الجمهورية الجزائرية الديمقراطية الشعبية
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة عمار ثليجي بالاغواط
UNIVERSITÉ AMAR TELIDJI LAGHOUAT
كلية العلوم
FACULTÉ DES SCIENCES
قسم الرياضيات
DÉPARTEMENT DE MATHÉMATIQUES

MEMOIRE DE MASTER

Domain: Mathématiques

Filiére : Mathématiques

Option : analyse foncionnelle et applications

PAR:

BEN MOHAMMED BENT DAOUD

Théme

La regression lineaire simple et multiple

Dtnave le jury composé de :

ABDESSALEM Nawel	Maitre de conférence B	Université de Laghouat	Président
MERRAD Boulrabah	Maitre assistant A	Université de Laghouat	Examinateur
GAGUI Abdel Malek	Maitre de conférence B	Université de Laghouat	Encadreur

Année Universitaire : 2024-2023

المحتويات

قائمة الأشكال

قائمة الجداول

مقدمة

الانحدار الخطي البسيط و المتعدد هو خوارزمية إحصائية تستخدم لنمذجة العلاقة بين متغيرين أو أكثر ، و قد تم ظهور الأول لمصطلح الانحدار عام 1866 ميلادي في أعمال فرانسيس غالتون ، و التي أظهرها في الانحدار الخطي البسيط ، و لكن قبل ذلك قام العالم يوسيب بوسكوفيتش بحساب معاملات الانحدار الخطي خلال أعماله في مجال الجغرافيا عام 1755 .

و تفترض هذه الطريقة أو الخوارزمية أن هناك علاقة خطية بين المتغير التابع و المتغيرات المستقلة المستخدمة في التنبؤ ، و تكمن أهمية الانحدار الخطي البسيط و المتعدد في التنبؤ بالمتغير التابع و الذي يكون أكثر دقة إذا تم استخدام أكثر من المتغيرات المستقلة ، و تكون الملاحظة أو الاختلاف بين مدى ارتباط المتغير التابع و المتغيرات المستقلة واضحة و غير متحيزة لمتغير مستقل واحد .

إن زيادة حجم البيانات الكبيرة الموضوعة في معادلة الانحدار الخطي البسيط و المتعدد تتطلب لغة برمجية أكثر استخداما و أكثر تموضعا للبيانات و في مثل هذه الحالة لغة برنامج R هي التي تفيدنا من ناحية تحليل البيانات و معالجتها في الحساب الاحصائي . و من هنا يمكننا أن نطرح تساؤلنا حول هل يمكن معرفة العلاقة الدقيقة بين المتغير التابع و المتغيرات المستقلة و كيف يتم التعامل معها في برنامج R ؟ و سنرى هذا في موضوعنا الذي يتكون من مقدمة و خاتمة و فصلين ، الفصل الأول نظري يشرح المفاهيم و التعاريف للانحدار الخطي البسيط و المتعدد و الفصل الثاني تطبيقي حيث نعمل على الانحدار الخطي المتعدد فقط بمثال توضيحي . أما الدراسة المتبعة هي دراسة تحليلية وصفية لجمع المعلومات حول الدراسة التطبيقية للانحدار الخطي البسيط و المتعدد .

كلمة شكر و عرفان

إن لله الحمد و الشكر على توفيقه و اعانته لي على تمام رسالتي العلمية ، و الصلاة و السلام على نبينا محمد صلى الله عليه و سلم ثم أما بعد :
إنني أقدم أسمى الشكر والعرفان بالجميل للدكتور قاقي عبد المالك الذي تفضّل بقبول الإشراف على رسالة الماجستير، والذي منحني من وقته الثمين ومن بحر معلوماته وخبراته الواسعة ما شكّل إضافة كبيرة للعمل البحثي، حيث كانت توجيهاته ونصائحه المنارة التي استعنت فيها في كامل عملي البحثي، فأسأل الله العزيز أن يجازيه خير الجزاء.
كما أتوجه بالشكر الجزيل على قبول مناقشة رسالة الماستر لكل أعضاء اللجنة الكريمة المؤلفة من الدكتور مراد بولرباح والدكتورة عبد السلام نوال .

الاهداء

الحمد لله حمدا كثيرا مباركا فيه و الصلاة و السلام على من لا نبي بعده صلى الله عليه و سلم و على اله و صحبه و التابعين إليهم باحسان إلى يوم الدين ثم أما بعد :
أهدي عملي إلى الوالدين الكريمين حفظهما الله و رعاهما برعايته ، و لكل من ساندني من اخوتي و صديقاتي في المسجد و الجامعة و كل أستاذ زرع في ثمرة من ثمار جهده ، و لا أنسى و لن أنسى أنسي بالقرآن الذي رافقني و باذن الله يرافقني في كل حياتي ، و في الأخير أتمنى أن يحوز عملي رضاكم و شكرا .

باب 1

مفاهيم أساسية

1.1 الإحصاء

1.1.1 تعريف الإحصاء

هو العلم الذي يهتم بوصف طرق متعددة لجمع البيانات و المشاهدات ، و من ثم يتم تنظيمها و عرضها باستخدام الاساليب العلمية لتحيلها و استخلاص النتائج منها . يشار للاحصاء إلى أنه مجموعة من الأساليب العلمية القياسية ، التي يمكن توظيفها لجمع البيانات عن الظواهر و تبويبها ، و تلخيصها و تقييمها و الخروج منها باستنتاجات حول مجموعة وحدات المجتمع من خلال اعتماد جزء صغير من هذا المجتمع . و قبل أن نتطرق إلى أن نغوص في بعض المفاهيم للاحصاء علينا أولا أن نذكر العلوم الأخرى ، التي تساهم في فهم استخدام هذه الأساليب .[1]

نظرا لإحتواء علم الاحصاء على مفاهيم عديدة من علوم أخرى كعلم الجبر و بالتحديد علم المصفوفات ، لذا وجب أولا علينا أن نتطرق لمعرفة احتياجاته منها .

2.1 المصفوفات

1.2.1 تعريف

هي عبارة عن مجموعة من البيانات مرتبة على شكل صفوف و أعمدة بشكل مستطيل ، و المصفوفة تكتب من الشكل :

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} .$$

2.2.1 المصفوفات الخاصة

1. المصفوفة الصفرية :

نقول عن المصفوفة الصفرية إذا كان جميع عناصرها مساوية للصفر .

2. المصفوفة المربعة :

نقول عن مصفوفة مربعة إذا كان عدد أسطرها n و عدد أعمدتها n .

3. المصفوفة المثلثية العلوية :

هي التي جميع عناصرها التي تقع تحت القطر الرئيسي معدومة أي:

$$a_{ij} = 0 , \qquad \forall i > j .$$

4. المصفوفة المثلثية السفلية :

هي التي جميع عناصرها التي تقع فوق القطر الرئيسي معدومة أي:

$$a_{ij} = 0 \ , \qquad \forall i < j \ .$$

5. المصفوفة القطرية :

هي التي جميع عناصرها معدومة ما عدا عناصر القطر الرئيسي أي :

$$a_{ij} = 0 \ , \qquad \forall i \neq j \ .$$

ملاحظة:

يمكن أن يكون بعض عناصر القطر الرئيسي معدوم .

6. المصفوفة الواحدية :

إذا كانت مصفوفة قطرية و جميع عناصر قطرها الرئيسي تساوي 1 و نرمز لها بالرمز : I_n .

7. مصفوفة سطر :

أو شعاع أفقي و هي التي تحتوي على سطر واحد .

8. مصفوفة عمود :

أو شعاع عمودي وهي التي تحتوي على عمود واحد .

9. تساوي مصفوفتين :

نقول إن $A = (a_{ij})$ مساوية $B = (b_{ij})$ ، إذا و فقط إذا كانا لهما عدد الأسطر m و نفس عدد الأعمدة n ، و العناصر المتقابلة متساوية أي : $a_{ij} = b_{ij}$ ، من أجل $1 \leq i \leq m$ و $1 \leq j \leq n$.

3.2.1 عمليات على الموصفوفات

(ا) منقول مصفوفة:

نحصل على منقول مصفوفة بتبديل صفوفها بأعمدتها , أي بجعل صفها الأول مكان العمود الأول و صفها الثاني مكان

العمود الثاني وهكذا ، فإن كان لدينا المصفوفة

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

فإن مبدول هذه المصفوفة ، A ويرمز لها بالرمز A^T يكون على الصورة الأتية :

$$A^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

خواص : لتكن A و B ، مصفوفتين لهما نفس عدد الأسطر و نفس الأعمدة ، و ليكن λ ، عددا حقيقيا أو مركبا . لدينا الخواص التالية :

i. $(A+B)^T = A^T + B^T$

ii. $(AB)^T = B^T A^T$

iii. $(\lambda A)^T = \lambda A^T$

vi. $(A^T)^T = A$

(ب) المقلوب :

لتكن A ، مصفوفة مربعة $n \times n$ ، نقول أن المصفوفة A ، عكوسة أو قابلة للقلب inversible إذا وجدت مصفوفة مربعة B ، $n \times n$ حيث : $AB = BA = I_n$ عندئذ نسمي B . مقلوب أو عكوس A ، inverse و نرمز له بالرمز A^{-1} .

خواص :

- لتكن $A , B \in \mathcal{M}_n(\mathbf{R})$ إذا كان A و B ،عكوسين فان BA ، عكوسة و لدينا : $(AB)^{-1} = B^{-1}A^{-1}$.
- لتكن $A \in \mathcal{M}_n(\mathbf{R})$ ، إذا كانت A ، عكوسة فإن A^T عكوسة و لدينا : $(A^T)^{-1} = (A^{-1})^T$.

4.2.1 قواعد اشتقاق المصفوفات

الاشتقاق بالنسبة للمتغير x

1. $\frac{\partial A(x)}{\partial x} = \left[\frac{\partial a_{ij}(x)}{\partial x}\right]$

2. $\frac{\partial}{\partial x}(A \pm B) = \frac{\partial A(x)}{\partial x} \pm \frac{\partial B(x)}{\partial x}$

3. $\frac{\partial}{\partial x}(\alpha * A) = \alpha \frac{\partial A(x)}{\partial x}$

4. $\frac{\partial (A*B)}{\partial x} = \frac{\partial A(x)}{\partial x} * B + A \frac{\partial B(x)}{\partial x}$

الاشتقاق بالنسبة للمصفوفة ذاتها

1. $\frac{\partial}{\partial X}(X^T * A * X) = 2AX$

2. $\frac{\partial}{\partial X}(C^T * X) = C$

3. $\frac{\partial |A|}{\partial A}() = |A| * (A^{-1})^T$

4. $\frac{\partial}{\partial A}[T_r|A^T M A|] = MA + M^T A$

3.1 بعض المقاييس الاحصائية

إن الاحصاء الوصفي هو أحد أنواع الاحصاء و هو العلم الذي يساعد في تصنيف و تلخيص و عرض البيانات ، و يتميز بمبادئ الاحصائية و التي تساعد في وصف النزعة المركزية و التشتت .

4.1 مقاييس النزعة المركزية

إن كلمة النزعة المركزية تعني الرغبة في التمركز و التكثف نحو رقم معين ، و هي إحدى الطرق الاحصائية التي بواسطتها نحاول الحصول على رقم واحد ، يمثل مجموعة من الأرقام أو البيانات كبديل عنها أو ممثلا لها ، و قد وجد باحثوا الاحصاء العديد من هذه المقاييس منها :

1.4.1 المتوسط الحسابي :

يعرف في الاحصاء بأنه القيمة التي تتجمع حولها مجموعة قيم ، و يرمز له بالرمز $\bar{X}$ ، و يعطى بالعلاقة :

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) .$$

ملاحظة : عند استخدام التكرار المطلق في حالة متغير كمي منفصل و متصل فيكون المتوسط الحسابي كمايلي :

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} ,$$

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i c_i}{\sum_{i=1}^{k} n_i} .$$

حيث : c_i هو مركز الفئات .

أما في حالة استخدام التكرار النسبي فيكون حساب المتوسط الحسابي كمايلي :

$$\bar{X} = \sum_{i=1}^{k} f_i x_i ,$$

$$\bar{X} = \sum_{i=1}^{k} f_i c_i .$$

1.5 مقاييس التشت

قبل الخوض في أهم مقاييس التشتت نرى لزما توضيح فكرة التشتت و إعطاء معنى واضح للتشتت .

معنى التشتت بشكل عام : هو التباعد القيم عن بعضها أو تقاربها عن وسطها الحسابي و لعدم تجانس البيانات في بعض الأوقات لذا تم الموافقة على أن يكون هناك نقطة ثابتة لقياس التباعد او التقارب عن هذه النقطة و وجد أن الوسط الحسابي خير ممثل لهذه النقطة . و لقياس التشتت عدة طرق احصائية منها : المدى ، الانحراف المتوسط ، التباين و الانحراف المعياري .

1.5.1 التباين :

هو عبارة عن متوسط الحسابي لمربعات الفروق بين قيم المتغير الاحصائي و متوسطها الحسابي ، ويرمز له بالرمز : σ^2 او $var(x)$

حساب التباين

1. في حالة بيانات أولية (غير مبوبة)

$$\sigma^2 = \frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2 .$$

2. في حالة بيانات مبوبة

$$\sigma^2 = \frac{1}{\sum_{i=1}^{k} n_i}\left(\sum_{i=1}^{k} n_i (X_i - \bar{X})^2\right) = \frac{1}{\sum_{i=1}^{k} n_i}\left(\sum_{i=1}^{k} n_i X_i^2\right) - \bar{X}^2 .$$

2.5.1 التباين المشترك أو التغاير :

هو مقياس لكمية تغير متغيرين مع بعضهما ، حيث تكون قيمة التغاير موجبة عندما يتغير المتغيران مع بعضهما البعض ، و العكس إذا تغير أحد المتغيرين فوق قيمته المتوقعة و بينما الآخر يكون دونها فهذا تغاير قيمته سالبة ، و يعرف بالعلاقة :

$$cov(X,Y) = \frac{1}{n}\left(\sum_{i=1}^{k}(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)\right) .$$

3.5.1 الانحراف المعياري :

هو الجذر التربيعي للتباين و يعتبر من أهم المقاييس الاحصائية للتشتت و أكثرها استخداما و يرمز له بالرمز σ .

6.1 التوزيعات الاحتمالية و الاختبارات

7.1 التوقع العشوائي

هو عبارة عن أخذ جميع النتائج الممكنة لذلك المتغير حيث :

$$E(X) = \mu .$$

و تمثل القيمة المتوقعة لمتغير عشوائي X ، للمجتمع .

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx .$$

1.7.1 الاختبار ستودنت T

ترجع نشأة اختبار T ، إلى العالم ستودنت ، ويستخدم لقياس دلالة الفروق بين المتغيرات المرتبطة و المستقلة للعينات الإحصائية.

شروط اختبار T

- مستوى قياس المتغير التابع كمي (نسبي أو فئة) .
- المعاينة العشوائية ، أي استخدام أسلوب العشوائي في اختيار العينات.
- استقلالية المشاهدات ، يعني كل فرد من أفراد العينة لا يرتبط إلا بمجموعة واحدة ، وليس له تأثير على الأفراد الأخرى.

- التوزيع الإعتدالي للمتغير التابع.
- تجانس العينات أو تماثل تشتت درجات العينات.

1.7.2 اختبار فيشر F

هو اختبار احصائي يستخدم لمقارنة المتغيرات.

شروط اختبار F

- يجب أن تتبع البيانات التوزيع الطبيعي.
- يجب أن يكون عدد المتغيرات المراد مقارنتها أكثر من متغيرين.

خطوات اختبار F

- تحديد فرضيات العدم و البديل و هما :

$$H_0 : \nu_1 = \nu_2 = \cdots = \nu_k \ ,$$

$$H_1 : \nu_1 \neq \nu_2 \neq \ldots \neq \nu_k \ .$$

- تحديد مستوى المعنوية $\alpha = 0.05$, $\alpha = 0.01$.
- ايجاد قيمة F المحسوبة.
- ايجاد قيمة f الجدولية من جدول f .
- مقارنة قيمة F المحسوبة مع f الجدولية .
- انشاء جدول التحليل التباين .

باب 2

الانحدار الخطي البسيط

1.2 الانحدار

مفهوم الانحدار :

هو ايجاد معادلة رياضية تعبر عن العلاقة بين المتغيريين y و x ، تستعمل للتنبؤ عن قيم سابقة و قيم مستقبلية ل y أو x ، حسب المعلوم منهما و تكون هذه المعادلة الرياضية خطية , و قد تكون بدرجة ثانية أو ثالثة .

1.1.2 إستعمالات الانحدار

يستعمل الانحدار لأغراض منها :[3]

1. وصف البيانات حيث يمكن تلخيص و وصف البيانات لدى الباحث بايجاد معادلة الإنحدار التي تصف بين المتغيرات المستقلة و المتغير التابع .
2. تقدير المعاملات يمكن تقدير المعملات المجهولة التي تصف البيانات و منها يمكن تقدير إتجاه و قوة العلاقة بين المتغيرات .
3. التنبؤ يمكن تقدير أو التنبؤ بالاستجابة للمتغير أو المتغيرات المستقلة و هذا يفيد كثيرا في اتخاذ القرارات اللازمة أو التخطيط المستقبلي .
4. السيطرة قد تستخدم النماذج الإنحدارية لغرض السيطرة حيث يمكن السيطرة على قيم المتغير التابع بتغير قيم المتغيرات المستقلة عند ايجاد المعادلة التي تصف البيانات .

2.1.2 أهداف الإنحدار

1. تقدير العلاقة بين متغيرين على شكل علاقة دالية $y = f(x)$ ، و التي عن طريقها يمكن معرفة التغير في أحد المتغيرين على أساس تأثره بالمتغير الآخر.
2. قياس مدى الارتباط الكلي بين المتغير التابع و المتغير المستقل .
3. تقدير نسبة تفسير كل متغير مستقل لاختلاف المتغير التابع . معرفة إتجاه تأثير المتغير المستقل على المتغير التابع ، هل هو تأثير إيجابي أم سلبي من خلال قيمة اشارة (b) ، و دلالتها الاحصائية .

3.1.2 أنواع الانحدار

1. الانحدار البسيط : يستخدم لدراسة العلاقة بين متغيرين متغير مستقل و متغير تابع .
2. الانحدار الجزئي : يدرس العلاقة بين المتغير التابع و واحد فقط من المتغيرات المستقلة بفرض أن العوامل الأخرى ثابتة .
3. الانحدار المتعدد : يدرس العلاقة بين المتغير التابع و المتغيرات المستقلة كلها .

2.2 تعريف الانحدار الخطي البسيط

هو أسلوب رياضي قدمه العالم جالتون بهدف الاستفادة من الارتباط في التنبؤ أو تقدير قيمة متغير ما ، فقد وجد من خلال بحثه حول وراثة طول القامة ، لأن الاطفال الذين يأتون من أباء طوال القامة يملون لأن يكونوا أقصر القامة من آبائهم ، و الذين يأتون من أباء قصار القامة يميلون لأن يكونوا أطول قامة من آبائهم ، أي أن طول الأبناء يميل إلى التراجع (الانحدار) نحو المتوسط العام ، لذا أطلق على هذه العلاقة اسم قاعدة الانحدار ، كما أطلق على الخط الذي يوضح هذه العلاقة اسم خط الانحدار ، فإذا كان لدينا اختبارين و قمنا بقياس العلاقة الارتباطية لهما ، فإنه من خلال معادلة الانحدار يمكننا أن نتنبأ بدرجة كل فرد في الختبار الثاني بناءا على درجاته في الاختبار الأول ، وقد سمي هذا الأسلوب الاحصائي بالانحدار لأنه يخدر في تقديره للدرجات المختلفة نحو المتوسط ، كما أنه يسمى بالإنحدار الثنائي ذلك أنه يتكون من متغيرين احدهما متغير مستقل و الآخر تابع .

3.2 شروط الانحدار الخطي البسيط

[4]

- العشوائية في اختيار العينة و استقلالية درجة كل فرد من الأفراد الآخرين في العينة المختارة .
- أن يكون المتغيرين المستقل و التابع مصنفين ضمن المقاييس الكمية (النسبي أو فئة) .
- التوزيع الاعتدالي (الطبيعي) لدرجات المتغيرين المستقل و التابع .
- وجود علاقة خطية بين المتغيرين المستقل و التابع .
- أن يكون تباين المتغيرين المستقل و التابع .
- أن يكون تباين المتغير المستقل أكبر من الصفر $(\sigma^2 > 0)$ ، و الغرض من هذا أن يسهم المتغير المستقل في تفسير التباين في درجة المتغير التابع .
- أن يكون متوسط البواقي أو الأخطاء العشوائية $(Y - \hat{Y})$ ،يساوي صفرا 0 و تباين يساوي σ^2 ، و الأخطاء العشوائية هي الفرق بين القيمة الحقيقية الفعلية (Y) ، و القيمة التقديرية أو المتنبأ بها $(\hat{Y})$ ، و تعرف هذه الأخطاء العشوائية البواقي .
- أن تكون الأخطاء العشوائية (البواقي) موزعة توزيعا طبيعيا (اعتداليا) .

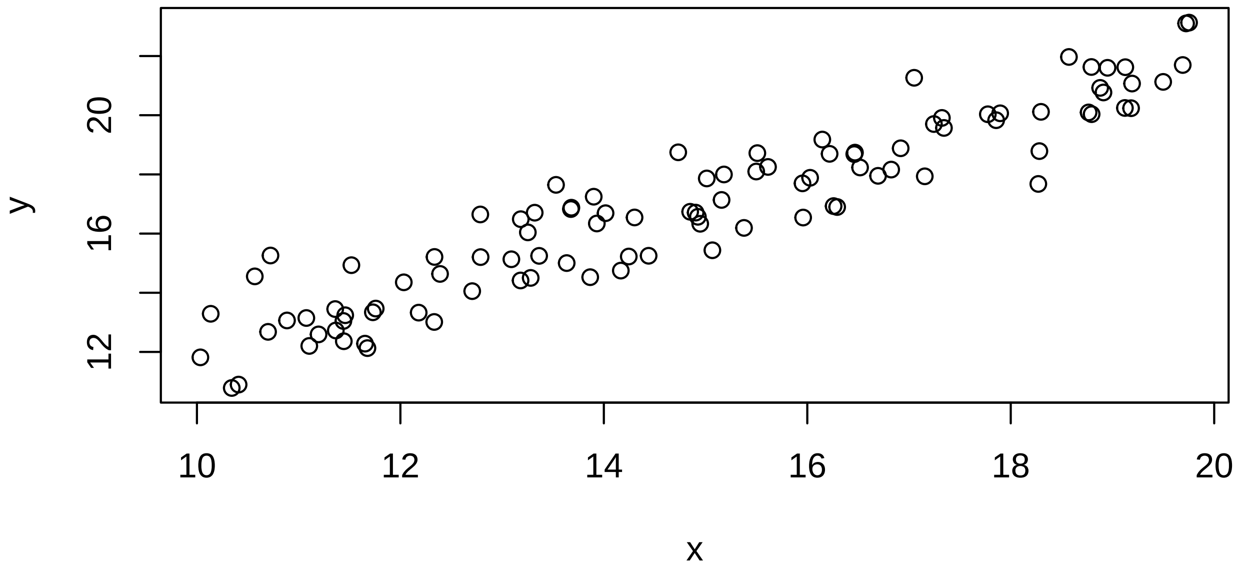
4.2 حساب الانحدار الخطي البسيط

- معادلة انحدار Y على X .
- معادلة انحدار X على Y.

هناك طريقتين رئيستين تتبعان في حساب معادلة الانحدار الخطي البسيط :

1.4.2 طريقة شكل الانتشار

قبل حساب معادلة الانحدار الخطي البسيط ، يتوجب علينا تمثيل نقاط كل من المتغيرين X و Y في رسم بياني ، من أجل معرفة إتجاه الدالة التي تأخذها قيم المتغيرين ، سواءا كانت خطية أو غير خطية (أسية ، لوغارتمية ، ... الخ .) ، و بناءا على شكل الإنتشار يمكن تحديد أفضل نموذج لحساب العلاقة بين المتغيرين (النموذج الخطي أو النموذج غير خطي) .



شكل 1.2: شكل الانتشار لبعض الدوال

2.4.2 طريقة المربعات الصغرى (SLO)

إن المستقيم المتشكل في شكل الإنتشار ، يسمى بمستقيم الإنحدار و الذي يربط بين المتغير المستقل و المتغير التابع ، لذا فإن معادلة الإنحدار الخطي البسيط للمجتمع تكتب من الشكل :

$$Y_i = \alpha + \beta X_i \, .$$

و لكن في كثير من الأحيان يصعب التعامل مع قيم المجتمع الاحصائي , لذا نستعين بالعينة و نقدر معاملي الإنحدار المجتمع (α) و

(β) بأفضل معاملي لتقديرهما بنقطة هما (a) و (b) على التوالي ، و بهذا تكون معادلة الانحدار : الخطي البسيط للعينة هي

$$Y_i = a + bX_i \; .$$

كما أنه من غير المتوقع أن تقع جميع نقاط المتغير المستقل و المتغير التابع على خط المعادلة فإننا نلجأ إلى إضافة الأخطاء العشوائية أو البواقي فتصير معادلة الانحدار كمايلي :

$$Y_i = a + bX_i + e_i \; .$$

حيث :

- Y_i : تمثل قيم المتغير التابع للمشاهدة i .
- X_i : تمثل قيم المتغير المستقل للمشاهدة i .
- a : تمثل قيم Y عندما $X = 0$ ، و تسمى ثابت الانحدار .
- b : تسمى ميل الانحدار و تدل على مقدار التغير المقدرة للمتغير Y .
- e_i : تمثل البواقي او الاخطاء العشوائية و هي تساوي :

$$e_i = (Y_i - \hat{Y}) \; .$$

حساب معادلة الانحدار الخطي البسيط وفق طريقة المربعات الصغرى :

1. لتقدير معادلة انحدار Y على X ، فهي تساوي :

$$\hat{Y} = a + bX \; ,$$

حيث (a) ، يمكن حسابها من خلال إحدى الطرق التالية :

- طريقة 1:

$$a = \bar{Y} - b\bar{X} \; ,$$

• طريقة 2 :

$$a = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b(\sum_{i=1}^{n} X_i)\right) .$$

و (b) ، يمكن حسابها من خلال إحدى الطرق التالية :

• طريقة 1:

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - (n\bar{X}\bar{Y})}{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2} ,$$

• طريقة 2 :

$$b = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{nX^2 - \left(\sum_{i=1}^{n} X_i\right)^2} ,$$

• طريقة 3 :

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - \left(\frac{1}{n}\left(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i\right)\right)}{X^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)} ,$$

• طريقة 4 :

$$b = \frac{1}{\sigma^2}\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i Y_i)\right) - \bar{X}\bar{Y}\right) ,$$

• طريقة 5 :

$$b = \frac{1}{n\sigma^2}\left(\sum_{i=1}^{n}(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)\right) ,$$

• طريقة 6 :

$$b = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_i)(X_i - \bar{Y}_i)}{\sum_{i=1}^{n}(X_i - \bar{X}_i)^2} \ .$$

2. لتقدير معادلة انحدار X على Y ، فهي تساوي :

$$\hat{X} = a + bY \ ,$$

حيث (a) ،يمكن حسابها من خلال إحدى الطرق التالية :

• طريقة 1:

$$a = \bar{X} - b\bar{Y} \ ,$$

• طريقة 2 :

$$a = \frac{1}{n}\left(\sum_{i=1}^{n} X_i - b\left(\sum_{i=1}^{n} Y_i\right)\right) \ .$$

و (b) ، يمكن حسابها من خلال إحدى الطرق التالية :

• طريقة 1:

$$b = \frac{\sum_{i=1}^{n} Y_i X_i - (n\bar{X}\bar{Y})}{\sum_{i=1}^{n} Y_i^2 - n(\bar{Y})^2} \ ,$$

• طريقة 2 :

$$b = \frac{n\sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i}{nY^2 - \left(\sum_{i=1}^{n} Y_i\right)^2} \ ,$$

- طريقة 3 :

$$b = \frac{1}{\left(Y^2 - \frac{(X)^2}{n}\right)} \left(\sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i \right) \right) ,$$

- طريقة 4 :

$$b = \frac{1}{\sigma^2} \left(\left(\frac{1}{n} \sum_{i=1}^{n} (YX) \right) - \bar{Y}\bar{X} \right) ,$$

- طريقة 5 :

$$b = \frac{1}{n\sigma^2} \left(\sum_{i=1}^{n} (Y_i - \bar{X}_i)(X - \bar{Y}) \right) ,$$

- طريقة 6 :

$$b = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)(Y - \bar{X})}{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2} .$$

حساب معادلة الانحدار الخطي وفق طريقة معامل الارتباط :

1. لتقدير معادلة انحدار Yعلى X ، فهي تساوي :

$$\hat{Y} = r \times \frac{\sigma_Y}{\sigma_X}(X - \bar{X}) + \bar{Y} .$$

حيث :

- r: معامل الارتباط بيرسون بين المتغيرين (X و Y).
- σ_Y:الانحراف المعياري للمتغير (Y) .
- σ_X:الانحراف المعياري للمتغير (X) .
- $\bar{X}$:المتوسط الحسابي للمتغير (X) .
- $\bar{Y}$:المتوسط الحسابي للمتغير (Y) .

ذلك ان :

$$b = r \times \frac{\sigma_Y}{\sigma_X} \; .$$

و

$$a = \bar{Y} - b\bar{X} \; .$$

2. لتقدير معادلة انحدار X على Y ، فهي تساوي :

$$\hat{X} = r \times \frac{\sigma_X}{\sigma_Y}(Y - \bar{Y}) + \bar{X} \; .$$

حيث :

- r: معامل الارتباط بيرسون بين المتغيرين (X و Y).
- σ_Y:الانحراف المعياري للمتغير (Y) .
- σ_X:الانحراف المعياري للمتغير (X) .
- $\bar{X}$:المتوسط الحسابي للمتغير (X) .
- $\bar{Y}$:المتوسط الحسابي للمتغير (Y) .

ذلك أن :

$$b = r \times \frac{\sigma_X}{\sigma_Y} \; .$$

و

$$a = \bar{X} - b\bar{Y} \; .$$

حساب معادلة الانحدار الخطي البسيط بالاعتماد على الدرجات المعيارية :

1. لتقدير معادلة انحدار Yعلى X ، فهي تساوي :

$$Z_{\hat{Y}} = rZ_X \; .$$

حيث :

- r: معامل الارتباط بيرسون بين المتغيرين (X و Y).
- $Z_{\hat{Y}}$:قيمة Y المعيارية المتنبا بها .
- Z_X:الدرجات المعيارية للمتغير (X) .

2. لتقدير معادلة انحدار X على Y ، فهي تساوي :

$$Z_{\hat{X}} = rZ_Y \ .$$

حيث :

- r: معامل الارتباط بيرسون بين المتغيرين (X و Y).
- Z_Y:قيمة Y المعيارية المتنبا بها .
- $Z_{\hat{X}}$:الدرجات المعيارية للمتغير (X) .

تفسير قيمة معامل الانحدار :

يمكن تفسير قيمة معامل الانحدار ، المتحصل عليها من خلال المعادلة التالية

$$\hat{Y} = a + bX \ .$$

وفق الحالات التالية :

- إذا كانت قيمة معامل الانحدار موجبة يعني هنا أنه كلما كانت زيادة في المتغير المستقل X ، كلما تبعه زيادة في المتغير التابع Y بقيمة b .
- أما إذا كانت قيمة معامل الانحدار سالبة فهذا يعني أنه إذا كانت زيادة في المتغير المستقل X ، كلما تبعه نقصان في المتغير التابع Y بقيمة b .
- أما إذا كانت قيمة معامل الانحدار معدومة (يساوي الصفر) فإنه في هذه الحالة لا يوجد تأثير للمتغير المستقل X على المتغير التابع Y .

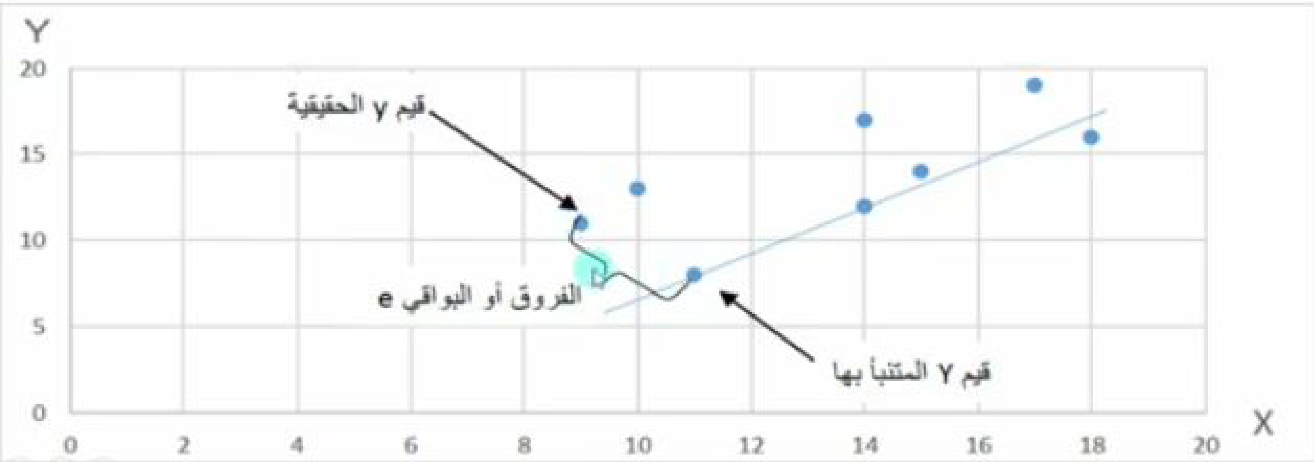
البواقي أو الأخطاء العشوائية :

هي إنحراف كل قيمة حقيقية للمتغير التابع $(\hat{Y}_i)$ ، عن القيمة التقديرية لها (Y_i) ويرمز للبواقي بالرمز (e_i) ، و يتم حسابها من خلال العلاقة التالية :

$$e_i = (Y_i - \hat{Y})$$

حيث متوسط البواقي يساوي الصفر أما تباين البواقي يتم حسابه كما يلي :

$$\sigma_e^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{n-2} \, .$$



شكل 2.2: شكل يبين الفروق أو البواقي

5.2 معامل التحديد

هو عبارة عن النسبة بين التغير في القيم المقدرة على التغير الكلي ، و يرمز له بالرمز R^2 ، و يطلق عليه بمربع معامل الارتباط و يتم حسابه في الحالات التالية :

1. في حالة معادلة انحدار Y على X ، يتم حسابه وفق الطرق التالية :

- طريقة 1:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} ,$$

- طريقة 2 :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} ,$$

- طريقة 3 :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{Y_i^2} ,$$

- طريقة 4 :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{Y_i^2} .$$

ثم القيمة المتحصل عليها يتم ضربها في 100% كي نتحصل على نسبة تفسير تباين درجات المتغير التابع .

2. في حالة معادلة انحدار X على Y ، يتم حسابه وفق الطرق التالية :

- طريقة 1:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{X}_i - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \ ,$$

- طريقة 2 :

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (X_i - \hat{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \ ,$$

- طريقة 3 :

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (X_i - \hat{X})^2}{X_i^2} \ ,$$

- طريقة 4 :

$$R^2 = 1 - \frac{1}{X_i^2} \sum_{i=1}^{n} (\hat{X}_i - \bar{X})^2 \ .$$

ثم القيمة المتحصل عليها يتم ضربها في 100% كي نتحصل على نسبة تفسير تباين درجات المتغير المستقل .

3. كما يمكن حساب معامل التحديد إذا تحصلنا على قيمة (b) ، فإنه يمكننا حساب معامل التحديد من خلال العلاقة التالية :

$$R^2 = b_{Y/X} \times b_{X/Y} \ .$$

- أما نسبة التباين الغير مفسر (غير متنبأ به) يتم حسابه من خلال العلاقة التالية :

$$1 - R^2 \ .$$

- و حيث كلما زادت R^2 ، كلما اقتربت القيم الحقيقية (Y_i) ، في مطابقة الفيم التقديرية $(\hat{Y}_i)$ ، و بالتالي تقل مسافات ابتعاد Y_i ، عن خط الانحدار .

6.2 العلاقة بين معاملي الانحدار و الارتباط الخطي البسيط

إن معامل الانحدار الخطي البسيط و معامل الارتباط الخطي يسيران في نفس الاتجاه ، و اشارة ميل خط الانحدار هي نفسها اشارة معامل الارتباط الخطي فإذا كانت العلاقة بين المتغيرين طردية فاشارتهما موجبة و إذا كانت العلاقة عكسية فاشارتهما سالبة .

1.6.2 حساب معامل الارتباط :

إذا تحصلنا على قيمة (b) ، فإنه يمكننا حساب معامل الارتباط من خلال العلاقة التالية

$$r = \sqrt{b_{Y/X} \times b_{X/Y}} \ .$$

و تكون اشارة معامل الارتباط (r) هي اشارة جداء كل من $(b_{X/Y})$ و $(b_{Y/X})$ ، حيث تدل الاشارة على اتجاه العلاقة الارتباطية موجبة طردية او سالبة عكسية .

2.6.2 الخطأ المعياري للتقدير :

هو الخطأ الذي يمكن ان يقع فيه الباحث عند محاولته لتقدير قيم المتغير التابع (Y) و انحراف القيم عن خط المعادلة ، حيث هناك احتمال لأن تكون قيمة $(\hat{Y})$ المتنبا بها مساوية لقيمة (Y) الحقيقية، و يمكن حساب هذا الخطأ في عملية التقدير كما يلي :

1. في حالة معادلة انحدار Y على X ، يمكن حسابه من خلال الطرق التالية :

- في حالة التعامل مع القيم الخام

$$\sigma_{YX} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \ ,$$

او

$$\sigma_{YX} = \sqrt{\frac{Y^2 - a\sum_{i=1}^{n} Y - b\sum_{i=1}^{n} XY}{n-2}} \ .$$

- في حالة التعامل مع معامل الارتباط

$$\sigma_{YX} = \sigma_Y \sqrt{1-(r)^2} \ .$$

حيث r هو معامل الارتباط بين المتغيرين (X و Y) ، σ_Y هو الانحراف المعياري للمتغير (Y)

- في حالة التعامل مع الدرجات المعيارية

$$\sigma_{YX} = \sqrt{1-(r)^2} \, .$$

2. في حالة معادلة انحدار X على Y يمكن حسابه من خلال الطرق التالية :

- في حالة التعامل مع القيم الخام

$$\sigma_{XY} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\hat{X})^2}{n-2}} \, ,$$

او

$$\sigma_{XY} = \sqrt{\frac{X^2 - a\sum X - b\sum_{i=1}^{n} YX}{n-2}} \, .$$

- في حالة التعامل مع معامل الارتباط

$$\sigma_{XY} = \sigma_X \sqrt{1-(r)^2} \, .$$

حيث r هو معامل الارتباط بين المتغيرين (X و Y) ، σ_Y هو الانحراف المعياري للمتغير (Y)

- في حالة التعامل مع الدرجات المعيارية

$$\sigma_{XY} = \sqrt{1-(r)^2} \, .$$

7.2 العلاقة بين التنبؤ و الارتباط

كثيرا ما يتبادر إلى أذهاننا تساؤل مفاده ما قدرة المتغير المستقل (X) على التنبؤ بدرجات و قيم المتغير التابع (Y) ، و للاجابة عن هذا التساؤل فإننا نلجأ إلى حساب العلاقة الارتباطية بين المتغيرين و بالتالي يمكن للمتغير المستقل (X) التنبؤ بدرجات و قيم التابع (Y) بدرجة قوية ، و إذا كانت قيمة العلاقة الارتباطية ضعيفة بينهما ، فانه يمكن للمتغير المستقل (X) التنبؤ بدرجات و قيم المتغير

التابع (Y) بدرجة ضعيفة .

1.7.2 العلاقة بين الخطأ المعياري للتقدير و معامل الارتباط

هناك علاقة عكسية بين معامل الارتباط و الخطأ المعياري للتقدير ، حيث كلما زادت قيمة معامل الارتباط ، فإن مجموع مربع الانحرفات الخطا المعياري سيكون قليل ، و هذا يعني أنه كلما كانت العلاقة الارتباطية للمتغيرين (X و Y) قوية كلما كانت هناك إحتمالية للتنبؤ بقيم المتغير (Y) بأقل خطأ .

2.7.2 الدلالة الاحصائية للانحدار الخطي البسيط

من أجل معرفة الدلالة الاحصائية للانحدار الخطي البسيط يجب علينا تطبيق اختبار (T) ، و معرفة مدى تأثير المتغير المستقل (X) على المتغير التابع (Y) و ذلك من خلال حساب قيمة (T) المحسوبة و مقارنتها بقيمة (t) الجدولية عند مستوى الدلالة $\alpha = 0.05$ و تطبيق اختبار (F) المسمى بتحليل التباين (ANOVA) لمعرفة مدى قدرة نموذج الانحدار على التنبؤ بقيم المتغير التابع (Y) ، وكذا اختبار الدلالة الاحصائية لمعمل التحديد ، و ذلك من خلال حساب قيمة (F) المحسوبة و مقارنتها بقيمة (f) الجدولية عند مستوى الدلالة $\alpha = 0.05$ و ذلك كما يلي :

1. تطبيق اختبار (T) : يستخدم هذا الاختبار لمعرفة الدلالة الاحصائية لتأثير المتغير المستقل (X) على المتغير التابع (Y) ، حيث مع هذا الاختبار يتم اختبار الفرضيات التالية :

   - الفرضية الصفرية (H_0) : $H_0 : b = 0$
   - الفرضية البديلة (H_1) : $H_1 : b \neq 0$

   أي الفرضية (H_0) تنص على أن المعامل (b) للمتغير المستقل ليس له تأثير على المتغير التابع .
   أما الفرضية البديلة (H_1) تنص على أن المعامل (b) للمتغير المستقل له تأثير على المتغير التابع .
   و يتم حساب قيم (T) المحسوبة من خلال الطرق التالية :

   (ا) الطريقة الاولى : $T = \frac{b}{\sigma_b}$ حيث :

$$\sigma_b = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\left(\sum_{i=1}^{n} X^2 - n(\bar{X})^2\right) \times (n-2)}},$$

   (ب) الطريقة الثانية : $T = \frac{b}{\sigma_b}$ حيث :

$$\sigma_b = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{(n-2)(\sigma_X)^2 \times n \times (n-1)}},$$

(ج) الطريقة الثالثة :

$$T = \frac{b\sigma_X\sqrt{n-1}}{\sqrt{\frac{n\sigma_Y^2(1-r^2)}{(n-2)}}} \,,$$

حيث :

- (b) : معامل الانحدار .
- (n) : حجم العينة .
- (r) : معامل الارتباط بين المتغيرين (Y و X) .
- (σ_X) : الانحراف المعياري للمتغير X .
- (σ_Y^2) : تباين المتغير Y.

(د) الطريقة الرابعة : لاختبار الدلالة الاحصائية للمعامل a فيكون :

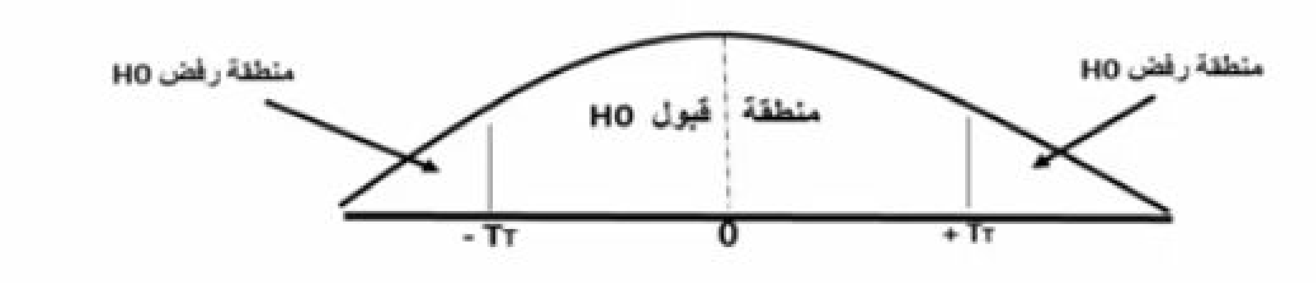
$$T = \frac{a}{\sqrt{\frac{1}{n-2}\left(\sum_{i=1}^{n}(Y_i-\hat{Y})^2\right)\times\frac{1}{nX_i^2}\left(\sum_{i=1}^{n}X_i^2\right)}} \,.$$

و لاختبار الدلالة الاحصائية للمعامل b فيكون :

$$T = \frac{b}{\sqrt{\frac{1}{n-2}\left(\sum_{i=1}^{n}(Y_i-\hat{Y})^2\right)\times\frac{1}{X_i^2}}} \,.$$

فإذا كانت القيمة المطلقة ل(T) المحسوبة أكبر من أو تساوي قيمة (t) الجدولية عند درجة الحرية (n-2) عند مستوى الدلالة $\alpha = 0.05$ و باختبار بطرفين (لأن الفرضية البديلة تنص على $(H_1 : b \neq 0)$ ، أما إذا كانت الفرضية البديلة تنص على $(H_1 : b > 0)$ او $(H_1 : b < 0)$ فيكون الاختبار بطرف واحد) فاننا نرفض الفرض الصفري (H_0) و نقبل بالفرض البديل (H_1) و نقول إن معامل الانحدار دال احصائيا أي أن المتغير المستقل (X) له تأثير على المتغير التابع (Y) و يمكننا التنبؤ بدرجات المتغير (Y) من خلال درجات المتغير (X) ، أما إذا كانت القيمة المطلقة ل(T) المحسوبة اقل تماما من قيمة (t) الجدولية فإننا نقبل الفرض الصفري (H_0) و نرفض الفرض البديل (H_1) و نقول أن معامل الانحدار غير دال احصائيا أي أن المتغير المستقل (X) ليس له تأثير على المتغير التابع (Y) ، و لا يمكننا التنبؤ بدرجات المتغير (Y) من خلال درجات المتغير (X).

2. تطبيق اختبار (F) : يستخدم هذا الاختبار لمعرفة مدى صلاحية نموذج الانحدار و قدرته على التنبؤ بقيمة المتغير التابع (Y) حيث يتم اختبار الفرضيات التالية :



شكل 3.2: اختبار الفرضيات

- الفرضية الصفرية (H_0) : نموذج الانحدار ليس له الصلاحية و قدرة التنبؤ .
- الفرضية البديلة (H_1) : نموذج الانحدار له الصلاحية و قدرة التنبؤ .

و ذلك من خلال حساب قيمة (F) المحسوبة و مقارنتها بقيمة (f) الجدولية عند مستوى الدلالة $\alpha = 0.05$ و معرفة مدى صلاحية نموذج الانحدار للتنبؤ ، فإذا كانت (F) المحسوبة أكبر أو تساوي (f) الجدولية ، فإننا نقول ان نموذج الانحدار له صلاحية و القدرة على التنبؤ بقيم المتغير التابع (Y) ، أما إذا كانت (F) المحسوبة أقل تماما من (f) الجدولية ، فإننا نقول أن نموذج الانحدار ليس له صلاحية أو القدرة على التنبؤ بقيم المتغير التابع (Y) ، و يتم حساب قيمة (F) المحسوبة من خلال العلاقة التالية :

التباين الكلي = تباين الانحدار + تباين الأخطاء العشوائية

أي أن

التباين الكلي = التباين المفسر + التباين الغير المفسر

أي أن

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2 \ .$$

و يتم حساب قيمة (F) كما في الحالات التالية :

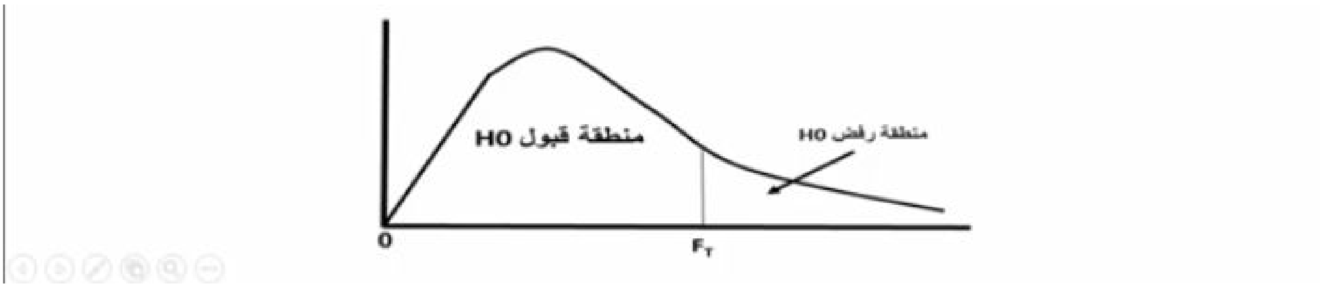
(ا) في حالة معادلة انحدار Y على X يمكن حسابها كما في الجدول التالي : و يتم استخراج (F) الجدولية عند درجة حرية البسط $(df = 1)$ و درجة حرية المقام $(df = n - 2)$ عند مستوى الدلالة $\alpha = 0.05$

(ب) في حالة معادلة انحدار X على Y و يمكن حسابها كما في الجدول التالي :

و يتم استخراج (F) الجدولية عند درجة حرية البسط $(df = 1)$ و درجة حرية المقام $(df = n - 2)$ عند مستوى الدلالة $\alpha = 0.05$

F المحسوبة	متوسط المربعات	درجات الحرية	مجموع المربعات	مصدر التباين
$\frac{MSR}{MSE}$	$MSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	df=1	$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	الانحدار Regression
	$MSE = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{n-2}$	df=n-2	$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$	الخطا Error
		df=n-1	$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	التباين الكلي Total

جدول 1.2: حساب قيمة F



شكل 4.2: اختبار الفرضيات

8.2 الدلالة العلمية للانحدار الخطي البسيط

حساب حجم التأثير للانحدار الخطي البسيط يكون من خلال حساب معامل التحديد (R^2) وفق الحالات التالية :

1. في حالة معادلة انحدار Y على X يمكن حسابها وفق القانون التالي :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} .$$

2. في حالة معادلة انحدار X على Y و يتم تفسير قيمة حجم التأثير المتحصل عليها من قيمة معامل التحديد وفق الحالات التالية :

9.2 مجال (فترة) التنبؤ

إن قيمة $(\hat{Y})$ المقدرة تساعدنا في تحديد فترة أو مجال التنبؤ للقيمة Y الحقيقية ، و يمكن تحديد فترة التنبؤ وفق الحالات التالية :

1. في حالة معادلة انحدار Y على X يتم حسابها وفق العلاقة التالية :

$$\hat{Y} - E \leq Y \leq \hat{Y} + E .$$

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F المحسوبة
Regression الانحدار	$SSR=\sum_{i=1}^{n}(\hat{X}_i-\bar{Y})^2$	df=1	$MSR=\sum_{i=1}^{n}(\hat{X}_i-\bar{X})^2$	$\frac{MSR}{MSE}$
Error الخطا	$SSE=\sum_{i=1}^{n}(X_i-\hat{X})^2$	df=n-2	$MSE=\frac{\sum_{i=1}^{n}(X_i-\hat{X}_i)}{n-2}$	
Total البتاين الكلي	$SST=\sum_{i=1}^{n}(X_i-\bar{X})^2$	df=n-1		

جدول 2.2: حساب قيمة F

$0.25\leq R^2$	$0.25>R^2\geq 0.09$	$0.09>R^2\geq 0.01$	R^2 معامل التحديد
كبير	متوسط	صغير	حجم التاثير (ES)

جدول 3.2: تفسير قيمة حجم التأثير المتحصل عليها من قيمة معامل التحديد

حيث :

- (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E=t_{\alpha/2}\times\sigma_{YX}\sqrt{1+\frac{1}{n}+\frac{n(X-\bar{X})^2}{n(X^2)-(\sum X)^2}}\ .$$

- $(t_{\alpha/2})$: قيمة (T) الجدولية التي يتم استخراجها من القيم الحرجة لاختبار T بطرفين (ذيلين) عند درجة حرية $(n-2)$ و مستوى الدلالة المعنوية $\alpha=0.05$
- (σ_{YX}) : الخطأ المعياري .
- n : حجم العينة .
- X: قيمة المتغير المستقل (X) التي يتم التنبؤ بواسطتها .
- $(\bar{X})$: المتوسط الحسابي للمتغير المستقل (X) .

2. في حاالة معادلة انحدار X على Y يتم حسابها وفق العلاقة التالية :

$$\hat{Y}-E\leq Y\leq\hat{Y}+E\ .$$

حيث :

- (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E = t_{\alpha/2} \times \sigma_{XY} \sqrt{1 + \frac{1}{n} + \frac{n(Y-\bar{Y})^2}{n(Y^2) - (\sum Y)^2}} \, .$$

- $(t_{\alpha/2})$: قيمة (T) الجدولية التي يتم استخراجها من القيم الحرجة لاختبار T بطرفين (ذيلين) عند درجة حرية $(n-2)$ و مستوى الدلالة المعنوية $\alpha = 0.05$
- (σ_{XY}) : الخطأ المعياري .
- n : حجم العينة .
- Y: قيمة المتغير التابع (Y) التي يتم التنبؤ بواسطتها .
- $(\bar{Y})$: المتوسط الحسابي للمتغير التابع (Y) .

01.2 الخطأ المعياري لمعاملي الانحدار (a) و (b)

[4]

إن ثابت انحدار العينة (a) و ميل انحدار العينة (b) سيتراوحان حول القيم الحقيقية للمجتمع الاحصائي (α)و (β) على التوالي ، ولقياس مقدار انحراف كل من ثابت (a) و ميل انحدار (b) العينة عن ثابت (α) و ميل المجتمع (β) نلجأ إلى حساب الخطأ المعياري لهما وفق الطرق التالية :

1. في حالة معادلة انحدار Y على X يمكن حسابه من خلال الطرق التالية :

- الطريقة الأولى :

$$\sigma_a = \sigma_{YX} \times \sqrt{\frac{1}{n} + \frac{(\bar{X})^2}{\sigma_X^2(n-1)}} \, ,$$

و

$$\sigma_b = \sigma_{YX} \times \sqrt{\frac{1}{\sigma_X^2(n-1)}} \, ,$$

حيث :

- (σ_{YX}) : الخطأ المعياري للتقدير و يتم حسابه كما يلي :

$$\sigma_{YX} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{n-2}} \ ,$$

- (σ_X^2): تباين المتغير المستقل (X) و يتم حسابه كمايلي :

$$\sigma_X^2 = \frac{\sum X^2 - n(\bar{X})^2}{n-1} \ .$$

- الطريقة الثانية :

$$\sigma_b = \frac{\sigma_{YX}}{\sqrt{X^2 - \frac{(\sum X)^2}{n}}} \ ,$$

- الطريقة الثالثة :

$$\sigma_b = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{(n-2)\times(\sigma_X^2)\times n \times (n-2)}} \ ,$$

- الطريقة الرابعة :

$$\sigma_b = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2 \times (n-2)}} \ .$$

2. في حالة معادلة انحدار X على Y يمكن حسابه من خلال الطرق التالية :

- الطريقة الأولى :

$$\sigma_a = \sigma_{XY} \times \sqrt{\frac{1}{n} + \frac{(\bar{Y})^2}{\sigma_Y^2(n-1)}} \ ,$$

و

$$\sigma_b = \sigma_{XY} \times \sqrt{\frac{1}{\sigma_Y^2(n-1)}} \,,$$

حيث :

- (σ_{XY}) : الخطأ المعياري للتقدير و يتم حسابه كما يلي :

$$\sigma_{XY} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X})^2}{n-2}} \,,$$

- (σ_Y^2): تباين المتغير التابع (Y) و يتم حسابه كمايلي :

$$\sigma_Y^2 = \frac{\sum Y^2 - n(\bar{Y})^2}{n-1} \,.$$

• الطريقة الثانية :

$$\sigma_b = \frac{\sigma_{XY}}{\sqrt{Y^2 - \frac{(\sum Y)^2}{n}}} \,.$$

11.2 تقدير معاملي انحدار المجتمع α و β بفترة الثقة

[4]

إن معاملي انحدار العينة (a) و (b) هما تقدير لمعاملي انحدار المجتمع α و β بنقطة ، و لكن إذا أردنا تقدير معاملي انحدار المجتمع بفترة ثقة ، فإنه يكون وفق العلاقة التالية :

1. في حاالة معادلة انحدار Y على X يمكن حسابه من خلال الطرق التالية :

(ا) تقدير معامل α من خلال العلاقة التالية :

$$a - E \leq \alpha \leq a + E \,.$$

حيث : (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E = t_{\alpha/2} \times \sigma_{YX} \times \sqrt{\frac{1}{n} + \frac{(\bar{X})^2}{\sigma_X^2(n-1)}} \ .$$

(ب) تقدير معامل β من خلال العلاقة التالية :

$$b - E \leq \beta \leq b + E \ .$$

حيث : (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E = t_{\alpha/2} \times \sqrt{\frac{1}{\sigma_X^2(n-1)}} \ .$$

- $(t_{\alpha/2})$: قيمة (T) الجدولية التي يتم استخراجها من القيم الحرجة لاختبار T بطرفين (ذيلين) عند درجة حرية $n-2$ و مستوى الدلالة المعنوية $\alpha = 0.05$
- (σ_{YX}) : الخطأ المعياري .
- $(\sigma_X)^2$ تباين المتغير التابع (X).
- n : حجم العينة .
- $(\bar{X})$: المتوسط الحسابي للمتغير المستقل (X) .

2. في حاالة معادلة انحدار X على Y يمكن حسابه من خلال الطرق التالية :

(ا) تقدير معامل α من خلال العلاقة التالية :

$$a - E \leq \alpha \leq a + E \ .$$

حيث : (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E = t_{\alpha/2} \times \sigma_{YX} \times \sqrt{\frac{1}{n} + \frac{(\bar{Y})^2}{\sigma_Y^2(n-1)}} \ .$$

(ب) تقدير معامل β من خلال العلاقة التالية :

$$b - E \leq \beta \leq b + E \ .$$

حيث : (E) : هامش الخطأ و يتم حسابه من القانون التالي :

$$E = t_{\alpha/2}\sigma_{XY} \times \sqrt{\frac{1}{\sigma_Y^2(n-1)}}\,.$$

- $(t_{\alpha/2})$: قيمة (T) الجدولية التي يتم استخراجها من القيم الحرجة لاختبار T بطرفين (ذيلين) عند درجة حرية $n-2$ و مستوى الدلالة المعنوية $\alpha = 0.05$
- (σ_{YX}) : الخطأ المعياري .
- $(\sigma_Y)^2$ تباين المتغير التابع (Y).
- n : حجم العينة .
- $(\bar{Y})$: المتوسط الحسابي للمتغير المستقل (Y) .

مثال فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع ، ومقدار الزيادة في وزن العجل بالكيلو غرام ، وذلك لعينة من العجول الرضيعة حجمها . **المطلوب**

الطاقة	10	11	14	15	20	25	46	50	59	70
الانتاج	20	16	15	19	13	13	12	12	10	10

جدول 4.2: كمية البروتين لعجل رضيع

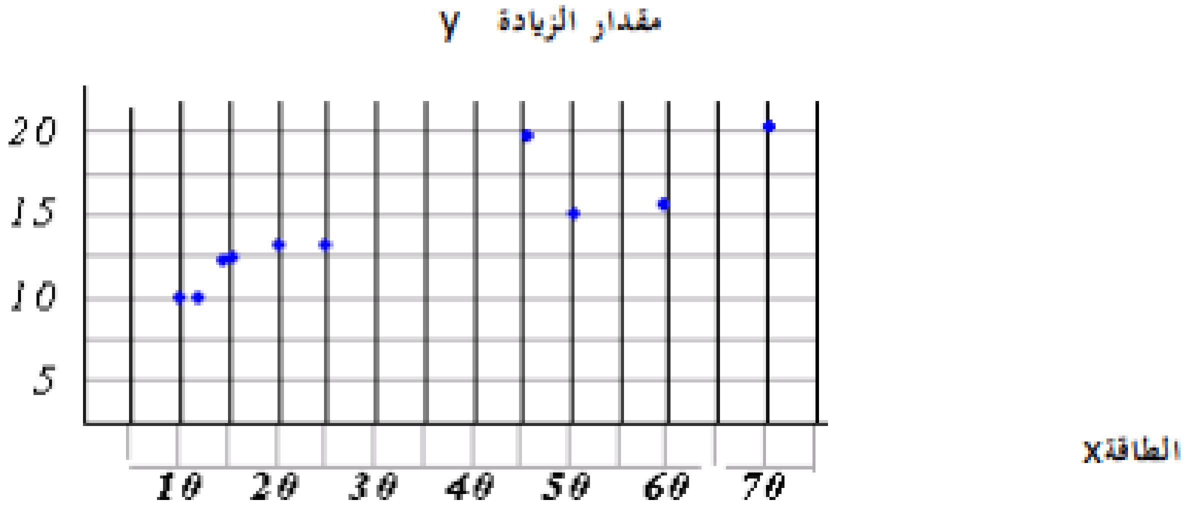
1. أرسم نقط الانتشار ، و ماهو توقعاتك لشكل العلاقة ؟ .
2. قدر معادلة انحدار الوزن على كمية البروتين .
3. فسر معادلة نموذج الانحدار
4. ماهو مقدار الزيادة في الوزن عند اعطاء العجل 05غرام من البروتين ؟ .وماهو مقدار الخطأ العشوائي ؟

الحل

1. رسم نقط الانتشار : من المتوقع أن يكون لكمية البروتين أثر طردي (ايجابي) على مقدار الزيادة في الوزن .
2. تقديم معادلة الانحدار بفرض أن X هي كمية البروتين ، و هي Y مقدار الزيادة في الوزن لدينا:

- لدينا:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{140}{10} = 14 \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{320}{10} = 32$$



شكل 5.2: شكل الانتشار

X	Y	XY	X^2
10	10	100	100
11	10	110	121
14	12	168	196
15	12	180	225
20	13	260	400
25	13	325	625
46	19	874	2116
50	15	750	2500
59	16	944	3481
70	20	1400	4900
320	140	5111	14664

جدول 5.2: حساب YX و X^2

$$b=\frac{n\sum_{i=1}^{n}X_iY_i-\sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}=\frac{10)(5111)(320)(140)}{(10)(14664)-(320^2)}=\frac{6310}{44240}=0.1426\ .$$

- حساب قيمة a :

$$a=\bar{Y}-b\bar{X}=14-(0.1426)(32)=9.4368\ .$$

- إذن معادلة الانحدار هي :

$$\hat{Y}=9.44+0.143X\ .$$

3. تفسير الاحصائي لنموذج الانحدار :

- معامل الانحدار : $b = 0.143$ يدل على أنه كلما زادت كمية الطاقة بوحدة واحدة ، يؤدي ذلك إلى زيادة في الوزن بمقدار 341.0 كيلوغرام .
- الثابت $a = 9.44$:يدل على أنه في حالة عدم استخدام البروتين في التغذية ، فإن الوزن يزيد ب44.9 غرام .

4. مقدار الزيادة في الوزن عند $X = 50$ هو :

$$\hat{Y} = 9.44 + 0.143(50) = 16.59 \ .$$

أما مقدار الخطأ العشوائي هو

$$e_{\hat{X=50}} = Y_{X=50} - \hat{Y_{X=50}} = 15 - 16.59 = -1.59 \ .$$

باب 3

الإنحدار الخطي المتعدد

1.3 تعريف

إن الانحدار الخطي المتعدد هو تعميم للانحدار الخطي البسيط ، و هو يشير إلى العلاقة الخطية التي تربط بين المتغير التابع (Y) واحد و المتغيرات المستقلة عديدة $(X_1, X_2, \ldots, X_k)$ و يمكن التعبير عن هذه العلاقة بالمعادلة الأتية :

$$Y_i = a + b_1 X_{1j} + b_2 X_{2j} + \cdots + b_k X_{kj} + e_i \ .$$

$$Y_i = a + \sum_{i=1}^{n} b_i X_{ij} + e_i \ .$$

حيث أن :

- $(a, b_1, b_2, \ldots, b_k)$ تعبر عن معاملات الانحدار .
- (e_i) يعبر عن الخطأ العشوائي أو البواقي .
- n : عدد المشاهدات .

لذا يكون لدينا n من المعادلات يمكن صياغتها في صورة مصفوفات :

$$Y_1 = a + b_1 X_{11} + b_2 X_{12} + \cdots + b_k X_{k1} + e_1 \ ,$$

$$Y_2 = a + b_1 X_{12} + b_2 X_{22} + \cdots + b_k X_{k2} + e_2 \ ,$$

$$\vdots$$

$$Y_n = a + b_1 X_{1n} + b_2 X_{2n} + \cdots + b_k X_{kn} + e_n \ .$$

و لتسهيلها نضعها في مصفوفة كمايلي :

$$\begin{pmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X12 & \dots & X_{1k} \\ 1 & X_{21} & X22 & \dots & X_{2k} \\ \vdots & \vdots & & & \\ 1 & X_{n1} & Xn2 & \dots & X_{nk} \end{pmatrix} \times \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} .$$

و نعبر عنها ايضا ب:

$$Y = XB + a + e_i \ .$$

حيث X تمثل مصفوفة المتغيرات المستقلة و هي من الدرجة $(n \times (k+1))$.

2.3 أهمية الانحدار الخطي المتعدد

تكمن أهميته في معرفة أكثر المتغيرات المستقلة تأثيرا على المتغير التابع ، حيث إذا كنا بصدد دراسة مجموعة متغيرات مستقلة و مدى تأثيرها على المتغير التابع ، و نود معرفة ترتيب المتغيرات المستقلة في تأثيرها على المتغير التابع ، بدءا بالمتغيرات الأكثر تأثيرا فالأقل فإننا نستعين بالانحدار الخطي المتعدد الذي من خلال حساب قيمة (b) و معرفة اشارتها و اختبار دلالتها الاحصائية ، فإنه يمكننا معرفة مدى تأثير كل متغير مستقل على التابع و إتجاه هذا التأثير هل هو سلبي أم تأثير إيجابي ، حيث إذا كانت قيمة (b) موجبة دل هذا على تأثير موجب للمتغير المستقل على المتغير التابع ، أما إذا كانت قيمة (b) سالبة دل هذا على سلبي للمتغير المستقل على المتابع .

3.3 الافتراضات الموضوعة على نموذج الانحدار الخطي المتعدد

تصنف هذه الافتراضات إلى نوعين هيكلية و عشوائية

1. الهيكلية : و هي الشروط التي يجب أن تتوفر في المتغيرات المستقلة و المتغير التابع .

- أن تكون المتغيرات المستقلة مستقلة عن بعضها البعض .
- ان تكون العلاقة بين Y و المتغيرات السنقلة $(X_1, X_2, \dots, Xp)$
- إن المتغيرات المستقلة تكون متغيرات غير عشوائية و يكون لها قيم ثابتة $cov(X, e_i) = E$

$$(X^T e_i) - [E(X)]^T [E(e_i)] = 0 \ .$$

- إن المتغير التابع (Y) يكون دالة خطية في المتغيرات المستقلة .

2. العشوائية : و هي الشروط الموضوعة على الخطأ العشوائي e و هي :

(ا) أن يكون التوقع الرياضي لكل e_i معدوما .

$$E(e_i) = 0 \ .$$

(ب) أن يكون تباين e_i مساويا لمقدار ثابت σ^2 أي أن :

$$var(e_i) - E(e_i - 0)^2 = E(e_i^2) = \sigma^2 \ .$$

(ج) أن يكون التباين المشترك لأي حدين مختلفين e_k e_l معدوما أي أن يكون :

$$cov(e_k, e_l) = E(e_k, e_l) = 0 \ , \qquad k \neq l \ .$$

(د) يمكن دمج الشرطين (ب) و (ج) في شرط واحد و كتابتهما ضمن مصفوفة التباين المشترك لـe كمايلي :

$$v = cov(e) = E[(e-0)(e-0)^T] = E(e * e^T) = \sigma^2 I_n \ ,$$

$$v = cov(e) = E(e * e^T) = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} * \begin{pmatrix} e_1 & e_2 & \dots & e_n \end{pmatrix} \ ,$$

$$= \begin{pmatrix} E(e_1e_1) & E(e_1e_2) & \dots & E(e_1e_n) \\ E(e_2e_1) & E(e_2e_2) & \dots & E(e_2e_n) \\ \vdots & \vdots & \dots & \vdots \\ E(e_ne_1) & E(e_ne_2) & \dots & E(e_ne_n) \end{pmatrix} \ . \quad (1)$$

و بناءا على الافتراضات السابقة نجد

$$v = cov(e) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 * I_n \ .$$

أي أن المصفوفة التباينات المشتركة لـe هي مصفوفة قطرية و عناصر قطرها الرئيسي متساوية و تساوي σ^2.

4.3 تقدير معلمات النموذج بطريقة المربعات الصغرى

[6]

لتقدير معلمات النموذج نسحب عينة من عناصر المجتمع بحجم n ، ونأخذ منها القياسات المتقابلة لكل من Y و المتغيرات $(X_1, X_2, \dots, X_p)$ ثم نفترض ان $(b_0, b_1, b_2, \dots, b_p)$ هي تقديرات معينة للمعلمات $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ في المجتمع و بذلك نجد أنه يمكننا حساب قيمة المشاهدة Y_i و لكن بخطأ آخر و هو خطأ التقدير و نكتب ذلك كمايلي :

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_p X_{ip} + e_i \ , \qquad i = 1, \dots, n \ . \tag{2}$$

و إذا أخذنا جميع المشاهدات فإنه يمكننا كتابتها مصفوفيا كمايلي :

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \dots & \vdots & \\ i & X_{i1} & X_{i2} & \dots & X_{ip} \\ \vdots & \vdots & \dots & \vdots & \\ i & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \\ \vdots \\ e_n \end{pmatrix} \ . \tag{3}$$

و باستخدام الرموز المصفوفية يمكننا كتابة النموذج السابق كمايلي :

$$Y = X * b + e \ , \tag{4}$$

و ذلك بوضع التقدير يساوي $X * b$ و نكتب

$$\hat{Y} = X * b \ , \tag{5}$$

و من العلاقتين(3) و (4) نحصل على أن Y الفعلية تساوي و نجد أن شعاع البواقي يساوي

$$e = Y - \hat{Y} = Y - Xb \ , \tag{6}$$

و أنه بالمقابل القيم (i) يكون لدينا الخطأ يساوي

$$e_i = Y_i - \hat{Y}_i \ , \qquad i = 1, \ldots, n \ . \tag{7}$$

حيث أن $\hat{Y}_i$ هي ناتج جداء السطر i من المصفوفة X في العمود b و هو يساوي

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_p X_{ip} \ . \tag{8}$$

و علينا الآن ايجاد قيم المعلمات $(b_0, b_1, b_2, \ldots, b_p)$ بحيث يكون مجموع مربعات الأخطاء التقديرية في العينة أصغر ما يمكن أي بحيث يكون المقدار

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \min \ (Y, \hat{Y}) \ , \tag{9}$$

و الذي يمكن كتابته مصفوفيا كمايلي

$$Q = e^T * e = [(Y_i - Xb)^T (Y - Xb)] = \min \ (Y, \hat{Y}) \ , \tag{10}$$

و لحساب عناصر الشعاع b نشتق Q بالنسبة لـb (حسب قواعد الاشتقاق في المصفوفات) و نضعه مساويا للصفر فنجد أن :

$$\partial Qb = \left[\partial b (Y - Xb)^T\right] (Y - Xb) + (Y - Xb)^T \left[\partial b (Y - Xb)\right] \ , \tag{11}$$

$$\partial Qb = [-X^T](Y - Xb) + (Y - Xb)^T[-X] = 0 \ , \tag{12}$$

$$\partial Qb = -X^T(Y - Xb) - (Y - Xb)^T X = 0 \ . \tag{13}$$

و لكن بما أن الجداء الأخير $(Y - Xb)^T X$ هو شعاع من الرتبة (1+p) لذلك فانه يساوي منقوله . أي يكون لدينا مايلي :

$$(Y - Xb)^T X = [(Y - Xb)^T X]^T = X^T (Y - Xb) \ , \tag{14}$$

(31) و بالتعويض في المشتق الأخير :

$$\partial Qb = -X^T(Y - Xb) - X^T(Y - Xb) = -2X^T(Y - Xb) = 0 \ , \tag{15}$$

$$- X^T Y + XX^T b - X^T Y + XX^T b = 0 \ , \tag{16}$$

و بعد الاصلاح نحصل على المعادلة التالية :

$$X^T Y = X^T X b \ , \tag{17}$$

أو على المعادلة

$$X^T X b = X^T Y \ , \tag{18}$$

فإذا كانت المصفوفة $X^T X$ نظامية (محددها غير معدوم $|X^T X| \neq 0$) فإن مقلوبها يكون موجودا و عندها نضرب من اليسار طرفي المعادلة ب$(X^T X)^{-1}$ فنحصل على

$$(X^T X)^{-1} * XX^T b = (X^T X)^{-1} * X^T Y \ , \tag{19}$$

$$I * b = (X^T X)^{-1} * X^T Y \ , \tag{20}$$

و منها نجد أن b يساوي

$$b = (X^T X)^{-1} * X^T Y \ , \tag{21}$$

و منها يمكننا حساب عناصر الشعاع b باجراء مطابقة بين عناصر الطرفين ، ثم نعوض b في العلاقة (81) فنحصل على الصغية المقدرة لتابع الانحدار التالية

$$\hat{Y} = Xb \ . \tag{22}$$

و لدراسة جودة تمثيل النموذج نحلل مجموع مربعات الانحرافات التابع Y كمايلي :

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i+\hat{Y}_i-\bar{Y})^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2+0\,. \tag{23}$$

و نرمز لهذه المجاميع بالرموز المعروفة التالية :

$$SST=SSE+SSR\,. \tag{24}$$

5.3 معامل التحديد R^2

$$R^2=\frac{SSR}{SST}=\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=1-\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=1-\frac{SSE}{SST}\,. \tag{25}$$

وهو يشير إلى نسبة تأثير كل المتغيرات المستقلة على المتغير التابع .

1.5.3 خواص معامل التحديد R^2

1. من معادلة تحليل التباين يكون لدينا $0\leq R^2\leq 1$

2. يمكن تمييز بين حالتين

(ا) إذا اقتربت قيمة R^2من الصفر فهو دلالة على ضعف العلاقة بين المتغير التابع و المتغيرات المستقلة ، و إذا كان $R^2=0$ فلا وجود للعلاقة بين المتغير التابع و المتغيرات المستقلة الداخلة في النموذج . لذلك كلما اقتربت قيمة R^2 من الواحد الصحيح كلما زادت درجة الثقة في النموذج المقدر .

(ب) اذا كان$R^2=1$ فإن ذلك يشير إلى أن (100%) من التغير في المتغير التابع يعود إلى التغير في التغيرات المستقلة المدرجة في النموذج .

3. يعاب على المعامل R^2 أنه لا يأخذ في الاعتبار عدد المشاهدات و عدد المتغيرات .

2.5.3 معامل التحديد المعدل $\bar{R}^2$

إن إضافة متغير مستقل أو أكثر إلى النموذج المقدر يؤدي إلى زيادة في قيمة R^2 نتيجة لزيادة قيمة البسط في معادلة R^2 و بقاء المقام على حاله دون تغيير ، و بالتالي قيمة R^2 لا تعبر عن القيمة الحقيقية له و لتفادي هذا الإشكال أقترح مؤشر آخر يأخذ بعين الاعتبار درجة الحرية في النموذج و يسمى معامل التحديد المعدل $\bar{R}^2$ حيث

$$\bar{R}^2 = 1 - (\frac{n-1}{n-k-1})(1-R^2) . \tag{26}$$

1. في العينات الصغيرة $k > l$ فان

$$\bar{R}^2 \leq R^2 .$$

أما في العينات الكبيرة

$$\lim_{n+\to\infty} (\frac{n-1}{n-k-1}) = 1 \Rightarrow \bar{R}^2 \approx R^2 .$$

2. يمكن أن تنخفض قيمة R^2 و ذلك بادخال متغيرات خارجية جديدة في النموذج .

3. يمكن أن يأخذ $\bar{R}^2$ قيمة سالبة .

6.3 العلاقة بين R^2 و F

$$F_c = (\frac{n-k-1}{k})(\frac{R^2}{1-R^2}) . \tag{27}$$

يمكن تحديد طبيعة العلاقة بين R^2 و F من خلال بناء جدول تحليل التباين (ANOVA)

F_c	MS	df	SS	Source
$F_c = (\frac{n-k-1}{k})(\frac{R^2}{1-R^2})$	$R^2Y^TY/(k)$	k	$SSE = \hat{Y}^T\hat{Y}$	Regression
	$(1-R^2)Y^TY/(n-k-1)$	n-k-1	$SSR = \hat{e}^T\hat{e}$	Residuel
		n-1	$SST = Y^TY$	Total

جدول 1.3: يوضح جدول (ANOVA) العلاقة بين R^2 و F

1.6.3 الخطأ المعياري للتقدير

هو الخطا الذي يمكن أن يقع فيه الباحث عند محاولته لتقدير قيم المتغير التابع (Y) و انحراف القيم عن خط المعادلة ، حيث يمكن حسابه من خلال العلاقة التالية :

$$\sigma_{Y.X1.X_2...Xk} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}} .$$

2.6.3 حساب معامل الارتباط المتعدد

يمكن حسابه وفق الطريقة التالية :

$$r_{Y.X_1.X_2...X_k} = \sqrt{\left(\beta_1 \times \frac{\sigma_{X_1}}{\sigma_Y} \times r_{X_1Y}\right) + \left(\beta_2 \times \frac{\sigma_{X_2}}{\sigma_Y} \times r_{X_2Y}\right) + \cdots + \left(\beta_i \times \frac{\sigma_{X_i}}{\sigma_Y} \times r_{X_iY}\right)} .$$

و لحساب الدلالة الاحصائية لمعامل الارتباط فإننا نستخدم اختبار (F) من خلال حساب قيمة (F) المحسوبة و هو نفس قانون المطبق في معامل التحديد :

$$F = \frac{r^2(n-k-1)}{k(1-r^2)} .$$

حيث :

- k : تمثل عدد المتغيرات المستقلة .
- n : حجم العينة .
- r^2 : مربع معامل الارتباط المتعدد .

و يتم استخراج (f) الجدولية عند درجة حرية البسط $(df = k)$ و درجة حرية المقام $(df = n-k-1)$ عند مستوى الدلالة $\alpha = 0.05$ و يتم مقارنة قيمة (F) المحسوبة مع قيمة (f) الجدولية ، فإذا كانت (F) المحسوبة أكبر من أو تساوي (f) الجدولية ، فان معامل الارتباط المتعدد يكون دال احصائيا ، و العكس صحيح .

3.6.3 تقدير معاملات الانحدار (β_i) بفترة ثقة بنسبة 95%

إذا أردنا تقدير معاملات الانحدار بفترة الثقة فانه يكون وفق العلاقة التالية :

$$\beta_i - E \geq \beta_i \geq \beta_i + E .$$

حيث :

- E هامش الخطأ و يتم حسابه وفق :

$$E = t_{\alpha/2} \times \sigma_{\beta_i} \ .$$

حيث :

- $t_{\alpha/2}$: قيمة (t) الجدولية التي يتم استخراجها من القيم الحرجة لاختبار (T) بطرفين عند درجة حرية $(n-k-1)$ و مستوى الدلالة المعنوية $\alpha = 0.05$.
- σ_{β_i} : الخطأ المعياري للانحدار .

4.6.3 الدلالة الاحصائية للانحدار الخطي المتعدد

من أجل معرفة الدلالة الاحصائية للانحدار الخطي المتعدد يجب علينا تطبيق اختبار (F) أو تحليل التباين (AVONA) لمعرفة الدلالة الاحصائية لمعاملات $(\beta_1, \beta_2, \dots, \beta_k)$ و كذا الدلالة الاحصائية لمعامل التحديد ، و ذلك من خلال حساب قيمة (F) المحسوبة و مقارنتها بقيمة (f) الجدولية عند مستوى الدلالة $\alpha = 0.05$ ثم تطبيق اختبار (T) لمعرفة مدى تأثير كل معامل على المتغير التابع (Y) ، و ذلك من خلال حساب (T) المحسوبة و مقارنتها بقيمة (t) الجدولية عند مستوى الدلالة $\alpha = 0.05$

1. تطبيق اختبار (F): [5] اختبار الفرضيات التالية :

- الفرضية الصفرية (H_0) : $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$
- الفرضية البديلة (H_1) : $H_1 : \beta_i \neq 0$

أي الفرضية الصفرية (H_0) تنص على أن المعاملات $n = 1, 2, \dots, k$ (β_n) ليس لها تأثير على المتغير التابع Y.
أما الفرضية البديلة (H_1) تنص على أن هناك على الأقل معامل واحد (β_i) له تأثير على المتغير التابع Y.
و يتم حساب قيمة (F) المحسوبة من الجدول التالي :

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F المحسوبة
الانحدار Regression	$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	df=1	$MSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	$\frac{MSR}{MSE}$
الخطا Error	$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$	df=n-2	$MSE = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{n-2}$	
التباين الكلي Total	$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	df=n-1		

جدول 2.3: حساب قيمة F

2. تطبيق اختبار (T) : اختبار الفرضيات التالية :

- الفرضية الصفرية (H_0) : $H_0 : \beta_i = 0$.
- الفرضية البديلة (H_1) : $H_1 : \beta_i \neq 0$.

أي أن الفرضية الصفرية (H_0) تنص على أن معامل (β_i) ليس له تأثير على المتغير التابع (Y).
أما الفرضية البديلة (H_1) تنص على أن معامل (β_i) له تأثير على المتغير التابع (Y).
و يتم حساب قيمة (T) المحسوبة كمايلي :

$$T_{\beta_i} = \frac{\beta_i}{\sigma_{\beta_i}} .$$

أي

$$T_{\beta_i} = \frac{\text{قيمة معالم الانحدار}}{\text{الخطأ المعياري له}} .$$

فإذا كانت القيمة المطلقة ل(T) المحسوبة أكبر من أو تساوي قيمة (t) الجدولية عند درجة الحرية (n-2) عند مستوى الدلالة $\alpha = 0.05$ و باختبار بطرفين (لأن الفرضية البديلة تنص على $(H_1 : \beta \neq 0)$ ، أما إذا كانت الفرضية البديلة تنص على $(H_1 : b > 0)$ أو $(H_1 : b < 0)$ فيكون الاختبار بطرف واحد) فإننا نرفض الفرض الصفري (H_0) و نقبل بالفرض البديل (H_1) و نقول إن معامل الانحدار دال احصائيا أي أن له تأثير على المتغير التابع (Y) ، أما إذا كانت القيمة المطلقة ل(T) المحسوبة أقل تماما من قيمة (t) الجدولية فإننا نقبل الفرض الصفري (H_0) و نرفض الفرض البديل (H_1) و نقول إن معامل الانحدار غير دال احصائيا و بالتالي ليس له تأثير على المتغير التابع (Y) ، و لا يمكننا التنبؤ بدرجات المتغير (Y).

7.3 الدلالة العلمية للانحدار الخطي المتعدد

[5] حساب حجم التاثير للانحدار الخطي المتعدد يكون من خلال حساب معامل التحديد (R^2) و يتم حسابه كمايلي :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} .$$

و يتم تفسير قيمة حجم التأثير المتحصل عليها من قيمة معامل التحديد وفق الحالات التالية :

R^2 معامل التحديد	$0.09 > R^2 \geq 0.01$	$0.25 > R^2 \geq 0.09$	$0.25 \leq R^2$
(ES) حجم التاثير	صغير	متوسط	كبير

جدول 3.3: تفسير قيمة حجم التأثير المتحصل عليها من قيمة معامل التحديد

مثال

لدينا 10 مشاهدات لمتغير تابع واحد مع متغيرين مستقلين كالتالي :

Y	X_1	X_2
10	5	22
11	8	19
9	4	17
7	2	24
13	8	22
15	7	20
12	7	19
10	5	18
6	5	25
14	11	23
$\sum$ 107	62	209

جدول 4.3: المشاهدات

المطلوب :

1. حساب مقدرات معلمات بطريقة المربعات الصغرى مع التفسير .
2. حساب التباينات و الانحرافات المقدرة لمعلمات .
3. اختبر معنوية الميول الحدية (الدلالة) مع تحديد مجالات الثقة عند مستوى 95% .

الحل :

1. تقدير المعلمات بطريقة المربعات الصغرى : بداية علينا استخراج المصفوفتين $(X^TX)^{-1}$ و (X^TY) على النحو التالي :

$$(X^TX) = \begin{pmatrix} 10 & 62 & 209 \\ 62 & 442 & 1225 \\ 209 & 1225 & 4433 \end{pmatrix},$$

Y	X_1	X_2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
10	5	22	25	484	110	50	220
11	8	19	64	361	152	88	209
9	4	17	16	289	68	36	153
7	2	24	4	576	48	14	168
13	8	22	64	484	176	104	286
15	7	20	49	400	140	105	300
12	7	19	49	361	133	84	228
10	5	18	25	324	90	50	180
6	5	25	25	625	125	30	150
14	11	23	121	529	253	154	322
$\sum$ 107	62	209	441	4433	1295	717	2216

جدول 5.3: حساب X_1^2 و X_2^2 و X_1X_2 و X_1Y X_2Y

$$(X^TY) = \begin{pmatrix} 107 \\ 715 \\ 2216 \end{pmatrix},$$

$$(X^TX)^{-1} = \frac{1}{37376}\begin{pmatrix} 282361 & -4191 & -12088 \\ -4191 & 649 & 8 \\ -12088 & 8 & 576 \end{pmatrix},$$

$$\hat{\beta} = (X^TX)^{-1}(X^TY) = \frac{1}{37376}\begin{pmatrix} 282361 & -4191 & -12088 \\ -4191 & 649 & 8 \\ -12088 & 8 & 576 \end{pmatrix}\begin{pmatrix} 107 \\ 715 \\ 2216 \end{pmatrix} = \begin{pmatrix} 11.4794 \\ 0.8916 \\ -0.3018 \end{pmatrix}.$$

و بالتالي تكون معادلة الانحدار المقدرة كمايلي :

$$\hat{Y} = 11.4794 + 0.8916X_1 - 0.3018X_2 + e\ .$$

التفسير :

من خلال معادلة الانحدار المقدرة نلاحظ أن المتغير Y له علاقة طردية بالمتغير X_1 حيث كلما ارتفع ب 10% سيؤدي ذلك إلى ارتفاع Y ب 8.916% و نلاحظ وجود علاقة عكسية بين Y و X_2 حيث إذا ارتفع X_2 ب 10% أدى ذلك إلى انخفاض Y ب 3.018% .

2. حساب التباينات و الانحرافات

أولا حساب e^2 و e

$$e_i = \sum_{i=1}^{n}(Y_i - \hat{Y}_i) \ .$$

Y	$\hat{Y}$	e	e^2
10	9.2978	0.7022	0.49308484
11	12.878	-1.878	3.526884
9	9.9152	-0.9152	0.8375910
7	6.0194	0.9806	0.96157636
13	11.9726	1.0274	1.05555076
15	11.6846	3.3154	10.9918496
12	11.9864	0.0136	0.000187716
10	10.505	-0.505	0.255025
6	8.3924	-2.3924	5.72357776
14	14.3456	-0.3456	0.11943936
$\sum$			23.96479124

جدول 6.3: حساب $\hat{Y}$ و e و e^2

و منه يكون التباين المقدر للاخطاء :

$$\sigma_e^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(e_i^2) = \frac{23.96479124}{10-2-1} = 3.4235 \ .$$

التباينات و الانحرافات المقدرة للمعالم : حيث $i = 0, 1, 2$, $b_i = \hat{\beta}$.

$$\sigma_{\hat{b_0}}{}^2 = 3.4235\frac{23.96479124}{37376} = 25.8632 \ .$$

$$\sigma_{\hat{b_0}} = 5.0855 \ .$$

$$\sigma_{\hat{b_1}}{}^2 = 3.4235\frac{649}{37376} = 0.0594 \ .$$

$$\sigma_{\hat{b_1}} = 0.2437 \ .$$

$$\hat{\sigma_{b_2}}^2 = 3.4235\frac{576}{37376} = 0.0527 \ .$$

$$\hat{\sigma_{b_2}} = 0.2295 \ .$$

3. اختبار معنوية الميولة الحدية (الدلالة)

حساب قيم T المحسوبة لاختبار الفرضيات

$$T_{b_1} = \frac{|b_1|}{\hat{\sigma_{b_2}}} = \frac{0.8916}{0.2437} = 3.65 \ .$$

$$T_{b_2} = \frac{|b_2|}{\hat{\sigma_{b_1}}} = \frac{0.3018}{0.2295} = 1.31 \ .$$

لاختبار الفرضيات التالية :

$H_0 : b_1 = 0 \quad , \quad H_0 : b_2 = 0$.

$H_1 : b_1 \neq 0 \quad , \quad H_1 : b_2 \neq 0$. أما قيمة t فهي :

$$t^{0.05}_{n-k-1} = t^{0.05}_{7} = 2.356 \ .$$

فيما يخص قيمة b_1 فإن قيمة T المحسوبة أكبر من قيمة t الجدولية مما يعني قبول الفرضية البديلة و رفض الفرضية الصفرية و نقول أن b_1 معنوية عند مستوى الدلالة $\alpha = 0.05$ أي المتغير X_1 يؤثر معنويا على المتغير التابع Y ، أما المعلمة b_2 فإن قيمة T المحسوبة أصغر من قيمة t الجدولية مما يعني قبول الفرضية الصفرية و القول أن b_2 غير معنوية عند مستوى الدلالة أي المتغير X_2 لا يؤثر معنويا على المتغير التابع Y.

في حين يتم تحديد مجالات الثقة كالتالي :

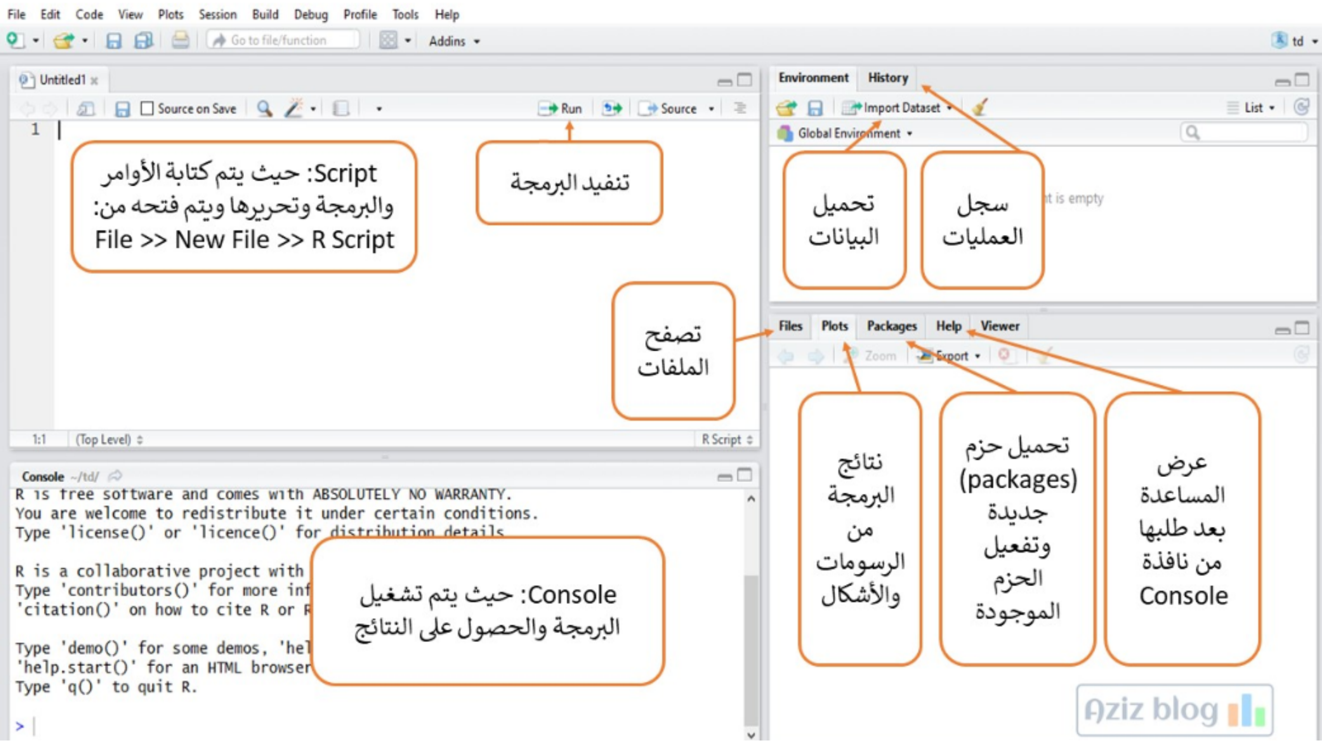
$$I_{b_1} = \hat{b_1} \pm t^{0.05}_{n-k-1} * \hat{\sigma_{b_1}} = 0.8916 \pm 2.365 * 0.2437 = [0.3152, 1.4679] \ ,$$

$$I_{b_2} = \hat{b_2} \pm t^{0.05}_{n-k-1} * \hat{\sigma_{b_2}} = -0.30108 \pm 2.365 * 0.2295 = [-0.8445, 0.2409] \ .$$

أهم ملاحظة من مجالي الثقة هو أن المجال الأول لا يتضمن القيمة صفر مما يؤكد أن b_1 قيمة المعامل لا يمكن أن تساوي صفر عند مستوى الثقة 95% و العكس في المجال الثاني .

باب 4

تطبيق



شكل 1.4: بيئة عمل لبرنامج R studio

1.4 برنامج R

البرنامج R هو واحد من أشهر البرامج الاحصائية ، يعتمد على لغة برمجة R التي تسمح ببناء البرامج و التطبيقات الاحصائية . هذا البرنامج يحوي على تشكيلة واسعة من التطبيقات الاحصائية مثل الانحدار الخطي البسيط و المتعدد و الرسومات الاحصائية ، و هناك بيئة احصائية لبرنامج R و التي تحتوي على نوافذ و أدوات تساعد على زيادة الانتاج و الابداع وهي R studio و التي تدعم اللغة الانجليزية فقط .

عند الدخول إلى شاشة العمل لا تحتاج لفتح ملف أو مشابه ذلك فقط إدخال الكود المراد العمل به . لنأخذ على سبيل المثال و نقوم بالعمل على برنامج المثال الذي أخذناه سابقا في الانحدار الخطي المتعدد : ندخل أولا البيانات

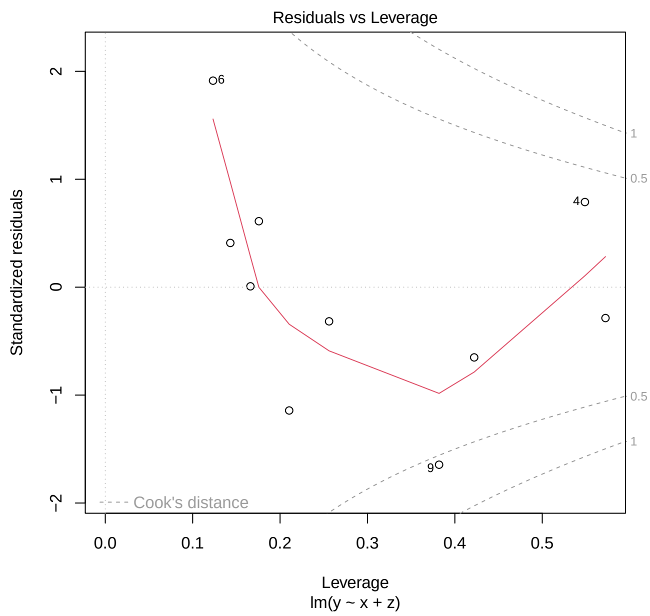
و لتقدير المعلمات و معادلة الانحدار الخطي المتعدد و الانحرافات و التباينات و جدول التحليل التبايني فما عليك سوى بإدخال بعض

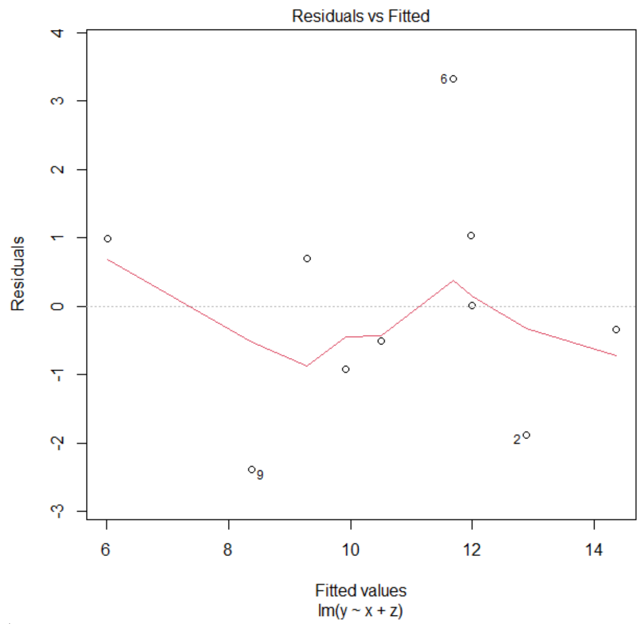
```
> x<-c(5,8,4,2,8,7,7,5,5,11)
> z<-c(22,19,17,24,22,20,19,18,25,23)
> y<-c(10,11,9,7,13,15,12,10,6,14)
```

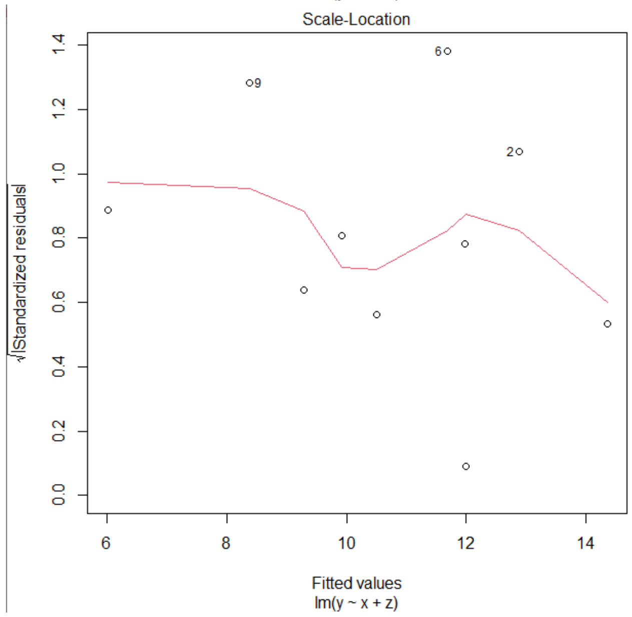
الاكواد أو الشيفرات التالية : ثم الرسومات .

```
> lm(y~x+z)

Call:
lm(formula = y ~ x + z)

Coefficients:
(Intercept)            x            z
    11.4794       0.8916      -0.3018

> summary(lm(y~x+z))

Call:
lm(formula = y ~ x + z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3927 -0.8129 -0.1664  0.9108  3.3151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.4794     5.0856   2.257   0.0586 .
x             0.8916     0.2438   3.657   0.0081 **
z            -0.3018     0.2297  -1.314   0.2303
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.85 on 7 degrees of freedom
Multiple R-squared:  0.6851,    Adjusted R-squared:  0.5951
F-statistic: 7.614 on 2 and 7 DF,  p-value: 0.01753
```

```
> coefficients(lm(y~x+z))
(Intercept)           x           z
 11.4793985   0.8916417  -0.3017979
> confint(lm(y~x+z))
                 2.5 %     97.5 %
(Intercept) -0.5461773 23.5049744
x            0.3151064  1.4681770
z           -0.8449417  0.2413458
> residuals(lm(y~x+z))
          1           2           3           4           5           6           7
 0.70194777 -1.87837115 -0.91540026  0.98046875  1.02702269  3.31506849  0.01327055
          8           9          10
-0.50524401 -2.39265839 -0.34610445
> anova(lm(y~x+z))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x          1 46.225  46.225 13.5021 0.007917 **
z          1  5.910   5.910  1.7263 0.230298
Residuals  7 23.965   3.424
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(y ~ x + z)


Residuals vs Fitted
Residuals
Fitted values
lm(y ~ x + z)


Scale-Location
√|Standardized residuals|
Fitted values
lm(y ~ x + z)

الخاتمة

بالنظر إلى ما قمنا به في حل المثالين للانحدار الخطي البسيط و المتعدد ، فقد تبين لنا مدى تأثير و قوة علاقة المتغيرات المستقلة بالمتغير التابع ، و كيف أن البرنامج studio R قام من الناحية التنظيمية و الزمنية (اختصار الوقت و الجهد في البيانات الكبيرة) التقليل على الباحث في بحوثه التطبيقية .

ملخص

الانحدار الخطي البسيط و المتعدد هو معادلة ذات متغيرات مستقلة و متغيرات تابعة ترتبط فيما بينها بعلاقة ، و هذه العلاقة تحدد عن طريق مدى تاثر المتغيرات ببعضها (المستقلة بالتابعة) ، لتفادي الوقوع في الاخطاء عند تقدير المعلمات بطريقة المصفوفات يمكننا استخدام عندئذ برنامج حاسوبي R لحساب المعلمات دون جهد و عناء و يمكن استخلاص جدول تحليل التباين منه ايضا لذا هو يسهل للباحث عمله في ابحاثه الاحصائية .

La régression linéaire simple et multiple est une équation avec des variables indépendantes et des variables dépendantes liées les unes aux autres par une relation. Cette relation est déterminée par la mesure dans laquelle les variables s'influencent mutuellement (indépendantes et dépendantes). paramètres en utilisant la méthode matricielle, nous pouvons utiliser un programme informatique R pour calculer les paramètres sans C'est un effort et un effort, et un tableau d'analyse de variance peut également en être extrait, cela facilite donc le travail du chercheur dans sa recherche statistique .

Simple and multiple linear regression is an equation with independent variables and dependent variables that are linked to each other by a relationship. This relationship is determined by the extent to which the variables affect each other (independent and dependent). To avoid making mistakes when estimating parameters using the matrix method, we can use an R computer program to calculate the parameters without It is an effort and effort, and an analysis of variance table can also be extracted from it, so it facilitates the work of the researcher in his statistical research.

الاختصارات و الرموز

الرمز أو الاختصار	الدلالة باللغة الفرنسية	الدلالة باللغة العربية
σ^2 أو var	Variance	التباين
σ	L'écart type	الانحراف المعياري
cov	Covariance	التباين المشترك أو التغاير
E	Le prédiction aléatoire	التوقع العشوائي
T	Test étudient	اختبار ستودنت
F	Test de Fisher	اختبار فيشر
OLS	Méthode des moindres carrés	طريقة المربعات الصغرى
$ANOVA$	Analyse de variance	تحليل التباين
MS	Carrés Moyen	متوسط مربعات
SS	Somme des carrés	مجموع مربعات
MSE	Carrés moyen des résidus	متوسط مربعات البواقي
MST	Carrés moyen de la variance totale	متوسط مربعات التباين الكلي
MSR	Régression carrée moyenne	متوسط مربعات الانحدار
SSE	Somme des carrés des restes	مجموع مربعات البواقي
SST	Somme des carrés de variance totle	مجموع مربعات التباين الكلي
SSR	Somme des carrés de Régression	مجموع مربعات الانحدار
df	Le degré de liberté	درجة الحرية

المصادر

[1] د.محمد بداوي ، الاحصاء الوصفي ، دار هومه . الجزائر 2017.

[2] د.شاكر مصلح محمدي ، أ.د.فاضل مصلح محمدي ، الاحصاء و تصميم التجارب ، دار أسامة للنشر و التوزيع ، الأردن - عمان الطبعة الأولى 2011.

[3] د.وليد إسماعيل السيفو ، د. عيد أحمد أبو بكر ، أ.د. غالب عوض الرفعي ، اساسيات الأساليب الاحصائية للاعمال و تطبيقاتها في العلوم المالية و الادارية و الاقتصادية ،ناشرون و موزعون ، الطبعة الأولى 2010.

[4] رينيه فيسيير، مذكرة مساعدة، الإحصاء والاحتمالات، للمهندسين، دونود، باريس، 2001، 2006.

[5] الانحدار الخطي التطبيقي ، جامعة مينيسوتا، كلية الإحصاء، مينيابوليس، مينيسوتا، 2005.

[6] نماذج الانحدار الخطي، ماجالي فرومونت رينوار، ماجستير في الإحصاء التطبيقي.