أسلوب تحليل الانحدار الخطي المتعدد في دراسة أهم المتغيرات الاقتصادية والاجتماعية والديمغرافية المؤثرة في معدل الولادات الكلية

(دراسة تطبيقية من واقع بيانات تقرير التنمية البشرية لعام 2006 لـ 177 دولة)

الدكتور فريد خليل الجاعوني

قسم الإحصاء

كلية الاقتصاد

جامعة دمشق

الملخص

الهدف الأول لهذه الورقة هو دراسة أهم العوامل الاقتصادية والاجتماعية والديمغرافية المؤثرة في معدل الولادات الكلية Total Fertility Rate (T..F.R) للمرأة طوال مدة الإخصاب للفئة العمرية (15-49) وذلك على مستوى العالم تقريباً. شملت الدراسة بيانات (177) دولة مضمنة في تقرير التنمية البشرية لعام 2006، وقد طبقنا أسلوب الانحدار الخطي المتعدد في اختيار أهم العوامل المؤثرة (المتغيرات المستقلة) (Independents) والمرتبطة معنوياً مع متغير الدراسة (T.F.R) المتغير التابع (dependent variable). وفي سياق البحث ركزنا على شروط وفرضيات البيانات (Data) عند استخدام هذا الأسلوب وكان هذا هو هدفنا الثاني من البحث وقد توصلنا إلى أهم النتائج الآتية:

بدايةً تم تقسيم الدول إلى أربع مجموعات حسب قيمة مؤشر دليل التنمية البشرية *(HDI) على النحو الآتي: مجموعة (1) : مجموعة الدول المتقدمة اقتصادياً ومعظمها من الدول الأوروبية وأمريكا الشمالية حيث تقع قيمة مؤشر دليل التنمية فيها بين (1- 0.801) ، مجموعة رقم (2) : حيث تقع قيمة الدليل بين (0.8 - 0.601) مجموعة (3) : قيمة الدليل (0.6 - 0.401) مجموعة رقم (4) : مجموعة الدول الفقيرة تيمور الشرقية ـ الكونغو ـ مدغشقر ـ اليمن ـ زامبيا ـ ملاوي ـ سيراليون ـ هوندورا ـ ارتيريا ـ تنزانيا ـ أثيوبيا ـ مالي ـ النيجر ـ الكونغو وغيرها حيث تقع قيمة الدليل (0.4-0.201) .

قد خلصنا بنتيجة التحليل إلى أهم النتائج الآتية :

1- معنوية متغير نسبة السكان للفئة العمرية أقل من 15 عاماً في مجموعات الدول كلها.

2- إن متغير نسبة السكان للفئة العمرية أقل من 15 عاماً هو المتغير الوحيد المؤثر في مجموعة الدول (1) ومجموعة الدول (4) وعدم معنوية جميع المتغيرات الأخرى على هاتين المجموعتين.

3- معنوية متغير التعليم وارتباطه بعلاقة عكسية مع (T.F.R) في مجموعات الدول (3)+(2) وعدم معنويته لمجموعات الدول (4)+ (1)، وتفسيرنا لذلك أن مجموعة الدول المتقدمة (1) قد تجاوزت طور تأثير متغير التعليم في (T.F.R) في حين إن المجموعات (3) + (2) واقعة في طور التأثير، والمجموعة (4) لم تدخل بعد في طور تأثير متغير التعليم في (T.F.R) نظراً لتدني نسبة التعليم في هذه المجموعة من الدول .

4- إن المتغيرات المعنوية والمؤثرة على مجموعة الدول (2) هما التعليم ونسبة السكان للفئة العمرية أقل من 15 عاماً، وقد دخل متغير التعليم كمؤثر جديد بالمقارنة مع مجموعة الدول (1) .

5- أما بالنسبة للمجموعة رقم (3) فازداد فيها عدد المتغيرات المؤثرة ليصل إلى أربعة متغيرات (التعليم ـ نسبة السكان للفئة العمرية أقل من 15 عاماً ـ الإنفاق الحكومي على الصحة ـ معدل وفيات الأمهات لكل 100000 ولادة حية) .

6- المجموعة الأخيرة (4) مجموعة الدول الفقيرة ، كان المتغير الطبيعي الوحيد المؤثر هو نسبة السكان للفئة العمرية أقل من 15 عاماً، على اعتبار أن جميع المتغيرات المؤثرة الأخرى هي في أدنى حدودها في هذه الدول .

* HDI : Human Development index

مقدمة:

يعد أسلوب تحليل الانحدار الخطي المتعدد Method of multiple linear regression analysis من أهم الأدوات البحثية في شتى المجالات ويهدف إلى دراسة العلاقة البينية بين متغير متأثر أو تابع (y) dependent variable وأكثر من متغير مؤثر مستقل (X_i) independent variables بواسطة معادلة رياضية تسمى نموذج الانحدار الخطي المتعدد multiple linear regression model، ومن الضروري التفريق بين نوعين من العلاقات التي تربط المتغير التابع بالمتغيرات المستقلة ، النوع الأول يسمى بالعلاقة الدالية والثاني يسمى بالعلاقة الإحصائية[1] .

يمكن تمثيل العلاقة الدالية بالعلاقة الرياضية $y = f(X)$ وهذا يعني أن تغير قيم المتغير التابع يعتمد فقط على تباين المتغيرات المستقلة في النموذج ، وذلك لعدم إمكانية تضمين النموذج جميع المتغيرات المؤثرة في الظاهرة المدروسة ، إما بسبب عدم توافر البيانات عن بعض المتغيرات من جهة، أو لعدم أهمية بعضها الآخر من جهة أخرى ، وهذا ما لا يمكن حدوثه في الدراسات الاجتماعية والاقتصادية، ومن هنا كان لا بد من تضمين النموذج فضلاً عن المتغيرات المستقلة (X_i) حداً يمثل متغير الخطأ العشوائي ($U_i = (y - \hat{y})$ Residuals والذي يقيس الجزء من التغير في المتغير التابع والذي سببه المتغيرات المستقلة غير المدرجة في نموذج الانحدار ، ويمكن تفسير إضافة المتغير العشوائي (U_i) إلى النموذج بالأسباب الآتية:

1- عدم احتواء النموذج على جميع المتغيرات المستقلة المؤثرة في الظاهرة .

2- عشوائية الاستجابات الأساسية.

وعلى الرغم من أهمية هذا الأسلوب في جميع الدراسات البحثية، إلا أنه يجب توخي الحذر الشديد عند تطبيقه، لأنه غالباً ما يساء فهمه واستخدامه بالشكل الصحيح بسهولة، ويمكن أن يعطي عندها نتائج غير صحيحة ومضللة ، خصوصاً إذا لم تراع الشروط أو الافتراضات على البيانات عند تطبيقه ويصبح غير ذي مصداقية في النتائج والتفسير[2] .

ومن المفيد قبل الشروع بحل المسائل الكبيرة باستخدام هذا الأسلوب وضع خطة للمشروع بعيدة المدى (as far as possible) ثم تحديد الأهداف وفحص النتائج في كل خطوة في أثناء سير العمل .

([1]) أبو سدرة فتحي صالح ، الكيخيا نجاة رشيد (1999) : الإحصاء والاقتصاد القياسي ، منشورات المركز القومي للدراسات والبحوث العلمية ، دار الكتب الوطنية ، بنغازي ليبيا ص 47-45

([2]) Draper,N.R.and smith , H.(1981) : Applied Regression anlysis , new Yourk , P.413

فروض البحث:

لخصت فروض البحث في الإجابة عن التساؤلات أدناه حول مدى معنوية تأثير المتغيرات المستقلة المضمنة في نموذج الانحدار على المتغير التابع (T.F.R)

- معنوية الأجل المتوقع عند الولادة في تأثيره في (T.F.R) ؟
- معنوية متغير نسبة المتعلمين بين السكان بعمر 15 ≥ عاماً على T.F.R ؟
- معنوية متغير نسبة القيد المدرسي في الصفوف الابتدائية والإعدادية والثانوية ؟
- معنوية متغير نصيب الفرد من الناتج القومي الإجمالي على T.F.R ؟
- معنوية متغير نسبة السكان الحضر في مجموع السكان ؟
- معنوية متغير نسبة السكان بعمر أقل من 15 عاماً في مجموع السكان ؟
- معنوية متغير وفيات الأطفال بعمر أقل من 5 سنوات لكل 1000 ولادة حية ؟
- معنوية متغير معدل وفيات الأمهات لكل 100000 ولادة حية ؟
- معنوية متغير الإنفاق الحكومي على التعليم كنسبة مئوية من الناتج المحلي الإجمالي ؟
- معنوية الإنفاق الحكومي على الصحة كنسبة مئوية من الناتج القومي الإجمالي ؟

الدراسات السابقة:

باعتقادنا أن هذا الموضوع هو الأول من الناحية التطبيقية ولم يسبق أن تمت معالجة هذا الموضوع في دراسات سابقة.

مشكلة البحث:

إن بناء نموذج الانحدار وتقدير معالمه دون دراية أو دون التحقق من الشروط أو الافتراضات الواجب توافرها عند تطبيقه يقود إلى نتائج غير صحيحة ، ومن الضروري الإشارة إلى أن متغيرات أو بيانات أي دراسة غالباً أو في معظمها تعاني من مشاكل" problems with messy data" كمشكلة الارتباط الذاتي بين الأخطاء العشوائية autocorrelation ومشكلة التعدد الخطي multicollinearity problemوهذا ما يتطلب من الباحث معرفة بتشخيص المرض وأسلوب معالجته .

هدف البحث:

إن هدفنا من هذا البحث هو هدف مزدوج : الأول هو هدف نظري (تطبيق النظرية) من خلال مناقشة الشروط المطلوبة من البيانات عند تطبيق هذا الأسلوب وكيفية معالجة أو مواجهة جميع المشاكل المحتمل مصادفتها في أثناء سير العمل؛ وذلك باستخدام بعض التقنيات الإحصائية السهلة والبسيطة ، والثاني هو تطبيق النظرية على حالة عملية لها أهمية استثنائية وهي مشكلة النمو السكاني المرتفع لدى العديد من الدول وخصوصاً الفقيرة، وتأثير هذه المعدلات المرتفعة للنمو السكاني في معدلات النمو الاقتصادي من خلال التعرف على أهم العوامل المعنوية المؤثرة فيها واختلافها بين مجموعات الدول ، ووضع تفسيرات لأسباب هذا الاختلاف .

متغيرات الدراسة[1] Variables study

نظراً لكثرة المتغيرات التي تجاوز عددها 250 متغيراً لـ 177 دولة ، فقد اقتصرت دراستنا على عدد منها ، على ما هو وثيق الصلة بموضوع بحثنا ، ومن الضروري الإشارة إلى أن عدد الدول ليس ثابتاً بالنسبة لجميع متغيرات الدراسة نظراً لعدم توافر البيانات لبعض المتغيرات في بعض الدول، وأهم المتغيرات التي تم إدراجها بالدراسة هي على النحو :

(χ_{12}) : متوسط العمر المتوقع عند الولادة بالسنوات ، ويحسب من جداول الحياة life table وفق العلاقة $ex=Tx/\ell_x$ وتشير Tx : إلى مجموع السنوات التي سوف يعيشها الجيل الفرضي ℓ_x ابتداء من العمر (0) وحتى نهاية العمر (N-1) ، وتشير ℓ_x إلى أساس الجدول وعادة ما يكون 10000 أو 100000 والجيل عادة يتكون من عدد مواليد أساس، وتشير N إلى أعلى عمر لأي شخص باق على قيد الحياة .

(χ_{13}) :نسبة المتعلمين الذين يعرفون القراءة والكتابة بين السكان بعمر 15 عاماً فما فوق.

(χ_{14}) : نسبة القيد المدرسي في الصفوف الابتدائية والإعدادية والثانوية.

(χ_{15}) : نصيب الفرد من الناتج القومي الإجمالي مقدراً بالدولار الأمريكي.

(χ_{23}) : نسبة السكان الحضر إلى مجموع السكان.

(χ_{24}) : نسبة السكان بعمر أقل من 15 عاماً .

1 وقد تم الحصول على البيانات عن طريق شبكة network من موقع UNDP على العنوان :
www.undp.org/hadr2006/statistics/indicators

(χ_{245}) : معدل وفيات الأطفال بعمر أقل من 5 سنوات لكل 1000 ولادة حية

(χ_{26}) : معدل الخصوبة الكلية لكل امرأة طوال مدة الإخصاب عندها وهو المتغير التابع في دراستنا.

(χ_{249}) : معدل وفيات الأمهات لكل 100000 ولادة حية.

(χ_{31}) : نسبة الإنفاق الحكومي على التعليم من الناتج المحلي الإجمالي.

(χ_{61}) : نسبة الإنفاق الحكومي على الصحة من الناتج المحلي الإجمالي.

(χ_{62}) : نسبة الإنفاق الحكومي الخاص على الصحة من الناتج المحلي الإجمالي .

الإطار العملي للدراسة:

كان من الصعب التوصل إلى أفضل نموذج لانحدار قبل ظهور الرزم الإحصائية الجاهزة ، لما تتطلبه عمليات الحسابات من وقت وجهد ، أما الآن ومع وجود هذه البرمجيات فقد أصبح من اليسير التوصل إلى أفضل نموذج لانحدار من بين العديد من النماذج التي يمكن بناؤها والتي تصف العلاقة المتوسطة بين المتغيرات المستقلة والمتغير التابع ، وبالاعتماد على الحزمة البرمجية (Spss) statistical package for social sciencees تم التوصل إلى نموذج لانحدار متعدد لدراستنا

حيث المتغير التابع (y) يمثل (T.F.R) ورمزه X_{26} على المتغيرات المستقلة : X_{12} , X_{13} , X_{14} , X_{15} , X_{23} ,X_{245} , X_{24} ,X_{62} ,X_{61} ,X_{31} ,X_{249}

اتبعنا في البداية الطريقة الاعتيادية (Enter) ، أي إدخال المتغيرات كلّها في النموذج ، علماً بأن الحزمة توفر خيارات أخرى لاختيار أفضل نموذج لانحدار متعدد وسنتعامل مع أحدها فيما بعد وهي طريقة (Stepwise)، وهي الطريقة المقترحة من قبل (Efroymson 1960) والتي تقترح إدخال المتغيرات واحداً تلو الآخر على عكس الطريقة (enter) مع استبعاد المتغيرات التي تصبح غير مؤثرة بوجود بقية المتغيرات، وبالعودة إلى الطريقة الاعتيادية حصلنا على النتائج الملخصة في الجداول أدناه:

جدول رقم 1

Model Summary(c,d)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.995(b)	.991	.989	.3524	2.126

جدول رقم 2

ANOVA(c,d)

Sig.	F	Mean Square	df	Sum of Squares		Model
.000(a)	616.083	76.498	11	841.483	Regression	1
		.124	64	7.947	Residual	
			75	849.430(b)	Total	

الجدول رقم 3

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	X12	7.290E-03	.005	.145	1.417	.161	.812	.174	.017	.014	71.511
	X13	-.008	.004	-.200	-1.754	.084	.810	-.214	-.021	.011	88.733
	X14	-.007	.006	-.148	-1.212	.230	.809	-.150	-.015	.010	102.327
	X15	2.900E-05	.000	.093	3.218	.002	.544	.373	.039	.174	5.760
	X23	-.004	.003	-.078	-1.652	.103	.731	-.202	-.020	.066	15.252
	X245	3.531E-03	.002	.085	1.951	.055	.885	.237	.024	.077	13.018
	X249	9.197E-04	.000	.120	3.787	.000	.802	.428	.046	.146	6.861
	X31	1.779E-02	.023	.027	.768	.445	.802	.096	.009	.121	8.283
	X61	8.885E-03	.033	.009	.268	.790	.692	.033	.003	.117	8.539
	X62	7.951E-03	.036	.006	.218	.828	.794	.027	.003	.170	5.897
	X24	.102	.006	.976	17.117	.000	.980	.906	.207	.045	22.246

a. Dependent Variable: X26

b. Linear Regression through the Origin

نلاحظ من خلال جدول تحليل التباين (Anova) جدول رقم (2) رفض الفرضية المختبرة $H_0 : B_1 = B_2 = ...B_q = 0$

حيث إن قيمة $F_{cal} = 616.083$ من أجل (d.F(11,64 وإن قيمة p-value=0 المرافقة لإحصاءة (F) مما يشير إلى معنوية معاملات الانحدار الجزئية ، وهذا يعني أن المتغيرات المستقلة مجتمعة لها تأثير معنوي في الانحدار أو أن واحدة على الأقل من معالم النموذج لها تأثير معنوي ، لكن المهم بالنسبة لنا أن تكون جميع معالم النموذج معنوية، ولهذا السبب يتم اختبار معنوية جميع معالم النموذج كل على حدة حتى يتم قبوله، و السؤال الذي يطرح نفسه: ما الفائدة إذاً من إجراء اختبار(F) إذا كنا سنقوم باختبار كل معلمة على حدة، والإجابة: أننا لو قبلنا الفرضية في اختبار(F) سنتوقف عند هذا الحد ولن نتابع لأننا سنرفض النموذج بالكامل وفي حال رفضنا للفرضية يستوجب الأمر اختبار المعنوية لكل معلمة جزئية (Bj) .

ومن الجدول رقم (3) coefficients نلاحظ عدم معنوية ثماني معاملات جزئية للنموذج ومعنوية ثلاث معاملات فقط هي على التوالي X_{15}, X_{249}, X_{24} عند مستوى دلالة $\alpha = 0.01$ حيث p-value للمتغيرات على التوالي :(0,0, 0.002)

من الجدول رقم (1) نلاحظ ارتفاعاً في قيمة مربع معامل الارتباط المتعدد $R^2 = 0.991$، وكذلك المعدل $\overline{R}^2 = 0.989$، وكذلك ارتفاع لقيمة معاملات الارتباط البسيط بين المتغير التابع والمتغيرات المستقلة جميعها في الوقت نفسه انخفاض معاملات الارتباط الجزئية بين المتغير التابع والمتغيرات المستقلة جدول رقم (3) Coefficients ، وهذه من مؤشرات الارتباط الخطي المتعدد وللتحقق من الأمر تم حساب معامل تضخم التباين(VIF) factor جدول رقم (3) وهي تشير إلى تجاوز ستة متغيرات للقيمة (10) مما يؤكد أن هذه المغيرات ترتبط ببعضها بعلاقة خطية وهي المتغيرات X_{23} $X_{245}, X_{24}, X_{12}, X_{13}, X_{14}$ ، وإذا ما نظرنا بعمق إلى جوهر هذه المتغيرات وجدنا المنطق في التعدد بين هذه المتغيرات ، فمثلاً إن المتغير X_{12} يشير إلى متوسط العمر المتوقع عند الولادة، ومن الطبيعي أن يكون مرتبطاً مع المتغير X245 معدل وفيات الأطفال بعمر أقل من 5 سنوات على اعتبار أن الأول يعتمد جزئياً عند حسابه على الثاني، وكذلك X_{13} نسبة المتعلمين الذين يجيدون القراءة والكتابة مرتبط مع المتغير X_{14} نسبة القيد المدرسي، وبالرجوع إلى مصفوفة الارتباط وجدنا أن معامل الارتباط البسيط بين $rx_{13}.x_{245} = -0.903$ وهو مرتبط بعلاقة عكسية مع متوسط العمر المتوقع وكذلك $rx_{13}.x_{14} = -0.795$ وكذلك $rx_{24}.x_{13} = 0.724$ وجميعها معنوية عند مستوى دلالة(0.01) وتستخدم الجذور المميزة لمصفوفة الارتباط ($X'X$) بين المتغيرات المستقلة في تشخيص مشكلة التعدد الخطي ، ففي حال وجود عدة جذور مميزة قريبة من الصفر فهذا مؤشر آخر على وجود مشكلة التعدد الخطي ، وهذا ما لاحظناه على (9) متغيرات كانت جذورها المميزة (Eigen value) قريبة من الصفر جدول رقم (4) Dimension (3,4,5,6,7,8,9,10,11)، أما بالنسبة لدليل الشرط (condition index) فهو عبارة عن الجذر التربيعي لحاصل قسمة أكبر جذر مميز على كل جذر مميز المقابل للأبعاد $D_j, j = 1,2....11$

ويحسب هذا الدليل لأكبر جذر مميز مثلاً وفق العلاقة : $\text{condition } index = \sqrt{\frac{8.656}{8.656} = 1}$

فإذا زادت قيمة الدليل على القيمة (15) فهذا مؤشر للشك بوجود المشكلة وإذا زادت قيمة الدليل على 30 فهذا دليل على وجود المشكلة وخطورتها ، وقد زادت قيمة دليل الحالة على 15 في المتغير

(X_{61})، وزادت أو اقتربت من (30) في المتغيرين (X_{24} ,X_{26}) ، جدول رقم (4) collinearity diagnostics.

الجدول رقم 4

Collinearity Diagnostics[a,b]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions											
				X12	X13	X14	X15	X23	X245	X249	X31	X61	X62	X24
1	1	8.656	1.000	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	2	1.511	2.394	.00	.00	.00	.01	.00	.02	.04	.00	.00	.00	.00
	3	.278	5.578	.00	.00	.00	.24	.00	.00	.05	.01	.03	.21	.01
	4	.149	7.629	.00	.00	.00	.31	.00	.00	.02	.00	.15	.40	.03
	5	.130	8.150	.00	.00	.00	.01	.00	.01	.04	.50	.20	.19	.00
	6	8.915E-02	9.854	.00	.00	.00	.15	.02	.18	.44	.27	.07	.00	.06
	7	7.550E-02	10.708	.01	.00	.00	.00	.22	.09	.12	.21	.45	.11	.01
	8	5.814E-02	12.201	.02	.02	.00	.09	.64	.15	.14	.01	.06	.00	.00
	9	3.761E-02	15.171	.00	.05	.04	.16	.00	.25	.02	.00	.03	.03	.62
	10	9.810E-03	29.703	.94	.10	.16	.00	.00	.28	.04	.00	.00	.01	.26
	11	6.271E-03	37.153	.02	.83	.79	.01	.12	.01	.09	.00	.00	.04	.01

a. Dependent Variable: X26

b. Linear Regression through the Origin

ولكل هذه الأسباب مجتمعةً وبدلاً من عملية الحذف لبعض المتغيرات ثم اختبار جودة توفيق النموذج اتبعنا استراتيجية أخرى ، وهي أسهل في الوصول لأفضل نموذج انحدار وهو أسلوب stepwise Regression method وهذا الأسلوب يعد تطويراً لأسلوب (الحذف الأمامي) Forward selection method ، ومن الضروري التعرض لهذا الأسلوب بالشرح قبل البحث في أسلوبstepwise ، لأنه وكما ذكرنا فإن أسلوب stepwise يعد تطويراً لأسلوب الحذف الأمامي ، وتعتمد استراتيجية الحذف الأمامي على إدخال المتغيرات المستقلة للنموذج واحداً تلو الآخر ، والخطوة الأولى تبدأ بحساب معامل الارتباط البسيط أو مصفوفة الارتباط بين المتغير التابع وكل من المتغيرات المستقلة ، ويتم اختيار المتغير المستقل الذي له أعلى معامل ارتباط مع المتغير (y) من بين جميع المتغيرات .

حيث وجدنا أن أعلى معامل ارتباط للمتغير التابع هو مع المتغير المستقل X_{24} $r_{yx\,24} = 0.980$ وهو معنوي عند مستوى دلالة $\alpha = 0.01$ جدول رقم(3) zero-order، ومن الضروري التنويه إلى أن هذا الأسلوب يحتاج إلى تحديد مستوى المعنوية لإدخال واستبعاد المتغيرات من النموذج ويجب أن يكون مستوى المعنوية لإدخال المتغيرات أقل من مستوى المعنوية للاستبعاد remove حيث حددنا مستوى الدلالة للإدخال enter $\alpha = 0.05$ وللاستبعاد بـ $\alpha = 0.1$بعد ذلك يتم حساب إحصاءه (F) للنموذج الذي تم توفيقه في هذه المرحلة ، فإذا كانت معنوية يتم حساب إحصاءه (t) لكل معلمة من معالم النموذج ، وعند حساب أول معلمه لا توجد ضرورة لحساب إحصاءه (t) لأن إحصاءة (F) تقوم بالغرض على اعتبار أنه لا يوجد لدينا إلا معلمة واحدة (a_1)حتى هذه المرحلة في المرحلة التالية ندخل المتغير المستقل الثاني والذي له أعلى معامل ارتباط جزئي مع المتغير التابع ثم نقوم بالاختبارات السابق ذكرها فإذا كانت معنوية ، نتابع بإدخال

متغير مستقل جديد إلى النموذج حتى نحصل على قيمة لـ (F) غير معنوية، عندئذ نتوقف، وعند هذا الحد يتوقف الاختيار والنموذج ما قبل الأخير يكون أفضل نموذج انحدار.

أما أسلوب stepwise فهو أسلوب يسمح باستبعاد بعض المتغيرات التي تم اختيارها في المراحل السابقة عند كل مرحلة من المراحل الآتية : إذا تم اكتشاف علاقة قوية بينه وبين أحد المتغيرات المستقلة التي تم ضمها للنموذج في المراحل اللاحقة وذلك على عكس طريقة الحذف الأمامي التي لا تستبعد متغيراً تم اختياره ، حتى لو تبين أن معامل ارتباطه الجزئي مع المتغير التابع أقل من معامل الارتباط الجزئي للمتغير المرتبط معه بشدة.

وبتطبيق هذه الاستراتيجية على دراستنا كان المتغير المختار الأول في الخطوة الأولى X_{24} لأن له أعلى معامل ارتباط بسيط مع المتغير التابع ، وهذا متوقع على اعتبار أن فتوة السكان هي من أهم المتغيرات المؤثر في T.F.R ، وأن قيمة$B_1=0.102$ وإن النموذج الأول الذي بين أيدينا الآن هو

$$\hat{y} = 0.102X_{24}$$

إن إحصاءة F لهذا النموذج$F_{cal} = 1806.464$ $model_1$جدول رقم (6) وهي معنوية حيث إن p-value=sig=0 وإن إحصاءة (t) لمعلمة النموذج $t_{cal}=42.503$ جدول رقم (7) وهي معنوية sig=0 وتفسر هذه المعلمة بأن كل زيادة بمقدار (1%) في نسبة السكان للفئة العمرية أقل من 15 عاماً سوف يؤدي إلى زيادة معدل الخصوبة الكلية بـ 0.102 طفلاً بالمتوسط لكل امرأة طول فترة حياة الإخصاب عندها (15-49) عاماً .

جدول رقم 5

Model Summary[g,h]

Model	R	R Square[a]	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.980[b]	.961	.960	.6721	
2	.992[c]	.984	.984	.4287	
3	.994[d]	.988	.987	.3790	
4	.995[e]	.989	.988	.3612	
5	.995[f]	.990	.989	.3525	1.963

جدول رقم 6

ANOVA[g,h]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	816.003	1	816.003	1806.464	.000[a]
	Residual	33.427	74	.452		
	Total	849.430[b]	75			
2	Regression	836.012	2	418.006	2274.209	.000[c]
	Residual	13.418	73	.184		
	Total	849.430[b]	75			
3	Regression	839.086	3	279.695	1946.749	.000[d]
	Residual	10.344	72	.144		
	Total	849.430[b]	75			
4	Regression	840.165	4	210.041	1609.561	.000[e]
	Residual	9.265	71	.130		
	Total	849.430[b]	75			
5	Regression	840.732	5	168.146	1353.262	.000[f]
	Residual	8.698	70	.124		
	Total	849.430[b]	75			

جدول رقم 7

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	X24	.102	.002	.980	42.503	.000	.980	.980	.980	1.000	1.000
2	X24	8.644E-02	.002	.827	39.740	.000	.980	.978	.585	.500	2.000
	X249	1.664E-03	.000	.217	10.434	.000	.802	.774	.153	.500	2.000
3	X24	.105	.004	1.006	23.417	.000	.980	.940	.305	.092	10.918
	X249	1.235E-03	.000	.161	7.315	.000	.802	.653	.095	.349	2.866
	X13	-.006	.001	-.157	-4.625	.000	.810	-.479	-.060	.146	6.854
4	X24	.110	.005	1.049	24.074	.000	.980	.944	.298	.081	12.361
	X249	1.224E-03	.000	.160	7.607	.000	.802	.670	.094	.349	2.867
	X13	-.010	.002	-.255	-5.435	.000	.810	-.542	-.067	.070	14.292
	X15	2.238E-05	.000	.072	2.876	.005	.544	.323	.036	.245	4.085
5	X24	.111	.004	1.061	24.737	.000	.980	.947	.299	.079	12.585
	X249	1.189E-03	.000	.155	7.528	.000	.802	.669	.091	.345	2.899
	X13	-.008	.002	-.203	-3.927	.000	.810	-.425	-.047	.055	18.280
	X15	3.015E-05	.000	.097	3.581	.001	.544	.394	.043	.199	5.023
	X23	-.005	.002	-.086	-2.137	.036	.731	-.247	-.026	.090	11.124

في الخطوة الثانية تم إدخال المتغير المستقل X_{249} والذي له أعلى معامل ارتباط جزئي مع المتغير التابع $r_{y.x\,249} = 0.774$ جدول excluded variables ، وهو معنوي وإن p-value=0 ، بثبات المتغير المختار في الخطوة السابقة حصلنا على نموذج الانحدار الآتي model(2)

$$\hat{y}= 0.102X_{24}+1.664E\text{-}0.3X_{249}$$

وتفسر معلمة النموذج السابق (a_2) بأن زيادة معدل وفيات الأمهات بمقدار 0.000001 سيؤدي إلى زيادة T.F.R بمقدار 0.001664 طفلاً بالمتوسط ، وقد أظهرت إحصاءة (F) من جدول ANOVA معنوية النموذج وإن كلا معلمتي النموذج (a_1,a_2) معنوية علماً بأن قيمة معامل تضخم التباين للمتغير (X_{249}) VIF=2 وإن تفسير العلاقة الطردية بين معدل وفيات الأمهات و (T.F.R) والتي تبدو للوهلة الأولى غير منطقية سببه زواج الأب بعد وفاة الأم من فتاة غالباً ما تكون قادرة على الإنجاب حتى ولو كان الأب متقدماً بالعمر ولديه الكفاية من الأطفال من الزوجة المتوفاة .

في الخطوة الثالثة تم إدخال المتغير X_{13} الإنفاق الحكومي على التعليم ، كنسبة مئوية من الناتج القومي الإجمالي وإن قيمة معامل الارتباط r=-0.479 وهو ثاني أكبر معامل ارتباط جزئي مع المتغير التابع وقيمته معنوية ويرتبط بعلاقة عكسية مع المتغير التابع فحصلنا على النموذج model(3)

$$\hat{y} = 0.102X_{24} + 1.664E-03X_{249} - 6.186E-03X_{13}$$

إن زيادة الإنفاق على التعليم بمقدار (1%) من الناتج القومي الإجمالي سيؤدي إلى انخفاض في T.F.R بمقدار 0.00186 طفلاً بالمتوسط ، المتغير المستقل التالي الذي تم إدخاله في المرحلة الرابعة (X_{15}) نصيب الفرد من الناتج القومي الإجمالي السنوي مقدراً بالدولار الأمريكي وكان تأثيره إيجابياً وجميع معاملاته معنوية :

$$\hat{y} = 0.102X_{24} + 1.664E-03X_{249} - 6.186E-03X_{13} + 2.238E-05X_{15}$$

إن زيادة متوسط الدخل السنوي بمقدار 1000$ أمريكي سنوياً ستؤدي إلى زيادة في T.F.R بمقدار (0.00223) طفلاً بالمتوسط ، إن المتغير المستقل الأخير والذي له أعلى معامل ارتباط جزئي مع المتغير التابع يثبات تأثير جميع المتغيرات المستقلة المدخلة في المراحل السابقة هو نسبة السكان الحضر في مجموع السكان (X_{23})، حيث كان له تأثير عكسي مع المتغير التابع وإن هذه الإشارة منطقية في تأثيرها ، ونلاحظ من خلال قيمة p-value=0 المرافقة لإحصاءة (t) أن معلمة المتغير X_{23} قد أظهرت معنوية عند مستوى دلالة $\alpha = 0.05$ هذا فضلاً عن معنوية النموذج بالكامل حيث تم اختبار الفرضية:

$$H_0 : B_1 = B_2 = B_3 = B_4 = B_5 = 0$$
$$H_1 : B_1 \neq B_2 \neq B_3 \neq B_4 \neq B_5 \neq 0$$

فقد أشارت قيمة p-value=0 المرافقة لقيمة F=1353.2 إلى معنوية النموذج ورفض الفرضية الصفرية هذا فضلاً عن معنوية معاملات الانحدار الجزئية كل على حدة، وإن النموذج الأخير ظهر على النحو :

$$\hat{y}= 0.102X_{24} + 1.664E-03X_{249} - 6.186E-03X_{13} + 2.2385-05X_{15} - 4.945E-03X_{23}$$

إذ إن زيادة نسبة السكان الحضر في مجموع السكان بمقدار (1%) سيؤدي إلى انخفاض في T.F.R بمقدار(0.004945-) طفلاً بالمتوسط .

إن المرحلة السادسة والأخيرة والتي عندها توقف إدخال متغيرات مستقلة جديدة للنموذج ، كان اختيار المتغير المستقل X_{245} والذي له أعلى معامل ارتباط جزئي بعد المتغيرات المختارة السابقة ، وقد أظهر اختبار (t) عدم معنوية هذه المعامل p-value=0.122>0.05 ومن ثم يتم التوقف في إدخال المتغيرات عند هذه المرحلة، وإن النموذج السابق كان أفضل نموذج يمثل العلاقة الخطية بين المتغيرات المستقلة المختارة والمتغير التابع .

وبالنظر إلى جدول collinearity diagnostics رقم 8 نلاحظ أن أكبر قيمة لدليل الحالة condition index كان للمتغير (X_{23}) حيث بلغت قيمته 10.347؛ مما يشير إلى عدم وجود مشكلة للتعدد الخطي في النموذج المختار .

جدول رقم 8

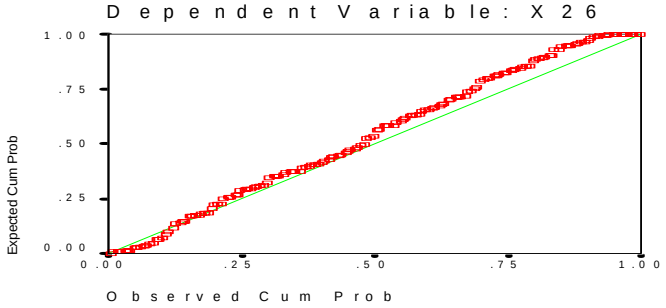
Collinearity Diagnostics [a,b]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions				
				X24	X249	X13	X15	X23
1	1	1.000	1.000	1.00				
2	1	1.707	1.000	.15	.15			
	2	.293	2.414	.85	.85			
3	1	2.380	1.000	.02	.04	.02		
	2	.565	2.052	.00	.40	.09		
	3	5.420E-02	6.627	.98	.56	.89		
4	1	2.902	1.000	.01	.02	.01	.02	
	2	.871	1.826	.00	.23	.00	.09	
	3	.190	3.912	.11	.41	.05	.51	
	4	3.726E-02	8.826	.88	.33	.94	.38	
5	1	3.765	1.000	.00	.01	.00	.01	.01
	2	.940	2.001	.01	.24	.00	.04	.01
	3	.193	4.416	.08	.37	.03	.49	.01
	4	6.682E-02	7.506	.23	.15	.03	.31	.86
	5	3.517E-02	10.347	.68	.24	.94	.14	.12

وللتحقق من التوزيع الطبيعي للأخطاء قمنا بتمثيل القيم الاحتمالية التراكمية للأخطاء على المحور الأفقي (observed cum prob) والقيم التراكمية المتوقعة للأخطاء (expected cum prob) على

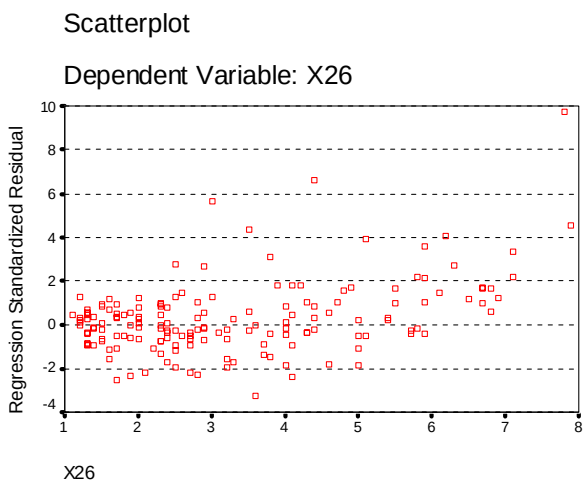
المحور الرأسي ، ولاحظنا من الشكل رقم (1) الناتج أن معظم النقاط تتجمع تقريباً بمحـاذاة الخـط المستقيم مما يشير إلى أن البواقي تتوزع طبيعياً بمتوسط يساوي الصفر .

شكل رقم 1



وللتحقق من فرضية تجانس تباين الخطأ العشوائي ، قمنا بتمثيل قـيم (y) علـى المحـور الأفقـي والبواقي المعيارية standardized residual على المحور الرأسي ولاحظنا من الشكل الناتج شكل رقم (2) أن النقاط تتوزع بشكل شريط أفقي متساوٍ حول الصفر؛ مما يدل على توافر فرضـيات التحليـل بصورة عامة ، حيث لا يعاني النموذج من مشكلة عدم تجانس تباين الخطأ العشوائي .

شكل رقم 2



ويهدف المقارنة بكيفية اختلاف تأثير المتغيرات المستقلة في (T.F.R) بحسب ارتفاع المستوى الاقتصادي للدول فمنا بتقسيم الـ (177) دولة إلى أربع مجموعات وذلك حسب مؤشر التنمية البشرية في كل دولة (HDI)Human Development Index والذي يحسب من العلاقة :

$$HDI = 1/3 \text{ (life expectancy index) + 1/3 (education index) +1/3 (GDP index)}$$

حيث إن قيمة المؤشر (HDI) تتراوح نظرياً بين (0-1) وأعطينا ترتيب كل مؤشر على النحو الآتي :

مجموعه	المؤشر HDI	الترتيب	عدد الدول
1	0.801-1	1	41
2	0.601-0.8	2	62
3	0.401-0.6	3	50
4	0.201-0.4	4	24

تم توفيق نموذج انحدار خطي لكل مجموعة من الدول بأسلوب STEPWISE فحصلنا على النتائج الملخصة في الجدول أدناه:

جدول رقم (10)

	مجموعة (1)	مجموعة (2)	مجموعة (3)	مجموعة (4)
المتغيرات المؤثرة	X_{24}	X_{24} , X_{13}	X_{24} , X_{13} X_{249} X_{61}	X_{24}
مربع معامل الارتباط المتعدد	0.980	0.995	0.992	0.999
مربع معامل الارتباط المتعدد المعدل	0.979	0.994	0.99	0.998
إحصاءة Durbin Watson	1.941	1.668	2.253	0.808
معنوية النموذجsig	0	0	0	0
قيم معاملات الانحدار الجزئي	+8.597E-02	+0.112 -6.933E-03	+0.139 -1.483E-02 +8.349E-04 -0.237	+ 0.118
p-value للمعالم باستخدام إحصائية (t)	0	0 , 0	0 0.01 0.008 0.019	0

تفسير النتائج:

أظهرت النماذج جميعها تقريباً أن متغير التعليم X_{13} يرتبط عكسياً مع T.F.R وهذا يشير إلى أن الجهل وزيادة نسبة الأمية هي أحد الأسباب الرئيسية للزيادة السكانية كما لم يكن لهذا المتغير أي دور في البلدان ذات مؤشر الدخل المرتفع مجموعة (1) وكذلك بلدان المجموعة (4)، فبالنسبة

للمجموعة الأولى فقد خرج مؤشر التعليم وجميع المتغيرات الأخرى باستثناء نسبة السكان بعمر أقل من 15 عاماً X24 من طور التأثير في الزيادة السكانية، وأصبح وعي المواطن في هذه المجموعة واهتماماته ونظرته للحياة والزواج أكبر من حدود تكوين أسرة وأطفال والعمل من أجل تأمين لقمة عيشهم وتعليمهم فقط ، أما بالنسبة لتأثير هذا المتغير (التعليم) في بلدان المجموعة (4) ، فلم يدخل في التأثير بعد ولا جميع المتغيرات الأخرى باستثناء X_{24} أي نسبة السكان بعمر أقل من 15 عاماً ، وكأن هذه المجتمعات مازالت مجتمعات بدائية ولا تعرف أياً من المؤثرات الخارجية التي تؤثر في هذه الظاهرة، وهذه الدول هي : تيمور الشرقية ـ الكونغو ـ مدغشقر ـ اليمن ـ زامبيا ـ ملاوي ـ سيراليون ـ موناموز ـ ارتيريا ـ تنزانيا ـ اثيوبيا ـ مالي ـ النيجر ـ الكونغو وغيرها .

وأما بالنسبة إلى المجموعة (2) فقد بدا واضحاً أثر التعليم في T.F.R فضلاً عن المتغير (X_{24}) نسبة السكان بعمر أقل من 15 عاماً وإلى المجموعة (3) أضف متغيرين جديدين في تأثيرهما وهما معدل وفيات الأمهات ومتغير الإنفاق الحكومي على الصحة وننوه في النهاية إلى أننا لم نستطع تضمين هذه الدراسة جميع النتائج والجداول والرسوم لأننا محكومون بعدد الصفحات .

ومهما كانت قيمة هذه الدراسة فلن تخرج عن كونها محاولة لوضع حجر إضافي في البناء الذي بدأ بتشييده آلاف قبلنا ولم يكتمل بعد ولن يكتمل .

المراجع

المراجع العربية

أبو سدرة فتحي صالح ، الكيخيا نجاة رشيد (1999) : الإحصاء والاقتصاد القياسي ، منشورات المركـز القومي للدراسات والبحوث العلمية ، دار الكتب الوطنية ، بنغازي ليبيا ص 47-45

إسماعيل ، محمد عبد الرحمن (2001) تحليل الانحدار الخطي معهد الإدارة العامة ، المملكة العربية السعودية ، الرياض ، ص 364

الصديقي، عبد اللطيف يوسف، الدوسري راشد حماد (دون تاريخ) معجم الإحصاء إنجليزي-فرنسي- عربي.

توماس .س . كنير، وجيمس أر. تايلور (1983): بحوث التسويق، مدخل تطبيقي، الجزء الأول، ترجمة عبد الرحمن دعالة، وعبد الفتاح السيد، دار المريخ للنشر، المملكة العربية السـعودية، الريـاض، ص 393

نترجون وازمان ويليام وكنتر ميخائيل (1990) ، نماذج إحصائية خطية تطبيقية، انحدار تحليـل تبـاين وتصميم تجارب الجزء الأول (الانحدار) ترجمة كنجوانيس إسماعيل، والزير عبد الحميـد بـن عبـد الله وآخرون النشر العلمي والطابع ـ جامعة الملك سعود ـ المملكة العربية السعودية الرياض، ص 382

المراجع الأجنبية

- Brown , C.E (1998) : Applied Multivariate , statistics in Geohydrology and related sciences , Springer – verlag . Berlin Heidelberg , chapter 6 , multiple regression . pp . 62-66
- Draper,N.R.and smith , H.(1981) : Applied Regression anlysis , New York , P.413
- kleinbaum , D.G, Kupper , L.L. and muller . k .E . (1988) : Applied Regression Analysis and other Multibariable methods DWS- KENT publishing company , a division of wadsworth P 317
- Koutsoyiannis , A., (1977) : Theory of Econometrics . , An introduction Exposition of Econometric Methods p129.

تاريخ ورود البحث إلى مجلة جامعة دمشق 2007/8/8.