دار المنظومة
DAR ALMANDUMAH
الرواد في قواعد المعلومات العربية

العنوان:	اختيار المتغيرات في انحدار الحرف
المؤلف الرئيسي:	النعيمي، أسوان محمد طيب رشيد
مؤلفين آخرين:	دبدوب، مروان عبدالعزيز(مشرف)
التاريخ الميلادي:	2005
الصفحات:	1 - 120
رقم MD:	552409
نوع المحتوى:	رسائل جامعية
اللغة:	Arabic
الدرجة العلمية:	رسالة ماجستير
الجامعة:	جامعة الموصل
الكلية:	كلية علوم الحاسبات والرياضيات
الدولة:	العراق
قواعد المعلومات:	Dissertations
مواضيع:	أنحدار الحرف، الاحصاء، المتغيرات الرياضية
رابط:	http://search.mandumah.com/Record/552409

© 2020 دار المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو دار المنظومة.

اختيار المتغيرات في انحدار الحرف

رسالة مقدمة

الى

مجلس كلية علوم الحاسبات والرياضيات في جامعة الموصل

وهي جزء من متطلبات نيل شهادة الماجستير علوم في الإحصاء

من قبل الطالبة

أسوان محمد طيب رشيد النعيمي

بإشراف

الأستاذ المساعد

مروان عبد العزيز دبدوب

١٤٢٦هـ ٢٠٠٥ م

Variables Selection in Ridge Regression

A Thesis Submitted To
Council of Collage of Computers Sciences And Mathematics
University of Mosul

As
Partial Fulfillment of The Requirements for the
Degree of Master of Science in Statistics

By
Aswan M.T.R. Al-Nu'eimy

Supervised by
Assistant professor
Marwan A. Dabdoub

1426 A.H **2005 A.c**

بسم الله الرحمن الرحيم

يرفع الله الذين آمنوا منكم والذين

أوتوا العلم درجات والله بما

تعملون خبير

صدق الله العظيم

سورة المجادلة

الآية (١١)

الإهداء

- إلى من عبدا لي الدرب وذللا لي الصعاب وغمراني بفيض من الحنان

والدي الكريمين...

- إلى سنـــد الـــدهـــــر

أخي محمـــد

واخوتي وأخواتي

- إلى كل من ينشد الوقوف على الحقيقة التي ما فوقها غاية ولا من دونها درايـــــة

أهـــدي ثمـــرة جهـــدي

أســوان

شكر وتقدير

يطيب لي وانا انهي رسالتي هذه فالحق يدعوني ويملي علي ان اتوجه بالشكر والتقدير الى استاذي الفاضل الاستاذ المساعد مروان عبد العزيز دبدوب لمتابعته العلمية طيلة مدة اعداد الرسالة ولما قدمه من ارشادات وتوجيهات قيمة كان لها الاثر الكبير في اغناء الرسالة فجزاه الله عني كل الخير .

واقدم شكري وتقديري الى الدكتور ظافر رمضان مطر عميد كلية علوم الحاسبات والرياضيات لدعمه المتواصل لطلبة الدراسات العليا . ويسرني ان أتقدم بفائق الاحترام والتقدير للدكتور حسن محمد الياس رئيس قسم الإحصاء لما أبداه من مساعدة وتشجيع ودعم معنوي خلال فترة الدراسة .

كما يدفعني واجب العرفان ان أسجل فائق شكري وتقديري للاستاذ المساعد الدكتور صفاء يونس الصفاوي حيث كان خير مثال في الارشاد العلمي طوال مدة اعداد الرسالة ولجهوده المتميزة فيما يتعلق بتأمين جانباً من المصادر . كما اقدم بالغ شكري وتقديري الى أستاذتي الأفاضل في قسم الإحصاء لابداء المساعدة العلمية والتفقد طيلة فترة الدراسة والبحث فجزاهم الله عني كل الخير .

كما أتقدم بجزيل الشكر والتقدير الى السيد رئيس لجنة المناقشة واعضاؤها لتفضلهم بالموافقة على مناقشة الرسالة .

ان واجب الوفاء والعرفان ان أسجل شكري وتقديري الى الدكتور محمد حسن عواد لما ابداه من مساعدة في جمع البيانات فضلاً على توجيهاته السديدة فيما يخص الجانب الطبي المتعلق بموضوع الرسالة . ويلزمني واجب الاعتراف بالجميل ان أسجل فائق شكري وتقديري الى الأستاذ الدكتور مقداد رحمة الله الجواري مديري في الدائرة (ديوان رئاسة جامعة الموصل) لما أبداه من مساعدة وتشجيع مستمر في اعتماد بيانات مرضى الثلاسيميا في الرسالة وكان لي خير عون من الناحية الطبية وتشجيعه لكونه مدير جمعية الثلاسيميا على المواصلة في هذا الموضوع من الناحية الإنسانية على أمل القضاء على هذا المرض في العراق كما في بعض الدول اذ لم يبخل علي باي معلومة طبية او مصدر طبي او لقاء مع أي جهة تعنى بالموضوع فجزاه الله عني كل الخير .

كما اقدم بالغ الشكر والتقدير الى الست فاتن علي احمد والسيد ثائر ابراهيم خليل الموظفين المتطوعين مجاناً في جمعية الثلاسيميا لما ابدياه من مساعدة باغناء الرسالة طبياً بكل ما يستطيعون .

كما اتقدم بالشكر والتقدير للدكتور مزاحم سعيد الحمداني مدير مركز الثلاسيميا في الموصل لمساعدته الطبية ومتابعته للنتائج الإحصائية . وكذلك اتقدم بشكري للدكتور ناصح عبد اللطيف القزاز الطبيب في مركز الثلاسيميا للمشورات والمساعدة الطبية .

ولابد ان اقدم اعتزازي وتقديري الى اخوتي واخواتي طلبة الدراسات العليا بالصحبة الطيبة والمودة التي جمعتنا في الدراسة . كما اتقدم بجزيل الشكر والتقدير لكل من أسهم بجهد او مشورة في انجاز الرسالة .

ختاماً نسأله تعالى ان ينفعنا بما علمنا ومنه التوفيق والسداد .

أسوان

الملخـــص

كشف عن وجود مشكلة التعدد الخطي بطرائق عدة اشار معظمها الى وجـود هـذه المشكلة التي عولجت بطريقة انحدار الحرف وقد اتبعت عدة طرائق لتعيين قيمة الثابـت k المحصورة بين (0,1) ، وهي : ملاحظة القيم التقديريـة لمعلمـات الانحـدار بمـرحلتين ، المخطط البياني لاثر الحرف Ridge trace وتدوير هذا المخطط ، الاسـتعانة بالمقياسـين : معامل تضخم التباين VIF ومقياس العدد الشرطي .c.n .

جزئ مخطط اثر الحرف ليحتوي كل مخطط على احد المتغيرات التوضيحية منفرداً والاستعانة بها في تعيين المتغيرات التوضيحية التي تؤثر في المتغير المعتمد .

اتبعت طرائق عدة لتعيين المتغيرات التي تدخل الى معادلة انحـدار الحـرف وكـان للمصفوفة الممددة Augmented Matrix أهمية واضحة في تسهيل كافة العمليات التحليليـة لانحدار الحرف واختيار افضل المتغيرات باتباع اساليب : الاختيار الامامي ، الاختيار الخلفي وانحدار الخطوات المتسلسلة .

لذا يفضل عند تعيين الثابت k إستخدام طريقة اثر الحرف بمـرحلتين ، والاسـتعانة بالعديد من الطرائق عند الكشف عن وجود مشكلة تعدد العلاقة الخطية ، والاستفادة ايضا من هذه الطرائق في تعيين الثابت k . تدوير رسم اثر الحرف فـي انحـدار الحـرف لتعيـين الثابت k.

تم تجزئة مخطط كل متغير في اثر الحرف باختيار افضل معادلة انحدار والاسـتعانة بالمصفوفة الممددة Augmented Matrix عند اختيار افضل معادلة انحدار لتعزيـز نتـائج طريقة اثر الحرف للقيم التقديرية لمعلمات معامل الانحدار .

جمعت البيانات من مستشفى ابن الاثير التعليمي للـولادة والاطفـال فـي الموصـل لمرضى مصابين بالثلاسيميا واحتوت الدراسة على متغير الاستجابة هو عمر العظم الذي لـه اهمية في تشخيص هذا المرض واستخدمت عشرة متغيرات توضيحية لدراسة تأثيرهـا فـي المتغير المعتمد واعطى الاسلوب المتبع في اختيار المتغيرات في انحـدار الحـرف اهميـة احصائية لثلاث متغيرات توضيحية هي X_1 عمر المريض الحقيقي مقاسـاً (بالشـهر) ، X_4 هيموكلوبين الدم ، X_7 آرومة حمراء .

قائمة المحتويات

الموضوع رقم الصفحة

الفصل الاول: المقدمة والنبذة التاريخية والهدف ١-٧

1-1 المقدمة ١

1-2 نبذة تاريخية ٣

1-3 الهدف ٧

الفصل الثاني : تحليل الانحدار ٨-٢٩

2-1 مفهوم تحليل الانحدار ٨

2-1-1 فروض تحليل الانحدار ٩

2-2 طريقة المربعات الصغرى الاعتيادية ١٠

2-2-1 التحيز والتباين للمعلمات المقدرة في طريقة المربعات الصغرى الاعتيادية... ١٢

2-3 مشكلة تعدد العلاقة الخطية ١٣

2-3-1 انواع تعدد العلاقة الخطية..... ١٤

2-3-2 الكشف عن وجود مشكلة تعدد العلاقة الخطية ١٥

2-4 اختيار افضل معادلة انحدار ٢٢

2-5 طرائق اختيار افضل معادلة انحدار ٢٥

2-5-1 اسلوب الاختيار الامامي ٢٥

2-5-2 اسلوب الاختيار الخلفي ٢٦

2-5-3 اسلوب انحدار الخطوات المتسلسلة ٢٦

2-5-4 طريقة كل الانحدارات الممكنة ٢٧

2-6 معالجة مشكلة تعدد العلاقة الخطية ٢٨

الفصل الثالث : انحدار الحرف ٣٠-٤٦

3-1 استخدام انحدار الحرف ٣٠

3-2 التقدير بطريقة انحدار الحرف ٣١

3-3 خواص مقدرات انحدار الحرف ٣٣

3-4 متوسط مربعات الاخطاء لمقدرات الحرف ٣٤

3-5 مقياس متوسط مربعات الخطأ العمومي لمقدرات انحدار الحرف ٣٨
3-6 معلمة الحرف ٤٠
3-7 طرائق اختيار معلمة الحرف ٤١
3-7-1 طريقة Horel , Kennard & Baldwin ٤١
3-7-2 طريقة Thistal ٤١
3-7-3 الطريقة التكرارية ٤٢
3-7-4 اثر الحرف ٤٣
3-8 اختيار المتغيرات في انحدار الحرف ٤٥

الفصل الرابع : الجانب التطبيقي ٤٦-١١٢

4-1 تمهيد ٤٧
4-2 نظرة في الثلاسيميا ٤٧
4-3 جمع البيانات ٥١
4-4 تحاليل احصائية لمتغيرات داخلة في النموذج ٥٢
4-5 متغيرات اخرى ٦٠
4-6 تقدير معلمات نموذج الانحدار بطريقة المربعات الصغرى الاعتيادية ٦٧
4-7 الكشف عن وجود مشكلة التداخل الخطي المتعدد ٦٨
4-8 اختيار افضل معادلة انحدار ٧٦
4-9 معالجة مشكلة تعدد العلاقة الخطية بطريقة انحدار الحرف ٨٢
4-10 طريقة الكشف عن وجود مشكلة التعدد الخطي في
انحدار الحرف وعلاقتها بقيمة k المختـــــــارة ٩٠
4-11 اختيار المتغيرات في انحدار الحرف ٩٤
4-11-1 اختيار افضل المتغيرات بالاعتماد على طريقـــة
اثر الحرف في انحدار الحـــــــــــــــرف ٩٤
4-11-2 تجزئة اثر الحرف في انحدار الحرف ٩٥
4-11-3 اختيار افضل المتغيرات بالاعتماد على اسلـــــــــــــــــوب
الخطوات المتسلسلة واسلوب الاختيار الامامـــــــــــــــــي
واسلوب الاختيار الخلفي للمصفوفة الممددة Augmented Matrix..... ١٠١

الاستنتاجات ١١١
التوصيـــات ١١٢
المصـــــادر ١١٣
الملاحـــــق

قائمة الأشكال

الشكل رقم الصفحة

2-1 اشكال توضح رسم الانتشار لمعامل التحديد R^2 ..... ١٨
4-1 X_1 العمر الحقيقي (بالشهر) لمرضى الثلاسيميا ٥٣
4-2 X_2 عمر المريض عند ظهور المرض (بالشهر) لمرضى الثلاسيميا ٥٣
4-3 X_3 تضخم الكبد (بالسنتمتر) لمرضى الثلاسيميا ٥٤
4-4 X_4 هيموكلوبين الدم لمرضى الثلاسيميا ٥٥
4-5 X_5 مكداس الدم لمرضى الثلاسيميا ٥٥
4-6 X_6 الخلايا الشبكية لمرضى الثلاسيميا ٥٦
4-7 X_7 الارومة الحمراء لمرضى الثلاسيميا ٥٦
4-8 X_8 الهيموكلوبين الجيني لمرضى الثلاسيميا ٥٧
4-9 X_9 عدد وحدات الدم لمرضى الثلاسيميا ٥٨
4-10 X_{10} بداية نقل الدم للمريض وحسب العمر مقاساً (بالشهر)
لمرضى الثلاسيميا ٥٨
4-11 Y العمر من العظم مقاساً (بالشهر) لمرضى الثلاسيميا..... ٥٩
4-12 نسبة الذكور والاناث لعينة مرضى الثلاسيميا ٦١
4-13 انتشار الاوساط الحسابية لكل من الجنسين (ذكور واناث) لمرضى الثلاسيميا .. ٦١
4-14 نسبة الشكل غير المنغولي والشكل المنغولي لمرضى الثلاسيميا ٦٣
4-15 انتشار الاوساط الحسابية للشكل غير المنغولي والشكل
المنغولي لمرضى الثلاسيميا ٦٤
4-16 X_{03} الطول مقاساً (بالسنتمتر) لمرضى الثلاسيميا ٦٥
4-17 X_{04} الوزن مقاساً (بالكغم) لمرضى الثلاسيميا ٦٥
4-18 الوزن والطول لمرضى الثلاسيميا ٦٦
4-19 خط الانحدار (الطول والوزن) لمرضى الثلاسيميا..... ٦٦
4-20 اثر الحرف في انحدار الحرف لجميع متغيرات مرضى الثلاسيميا ٨٩
4-21 اثر الحرف المدور في انحدار الحرف لجميع متغيرات مرضى الثلاسيميا ٩٠
4-22 اثر الحرف في المتغير X_1 العمر مقاساً (بالشهر) لمرضى الثلاسيميا..... ٩٥

4-23 اثر الحرف في المتغير X_2 بداية ظهور المرض مقاساً بالشهر لمرضى الثلاسيميا..... ٩٦
4-24 اثر الحرف في المتغير X_3 تضخم الكبد مقاساً (بالسنتمتر) لمرضى الثلاسيميا.. ٩٦
4-25 اثر الحرف في المتغير X_4 هيموكلوبين الدم لمرضى الثلاسيميا ٩٧
4-26 اثر الحرف في المتغير X_5 مكداس الدم لمرضى الثلاسيميا ٩٨
4-27 اثر الحرف في المتغير X_6 خلايا الشبكية لمرضى الثلاسيميا ٩٨
4-28 اثر الحرف في المتغير X_7 آرومة حمراء لمرضى الثلاسيميا ٩٩
4-29 اثر الحرف في المتغير X_8 الهيموكلوبين الجيني لمرضى الثلاسيميا..... ٩٩
4-30 اثر الحرف في المتغير X_9 عدد وحدات الدم لمرضى الثلاسيميا ١٠٠
4-31 اثر الحرف في المتغير X_{10} بداية نقل الدم للمريض حسب العمر مقاساً (بالشهر) لمرضى الثلاسيميا ١٠١
4-32 اثر الحرف في انحدار الحرف للمتغيرات المعتمدة في اختيار افضل معادلة انحدار لمتغيرات مرضى الثلاسيميا ١٠٨

قائمة الجداول

الجدول رقم الصفحة

4-1 تحليل تباين الانحدار (الجنس) ٦٠
4-2 تحليل تباين الانحدار (الشكل المنغولي) ٦٣
4-3 تباين الانحدار لمتغيرات مرضى الثلاسيميا ٦٨
4-4 مقياس تضخم التباين VIF لمتغيرات مرضى الثلاسيميا ٦٩
4-5 الجذور المميزة لمصفوفة $(X'X)$ غير المصححة ٧٠
4-6 الجذور المميزة لمصفوفة $(X'X)$ المصححة ٧١
4-7 الجذور المميزة لمصفوفة الارتباط (R)..... ٧١
4-8 ارتباط المتغيرات التوضيحية والمتغير المستجيب ٧٣
4-9 نتائج تحليل الانحدار المتعدد ٧٤
4-10 تحليل الانحدار البسيط لكل متغير ٧٥
4-11 اختيار افضل معادلة انحدار باسلوب الاختيار الامامي ٧٧
4-12 اختيار افضل معادلة انحدار باسلوب الاختيار الخلفي ٧٩
4-13 اختيار افضل معادلة انحدار باسلوب الخطوات المتسلسلة ٨١
4-14 معامل تضخم التباين والمعلمات المقدرة لقم k المختارة لمتغيرات مرضى الثلاسيميا لطريقة HKB..... ٨٢
4-15 معامل تضخم التباين والمعلمات المقدرة لقيم k المختارة لمتغيرات مرضى الثلاسيميا بطريقة Thistal ٨٣
4-16 المعلمات المقدرة لقيم k ومتوسط مربعات الخطأ ومعامل التحديد ٨٦
4-17 المعلمات المقدرة للفترة المختارة لقيم k ومتوسط مربعات الخطأ ومعامل التحديد ٨٨
4-18 معامل تضخم التباين ومقياس العدد الشرطي للبيانات القياسية في انحدار الحرف ولقيم k المختلفة ٩٢
4-19 معامل تضخم التباين ومقياس العدد الشرطي للبيانات القياسية في انحدار الحرف ولقيم k ضمن الفترة المختارة ٩٤
4-20 انحدار الخطوات المتسلسلة عندما k=0.05 ومستوى معنوية 0.01 ..... ١٠٣
4-21 انحدار الخطوات المتسلسلة عندما k=0.05 ومستوى معنوية 0.05 ..... ١٠٣

4-22 انحدار الخطوات المتسلسلة عندما k=0.05 ومستوى معنوية 0.1 ..... ١٠٤
4-23 انحدار الخطوات المتسلسلة عندما k=0.04 ومستوى معنوية 0.01 ..... ١٠٤
4-24 انحدار الخطوات المتسلسلة عندما k=0.04 ومستوى معنوية 0.05 ..... ١٠٥
4-25 انحدار الخطوات المتسلسلة عندما k=0.04 ومستوى معنوية 0.1..... ١٠٦
4-26 انحدار الخطوات المتسلسلة عندما k=0.045 ومستوى
معنوية 0.01 ..... ١٠٧
4-27 انحدار الخطوات المتسلسلة عندما k=0.045 ومستوى معنوية 0.05 ..... ١٠٧
4-28 انحدار الخطوات المتسلسلة عندما k=0.045 ومستوى معنوية 0.1 ..... ١٠٨
4-29 طريقة الاختيار الامامي عندما k=0.045 ومستوى معنوية 0.05 ..... ١٠٩
4-30 طريقة الاختيار الخلفي عندما k=0.045 ومستوى معنوية 0.05 ..... ١١٠

المقدمة والنبذة التاريخية والهدف

1-1 المقدمة Introduction

يعد علم الاحصاء بحد ذاته وسيلة وليس غاية ، وهذا يعني امكانية استخدام اصول هذا العلم وقواعده وقوانينه اينما وجد البحث العلمي سواء كان ذلك في مجال الاقتصاد والزراعة والصناعة والطب وغيرها من المجالات .

يعد تحليل الانحدار من اوسع الطرائق الاحصائية استخداماً في مختلف العلوم حيث يوضح العلاقة بين المتغيرات على هيئة معادلة يستدل من تقدير معلماتها على اهمية وقوة واتجاه هذه العلاقة كما يبين تقدير الاستجابة والتنبؤ بها بما يفيد كثيراً في التخطيط واتخاذ القرارات الرصينة حولها .(الراوي ،١٩٨٧)

قد تواجه الباحث مشكلات كثيرة اغلبها عدم توفر فروض التحليل عند استخدام طريقة المربعات الصغرى الاعتيادية ومنها مشكلة تعدد العلاقة الخطية التي تؤثر على نتائج التقديرات والاختبارات وتظهر هذه المشكلة نتيجة وجود ارتباط بين المتغيرات التوضيحية التي تؤدي الى اعطاء تقديرات ضعيفة لا يمكن ان يعول عليها كما ان تباينات هذه المقدرات تكون متضخمة وغير مقبولة وعدم قدرة طريقة المربعات الصغرى الاعتيادية (OLS) من اعطاء تقديرات جيدة عند وجود علاقة خطية بين المتغيرات التوضيحية ، وهناك العديد من الطرائق التي بواسطتها يتم الكشف عن مشكلة التعدد الخطي بين المتغيرات التوضيحية ، ومن الطرائق المهمة المستخدمة في تقدير معلمات الانحدار في حالة وجود تداخل خطي هي طريقة المربعات الصغرى المقيدة وطريقة دمج السلاسل مع البيانات المقطعية وطريقة المكونات الرئيسة وطريقة انحدار الحرف وهي اكثر الطرائق المتحيزة شيوعاً ومن المعضلات التي تعاني منها طريقة انحدار الحرف هي كيفية اختيار الثابت الذي يعد العامل المهم لهذه الطريقة وعلى اساس اختيار الثابت k يتم تقدير معلمات المتغيرات التوضيحية .

نظراً لكون إدخال أعداد كبيرة من المتغيرات التوضيحية في المعادلة يكلف جهداً ووقتاً وثمناً كثيراً وقد تحتوي معادلة الانحدار على متغيرات ليس بذي اهمية في تفسير التغيرات في متغير الاستجابة فانه من الافضل ان نختار معادلة تحوي على اقل عدد ممكن من المتغيرات التوضيحية ذات الاهمية الاحصائية في تفسير الظاهرة المدروسة ، وهنالك طريقة تعتمد على طريقة اثر الحرف (الرسم البياني) في انحدار الحرف في اختيار افضل معادلة انحدار ، وقد استخدمنا طرائق اخرى بالمقارنة مع الطريقة المتبعة لاختيار افضل معادلة انحدار هي طريقة الاختيار الامامي وطريقة الاختيار الخلفي وطريقة انحدار الخطوات

المتسلسلة للمصفوفة الممددة، ولغرض تقديم صورة واضحة عن طبيعة الرسالة ومحتوياتها قسمت إلى اربعة فصول :

تضمن الاول منه استعراض لاهم البحوث والدراسات التي تناولت موضوع تعدد العلاقة الخطية وانحدار الحرف واختيار افضل معادلة انحدار فضلاً عن تحديد هدف الرسالة.

تضمن الفصل الثاني شرحاً عن فروض النموذج الخطي العام وخواص مقدرات المربعات الصغرى الاعتيادية ، كما تضمن شرحاً عن مشكلة تعدد العلاقة الخطية بين المتغيرات التوضيحية والعديد من اساليب الكشف عن هذه المشكلة وطرائق اختيار افضل معادلة انحدار في ظل وجود مشكلة التعدد الخطي .

الفصل الثالث تضمن شرحاً عن استخدام انحدار الحرف والتقدير بطريقة انحدار الحرف وخواص مقدرات انحدار الحرف ومتوسط مربعات الاخطاء لمقدرات الحرف ومقياس متوسط مربعات الخطأ العمومي لمقدرات انحدار الحرف ، كما تضمن معلمة الحرف وطريقة اختيار معلمة الحرف ، كما تضمن اختيار افضل معادلة انحدار بالاعتماد على طريقة اثر الحرف في انحدار الحرف وفق الضوابط والقوانين .

تناول الفصل الرابع وهو الاخير نظرة في الثلاسيميا للبيانات المعتمدة في الرسالة وجمع البيانات وتحاليل احصائية لمتغيرات داخلة في النموذج ومتغيرات اخرى وتطبيق طريقة المربعات الصغرى الاعتيادية والكشف عن وجود مشكلة التعدد الخطي بالطرائق المختلفة واختيار افضل معادلة انحدار بالطرائق المختلفة ، وتضمن ايضاً اختيار الثابت k بعدة طرائق واعتماداً على طريقة اثر الحرف في تقدير المعلمات بطريقة انحدار الحرف واختيار افضل معادلة انحدار في طريقة اثر الحرف للفترة المعتمدة في اثر الحرف $0.04 \le k \le 0.05$ والقيمة المثلى $k=0.045$ المطورة في اثر الحرف لانحدار الحرف . وكما تم الاستعانة بطرائق الكشف عن وجود مشكلة تعدد العلاقة الخطية بعد التقدير بطريقة انحدار الحرف في تعيين الثابت k ومن ثم تأكيد هذا الاختيار باتباع اساليب في اختيار افضل معادلة انحدار بطريقة الاختيار الامامي وطريقة الاختيار الخلفي وطريقة انحدار الخطوات المتسلسلة للمصفوفة الممددة Augmented Matrix.

اذ تم التطبيق اعتماداً على بيانات مرضى الثلاسيميا في مستشفى ابن الاثير التعليمي للولادة والاطفال في الموصل .

1-2 نبذة تاريخية Review of Literature

لم تأخذ مشكلة تعدد العلاقة الخطية حقها في الدراسة الا في عام (١٩٣٤) عندما بدأ العالم الاحصائي النرويجي R. Frisch في دراسة تلك الظاهرة ومناقشتها عند تحليله لظاهرة السلاسل الزمنية اذ أوضح مخاطر وجود عدد كبير جداً من المتغيرات التوضيحية ، مما يتبع بالضرورة ظهور مشكلة تعدد العلاقات الخطية ، وتأثيرها في تقديرات المعالم وتبايناتها ، ولقد تبعه الكثير من الباحثين الذين درسوا الجوانب المختلفة للمشكلة ، ففي عام (١٩٥٣) اقترح Kendall استخدام طريقة المكونات الرئيسة Principal Component Method كحل للتخلص من تعدد العلاقات الخطي ، اما في عام (١٩٦٠) فقد قدم Tobin طريقة لحل مشكلة التعدد الخطي او التخفيف من حدتها عن طريق مزج بيانات السلاسل الزمنية مع بيانات القطاعات المستعرضة Pooting Time series and cross section data .

قد اقترحت طريقة المربعات الصغرى المتدرجة Stepwise least square عام (١٩٦١) من قبل Goldberger للتخلص من تأثير تعدد العلاقات الخطية ، وفي العام نفسه اوضح Johnston من خلال بحث له ، تأثير زيادة عدد المشاهدات على قيم التقديرات عند وجود علاقة خطية .

في عام (١٩٦٢) قدم Klein فكرته التي يستدل من خلالها على مدى خطورة تعدد العلاقات الخطية ، وعلى الرغم من وجود انتقادات عديدة عليها الا انها تستعمل بكثرة كونها مؤشراً بسيطاً . اذ قدم Goldberger تعريفاً في عام (١٩٦٤) عن المفهوم الضمني لتعدد العلاقات الخطية ، حيث وضح ان التعدد الخطي يعطي اخطاء معيارية كبيرة للمعلمات اذ يمكن الحصول على علاقة متطابقة بصورة جيدة بين المتغيرات أي ان معامل التحديد (R^2) يكون عالياً جداً بينما لا تكون اختبارات المعلمات مختلفة معنوياً عن الصفر . (Arthers,1970) .

أشار في عام (١٩٦٧) كل من Farrar & Glauber الى ان التعدد الخطي يحصل عندما تكون المصفوفة ($X'X$) منحرفة انحرافاً شديداً عن التعامد أي عندما تكون هنا رتبة المصفوفة اقل من رتبة عمود كامل ، كما اشار الى ان تحديد الاعتماد في المتغير التوضيحي يعكس من خلال تباينات معلمات الانحدار المتعددة (Farrar & Glauber,1967) . وبمحاولة لمعرفة تأثير العلاقات الخطية قام Fox في عام (١٩٦٨) بدراسة لمدى تأثير العلاقات الخطية في التقديرات مستخدماً معاملات الارتباط البسيطة والجزئية ، وفي نفس العام قام Silvey بدراسة لمعرفة الاسباب التي تؤدي الى الحصول على تقديرات غير دقيقة

من خلال التباينات الكبيرة بزيادة تعدد العلاقات الخطية واستخدم في الوصول الى هذه النتائج الجذور والمتجهات المميزة Characteristic Root & Characteristic Vectors .

قد اوضح Silvey في عام (١٩٦٩) ان التعدد الخطي يكون تاماً عند وجود علاقة خطية تامة بين المتغيرات التوضيحية وتكون العلاقة غير تامة عند وجود علاقة تقريبية بين المتغيرات التوضيحية (Al-Timimi , 1976) ، وفي العام نفسه اقترح Yoel Haitarsky استبدال الارتباطات المتعددة لمجموعة المتغيرات التوضيحية (Haitorsky , 1969b) .

في عام (١٩٧١) اوضح Kmenta ان التعدد الخطي يشير الى شروط المتغيرات التوضيحية التي يفترض ان تكون غير ثابتة لذلك فهو يكون بصيغة العينة وليست المجتمع (Kementa , 1971) .

في عام (١٩٧٢) بين كل من Quandt & Goldfeld ان الصفة المميزة لـ (X′X) تتضمن ان الدالة الاحتمالية (Probability Function) لها حد اعلى وحيد لذلك فان مقدر المعامل غير محدد . (Al-Timimi , 1976)

أوضح في العام (١٩٧٧) كل من Betram & Samprit ان التعدد الخطي يكون صعب الاكتشاف الى ابعد حد فهو ليس خطأ معيناً يمكن اكتشافه عن طريق استخراج باقي الانحدار وهو ليس خطأ صياغة أي انه حالة من العجز في البيانات . (سامبرت ، برايس ، ١٩٧٧) .

في عام (١٩٨٠) فقد اوضح Belsley ان التعدد الخطي يظهر بين (m) من المتغيرات التوضيحية عندما تكون المتجهات الممثلة لهذه المتغيرات مرتبطة في مجال ذي بعد اقل من (P) وفي هذه الحالة يكون احد المتجهات في الاقل عبارة عن توفيق خطي للموجهات الاخرى . (Belsely , 1980) .

في العام (١٩٨١) اوضح Suny H. Park ان التعدد الخطي بين المتغيرات في نموذج الانحدار الخطي العام له تأثير في دقة تقدير الاستجابة وان حالة التعدد الخطي التام هي نماذج ذات رتبة اقل من الكاملة .

في عام (١٩٨٤) اوضح Johnston ان زيادة الارتباط بين أي اثنين من المتغيرات التوضيحية هو دليل على زيادة القيمة العددية محدد خارج القطر وهذا ايضاً يعكس انخفاض قيمة المحدد .

فضلاً على ما ذكر سابقاً ، فان هناك دراسة اهتمت بمعالجة مشكلة التعدد الخطي عام (١٩٩٠) باسلوب المحاكاة وعلى الرغم مما يتضمنه هذا الاسلوب من مساوئ في معالجته

لهذه المشكلة (كمال ، ١٩٩٠) وهناك طرائق عديدة لمعالجة مشكلة التعدد الخطي ومنها انحدار الحرف Ridge Regression .

الاستعراض التاريخي لانحدار الحرف ففي عام (١٩٧٠) اقترح كل من Hoerl & Knnard مقدراً جديداً سمياه بـ Ridge بديلاً عن مقدرات المربعات الصغرى الاعتيادية عند وجود مشكلة التعدد الخطي وذلك باضافة مقدرا صغير موجب k الى العناصر القطرية لمصفوفة المعلومات (X′X) ، (Horel , 1970) ، وفي العام نفسه أوضح Marguardt ان طريقة Ridge هي احد طرق التقدير المتحيزة وانها تستخدم في حالة وجود متغيرات غير متعامدة .

في عام (١٩٧٤) قام Theobeld بدراسة متوسط مربعات الخطأ لمقدرات الـ Ridge على نحو عام ، وتقديم اثبات قوي حول اهمية هذه الطريقة في حالة وجود التعدد الخطي . (Theobeld , 1974) . كما استخدم عام (١٩٧٣) Mallows تقنية تعرف بـ (Cp-plot) نسبة الى اسم الباحث لمعالجة المقدرات الخطية العامة التي تكون مقدرات الحرف حالة خاصة فيها . (Mallows , 1973)

أما في عام (١٩٧٥) فقد نبه كل من Moson , Gunstالى خطورة محاولة التنبؤ بالمتغيرات في حالة وجود مشكلة تعدد العلاقة الخطية بطريقة المربعات الصغرى الاعتيادية ووجوب استعمال احد البدائل المتحيزة ، وركزوا على طريقة انحدار الحرف بوصفها الطريقة الانسب ، وفي عام (١٩٧٥) تطرق كل من Marquardt & Snee الى استعمال التقديرات المتحيزة في تحليل البيانات وبناء النموذج ، وفي العام نفسه كانت وجهة نظر كل من Mcdonald & Galarnean لطريقة انحدار الحرف على انها صنف للمعلمات المقدرة ، ويشار اليها بمقياس المعلمة k ، وان الانجاز لمقدرات الحرف المقدرة يعتمد على تباين الخطأ العشوائي والارتباطات بين المتغيرات التنبؤية ومتجه المعاملات غير المعروف ، وعد ان مقدر الحرف يقلل من مجموع مربعات الاخطاء ويتوقف ذلك على طول المقدر .

قام كل من Gunst & Mason عام (١٩٧٦) بمقارنة المربعات الصغرى الاعتيادية مع مقدرات انحدار الحرف باستخدام مقياس يدعى متوسط مربع الخطأ العمومي (Generalized mean Squared Error).

في دراسة عام (١٩٧٨) لـ Lawless انتقد طريقة المربعات الصغرى الاعتيادية ورفضها تماماً في حالة وجود تعدد علاقة خطية بين المتغيرات التوضيحية ، وعدت مقدرات الحرف بانها تملك خواصاً جيدة لمتوسط مربع الخطأ .

عرف عام (١٩٨٦) كل من Mendenhall & Sinich انحدار الحرف لبيانات عينة الانحدار التي تعاني من مشكلة تعدد العلاقة الخطية ، فان مقدرات المربعات الصغرى الاعتيادية لمعلمات الانحدار تعطي خطأ كبير (تضخم الاخطاء القياسية) وان كبر المعاملات وتغير الاشارة من عينة لاخرى يجعل مقدرات المربعات الصغرى الاعتيادية غير ثابتة ، لذلك فان التقنيات المتطورة لتثبيت معاملات الانحدار بوجود تعدد العلاقة الخطية تدعى (انحدار الحرف)، فانحدار الحرف هو تحويل خطي لطريقة المربعات الصغرى الاعتيادية وهذا التحويل يجعل المقدرات لمعاملات الانحدار متحيزة ، وفي هذه الحالة نفضل المقدرات المتحيزة على المقدرات غير المتحيزة لان ذلك سيؤدي الى تقديرات اكثر دقة لـ β ، بعبارة اخرى تكون حدود ثقة ضيقة لـ β.

أشار Sundbery عام (١٩٩٣) ، (Sundbery,1993) الى انحدار معين سماه الانحدار المتصل (Continuum regression) وعلاقة هذا الانحدار بانحدار الحرف ، اذ اثبت وضوح العلاقة المغلقة بين الانحدار المتصل لنموذج خطي بسيط من الدرجة الاولى (First – factor continuum regression) وانحدار الحرف القياسي (Standard ridge regression) وبين ان التجارب تشير الى انه مع الانحدار المتصل يمكن ان نحصل على المتنبئات التي تمتلك متوسط مربع خطأ قليل مقارنة بانحدار الحرف ولكن مع اقل حساسية لاختيار ثابت الحرف .

في عام (١٩٩٦) أشار Bjorkstrom & Sundberg في دراسة موسعة حول الانحدار المتصل ، وفي عام (١٩٩٩) قام الباحثان نفساهما بإجراء بحث تناولا فيه بنظرة عامة الانحدار المتصل وقاموا باعمام العلاقة بين الانحدار المتصل وانحدار الحرف ، حيث ان هذه العلاقة تبين بان يعرف الانحدار المتصل بواسطة انحدار الحرف يكون اكثر عمومية .

أشارت دراسات كثيرة حول تعدد العلاقة الخطية وانحدار الحرف منها (المشهداني ، ١٩٩٤) و (محمد علي ، ١٩٩٤) الى استخدام اسلوبين لتقدير معلمات الحرف ، وكذلك (سعيد، ١٩٩٦) اذ استخدم طرائق متحيزة (مقدرات المكونات الرئيسة ومقدرات الجذور الصماء ومقدرات انحدار الحرف) لمعالجة التعدد الخطي بعد الكشف عنها .

استخدم (الطالب ، ١٩٩٧) مقارنة البرمجة الهدفية مع المربعات الصغرى وانحدار الحرف ، وايضاً (شاكر ، ١٩٩٨) العلاقة بين انحدار الحرف والتحليل الذاتي لمصفوفة الارتباط المتضخمة و (القصيمي ، ٢٠٠٠) استخدم اسلوب المحاكاة في مقارنة مقدرات انحدار الحرف ، و (الكاتب ، ٢٠٠٤) استخدم تحليل الاتجاهات في التجارب العاملية وقام بمعالجة مشكلة التعدد الخطي التي صادفته بطريقة انحدار الحرف .

1-3 الهدف The Aim

لكل دراسة علمية هدف تسعى الى تحقيقه بالاستقصاء والتخطيط والتحليل كي تضاف معرفة علمية جديدة قد يستنار بها . وهذه الرسالة تسعى الى تحقيق الاهداف الآتية :

١- تعيين طرائق اكثر دقة للوصول الى قيمة معلمة انحدار الحرف k .
٢- التوسع في استخدام طريقة اثر الحرف Ridge trace (الرسم البياني) في انحدار الحرف وكيفية الاستفادة منها في تعيين المتغيرات التوضيحية ذات الاهمية الاحصائية في تأثيرها على المتغير المعتمد .
٣- كما هو معروف فان انحدار الحرف من الطرائق المتحيزة ولجأنا الى بعض الاساليب في الطرائق غير المتحيزة (المصفوفة الممددة) لتعيين تلك المتغيرات التوضيحية التي يمكن اختيارها لتكون في معادلة الانحدار .

تحليـــل الانـحـــدار

٢.١ مفهوم تحليـل الانحـدار (الـــراوي، ١٩٨٧) (الـــدليمي وكـــاظم، ١٩٨٨) (الكاتب ، ٢٠٠٤)

يعرف مفهوم الانحدار عامة بانه مقياس رياضي لمتوسط العلاقة بين المتغير التـــابع (متغير الاستجابة) Response variable ومتغير او متغيرات اخرى تســـمى المتغيـــرات التوضيحية Predictor Variable .

يختص تحليل الانحدار بوصف العلاقة بين المتغيرات على هيئة نموذج ، فقد يكـــون النموذج يحتوي على متغير توضيحي واحد فيسمى في هذه الحالة بنموذج الانحدار الخطـــي البسيط Simple Regression Liner Model ويكتب بالصيغة الآتية :

$$Y_i = \beta_0 + \beta_1 Xi + \varepsilon_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (2\text{-}1)$$

$$i = 1,2,\ldots\ldots,n$$

حيث ان :

Y_i : المتغير التابع

X_I : المتغير التوضيحي

β_0, β_1 : ثوابت وهي معلمات الانحدار

ε_i : الخطأ العشوائي

او يكون النموذج البسيط يحتوي على متغير توضيحي واحد مرفوع الى أس معـــين (درجـــة معينة r) فيسمى النموذج في هذه الحالة بنموذج انحدار بسيط من الدرجة (r) ويكتب بالصيغة الآتية :

$$Y_i = \beta_0 + \beta_1 X_i^r + \varepsilon_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (2\text{-}2)$$

$$i = 1,2,\ldots\ldots, n$$
$$r = 2,3,\ldots\ldots$$

اما في حالة كون النموذج يحتوي على متغيرات توضيحية عدة فانه يسمى بنموذج الانحدار الخطي المتعدد (Multiple Liner Regression Model) ويكتب بالصيغة الاتية:-

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ij} + \varepsilon_i \quad \ldots\ldots\ldots (2\text{-}3)$$

$$i = 1,2,\ldots\ldots, n$$

حيث ان

Y_i : متغير الاستجابة، ان n حجم العينة وان: $i=1,2,\ldots,n$

X_{ij} : المتغيرات التوضيحية ،حيث m تمثل عدد المتغيرات التوضيحية وان $j=1,2,\ldots m$

ε_i : الخطأ العشوائي

β_0, β_j : ثوابت وهي معلمات الانحدار

يمكن تلخيص اهداف تحليل الانحدار بـــ : (الدليمي وكاظم ، ١٩٨٨)

١ـ وصف البيانات *Data Description*

يعني ايجاد معادلة نموذج الانحدار التي تصف تلك البيانات ، لغرض تحليل العلاقة بين المتغيرات وتقدير معالم النموذج للمتغيرات التي تصف البيانات ، ومنها يمكن الاستدلال على اهمية العلاقة بين المتغيرات قيد الدراسة والبحث واتجاهها .

٢ـ السيطرة Control

فعند ايجاد معادلة نموذج الانحدار التي تصف البيانات فانه يمكن السيطرة على قيم المتغير المستجيب بتغير قيم المتغير التوضيحي .

٣ـ التنبؤ Prediction

ان تحليل الانحدار بالاساس يعمل على التنبؤ عند تقدير المعلمات وذلك لاتخاذ القرار والتخطيط المستقبلي بما يفيد الدراسات نحو نتائج افضل .

١ـ١ـ٢ فروض تحليل الانحدار

ان نماذج الانحدار الخطية بنيت اعتماداً على مجموعة من الفروض منها :
(الراوي ، ١٩٨٧) ، (الكاتب ، ٢٠٠٤)

١- وجود علاقة خطية بين $\hat{Y}$ و $X_1,X_2,\ldots\ldots,X_m$ تتمثل بمعادلة خط مستقيم .

٢- ان Y هو متغير عشوائي وقيمه مستقلة احصائياً الواحدة عن الاخرى ويتوزع توزيعـاً طبيعياً بوسط حسابي قدره $\mu_{Y|x_1\,x_2\ldots.x_m}$ وتباين $\sigma^2{}_{Y|x_1\,x_2\ldots.x_m}$ أي :

$$Y \sim N(\mu_{Y|x_1\,x_2\ldots.x_m}, \sigma^2{}_{Y|x_1\,x_2\ldots.x_m})$$

٣- ان ε_i هو خطأ عشوائي يتبع التوزيع الطبيعي بمتوسط حسابي $E(\varepsilon_i)=0$ وتباين ثابت مقداره σ^2_ε.

٤- ان قيمة الخطأ ε_i في فترة معينة غير مرتبطة مع قيمتها في فترة اخرى لذلك فان التباين المشترك بينهما يساوي صفراً أي :

$$Cov(\varepsilon_i \varepsilon_i') = 0$$

٥- ان قيم التنبؤ X_j ثابتة وتقاس بدون خطأ وقيمة غير مرتبطة بقيم الخطأ أي ان

$$E(X_j \varepsilon_i) = 0$$

٦- عدم وجود علاقة خطية تامة او شبه تامة بين المتغيرات التوضيحية أي غياب مشكلة تعدد العلاقة الخطية multicollinearity .

وعند توفر هذه الفروض يمكن تقدير معالم النموذج باستخدام طريقة المربعات الصغرى الاعتيادية وهي إحدى طرائق التقدير غير المتحيزة . اذ يلاحظ في معادلات نماذج الانحدار ان متغير الاستجابة والمتغيرات التوضيحية تكون متغيرات أصلية ونتيجة لحاجتنا الى تحويل المتغيرات الى الصيغة القياسية وذلك لكوننا عند استخدام طرائق التقدير وبخاصة المتحيزة منها نحتاج الى مصفوفة الارتباط وهي المصفوفة التي يمكن الحصول عليها مباشرةً عن طريق تحويل المتغيرات الى الصيغة القياسية .

٢-٢ طريقة المربعات الصغرى الاعتيادية Ordinary Last Squares Method (OLS) (الراوي ، ١٩٨٧) ، (الطالب ، ١٩٩٧)

لتقدير معالم النموذج في ظل الفروض المشار اليها ، يمكن استخدام طريقة المربعات الصغرى الاعتيادية الاتية :

$$\hat{\underline{\beta}}_{ols} = (X'X)^{-1}X'\underline{Y} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2\text{-}5)$$

اذ تعد طريقة المربعات الصغرى الاعتيادية افضل اسلوب لتوفيق خط مستقيم لعينة المشاهدات (X,Y) وهي تتضمن تصغير مجموع مربعات البواقي لانحراف النقاط الرأسية عن الخط الى ادنى حد ممكن (Robort & James,1980) , (Balestra,1970) أي :

$$\min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2\text{-}6)$$

اذ ان $\hat{Y}_i$ تمثل القيمة التقديرية لمتغير الاستجابة الحقيقي Y .

وتتصف طريقة المربعات الصغرى الاعتيادية بانها غير متحيزة ولها اقل تباين ، لذلك يطلق على مقدرات هذه الطريقة بان لها افضل تقدير خطي غير متحيز Best Linear unbiased Estimations ومختصرها (BLUE) ، وان خواص هذه المقدرات كالاتي :