الفصل التاسع

التراجع الخطي البسيط

٩ـ١ مقدمة : إحدى مسائل التقدير التي يهتم بها عادة طلبة الشهادة الثانوية ، وأولياؤهم ، وإدارات الجامعات ، هي مسألة الإنجاز المتوقع لطالب في نهاية سنته الجامعية الأولى . فقد نرغب ، مثلاً ، في تقدير ما سيكون عليه معدل طالب في سنته الجامعية الأولى وذلك قبل تسجيل الطالب في الجامعة . وتبدو مثل هذه المسألة للوهلة الأولى عملاً صعباً .

والطريقة الإحصائية لمعالجة هذه المسألة هي من وجوه عدة صياغة للطريقة التي تمليها البداهة ، فإذا توفر لنا بيان إحصائي يعطي الدرجات التي حصل عليها الطالب في الشهادة الثانوية ، معلومات نفسية واجتماعية واقتصادية ، بالإضافة إلى درجات العام الجامعي الأول بالنسبة لعدد كبير من الطلاب الذين سُجلوا في الجامعة وأنهوا عامهم الأول فيها ، فيمكننا تصنيف هؤلاء الطلاب إلى جماعات تمتلك خواصاً متشابهة أو متقاربة ، ومن المؤكد أن الطلاب الذين توفرت لهم حوافز عالية ، وتخرجوا من المرحلة الثانوية بمعدلات عالية وبمواصفات دراسية عامة متميزة ، ينبغي أن يحصلوا ، في المتوسط ، على معدل عال في نهاية عامهم الجامعي الأول . وعلى الوجه الآخر يجب ألا نتوقع حصول الطلاب ، الذين لم تتوفر لهم الحوافز المناسبة ، ولم تكن معدلات نجاحهم في الشهادة الثانوية عالية ، على معدلات في نفس مستوى معدلات الزمرة الأولى من الطلاب . وإذا دفعنا هذه الأفكار إلى نهاياتها النظرية ، فسنتوقع

أن يكون معدل طالب تابعاً في عدة **متحولات تعرّف الخواص** النفسية والفيزيولوجية للطالب . كما تعرّف الشروط المحيطية ، دراسية كانت أم إجتماعية أم اقتصادية ، التي تعرَّض لها الطالب . وبصورة نموذجية نقول أننا نرغب في امتلاك معادلة رياضية تربط معدل الطالب بجميع هذه المتحولات المستقلة ، وبحيث يمكن إستخدام مثل هذه المعادلة للتنبوء .

ويلاحظ القارىء أن مثل هذه المسألة هي من طبيعة عامة جداً . إذ نهتم بالعلاقة بين متحول عشوائي y ، مثلاً ، وبين عدد من المتحولات المستقلة (غير عشوائية في طبيعتها) $x_1, x_2, x_3, \ldots$ وفي مثالنا كان المتحول العشوائي y هو معدل الطالب ، والمتحولات المستقلة يمكن أن تكون :

مرتبة النجاح في الثانوية $= x_1$

درجة اختبار مادة الرياضيات $= x_2$

درجة اختبار شفهي $= x_3$

وهكذا . والهدف النهائي هو أن نقيس x_1, x_2, x_3 من أجل طالب معين ، ثم نبدل هذه القياسات في معادلة التنبوء لنحصل أخيراً على تنبؤ بما سيكونه معدل الطالب في نهاية سنته الجامعية الأولى . وللوصول إلى مثل هذه المعادلة لا بد أن نحدد أولاً المتحولات المستقلة $x_1, x_2, x_3, \ldots$ ثم نحصل على قياس لمدى علاقتها بالمتحول y . وعندئذ نضع معادلة تنبؤ تعبر عن y كتابع في المتحولات المستقلة التي أخترناها .

وسندرس في هذا الفصل ، بصورة رئيسية ، مسألة التنبؤ بقيمة y كتابع خطي في متحول مستقل واحد x . ونعمم في الفصل القادم إلى حالة تابع خطي في عدة متحولات مستقلة .

٩ـ٢ نموذج احتمالي خطي بسيط : سنعتبر مسألة التنبؤ بالدرجة النهائية لمادة الحساب التفاضلي لطالب في سنته الأولى وذلك بالاعتماد على معدله في اختبار الرياضيات في مسابقة تجري للقبول في الجامعة ، وسنعتبر هذه المسألة

obeikandl.com

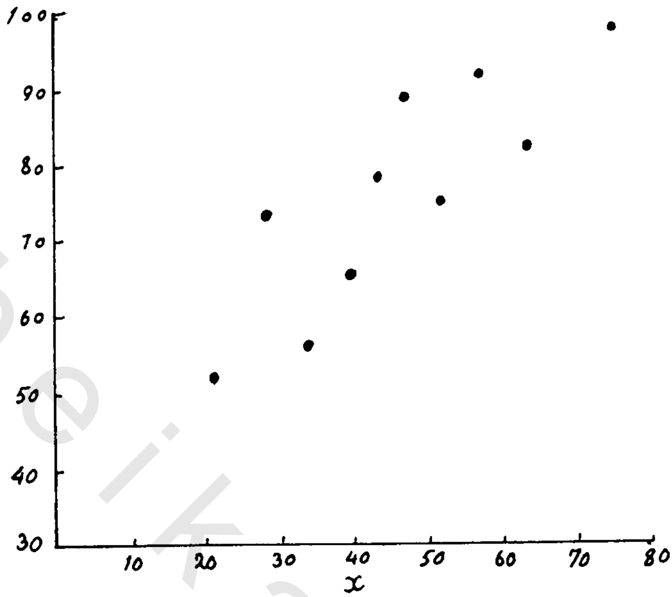
كمدخل إلى موضوعنا هنا . وكما رأينا في الفقرة السابقة فإننا نريد تحديد ما إذا كانت درجة اختبار الرياضيات ذات قيمة جدية في تحديد الدرجة النهائية لمادة الحساب التفاضلي ، بالإضافة إلى رغبتنا في الحصول على معادلة ذات فائدة في مجال التنبؤ . ويقدم الجدول (٩ـ١) عشر درجات في اختبار الرياضيات والمعدلات النهائية الموافقة لها وذلك من أجل عينة من عشرة طلاب ممن أنهوا سنتهم الجامعية الأولى .

جدول ٩ـ١

الطالب	درجة اختبار الرياضيات	الدرجة النهائية في مادة الحساب التفاضلي
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

لنمثل هذه المعلومات بيانياً معتبرين الدرجة النهائية في مادة الحساب المتحول y ، ودرجة اختبار الرياضيات كمتحول x . فنحصل على الشكل (٩ـ١) . ونلاحظ من الشكل مباشرة أن y تميل إلى الزيادة عندما تزداد x . (فهل يكون حصول مثل هذه الظاهرة عائداً إلى الصدفة فقط ، وأنه لا توجد في الحقيقة أية علاقة بين x و y ؟) .

إن إحدى الطرق الممكنة للحصول على معادلة تنبؤ تربط بين y و x هو وضع مسطرة فوق التمثيل البياني وتحريكها حتى تبدو وكأنها تمر عبر أكبر عدد ممكن من النقاط ، ويقدم لنا مثل هذا الوضع ما نسميه « بأفضل تلاؤم »



الشكل ٩ـ١ التمثيل البياني للمعلومات الاحصائية في الجدول ٩ـ١

مع المعلومات الإحصائية المتوفرة . وفي الحقيقة إذا رسمنا مستقيماً عبر النقاط تكون مسألة التنبؤ المطروحة قد انتهت . إذ يمكننا استخدام هذا الخط البياني للتنبؤ بدرجة الطالب في الحساب ، y ، من أجل قيمة معينة لمعدله في اختبار الرياضيات x . وفضلاً عن ذلك نلاحظ أننا اخترنا بهذه الطريقة نموذجاً رياضياً يعبر عن العلاقة التابعية المفترضة بين y و x .

ومن المعروف أنه يمكن التعبير عن معادلة مستقيم على الشكل :

$$y = \beta_0 + \beta_1 x$$

حيث β_0 ترتيب نقطة تقاطع المستقيم مع محور التراتيب (المحور الشاقولي) و β_1 هو ميل المستقيم (أي ظل الزاوية التي يصنعها مع الاتجاه الموجب لمحور الفواصل) . كما نعلم أنه يوجد مستقيم واحد وواحد فقط موافق لمعادلة خطية بسيطة من هذا النوع ، كما توجد ، على العكس ، معادلة واحدة وواحدة فقط موافقة لكل مستقيم . وهكذا فإننا عندما نرسم مستقيماً عبر هذه النقاط

نكون قد اخترنا بشكل آلي معادلة رياضية :

$$Y = \beta_0 + \beta_1 x \qquad (1)$$

وبذلك تتحدد قيمتا β_0 و β_1 بصورة وحيدة .

ويُقال عن النموذج الخطي $y = \beta_0 + \beta_1 x$ أنه نموذج رياضي حتمي لأننا عندما نبدل قيمة x في المعادلة نحصل على قيمة محددة لِ y دون أن يكون هناك أي مجال لخطأ .

والنماذج الحتمية ملائمة لشرح ظواهر فيزيائية والتنبؤ بها عندما يكون خطأ التنبؤ مهملاً من وجهة النظر العملية . وهكذا فإن قانون نيوتن $F = ma$ حيث F هي القوة التي ينتجها جسم متحرك كتلته m وتسارعه a ، يسمح لنا بالتنبؤ بالقوة في معظم الحالات العملية بخطأ صغير يمكن إهماله عملياً .

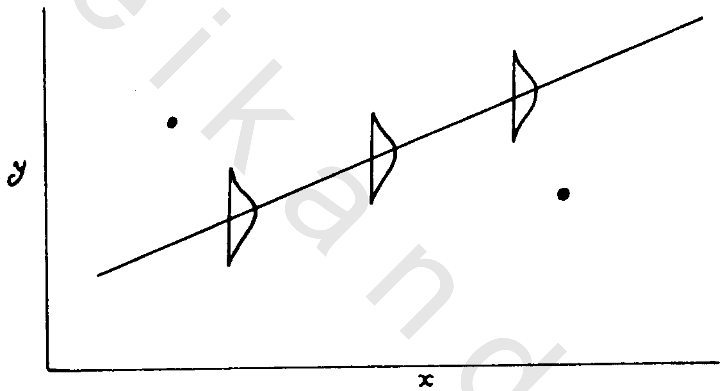
وبالطبع فإن كون الخطأ صغيراً أو كبيراً مسألة نسبية . فقد يكون خطأ قدره بوصة صغيراً جداً عند إقامة دعامة جسر إلا أنه خطأ كبير جداً بالنسبة لصناعة أجزاء من ساعات اليد . وفي العديد من الحالات الفيزيائية لا يمكننا تجاهل خطأ التنبؤ . وفي الحقيقة سنتردد كثيراً في منح الكثير من الثقة لتنبؤ غير مصحوب بقياس لجودة هذا التنبؤ ولهذا السبب فإن الطريقة التي استخدمنا فيها النظر والمسطرة لاختيار خط مستقيم يربط بين x و y هي طريقة غير مقبولة وفائدتها محدودة جداً .

وفي مقابل النماذج الحتمية يمكننا استخدام النماذج الرياضية الاحتمالية لتفسير بعض الظواهر الفيزيائية . وكما نتوقع فإن النموذج الاحتمالي الرياضي يحوي عنصراً واحداً أو أكثر من العناصر ذات الطبيعة العشوائية والتي لها توزيعات احتمالية محددة . وفي مثالنا سنربط بين x و y بالعلاقة :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (2)$$

حيث المقادير ε_i هي متحولات عشوائية توقع كل منها يساوي الصفر وتشتته σ^2. كما نفرض أن أي زوج من المتحولات ε_i ، ε_j ، $i \neq j$ مستقلان عن

بعضهما . وبعبارة أخرى نفرض أن توقع y يرتبط بِـ x خطياً وأن القيم الملحوظة لِـ y ستحيد عن الخط المستقيم الذي يمثل هذه العلاقة الخطية ، أي ستقع فوق المستقيم أو تحته ، بمقدار عشوائي ε يسمى بالخطأ . وعلاوة على ذلك نفرض أن لكل من الأخطاء ε نفس التوزيع الاحتمالي حول هذا المستقيم وذلك بصرف النظر عن قيمة x . ويبين الشكل (٩ـ٢) الخط الذي يمثل توقع y من أجل قيمة معينة لِـ x . كما يبين التوزيع الاحتمالي للخطأ العشوائي ε من أجل عدة قيم لِـ x .



شكل ٩ـ٢ نموذج احتمالي خطي

لنناقش الآن مسألة إيجاد معادلة التنبؤ أو خط التراجع كما يسمى عادة في الإحصاء .

١٠ـ٣ طريقة المربعات الدنيا : إذا رمزنا بِـ $\hat{y}$ للقيمة التي نتنبؤها لـ y فيمكن كتابة معادلة التنبؤ على الشكل :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (3)$$

حيث $\hat{\beta}_0$ و $\hat{\beta}_1$ تقديران لِـ β_0 و β_1 على الترتيب . ويبين الشكل (٩ـ٣) خط التنبؤ الموافق للبيان الإحصائي المعطى في الجدول (٩ـ١) . والخط الشاقولي المرسوم من كل نقطة إلى خط التنبؤ يمثل إنحراف أو حيدان هذه النقاط (وهي تمثل القيم الملحوظة لِـ y) عن خط التنبؤ . وهكذا يمكن تمثيل حيدان النقطة i

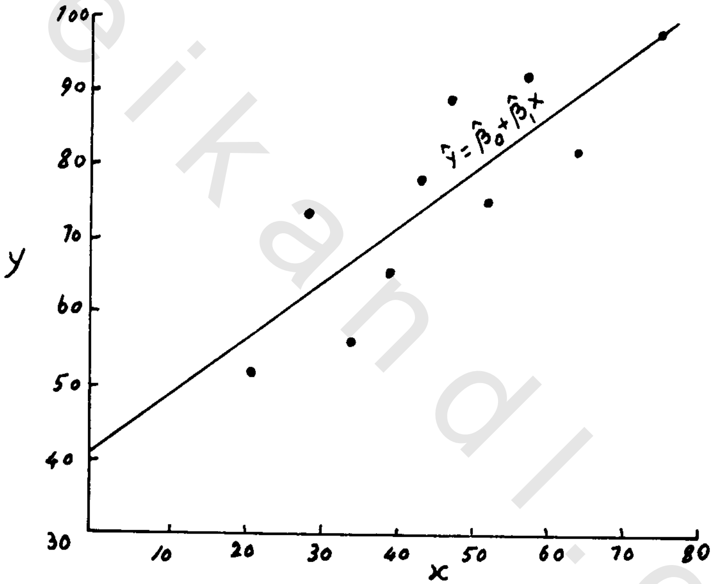
بالفرق :

$$y_i - \hat{y}_i = e_i$$

حيث :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

ولكي يمثل خط التنبؤ $\hat{y}$ أفضل ملاءمة ممكنة للقيم الملحوظة لا بد لنا من أن نجعل هذه الانحرافات أصغر ما يمكن :



شكل ٩ـ٣ معادلة التنبؤ الخطية

وسنعرف قاعدة للحصول على « أفضل ملاءمة » تعتمد على مبدأ «المربعات الدنيا » وستعطي هذه القاعدة تحت شروط معينة أفضل تنبؤ لِ y من أجل قيمة معينة لِ x . وينص مبدأ المربعات الدنيا على أن أفضل ملاءمة هي تلك التي تجعل مجموع مربعات الانحرافات في نهايتها الصغرى. أي نختار تقديري β_0 و β_1 بحيث تكون الكمية : $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$

في نهايتها الصغرى . وترمز SSE لمجموع مربعات الانحرافات أو مجموع مربعات الخطأ . وبتبديل $\hat{y}_i$ نجد :

$$SSE = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \qquad (4)$$

ولحساب قيم β_0 و β_1 التي تجعل هذه الكمية في نهايتها الصغرى نحسب وفقاً لقواعد الحساب التفاضلي المشتقين الجزئيين بالنسبة لِ β_0 و β_1 ونساويهما للصفر وبحل المعادلتين الناتجتين نحصل على β_0 و β_1 . (نعني بالمشتق الجزئي لـ SSE بالنسبة لِ β_i ، ونرمز له بـ $\frac{\partial SSE}{\partial \beta_i}$ ، مشتق الكمية SSE بالنسبة لِ β_i معتبرين جميع المقادير الأخرى في عبارة SSE وكأنها مقادير ثابتة) .

وهكذا نكتب :

$$\frac{\partial SSE}{\partial \beta_1} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} [y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i]^2$$

$$= \sum_{i=1}^{n} (-x_i)(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

أو :

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n} [-(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)] = 0$$

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

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (6)$$

وبحل هاتين المعادلتين نجد :

$$\hat{\beta}_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \qquad (7)$$

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} \qquad (8)$$

ولحساب β_1 و β_0 من أجل البيان الاحصائي في الجدول (٩-١) نقوم أولاً بحساب الجدول التالي :

جدول ٩ـ٢ حساب $\hat{\beta}_0$ و $\hat{\beta}_1$ من أجل البيان الاحصائي في الجدول (٩ـ١)

	y_i	x_i	x_i^2	$x_i y_i$	y_i^2
	65	39	1521	2535	4225
	78	43	1849	3354	6o84
	52	21	441	1092	2704
	82	64	4096	5248	6724
	92	57	3249	5244	3464
	89	47	2209	4183	7921
	73	28	784	2044	5329
	98	75	5625	7350	9604
	56	34	1156	1904	3136
	75	52	2704	3900	5625
المجموع	760	460	23634	36854	59816

وبالتبديل في المعادلتين (7) و (8) نجد :

$$\hat{\beta}_1 = \frac{10(36854) - (460)(760)}{10(23634) - (460)^2} = .76556 = .77$$

و :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 76 - (.76556)(46) = 40.78424 = 40.78$$

وهكذا يكون الخط الأكثر ملاءمة وفقاً لمبدأ المربعات الدنيا والذي يربط بين درجة الطالب في الحساب y ودرجته في اختبار القبول x هو :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

أو :

$$\hat{y} = 40.78 + .77x. \qquad (9)$$

والتمثيل البياني لهذه المعادلة هو المستقيم المبين في الشكل (١٠ـ٣) . وبالعودة إلى هذا المستقيم يمكن الآن أن نتنبأ بقيمة y من أجل قيمة معينة لِ x وذلك بتبديل قيمة x هذه في معادلة التنبؤ (9) . فمثلاً نتنبأ بأن تكون درجة الحساب

obeikandl.com

للطالب الذي نال درجة 50 في اختبار القبول هي :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 40.78 + (.77)(50) = 79.28$$

ونفسر هذا العمل على الشكل التالي : عندما نختار قيمة لِـ x وهي هنا 50 فإننا نتصور مجتمعاً من الطلاب الذين نال كل منهم درجة 50 في اختبار الرياضيات في مسابقة القبول ، ولكل فرد في هذا المجتمع قيمة y هي درجته في مادة الحساب في نهاية السنة الجامعية الأولى . وللقيم y هذه توزيع احتمالي ، باعتبارها قيماً لمتحول عشوائي ، ولهذا التوزيع الاحتمالي متوسط وتشتت ، والقيمة 79.28 التي حصلنا عليها هي تقدير لمتوسط هذا التوزيع الاحتمالي . أي لو أمكننا الحصول على الدرجات النهائية في مادة الحساب لجميع أفراد هذا المجتمع وحسبنا متوسط هذه الدرجات فإننا نقدر القيمة التي سنحصل عليها بِـ 79.28 .

ولكي نتمكن من وضع حدود خطأ لمثل هذا التقدير لا بد لنا من معرفة تشتت التوزيع الاحتمالي σ^2 أو تقدير هذا التشتت . (وهو نفسه تشتت الخطأ العشوائي ε) . وهذا ما سنناقشه في الفقرات التالية .

٩ـ٤ حساب S^2 وتقدير σ^2 : بما أن التشتت σ^2 وهو تشتت الخطأ العشوائي ε غير معروف فلا بد من تقديره . ويمكن البرهان على أن التقدير :

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n-2} \qquad (10)$$

هو تقدير جيد ومنصف لِـ σ^2 وهو مبني على $n - 2$ درجة من الحرية .

ويمكن حساب مجموع مربعات الخطأ SSE مباشرة من التعريف ، أي بحساب $\hat{y}_i$ من أجل كل من القيم x_i حساب الانحرافات $(y_i - \hat{y})$. وبعدها حساب :

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2 .$$

ولكن هذه الطريقة شاقة وغير دقيقة من وجهة النظر الحسابية بسبب العدد الكبير من عمليات الطرح التي تؤدي إلى تضخيم الخطأ الحسابي الناتج عن تدوير الرقم العشري الأخير . والطريقة الأسهل والأدق حسابياً هي التي تستخدم

العلاقة :

$$SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2 - \frac{\hat{\beta}_1}{n} \left[n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \right] \quad (11)$$

ونلاحظ أن الكمية بين قوسين هي الصورة في العلاقة (7) التي تعطي $\hat{\beta}_1$ ، وبالتالي فهي محسوبة من قبل ، ولا تكلفنا عناء جديداً . ومن المستحسن الاحتفاظ عند الحسابات ، بعدد كبير من الأرقام العشرية لتجنب أخطاء كبيرة في تدوير الرقم العشري الأخير في الجواب النهائي .

وفي مثالنا هنا نجد عند التعويض في العلاقة (11) :

$$SSE = 2056 - \frac{.76556}{10} [(10)(36854) - (460)(760)] = 606.03$$

ومنه :

$$s^2 = \frac{SSE}{n-2} = \frac{606.03}{8} = 75.754.$$

ويمكن تلخيص هذه الحسابات في جدول تحليل التشتت وذلك بتقسيم مجموع المربعات الكلي $\sum_{i=1}^{n} (y_i - \bar{y})^2$ إلى جزئين ، الأول يمثل مجموع المربعات ، العائد لخط التراجع ، والآخر يمثل مجموع المربعات العائد للانحرافات حول خط التراجع . والجدول ٩ـ٣ يوضح الفكرة .

٩ـ٥ استقراءات تتعلق بميل خط التراجع β_1 : لا بد أن نستقرىء أولاً ما إذا كانت توجد أصلاً علاقة بين y و x . وبمعنى آخر هل تقدم المعلومات الإحصائية المتوفرة لنا دليلاً كافياً على وجود علاقة خطية بين y و x فوق مجال معين لقيم x ؟ أو هل يكون من المحتمل أن تقع النقاط في التمثيل البياني وفقاً لما نراه في الشكل (٩ـ١) مع أنه لا توجد أية علاقة أو رابطة بين y و x ؟

والسؤال العملي الذي يطرح نفسه هنا يتعلق بقيمة المقدار β_1 الذي يمثل التغير الوسطي في y من أجل تغير في x قدره الواحد . والقول بأن y و x لا يرتبطان ببعضهما خطياً يكافيء القول بأن $\beta_1 = 0$. وهكذا نرغب في اختبار

جدول ٩ـ٣ تحليل التشتت من أجل التراجع الخطي البسيط

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
التراجع	1	$\frac{\hat{\beta}_1}{n}[n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)]$ $= [\Sigma(x_i-\bar{x})(y_i-\bar{y})]^2/\Sigma(x_i-\bar{x})^2$ $= SSR$	SSR	$\sigma^2+\beta_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2$
الخطأ أو الانحرافات حول خط التراجع	$n-2$	$SSE=\sum_{i=1}^{n}(y_i-\hat{y})^2 = SST-SSR$	$\frac{SSE}{(n-2)}$	σ^2
المجموع	$n-1$	$SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$		

الفرضية $\beta_1 = 0$ ضد البديل $\beta_1 \neq 0$. وكما نتوقع فإن التقدير $\hat{\beta}_1$ يلعب دوره الهام كإحصاء اختبار . ولذلك تهمنا معرفة التوزيع الاحتمالي للتقدير $\hat{\beta}_1$. وإذا فرضنا الآن أن الخطأ العشوائي يتبع التوزيع الطبيعي بالإضافة إلى الشروط التي فرضناها سابقاً فيمكن البرهان على أن كلاً من $\hat{\beta}_0$ و $\hat{\beta}_1$ يتبع التوزيع الطبيعي أيضاً وأن :

$$E(\hat{\beta}_1) = \beta_1 \tag{12}$$

$$V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{13}$$

obeikandi.com

وبما أن التشتت σ^2 غير معروف فإننا نستخدم تقديراً له هو s^2 وبالتالي يكون تقدير $V(\hat{\beta}_1)$ هو $s^2/\sum_{i=1}^{n}(x_i-\bar{x})^2$ وبالاستناد إلى ما رأيناه في الفصل الثامن فإن إحصاء الاختبار :

$$t = \frac{\hat{\beta}_1 - \beta_1}{s/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{\hat{\beta}_1 - \beta_1}{s}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (14)$$

يتبع التوزيع t بعدد من درجات الحرية يساوي تلك الموافقة لِ s^2 أي $(n-2)$ درجة من الحرية . وهكذا يكون الاختبار المناسب للفرضية بأن β_1 تساوي قيمة معينة b مثلاً هو الاختبار t الذي درسناه في الفصل الثامن .

لنفرض الآن أننا نرغب في اختبار الفرضية الابتدائية :

$$H_0 : \beta = 0$$

فعندئذ يكون إحصاء الاختبار هو :

$$t = \frac{\hat{\beta}_1 - 0}{s}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (15)$$

وإذا اخترنا مستوى الأهمية α مساوياً $.05$ فإننا نرفض الفرضية H_0 إذا كان $t > t_{\alpha/2}(n-2)$ أو $t < -t_{\alpha/2}(n-2)$ .. وفي مثالنا هنا المبين في الجدول (١٠-١) حيث $n-2=8$ نرفض الفرضية H_0 إذا كان $t > 2.306$ أو $t < -2.306$. وبالتبديل في العلاقة (15) نجد :

$$t = \frac{\hat{\beta}_1}{s}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{.76556}{8.71}\sqrt{2474} = 4.373$$

أي أننا نرفض الفرضية H_0 عند المستوى $\alpha = .05$ ونستنتج أن هناك دلالة تشير إلى وجود علاقة خطية بين y و x .

وبعد أن نقرر وجود علاقة خطية بين y و x نرغب في دراسة هذه العلاقة

obeikandl.com

بتفصيل أكبر . فما هو التغير الذي نتنبؤه في قيمة y عندما تتغير قيمة x بمقدار الواحد ؟ وما هو مدى الثقة الذي يمكن منحه لمثل هذا التنبؤ أو التقدير ؟ وبعبارة أخرى فإننا نريد تقديراً مجالياً لِ β_1 . ويمكن ، قياساً على ما رأيناه في الفصل الثامن ، أن نبين أن مجال الثقة الموافق لِ β_1 ، حيث أمثال الثقة تساوي $(1-\alpha)$ هو :

$$\hat{\beta}_1 \pm t_{\alpha/2} \cdot [\widehat{V(\hat{\beta}_1)}]^{\frac{1}{2}}$$

حيث $\widehat{V(\hat{\beta}_1)}$ هو تقدير تشتت β_1 . أو :

(16)

$$\hat{\beta}_1 \pm \frac{t_{\alpha/2} \cdot s}{\sqrt{\sum_{i=1}^{n}(x_i - x)^2}}$$

وفي حال $1-\alpha = .95$ نجد مجال الثقة :

(17)

$$\hat{\beta}_1 \pm \frac{t_{.025} \cdot s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

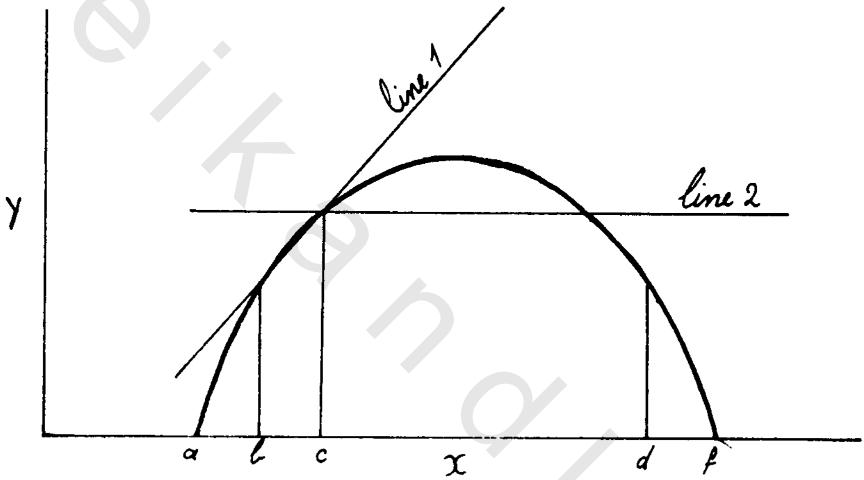
وفي مثالنا هنا نحصل عند التعويض في (17) على :

$$.77 \pm \frac{(2.306)(8.70)}{\sqrt{2474}} = .77 \pm .40$$

والجدير بالملاحظة هو أننا لو لم نرفض الفرضية الابتدائية $\beta = 0$ لما كان تفسير ذلك بالضرورة أنه لا توجد علاقة بين y و x . ففي المقام الأول يجب أن نتذكر أنه يمكن أن نرتكب خطأً من النوع II أي قبول فرضية خاطئة ، وفي المقام الثاني يمكن أن تكون هناك علاقة بين y و x ولكنها علاقة غير خطية أي علاقة من الدرجة الثانية أو أكثر . وهكذا يعني قبول الفرضية $\beta_1 = 0$ أننا نميل إلى الاعتقاد بعدم وجود علاقة خطية بين y و x ، ويصبح من الضروري

تقصي وجود علاقة من الدرجة الثانية أو أكثر ، هذا في حال غياب أية معلومات سلفية عن شكل وطبيعة العلاقة بين y و x .

ويمثل الشكل (٩ـ٤) علاقة من الدرجة الثانية بين y و x فوق المجال $a \leqslant x \leqslant f$. ونلاحظ أن الخط المستقيم المبين في الشكل يقدّم معادلة تنبؤ جيدة إذا كانت الملاءمة مقتصرة على المجال الصغير $b \leqslant x \leqslant c$ ، أي إذا كانت قيم x التي نهتم بدراستها هي قيم المجال بين b و c فقط .



شكل ٩ـ٤ علاقة منحنية بين y و x

وهكذا يجب الحذر في استخلاص النتائج عندما لا توجد دلالة كافية على أن $\beta_1 \neq 0$. فقد يكون الخطأ في شكل النموذج الاحتمالي الذي اخترناه لتمثيل الحالة الفيزيائية المدروسة . وسنعود لدراسة هذه المسألة بصورة مفصلة في نهاية هذا الفصل .

٩ـ٦ التنبؤ بقيمة توقع y من أجل قيمة معينة لِ x : لنفرض أن x و y ترتبطان خطياً وفقاً للنموذج الاحتمالي المذكور في الفقرة (٩ـ٢) ، فعندئذ تمثل العلاقة $E(y/x) = \beta_0 + \beta_1 x$ توقع y من أجل قيمة معينة لِ x . وبما أن خط الملاءمة :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

يمثل تقديراً للعلاقة الخطية الحقيقية بين y و x ، أي أنه يشتمل على تقدير لكل من β_0 و β_1 ، فيمكن استخدام $\hat{y}$ لتقدير قيمة توقع y ، وتقدير قيمة بعينها لـ y من أجل قيمة معطاة لِ x . ويبدو من الطبيعي أن يكون خطأ التنبؤ أو التقدير مختلفاً في هاتين الحالتين . وسنناقش في هذه الفقرة الحالة الأولى أي تقدير قيمة توقع y من أجل قيمة معطاة لِ x .

ويلاحظ القارىء في الشكل (٩ـ٥) خطين يمثل أحدهما الخط

$$E(y/x) = \beta_0 + \beta_1 x$$

ويمثل الآخر خط الملاءمة أو معادلة التنبؤ :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

ونلاحظ لتوّنا أن الخطأ في التنبؤ بقيمة توقع y عندما تكون $x = x_p$ ، مثلاً ، هو المسافة بين الخطين من أجل $x = x_p$ ، وأن هذا الخطأ يتزايد كلما ابتعدنا عن نقطة التقاطع في اتجاه أي من طرفي المجال الذي يتحول فوقه x .

ويمكن البرهان على أن قيمة التنبؤ :

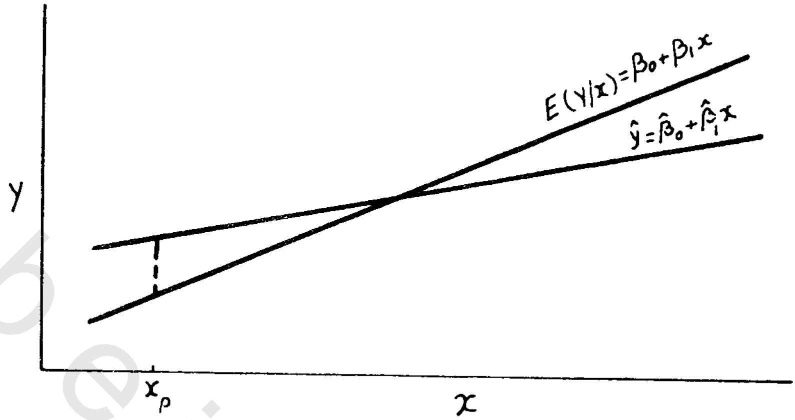
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

هو تقدير منصف لِ $E(y|x)$ ، أي أن $E(\hat{y}) = \beta_0 + \beta_1 x$ ، وأن هذا التنبؤ $\hat{y}$ يتبع التوزيع الطبيعي بتشتت هو :

(18)

$$V(\hat{y}) = \sigma^2 \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]$$

ولتقدير هذا التشتت نضع في عبارته S^2 بدلاً من σ^2 .



شكل ٩ـ٥ قيمتا توقع y والتنبؤ بـ y .

ويمكن إستخدام النتائج السابقة لاختبار فرضية تتعلق بتوقع y من أجل قيمة معينة لِ x ولتكن x_p مثلاً . وهذا بالطبع يسمح لنا باختبار فرضية حول، β_0 ، تقاطع خط التراجع مع محور التراتيب . وهي قيمة توقع y من أجل الحالة الخاصة $x_p = 0$. وهكذا نتمكن من اختبار الفرضية :

$$H_0 : E(y | x = x_p) = E_0$$

حيث E_0 هي القيمة العددية المفترضة لِ $E(y)$ من أجل $x = x_p$. ويمكن البرهان على أن الكمية :

(19)

$$t = \frac{\hat{y} - E_0}{(\text{تقدير } \sigma_{\hat{y}})}$$

$$= \frac{\hat{y} - E_0}{s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_1^n (x_i - \bar{x})^2}}}$$

تتبع التوزيع t بـ $(n-2)$ درجة من الحرية . مما يسمح لنا بتطبيق الاختبار t كما وصفناه سابقاً .

ومجال الثقة الذي أمثاله $(\alpha-1)$ لتوقع y علماً أن $x = x_p$ هو :

$$\hat{y} \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{1}^{n} (x_i - \bar{x})^2}} \qquad (20)$$

وعلى سبيل المثال ، إذا أردنا تقدير معدل مادة الحساب لطالب كانت علامته في مسابقة القبول $x_p = 50$ نجد :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

أو :

$$\hat{y} = 40.78 + (.77)\ (50) = 79.28.$$

والـ 95% مجال ثقة لِـ $E(Y)$ علماً أن $x = 50$ هو :

$$79.28 \pm (2.306)(8.70)\sqrt{\frac{1}{10} + \frac{(50-46)^2}{24}}$$

أو :

$$79.28 \pm 6.54.$$

٩ـ٧ التنبؤ بقيمة y من أجل قيمة معينة لِـ x : لنفرض أن معادلة التنبؤ التي حصلنا عليها من القياسات العشرة في الجدول (٩ـ١) قد استُخدمت للتنبؤ بالدرجة النهائية في مادة الحساب لطالب جديد اخترناه بصورة عشوائية .

ومع أهمية التنبؤ بتوقع y من أجل قيمة معطاة لِـ x ، فإننا نهتم بصورة رئيسية ، في مثال كالمثال المعطى في الجدول (٩ـ١) ، بالتنبؤ بما ستكون عليه الدرجة النهائية في الحساب لطالب معين اخترناه عشوائياً من المجتمع الذي تتناوله الدراسة ، وذلك بالاستفادة من معادلة التنبؤ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ التي حصلنا عليها بالاستناد إلى البيان الإحصائي الملحوظ . وإذا كانت درجة الطالب في مسابقة القبول هي x_p فإننا نرى بوضوح أن خطأ التنبؤ (وهو الفرق بين $\hat{y}$

والدرجة الفعلية y التي سينالها الطالب) مؤلف من مركبتين . ذلك لأنه يمكن كتابة درجة الطالب y على الشكل :

$$y = \beta_0 + \beta_1 x_p + \varepsilon$$

حيث يساوي $(\hat{y} - y)$ إنحراف $\hat{y}$ عن توقع y وهو الانحراف الذي تعرضنا له في الفقرة السابقة بالإضافة إلى الكمية العشوائية ε التي تمثل إنحراف درجة الطالب الفعلية عن قيمة التوقع . وهكذا فإن تشتت الخطأ عند التنبؤ بقيمة y سيتجاوز حتماً ذلك المتعلق بالتنبؤ بتوقع y .

ويمكن البرهان على أن تشتت الخطأ عند التنبؤ بقيمة y من أجل $x = x_p$ أي تشتت المقدار $(\hat{y} - y)$ هو :

$$\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right] \qquad (21)$$

وعندما تكون n كبيرة جداً فإن الحدين الثاني والثالث من العبارة ضمن قوسين سيصبحان صغيرين جداً ويقترب تشتت $(\hat{y} - y)$ من σ^2. ويمكننا الاستفادة من هذه النتائج لوضع التقدير المجالي التالي لِ y علماً أن $x = x_p$:

$$\hat{y} \pm t_{\alpha/2} \cdot s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \qquad (22)$$

وعلى سبيل المثال . إذا أحرز طالب معين درجة $x_p = 50$ في مسابقة القبول فيمكننا التنبؤ بأن معدله النهائي في الحساب سيكون ضمن المجال :

$$79.28 \pm (2.306)(8.70) \sqrt{1 + \frac{1}{10} + \frac{(50 - 46)^2}{2474}}$$

أو :

79.28 21.10.

ولنلاحظ أن عدد الملاحظات n يمكن أن يكون أكبر بكثير من الملاحظات العشر المبينة في الجدول (٩ـ١) مما سيجعل عرض المجال أقل مما نراه هنا .

٩ ـ ٨ أمثال الترابط : كثيراً ما نرغب في الحصول على مؤشر يدل على قوة العلاقة الخطية بين المتحولين y و x ، ويكون مستقلاً عن سلّمي القياس المستخدمين لقياس x و y . ويدعى مثل هذا المؤشر « الترابط الخطي بين x و y » .

والقياس المستخدم في الإحصاء بصورة عامة من أجل الترابط الخطي يدعى أمثال الترابط بين x و y . ونعرفه على الشكل التالي :

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{23}$$

$$= \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\left[n\sum x_i^2-(\sum x_i)^2\right]\left[n\sum y_i^2-(\sum y_i)^2\right]}}$$

والشكل الأخير لأمثال الترابط أسهل حسابياً باعتبار أننا حصلنا على العديد من الكميات المذكورة عند حساب $\hat{\beta}_1$. وإذا رمزنا للصورة $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$ بـ s_{xy} فيمكن كتابة r على الشكل :

$$r = \frac{s_{xy}}{s_x \cdot s_y} \tag{24}$$

كما يمكن كتابة عبارة $\hat{\beta}_1$ المذكورة في (7) على الشكل :

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} \qquad (25)$$

ومن (24) و (25) نجد بسهولة أن :

$$\hat{\beta}_1 = r . \frac{s_y}{s_x} \qquad (26)$$

وبما أن كلاً من الانحرافين المعياريين s_x و s_y موجب بالتعريف فإن r و $\hat{\beta}_1$ ، أي أمثال الترابط بين x و y ، وتقدير β_1 (ميل خط التراجع) ، هما من إشارة واحدة .

ويمكن حساب أمثال الترابط r بين x و y في الجدول (٩-١) باستخدام (23) والاستفادة من أن حسابات $\hat{\beta}_1$ أعطتنا :

$$[n \Sigma x_i y_i - (\Sigma x_i)] = 18940$$

و :

$$[n \Sigma x_i^2 - (\Sigma x_i)^2] = 24740.$$

ومنه :

$$r = \frac{18940}{\sqrt{(24740)(20560)}} = 0.84$$

وبالعودة إلى العلاقة (26) نجد أن r = 0 عندما تكون $\hat{\beta}_1 = 0$. وهكذا يتضمن كون r = 0 عدم وجود علاقة خطية بين x و y . أما كون r موجباً فيتضمن أن خط التراجع يميل في إتجاه إلى الأعلى وإلى الأيمن ، بينما يتضمن كون r سالباً أن خط التراجع يميل في إتجاه إلى الأدنى وإلى الأيسر .

وإذا عدنا الآن إلى جدول تحليل التشتت وهو الجدول (٩-٣) فيمكن بعد استخدام r^2 إعادة كتابته على الشكل التالي :

جدول ٩ـ٤ تحليل التشتت من أجل التراجع الخطي البسيط

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
التراجع	1	$SSR = r^2 \sum_{i=1}^{n}(y_i-\bar{y})^2$	SSR	$\sigma^2 + \beta_1^2 \sum_{i=1}^{n}(x_i-\bar{x})^2$
الخطأ أو الانحرافات حول خط التراجع	$n-2$	$SSE = (1-r^2)\sum_{i=1}^{n}(y_i-\bar{y})^2$	$\frac{SSE}{(n-2)}$	σ^2
المجموع	$n-1$	$SST = \sum_{i=1}^{n}(y_i-\bar{y})^2$		

ومن عبارة SSE في هذا الجدول يتضح أن :

(27)

$$r^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - SSE}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

وبما أن مجموع مربعات الخطأ SSE لا يمكن أن يكون سالباً فإن :

$$0 \leqslant r^2 \leqslant 1$$

وسيكون أمثال الترابط r مساوياً -1 أو $+1$ فقط عندما تقع جميع نقاط العينة على استقامة واحدة تكون هي بالطبع خط الملاءمة أي عندما يكون $SSE = 0$.
وفي الواقع نرى أن :

(28)

$$r^2 = \frac{SST - SSE}{SST} = \frac{SSR}{SST}$$

يمثل النسبة بين التخفيض في مجموع مربعات الانحرافات الناتج عن استخدامنا

للنموذج الخطي إلى مجموع مربعات الانحرافات حول المتوسط $\bar{y}$ الذي كان سيمثل تقدير توقع y فيما لو تجاهلنا المتحول x كلية . وهكذا يبدو أن r^2 يقدم تفسيراً أجدى لقوة أو مدى العلاقة بين x و y من التفسير الذي يقدمه أمثال الترابط r .

وتجدر ملاحظة أن أمثال ترابط العينة r يمثل تقديراً لأمثال ترابط المجتمع ρ الذي نحصل عليه فيما لو أمكن لحساباتنا المعتمدة على العلاقة (23) أن تشمل كل النقاط في المجتمع .

٩ـ٩ منحنيات التراجع من نوع كثيرات الحدود من درجة $k \geq 2$:

كثيراً ما نحصل على بيان إحصائي حول متحولين y و x لا تكون العلاقة المفترضة بينهما من النوع الخطي . وغالباً ما نفترض في مثل هذه الحالات وجود علاقة بين المتحولين على شكل كثيرة حدود غير خطية ، على أساس أنها تقدم الصورة الحقيقية للعلاقة القائمة بين x و y . وعلى سبيل المثال ، يمكن أن نفترض :

(29) (كثيرة حدود من الدرجة الثانية) ، $y = \alpha + \beta x + \gamma x^2$

أو :

(30) (كثيرة حدود من الدرجة الثالثة) ، $y = \alpha + \beta x + \gamma x^2 + \delta x^3$

وهكذا يمكن أن نكتب بصورة عامة :

(31) (الدرجة k) ، $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k$

حيث k عدد صحيح موجب . وكلما رفعنا الدرجة بمقدار 1 فإننا نفترض ، كقاعدة ، وجود انحناء جديد في شكل المنحني . وإذا أردنا ملاءمة كثيرة حدود من الدرجة k للبيان الاحصائي المتوفر بين أيدينا ، فإننا نلجأ إلى تطبيق مبدأ المربعات الدنيا أي نحسب ، كالمعتاد ، النهاية الصغرى للتابع :

(32)

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - \ldots\ldots - b_k x_i^k)^2$$

ونجد المعادلات التالية من أجل الحالتين $k = 2$ و $k = 3$ (والتعميم إلى الحالات $k = 4$ أو أكثر يجب أن يصبح واضحاً) . وتسمى هذه المعادلات عادة بالمعادلات الناظمية :

حالة $k = 2$:

$$\sum y = na + b\sum x + c\sum x^2$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \qquad (33)$$

حالة $k = 3$:

$$\sum y = na + b\sum x + c\sum x^2 + d\sum x^3 \qquad (34)$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 + d\sum x^4$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 + d\sum x^5$$
$$\sum x^3 y = a\sum x^3 + b\sum x^4 + c\sum x^5 + d\sum x^6$$

حيث d, c, b, a هي تقديرات لِـ $\delta, \gamma, \beta, \alpha$ على الترتيب ، وكل المجاميع $\sum$ تمتد من $i = 1$ إلى $i = n$. (وقد حذفنا للسهولة الدليل i على كل x و y ، في الحقيقة كل x هي x_i وكل y هي y_i) . وسنأخذ المثال التالي لتوضيح حالة علاقة مفترضة من الدرجة الثالثة بين y و x .

جدول ٩ـ٥ اختبار شابّر ـ ريجلر لمعجونة الورق عند خلطها

عدد ساعات الخلط	قياس شابّر ـ ريجلر بالدرجات
x	y
1	17
2	21
3	22
4	27
5	36
6	49
7	56
8	64
9	80
10	86
11	88
12	92
13	94

والمعادلات الناظمية في (4 3) تصبح في هذا المثال :

$$732 = 13a + 91b + 819c + 8281d$$
$$6485 = 91a + 819b + 8281c + 89271d$$
$$65097 = 819a + 8281b + 89271c + 1002001d$$
$$696509 = 8281a + 89271b + 1002001c + 11462759d$$

وبحل هذه المعادلات (سنقدم في الفصل القادم طريقة لحل مثل هذه المعادلات) نحصل على معادلة منحني التراجع :

$$\hat{y} = 21.7273 + 5.8396x + 2.3438x^2 + 0.1133x^3 \qquad (35)$$

وإذا فرضنا أنه يمكن تمثيل كل ملاحظة على الشكل :

$$y_i = \alpha + \beta x_i + \gamma x_i^2 + \delta x_i^3 + \varepsilon_i \qquad (36)$$

حيث المقادير ε_i هي متحولات عشوائية مستقلة فيما بينها ويتبع كل منها التوزيع الطبيعي بمتوسط يساوي الصفر وتشتت يساوي σ^2 ، فيمكن

obeikandi.com

وضع تحليل التشتت المبين في الجدول (٩ـ٦) واختبار أهمية علاقة التراجع ككل.

جدول ٩ـ٦ تحليل التشتت لملاءمة كثيرة حدود من الدرجة الثالثة للبيان الإحصائي في الجدول ٩ـ٥ .

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	بنسبة F
منحني التراجع	3	10444.78	3481.59	
الانحرافات حول منحني التراجع أوالخطأ .	9	49.99	5.55	627.31
المجموع	12	10494.77		

وتتضح من حساب النسبة $= 627.31$ $F = \dfrac{3481.59}{5.55}$ بدرجات من الحرية $\nu_1 = 3$ و $\nu_2 = 9$ أهمية التراجع . أي أنه يمكن تفسير معظم التغيرات في y باستخدام علاقة التراجع المفترضة من الدرجة التكعيبية . ومن التمثيل البياني المبدئي للمعلومات الإحصائية في الجدول (٩ـ٥) سيتوقع الباحث العلمي بدون شك مثل هذه النتيجة .

لنقف الآن برهة ولنستعرض إمكانية أخرى . فلنفرض أنه بعد ملاءمة خط مستقيم برزت نتائج تدعونا للاعتقاد بأن ملاءمة منحن تراجع حذف من الدرجة الثانية سيكون مناسباً أكثر للمعلومات الإحصائية المتوفرة بين أيدينا فعندئذ سنحل مجموعة المعدلات الناظمية التالية الموافقة لحالة $k = 2$:

$$732 = 13a + 91b + 819c$$
$$6485 = 91a + 819b + 8281c$$
$$65097 = 819a + 8281b + 89271c$$

وبحلّها نحصل على معادلة التراجع :

$$\hat{y} = 2.6853 + 7.9885x - .0365\,x^2. \quad (37)$$

ومع افتراض توفر الشروط الأساسية من توزيع طبيعي وغيره مما ذكرناه أعلاه يمكن وضع جدول تحليل التشتت التالي :

جدول ٩ـ٧ تحليل التشتت لملاءمة كثيرة حدود من الدرجة الثانية للبيان الإحصائي في الجدول ٩ـ٥.

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات
منحني التراجع	2	10180.22	5090.11
الانحرافات حول منحني التراجع	10	314.55	31.46
المجموع	12	10494.77	

أما المعادلات الناظمية ومعادلة التراجع وتحليل التشتت الذي سنحصل عليه في حالة ملاءمة علاقة خطية للبيان الاحصائي في الجدول (٩ـ٥) فستكون كما يلي :

$$732 = 13a + 91b$$
$$6485 = 91a + 819b$$

وبحلها نجد معادلة التراجع :

$$\hat{y} = 3.962 + 7.478x \qquad (38)$$

وتحليل التشتت هو :

جدول ٨ـ٩ تحليل التشتت لملاءمة علاقة خطية للبيان الاحصائي في الجدول ٩ـ٥ .

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات
خط التراجع	1	10177.59	10177.59
الانحرافات حول خط التراجع	11	317.18	28.83
المجموع	12	10494.77	

وبمقارنة مجموع المربعات العائد للتراجع في الجدولين (٩ـ٧) و(٩ـ٨) ، نجد أن التغير الذي تفسره العلاقة التربيعية أكبر من التغير الذي تفسره العلاقة

obeikandl.com

الخطية . والفرق بين مجموعي المربعات وهو 2.63 = 10177.59 - 10180.22 . يسمى « التخفيض في مجموع المربعات الكلي » الناشيء عن إضافة الحد التربيعي إلى علاقة التراجع الخطية . وبصورة مشابهة إذا حسبنا الفرق بين مجموعي مربعات التراجع الناتجين عن ملاءمة منحن تكعيبي ومنحن تربيعي على الترتيب نحصل على التخفيض في مجموع المربعات الكلي الناشيء عن إضافة الحد التكعيبي إلى علاقة التراجع من الدرجة الثانية . ونجد ، في مثالنا ، أن هذا الفرق يساوي :

1044.78 - 10180.22 = 246.56

وبالعودة إلى الجدول (٩ـ٦) نتمكن الآن من تجزئة مجموع مزبعات التراجع على الشكل المبين في الجدول (٩ـ٩) التالي :

جدول ٩ـ٩ تحليل التشتت لملاءمة علاقة تكعيبية للبيان الإحصائي في الجدول ٩ـ٥.

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات
الحد الخطي	1	10177.59	10177.59
الحد التربيعي	1	2.63	2.63
الحد التكعيبي	1	264.56	264.56
الانحرافات حول منحني التراجع	9	49.99	5.55
المجموع	12	10494.77	

وبالقيام بالاختبارات F المناسبة نجد أن الحدين الخطي والتكعيبي مهمان ، أما الحد التربيعي فغير مهم . ولا بد أن يقف القارىء أمام مثل هذه النتيجة ويفكر برهة . فلنفرض أننا بعد ملاءمة كثيرة الحدود من الدرجة الثانية في العلاقة (37) . جزأنا مجموع مربعات التراجع في الجدول (٩ـ٧) إلى مركبة خطية ومركبة تربيعية وقمنا باختبار أهميتها وفقاً للاختبار F ، فسنجد بالطبع أن المركبة الخطية هامة . ولكن المركبة التربيعية غير هامة . وقد تقودنا مثل هذه النتيجة إلى القول بأننا لا نستطيع الافتراض بأن العلاقة بين x و y منحنية

بل خطية . ومن الواضح خطأ مثل هذه النتيجة فالحقيقة هي أن الحد التربيعي فقط غير هام ، وعدم أهمية الحد التربيعي لا تعني بالضرورة أن الحد التكعيبي أو حدود من درجة أعلى سوف لا تكون هامة . ولذلك فإنه يجب ألا نوقف عملية ملاءمة حدود إضافية لمجرد أن نعثر على حد غير هام . ولكن كم من الحدود غير الهامة المتتالية يجب أن نعثر عليها قبل إيقاف عملية التلاؤم ؟ لا يوجد جواب دقيق على هذا السؤال إلا أنه ، كقاعدة تجريبية جيدة ، يمكن القول بأنه يجب الإصرار على حدين متتاليين غيرها مّين قبل التوقف عن ملاءمة حدود إضافية . ولا يعني هذا أنه يجب الاستمرار في ملاءمة حدود إضافية حتى العثور على حدين متتاليين غير هامين إذ يجري ، بصورة عامة ، تحديد درجة كثيرة الحدود التي سنقوم بملاءمتها سلفاً وبدءاً من اعتبارات تفرضها طبيعة المسألة المدروسة .

١٠ـ١٠ كثيرات الحدود المتعامدة : عندما تكون قيم x (المتحول المستقل) متساوية البعد عن بعضها البعض فيمكن اللجوء إلى طريقة أخرى في الملاءمة لها من الميزات العديدة ما يزكيها بالمقارنة مع الطرق الأخرى ، وهي طريقة كثيرات الحدود المتعامدة . وقد لاحظنا عند ملاءمة كثيرات الحدود المختلفة في الفقرة السابقة أنه كان من الضروري أن نبدأ ثانية ومن جديد في حل مجموعة جديدة من المعادلات الناظمية . ولكن طريقة كثيرات الحدود المتعامدة تعفينا من ذلك إذ تمكّننا من الاحتفاظ بقيم الوسطاء التي حسبناها في كل مرحلة سبقت وان نكتفي ببساطة بحساب الوسيط المتعلق بالحد الجديد الذي أضفناه إلى علاقة تمّ تحديدها وذلك لرفع درجة كثيرة الحدود بمقدار الواحد . وسنبين فيما يلي آلية هذه الطريقة التي سيجدها القارىء أسهل استخداماً بكثير من الطريقة المبينة في الفقرة السابقة ، وخاصة عندما تكون درجة كثيرة الحدود أكبر من 2 .

وسنكتفي هنا بتوضيح الطريقة في الحالة التي تختلف فيها قيم x المتتالية

بمقدار الواحدة ، وحيث يكون لكل قيمة لِ x قيمة واحدة لِ y . وإذا لم تختلف قيم x عن بعضها بمقدار الواحدة فيمكن تغيير سلّم القياس بحيث نعتبر واحدة القياس مساوية للمجال الذي يفصل بين كل قيمتين متتاليتين أي نقسم كل قيمة لِ x على طول هذا المجال ثم نبدأ بتطبيق الطريقة التي سنشرحها فيما يلي والتي توفر على الباحث الوقت وتسمح له بأن يحسب ويقيّم لتوه ، وخطوة فخطوة ، ما ستقدمه ملاءمة كل حد إضافي في عبارة تابع التراجع .

يمكن البرهان على أنه يمكن التعبير دائماً عن أية كثيرة جدود عادية من النوع :

$$y = \beta_0 + \beta_1 x + \dots + \beta_k x^k \qquad (39)$$

على الشكل :

$$y = A_0 + A_1 \xi_1' + \dots + A_k \xi_k' \qquad (40)$$

حيث المقادير A_i $(i = 1,2,\dots,k)$ وسطاء ثابتة و ξ_i' $(i = 1,2,\dots,k)$ هي كثيرات حدود متعامدة ، (ونقول أن كثيرتي الحدود ξ_i' ، ξ_k' ، $i \neq k$ ، متعامدتان إذا كان $\Sigma \xi_i' \xi_k' = 0$ عندما تأخذ x مجموعة . معينة من القيم ، وحيث تعني إشارة الجمع أننا نحسب أولاً الجداء $\xi_i' \xi_k'$ من أجل كل قيمة من قيم x في تلك المجموعة ثم نجمع قيم هذه الجداءات) . وفي حالتنا حيث تأخذ x القيم $1,2,\dots,n$ يمكن التعبير عن أول ثلاثة من كثيرات الحدود المتعامدة على الشكل :

$$\xi_1' = \lambda_1 (x - \bar{x}) \quad , \qquad (41)$$

$$\xi_2' = \lambda_2 \left[(x - \bar{x})^2 - \frac{n^2 - 1}{12} \right] \quad , \qquad (42)$$

$$\xi_3' = \lambda_3 \left[(x - \bar{x})^3 - (x - \bar{x}) \frac{3n^2 - 7}{20} \right] \quad , \qquad (43)$$

حيث λ_i $(i = 1,2,3)$ أعداد ثابتة تعتمد على n ونختارها بحيث تكون قيم ξ_i' أعداداً صحيحة مختزلة إلى أدنى حد ممكن . وللتوضيح نقدم في الجدول (٩ــ١٠) قيم ξ_i' من أجل كثيرات حدود من الدرجة $k = 1,2,3,4$

وذلك في كل من الحالات التي تكون فيها n = 1,2,3,4,5 . وتوجد جداول معدّة لقيم ξ'_i تصل حتى n = 104 يمكن العثور عليها في عدد من المراجع الخاصة بالجداول الإحصائية .

جدول ٩ـ١٠ بعض قيم ξ'_i من أجل $1 \leqslant n \leqslant 5$, $1 \leqslant k \leqslant 4$

درجة كثيرة الحدود

n	k = 1	k = 2		k = 3			k = 4			
	ξ'_1	ξ'_1	ξ'_2	ξ'_1	ξ'_2	ξ'_3	ξ'_1	ξ'_2	ξ'_3	ξ'_4
1	− 1	− 1	+ 1	− 3	+ 1	− 1	− 2	+ 2	− 1	+ 1
2	+ 1	0	− 2	− 1	− 1	+ 3	− 1	− 1	+ 2	− 4
3		+ 1	+ 1	+ 1	− 1	− 3	0	− 2	0	+ 6
4				+ 3	+ 1	+ 1	+ 1	− 1	− 2	− 4
5							+ 2	+ 2	+ 1	+ 1

وتعود أفضلية هذه الطريقة إلى أن ξ_i متعامدة ، مما يجعلها غير مترابطة فيما بينها ، وعدم الترابط هذا يسمح لنا بحساب كل من أمثال التراجع بصورة مستقلة عن جميع الأمثال الأخرى ، ويؤدي إلى حساب مجموع المربعات العائدة إلى ملاءمة حدود إضافية بطريقة سهلة ، كما يقود إلى طريقة مريحة لاختبار أهمية كل من أمثال التراجع أي أهمية الحد الخطي ، التربيعي ، التكعيبي ، الخ . وذلك فور حساب هذه الحدود .

لنستعرض الآن الطريقة العامة لتحديد المقادير A_i $(i = 0, 1, \ldots, k)$.

obeikandi.com

ومجموع المربعات المتعلقة بالحدود المختلفة لتابع التراجع . ونحسب المقادير A_i كما يلي :

$$A_0 = \sum_{j=1}^{n} y_j \,/ n = \bar{y} \qquad (44)$$

و

$$A_i = \sum_{j=1}^{n} y_j (\xi_i')_j \Big/ \sum_{j=1}^{n} (\xi_i')_j^2 \;,\; j = 1, \ldots, n \;;\; i = 1, \ldots k \qquad (45)$$

بينما نحسب مجاميع المربعات كما يلي :

(46) مجموع المربعات الكلي = $\sum_{j=1}^{n} y_j^2 - A_0 \sum_{j=1}^{n} y_j$

(47) التخفيض في مجموع المربعات العائد لملاءمة الحد من الدرجة i = $A_i \sum_{j=1}^{n} y_j (\xi_i')_j$ ، $i = 1, 2, \ldots, k$

لنعد إلى البيان الإحصائي في الجدول (٩ـ٥) باعتباره مناسباً جداً لطريقة كثيرات الحدود المتعامدة . فعندما نلقي نظرة أولية على التمثيل البياني للنقاط الثلاثة عشر التي يحويها البيان الاحصائي في الجدول (٩ـ٥) نجد أن ملاءمة تابع من الدرجة الثالثة سيكون مناسباً . وللقيام بذلك سيجد الباحث أنه من المفيد أن يرتب المعلومات وفقاً لما هو مبين في الجدول (٩ـ١١) وهكذا نجد :

$$A_0 = \frac{732}{13} = 56.3$$

$$A_1 = \frac{1361}{182} = 7.4780$$

$$A_2 = \frac{-73}{2002} = -.0365 \qquad (48)$$

$$A_3 = \frac{-389}{572} = -.6801 \,,$$

وهذا يعطينا معادلة التراجع .

$$\hat{y} = 56.3 + 7.4780\,\xi_1' - .0365\,\xi_2' - .6801\,\xi_3' \qquad (49)$$

obeikandi.com

جدول ٩ـ١١ لحساب المقادير A_i ومجاميع المربعات الموافقة

y	ξ'_1	ξ'_2	ξ'_3	$y\xi'_1$	$y\xi'_2$	$y\xi'_3$
17	–6	22	–11	–102	374	–187
21	–5	11	0	–105	231	0
22	–4	2	6	–88	44	132
27	–3	–5	8	–81	–135	216
36	–2	–10	7	–72	–360	252
49	–1	–13	4	–49	–637	196
56	0	–14	0	0	–784	0
64	1	–13	–4	64	–832	–256
80	2	–10	–7	160	–800	–560
86	3	–5	–8	258	–430	–688
88	4	2	–6	352	176	–528
92	5	11	0	460	1012	0
94	6	22	11	564	2068	1034
732	0	0	0	1361	–73	–389
λ	1	1	$\frac{1}{6}$			
$\sum_{j=1}^{n}(\xi'_i)_j^2$	182	2002	572			

opeikandi.com

وفي العديد من الحالات نرغب في الحصول على تابع تراجع بدلالة x بدلاً من ξ' . ويمكن القيام بذلك بالتعويض المباشر :

$$\xi'_1 = \lambda_1(x-\bar{x}) = 1(x-7) = x-7$$

$$\xi'_2 = \lambda_2\left[(x-\bar{x})^2 - \frac{n^2-1}{12}\right] = (x-7)^2-14 \qquad (50)$$

$$\xi'_3 = \lambda_3\left[(x-\bar{x})^3 - (x-\bar{x})\frac{3n^2-7}{20}\right]$$

$$= \frac{1}{6}\left[(x-7)^3 - 25(x-7)\right]$$

وبالتعويض في (49) نجد :

$$\hat{y} = 21.7277 - 5.8458x + 2.3449x^2 - .1134x^3 \qquad (51)$$

وتتفق هذه المعادلة مع المعادلة (35) . ولو أخذنا بعين الاعتبار ξ'_1 فقط أو ξ'_1 و ξ'_2 وقمنا بالتعويض المناسب من المعادلات (50) في المعادلة (49) بعد حذف ξ'_2 ، ξ'_3 في الحالة الأولى ، أي الاكتفاء بِ ξ'_1 في الطرف الأيمن ، وبعد حذف ξ'_3 أي الاكتفاء بِ ξ'_1 و ξ'_2 في الحالة الثانية ، فسنجد ، على الترتيب ، معادلة التراجع الخطية :

$$\hat{y} = 3.954 + 7.4780x, \qquad (52)$$

ومعادلة التراجع من الدرجة الثانية :

$$\hat{y} = 2.6765 + 7.9890x - .0365x^2 \qquad (53)$$

وهما تتفقان مع المعادلتين (38) و (37) على الترتيب ، في حدود الأخطاء الناتجة عن تدوير الأرقام العشرية في الحسابات .

ويمكن الآن حساب التخفيض في مجموع المربعات العائد لملاءمة الحدود المختلفة في تابع التراجع وذلك باستخدام المعادلة (47) .

٩-١١ التراجع الأسّي : في العديد من الحالات لا تكون العلاقة المفترضة بين المتحوّلين على شكل كثيرة حدود . وإحدى الأشكال الأكثر ظهوراً ينتمي إلى أسرة من التوابع تسمى التوابع الأسية . وبعض هذه التوابع التي تظهر

obeikandi.com

بتواتر كبير في التطبيقات العملية هي من الأنواع التالية :

$$y = \alpha \beta^{x} \quad , \quad \alpha > 0, \quad \beta > 0 \qquad (54)$$

$$\ln y = \ln \gamma + (\ln \alpha)\beta^{x} ; \quad \alpha > 0, \ \beta > 0, \ \gamma > 0 \qquad (55)$$

$$\frac{1}{y} = \gamma + \alpha \beta^{x} : \quad \alpha > 0, \ \beta > 0, \ \gamma > 0 \qquad (56)$$

$$y = \gamma \left[1 - e^{\alpha - \beta x} \right] ; \quad \beta > 0, \ \gamma > 0 \qquad (57)$$

وتدعى هذه التوابع عادة ، وعلى الترتيب (i) التابع الأسي البسيط أو منحني الفائدة المركّبة (ii) منحني جومبرتز ، (iii) المنحني اللوجيستيكي ، و (iv) منحني ميتشيرليش . وللملاءمة أي من هذه التوابع لبيان إحصائي معين نستخدم بصورة عامة مبدأ المربعات الدنيا . ولكن المعادلات الناظمية الناتجة عن تطبيق هذه الطريقة ستكون ، على الأغلب ، معقدة جداً وصعبة الحل ، وتوجد طرق تقريبية لحل مثل هذه المعادلات . وكمثال على إحدى هذه الطرق التقريبية المستخدمة لنأخذ المعادلة (54) . فبدلاً من اعتبار المعادلات الناظمية كما نحصل عليها من تطبيق طريقة المربعات الدنيا مباشرة ، فسيكون من المريح ، في الغالب ، أن يقوم الباحث بتحويل لوغاريتمي فيأخذ لوغاريتم الطرفين بالنسبة للأساس عشرة . وهكذا يأخذ النموذج في (54) الشكل التالي :

$$\log y = \log \alpha + (\log \beta)\, x \qquad (58)$$

وبإجراء التغيير التالي في المتحولات :

$$\log y = z, \quad \log \alpha = a, \quad \log \beta = b, \quad x = w$$

تصبح المعادلة (58) على الشكل :

$$z = a + bw \qquad (59)$$

ومن الطبيعي عندئذ أن نحصل على معادلة التراجع الموافقة لِ (59) بالطريقة التي استعرضناها في بداية هذا الفصل مستخدمين البيان الاحصائي الذي يضم

obeikandi.com

قيم المتحولين الجديدين w و z لنجد معادلة التراجع $\hat{z} = \hat{a} + \hat{b}w$. ونحصل على تقدير y بدءاً من قيمة $\hat{z}$ ، ولا بد من التأكيد على أن مثل هذا الحل سوف لا يعطي نتائج مطابقة لتلك التي كنا سنحصل عليها لو طبقنا الطريقة بصورة مباشرة أي اتبعنا المسلك الصعب ، ولكن يمكن القول ، على أي حال ، بأن الخطأ ليس كبيراً جداً مما يسمح بتطبيق الطريقة التقريبية غير المباشرة نظراً لسهولتها .

وسنوضح هذه الطريقة بالبيان الاحصائي المعطى في الجدول (٩ــ١٢) . ولو قمنا بتمثيل بياني لهذه المعلومات على ورقة نصف لوغاريتمية (أي نتبع سلماً حسابياً عادياً على محور الفواصل وسلّماً لوغاريتمياً على محور التراتيب) لرأينا أن الخط المستقيم يمكن أن يكون تابعاً مقبولاً لتمثيل العلاقة بين $z\ (= \log y)$ و$w\ (= x)$. وهكذا نحصل على $\hat{a}$ و $\hat{b}$ بحل المعادلتين الناظميتين :

$$22.389 = 20\,\hat{a} + 1340\hat{b}$$

$$1545.888 = 1340\hat{a} + 107566\hat{b}$$

جدول (٩-١٢) محتوى البروتين ونسبة النويات الزجاجية في عينات من القمح

رقم العينة	نسبة النويات الزجاجية $x\ (= w)$	محتوى البروتين y	$z = \log y$
1	6	10.3	1.013
2	75	12.2	1.086
3	87	14.5	1.161
4	55	11.1	1.045
5	34	10.9	1.037
6	98	18.1	1.258
7	91	14.0	1.146
8	45	10.8	1.033
9	51	11.4	1.057
10	17	11.0	1.041
11	36	10.2	1.009
12	97	17.0	1.230
13	74	13.8	1.140
14	24	10.1	1.004
15	85	14.4	1.158
16	96	15.8	1.199
17	92	15.6	1.193
18	94	15.0	1.176
19	84	13.3	1.124
20	99	19.0	1.279

فنجد :

$$\hat{b} = 0.002576$$

$$\hat{a} = 0.9469$$

وأخيراً :

$$\hat{z} = 0.9469 + 0.002576w$$

أو :

$$\widehat{\log y} = 0.9469 + 0.002576x$$

obeikandl.com

٩ـ١٢ مسألة الفرز : لنفرض أن اختصاصياً في التغذية يحلل عينات من الدم لتحديد مستوى الفوسفات y مقاساً بوحدات الفينول ، وقيمة الفيتامين c ولنرمز لها بـ x مقاسة بالملغم . ولنفرض أن علاقة تراجع خطي قد حُددت لتقدم أفضل تمثيل للعلاقة التابعية بين x و y . فغالباً ما يواجه الباحث مسائل من النوع التالي : لنفرض أن اختصاصياً في التغذية يحلل عينة من دم شخص ينتمي إلى نفس المجموعة التي تنطبق عليها العلاقة الخطية المتوفرة بين أيدينا . ولكن حدث أنه لم يستطع الحصول على قيمة الفيتامين c أو أن مثل هذه المعلومات مفقودة ، وكل ما لدينا هو قيمة مستوى الفوسفات y_0 والسؤال الذي يطرح نفسه هو هل يمكننا تقدير x_0 ؟ وهذه الحالة تشكل مثالاً بسيطاً من مجموعة أعم من المسائل تُعرف بمسائل التصنيف أو الفرز . ولنر كيف يمكن أن نضع جواباً لتساؤلنا ؟

بتطبيق مبدأ المربعات الدنيا على بيان إحصائي يحوي كلاً من قيم x و y نحصل على معادلة التنبؤ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ بالطريقة المعتادة . ومن هذه المعادلة يمكننا كتابة :

$$x = \frac{\hat{y} - \hat{\beta}_0}{\hat{\beta}_1} \qquad (60)$$

وكما نتوقع فإن أفضل تقدير نقطي لِـ x_0 معطى بالعلاقة :

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \qquad (61)$$

ويمكن الحصول على تقدير مجالي لِـ x_0 بحساب مجال الثقة :

$$\bar{x} + \frac{b(y_0 - \bar{y})}{D} \pm \frac{t_\alpha \Delta}{D} \sqrt{B(y_0 - \bar{y})^2 + D\left(1 + \frac{1}{n}\right)} \qquad (62)$$

حيث

$$B = \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (63)$$

$$D = b^2 - t_\alpha^2 \quad s^2 B = b^2 - t_\alpha^2 s_b^2 \tag{64}$$

وللتوزيع t هنا n − 2 درجة من الحرية .

وغالباً ما يحصل الباحث ، خاصة في التجارب البيولوجية ، على عدة قيم لِ y ، ولتكن m مثلاً ، مقابلة للقيمة x_o الغير معروفة . وفي مثل هذه الحالات لا بد من إدخال التعديلات التالية :

(65)

$$\hat{x}_o = \frac{\bar{y}_o - a}{b}$$

ويصبح مجال الثقة :

(66)

$$\bar{x} + \frac{b(\bar{y}_o - \bar{y})}{D} \pm \frac{t_\alpha S'}{D}\sqrt{B(\bar{y}_o - \bar{y})^2 + D\left(\frac{n+m}{nm}\right)},$$

حيث

(67)

$$\bar{y}_o = \frac{1}{m}\sum_{i=1}^{m} y_{oi} \quad ,$$

و

(68)

$$(S')^2 = \frac{(n-2)s^2 + \sum_{i=1}^{m}(y_{oi} - \bar{y}_o)^2}{n+m-3} \quad ,$$

وعدد درجات الحرية الموافقة لِ t هو هنا n + m − 3. وبما أن m عملياً صغيرة بالنسبة لِ n فإن الجهد الحسابي ينخفض بصورة ملحوظة باستخدام s^2 بدلاً من $(S')^2$. وهذا يقود بالطبع إلى حل تقريبي إلا أنه تقريب جيد من وجهة النظر العملية .

ولتوضيح هذه الطريقة نعود مرة أخرى إلى البيان الاحصائي في الجدول (٩ــ٥) . ولنفرض أن $y_o = 60$ ولكن ساعات الخلط x_o الموافقة لهذه القيمة

لِـ γ غير معروفة . ونجد بعد التعويض في (63) و(64) و(61) و(62) أن :

$$B = \frac{1}{182} = 0055$$

$$D = (7.478)^2 - (2.201)^2\ (28.83)\ (.005) = 55.152$$

$$\hat{x}_0 = 7.494$$

ومجال الثقة هو :

$$(5.85,\ 9.15)$$

وتجدر الاشارة إلى أن حساب حدي مجال الثقة ليس ممكناً إذا كانت الكمية تحت الجذر التربيعي في (62) أو (66) سالبة ويحصل ذلك ، بصورة عامة ، إذا كان s^2 كبيراً و b صغيراً جداً ، وفي مثل هذه الحالة تكون . جدوى معادلة التراجع من أساسها موضع شكّ .

٩ـ١٣ ملاحظات عامة تتعلق بالتراجع الخطي البسيط : لنلق نظرة على الدقة التي نستطيع بموجبها تقدير β_0 و β_1 . فمن العلاقتين (13) و (8) يمكن أن نكتب :

(69)

$$V(\hat{\beta}_0) = \sigma^2 \sum_{i=1}^{n} x_i^2 / n\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right),$$

$$V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$$

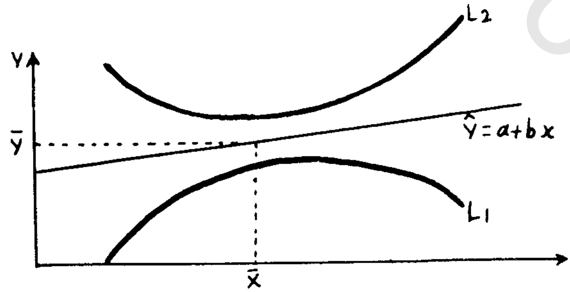
(70)

وإذا وضعنا بدلاً من σ^2 تقديره s^2 نحصل على تقدير لكل من $V(\hat{\beta}_0)$ و $V(\hat{\beta}_1)$. ونلاحظ أن تشتت $\hat{\beta}_0$ سيكون صغيراً بالقدر الممكن إذا اخترنا المقادير x_i $(i = 1, 2, \ldots, n)$ بحيث يكون $\bar{x} = 0$ ، وهذا يعني أن مجال الثقة الخاص بِـ β_0 سيكون أقصر ما يمكن ، بصورة وسطية ، عندما يكون $\bar{x} = o$. وإذا تأملنا العلاقة (70) نجد أنه يمكن جعل تشتت $\hat{\beta}_1$ صغيراً باختيار قيم لِـ x_i يكون

obeikandi.com

تشتتها كبيراً ، أي متباعدة عن بعضها البعض . وكمسألة تلفت النظر نجد أننا إذا اخترنا أصغر قيمة ممكنة لِ x (ممكنة ضمن حدود تغير x والقياسات المتوفرة عملياً) ، وأكبر قيمة ممكنة لِ x ، وبعبارة أخرى إذا كانت جميع قيم المتحول y موافقة إما لهذه القيمة أو تلك لِ x فإن تشتت b سيكون صغيراً بالقدر الممكن . وبما أن مثل هذا الاختيار يؤدي بنا إلى صعوبة هي أننا سوف لا نستطيع ، في مثل هذه الحالة ، القيام باختبار كون العلاقة بين x و y خطية أم لا ، هذا في حال وجود شكّ حول شكل هذه العلاقة ، ولذلك فإننا نختار عادة ثلاث قيم لِ x على الأقل ونضحي ببعض الدقة في تقدير β_1 .

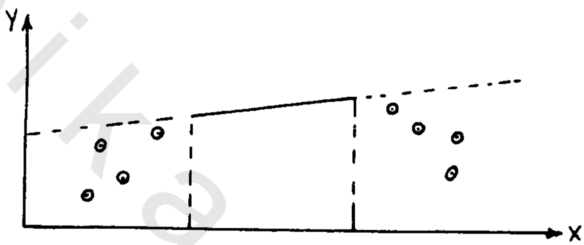
لنعتبر الآن الدقة التي يمكن أن تتصف بها تنبؤاتنا المبنية على معادلة تراجع خطي بسيط قمنا بتحديدهما وفق الطرق الموصوفة سابقاً . فنرى أن مجال الثقة أو مجال التنبؤ الذي نحصل عليه . سيختلف حجمه ، حتى من أجل نفس العينة ، وذلك وفقاً لقيمة x_p (أنظر العلاقتين (20) و (22)) . وسيكون طول المجال أكبر كلما كانت x_p بعيدة عن $\bar{x}$ ، كما سيكون المجال أصغر عندما يكون $x_p = \bar{x}$. وهذا يعني أن أكبر دقة ممكنة لتقدير $E(y/x) = \beta_0 + \beta_1 x$ ، مبنياً على العينة أو البيان الاحصائي الموجود بين أيدينا ، سنحصل عليها من أجل $x = \bar{x}$ ، وبالطبع تبقى نفس الفكرة صحيحة فيما يتعلق بالتنبؤ بقيمة y الموافقة لقيمة معينة لِ x . ويمكن عرض الفكرة بيانياً كما في الشكل (٩ـ٦) .



شكل ٩ـ٦ توضيح مجالات الثقة من أجل $\beta_0 + \beta_1 x$.

وما يجب ألا يغيب عنا ، هو أن التنبؤ بقيمة y من أجل قيمة معينة لِ x ،

سيكون عرضة لمخاطر جديدة إذا جرى هذا التنبؤ من أجل قيمة لِ x خارج المجال الذي اخترناه لقيم x المستخدمة فعلاً عند الحصول على عينة . أي أن التعميم إلى ما وراء المجال الملحوظ للمتحول المستقل لا يخلو من المغامرة ما لم يكن لدينا نوع من الثقة بأن تابع التراجع الخطي يبقى موجوداً أو قائماً فوق مجال لقيم x أعرض من المجال المستخدم عند الحصول على العينة أو البيان الاحصائي . ويكفي توضيح بسيط للإشارة إلى المشاكل الممكنة . فلنفرض أن لدينا قيماً لِ x و y تمثيلها البياني هو مجموعة النقاط المبينة في الشكل (٩ـ٧) .



شكل ٩ـ٧ توضيح مخاطر التعميم إلىما وراء مجال قيم x المستخدمة في العينة

ويبدو الخط المستقيم ملاءمة جيدة للبيان الاحصائي ضمن المجال المعطى لـ x ، مما يمكن أن يغرينا بتمديد خط التراجع إلى أبعد من هذا المجال في كل من الاتجاهين ، وكان سيكون من المحتمل جداً ، على أي حال ، لو أننا اخترنا منذ البداية مجالاً أعرض لقيم x (أنظر النقاط المحاطة بدائرة) . ولاحظنا قيم y الموافقة ، أن تشير النتائج إلى ضرورة ملاءمة منحن من الدرجة الثانية باعتباره الشكل الأنسب لتابع التراجع وليس الخط المستقيم الذي قمنا برسمه . ونرى بوضوح من الشكل (٩ـ٧) أن التنبؤ بقيم y مستخدمين القسم المنقط من خط التراجع يمكن أن يقودنا إلى أخطاء كبيرة . ولذلك فإنه من المستحسن أن يكون الباحث شديد الحذر عند قيامه بتنبؤات خارج المجال الملحوظ للمتحول المستقل .

تمارين

١ ـ أحسب (أ) : $\hat{y} = \hat{\alpha} + \hat{\beta} x$ ، (ب) SSE ، (جـ) $s_{\hat{\beta}}$

إذا علمت أن :

$$\sum_{i=1}^{n} x_i^2 = 121 , \quad \sum_{i=1}^{n} x_i = 20 , \quad \sum_{i=1}^{n} x_i y_i = 82$$

$$\sum_{i=1}^{n} y_i^2 = 516 , \quad \sum_{i=1}^{n} y_i = 40 , \quad n = 10 .$$

٢ ـ إذا علمت أن :

$$n = 277 , \quad \bar{x} = 65 , \quad \bar{y} = 72 , \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 1600,$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = 3600 , \quad \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 2000$$

فاحسب قيم (أ) : $\sqrt{SSE} = \hat{\sigma}$ ، (ب) $S_{\hat{\beta}}$ ، (جـ) $S_{\hat{y}}$ من أجل $x = 45$

٣ ـ إذا علمت أن :

$$n = 62 , \quad \bar{x} = 10 , \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 40 , \quad \sum_{i=1}^{n} (y_i - \bar{y})^2 = 250 ,$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = - 80 , \quad \bar{y} = 20$$

فالمطلوب :

أ ـ أحسب التراجع الخطي لِـ y على x .

ب ـ أحسب 99% مجال ثقة لأمثال التراجع . واعرض جميع الشروط التي تجدها ضرورية للقيام بذلك .

obeikandi.com

جـ ـ قدّر الربح في المعلومات بالنسبة لتقدير متوسط المجتمع المتعلق بالمتحول y وذلك كنتيجة لاستخدام المتحول x .

٤ ـ يمثل الجدول التالي الطول (x) والوزن (y) لعدة رجال . وقد اخترنا الأطوال سلفاً ثم لاحظنا أوزان مجموعة عشوائية من الرجال الذين يتصفون بهذه الأطوال :

y (بالرطل الإنجليزي)	x (بالبوصة)
110	60
135	61
120	60
126	61
140	62
130	60
135	62
158	65
145	64
170	70
185	72
180	70

أ ـ مثّل هذا الجدول من القيم بيانياً .

ب ـ أحسب خط التراجع الخطي البسيط لِ y على x : $\hat{y}=\hat{\alpha}+\hat{\beta}x$

جـ ـ أحسب وفسّر 90% مجال ثقة من أجل β.

د ـ أحسب وفسّر 98% مجال ثقة من أجل α .

هـ ـ أحسب وفسر 95% مجال ثقة من أجل $E(y/x = 65)$.

و ـ اختبر الفرضية $H_0 : \beta = 0$.

ز ـ اختبر الفرضية $H_1 : \beta = 6$.

ح ـ اختبر الفرضية $H_2 : \alpha = -30$

ط ـ تنبأ بوزن شخص طوله 66 بوصة . أعط تقديراً مجالياً .

ي ـ قدر طول رجل وزنه المسجل 170 رطلاً إنكليزياً . أعط كلاً

من التقديرين النقطي والمجالي .

ك ـ اختبر خطية التراجع .

ل ـ ما هي النسبة من تغير y التي يفسرها تراجع الوزن على الطول ؟

٥ ـ مستخدماً البيان الاحصائي التالي ، لائم منحن من الدرجة الثانية (قطع مكافىء) للربح الاجمالي لمزرعة (y) في مقابل عدد أشهر العمل (x) :

رقم المزرعة	الربح الإجمالي y	أشهر العمل x	رقم المزرعة	الربح الإجمالي y	أشهر العمل x
1	16.7	20	18	11.2	14
2	17.9	19	19	9.4	15
3	17.4	24	20	8.7	12
4	14.9	15	21	12.2	17
5	16.2	24	22	7.7	14
6	14.0	15	23	11.5	14
7	15.1	24	24	7.3	13
8	18.3	24	25	11.8	16
9	11.3	16	26	15.1	23
10	18.3	26	27	10.5	33
11	17.1	24	28	17.0	29
12	12.0	16	29	15.6	30
13	15.2	25	30	13.2	31
14	16.2	28	31	17.2	22
15	14.9	24	32	14.6	32
16	10.5	15	33	12.2	34
17	16.5	27	34	9.8	36

obeikandi.com

٦ ـ ترمز x في الجدول إلى وزن الجسم بالكيلوغرام ، و y لحجم الدم بالسنتمترات المكعبة ، وذلك من أجل الماعز :

x	y	x	y	x	y	x	y	x	y
34	2370	21	1480	38	2980	18	1070	35	2410
28	2100	39	2450	30	2020	40	2300	38	2900
19	1120	37	2560	26	1710	66	4230	21	1580
41	2810	23	1550	19	1240	34	2440	52	3600
21	1500	17	1100	60	3990	16	1050	28	1850
20	1660	48	3550	45	2940	30	2360	45	3010

أ ـ ارسم التمثيل البياني لهذه المعلومات الاحصائية .

ب ـ أحسب $\bar{x}$ ، $\bar{y}$ ، s_x ، s_y ، r ، $\hat{\sigma}$ وميل خط التراجع .

جـ ـ ارسم خط التراجع الناتج .

د ـ قدّر متوسط y من أجل $x = 50$ ، ضع 95% مجال ثقة حول هذا التقدير .

٧ ـ أخذنا قياسات تتعلق بمقدرة فئران على اجتياز متاهة قبل تناول منشط (x) وبعد تناول منشط (y) . ووجدنا في عينة من ثلاثمائة فأر أن $\bar{x} = 16.0$ ، $\bar{y} = 12.8$ ، $s_x^2 = 4.0$ ، $S_y^2 = 3.7$ ، $r = .4$

أ ـ ارسم خط تراجع y على x .

ب ـ قدّر متوسط y من أجل $x = 17.0$.

جـ ـ ضع 95% مجال ثقة حول هذا التقدير .

د ـ إذا كان $x = 17$ من أجل فأر ، فالمطلوب وضع 90% مجال ثقة من أجل قيمة y الموافقة لهذا الفأر .

obeikandl.com