تحليل الانحدار الخطي المتعدد

Multiple Linear Regression Analysis

المحاضـــــــرة الخامسة

مثال (03): اعتمادا على بيانات مثال نموذج وزن الطفل، أحسب معامل التحديد ومعامل التحديد المعدل وفسر معناهما.

الحل:

لإيجاد **معامل التحديد** نستخدم الصيغة التالية:

$$\mathbf{R^2 = \frac{b^T.X^T.Y - n\overline{Y}^2}{Y^T.Y - n\overline{Y}^2}}$$

لدينا:

$$b^T.X^T.Y - n\overline{Y}^2 = 6596.6948 - 5155.19 = 1441.50$$

$$Y^T.Y - n\overline{Y}^2 = 6653.3853 - 5155.19 = 1498.20$$

ومنه نجد:

$$\mathbf{R^2 = \frac{b^T.X^T.Y - n\overline{Y}^2}{Y^T.Y - n\overline{Y}^2} = \frac{1441.50}{1498.20} = 0.9621}$$

وبهذا نجد أن 96.21 % من التباين أو التغير في أوزان الأطفال قد تم تفسيره من خلال عمر الطفل (X_1) وطوله (X2).

أما **معامل التحديد المعدل** فيحسب وفق الصيغة التالية:

$$R_a^2 = 1 - \left[\frac{n-1}{n-k-1}\right].(1 - R^2) = 1 - \left[\frac{50-1}{50-2-1}\right].(1 - 0.9621) = 0.96048$$

أي أن 96 % تقريبا من التباين في أوزان الأطفال يفسر من خلال التغيرات الحاصلة في عمر الطفل وطوله.

II-8. معامل الارتباط المتعدد

ان معامل الارتباط المتعدد يقيس العلاقة الخطية بين المتغير التابع والمتغيرات المستقلة، ويتم حسابه وفق الصيغة التالية:

$$R = +\sqrt{R^2} \qquad (24)$$

والشيء الملاحظ على معامل الارتباط المتعدد هو:

-كلما اقتربت القيم المقدرة من قيم المشاهدات الفعلية للمتغير التابع ارتفعت قيمة معامل الارتباط المتعدد.
-يأخذ معامل الارتباط المتعدد قيما **غير سالبة**، وبهذه الصفة فهو يختلف عن معامل الارتباط الخطي البسيط الذي يمكن أن يأخذ قيما سالبة، أي:

$$0 \leq R \leq 1$$

II-9. توزيع المعاينة لمقدرات المربعات الصغرى

باستيفاء فرضية التوزيع الطبيعي لحد الخطأ العشوائي (ε) الذي له توقع قدره **صفر** وتباين $(\boldsymbol{\sigma^2})$ فإن:

$(\mathbf{b_i})$ لها **توزيع طبيعي** بتوقع قدره $(\boldsymbol{\beta_i})$ وتباين قدره $[\boldsymbol{\sigma^2}.(\mathbf{X^TX})^{-1}]$.

وبالتالي فإن المقدار:

$$Z = \frac{b_i - \beta_i}{\sigma_{b_i}} \qquad (25)$$

له توزيع طبيعي معياري، هذا إذا كان تباين النموذج (σ_ε^2) معلوما. أما إذا كان التباين غير معلوم فإننا نقدره بتباين البواقي (S_e^2). وبالتالي فإن المقدار:

$$T = \frac{b_i - \beta_i}{\hat{\sigma}(b_i)} \qquad ; i = 0, 1, 2, \dots, k \qquad (26)$$

له توزيع ستيودنت (t) بدرجات حرية $(n\text{-}k\text{-}1)$.

II-10. التقدير بفترة للمعامل (β_i) واختبار الفرضيات حوله

1-فترة الثقة للمعامل (β_i)

إذا كانت درجة الثقة المطلوبة هي: $100\%\ (1-\alpha)$ فان المعامل (β_i) يُقدر بفترة الثقة التالية:

$$\mathbf{b_i} \mp \mathbf{t}_{\left(1-\frac{\alpha}{2};\ n-k-1\right)} \cdot \hat{\sigma}(\mathbf{b_i}) \qquad (27)$$

2- اختبار الفرضيات حول المعامل (β_i)

ان الشكل الشائع للاختبار في هذه الحالة هو:

$$\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases}$$

وبالتالي فإن إحصاء الاختبار المناسب هو:

$$T = \frac{b_i - 0}{\hat{\sigma}(b_i)} = \frac{b_i}{\hat{\sigma}(b_i)} \qquad (28)$$

وبمقارنة قيمة (T) المحسوبة مع القيمة الجدولية (t) عند مستوى المعنوية أو الدلالة المطلوب (α)، فإن قاعدة اتخاذ القرار تكون على النحو التالي:

-إذا كانت القيمة المطلقة لـ (T) المحسوبة أكبر من القيمة الجدولية (t) نرفض الفرضية الصفرية ونقبل بديلتها، وهذا يعني أن **قيمة المعلمة (β_i) تختلف عن الصفر**؛ بعبارة أخرى المتغير المستقل المرتبط بهذه المعلمة يؤثر على المتغير التابع ويُساهم في تفسيره.

-إذا كانت القيمة المطلقة لـ (T) المحسوبة أقل من القيمة الجدولية (t) نقبل الفرضية الصفرية ونرفض بديلتها، لأن (T) المحسوبة وقعت في منطقة قبول الفرضية الصفرية، وهذا يعني أن **قيمة المعلمة (β_i) تساوي الصفر**؛ بعبارة أخرى المتغير المستقل المرتبط بهذه المعلمة لا يؤثر على المتغير التابع ولا يُساهم في تفسيره وبالتالي لا توجد علاقة إرتباط بينهما.

مثال (04): حدد فترة الثقة 95% للمعلمتين (β_1) و (β_2) لنموذج انحدار وزن الطفل على عمره وطوله، ثم بعد ذلك إختبر معنوية متغير العمر والطول كل على حدة.

الحل:

أولا. فترة الثقة للمعلمة (β_1) تأخذ الشكل التالي:

$$\mathbf{b_1} \mp \mathbf{t}_{\left(1-\frac{\alpha}{2};\ n-k-1\right)} \cdot \hat{\sigma}(\mathbf{b_1})$$

لدينا: $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$

ومنه نجد:

$$\mp \mathbf{t}_{\left(\mathbf{1}-\frac{\alpha}{2};\ \mathbf{n-k-1}\right)} = \mp t_{\left(1-\frac{0.05}{2};\ 50-2-1\right)} = \mp t_{(0.975;\ 47)} = \mp 2.014$$

ومنه فان فترة الثقة 95% لمعامل الانحدار (β_1) هي:

$$1.2008 \mp 2.014.(0.21107)$$

$$\Rightarrow (\beta_1) \in [0.7757 \ : \ 1.6259]$$

وهذا يعني أننا نثق بدرجة 95% أن المعلمة الحقيقية (β_1) تقع في المدى ما بين: (0.7757 و 1.6259).

ثانيا. فترة الثقة للمعلمة (β_2) تأخذ الشكل التالي:

$$\mathbf{b_2} \mp \mathbf{t}_{\left(\mathbf{1}-\frac{\alpha}{2};\ \mathbf{n-k-1}\right)} . \hat{\sigma}(\mathbf{b_2})$$

بنفس الطريقة السابقة نجد أن فترة الثقة 95% لمعامل الانحدار (β_2) هي:

$$0.12457 \mp 2.014.(0.018466)$$

$$\Rightarrow (\beta_2) \in [0.08737 \ : \ 0.16177]$$

وهذا يعني أننا نثق بدرجة 95% أن المعلمة الحقيقية (β_2) تقع في المدى ما بين: (0.08737 و 0.16177).

ثالثا. اختبار معنوية (β_1):

نريد اختبار ما يلي:

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

إحصاء الاختبار المناسب هو:

$$\mathbf{T} = \frac{\mathbf{b_1}}{\hat{\sigma}(\mathbf{b_1})} = \frac{1.2008}{0.21107} = \mathbf{5.69}$$

وبمقارنة قيمة (T) المحسوبة مع القيمة الجدولية (t) عند مستوى المعنوية (0.05)، نجد أن (T) المحسوبة (5.69) أكبر من (t) الجدولية (2.014) وبذلك **نرفض** الفرضية الصفرية و**نقبل** بديلتها، وهذا يعني أن **قيمة المعلمة (β_1) تختلف عن الصفر**؛ بعبارة أخرى متغير عمر الطفل يساهم في تفسير تباين وزن الطفل.

رابعا. اختبار معنوية (β_2):

أي نريد اختبار الفرضيتين التاليتين:

$$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

بنفس الطريقة السابقة نجد إحصاء الاختبار المناسب هو:

$$\mathbf{T} = \frac{\mathbf{b_2}}{\hat{\sigma}(\mathbf{b_2})} = \frac{0.12457}{0.018466} = \mathbf{6.746}$$

بما أن (T) المحسوبة (6.746) أكبر من (t) الجدولية (2.014) فإننا **نرفض** الفرضية الصفرية و**نقبل** بديلتها عند مستوى الدلالة أو المعنوية (0.05)، وهذا يعني أن **قيمة المعلمة (β_2) تختلف عن الصفر**؛ بعبارة أخرى متغير طول الطفل يساهم في تفسير تباين وزن الطفل.