحية لكل من X_6,X_5 اكبر من (10) فان هذا يعني ان هذه المتغيرات تعاني من مشكلة تعدد العلاقة الخطية بين المتغيرات التوضيحية أي وجود علاقة او ارتباط بين المتغيرات التوضيحية .

3- طريقة المركبات الرئيسة

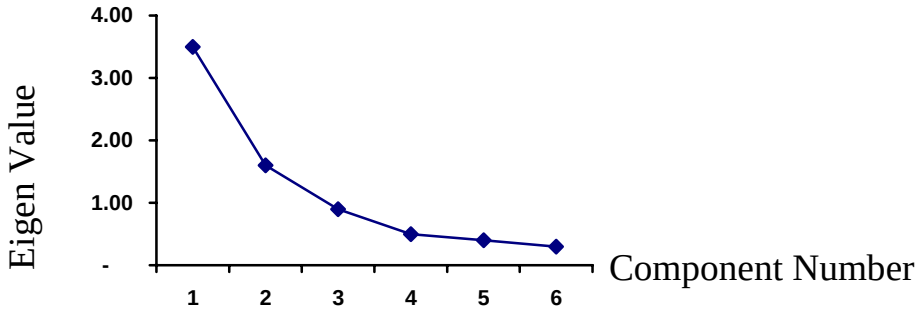
بعد تحليل مصفوفة الارتباط البسيط تمكنا من تمييز مكونين رئيسيين فسرا ان نسبة (84%) من اجمالي التباين الكلي . كما تم تمييز المتغيرات المؤثرة في كل مكون رئيس من خلال مصفوفة المتجهات المميزة كما ياتي : عكس المكون الرئيسي الأول أهمية كبيرة بتفسير ما نسبته (57.8%) من التباين الكلي واشتمل هذا المكون على متغيرين ذوي تأثير معنوي وهما المتغير (X_5) والمتغير (X_6) كما ظهر المتغيران (X_1) و (X_3) ذوي تأثير معنوي في المكون الرئيس الثاني والذي فسر ما نسبته (26.5%) من التباين الكلي فضلاً عن ذلك لوحظ عدم أهمية المتغيرين (X2) و (X4) في كلا المكونين الرئيسين ويمكن توضيح ما سبق في الجدول (2) الجذور المميزة للمتغيرات . اذ يوضح هذا الجدول أي الجذور المميزة الأكبر من الواحد التي على أساسها تم اختبارهما ثم استخدام المكونات الرئيسة .

الجدول (2)

الجذور المميزة للمتغيرات

الجذور المميزة	قيمة الجذر المميز
λ_1	3.4654**
λ_2	1.5896**
λ_3	0.6137
λ_4	0.1933
λ_5	0.1298
λ_6	0.0081

** يمثل الجذور المميزة والتي قيمتها اكبر من الواحد .



الشكل (1) عدد الجذور المميزة التي اختيرت

يوضح الشكل (1) الجذور المميزة والتي قيمها اكبر من الواحد التي اعتمـدت فـي طريقـة المركبات الرئيسة .

وبعد اعتماد طريقة المركبات الرئيسة تم الكشف مرة اخرى عن مشكلة التعدد الخطي بطريقة (VIF) فلوحظ أن المتغيرات التي كانت تعاني من تضخم في تباين معاملاتها وهـي X_5 ، X_6 ، أصبح معامل تضخم التباين على التوالي للمتغيرات (1.0,1.0) وهذا يعني زوال مشكلة تعدد العلاقة الخطية للمتغيرات قيد الدراسة لاعتمادنا طريقة المركبات الرئيسة للبيانات التي كانت غير تامة ومعالجتها وتحويلها الى بيانات تامة وتقديرها بطريقة المركبات الرئيسة.

الاستنتاجات

يمكن توضيح أهم الاستنتاجات بما يلي :

1- ان حل مشكلة البيانات المفقودة تغني عن البحث وبذل المزيد مـن الوقـت والجهـد والتكلفة في إيجاد طرائق او بيانات أخرى لحل مشكلة الفقدان في البيانات قيد الدراسة او البحث .
2- ان التعرف على آلية الفقدان للبيانات يعتبر المفتاح لتشخيص أسلوب التحليل المعتمـد للظاهرة . كما وان الاستلال حول البيانات غير التامة قد يكون توزيعـاً مجهـولاً او توزيعاً طبيعياً للمتعدد المتغيرات خاصة اذا كان حجم العينة كبيراً .
3- كما وان البيانات غير التامة بعد معالجتها بالأسلوب المناسب لنوع الفقدان يمكـن ان تظهر مشكلة التعدد الخطي لخرق إحد فروض التحليل كما ويمكـن الكـشف عنهـا ومعالجتها بإحدى طرائق التقدير المتحيزة كما ورد في بحثنا وتقديرها بطريقة انحدار المركبات الرئيسة.
4- يمكن تقدير مقدار التحيز للبيانات غير التامة بعد معالجتها مهما كانت آليـة الفقـدان واعتمادها كبيانات تامة في التطبيقات الإحصائية كافة.

المصادر

أ- المصادر العربية

- البلداوي ، عبد الحميد عبد المجيد (2004) : " الأسـاليب الإحـصائية التطبيقيـة " ، دار الشرق الأوسط ، التوزيع الأردن .

- الشكرجي ، ذنون يونس ذنون (2005) : "استخدام مصفوفتي Q-Mode – R-Mode في التحليل العاملي" ، رسالة ماجستير ، كلية الادارة والاقتصاد ، جامعة الموصل .

- الطائي ، خالد ضاري عباس (1998) : "التحليل الإحصائي للبيانات غير التامة في نماذج الانحدار المتعدد" ، أطروحة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة الموصل .

- المشهداني ، ايمان محمد عبد الله (1994) : "استخدام المركبات الرئيـسة فـي تـشخيص ومعالجة مشكلة التعدد الخطي مع تطبيق عملي لبعض الظواهر الاقتـصادية" ، رسـالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة بغداد .

- سعيد ، هيفاء عبد الجواد (1996) : "طرق التعرف على تعدد العلاقـة الخطيـة وكيفيـة معالجتها بطرائق التقدير المتحيزة " ،رسالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة الموصل.

- سلو ، عمار محمد (2006) : "السياسة الإنتاجية والسعرية للملكة العربية الـسعودية فـي مجال النفط الخام" ، رسالة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة الموصل .

- كاظم ، أموري هادي (1998) : "طرق القياس الاقتصادي" ، جامعة بغداد ، مطبعة التعليم العالي ، الطبعة الأولى.

- كمال ، غفران إسماعيل (1990) : "دراسة تقويمية لطرق التقـديرات المتحيـزة لمعامـل الانحدار المتعدد عند مخالفة بعض الفرضيات باستخدام أسـلوب المحاكـاة" ، رسـالة ماجستير في الإحصاء ، كلية الإدارة والاقتصاد ، جامعة بغداد.-

ب- المصادر الأجنبية

- Gourieroux, G. and Mont Fort (1981) " on the problem of Missing Data an Linear Models" Review of Economic Studies, XL VIII, P: 579-586.

- Little, and Rubin, (1987) "Statistical Analysis with Missing Data", New York: John-Wiley, P: 88,125-134.

- Little R.J.A. (1988) "Robust Estimation of the Mean Covariance Matrix from Data with Missing Values" Applied Statistics, P: 37,33-29.

- Theil, H. (1971) "Principal of Econometrics" John Wiley & Sons. Inc. P: 90-94