Journal of Education and Science
Vol. 27, No.2, 2018, pp. 118-129
ISSN 1812-125X
http://edusj.mosuljournals.com

تصغير مجموع مربعات الخطأ في نموذج الانحدار الخطي المتعدد باستخدام الخوارزمية الجينية

همسة معن محمد ثابت

كلية الآداب/ قسم المعلومات والمكتبات/جامعة الموصل

تأريخ الاستلام 2013/05/06 تأريخ القبول 2013/09/10

Abstract

Regression models are regarded as the most important ones used in statistical models in defining the relation among variables through the available data which can be applied to various sciences . In most of these applications , there is a dependent variable and independent variables and the linear relation between them represents multiple linear regression function . The values of the dependent variable can be predicted when the independent variables take definite values and the model's parameters are estimated through minimizing deviation of squares among the values of the real and estimated data. In this research we use the genetic algorithm to find the minimum sum of error squares, where they are applied to different data in many applications . The genetic algorithm is able to achieve the minimum sum of error squares and estimating the parameters of the model.

Keywords : *Multiple Linear Regression ; Genetic Algorithm ; Statistics*

المستخلص

تعد نماذج الانحدار من أكثر وأهم النماذج الإحصائية استخداما في معرفة العلاقة بين المتغيرات من خلال البيانات المتاحة والتي يمكن تطبيقها في مجالات العلوم المختلفة . في كثير من هذه التطبيقات يكون هنالك متغير معتمد ومتغيرات مستقلة والعلاقة الخطية بينهم تم تمثليها من خلال دالة انحدار خطي متعدد , حيث يمكن التنبؤ بقيم المتغير المعتمد عندما تأخذ المتغيرات المستقلة قيما معينة , أما معلمات النموذج فإنها تقدر من خلال تقليل مربعات الانحراف بين قيم التغير المعتمد الفعلية والمقدرة . في هذا البحث قمنا باستخدام الخوارزمية الجينية لإيجاد أقل مجموع مربعات انحرافات , حيث تم تطبيقها على بيانات مختلفة في تطبيقات

عدة , تمكنا بواسطة الخوارزمية الجينية أن نحقق أقل مجموع لمربعات الأخطاء وتقدير قيم المعلمات للنموذج .

المقدمة

في كثير من التطبيقات العملية يود الباحث معرفة قيمة أحد المتغيرات عند معرفة قيمة للمتغير الأخر ولابد للوصول إلى إيجاد علاقة جبرية محددة بين المتغيرين , تسمى عادة العلاقة الرياضية التي تفرض التوقع أو التنبؤ بسلوك أحد المتغيرين عند معرفة الأخر بمعادلة (خط الانحدار) [7] , من اوائل من استخدم نماذج الانحدار الخطي كل من(Gauss و Legendre) مطلع القرن التاسع عشر [10].

إن معادلة خط الانحدار تتكون من المتغير المعتمد Y (dependent variable), والمتغير المستقل X (independent variable) ,أما الخطأ في تفسير Yi , يمكن كتابة معادلة إيجاده :

$$ei = Yi - \beta_0 - \beta_1 Xi$$

اذ ان :

β 0: تمثل الجزء المقطوع من محور Y

β 1: تمثل ميل خط الانحدار

ان معلمات النموذج β 0و β 1هي مقادير ثابتة يمكن حسابها من واقع البيانات المشاهدة من العينة , وبمعرفة قيمتهما يمكن التنبؤ بقيمة Yi عندما تاخذ Xi قيما معينة [1] .

ويرجع وجود الخطأ (ei) إلى عدة أسباب هي :

- إهمال بعض المتغيرات المستقلة التي يمكن ان تؤثر على المتغير المعتمد في النموذج .
- الصياغة الرياضية غير السليمة للنموذج .
- حدوث خطأ في كل من تجميع البيانات وقياس المتغيرات قيد الدراسة, [3].

لتحليل نموذج الانحدار هنالك فروض يجب توفرها وهي :

1- وجود علاقة خطية بين Y و X تمثل بمعادلة الخط المستقيم :

$$Yi = \beta_0 + \beta_1 Xi + ei$$

2- إن Y متغير عشوائي يتبع التوزيع الطبيعي بوسط حسابي قدره :

$$E(\hat{Y}) = \mu_{r/x}$$

وتباين ثابت قدره :

$$\sigma^2{}_{r/x} = \sigma^2$$

3- إن حد الخطأ ei متغير عشوائي يتبع التوزيع الطبيعي بوسط حسابي:

$$E(e_i) = 0$$

وتباين ثابت قدره :

$$E(e_i^2) = \sigma^2$$

4- إن قيمة الخطأ ei في فترة معينة غير مرتبطة مع قيمتها في فترة أخرى , أي أن ei's غير مرتبطة ذاتيا ولذا فتباينهما المشترك يساوي صفر أي :

$$E(e_i e_i') = 0$$

5- إن قيم المتغير X ثابتة وتقاس بدون أخطاء وهي قيمة غير مرتبطة مع قيم ei أي أن:

$$E(X_i e_i') = 0$$

ومن هذا نرى بأنه من المستحسن عمل اختبار لهذه الفروض واكتشاف إن كان هناك خلل في احد الفروض أو أكثر(تحليل الأخطاء أو المتبقيات) في نموذج الانحدار [2] في الجانب العملي من هذا البحث تم اختبار أكثر من خمسة نماذج تم اختيار ثلاث نماذج منها لأنها حققت فروض التحليل .

1- طرائق تقدير معلمات نموذج الانحدار الخطي المتعدد

(Methods of estimating the parameters of the multiple linear regression):

في كثير من التطبيقات العملية يكون هناك عدة متغيرات مستقلة ومتغير تابع واحد , فمثلا إيجار الوحدات السكنية والعوامل المؤثرة في الإيجار مثل المساحة , المنطقة , الدور , بعدها عن الخدمات ..... الخ , أيضا معدل نمو النبات يعتمد على نوع التربة ودرجة الحرارة وغيرها من العوامل [7] . إن دالة الانحدار لمثل هذه التطبيقات تكتب بالصيغة التالية :

$$Yi = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + ei \qquad i=1,\ldots,n$$

. [2]

هنالك عدة طرائق لتقدير قيم معلمات النموذج (β) بشكل يجعل $\hat{Y}$ أقرب ما يمكن للمتغير المعتمد Y , ومن هذه الطرائق :

أ- طريقة المربعات الصغرى

ب-طريقة الامكان الأعظم [3].

أ- طريقة المربعات الصغرى : تتميز طريقة المربعات الصغرى بأنها تختار أفضل نموذج مطابق للبيانات بحيث يجعل مجموع مربعات الخطأ ($e'e$) أقل ما يمكن وكالاتي :

$$\hat{e} = Y - \hat{Y}$$

$$\min \sum_{n=1}^{n} ei^{2} =$$

$$\min \sum_{n=1}^{n} (Y - \hat{Y})^{2} =$$

$$\min(Y - \hat{Y})'(Y - \hat{Y}) = \min(\hat{e}'\hat{e})$$

ان قيم المقدرات ($\hat{\beta}$) بطريقة المربعات الصغرى يمكن إيجادها عن طريق المعادلة التالية :

$$[\ \underline{\hat{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \qquad \text{الراوي, 1987}\ ]$$

ب-طريقة الامكان الأعظم : يتم تقدير معالم النموذج عن طريق تعظيم لوغاريتم دالة المعقولية العظمى , بفرض أن الأخطاء تتوزع توزيعا طبيعيا , يتم تقدير معالم النموذج ومن ثم الخطأ عن طريق تعظيم لوغارتيم دالة المعقولية حيث نشتق بالنسبة ل(β و σ^2) فنحصل على :

$$\underline{\hat{\beta}}_{MLE} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$$

$$\hat{\sigma}^{2}{}_{MLE} = \frac{1}{n}\left(\underline{Y} - \underline{X}\,\underline{\hat{\beta}}_{MLE}\right)'\left(\underline{Y} - \underline{X}\,\underline{\hat{\beta}}_{MLE}\right)$$

[3] في هذا البحث تم استحداث خوارزمية جينية جديدة لتقليل مجموع مربعات الخطأ في نموذج الانحدار الخطي المتعدد أقل ما يمكن .

2- تقليل مجموع مربعات الخطأ في نموذج الانحدار الخطي المتعدد عن طريق الخوارزمية الجينية

(Minimizing sum of squares error in multiple linear regression model by genetic algorithm):

تعد الخوارزمية الجينية (GA) إحدى التقنيات الذكائية التي تبحث في عدد هائل من الحلول لإيجاد الحل الأمثل للمسائل , أن دالة الهدف في الخوارزمية وحسب نوع المسألة المراد تطبيقها هي إما دالة تقليل (Minimization) أو دالة تعظيم (Maximization) [Sivanaudom and] [12], تحسب دالة الهدف الخاصة بالمسألة لكل مقطع (كروموسوم) من مقاطع الجيل في الخوارزمية ثم تحسب قيمة الجودة لكل مقطع من مقاطع الجيل اعتماداً على قيمة دالة الهدف لذلك

المقطع حيث تحدد مقدار جودة المقطع قياساً بباقي مقاطع الجيل فإذا كانت دالة الهدف تقليل (Minimization) (كما هو الحال في دالة تقليل مربع الأخطاء لنموذج الانحدار الخطي المتعدد)، فإن قيم الجودة تحسب كالأتي:

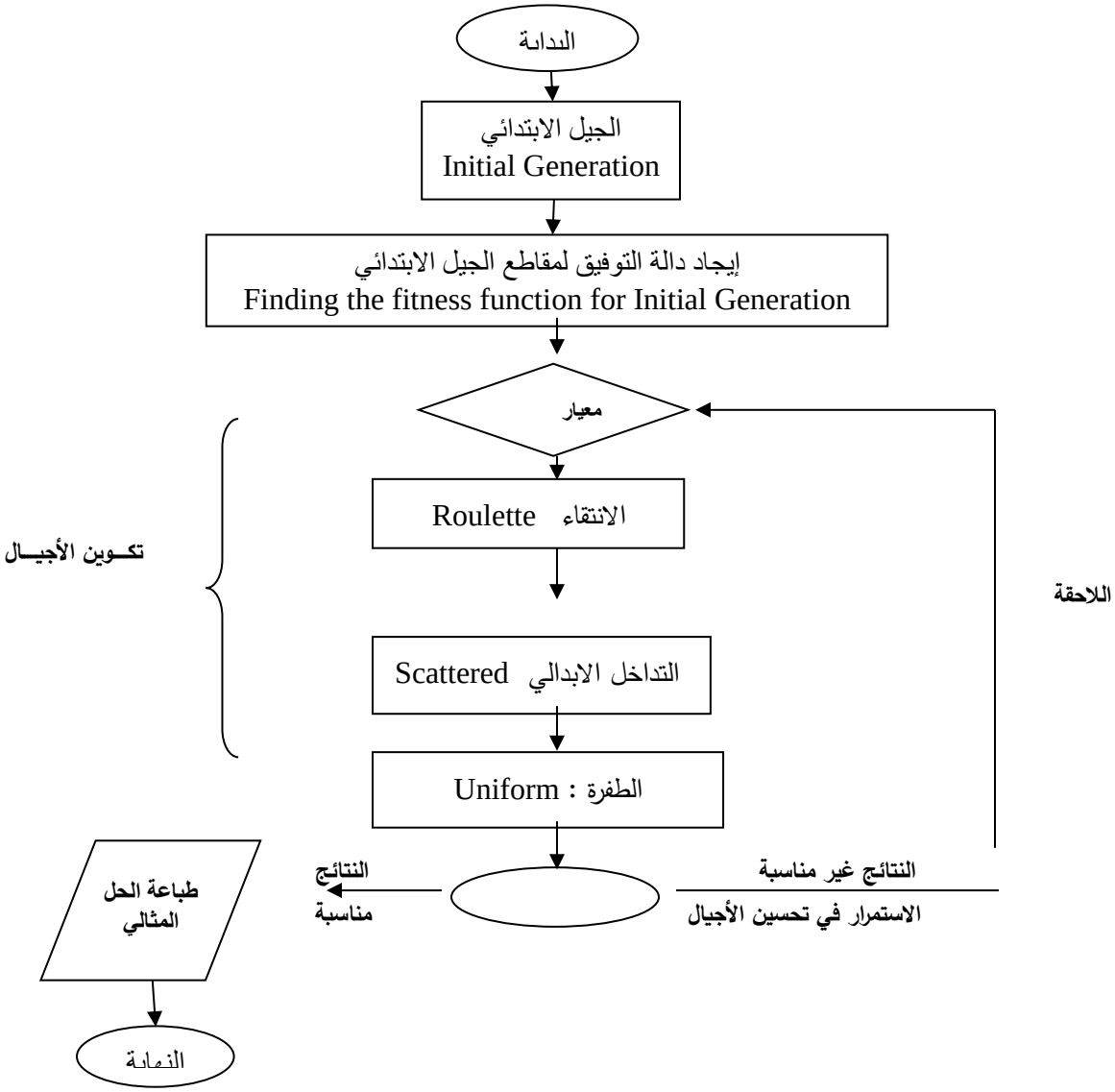
Fitness value (string i) = objective function (string i)*(-1) [GEN,2000].

بعد حساب قيم الجودة تجرى ثلاث عمليات تعتبر أساسية في الخوارزمية الجينية (GA) وهي :

- الانتقاء (Selection) .
- التداخل الابدالي (Crossover).
- الطفرة (Mutation) .

ان عملية التداخل الابدالي تجرى بين كروموسومين من الجيل الواحد في الخوارزمية الجينية (GA) لتحسين الأجيال اللاحقة , أما الطفرة فتجرى بين جينات الكرموسوم الواحد وبنسبة معينة تحدد حسب المسالة قيد الدراسة , أما العملية الثالثة فمهمة لاختيار أفضل الأفراد (الآباء) في الجيل الحالي لتوليد أفراد أفضل قدر الامكان في الأجيال اللاحقة [9].لإنشاء الجيل الابتدائي للخوارزمية الجينية المقترحة في بحثنا تم استخدام القيم الحقيقية للمسائل لتشفير الكروموسومات في الجيل الابتدائي تشفيرا حقيقيا , أما قيمة الجودة (Fitness Value) في هذه الخوارزمية تمثل (دالة تقليل مربع الأخطاء لنموذج الانحدار الخطي المتعدد) بعد أن تم برمجة قيمة التوفيق والمتغيرات بلغة (MATLAB) [10], تم استخدام دالة الهدف عن طريق برمجة الخوارزمية الجينية (GA) الجاهزة في برنامج (MATLAB9b) إذ تم تحديد حجم الجيل الابتدائي واختيار أنواع كل من عمليات (التداخل ألإبدالي : Scattered حيث تكون فيه الكروموسومات بشكل متجهات وباستخدام القناع يتم انتاج الاجيال الجديدة ، الطفرة : Uniform في هذا النوع من الطفرات يتم ابدال قيمة جين مختار من الكروموسومات بقيمة عشوائية تكون محصورة بين أعلى قيمة وأقل قيمة في الكرموسوم بنسب معينة وحسب كل مسالة والانتقاء :Roulette تعتمد هذه العملية على اختيار أفضل مافي المجتمع من افراد وذلك بعد حساب مجموع الجودة للمجتمع وحساب نسبة جودة كل فرد الى المجموع الكلي وتستخدم القيم الناتجة كاحتمالية لاختيار الافراد في الاجيال اللاحقة [9] ثم حدد عدد مرات توليد الخوارزمية , بعد كل تحديد لعدد الأجيال تم ملاحظة النتائج واعتماد التحديد المسبق

لعدد الأجيال لبيان مدى التقرب من الحل الأمثل , والشكل (1) يمثل مخططا انسيابيا للخوارزمية المستحدثة .



الشكل (1) مخطط انسيابي يوضح الخطوات الخوارزمية الجينية لتقليل مربع الأخطاء لنموذج الانحدار الخطي المتعدد

5- الجانب العملي (Application Part):

تم تنفيذ الخوارزمية الجينية بعد تحديد حجم المجتمع 20 كروموسوم ونوع التشفير لبيانات نماذج الانحدار والذي كان تشفيرا حقيقيا وأنواع كل من العمليات الأساسية الثلاثة وعدد الأجيال ,

نماذج للانحدار الخطي المتعدد في مجالات الحياة المختلفة , وظهرت النتائج خلال على ثلاثة ثواني معدودة وكما يلي :

1- في مجال الطب : البيانات في الجدول (1) تمثل مكونات دم الإنسان مع العمر بوحدات القياس المناسبة , [6] , كان عدد المتغيرات المستقلة لهذه البيانات خمس متغيرات وبعد إجراء اختبار فروض التحليل على البيانات بواسطة اختبار(Ryan-Joiner) في برنامج Minitab تم حذف المتغيرين الأول (الكريات الحمر) والثاني (الهيموكلوبين) لأنهما لم يحققا الفرضين الاول والثاني من فروض التحليل :

الجدول (1) يبين بيانات مكونات دم الإنسان مع العمر بوحدات القياس المناسبة

مكون الدم / العمر Y	حجم الكريات الحمر x1	الخلايا البيض x2	ترسيب الكريات الحمر x3
18	1.55	0.76	6.45
8	4.03	1.04	5.55
6	2.91	2.48	4.58
5	5.17	0.96	1.75
12	3.95	1.43	4.74
13	4.46	1.62	4.82
17	4.96	2.38	4.93
7	2.93	0.79	2.01
10	4.02	1.93	3.85
19	5.11	2.47	5.31
20	1.77	1	5.06
27	1.96	1.54	7.61
33	2.67	2.12	2.43
38	5.96	1.43	2.48
30	3.52	1.51	5.52
25	2.34	1.2	6.32
31	3.73	1.62	5.73
29	3.12	1.48	5.34
34	4.07	1.92	3.82
36	3.26	2.04	8.14
35	4.25	1.62	6.22
28	2.16	1.96	4.85
14	2.13	2.43	5.56
7	3.18	1.12	3.94
5	4.81	0.82	2.91
40	3.01	1.18	8.52

8.87	1.37	3.28	45
2.64	1.52	2	49
1.88	0.59	6.43	57
1.94	0.6	2.66	41
4.87	0.94	4.85	52
5.02	1.26	5.12	54
2.25	0.82	3.24	43
3.14	0.86	3.52	44
7.8	1.34	5.73	55
3.55	1.07	3.81	40
4.06	0.9	4.23	46
8.13	1.18	5.51	52
7.56	3.1	2.02	62
9.04	1.77	3.33	66
4.32	3.38	5.56	75
1.97	0.44	5.39	60
3.81	1.16	3.46	61
8.6	3.22	3.82	72
7.83	2.35	4.46	68
8.26	3.14	4.15	71
9.21	3.42	3.38	76
7.31	1.84	3.82	64
3.92	1.86	3.42	63
7.44	3.18	4.29	70
8.62	2.81	3.92	67

إن دالة الانحدار لبيانات الجدول (1) تكتب بالصيغة التالية :

$$Yi = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ei$$

بعد تحليل البيانات من خلال Minitab والخوارزمية الجينية حصلنا على النتائج في الجدول(2) برنامج

الجدول (2) يوضح قيم MSE لبيانات الجدول (1) من خلال برنامج Minitab والخوارزمية الجينية

	MSE	الانتقاء	التداخل الابدالي	الطفرة	عدد توليدالأجيال
GA	2102.377	Roulette	Scattered	بنسبة 1.0 Uniform	100
Reg.	2196.3				

من خلال الجدول (2) نلاحظ بان قيمة مجموع مربعات الاخطاء الناتجة من الخوارزمية الجينية أصغر .

3- في مجال النفط : البيانات في الجدول (3) تمثل متغيرات دالة الاستهلاك العالمي للمشتقات البترولية للمدة (1970-2004) [5] . كان عدد المتغيرات المستقلة للبيانات أربع متغيرات وبعد إجراء اختبار فروض التحليل على البيانات بواسطة اختبار(Ryan-Joiner) في برنامج Minitab تم حذف المتغير(السعر الاسمي) لأنه لم يحقق الفرض الاول من فروض التحليل:

الجدول (3) يبين متغيرات دالة الاستهلاك العالمي للمشتقات البترولية للمدة (1970-2004)

السنوات	الاستهلاك العالمي للمشتقات البترولية Y	إنتاج السعودية X1	إنتاج أوبك X2	إنتاج العالم X3
1970	43392.6	2786.5	23413.0	45200.0
1971	46599.2	2743.4	25326.3	47800.0
1972	50033.3	2712.2	27094.4	50600.0
1973	54102.1	2893.2	30988.5	55300.0
1974	53403.6	2977.7	30729.2	55700.0
1975	53304.8	3075.9	27155.0	52600.0
1976	56682	3125.3	30737.8	57500.0
1977	59166.6	3001.5	31253.5	59700.0
1978	61330	2905.7	29805.3	60200.0
1979	62290.3	2818.8	30928.9	62700.0
1980	59925.6	2714.2	26850.7	60047.1
1981	58102.5	2766.4	22485.3	56283.0
1982	56817.7	2734	18734.4	53579.0
1983	56453.4	2846.7	16615.6	52405.2
1984	57415.8	2947.2	15933.7	52992.6
1985	57334.6	3037.9	14921.0	52272.5
1986	58654.4	3087.8	17660.2	55091.3
1987	60118.4	3135.2	16741.5	54605.6
1988	61883.5	3134.8	18841.4	56887.9
1989	62735.7	3163.1	20406.3	57784.1
1990	62982.5	3136.2	22021.1	59105.8
1991	63305.6	3204	22253.6	58736.4
1992	63689.6	3251	23845.1	59332.1
1993	63976.7	3329.4	24230.5	59130.0
1994	64910	3405.5	24609.4	59889.5
1995	66148.6	3425.9	24600.1	60322.1
1996	68015.8	3493.9	24769.2	61493.5
1997	70048	3538.7	25431.8	62880.3
1998	70215.6	3552.2	27739.7	65137.0
1999	71629.8	3580.5	26227.8	63402.7
2000	72332.5	3641.8	27745.0	65663.2
2001	73128.7	3767.1	26873.5	65302.3

63933.8	24322.5	2786.5	73899.1	2002
67090.0	26885.4	2743.4	75585.7	2003
70575.4	29577.7	2712.2	77942.1	2004

إن دالة الانحدار للبيانات في الجدول (3) تكتب بالصيغة التالية :

$$Yi = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ei$$

بعد تحليل بيانات الجدول (3) توصلنا الى النتائج الموضحة في الجدول (4) خلال استخدامنا الجينية وبرنامج Minitab.للخوارزمية

الجدول (4) يوضح قيم MSE لبيانات الجدول (3) من خلال برنامج Minitab والخوارزمية الجينية

	MSE	الانتقاء	التداخل الابدالي	الطفرة	عدد توليدالأجيال
GA	604594152.22	Roulette	Scattered	بنسبة 1.0 Uniform	75
Reg.	717544268				

من خلال الجدول (4) نلاحظ بان قيمة مجموع مربعات الاخطاء الناتجة من الخوارزمية الجينية أصغر لبيانات الجدول (3).

3- في مجال الهندسة : البيانات في الجدول (5) تمثل خمسة متغيرات , الأول يمثل المتغير المعتمد (كلفة التدفئة) والمتغيرات المستقلة الأربع تتمثل في (درجة الحرارة الصغرى خارج الدار , سمك العازل بالانج , عمر جهاز التدفئة , احتواء الدار على مراب) [4]. بعد إجراء اختبار فروض التحليل على البيانات بواسطة اختبار(Ryan-Joiner) في برنامج Minitab تم حذف المتغير المستقل (سمك العازل) لانه لم يحقق الفرض الاول من فروض التحليل:

الجدول (5) يبين متغير كلفة التدفئة والعوامل المؤثرة عليها

احتواء الدار على مراب x3	عمر جهاز التدفئة x2	درجة الحرارة الصغرى خارج الدار x1	كلفة التدفئة Y
0	6	35	250
1	10	29	360
0	3	36	165
0	9	60	43
0	6	65	92
0	5	30	200
1	7	10	355
1	10	7	290
0	11	21	230

0	5	55	120
0	4	54	73
1	1	48	205
1	15	20	400
1	7	39	320
0	6	60	72
1	8	20	272
0	3	58	94
1	11	40	190
0	8	27	235
0	5	30	139

إن دالة الانحدار للبيانات في الجدول (5) تكتب بالصيغة التالية :

$$Yi = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ei$$

اما النتائج التي توصلنا اليها من Minitab والخوارزمية الجينية موضحة من خلال الجدول خلالبرنامج
(6).

الجدول (6) يوضح قيم MSE لبيانات الجدول (5) من خلال برنامج Minitab والخوارزمية الجينية

	MSE	الانتقاء	التداخل الابدالي	الطفرة	عدد توليدالأجيال
GA	4931.25684086	Roulette	Scattered	0.448387 بنسبة Uniform	100
Reg.	57596				

من خلال الجدول (6) نلاحظ بان قيمة مجموع مربعات الاخطاء الناتجة من الخوارزمية الجينية أصغر لبيانات الجدول (5).

6- الاستنتاجات :

1. إن الخوارزمية الجينية (GA) وجدت الحل الأمثل للمسائل التي تناولها البحث , أي وجدت أقل مجموع لمربعات الأخطاء , مقارنة مع النتائج التي استخرجت من خلال برنامج Minitab.
2. تم استخدام أنواع محددة من العمليات الثلاث (التداخل الابدالي والطفرة والانتقاء) في الخوارزمية الجينية (GA) لأنها أعطت أفضل النتائج , وأعداد مختلفة من الأجيال لتوليد الخوارزمية وحسب حجم مجتمع الدراسة لكل مسالة .

7- التوصيات :

نوصي باستخدام الخوارزمية الجينية في حل مسائل علم الإحصاء الامثلية ، لما توفره من حلول مثالية بوقت قصير جدا مقارنة ببقية التقنيات الذكائية والبرامج الحاسوبية.

المصادر

1- أبو عمة , عبد الرحمن بن محمد سليمان و محمود محمد إبراهيم هندي (2007):" الإحصاء التطبيقي " , العبيكان للأبحاث والتطوير , الرياض .

2- الراوي, خاشع محمود (1987) : " المدخل الى تحليل الانحدار " , مطبعة دار الكتب للطباعة والنشر , جامعة الموصل.

3- شيخي محمد (2012) :"طرق الاقتصاد القياسي : محاضرات وتطبيقات " , دار الحامد للنشر والتوزيع , عمان .

4- طارق شريف يونس و رائد عبد القادر (2010) : " التحليل الإحصائي : مفاهيم – منهجية – تطبيقات استخدام البرمجية Minitab " , فخراوي للدراسات والنشر , مملكة البحرين .

5- عمار محمد سلو أحمد (2006) : " السياسة الإنتاجية والسعرية للمملكة العربية السعودية في مجال النفط الخام " , أطروحة دكتوراه , كلية الإدارة والاقتصاد , جامعة الموصل .

6- فاتن عبد الواحد مجيد (2005) : " دراسة مقارنة لأنواع اليرقان وتأثيرها على بعض مكونات الدم في مدينة كركوك " , رسالة ماجستير , كلية التربية للبنات , جامعة تكريت.

7- محمود محمد إبراهيم هندي و خلف سلمان سلطان سلمان (2007) : " مفاهيم لطرق التحليل الإحصائي " , مكتبة الرشد , الرياض .

8-GEN,M.(2000):"Genetic Algorithms and Engineering Optimization", John Wiley and Sons, Inc.

9-Goldberg, D.E. (1989): "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley.

10-Julian, J.(2006) "Extending the linear model with R" , Taylor & Francis Group , LLC .

11-Rutkowski, L. Scherer,R. Tadeusiewicz,R. Zadeh,L. Zurada,J.nb(2010)" Artificial Intelligence and Soft Computing", Springer– Verlag Berlin .

12-Sivanaudom, S.N. Deepa , S.N. (2008): "An Introduction to Genetic Algorithms ", Springer – Verlag Berlin Heidelberg.

13- Oliver, S .Francis J.P.(2002) " Contemporary Statistical Models", CRC Press LIC.