2−1: تمهيد:

فى هذا الفصل سوف يتم التطرق إلى الأساليب والنماذج الإحصائية التى ستستخدم فى تحليل بيانات البحث.

2−2: نماذج الإنحدار:

تعتمد العديد من الدراسات والبحوث على أساليب متطورة من أجل الحصول على نتائج تتصف بالفعالية والدقة العاليتين، وقد كان لعلم الإحصاء وفروعه المرتبطة به الأثر الكبير في بناء النماذج الرصينة وتحليل البيانات من خلالها وصولاً للقرارات السليمة يعد تحليل الانحدار أهم فروع علم الاحصاء، والذي يهتم ببناء العلاقة الرياضية بين متغير الاستجابة والمتغيرات التوضيحية وتتمثل هذه العلاقة كتركيبة خطية تدعى معادلة الانحدار إذ دقتها تعتمد على صحة تقدير معلماتها والتي تشترط توفر فروض التحليل، وتعتبر طريقة المربعات الصغرى الاعتيادية أحد أكثر طرق تقدير المعلمات استخدامًا وتتصف هذه الطريقة بكفاءتها العالية بتقدير معلمات نموذج الانحدار عند توفر فروض التحليل، أما عند عدم توفر فروض التحليل فتصبح مقدراتها غير متسقة ولا متكاملة الشروط، إن الفرض الأساس في طريقة المربعات الصغرى هو أن يكون متغير الاستجابة يتبع التوزيع الطبيعي، ولكن قد يندر هذا الفرض في كثير من المجالات التطبيقية وعليه فيجب البحث عن طرق وأساليب إحصائية أخرى لدراسة نماذج الانحدار.

2-2-1: الانحدار الخطي البسيط: Simple Linear Regression

ينقسم تحليل الانحدار إلى قسمين رئيسيين هما الانحدار **(Linear and non-linear regression)** الخطي والانحدار اللاخطي سوف نأخذ القسم الأول (الانحدار الخطي). يختص الانحدار الخطي بدراسة العلاقة بين المتغيرات على هيئة نموذج، فقد يحتوي النموذج على متغير توضيحي واحد فيسمى في هذه الحالة بنموذج الانحدار الخطي البسيط ويكتب بالصيغة الآتية:

$$Y_i = B_0 + B_1 X_1 + e_i \quad .... \quad (2-1)$$

حيث أن:

Y_i : *المتغير التابع.*

X_i : المتغير التوضيحى.

$B_0 B_1$: ثوابت وهي معلمات نموذج الانحدار

e_i : الخطأ العشوائي.

2–2–2: نموذج الانحدار الخطي المتعدد: (Multiple Liner Regression Model)

أما في حالة كون النموذج يحتوي على متغيرات توضيحية عدة فانه يسمى (**Multiple Liner Regression Model**) بنموذج الانحدار الخطي المتعدد، ويكتب بالصيغه الرياضية:

$$\underline{Y} = \underline{X}\ \underline{B} + \underline{e} \quad .... \quad (2\text{-}2)$$

حيث أن:

$\underline{Y}$: متجه متغير الاستجابة.

$\underline{X}$: مصفوفة المتغيرات التوضيحي.

$\underline{B}$: متجه ثوابت وهي معلمات الانحدار.

$\underline{e}$:متجه الخطأ العشوائي.

3–2: معامل التحديد (R^2):

مهم جداً عند اجراء تحليل الانحدار التعرف على نسبة مساهمة المتغير اوالمتغيرات المستقلة فى إحداث التغيرات علي المتغير المعتمد، معامل التحديد هو المعيار الذى يبين ذلك، ويعرف معامل التحديد بكونه نسبة اختلاف المفسر فى المتغير المعتمد من قبل المتغير المستقل وهو يمثل مربع معامل الارتباط ويرمز بالرمز R^2 اي ان:

$$r^2 = \frac{SSR}{SST} \quad .... \quad (2\text{-}3)$$

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$$

حيث ان:

$-1 \le r \le 1$ لذلك فإن $0 \le R^2 \le 1$

وفى حالة إستخدام المصفوفات نلاحظ أن هذه المجاميع تحسب كالآتى:

$$\mathrm{SST} = \sum\left(Y - \overline{Y}\right)^2 = Y'Y - \frac{(\sum Y)^2}{n}$$

$$SSR = \sum\left(\hat{Y} - \overline{Y}\right)^2 = \hat{B}'X'Y - \frac{(\sum Y)^2}{n}$$

$$SSE = = \sum\left(Y - \hat{Y}\right)^2 = SST - SSR = Y'Y - \hat{B}'X'Y$$

4–2: إختبار t:

لإختبار الفرضية:

فرضية العدم $H_0: = B_0 = 0$

إختبار من طرفين $H_1: = B_0 \neq 0$

إختبار من طرف واحد (الأيمن) $H_1: = B_0 > 0$

إختبار من طرف واحد (الأيمن) $H_1: = B_0 < 0$

نستخدم إختبار t بالصيغة الآتية:

$$t = \frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{v(\hat{\beta}_1)}} \qquad .... (2-4)$$

وهذه الصيغة لاتصح إلا إذا كان توزيع المتغير العشوائى فى البسط هو توزيع طبيعى حيث أن:

$$\hat{\beta}_1 \sim N\left(\hat{\beta}_1 \frac{\sigma_u^2}{S_{xx}}\right)$$

لذلك فإن الإحصائية أعلاه تتوزع بتوزيع t بدرجة حرية (n-2) كما سنرى لذلك:

$$\therefore t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma_u^2}{S_{xx}}}}$$

أن فرضية العدم:

$$H_0: \beta_1 = 0$$

$$\therefore t = \frac{\sqrt{S_{xx}\beta_1}}{\sigma_u}$$

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$$

ولإتخاذ القرار فى شأن قبول أو رفض H_0 تقارن القيمة المطلقة ل t المحسوبة مع قيمة t الجدولية فإن كانت المحسوبة أكبر من الجدولية فيتم رفض فرضية العدم وقبول البديلة أى أن المتغير المستقل يؤثر معنوياً على المعتمد أما إذا كانت t المحسوبة أصغر أو تساوى t الجدولية فيتم قبول فرضية العدم اى أن المتغير المستقل لايؤثر معنوياً على المتغير المعتمد.

ويستخدم إختبار t فى حالة :

1- إذا كان تباين الخطأ للمجتمع مجهول ولذلك نستخدم التقدير له σ_u^2.

2- إذا كان حجم العينة أقل من 30.

2-5: إختبار F:

إختبار F المقصود هنا هو إختبار F الناتج عن تكوين جدول تحليل التباين ومن خلال المعادلة:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

نلاحظ أن قيمة Y_i يتم الحصول عليها من خلال جزئين الجزء الأول الأنحدار والجزء الثانى هو تأثير المتغير ات المستقلة التى تعرف بالخطأ نستنتج من ذلك أن جدول تحليل التباين يتضمن جزئين هما الإنحدار Regression والخطأ ومجموع هذين يعطينا الكلى Total أما درجة الحرية فهى كالآتى:

درجة حرية الإنحدار هى عبارة عن عدد المتغيرات المستقلة فى النموذج كذلك فإن درجة حرية الإنحدار تمثل عدد المعلمات فى النموذج بإستثناء معلمة المقطع والسبب فى هذا الإستثناء هو أن الإختبار خاص بتأثير المتغيرات المستقلة على المتغير المعتمد.

ولما كانت درجة الحرية للكلى هى (n-1) وبإستخدام خاصية الجمع لدرجات الحرية نستنتج أن درجة حرية الخطأ هى (n-2) ويكون شكل جدول تحليل التباين كالآتى:

مصدر التباين	درجة الحرية	مجموع المربعات	متوسط المربعات	قيمة f
الإنحدار	1	SSR	MSR	$\frac{MSR}{MSE}$
الخطأ	n-2	SSE	MSE	
الكلى	n-1	SST		

المصدر: إبراهيم ، بسام يونس ، حاجي ، انمار امين ، يونس , عادل موسي (2001) ، "الاقتصاد القياسي " ، دار عزه للنشر ، الخرطوم ، السودان

6-2: نموذج توبت (Tobit Model):

نحن نعلم ان تحليل الإنحدار (Regression analysis) هو أحد الوسائل الإحصائية التى تصف شكل العلاقة مابين المتغيرات التفسيرية والمتغير المعتمد، فاذا كانت قيم المتغيرات التفسيرية معلومة فان تحليل الانحدار يساعدنا فى التوقع (التنبؤ) بقيمة المتغير التابع لذلك فان اختيار النموذج الملائم للبيانات المتوفرة هى من ضروريات هذا التحليل، وبهدف الوصول إلى نتائج دقيقة عن الظاهرة المدروسة فيجب ان ينسجم النموذج المختار مع البيانات المتوفرة بأفضل صورة ممكنة، وكذلك فان إقتراح نموذج الانحدار الامثل للبيانات قيد البحث سيؤدى إلى الحصول على نتائج تكون قريبه من الواقع الحقيقى إذاً لكل نوع من البيانات يوجد نموذج أمثل يتناسب معها فمثلاً إذا توفر لدينا بيانات كمية للمتغير المعتمد وبتوفر الافتراضات الخاصة بالنموذج يمكن التعامل معها بإستخدام نموذج الإنحدار التقليدى (Conventional regression Model) وكذلك إذا توفرت بيانات ثنائية للمتغير المعتمد يمكن التعامل معها بإستخدام نموذج الإنحدار اللوجستى (Logistic Regression Model)، لكن فى حالة توفر مشاهدات تكون مقيدة فى جزء (محددة) وحره فى الجزء الآخر (غير محددة) حيث تسمى هذه البيانات بالبيانات المراقبة (censored data)، فإن إستخدام نموذج الإنحدار التقليدى مع هكذا نوع من البيانات سيؤدى إلى معالم مقدرة متحيزة (biased) من جهة ومن جهة أخرى غير متسقة (Inconsistent) وكذلك قابلية هذه المقدرات على التعميم ضعيفة جداً، اذا لابد من تحديد نموذج يكون متناسب مع هذه البيانات وهذا النموذج هو نموذج الإنحدار المراقب (censored regression model) (نموذج توبت) والذى يعتبر أفضل نموذج، إن نموذج الإنحدار المراقب يشبه إلى حد ما نموذج الإنحار المبتور (truncated regression model) وذلك عند وجود نسبة معينة من البيانات المتطرفة (extreme data) فى المتغير المعتمد إن إستبعاد هذه القيم المتطرفة من النموذج أفضل من بقائها فى حالة تقدير المعالم الخاصة بالنموذج.

أى بتعبير آخر إذا كان هنالك نسبة معينة من بيانات المتغير المعتمد هى بيانات مفقودة او متطرفة فإن النموذج الملائم لهذه البيانات هو نموذج الإنحدار المبتور، لكن عملية البتر هذه لاتكون صحيحة دائماً وذلك لأن هذه البيانات قد تمتلك معلومات ذات أهمية، فإهمالها ربما يؤثر على مقدرات المعالم وبالتالى فإن نموذج الإنحدار المبتور لايعتبر النموذج الأفضل (لان بتر هذه النسبة من البيانات يتسبب بخسارة كمية لا يستهان

بها من المعلومات)، لذلك فإن النموذج الملائم لمثل هذه البيانات هو نموذج انحدار توبت، الذى يتعامل مع بيانات مكونة من جزأين وتكون دالة هذا النموذج دالة مختلطة (Mixed function) إذ إن كل جزء من بيانات المتغير المعتمد فى هذه الحالة ستأخذ توزيع معين فالمشاهدات ذات القيم المساوية إلى الصفر ستأخذ الدالة التجميعية (cdf) للتوزيع الطبيعى والمشاهدات التى تأخذ كميات موجبة ستأخذ داله الكتلة الإحتمالية (pdf) للتوزيع الطبيعى.

أن الصيغة العامة لنموذج إنحدار توبت هى:

$Y=a \quad \text{if } y^* \leq a$

$\text{If } Y^* > a \quad Y= y^*$

حيث أن:

$$y^* = b_0 + b_1x_1 + \ldots\ldots + b_kx_k + e_i \quad \ldots\ldots\ldots\ldots (2\text{-}5)$$

$e_i \sim N(0,\sigma^2)$

$y^* \sim N(XB,\sigma^2)$

حيث أن:

a: هى نقطة التقيد.

Y: هى المتغير المعتمد.

y^*: هى المتغيرات الكامنة.

b_j: هى معالم النموذج (j=0,1,2,....k)

x_j: هى المتغيرات التفسيرية (j=0,1,2,....k)

e_i: حد الخطأ العشوائى.

تؤخذ قيمة a (نقطة التقيد) مساوية الى أى قيمة محددة حسب بيانات الدراسة، وفى بيانات الظاهرة قيد الدراسة كانت نقطة التقيد مساوية الى الصفر لذا فأن شكل الدالة سيأخذ الشكل التالى:

$Y=0$ if $y^* \leq 0$

$Y=y^*$ if $y^* > 0$

$a=0$

من خلال صيغة دالة نموذج توبت المبينة أعلاه سوف يتم التعامل في هذا البحث مع جزأين من البيانات الجزء الأول من المشاهدات عندما تكون قيم المشاهدات مساوية إلى الصفر والجزء الثاني من تلك المشاهدات عندما تكون مساوية إلى كميات موجبة، لذا فلكل جزء من هذه المشاهدات دالة رياضية خاصه بها ومن ضرب هذه الدوال مع بعضها سنحصل على دالة مختلطة لنموذج توبت وكما ذكر سابقًا فأن هذه الدالة المختلطة للنموذج تأخذ الصيغة الآتية:

$$P(Y) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i - XB)^2}{2\sigma^2}}\right]\left[1 - \Phi\left(\frac{a - XB}{\sigma}\right)\right] \quad .... (2\text{-}6)$$

وعندما تكون قيمة a=0 تكون صيغة الدالة كما يأتى:

$$P(Y) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i - XB)^2}{2\sigma^2}}\right]\left[1 - \Phi\left(\frac{-XB}{\sigma}\right)\right] \quad .... (2\text{-}7)$$

كذلك يمكن التعبير عن الدالة أعلاه بشكل آخر:

$$P(Y) = \left[\frac{1}{\sigma} \emptyset\left(\frac{y_i - XB}{\sigma}\right)\right]\left[1 - \Phi\left(\frac{-XB}{\sigma}\right)\right] \quad .... (2\text{-}8)$$

حيث أن $\emptyset$ هى دالة الكثافة الإحتمالية (p.d.f)، و Φ هى الدالة التجميعية (c.d.f) ومن خلال ماذكر أعلاه يمكن تقدير معالم نموذج توبت بالإعتماد على طريقة الإمكان الأعظم سنحصل على:

$$\text{Ln } P(Y) = \sum_{i=1}^{n}\left[\left[-ln\sigma + ln\emptyset\left(\frac{y_i - XB}{\sigma}\right)\right] + ln\left[1 - \Phi\left(\frac{-XB}{\sigma}\right)\right]\right] \quad .... (2\text{-}9)$$

ويتم إستخدام الطرق العددية فى الحصول على مقدرات معلمات نموذج توبت فى المعادلة (3-10).