تحليل الانحدار الخطي البسيط

Simple Linear Regression Analysis

٥-١ مقدمة:

عندما يهتم الباحث بدراسة وتحليل أثر أحد المتغيرين على آخر، كدراسة وتحليل أثر سلوك المتغير X على سلوك المتغير Y ، فإنه الأسلوب المستخدم في التحليل هو أسلوب تحليل الانحدار الخطي البسيط.

فإذا كان $f(\mathrm{Y},\mathrm{X})$ هي دالة كثافة الاحتمال المشترك، ولها توزيع طبيعي ثنائي متوسطه هو $\mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}$ ، وله مصفوفة تباين هي: $\Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}$ ، أي أن:

$$\begin{pmatrix} \mathrm{Y} \\ \mathrm{X} \end{pmatrix} \sim N(\mu, \Sigma) \qquad \textbf{(5-1)}$$

وإذا كنا نهتم بدراسة سلوكيات المتغير Y بمعلومية المتغير X ، يجب علينا في هذه الحالة استخدام دالة كثافة الاحتمال الشرطي $f(\mathrm{Y} \mid \mathrm{X})$ في دراسة هذا السلوك، وهذا التوزيع توزيع طبيعي أيضا، له متوسط وتباين على الشكل التالي:

$$\begin{aligned} \mu_{y|x} &= \left(\mu_y - \left(\rho_{xy} \frac{\sigma_y}{\sigma_x} \right) \mu_x \right) + \left(\rho_{xy} \frac{\sigma_y}{\sigma_x} \right) \mathrm{X} \\ &= \beta_0 + \beta_1 \mathrm{X} \end{aligned} \qquad \textbf{(5-2)}$$

حيث أن $\mu_{y|x}$ يعبر عن قيمة متوسط المتغير العشوائي Y عندما تكون قيمة $(\mathrm{X} = x)$ محددة ومعلومة.

والمعادلة أعلاه تأخذ شكل خط مستقيم، وعند كل قيمة محددة للمتغير X توجد قيمة لمتوسط المتغير Y يقع على هذا الخط، ومن ثم يمكن التعبير عن المتغير العشوائي Y بمعلومية المتغير X

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (5\text{-}3)$$

حيث أن:

$\beta_1 = \left(\rho_{xy} \dfrac{\sigma_y}{\sigma_x} \right)$ **: هو ميل الخط المستقيم، ويسمى بمعامل الانحدار في المجتمع، ويدل على مقدار التغير في المتغير التابع** Y **إذا تغيرت** X **بوحدة واحدة.**

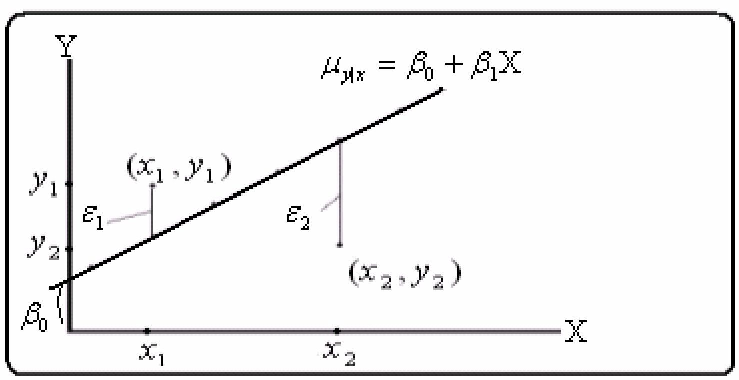
β_0 **: يعبر عن بالقاطع intercept وهو الجزء المقطوع من المتغير** Y **، ويعبر عنه من المعادلة (5-2) كما يلي:**

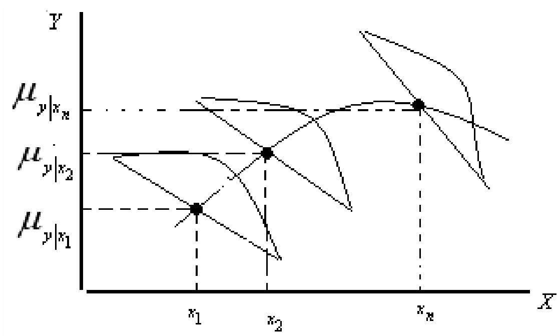
$$\beta_0 = \left(\mu_y - \left(\rho_{xy} \frac{\sigma_y}{\sigma_x} \right) \mu_x \right) = \mu_y - \beta_1 \mu_x \qquad (5\text{-}4)$$

ε **: يعبر عن الخطا العشوائي** Random Error**، وهو مقدار انحراف المتغير العشوائي** Y **عن متوسطة الشرطي** $\mu_{y|x}$ **، أي أن :**

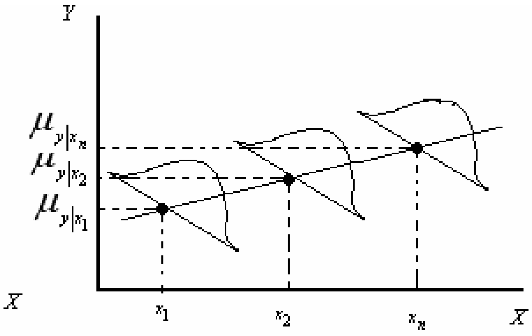
$$\varepsilon = Y - \mu_{y|x} = Y - \left(\beta_0 + \beta_1 X \right) \qquad (5\text{-}5)$$

والشكل التالي يبين قيم المتغير والمتوسطات الشرطية.





العلاقة غير خطية



العلاقة خطية

٥-٢ التقدير بنقطة لمعاملات الانحدار في المجتمع (β_0, β_1).

من المعادلة (5-2) يمكن التعويض عن تقدير كل معلمة من المعلمات $\mu_x, \mu_y, \rho_{xy}, \sigma_y, \sigma_x$ ، **ومن ثم نحصل على التقدير بنقطة للمعاملين** (β_0, β_1) **كما يلي:**

$$\hat{\beta}_1 = r_{xy} \frac{S_y}{S_x} \qquad (5\text{-}6)$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \qquad (5\text{-}7)$$

ويلاحظ أن إشارة المعامل $\hat{\beta}_1$ **تأخذ نفس إشارة معامل الارتباط** r_{xy} ، **وبإيجاد التقدير بنقطة للمعاملي** (β_0, β_1) **نجد أن تقدير معادلة الانحدار في المجتمع هي:**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \qquad (5\text{-}8)$$

وأما تقدير الخطأ العشوائي هو:

$$\hat{\varepsilon} = Y - \left(\hat{\beta}_0 + \hat{\beta}_1 X\right) \qquad (5\text{-}9)$$

تطبيق:

في المثال السابق، إذا رغب الباحث في دراسة أثر المساحة المنزرعة على كمية الإنتاج من

(4-7)، (4-6) كما يلي.

- $$S_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n-1}} = \sqrt{\frac{23850}{8-1}} = 58.3707 \qquad r_{xy} = -0.798$$
- $$S_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{12040}{8-1}} = 41.4729$$
- $$\hat{\beta}_1 = r_{xy}\frac{S_y}{S_x} = (-0.798)\frac{58.3707}{41.4729} = -1.1231$$
- $$\hat{\beta}_0 = \overline{\mathrm{Y}} - \hat{\beta}_1\overline{\mathrm{X}} = \frac{5264}{8} - (-1.1231)\frac{2108}{8} = 953.94$$

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

- **إذا تقدير معادلة الانحدار هي:**

$$\hat{Y} = 953.94 - 1.1231X$$

- **من المعادلة أعلاه يتضح الآتي:**

يدل المعامل $\hat{\beta}_1 = -1.1231$ **على أن أثر المساحة المنزرعة على كمية اللحوم أثر عكسي، وأن زيادة المساحة المنزرعة بمقدار ألف هكتار يؤدي إلى حدوث انخفاض في كمية اللحوم المنتجة بمقدار 1.1231 ألف طن.**

يدل المعامل $\hat{\beta}_0 = 953.94$ **على أن هناك إنتاج ثابت قدرة 953.94 ألف طن بصرف النظر عن التغير الذي يمكن أن يحدث في المساحة المنزرعة.**

٥-٢ منهجية التحليل

عند البدء في تحليل الانحدار ، يتم إتباع التالي:

١- **افتراض أن نموذج الانحدار يأخذ الشكل الخطي ، بمعنى أن** $\mu_{y|x_i} = \beta_0 + \beta_1 x_i$

٢- **يتم استخدام بعض طرق التقدير الإحصائي مثل طريقة المربعات الصغرى Ordinary Least Squares (OLS) أو طريقة الإمكان الأعظم Maximum Likelihood (ML) في تقدير معالم النموذج** (β_0, β_1) **، ومن ثم يتم تقدير النموذج وهو :**

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \textbf{(5-10)}$$

٣- **استخدام اطرق الإحصاء الاستدلالي في اختبار مدى صلاحية النموذج (6-3) في تمثيل العلاقة بين المتغير** y_i **كمتغير تابع ، والمتغير** x_i **كمتغير مستقل ، أو متغير مفسر للمتغير.**

٤- **إذا ثبت أن النموذج (6-4) جيد ومناسب ، يمكن الاعتماد عليه في التنبؤ ، وإذ أثبتت طريقة الاستدلال الإحصائي عدم صلاحية النموذج ، يتم اقتراح شكل آخر للنموذج غير شكل النموذج الخطي ، وتكرر نفس الخطوات من ٢ إلى ٤ وهكذا حتى يمكن التوصل إلى أفضل النماذج .**

٥-٣ شكل نموذج الانحدار الخطي البسيط

الانحدار الخطي البسيط هو الذي يهتم بدراسة وتحليل أثر متغير مستقل واحد على متغير تابع كمي ، وسمي انحدار خطي لأن المعادلة تأخذ الصورة الخطية في المعاملات (β_0, β_1) ، ووصف بأنه بسيط لأن عدد المتغيرات المستقلة محل الدراسة متغير واحد فقط ، وفيما يلي شكل نموذج الانحدار الخطي البسيط .

$$\begin{aligned} y_i &= \mu_{y|x_i} + \varepsilon_i \\ &= \beta_0 + \beta_1 x_i + \varepsilon_i \end{aligned} \qquad \textbf{(5-11)}$$

حيث أن:

y_i **: يعبر عن قيمة المشاهدة رقم** i **،** $i = 1,2,...,n$ **للمتغير التابع ، ويطلق عليه أحيانا بالمتغير المتنبأ به .**

x_i **: هي قيمة المشاهدة رقم** i **،** $i = 1,2,...,n$ **للمتغير المستقل، ويطلق عليه أحيانا ، المتغير المفسر ، أو المتنبأ منه .**

β_0 **: هو ثابت يعبر عن الجزء المقطوع (intercept) من المحور الرأسي، وهو هو عبارة عن قيمة متوسط المتغير التابع عند** $\mu_{y|x_i=0}$ **عندما** $x_i = 0$ **(انعدام قيمة المتغير المفسر).**

β_1 **: هو ميل الخط المستقيم** $\mu_{y|x_i} = \beta_0 + \beta_1 x_i$ **، ويعبر عن مقدار التغير في المتغير التابع ، إذا حدث تغير في المتغير المستقل بوحدة واحدة، ويطلق عليه أيضا معامل الانحدار ، ونوع إشارته تدل على ما إذا كان هناك تأثير طردي أو عكسي للمتغير المستقل على المتغير التابع.**

ε_i **: يعبر عن الخطأ العشوائي للمشاهدة التابعة رقم** i **،** $i = 1,2,...,n$

n **: هي عدد المشاهدات المتاحة من قيم المتغيرين** (y_i, x_i)

الافتراضات التي يستند عليها نموذج الانحدار الخطي البسيط

يستند نموذج الانحدار الخطي البسيط على عدد من الافتراضات هي:

١- أن المتغير المستقل x محدد Given ، وثابت Fixed.

٢- أن الأخطاء العشوائية ε_i ، $i = 1,2,...,n$ مستقلة ، وكل منها لها توزيع طبيعي متوسطه صفرا ، وتباينه σ^2 ، أي أن: $\varepsilon_i \sim N(0, \sigma^2)$

٣- أن كل مشاهدة من المشاهدات التابعة y_i عند القيمة المحددة x_i ، $i = 1,2,...,n$ لها توزيع طبيعي متوسطه هو الخط المستقيم $\mu_{y|x_i} = \beta_0 + \beta_1 x_i$ ، وتباينه هو تباين الخطأ العشوائي $\sigma^2_{y_i|x_i} = \sigma^2$ ، أي أن : $y_i \mid x_i \sim N(\beta_0 + \beta_1 x_i \; , \; \sigma^2)$.

٤- أن المتغير المستقل x_i ، مستقل عن الخطأ العشوائي ε_i ، أي أن : $Cov(x, \varepsilon) = E(x\varepsilon) = 0$.

٥-٤ الاستدلال الإحصائي حول نموذج الانحدار الخطي البسيط

لإجراء التحليل الإحصائي لنموذج الانحدار الخطي البسيط يمكن إتباع خطوات الاستدلال الإحصائي، وتلخص في الخطوات التالية:

١- تقدير معالم النموذج باستخدام بعض طرق التقدير الإحصائي مثل طريقة (OLS) ، وحساب مؤشرات جودة النموذج .

٢- اختبار جودة النموذج ، وكذلك اختبارات الفروض المتعلقة بمعالم النموذج ، وتقدير فترات ثقة لها.

٣- تقدير فترة تنبؤ لمتوسط المتغير التابع، وكذلك قيمة المتغير التابع.

أولا: تقدير معالم النموذج بطريقة (OLS)

من الملاحظ أن معالم نموذج الانحدار الخطي البسيط $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ تشمل

أ- معاملات الانحدار (β_0, β_1)

وتقدير المربعات الصغرى (OLS) لـ (β_0, β_1) هي القيم التي تجعل مجموع مربعات الأخطاء العشوائية أقل ما يمكن ، أي أن:

$$\text{(OLS) Estimates minimize :} \quad \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \qquad \text{(5-12)}$$

وهذا التقدير هو:

$$\hat{\beta}_1 = \frac{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)}}{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}} = \frac{\sum_{i=1}^{n} x_i y_i - \dfrac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad \text{(5-13)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \ \bar{x} \qquad \text{(5-14)}$$

وهذا التقدير هو نفس التقدير المبين بالمعادلتين (5-6) ، (5-7) ، ومن ثم يكون تقدير معادلة الانحدار هي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \text{(5-15)}$$

كما يمكن تعريف الفرق بين القيمة الفعلية للمشاهدة التابعة y_i ، والقيمة المقدرة $\hat{y}_i$ بالبواقيresiduals ، وهو في الوقت نفسه تقدير للأخطاء العشوائية ، أي أن البواقي $\hat{\varepsilon}_i$ هي :

$$\hat{\varepsilon}_i = (y_i - \hat{y}_i) = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \qquad \text{(5-16)}$$

ب- تباين الخطأ العشوائي σ^2

تقدير المربعات الصغرى لتباين الخطأ العشوائي هو $\hat{\sigma}^2$ ، ويحسب بالمعادلة التالية:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{(n-2)} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-2)}$$

$$= \frac{\sum_{i=1}^{n} y_i^2 - (\hat{\beta}_0 \sum_{i=1}^{n} y_i + \hat{\beta}_1 \sum_{i=1}^{n} y_i x_i)}{n-2} \qquad \textbf{(5-17)}$$

وتقيدرات المربعات الصغرى (OLS) Estimators $(\hat{\beta}_0، \hat{\beta}_1)$ تتصف بالآتي:

أ- أنها غير منحازة Unbiased ب- أنها متسقة Consistency .

جـ- لها أقل تباين . د- تحت تحقق افتراضات النموذج ، يكون لها توزيع طبيعي.

ومن ثم يتصف مقدر (OLS) بالمقدر الجيد غير المتحيز Best Linear Unbiased Estimator (BLUE) .

تطبيق (٢)

لدراسة أثر الوزن Weight على ضغط الدم B.P بالنسبة للمصابين بمرض ارتفاع الضغط ، قام باحث بسحب عينة من المرضي حجمها 10 وسجل الوزن والضغط لكل مريض ، وحصل على النتائج التالية:

No	1	2	3	4	5	6	7	8	9	10
B.P (y)	105	93	78	85	90	110	115	100	79	88
Weight (x)	90	85	60	63	75	85	90	85	58	65

بافتراض أن نموذج الانحدار الذي يعبر عن الضغط كدالة في الوزن يأخذ الصورة الخطية ، أوجد الآتي:

١- تقدير المربعات الصغرى (OLS) لمعاملات انحدار ضغط الدم علي الوزن، اكتب شكل النموذج المقدر، وفسره

٢- ما هو تقدير الضغط لشخص وزنه 63 كيلوجرام؟ وما هو تقدير الخطأ العشوائي عنده؟

٣- قدر تباين الخطأ العشوائي.

الحل :

١- حساب تقديرات (OLS) لمعاملات الانحدار.

يتم تكوين جدول المجاميع:

no	y	x	xy	x^2	y^2
1	105	90	9450	8100	11025
2	93	85	7905	7225	8649
3	78	60	4680	3600	6084
4	85	63	5355	3969	7225
5	90	75	6750	5625	8100
6	110	85	9350	7225	12100
7	115	90	10350	8100	13225
8	100	85	8500	7225	10000
9	79	58	4582	3364	6241
10	88	65	5720	4225	7744
Sum(Σ)	943	756	72642	58658	90393

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{72642 - \frac{(756)(943)}{10}}{58658 - \frac{(756)^2}{10}}$$

$$= \frac{1351.2}{1504.4} = 0.8982$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \ \bar{x} = \left(\frac{943}{10}\right) - 0.8982\left(\frac{756}{10}\right) = 26.396$$

إذا نموذج الانحدار تقديره هو:

$$\hat{y}_i = 26.396 + 0.8982 x_i$$

ويستدل من التقدير أعلاه ما يلي:

- $\hat{\beta}_1 = 0.8982$ **يدل على أن الوزن يؤثر طرديا على ضغط الدم، وأن زيادة الوزن بمقدار واحد كيلوجرام، يترتب عليه حدوث زيادة في الضغط بمقدار 0.9 درجة تقريبا.**
- $\hat{\beta}_0 = 26.396$ **يدل على أنه في حالة ثبات الوزن يكون متوسط الضغط هو 26.4 تقريبا** (هذا غير منطقي)

٢- **الضغط المقدر لشخص وزنه 63 كيلوجرام هو:**

$$\hat{y} = 26.396 + 0.8982(63) = 82.98$$

وأما الخطأ العشوائي عند هذا الوزن فتقديره هو:

$$\hat{\varepsilon} = (y - \hat{y}) = 85 - 82.98 = 2.02$$

٣- **تقدير تباين الخطأ العشوائي:**

أولا:- باستخدام الصيغة التعريفية

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{(n-2)} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{(n-2)}$$

no	y	x	$\hat{y} = 26.396 + 0.8982x$	$\hat{\varepsilon} = (y - \hat{y})$	$\hat{\varepsilon}^2$
1	105	90	107.2340	-2.2340	4.990756
2	93	85	102.7430	-9.7430	94.926050
3	78	60	80.2880	-2.2880	5.234944
4	85	63	82.9826	2.0174	4.069903
5	90	75	93.7610	-3.7610	14.145120
6	110	85	102.7430	7.2570	52.664050
7	115	90	107.2340	7.7660	60.310760
8	100	85	102.7430	-2.7430	7.524049
9	79	58	78.4916	0.5084	0.258471
10	88	65	84.7790	3.2210	10.374840
Sum(Σ)	943	756		0.000	254.4989

إذا تقدير تباين الخطأ العشوائي هو:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-2)} = \frac{254.4989}{(10-2)} = 31.812$$

ثانيا : باستخدام الصيغة الحسابية:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} y_i^2 - (\hat{\beta}_0 \sum_{i=1}^{n} y_i + \hat{\beta}_1 \sum_{i=1}^{n} y_i x_i)}{n-2}$$

$$= \frac{90393 - \left(26.396(943) + 0.8982(72642)\right)}{8} = \frac{254.5276}{8} = 31.816$$

ثانيا:- مؤشرات جودة النموذج:

من المعلوم أن القيمة الفعلية للمتغير التابع y تتكون من جزأين هما:

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

أي أن الاختلاف الكلي في المتغير التابع $(y_i - \overline{y})$ يمكن إرجاعه إلى جزأين هما:

$$(y_i - \overline{y}) \quad = \quad (\hat{y}_i - \overline{y}) \quad + \quad \hat{\varepsilon}_i$$

الاختلاف بسبب الخطأ + الاختلاف بسبب الانحدار = الاختلاف الكلي

ومن ثم يعبر عن مكونات مجموع المربعات الكلي بجزأين هما:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 \quad = \quad \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \quad + \quad \sum_{i=1}^{n}\hat{\varepsilon}_i$$

مجموع المربعات الكلي Total sum of squares	=	مجموع مربعات الانحدار Regression sum of squares	+	مجموع مربعات الأخطاء Error sum of squares
SST	=	SSR	+	SSE
$(n-1)$	=	1	+	$(n-2)$

ومن ثم يمكن استخلاص مؤشرين لجودة النموذج هما:

أ- مؤشر يعبر عن نسبة مجموع مربعات الانحدار (SSR) إلى مجموع المربعات الكلي SST ، ويطلق عليه معامل التحديد determinant coefficient ، أو (مربع معامل الارتباط الخطي البسيط R-Square) ويرمز له بالرمز (R^2) ، أي أن :

$$R^2 = \frac{SSR}{SST} \tag{5-18}$$

وهذا المعامل يدل على مدى قوة العلاقة بين القيم المقدرة اولقيم الفعلية ، فإذا كان هذا المعامل كبير ، دل ذلك على جودة النموذج ، لأن معنى ذلك أن المتغير المفسر x يشرح أو يفسر نسبة كبيرة من التغيرات الكلية التي تحدث في المتغير التابع.

ب-المؤشر الثاني وهو تباين الخطأ العشوائي وتقديره هو $\hat{\sigma}^2 = SSE/(n-2)$ ، ويقيس مدي انحرافات القيم الفعلية عن القيم المقدرة $(\hat{y}_i - \bar{y})$ ، فإذا كان هذا التقدير كبير دل ذلك إلى أن انحرافات القيم الفعلية للمتغير التابع عن القيم المقدرة لها كبير، ومن ثم يكون النموذج غير كفؤ.

تطبيق (٣)

في تطبيق (٢) احسب معامل التحديد ، وما هو مدلوله؟

الحل

لحساب معامل التحديد تطبق المعادلة **(5-18)**

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$= \frac{\hat{\beta}^2 \sum(x_i - \bar{x})^2}{\sum(y_i - \bar{y})^2} = \frac{\hat{\beta}_1^2\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)}{\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)} \qquad \textbf{(5-19)}$$

$$R^2 = \frac{(0.8982)^2\left(58658 - \frac{(756)^2}{10}\right)}{\left(90393 - \frac{(943)^2}{10}\right)} = \frac{1213.6946}{1468.1} 0.8267$$

أي أن الوزن يفسر 82.67% من الاختلافات التي تحدث في ضغط الدم ، وأن النسبة الباقية

(17.33%) ترجع لأخطاء عشوائية.

ثالثا :- الاختبارات المتعلقة بالنموذج ومعاملات الانحدار

قد يرغب الباحث في اختبار صلاحية النموذج المقترح من ناحية ، واختبار معنوية المعاملات من ناحية أخرى، وفيما يلي بيان ذلك.

أ- اختبار جودة (صلاحية) النموذج

الغرض من هذا الاختبار هو التوصل إلى قرار حول صلاحية النموذج في تمثيل العلاقة بين المتغير التابع والمتغير المستقل تمثيلا جيدا ، وفيما يلي خطوات إجراء الاختبار.

- **صياغة الفروض:**

يأخذ الفرض العدم والفرض البديل الصورة التالية.

الفرض العدم	H_o :**model is not suitable** **النموذج غير مناسب**
الفرض البديل	H_a :**model is suitable** **النموذج مناسب**

- **إحصائية الاختبار هي:**

$$F^* = \frac{\left(SSR/1\right)}{\left(SSE/(n-2)\right)} = \frac{MSR}{MSE} \qquad \textbf{(5-20)}$$

حيث أن :

MSR **: متوسط مربعات الانحدار ،** MSE **: متوسط مربعات الأخطاء ، وهو في نفس الوقت تقدير تباين الخطأ العشوائي**

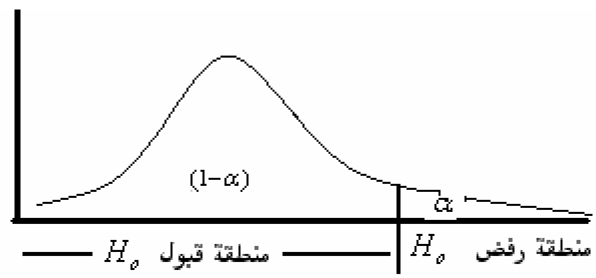
وإحصائية الاختبار F^* **تحت صحة الفرض العدم** H_o **تتبع توزيع** F **بدرجات حرية بسط ومقام هي:**

درجات حرية البسط = عدد المتغيرات المستقلة $(df_1 = 1)$

(1- عدد المتغيرات المستقلة - عدد المشاهدات) $= df_2 = (n-1-1 = n-2)$

- **مستوى المعنوية ، ومناطق الرفض والقبول:**

بفرض أن مستوى المعنوية الذي حدده الباحث هو α ، وأكثر مستوى المعنوية استخداما هو $(\alpha = 0.05 \;\; or \;\; \alpha = 0.01)$ ، فإنه عند درجات الحرية بسط $(df_1 = 1)$ ، ودرجة حرية مقام $df_2 = (n-2)$ ، يمكن من جدول توزيع مئويات F استخراج القيمة الحرجة (الجدولية) التي تفصل بين منطقي الرفض والقبول ، ويرمز لها بالرمز $F_{1,n-2}^{1-\alpha}$ ، وتظهر علي التوزيع كما يلي:



$F_{1,n-2}^{1-\alpha}$

- **القرار الذي يوصي به الباحث**

يتخذ القرار بخصوص الفرض العدم والبديل بناء على موقع إحصائية الاختبار F^* من مناطق الرفض والقبول ، كما يلي:

if $F^* > F_{1,n-2}^{1-\alpha}$ **we cannot accept** H_o

لا يمكن قبول الفرض العدم إذا كانت إحصائية الاختبار أكبر من القيمة الجدولية ، ويستدل من ذلك على مناسبة النموذج في تمثيل العلاقة بين المتغير التابع والمتغير المستقل.

if $F^* \leq F_{1,n-2}^{1-\alpha}$ **we can accept** H_o

يمكن قبول الفرض العدم ، ويكون افتراض الباحث لنموذج الانحدار الخطي غير صحيح.

ب - اختبار معنوية معامل الانحدار β

يقصد بذلك اختبار ما إذا كان المتغير المفسر (المستقل) X له أثر معنوي ذات دلالة إحصائية

Y

- تباين التقدير $\hat{\beta}_1$ ، يأخذ الصورة التالية:

$$\sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{\sum(x-\bar{x})^2} \quad \textbf{(5-21)}$$

- تباين الحد الثابت $\hat{\beta}_0$

$$\sigma^2_{\hat{\beta}_0} = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x-\bar{x})^2}\right) \quad \textbf{(5-22)}$$

وحيث أن تباين الخطأ العشوائي σ^2 غير معلوم، يتم التعويض عنه بتقديره (MSE) ، ومن ثم نحصل على تقدير تبايني $\hat{\beta}_1$ ، $\hat{\beta}_0$ وهما كالتالي:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{\sum(x-\bar{x})^2} = \frac{MSE}{\sum(x-\bar{x})^2} \quad \textbf{(5-23)}$$

$$\hat{\sigma}^2_{\hat{\beta}_0} = \hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x-\bar{x})^2}\right) = MSE\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x-\bar{x})^2}\right) \quad \textbf{(5-24)}$$

وبأخذ الجذر التربيعي للتباين ، نحصل على الخطأ القياسي : $\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$ ، $\hat{\sigma}_{\hat{\beta}_0} = \sqrt{\hat{\sigma}^2_{\hat{\beta}_0}}$.

وفيما يلي خطوات إجراء الاختبار:

- **صياغة الفروض:**

يأخذ الفرض العدم والفرض البديل الصورة التالية.

الفرض العدم	**المتغير المفسر ليس له أثر معنوي على المتغير التابع**	$H_o : \beta = 0$
الفرض البديل	**المتغير المفسر له أثر معنوي على المتغير التابع**	$H_a : \beta \neq 0$

- **إحصائية الاختبار هي:**

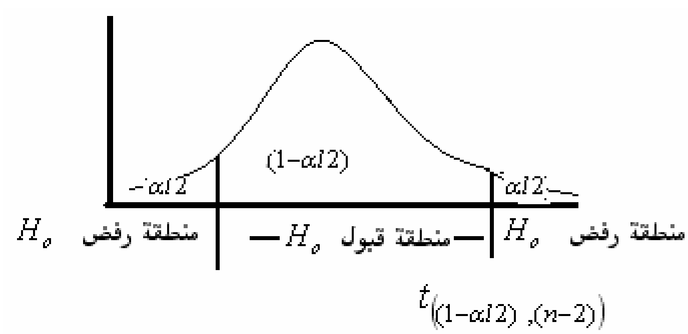
$$t^* = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \qquad \textbf{(5-25)}$$

حيث أن :إحصائية الاختبار t^* **تحت صحة الفرض العدم** H_o **تتبع توزيع** t **بدرجات حرية تساوي :**

$df = (n-1-1 = n-2)$

- **مستوى المعنوية ، ومناطق الرفض والقبول:**

بفرض أن مستوى المعنوية الذي حدده الباحث هو α **، وأن درجة الحرية هي** $df = (n-2)$ **،** **يمكن من جدول توزيع مئويات** t **استخراج القيمة الحرجة (الجدولية) التي تفصل بين منطقي الرفض والقبول ، ويرمز لها بالرمز** $t_{((1-\alpha/2)\ ,(n-2))}$ **، وتظهر علي التوزيع كما يلي:**



- **القرار الذي يوصي به الباحث**

يتخذ القرار بخصوص الفرض العدم والبديل بناء على موقع إحصائية الاختبار t^* من مناطق الرفض والقبول ، كما يلي:

If $|t^*| > t_{((1-\alpha/2)\ ,(n-2))}$ **we cannot accept** H_o

لا يمكن قبول الفرض العدم إذا كانت إحصائية الاختبار أكبر من القيمة الجدولية ، ويستدل من ذلك على أن المتغير المستقل له أثر معنوي على المتغير التابع.

If $|t^*| \le t_{((1-\alpha/2)\ ,(n-2))}$ **we can accept** H_o

يمكن قبول الفرض العدم ، ويكون المتغير المستقل ليس له أثر معنوي على المتغير التابع ، وذلك عند مستوى معنوية α .

تطبيق ٤

في التطبيق السابق أجري الاختبارات التالية:

- **اختبار صلاحية نموذج الانحدار الخطي المقترح لتمثيل العلاقة بين ضغط الدم كمتغير تابع، والوزن كمتغير مستقل، استخدم مستوى معنوية $\alpha = 0.05$.**
- **هل للوزن أثر ذات دلالة على الضغط ، $\alpha = 0.05$**

الحل ٤

أ- اختبار صلاحية نموذج انحدار ضغط الدم على الوزن ، استخدم مستوى معنوية $\alpha = 0.05$.

- **الفرض العدم والفرض البديل .**

H_o **:model is not suitable** نموذج الانحدار المقترح لتمثيل العلاقة بين الضغط والوزن غير مناسب

H_a **:model is suitable** النموذج مناسب

- **إحصائية الاختبار هي:**

$$F^* = \frac{\left(SSR/1\right)}{\left(SSE/(n-2)\right)} = \frac{MSR}{MSE}$$

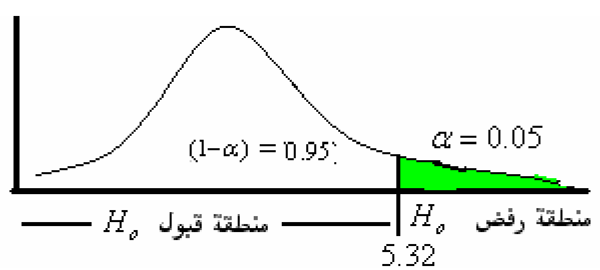
$$SSR = 1213.6646 \ , \ \ SSE = SST - SSR = 1468.1 - 1213.6946 = 254.4054$$

إذا إحصائية الاختبار قيمتها تساوي.

$$F^* = \frac{\left(1213.6646/1\right)}{\left(254.4054/(10-2)\right)} = \frac{1213.6646}{31.800675} = 38.165$$

- **مستوى المعنوية ، ومناطق الرفض والقبول:**

مستوى المعنوية الذي حدده الباحث هو $(\alpha = 0.05)$ **، درجات الحرية بسط** $(df_1 = 1)$ **، ودرجة حرية مقام** $df_2 = 8$ **، تكون قيمة** F **الجدولية هي:** $F_{((1-\alpha)\ ,(1,n-2))} = F^{(0.95)}_{(1,8)} = 5.32$ **، وتظهر علي التوزيع كما يلي:**



- **القرار الذي يوصي به الباحث**

بما أن $F^* = 38.165 > F_{((0.95)\ ,(1,8))} = 5.32$ **إذا لا يمكن قبول الفرض العدم ، ويستدل من ذلك على مناسبة النموذج المفترض لتمثيل العلاقة بين الضغط كمتغير تابع والوزن كمتغير مستقل.**

ب- اختبار معنوية أثر الوزن على الضغط

- **الفرض العدم والفرض البديل**

الفرض العدم	**الوزن ليس له أثر معنوي على الضغط** $H_o : \beta_1 = 0$
الفرض البديل	**الوزن له أثر معنوي على الضغط** $H_a : \beta_1 \neq 0$

ت- إحصائية الاختبار هي:

$$t^* = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}$$

$$\hat{\beta}_1 = 0.8982 \;,\; \hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{\sum (x-\bar{x})^2} = \frac{31.8007}{1504.4} = 0.0211$$

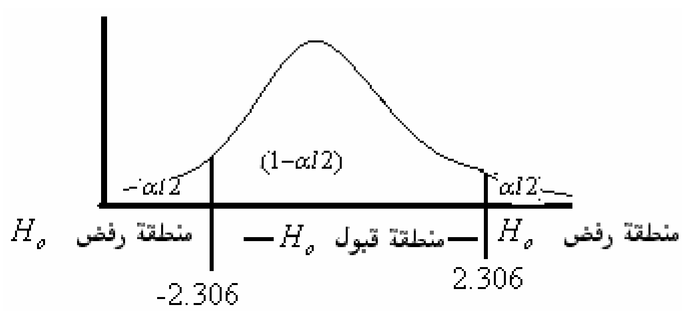
$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{0.0211} = 0.1453$$

$$t^* = \frac{0.8982}{0.1453} = 6.183$$

- **مستوى المعنوية ، ومناطق الرفض والقبول:**

مستوى المعنوية الذي حدده الباحث هو $\alpha = 0.05$ **، وأن درجة الحرية هي** $df = (n-2) = 8$ ،

ومن ثم تكون قيمة t **الجدولية هي:** $t_{((1-\alpha/2)\,,(n-2))} = t_{(0.975\,,8)} = 2.306$



- **القرار الذي يوصي به الباحث**

بما أن $|t^*| = 6.183 \succ t_{(0.975\ ,8)} = 2.306$ **، إذا لا يمكن قبول الفرض العدم ، ويستدل من ذلك أن الوزن له أثر معنوي على الضغط عند مستوى معنوية** $\alpha = 0.05$ **.**

واجب منزلي:

١- المطلوب حساب معامل الارتباط، واختبار معنويته.

٢- تقدير فترة ثقة لمعامل الانحدار β_1 **.**