الفصل الحادي عشر

الارتباط، الانحدار، والسلاسل الزمنية

Correlation, Regression, and Time Series

حتى الآن قمنا باستخدام المعلومات المتاحة في البيانات التي تم جمعها للمتغير العشوائي X لعمل استدلال حول القيم التي نرغب في معرفتها للمتغير. وبذلك تم اختبار المتوسط لـ X وذلك باستخدام المتوسط كمؤشر لمعظم القيم في مجموعة البيانات. الآن يتركز الاهتمام على المعلومات المتاحة في متغير آخر Y ونختبر لنرى ما إذا كان هناك علاقة بين المتغيرين X و Y والتي يمكن أن تعطي معلومات أكثر حول المتغير X بدلاً من المتوسط فقط. وفي حالة وجود زوجين من المتغيرات فإن لدينا نوعين من التحليل ممكنة. أحد تلك الأنواع، عشوائية كلا المتغيرين و يتركز الاهتمام في معرفة درجة العلاقه بينهما. فإذا كانت درجة العلاقة بينهما قويه بحيث إن زيادة قيمه أحدهما تؤدي لتوقع زيادة الآخر ولا يمكننا ملاحظة إلا القيم الخاصة بأحد تلك المتغيرين فإنه يمكن استخدامها للمساعدة في تقدير الكميات الخاصة للمتغير الذي لا نستطيع ملاحظته. أما النوع الآخر من التحليل فإن أحد تلك المتغيرات عشوائي بينما المتغير الآخر ثابت وعليه فإنه يمكننا معرفة القيم المفترضة بدرجه معينه من اليقين. وإذا كان بالإمكان وصف العلاقة بينهما رياضياً فإن لدينا طريقه تمكننا من تقدير القيم

المتوقعة للمتغير العشوائي باستخدام قيم معينة للمتغير الثابت. وتجدر الإشارة إلى أن النوع الأول من التحليل هو الارتباط بينما النوع الثاني من التحليل هو الانحدار وسيتم التطرق لهما في الأجزاء التالية.

تحليل الارتباط Correlation Analysis

ارتباط متغيرين عشوائيين X ، Y يعتمد على ما إذا كانت تتغير في نفس الاتجاه أو في الاتجاه المعاكس أو في اتجاه غير محدد. والحالة الأخيرة تعبّر عن استقلالية المتغيرات وليس ارتباطها . فعلى سبيل المثال أسعار القمح تكون مستقلة عن أسعار جهاز الحاسب المكتبي ؛ نظراً لأنها تتحدّد وفقاً لقوى سوقية مختلفة تماماً عن بعضها. وإذا حصل وأن تغيرت الأسعار معاً خلال فتره قصيرة فيمكن اعتبار أنه حدث عشوائي قصير الأمد. ولكن أسعار القمح وأسعار الشعير غالباً تميل للتغير في نفس الاتجاه ؛ بسبب أنها تتأثر بنفس القوى السوقية وبينها ارتباط موجب. وبالمثل ، فإن سعر وكمية القمح المستهلكة متغيرات عشوائية تتغير في الاتجاه المعاكس وبينها ارتباط سالب. ويمكن التفكير في العديد من الأمثلة المتعلقة بالمتغيرات العشوائية المرتبطة وغير المرتبطة.

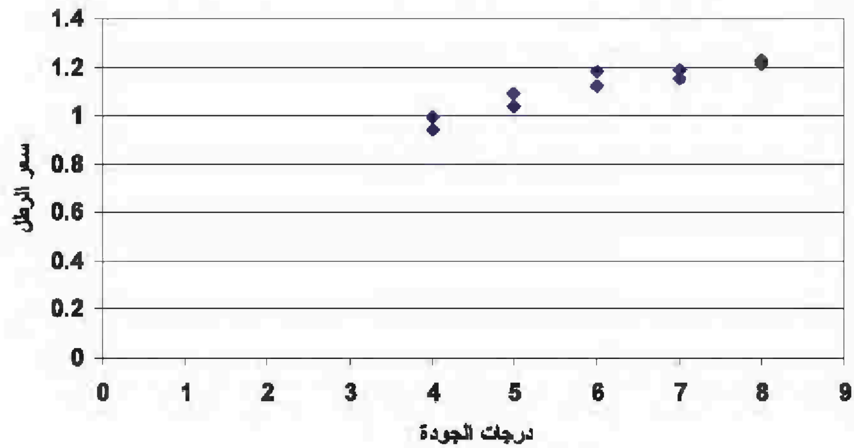
يقيس معامل الارتباط r قوة درجه الارتباط بين متغيرين عشوائيين. وتتراوح قيمته بين ±١ ، أي $-1 < r < +1$. وتكون قيمه هذا المعامل للمتغيرات العشوائية المستقلة قريبة من الصفر بينما تكون قيمته موجبة للمتغيرات المرتبطة في نفس الاتجاه وسالبة للمتغيرات التي بينها علاقة عكسية. وكلما كانت القيمة المطلقة لهذا المعامل r كبيره كانت درجة الارتباط بين المتغيرين قويه. ويمكن حساب معامل الارتباط باستخدام أي من الصيغ الرياضية التالية ، المعادلتين رقمي (11.1 و 11.2) :

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} \quad (11.1)$$

$$r = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{(\sum X^2 - n\overline{X}^2)(\sum Y^2 - n\overline{Y}^2)}} \quad (11.2)$$

ونظراً لأننا نأخذ القيمة الموجبة للجذر التربيعي للمقام دائما لكلا المعادلتين ، فإن r تكون سالبة في حالة كان حاصل طرح البسط سالب.

ويمكن استخدام شكل الانتشار لعرض الارتباط بين المتغيرين العشوائيين X ، Y بيانياً . في شكل الانتشار نرسم كل زوج (X_i, Y_i) بنقطة لكل البيانات ثم نختبر الشكل المتحصل عليه (الشكل رقم ١١.١) . وتقع كل النقاط على خط مستقيم إذا كانت $r = 1$ ويكون الخط موجب الميل إذا كانت r موجه وسالب الميل إذا كانت r



الشكل رقم (١١.١). شكل الانتشار لبيانات شرائح ثور لانجوس.

سالبة. وعندما تكون r أقل من ١ فإن النقاط تقع في نطاق ضيق أو على شكل قطع ناقص والتي تصبح أقرب للخط كلما قربت قيم r من الواحد. وعندما تكون قيمه r مساوية للصفر فإن النقاط تكون على شكل دائرة أو كتلة بدون اتجاه محدد. القيم الصغيرة لـ r تعطي شكل انتشار بنقط أقرب للدائرة منها للخط المستقيم.

ويمكن اختبار معامل الارتباط باستخدام توزيع t . حيث يمكن صياغة فرض العدم والفرض البديل كالتالي :

$$H_0 : \rho = 0$$
$$H_a : \rho < 0 \, (\text{or } \rho > 0)$$

حيث ρ (حرف إغريقي ينطق رو) يعبّر عن معامل الارتباط للمجتمع. درجه الحرية للاختبار هي $(n - 2)$ ؛ نظراً لأننا نخسر درجات حرية لتقدير متوسطات المتغيرين X و Y ، و n تعبر عن عدد أزواج البيانات في المجموعة وتحسب قيمة t باستخدام المعادلة رقم (11.3)، وحيث إن $\rho = 0$ في فرض العدم، فإننا عادة لا نضعها في البسط عند حساب قيمة t .

$$t = \frac{r - \rho}{S_r} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad (11.3)$$

ويمكن إيضاح ذلك من خلال المثال التالي :

تم تحديد درجات الجودة لنوعيه لحوم البقر المذبوحة بواسطة قسم الزراعة الامريكيه كالتالي :

جيد منخفض (slightly abundant marbling ؛ الدرجة=٨) ، اختيار مرتفع

(moderate marbling ؛ الدرجة =٧)، اختيار متوسط(modest marbling ؛ الدرجة = ٦)، اختيار منخفض (small amount of marbling ؛ الدرجة = ٥)، اختيار (slight marbling ؛ الدرجة = ٤)، مستوى عالي (trace of marbling ؛ الدرجة = ٣)، ومستوى متوسط (practically devoid of marbling ؛ الدرجة = ٢).ويرغب المختص في الإنتاج الحيواني في تحديد درجة العلاقة بين درجه نوعية اللحوم المحددّة من قبل إدارة الزراعة الأمريكية *(X)* وسعر البيع للرطل من تلك القطع *(Y)* وذلك لعينة عشوائية حجمها ١٠ قطع من لحم العجل والبيانات موضحة بالجدول رقم (١١.١). ويمكن

الجدول رقم (١١.١). درجات جودة أنواع شرائح لحوم البقر المذبوحة وأسعارها لشرائح ثور لانجوس.

درجة جودة النوع، X	سعر البيع للرطل بالدولار، Y	XY	X^2	Y^2
٥	١,٠٩	٥,٤٥	٢٥	١,١٨٨
٧	١,١٩	٨,٣٣	٤٩	١,٤١٦
٤	٠,٩٩	٣,٩٦	١٦	٠,٩٨٠
٥	١,٠٤	٥,٢٠	٢٥	١,٠٨٢
٨	١,٢١	٩,٦٨	٦٤	١,٤٦٤
٦	١,١٨	٧,٠٨	٣٦	١,٣٩٢
٤	٠,٩٤	٣,٧٦	١٦	٠,٨٨٤
٧	١,١٥	٨,٠٥	٤٩	١,٣٢٢
٨	١,٢٣	٩,٨٤	٦٤	١,٥١٣
٦	١,١٢	٦,٧٢	٣٦	١,٢٥٤
٦٠	١١,١٤	٦٨,٠٧	٣٨٠	١٢,٤٩٥

حساب قيمة r بالتعويض عن القيم الموجودة في الجدول في واحدة من المعادلات الحسابية وحلها لنجد أن قيمة $r = 0.94$.

$$r=\frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} = \frac{68.07-\frac{(60)(11.14)}{10}}{\sqrt{\left[380-\frac{(60)^2}{10}\right]\left[12.495-\frac{(11.14)^2}{10}\right]}} = 0.94$$

والتي تشير إلى أن درجه الارتباط بين درجه نوعية اللحم وسعر الرطل منها مرتفعة جداً. ويمكن اختبار هذه القيمة لمعرفة ما إذا كانت فعلاً أكبر من الصفر باستخدام اختبار t والفرض المناسب هو:

$$H_0 : \rho = 0$$

$$H_a : \rho > 0$$

ونستخدم اختبار من طرف واحد باتجاه اليمين لهذا المثال؛ نظراً لأننا نتوقع وجود ارتباط موجب بين المتغيرين العشوائيين، درجه جودة اللحوم وسعرها. درجة الحرية لهذا المثال هي:

$$v = (n - 2) = (10 - 2) = 8$$

وباختيار ٥٪ مستوى معنوية فإن قيمة t الحرجة (الجدوليه) من جدول رقم (٨) بالملحق هي ١,٨٦ . فإذا كانت قيمة t المحسوبة أكبر من ١,٨٦ فإننا نرفض فرض العدم ، خلاف ذلك لا نرفضه. وبحساب قيمه t باستخدام الصيغة الرياضية نجد أنها تساوي ٧,٧٩ كالتالي:

$$t=\frac{r-\rho}{\sqrt{\frac{1-r^2}{n-2}}}=\frac{0.94-0}{\sqrt{\frac{1-0.94^2}{10-2}}}=7.79$$

وحيث إن قيمة t المحسوبة $t=7.79$ أكبر من قيمة t الجدوليه $t=1.86$ فإننا نرفض الفرض القائل بعدم وجود ارتباط بين المتغيرين - درجه جودة اللحوم وسعرها للرطل. ولذلك فإن المتغيرين تتحرك في نفس الاتجاه ؛ كما في الشكل رقم (١١،١).

الانحدار الخطي البسيط Simple Linear Regression

صُمِّم تحليل الانحدار للإشارة إلى العلاقة الرياضية بين المتغير الثابت X والمتغير العشوائي Y في معادلة يكون فيها Y متغير تابع ، لذا فإن المعادلة التي لدينا تكون على الشكل $Y=a+bX$. وهذه المعادلة خطية في المعالم a ، b حيث a تشير إلى القاطع بينما b عبارة عن الميل إذا كانت المعادلة توضح خط مستقيم. ولمعرفة شكل الدالة يجب جمع بيانات على شكل أزواج مرتبة (X_i, Y_i) واستخدامها للحصول على تقدير لمعالم الدالة a ، b. فمثلاً الخط $Y=42-2X$ يختلف عن الخط $Y=-100+0.7X$ واختلافها واضح كلياً من قيم a ، b. رياضياً ، ربما نستخدم العديد من الطرق للحصول على قيم a ، b باستخدام البيانات الخاصة X ، Y . ولكن طريقة المربعات الصغرى والتي تركز على المتغير العشوائي Y تستخدم لتحليل الانحدار. وبصفة خاصة فإن طريقه المربعات الصغرى تعمل على تدنية مجموع مربعات الانحرافات لكل نقطه من البيانات (X_i, Y_i) من الخط الذي يتم تقديره في اتجاه Y . وحيث إنه من غير المتوقع أن تقع كل نقاط البيانات الأصلية (X_i, Y_i) على الخط المستقيم ، لكل قيم X_i في البيانات، لذا فإنه قد يكون لدينا قيمتين لـ Y. القيمة الأصلية من البيانات Y_i والقيمة المتحصل عليها من المعادلة Y_e

والتي تعتمد على قيم a ، b . والفرق بينهما $(Y_i - Y_e)$ هو الاختلاف محل النقاش. وحيث إن المجموع $\sum(Y_i - Y_e) = 0$ فإنه بذلك غير مفيد لنا، ولكن المجموع $\sum(Y_i - Y_e)^2$ لا يساوي الصفر ويمكن اختياره كمعيار لرسم الخط لمجموعة البيانات. وعليه فإن طريقه المربعات الصغرى تعمل على تدنية $\sum(Y_i - Y_e)^2$. رياضياً ، نحن نعمل على تصغير هذا المجموع بأخذ التفاضل لها بالنسبة لـ a ، b ثم مساواتها بالصفر وحل المعادلة . ولذلك نحصل على ما يسمى بالمعادلات الطبيعية المعادلة رقم (11.4) .

$$\begin{aligned} \sum Y_i &= na + b\sum X_i \\ \sum X_i Y_i &= a\sum X_i + b\sum X_i^2 \end{aligned} \qquad (11.4)$$

ويمكن حل المعادلة الأولى بالنسبة لـ a للحصول على الصيغة التعريفية والصيغة الحسابية (معادلتي رقم 11.5 و 11.6).

$$a = \overline{Y} - b\overline{X} \qquad (11.5)$$

$$a = 1/n(\sum Y_i - b\sum X_i) \qquad (11.6)$$

وبالتعويض عن الطرف الأيمن من المعادلة الأخيرة في المعادلة الطبيعة الثانية والمعالجة الرياضية لها نحصل على تعبيرين رياضيين لـ b (المعادلة رقم 11.7) :

$$b = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^{\,2} - \frac{(\sum X_i)^2}{n}} \qquad (11.7)$$

وتجدر الإشارة إلى أن المعادلة المشتملة على متوسط X ومتوسط Y أسهل في الاستخدام بالآلة الحاسبة اليدوية إذا كانت كلا المتوسطين أرقام صحيحة ، وما

عدا ذلك فإن التعبير الثاني لـ b يكون أسهل استخدام. وتحليل الانحدار يمكن عمله بكفاءة باستخدام البرامج الحاسوبية مثل برنامج اكسل وبرنامج SPSS بالحاسب الشخصي.

المعادلات السابقة لـ a ، b هي نتائج مباشرة لطريقة تقدير المربعات الصغرى، والتي تعطي أفضل [1] مقدرات؛ نظراً لأنها غير متحيزة وأكثر دقة مقارنة بأي مجموعة أخرى. لذا، فإننا نفضل استخدام هذه الصيغ عند إجراء تحليل الانحدار للحصول على أفضل معادلات ممكنة لشرح العلاقة الخطية بين المتغيرين X ، Y.

وبمجرد استخدام صيغ المربعات الصغرى لـ a ، b ومن ثمَّ الحصول على معادلة العلاقة الخطية بين X ، Y يظهر التساؤل كيف نعرف أن هذه المعادلة تعطي معلومات أكثر حول Y بدلاً من متوسط Y نفسها مع تجاهل X؟ ويمكن الإجابة على ذلك باختبار الفروض حول مقدّرة الميل b في معادلة المربعات الصغرى الخطية. فإذا كان الميل ، b فعلياً يساوي صفر فإنه لا يوجد علاقة خطية بين X ، Y وإنما فقط يوجد خط أفقي ارتفاعه يساوي متوسط Y نظرياً لأن الحد الثاني في صيغة المربعات الصغرى لـ a سيساوي الصفر عندما $b = 0$. وهذا مكافئ للقول بأن متوسط Y هو أفضل تقدير لـ Y . ومن جهة أخرى، إذا لم تكن b تساوي صفر فإن المتغير الثابت X سوف يضيف المزيد حول شرح التغير في Y بدلاً من متوسط Y نفسها، وعليه فإن الاختبار سيكون لـ b. في الانحدار الخطي البسيط حيث يوجد متغير X واحد و b واحده فإنه يمكن استخدام واحد من الاختبارين التاليين إما اختبار t لـ b أو تحليل التباين للمعادلة الخطية. وسيتم التطرق لذلك في الأجزاء التالية.

اختبار t حول b A t Test on

تتم صياغة فرض العدم والفرض البديل لهذا الاختبار من حيث معالم المجتمع لميل الخط β. والفرض البديل يمكن أن يكون اتجاهين إذا كان لا يوجد لدينا معلومات خارجية لتحديد اتجاه العلاقة بين المتغيرين X ، Y ، أو يكون اتجاه واحد إذا كان لدينا بيانات تحدد الاتجاه (ايجابي أو سلبي) للعلاقة . ويمكن كتابة الفرض باتجاهين على النحو التالي :

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

ويتم تحديد درجه الحرية لهذا الاختبار باستخدام المعادلة رقم (11.8) حيث الرمز k يساوي عدد المتغيرات المستقلة X في المعادلة رقم (11.9) :

$$v = (n - k - 1) \qquad (11.8)$$

ونلاحظ أننا نخسر درجه حرية لتقدير a ودرجات أخرى لكل b يتم تقديره. في حالة الانحدار الخطي البسيط فإن درجة الحرية هي $(n - 2)$ لأن لدينا مقدرين هما a ، b . وبمعرفة مستوى المعنوية يمكننا إيجاد القيمة الحرجة لـ t من الجدول رقم (٨) بالملحق ومقارنتها بقيمة t المحسوبة من المعادلة رقم (11.9) :

$$t = \frac{b - \beta}{S_\beta} = \frac{b - \beta}{\sqrt{\dfrac{\dfrac{\sum (Y_i - Y_e)^2}{n - 2}}{\sum (X_i - \overline{X})^2}}} \qquad (11.9)$$

فإذا كانت قيمة t المحسوبة أكبر من القيمة الحرجة (الجدوليه) فإننا نرفض فرض العدم ونقول بأن العلاقة الخطية تم وصفها بخط الانحدار لـ Y على X. وإذا

كانت القيمة المحسوبة اقل من الحرجة (الجدوليه) فإننا لا نرفض فرض العدم وعليه لا يوجد علاقة خطيه بين المتغيرين.

تحليل التباين لخط الانحدار

Analysis of Variance for the Regression Line

تتم تجزئة مجموع المربعات ودرجات الحرية المرتبطة بالمتغير العشوائي Y وذلك لإجراء تحليل التباين لخط الانحدار . وكما ناقشنا سابقاً في تحليل التباين باتجاه واحد في الفصل التاسع ، فإننا نحصل على مجموع المربعات الكلي كمجموع لمربع الاختلافات لـ Y عن متوسطها كما هو موضح في المعادلة رقم (11.10) :

$$Total\ SS = \sum (Y_i - \overline{Y})^2 \qquad (11.10)$$

وتتم تجزئته إلى مجموعي مربعات- هما مجموع مربعات الانحدار SSR ومجموع مربعات الخطأ SSE. ويمكن الحصول على مجموع مربعات الانحدار بجمع مربعات الفروق بين قيمة Y المقدرة من المعادلة، Y_e ، ومتوسط Y كما في المعادلة رقم (11.11) :

$$SSR = \sum (Y_e - \overline{Y})^2 \qquad (11.11)$$

وحساب مجموع مربعات الخطأ كمربع للانحرافات حول خط الانحدار (المعادلة رقم 11.12):

$$SSE = \sum (Y_i - Y_e)^2 \qquad (11.12)$$

ودرجات الحرية لمجموع المربعات الكلي تساوي (n-1) بينما درجه الحرية لمجموع مربعات الانحدار SSR تساوي k درجه حرية أما مجموع مربعات الخطأ فإن درجه الحرية لها هي (n-k-1). وبحساب مجموع المربعات يمكننا إنشاء جدول تحليل التباين كما في الجدول رقم (١١,٢) التالي:

الجدول رقم (١١.٢). جدول ANOVA للانحدار الخطي البسيط.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
الانحدار	*1*	*SSR*	*MRS=SSR/1*	*MSR/MSE*
الخطأ	*n-2*	*SSE*	*MSE=SSE/(n-2)*	
المجموع	*n-1*	*Total SS*		

معظم الحسابات تتم باستخدام الحاسب الشخصي والتي تعطي جدول لتحليل التباين مشابه للجدول رقم (١١,٢). ولكن إذا تطلب الأمر استخدام الآلة الحاسبة اليدوية فإنه يمكن إجراء ذلك باستخدام المعادلات الحسابية التالية. حيث مجموع المربعات الكلي يمكن حسابه كما في المعادلة رقم (11.13) ومجموع مربعات الانحدارات يمكن حسابه باستخدام واحده من المعادلات الثلاث الموضحة في المعادلة رقم (11.14):

$$\textit{Total}\ \mathrm{SS} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_i^2 - n\overline{Y}^2 \qquad (11.13)$$

$$\mathrm{SSR} = b^2 \sum (X_i - \overline{X})^2 = b\left[\sum (X_i - \overline{X})(Y_i - \overline{Y})\right] = b\left[\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right] \qquad (11.14)$$

أما مجموع مربعات الخطأ فإنه يمكن حسابه بالطرح أو استخدام الصيغة الموضحة في المعادلة رقم (11.5):

$$SSE = \textit{Total}\ SS - SSR = \sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i \qquad (11.5)$$

وتتم مقارنة قيمة F المحسوبة من جدول تحليل التباين ANOVA بقيمه F الحرجة (الجدوليه) باستخدام درجه حرية تساوي الواحد للبسط ودرجه حرية تساوي $(n - 2)$ للمقام ومستوى المعنوية المختار. فإذا كانت قيمه F المحسوبة أكبر من قيمة F الحرجة (الجدوليه) من الجدول رقم (١٠) بالملحق فإننا نرفض فرض العدم المتضمن عدم وجود علاقة خطية بين المتغيرين X ، Y وما عدا ذلك فلا يتم رفضه.

معامل التحديد The Coefficient of Determination, r^2

المقياس العددي لنسبة الاختلاف في المتغير العشوائي Y المفسر بواسطة خط الانحدار هو معامل التحديد و يكتب r^2. وفي حالة الانحدار الخطي البسيط فإنه عبارة عن مربع معامل الارتباط r والذي تم عرضه سابقاً ، ولكن r^2 يقيس كيفية تمثيل خط انحدار المربعات الصغرى للبيانات الفعلية. ولذا فإنه يحسب كنسبة لمجموع مربعات الانحدار إلى مجموع المربعات الكلي من تحليل التباين كما في المعادلة رقم (11.16) التالية:

$$r^2 = \frac{\text{SSR}}{\text{TotalSS}} = \frac{b^2 \sum \left(X_i - \overline{X}\right)^2}{\sum Y_i^2 - n\overline{Y}^2} \qquad (11.16)$$

وكذلك يمكن التعبير عن معامل التحديد بصيغه رياضية أخرى كما في المعادلة رقم (11.17) التالية:

$$r^2 = 1 - \frac{\text{SSE}}{\text{TotalSS}} \qquad (11.17)$$

وكلما كانت قيمة r^2 كبيرة كلما دل ذلك على تمثيل أفضل لخط الانحدار للبيانات. ولكن ، في حالة أنواع معينة من البيانات مثل الأسعار والدخل وبيانات

الخصائص الاجتماعية والاقتصادية والتي تستخدم من قبل المختصين في الاقتصاد الزراعي فإنه ليس من المعتاد الحصول على قيمة لمعامل التحديد تساوي ٠,٦٠ أو قريب من ذلك والتي تعتبر جيده جداً .

وقيمة r^2 التي تعتبر جيده تعتمد بشكل كبير على نوع البيانات المدروسة. وعليه فإن هذا ملخص إحصائي معتاد يتم عرضه مع معادلة الانحدار المقدرة.

الخطأ المعياري للتقدير $S_{y.x}$ The Standard Error of the Estimate

تستخدم معادلة الانحدار مبدئياً لتقدير المتغير العشوائي التابع باستخدام قيم معينة للمتغير X. وبالحصول على القيم المقدّرة للمتغير Y نرغب في معرفة مدى الاعتماد عليها. وهذا يقيس مدى قرب نقاط البيانات وانتشارها حول خط الانحدار المرسوم في شكل الانتشار. وكلما كانت تلك النقاط قريبة من الخط كلما كان الاعتماد على القيم المقدرة لـ Y باستخدام المعادلة أكثر.

والخطأ المعياري للتقــدير $S_{y.x}$ يعطي مقياس عددي لانتشار النقاط حول معادلة الانحدار كما في شكل الانتشار الذي يعطي تفسير مرئي أو مشاهد. وكلما كانت قيمة الخطأ المعياري للتقدير صغيره كلما كان الاعتماد أكثر على القيم المقدرة لـ Y باستخدام معادلة الانحدار ، أي كلما كانت Y_e قريبه من نقاط البيانات Y_i لقيمة X المعطاة وبذلك تكون البيانات أقل ابتعاد من خط الانحدار وعليه فإن الحالة النادرة والتي تقع فيها كل قيم Y_i على خط الانحدار فإنه لن يكون هناك اختلاف وتكون $r^2 = 1$ و $S_{y.x} = 0$ ويمكن حساب الخطأ المعياري للتقدير باستخدام المعادلة الرياضية رقم (11.18) التالية :

$$S_{y \cdot x} = \sqrt{MSE} = \sqrt{(SSE)/(n-2)} \tag{11.18}$$

حيث MSE هي متوسط مربعات الخطأ من جدول تحليل التباين ANOVA والخطأ المعياري للتقدير نظرياً مشابه للانحراف المعياري وعليه فإن تفسيره مشابه له أيضاً. فإذا كان انتشار قيم Y حول معادلة الانحدار يتوزع طبيعياً ولدينا عينة كبيرة من البيانات فإن حوالي ٦٨٪ من النقاط في شكل الانتشار سوف تكون ضمن حدود خطأ معياري واحد للتقدير أعلى وأسفل خط الانحدار ، و٩٥,٥٪ من النقاط سوف تكون في مدى خطائين معيارين للتقدير أعلى وأسفل الخط ، وتقريباً الكل، أو ٩٩,٧٪، من النقاط سوف تكون في مدى ثلاثة خطأ معياري للتقدير أعلى وأسفل الخط. وعليه فإنه يمكننا إيجاد فتره الثقة حول خط الانحدار بناء على الخطأ المعياري للتقدير وكذلك إيجاد حدود الثقة لمتوسط Y .

فترة الثقة لـ $E(Y_e)$ Confidence Interval for

في تحليل الانحدار، ربما نهتم أكثر بتقدير متوسط الاستجابة لـ Y لقيمة معينة من المتغير X. وندرس المتوسط بسبب أن مجموعة البيانات المستخدمة لتقدير معادلة الانحدار عبارة عن عينة مختارة من عينات عشوائية كثيرة ربما تكون جمعت لقيم المتغير العشوائي Y وفقاً لمجموعه معينه من قيم X.

فإذا تم أخذ جميع العينات العشوائية المحتملة وقدّر خط الانحدار لكل واحدة، فإن الخطوط ليس من الضروري أن يكون لها نفس القاطع a والميل b. وبالتالي فإننا ربما نقدر عدة خطوط انحدار مختلفة ولكن جميع هذه الخطوط ستمر بنقاط المتوسطات متوسط X ومتوسط Y. وعليه فإن قيم X القريبة من متوسطها تكون القيم المقدرة

لـ Y المتحصل عليها من كل خطوط الانحدار الممكنة $(Y_e\ 's)$ وتكون أكثر تشابهاً مقارنة بحالة X البعيدة من متوسطها. ويتم الحصول على متوسط Y لقيمة معطاة لـ X من متوسط توزيع المعاينة لـ Y_e. وفي حالة أخذ عينة واحدة وحساب خط الانحدار المفرد لها بناء على بيانات العينة ، فإنه يمكننا إيجاد فتره الثقة لمتوسط Y باستخدام توزيع t كما هو موضح في المعادلة رقم (11.19) :

$$Y_e - t_{(\alpha/2,v)} \cdot S(Y_e) \le E(Y_e) \le Y_e + t_{(\alpha/2,v)} \cdot S(Y_e) \qquad (11.19)$$

حيث $S(Y_e)$ يمكن حسابها من المعادلة رقم (11.20). ويجب ملاحظة أن القيمة $S(Y_e)$ المحسوبة من المعادلة رقم (11.20) تكون كبيره كلما كانت قيم X_e بعيدة من متوسطها كما في مربع الفرق في الحد الموضح تحت الجذر.

$$S(Y_e) = S_{y.x}\sqrt{\frac{1}{n} + \frac{(X_e - \overline{X})^2}{\sum(X_i - \overline{X})^2}} \qquad (11.20)$$

مثال توضيحي An Example Problem

ترغب مديرة مبيعات محل زراعي في التنبؤ بالمبيعات الشهرية Y لشركتها باستخدام النفقات الإعلانية X. قامت بجمع بيانات للأداء السابق للمحل لمدة عشرة أشهر (الجدول رقم ١١،٣) أحسب الانحدار لـ Y على X واختبر مدى اختلاف معامل الميل عن الصفر من عدمه باستخدام مستوى معنوية ١٪ ، أحسب معامل التحديد r^2 ثم فسره ، ثم أوجد فتره ثقـة ٩٥٪ لمتوسط Y عندما تكون قيمه X_e=1000 دولار.

الجدول رقم (٣.١١). النفقات الإعلانية و المبيعات الشهرية (بالألف دولار) للمحل الزراعي.

النفقات الإعلانية X	المبيعات الشهرية Y	XY	X^2	Y^2
١,٢	١٠١	١٢١,٢	١,٤٤	١٠٢٠١
٠,٨	٩٢	٧٣,٦	٠,٦٤	٨٤٦٤
١,٠	١١٠	١١٠,٠	١,٠٠	١٢١٠٠
١,٣	١٢٠	١٥٦,٠	١,٦٩	١٤٤٠٠
٠,٧	٩٠	٥٦,٠	٠,٤٩	٨١٠٠
٠,٨	٨٢	٦٥,٦	٠,٦٤	٦٧٢٤
١,٠	٩٣	٩٣,٠	١,٠٠	٨٦٤٩
٠,٦	٧٥	٤٥,٠	٠,٣٦	٥٦٢٥
٠,٩	٩١	٨١,٩	٠,٨١	٨٢٨١
١,١	١٠٥	١١٥,٥	١,٢١	١١٠٢٥
٩,٤	٩٥٩	٩٢٤,٨	٩,٢٨	٩٣٥٦٩

الحل

لحل هذا المثال نبدأ أولاً بصياغة فرض العدم والفرض البديل كالتالي:

$$H_0 : \beta = 0$$
$$H_a : \beta > 0$$

ومن ثم استخدام اختبار طرف واحد باتجاه اليمين نظراً لتوقعنا زيارة المبيعات الشهرية كلما زاد الإنفاق على الدعاية والإعلان. وسيتم اختبار الفرض أعلاه باستخدام اختبار t وتحليل التباين. ولكن أولاً سيتم حساب معالم خط الانحدار a ، b حيث نبدأ حساب b أولاً .

$$b = \frac{\Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}} = \frac{924.8 - \frac{(9.4)(959)}{10}}{9.28 - \frac{(9.4)^2}{10}} = 52.567$$

$$a = \overline{Y} - b\overline{X} = 95.9 - (52.567)(0.94) = 46.486$$

وعليه تكون معادلة الانحدار المقدرة بطريقة المربعات الصغرى هي :

$$Y_e = 46.49 + 52.57X$$

ولاختبار الفرض لـ β باستخدام تحليل التباين ANOVA فإننا نحسب أولاً إجمالي مجموع المربعات SS :

$$\text{Total SS} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 93{,}569 - \frac{(959)^2}{10} = 1{,}600.9$$

$$SSR = b\left[\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right] = 52.567\left[924.8 - \frac{(9.4)(959)}{10}\right] = 1{,}226.93$$

ثم نحصل على مجموع مربعات الأخطاء بالطرح كالتالي :

$$SSE = \text{TotalSS} - SSR = 1{,}600.9 - 1{,}226.93 = 373.97.$$

ويمكن الآن إنشاء جدول تحليل التباين ANOVA واختبار الفرض (الجدول رقم ١١،٤).

الجدول رقم (١١.٤). جدول ANOVA لتحليل الانحدار للنفقات الإعلانية والمبيعات الشهرية.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
الانحدار	١	١٢٢٦,٩٣	١٢٢٦,٩٣	٢٦,٢٥
الخطأ	٨	٣٧٣,٩٧	٤٦,٧٥	
المجموع	٩	١٦٠٠,٩		

ويتضح أن قيمة F المحسوبة تساوي F= 26.25 في حين أن قيمة F الحرجة تساوي ١١,٢٦ من الجدول رقم (١٠) بالملحق عند مستوى ١٪ ودرجات حرية ١ ، ٨ . وحيث إن F المحسوبة أكبر من الجدوليه فإن القرار برفض فرض العدم المتضمن أن ميل خط الانحدار يساوي صفر ونقول بأن الإنفاق على الدعاية والإعلان يفسر الاختلاف في المبيعات الشهرية أكثر من متوسط المبيعات نفسها. ويمكن تحديد المقدار المفسر من ذلك الاختلاف في Y باستخدام معامل التحديد والذي يساوي :

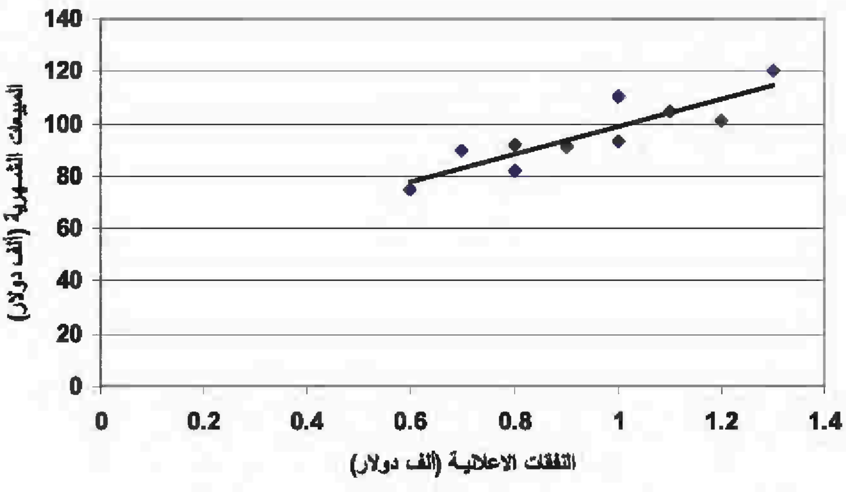
$$r^2 = \frac{SSR}{\text{Total SS}} = \frac{1{,}226.93}{1{,}600.90} = 0.77$$

وعليه فإن ٧٧٪ من الاختلافات في المبيعات الشهرية يتم تفسيرها بقيمة الإنفاق الشهري على الدعاية والإعلان. ولإجراء اختبار t لـ الميل β فإنه يجب أولاً حساب S_b ثم حساب قيمة t المحسوبة كالتالي :

$$S_b = \sqrt{\frac{MSE}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}} = \sqrt{\frac{46.75}{9.28 - \frac{(9.4)^2}{10}}} = 10.26$$

$$t = \frac{b - \beta}{S_b} = \frac{52.567 - 0}{10.26} = 5.12$$

ولاختبار الفرض فإننا نقارن t المحسوبة $t = 5.12$ بقيمه t الجدوليه والتي تساوي ٢,٨٩٦ من الجدول رقم (٨) بالملحق عند مستوى معنوية ١٪ ودرجة حرية ٨ درجات. وحيث إن قيمة t المحسوبة أكبر من قيمه t الجدوليه فإننا نرفض فرض العدم القائل بمساواة الميل للصفر (الشكل رقم ١١,٢) وهي نفس النتيجة التي تم التوصل إليها باستخدام تحليل التباين.



الشكل رقم (١١.٢). تقدير خط الإنحدار لبيانات الإعلان والمبيعات للمحل الزراعي.

ويلاحظ أن بعض الباحثين يكتبون قيمه S_b داخل قوسين وذلك تحت قيمة b عند كتابه معادلة الانحدار المقدرة، والبعض الآخر يفضل وضع قيمة t المحسوبة ووضع نجمة واحدة بجانب قيمة t إذا كانت معنوية عند مستوى ٥٪ أو نجمتين إذا كانت معنوية عند مستوى ١٪. وعند استخدام أي من الطريقتين فإنه يجب الإشارة إلى أي القيم التي تم استخدامها حتى يستطيع القارئ فهم ذلك. وعلى سبيل المثال يمكن

كتابة المعادلة المقدّرة آنفاً على النحو التالي:

$$Y_e = 46.49 + 52.57X$$
$$(5.12^{**})$$

و القيمة بين القوسين هي قيمة t المحسوبة.

ولاستخدام معادلة الانحدار لتقدير Y عندما X تساوي ١٠٠٠ دولار $X= \$1000$ فإنه يجب تحويل قيمة X أولاً حتى تتوافق مع الطريقة المستخدمة لكتابه البيانات الأصلية وبذلك تكون قيمه $X=\$1$ ثم نعوض عن قيمة X بـ ١ في المعادلة لإيجاد قيمة Y.

$$Y_e = 46.49 + 52.57X = 46.49 + 52.57(1) = 99$$

ولذا فإن المبيعات الشهرية تم تقديرها بالمعادلة وهي تساوي ٩٩٠٠٠ دولار عندما تكون النفقات الإعلانية تساوي ١٠٠٠ دولار. وهذا عبارة عن تقدير نقطة.

ولإيجاد فترة ثقة ٩٥٪ لمتوسط القيمة للمبيعات عندما تكون قيمه الإنفاق على الإعلان تساوي ١٠٠٠ دولار، يمكن استخدام صياغة فترة الثقة المشار لها سابقاً في المعادلة رقم (11.19) والمعادلة رقم (11.20) التالية:

$$Y_e - t_{(\alpha/2,v)} . S(Y_e) \leq E(Y_e) \leq Y_e + t_{(\alpha/2,v)} \cdot S(Y_e)$$

حيث:

$$S(Y_e) = S_{y \cdot x}\sqrt{\frac{1}{n} + \frac{(X_e - \overline{X})^2}{\sum(X_i - \overline{X})^2}}$$

ولحساب $S(Y_e)$ فأننا يجب أولاً أن نحسب $S_{y.x}$ والتي هي عبارة عن الجذر التربيعي لـ MSE كالتالي:

$$S_{y\cdot x} = \sqrt{MSE} = \sqrt{46.747} = 6.84$$

$$S(Y_e) = 6.84\sqrt{\frac{1}{10} + \frac{(1-0.94)^2}{0.444}} = 2.25$$

ومن الجدول رقم (٨) بالملحق نجد أن قيمة $t_{(0.025,8)}$ تساوي 2.306 . وبالتعويض عن تلك القيم في صيغة فتره الثقة آنفاً نجد أن :

$$99 - (2.306)(2.25) \le E(Y_e) \le 99 + (2.306)(2.25)$$

$$93.81 \le E(Y_e) \le 104.19$$

وعليه فإنه عندما يكون الإنفاق الإعلاني يساوي ١٠٠٠ دولار شهرياً، فإن المتوسط الشهري الحقيقي للمبيعات يتراوح بين ٩٣٨١٠ دولار و ١٠٤١٩٠دولار بفترة ثقة ٩٥٪ وهذا الفرق أكثر من ١٠٠٠٠ دولار والتي تعتبر كبيره. وفي الواقع العملي، إذا كان مدير المنشأة يرغب في عمل دراسات من هذا النوع فإنه يجب اختيار عينه كبيرة أكثر من عشرة والتي يجب أن تقلل الخطأ المعياري للتقدير $S(Y_e)$ والذي يجعل فترة الثقة أضيق.

الانحدار عندما تكون X عشوائية Regression when X Is Random

يفترض نموذج الانحدار للمربعات الصغرى العادية أن قيم X محددة و ثابتة. لذا فإن ثقة المعالم (المقدرات) ومخاطر الأخطاء تشير إلى العينات المعادة عندما تبقي قيم X نفسها من عينه إلى عينة.

في بعض الأحيان فإنه من غير المناسب اعتبار قيم X على أنها قيم ثابتة. فمثلاً، إذا حلّلنا أسعار القمح بناء على كمية القمح فإن كمية القمح أيضاً متغير

عشوائي لأن البائع الواحد لا يستطيع التحكم في هذا السوق لأنه يوجد فيه عدد كبير من البائعين وكميته قليله ولا يستطيع التأثير على الآخرين. وعليه فإنه من غير المنطق التفكير في العينات المعادة بحيث إن كميات القمح المباعة هي نفسها من عينة لأخرى.

لذا ففي حالة أن المتغيرين X ، Y عشوائية فهل نتائجنا السابقة غير ملائمة؟ الإجابة بالنفي حيث يمكن إيضاح أن جميع النتائج السابقة هي نفسها شريطه أن:

١- التوزيعات الاحتمالية الشرطية لـ Y_i بمعلوميه X_i طبيعية ومستقلة بمتوسط $\alpha + \beta X_i$ وتباين σ^2.

٢- X_i متغيرات عشوائية مستقلة والتي لها توزيعات احتماليه $f(X_i)$ لا تحتوي على المعالم α ، β ، σ^2 ، أي أن التوزيع لا يحتوي على معالم الانحدار.

والتغير الذي يجب عمله نظراً لأن المتغير X عشوائي هو في تفسير ثقة المعالم ومخاطر الخطأ. حيث إنها تشير الآن للعينات المعادة للأزواج (X_i, Y_i) حيث كلا المتغيرين تتغير من عينة إلى أخرى. أيضاً فإن قوة الاختبار تختلف عندما تكون X عشوائية.

البواقي Residuals

الباقي e_i يعرّف على أنه الفرق بين القيمة المشاهدة لـ Y والقيمة المقدرة من خط الانحدار (المعادلة رقم 11.21):

$$e_i \equiv Y_i - Y_e \qquad (11.21)$$

ويمكن اعتباره على أنه الخطأ المشاهد كعكس للخطأ الحقيقي ε_i في نموذج الانحدار، معادلة رقم (11.22):

$$\varepsilon_i = Y_i - E(Y_i) \qquad (11.22)$$

وفي نموذج الانحدار نحن نفترض أن ε_i متغيرات عشوائية طبيعية مستقلة بمتوسط $\mu = 0$ وتباين ثابت σ^2. وعند تحديد نموذج الانحدار بصورة صحيحة للبيانات، فإننا نتوقع أن يكون الخطأ المشاهد e_i له نفس الخصائص لـ ε_i. ويمكن استخدام هذه الفكرة كقاعدة لتحليل البواقي وهي طريقة جيدة لاختبار ملائمة نموذج الانحدار. والمتوسط للبواقي e_i والتي عددها n يساوي صفر كما هو موضح في المعادلة رقم (11.23):

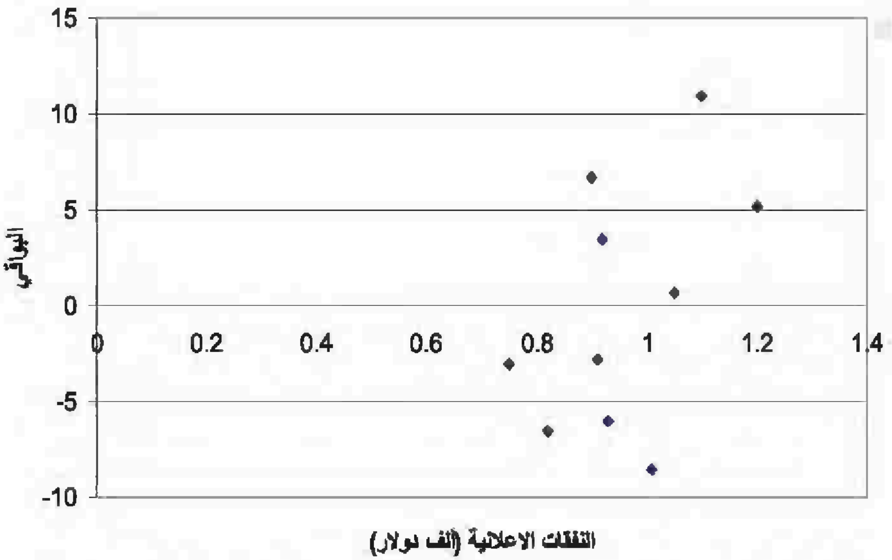
$$\bar{e} = \frac{\sum e_i}{n} = 0 \qquad (11.23)$$

وحيث إنها دائماً تساوي الصفر فإنها لا تعطي أي معلومة عما إذا كان الخطأ الحقيقي ε_i له متوسط يساوي الصفر. والتباين للبواقي e_i التي عددها n لنموذج الانحدار يساوي MSE كما هو موضح في المعادلة رقم (11.24):

$$\sigma^2 = \frac{\sum(e_i - \bar{e})^2}{n-2} = \frac{\sum e_i^{\,2}}{n-2} = \frac{SSE}{n-2} = MSE \qquad (11.24)$$

لذا فإنه إذا كان النموذج مناسب فإن المقدّر غير المتحيز لتباين الخطأ يساوي MSE.

إذا قمنا برسم البواقي على شكل انتشار مع X ، فإنه يمكننا معرفة مدى ملائمة نموذج الانحدار من شكل مجموع النقاط المتحصل عليها. فعندما يكون النموذج مقبول، فإن انتشار النقاط يتوزع بالتساوي إلى حد ما حول الصفر وتشكيل خط أفقي أو كرة بدون خط معين أو اتجاه. الشكل رقم (١١,٣) يوضح ذلك للبواقي للمثال السابق الخاص بالمبيعات والإنفاق الإعلاني.



الشكل رقم (١١.٣). شكل الانتشار للبواقي من نموذج الانحدار للنفقات الإعلانية والمبيعات الشهرية.

وفي حالة استخدام دالة انحدار غير خطية بدلاً من الخط المستقيم، فإن رسم البواقي سيأخذ شكل الخط المنحني. وفي هذه الحالة، فإننا نضيف متغير لـ X غير خطي مثل لوغارتيم X ($logX$)، أو قوى لـ X لمعادلة الانحدار كمتغير ثاني لـ X . وهذا لن يزيد من معامل التحديد فحسب بل تكون البواقي الجديدة للخط المعدّل لا تأخذ شكل المنحنى.

العامل الآخر الذي يمكن اختباره باستخدام شكل الانتشار للبواقي هو مدى ثبات التباين للخطأ. إذا أصبحت نقاط الانتشار في الاتجاه الرأسي بعيده عن الصفر واتخذت شكل مكبر الصوت أو النموذج الرباعي كلما زادت قيم X فإن تباين الخطأ كبير أو يزداد كلما زادت قيم X وليس ثابت. ويمكن أن نواجه حالات بحيث يقل التباين كلما زادت قيم X والتي تعطي شكل مشابه لشبه المنحرف بحيث تكون أكبر

للقيم الصغيرة لـ X. وكلا هذه الشروط تحدث بشكل متكرر في حالة البيانات العرضية. وفي حالة استخدام نموذج الانحدار المتعدد بأكثر من متغير X فإن رسم البواقي مقابل قيم Y_e على المحور الأفقي سوف تكون على شكل شبه منحرف إذا كان تباين الخطأ غير ثابت. نماذج المربعات الصغرى المرجحة، والتي خارج نطاق هذا الكتاب، تساعد على تصحيح عدم ثبات تباين الخطأ.

وعندما تكون بيانات نموذج الانحدار عبارة عن متغيرات عبر الزمن (سلاسل زمنيه) فإنه من المناسب رسم البواقي مقابل الزمن على المحور الأفقي. فإذا كان الشكل المتحصل عليه خط مستقيم بقيم سالبه للبواقي أولاً ثم قيم موجبة، أو إذا كانت على نمط محدد بحيث يمكن رسم خط خلال البواقي تتذبذب حول الصفر فإن حد الخطأ مرتبط مع الزمن.

ويمكن استخدام اختبار ديربن- واتسون للإشارة إلى مقدار الارتباط. وحدود الخطأ غير المستقلة جديرة بالاهتمام ولكن طرق المعالجة المقترحة لها مثل استخدام المتغيرات المبطأه خارج نطاق هذا الكتاب. وإذا كانت حدود الخطأ مستقلة عن الزمن فإنه لن يكون هناك لها نمط معين وسوف تقع أو تنتشر على شكل مجموعة عشوائية أفقية حول الصفر.

أخيراً ، إذا ما قررنا إضافة متغير مستقل معين لنموذج الانحدار لزيادة القوة التفسيرية له، فقد نرغب في رسم البواقي مقابل هذا المتغير. فإذا أوضحت البيانات اختلاف منهجي بحيث إنها في الغالب موجبه جميعها أو سالبة، أو تأخذ نمط معين مع المتغير فإنه قد يكون من المناسب إضافة ذلك المتغير للنموذج. وفي المقابل، إذا كانت البواقي تتوزع عشوائياً حول الصفر وتتخذ نموذج أفقي فإن إضافة ذلك المتغير للنموذج لن يضيف الكثير.

الانحدار المتعدد Multiple Regression

حيث إن نماذج الانحدار الخطي البسيط والتي تربط المتغير Y بمتغير مستقل X قويه ومفيدة، إلا أنه ربما نحتاج إلى إضافة متغيرات مستقلة X أخرى. وعموماً فإن هذا يتيح لنا شرح كامل للتغير في Y إضافة إلى تحسين قدرة النموذج التنبؤية وكذلك التقدير، وفي هذا الاتجاه يمكننا دراسة ثلاثة متغيرات أو أكثر معاً بدلاً من اثنين ولكن في هذه الحالة فإن الحسابات أكثر تعقيداً وكذلك تفسير النتائج، ويمكن تقليل الجهد الحسابي باستخدام البرامج الحاسوبية ولكن ذلك لن يساعد كثيراً في عملية التفسير.

نماذج الانحدار المتعدد Multiple Regression Models

يكون عرض نماذج الانحدار المتعدد بصورة أفضل باستخدام المصفوفات. ولكن جبر المصفوفات عموماً غير مطلوب للطلاب في العلوم الزراعية، لذا فإنه سيتم استخدام الطرق الجبرية في الشرح التالي. والطلاب الذين لديهم الرغبة في دراسة إضافية لهذا الموضوع عليهم تعلم جبر المصفوفات وأخذا مقررات متخصصة في الانحدار المتعلقة بهذا الموضوع.

وفيما يلي عرض لمعادلة الانحدار الخطي المتعدد. حيث إن متغيرات X في المعادلة يمكن أن تأخذ أي صورة، فمثلاً X_{3i} قد تكون عبارة عن حاصل ضرب X_{1i} في X_{2i} أو ربما تكون عبارة عن لوغاريثم X_2 ، أو ربما تكون مربع ، ... إلخ. في الغالب نختار نماذج متفاعلة وحدود مشروطة عليا كنماذج متكاملة ونسمي تلك النماذج الناقصة بنماذج الأثر المنخفض أو نماذج الأثر الرئيسي. وإذا كانت المعادلة خطية في معاملات الانحدار $(\beta_1, \beta_2, ...)$ فإننا نعتبرها معادلة خطية. وتستخدم β_0 كحد القاطع أو الثابت بدلاً من α وذلك لجعل العلامات أكثر تناسب (المعادلة رقم 11.25):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i \quad \textit{for } i = 1, 2, ..., n. \qquad (11.25)$$

وافتراضات نموذج الانحدار المتعدد هي تقريباً نفس شروط النموذج الخطي البسيط وهي:

١- متوسط الخطأ العشوائي ε_i يساوي صفر وتباينه σ^2 .

٢- الأخطاء العشوائية غير مرتبطة.

٣- β_0، β_1، .....، β_k ، هي $(k+1)$ معالم الانحدار و X_{1i} ، X_{2i} ، .... ، X_{ki} هي k متغيرات محدده ثابتة أو عشوائية معلومة.

٤- بهدف اختبار الفروض وتقدير فترات الثقة فإن ε_i يتوزع طبيعياً.

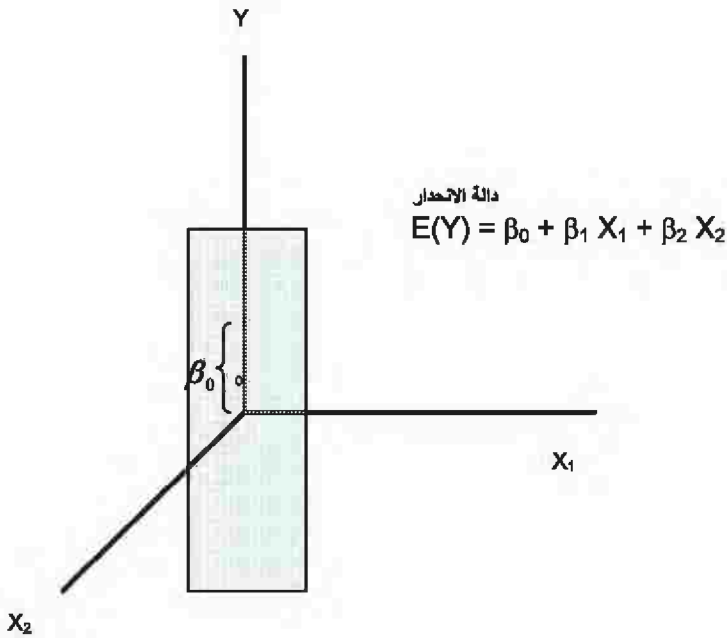
توضح الافتراضات أن متوسط Y أو القيمة المتوقعة $E(Y_i)$ لمجموعة معطاة من قيم X يمكن تعريفها بمعادلة الانحدار رقم (11.26):

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} \qquad (11.26)$$

والتي هي عبارة عن معادلة في مستوى أبعاده $(k+1)$ لأن Y تأخذ اتجاه مع عدد k لـ X . ولتبسيط ذلك يمكن خفض النموذج الخطي لنموذج ثلاثي الأبعاد وذلك بالتعامل مع متغيرين مستقلين X . في هذه الحالة، فإن النموذج يصف سطح المستوى كما في الشكل رقم (١١،٤). ويمكن اشتقاق المعادلات الطبيعية باستخدام المربعات الصغرى كما في الجزء الخاص بالانحدار الخطي البسيط، ماعدا أن هناك معادلة واحدة إضافية لكل X تم إضافته للنموذج. لذا ففي حالة وجود متغيرين مستقلين X فإن المعادلات الطبيعية محددة كما في المعادلة رقم (11.27):

$$\begin{aligned}
\sum Y_i &= nb_0 + b_1 \sum X_{1i} + b_2 \sum X_{2i} \\
\sum X_{1i} Y_i &= b_0 \sum X_{1i} + b_1 \sum X_{1i}^{\,2} + b_2 \sum X_{1i} X_{2i} \qquad (11.27) \\
\sum X_{2i} Y_i &= b_0 \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^{\,2}
\end{aligned}$$

ويمكن حل تلك المعادلات آنياً لإيجاد مقدرات المربعات الصغرى b_0 ، b_1 ، b_2 لـ β_i نظراً لأن لدينا ثلاث معادلات وثلاثة مجاهيل. ولعرض طريقة المربعات الصغرى في الانحدار المتعدد نستخدم المثال التالي:



الشكل رقم (١١.٤). مسطح دالة الانحدار في حالة متغيرين مستقلين.

مثال توضيحي An Example Problem

يرغب أحد المهندسين الزراعيين في تقدير العلاقة بين إنتاجيه القطن بالرطل للوبر لكل أيكر (Y) وكمية مياه الري المستخدمة بالأنش/ايكر (X_1) وكمية السماد المستخدمة (X_2) بالرطل/ايكر وذلك لمنطقه المزارع المروية. وقد تحصل على عينه

عشوائية لعدد ١٢ مزرعة والتي تعطي البيانات الموضحة في الجدول رقم (١١،٥). المطلوب تقدير نموذج الانحدار الخطي، واختبار نموذج الانحدار ومعالم الانحدار الجزئية لمعرفة ما إذا كانت تختلف عن الصفر، ثم أوجد تقدير لمعامل التحديد مع تفسير النتائج.

لحل هذه المسألة، فإننا ندخل البيانات في برنامج حاسوبي مثل برنامج اكسل. وبالتالي نحصل على معادلة الانحدار الموضحة بالمعادلة رقم (11.28).

$$Y_e = 248.063 + \underset{(3.88^{**})}{10.312X_{1i}} + \underset{(6.27^{**})}{4.285X_{2i}} \quad R^2 = 0.97 \quad (11.28)$$

الجدول رقم (١١.٥). بيانات الانحدار المتعدد لإنتاجية القطن وكمية مياه الري ومستوى التسميد.

إنتاجية القطن y	كمية مياه الري x_1	مستوى التسميد x_2
٤٢٠	٥	٣٠
٤٥٨	٩	٣٠
٤٠٠	٤	٢٥
٥١٠	٩	٤٠
٥٥٠	١٠	٤٠
٤٨٠	٧	٣٥
٥٧٥	١٢	٥٠
٤٦٠	٦	٣٥
٥٣٠	٩	٤٥
٦١٠	١٢	٥٥
٥٩٠	١٢	٥٥
٥٢٤	٩	٤٥

وقيمة t المحسوبة لكلا معاملات الانحدار الجزئية موضحة بين الأقواس تحت كل معامل في المعادلة رقم (11.28). وهي تختلف عن الصفر عند مستوى معنوية ١٪ ؛ نظراً لأن قيمة الاحتمال p اقل من ٠.٠١ ولذلك تم وضع نجمتين على القيمة. ويمكن إيجاد ٩٥٪ فتره ثقة لنماذج الانحدار والتي تساوي :

$$4.304 \le \beta_1 \le 16.319$$

$$2.740 \le \beta_2 \le 5.831$$

والخطأ المعياري للتقدير والذي نستخدمه لتحديد فترات الثقة تم حسابه بنفس الطريقة التي استخدمت في حاله الانحدار الخطي البسيط ولكن العلامات تختلف (معادلة رقم 11.29).

$$S_{y \cdot x_1 x_2} = \sqrt{MSE} \qquad (11.29)$$

وقد أوضح تحليل التباين لنموذج الانحدار بأن مستوى الانحدار ككل له ميل يختلف عن الصفر نظراً لأن قيمة F المحسوبة $F = 164.389$ معنوية عند مستوى احتمال ١٪ (انظر ملحق هذا الفصل للإطلاع على نتائج الحاسب لهذا النموذج).

معامل التحديد المتعدد $R^2 = 0.97$ يشير إلى أن نموذج الانحدار يفسر ٩٧٪ من التغير في إنتاجية القطن. ويتم حساب R^2 للانحدار المتعدد بنفس الطريقة المستخدمة في النموذج البسيط والموضحة في المعادلة رقم (11.30).

$$R^2 = SSR/TotalSS \qquad (11.30)$$

وكذلك فإن رسم البواقي هي جزء من نتائج الحاسوب. وتشير نتائج رسم البواقي مقابل كلا المتغيرين X_1 ، X_2 أن لها النمط الطبيعي. وكما ناقشنا في الجزء السابق فإن رسم البواقي هو إجراء مهم للتحقق من صحة النموذج الخطي.

ويمكن تفسير معالم الانحدار الجزئية مثلاً $b_1 = 10.312$ كالتالي. زيادة كمية المياه المستخدمة في الري بمقدار ١ بوصه/ ايكر تؤدي لزيادة إنتاجية القطن بمعدل ١٠.٣١٢ رطل للوبر/ايكر مع افتراض بقاء مستوى السماد (X_{2i}) ثابت. وإذا كانت إشارة المعالم b سالبه بدلاً من موجبه فإن زيادة X_{1i} بوحدة واحدة سيؤدي لنقص في قيمة Y_e بمقدار يساوي b_1 مع افتراض أن قيم X الأخرى ثابتة.

الارتباط الخطي المتعدد Multicollinearity

التفسير السابق لمعالم الانحدار الجزئية غير مناسب عندما تكون المتغيرات المستقلة X مرتبطة بقوة مع بعضها. وعندما تكون المتغيرات المستقلة في النموذج الخطي مرتبطة داخلياً بدرجه عاليه فأننا نقول بظهور الارتباط الخطي. في هذه الحالة من الارتباط الخطي تكون مقدرات المربعات الصغرى b_0 ، b_1 ، ... b_k لمجتمع من معالم الانحدار β_0 ، β_1 ، ... β_k غير كفؤة؛ نظراً لأن التباينات لتوزيعات المعاينة كبيره جدا. لذا فإن b_i لا تعطي إجابة مناسبة سواء بالنسبة لقيمتها نفسها أو إشارتها، أي أن المعرفة التي لدينا عن النماذج الخطية بناء على النظرية الخاصة بموضوعنا والتي نستخدمها لصياغة وتفسير النموذج تشير إلى أن قيم b المقدرة أو إشاراتها تقع خارج الخط. فإذا كانت درجه الارتباط الخطي حادة، فإن أحد المتغيرات المستقلة X يمكن كتابته كمربع خطي لمجموعة جزئية من الآخر ونموذج الانحدار يحتوي فعلاً مقدرين لنفس البيانات. والحل في مثل هذه الحالة هو حذف المتغير X المخالف من النموذج، ولكن لمعظم الحالات فإن العلاقة الخطية بين المتغيرات المستقلة X غير واضحة ونحتاج لطرق متقدمة لاستخدامها وليست من هدف موضوعنا هنا. وللحصول على معلومات إضافية عن هذا الموضوع يمكنك الرجوع لكتاب متقدم في الانحدار المتعدد أو الاقتصاد القياسي.

السلاسل الزمنية

Time series

يهتم الباحثين والمدراء ومسوقي المنتجات الزراعية بالسلاسل الزمنية لأنها تساهم في شرح ما حدث في الماضي والتنبؤ بالأحداث المستقبلية. وتحليل السلاسل الزمنية عبارة عن اختبار مبدئي للبيانات التي تم جمعها عبر الزمن مثل الأسعار الشهرية المدفوعة من قبل المزارعين للمدخلات منذ سنة ١٩٠٣. وبصفة عامة فإن السلاسل الزمنية عبارة عن مجموعه من البيانات الإحصائية والتي يتم جمعها وتسجيلها أو ملاحظتها خلال فتره زمنيه محدده. ويحدث التغير في السلاسل الزمنية نتيجة مجموعة من العوامل تؤثر عليها. فعلى سبيل المثال، في أوائل ١٩٠٠ كان المزارعين لا يشترون مدخلات للإنتاج ماعدا بعض العناصر مثل حيوانات الجر والآلات. ولكن مع تطور الزمن و توفر التسميد التجاري إضافة للبذور المهجنة والجرارات والتجهيزات الملحقة بها وجد المزارعين أنفسهم يشترون الكثير والكثير من العناصر المستخدمة في عملياتهم الزراعية. لذا أصبحت الأسعار المدفوعة للمدخلات أكثر أهمية كلما زادت قائمة تلك المدخلات. وهناك مجموعة من الحوافز لمراقبة البيانات واستخدامها كمقياس لرفاهية المزارع بالإضافة إلى البيانات الاقتصادية المهمة بعض منها على مستوى الأعمال الفردية المحلية والبعض الآخر يتم إصداره من قبل الصناعة أو المصادر الحكومية.

وإذا افترضنا أن هناك مكونات عادية ودورية تتفاعل معاً بطريقة يمكن التنبؤ بها لإعطاء السلاسل الزمنية الحالية، فإنه يمكن تحليلها بهدف التنبؤ بالسلسلة لفترة زمنيه مستقبلية. وتحليل السلاسل الزمنية طريقه تحاول فصل البيانات إلى مكوناتها التالية :

١- الاتجاه الزمني.

٢- التقلبات الدورية.

٣- التغيرات الموسمية.

٤- التحركات العرضية أو الفجائية.

الاتجاه الزمني يشير إلى الحركة الانسيابية باتجاه الأعلى أو الأسفل والتي تتميز بها السلسلة الزمنية خلال فتره من الزمن طويلة مثل عشر سنوات أو أكثر. ويعزى ذلك للتأثير الكامن للقوى الأساسية المؤثرة في البيانات مثل النمو السكاني، التغيرات التكنولوجية والتحوّل إلى السعات الكبيرة في أذواق وتفضيلات المستهلكين. لذا فإن الاتجاه يوضح حركه ثابتة ويتطلب وصفها نوع مختلف من النماذج عن تلك المستخدمة للمكونات الأخرى.

التقلبات الدورية هي التحركات في الدورات الاقتصادية والتي تعكس توسع الاقتصاد من فتره الكساد المنخفضة إلى الطفرة الاقتصادية وبالتالي انكماشها باتجاه كساد جديد. والدورة الاقتصادية تكون طويلة في العادة من سنة إلى ١٥ سنة للتحرك من قاع إلى قاع والتغيرات خلال فتره الانتعاش والانكماش. ويتم ذلك حول خط الاتجاه الزمني. وتعتمد فتره الانتعاش والانكماش على مدى تعافي الاقتصاد وتأثير الأحداث الاقتصادية على بيانات السلاسل الزمنية موضع الدراسة. وبسبب طبيعة الدورات الاقتصادية لم يتم استنباط طريقه ذات دقه عاليه للتنبؤ بالتقلبات الدورية في بيانات السلاسل الزمنية. ويتم توظيف طريقة البواقي، أي بعد استبعاد أثر الاتجاه الزمني والتغيرات الموسمية من بيانات السلسلة الزمنية، ويتبقى أثر التغيرات الدورية والعرضية، وبصفه عامه لا تنفصل هذين المكونين عن بعضها. ونظراً لأن التحركات الدورية كبيره بالنسبة للتغيرات العرضية فإن هذه الطريقة ممكنه.

التغيرات الموسمية هي دورات قصيرة الأجل داخل البيانات والتي تكتمل خلال السنة وتتكرر سنوياً. ويؤدي الطقس والعادات لحدوث التغيرات الموسمية. فمثلاً في

الزراعة فإن فتره الحصاد للمحاصيل الرئيسية تتطلب شراء أنشطه بواسطة المزارعين والتي تتكرر كل سنة في نفس الوقت تقريباً. وهذا يجعل مبيعات الإنتاج الحالي عالية في فتره معينه كل سنة يعني نمط موسمي. وبيانات السلاسل الزمنية الشهرية أو الربع سنوية عادة تحتاج لاختبارها حيال التغيرات الموسمية.

التحركات العرضية في بيانات السلاسل الزمنية لا تتبع نمط محدد لأنها تحدث بسبب تأثير الحرب، أخطار الطقس مثل الثلج والفيضانات، البراكين والعواصف، والإضرابات والحوادث وغير ذلك. لذا فإنه لا يوجد لدينا طريقه يمكن بها التنبؤ بمثل هذه الأحداث ولم يتم عمل نماذج لذلك. وتبقى التغيرات العرضية في البواقي بعد إزالة العوامل الأخرى من البيانات.

تحليل الاتجاه الزمني Secular Trend Analysis

نظراً لأن بيانات السلاسل الزمنية تمتد لعدد كبير من السنوات، فقد يكون هناك تغيرات مهمة خلال تلك الفترة نظراً لتغير السكان، تغير مستويات الأسعار، أو تغير في تعريفات بعض العناصر المهمة مثل تعريف المزرعة للبيانات الزراعية. وقد نحتاج لقسمه بيانات السلسلة على عدد السكان للحصول على البيانات للفرد وتحليل الاتجاه بالنسبة للفرد إذا كان التغير في السكان يؤثر على الاتجاه. وربما نرغب في قسمة بيانات النقود بمؤشر يعبّر عن التغير في القوة الشرائية للدولار عبر الزمن قبل تقدير الاتجاه الزمني. إذا كانت البيانات تتطلب العمل مع تعريف المزرعة المستخدم فإنه ربما يتطلب الأمر تعديل البيانات حسب التغير في التعريف؛ نظراً لأن إدارة الزراعة الأمريكية تعمل تغيرات دوريه لذلك. وبعد الاطمئنان بعدم وجود أي مشكلة منهجية في البيانات يمكن القيام بتقدير الاتجاه الزمني.

الاتجاه الخطي Linear Trend

يجب أولاً تحديد ما إذا كان الاتجاه خطي. ومن الوسائل المساعدة لذلك رسم البيانات مع الزمن على المحور الأفقي X باستخدام شكل الانتشار وملاحظة نمط النقاط لها. فإذا وقعت تقريباً على خط أفقي فإنه يمكننا رسم الاتجاه الخطي للبيانات باستخدام تحليل الانحدار. وعند استخدام الانحدار فنحن عادة نرمز لمتغير الزمن بـ X ونستخدم القيم الترميزية لتقدير المعادلة بدلاً من قيم السنوات. والطريقة المعتادة أن نرمز للسنة الأولى في السلسلة بـ صفر والسنة الثانية بـ ١ وهكذا حتى نصل إلى نهاية بيانات السلسلة. وعند ذلك نستخدم الترميز المناسب لـ X في معادلة الانحدار للتنبؤ بقيمه Y.

والطريقة الأخرى البديلة هي ترميز السنة الوسطى في البيانات بالرمز صفر و السنة التالية بعد ذلك بالرقم ١ وهكذا والسنوات السابقة -١ ، -٢ ، - ٣ وهكذا حتى نصل إلى نقطة البداية للبيانات. وبهذا الإجراء فإنه مجموع قيم X يساوي صفر. أو عندما يكون لدينا أعداد زوجية من السنوات يمكن ترميز البيانات الوسطية للسلسة بالرمز ١ ، -١ للرقمين التي تقع في المنتصف واستخدام الأرقام الفردية للتحرك للأمام والوراء مثل ١ ، ٣، ٥ ... إلخ و -١ ، -٣ ، -٥ ، حتى نصل إلى نهاية البيانات. وقيم X التي تم ترميزها بهذه الطريقة يكون مجموعها مساوي للصفر. ويتم اختيار أي من خيارات الترميز أكثر ملائمة مع تعريف القارئ بنوع طريقه الترميز المستخدمة للبيانات.

الاتجاه غير الخطي Nonlinear Trends

إذا كانت السلسلة الزمنية غير خطية ، فإننا نستخدم تحليل الانحدار المتعدد. ويجب أن لا يكون الاهتمام بإيجاد المعادلة الملائمة للبيانات ولكن أيضاً الصيغة الممكن استخدامها وفقاً للطبيعة الاقتصادية للسلسلة الزمنية موضع الدراسة.

ومن تلك الصيغ المنحنى الأسى والذي يفسر السلسلة الزمنية التي تزيد أو تنقص بمعدل ثابت عبر الزمن مثل نمو السكان، مبيعات المنتج الجديد، أو انتشار الأمراض المعدية (المعادلة رقم 11.31).

$$Y = ab^x \tag{11.31}$$

ويعتمد شكل الدالة على قيم a و b . حيث a هي القاطع لـ Y ولكن المنحنى يتناقص للقيم التي تكون عندها b بين الصفر والواحد ويتزايد عندما تكون قيم b أكبر من الواحد. ويتم تقدير المعادلة بأخذ اللوغاريثم للطرفين كما في المعادلة رقم (11.32) وهذه المعادلة لوغاريثميه خطية.

$$\log Y = \log (ab^x) = \log a + X \log b \tag{11.32}$$

الجدول رقم (١١.٦). بيانات السلسلة الزمنية لمبيعات الأسمدة من ١٩٨٩-١٩٩٦م.

الزمن x	مبيعات الأسمدة Y	ترميز x	لوغارثيم Y
١٩٨٩	٢٦,٠	٠	١,٤١٥٠
١٩٩٠	٢٨,١	١	١,٤٤٨٧
١٩٩١	٣٠,٣	٢	١,٤٨١٤
١٩٩٢	٣٢,٨	٣	١,٥١٥٩
١٩٩٣	٣٥,٤	٤	١,٥٤٩٠
١٩٩٤	٣٨,٣	٥	١,٥٨٣٢
١٩٩٥	٤١,٢	٦	١,٦١٤٩
١٩٩٦	٤٤,٥	٧	١,٦٤٨٤

ويمكن تقديره باستخدام طريقة المربعات الصغرى. ولتقدير المعادلة يتم ترميز X كما تم شرحه في الجزء السابق لنقل من صفر إلى n لعدد n من السنوات ثم نأخذ اللوغاريتم لقيم Y. وبالتالي فإن القيم المقدرة لمعالم الانحدار هي $\boldsymbol{loga}$ و $logb$. ونحصل على تقدير لقيم Y ثم نأخذ اللوغاريتم العكسي لـ $(logY)$ للحصول على قيم Y المقدرّة. ويمكن إيضاح ذلك من خلال المثال التالي.

تم الحصول على المبيعات السنوية للأسمدة في المنطقة الزراعية (الجدول رقم ١١,٦) ونرغب في تقدير الاتجاه الزمني باستخدام المنحنى الأسى.

لحل هذا المثال يتم أولاً ترميز السنوات قيم X وإيجاد اللوغاريثم لقيم Y ثم تقدير المعادلة الأسية باستخدام الانحدار الخطي. وتوضح المعادلة رقم (11.33) المعادلة المقدرة حيث تشير الأرقام بين الأقواس لقيم t المحسوبة والتي يتضح معنويتها عند مستوى ١٪.

$$\underset{\phantom{(334.6^{**})}}{\log Y = 1.4153 +} \underset{(334.6^{**})}{0.0334X} \qquad (11.33)$$

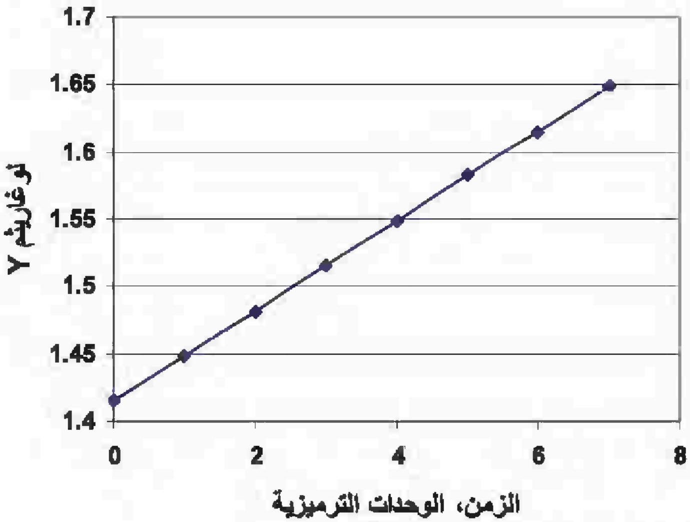
ولتقدير القيم الاتجاهية لسنه ١٩٩٧ فإننا نعوض بالترميز الخاص بالسنة ١٩٩٧ في المعادلة السابقة والحل لإيجاد قيمه $logY$ ثم إيجاد اللوغاريتم العكسي.

$$\log Y = 1.4153 + 0.0334(8) = 1.6825$$

$$\text{Antilog}(1.6825) = 48.1$$

لذا فإننا نتوقع أن تكون مبيعات الأسمدة ٤٨,١ مليون دولار لهذه المنطقة في عام ١٩٩٧ إذا كانت تتبع الاتجاه. وقد تم رسم خط الانحدار للبيانات في الشكل رقم

(١١,٥). ويجب أولاً ملاحظة أن نقاط البيانات تقريباً خطيه عندما حولّناها إلى اللوغاريثم بالنسبة لـ Y ورمّزنا قيم X وعليه فإنه ليس من المستغرب ملائمة خط الانحدار لرسم تلك البيانات.



الشكل رقم (١١.٥). تمثيل الخط للاتجاه غير الخطي.

بالإضافة لاستخدام المنحنى الأسى يمكن استخدام معادلة كثيرة الحدود من الدرجة الثانية لإيجاد الاتجاه غير الخطي للبيانات (المعادلة رقم 11.34).

$$Y = b_0 + b_1X + b_2X^2 \qquad (11.34)$$

والتي تعتبر مناسبة بصفة عامة عندما يكون هناك معدل ثابت للتغير في قيم Y عبر الزمن. ولتسهيل تحديد معدل التغير في Y يتم أولاً حساب التغير في Y من سنه إلى أخرى بالطرح. بعد ذلك يتم طرح البيانات السابقة (معدل التغير) بطرح السنة من التي بعدها. فإذا كان معدل التغير والناتج من الطرح الثاني هو تقريباً نفسه لكامل

الفترة فإن معادلة الدرجة الثانية تكون ملائمة لتقدير هذه البيانات.

ومن الصيغ غير الخطية والتي تستخدم غالباً لرسم الاتجاه دالة النمو أو الجيومتريه، المنحنى، المنحنى اللوجستي والمنحنى الأسى المعدّل ولكن لن تتم دراسة هذه الصيغ هنا.

التغيرات الموسمية Seasonal Variation

أكثر الطرق شيوعاً لتنعيم البيانات أو تقدير التغيرات الموسمية هي طريقه النسبة إلى المتوسط المتحرك. وتتطلب هذه الطريقة بأن يكون لدينا بيانات شهريه أو ربع سنوية للسلسلة الزمنية. وسيتم مناقشة الخطوات اللازمة في حالة البيانات الربع سنوية، ولكن يمكن استخدام نفس المنطق بالنسبة للبيانات الشهرية.

الخطوات هي:

١- حساب المتوسط المتحرك للأرباع الأربعة والتي تشتمل على الاتجاه الزمني والتغيرات الدورية الموجودة في البيانات الأساسية. والمتوسط المتحرك هو عبارة عن المتوسط السنوي لبيانات الأرباع الأصلية بما فيها ربع سابق. هذا يعني أننا نأخذ المتوسط للأربعة الأرباع الأولى في السلسلة الزمنية ثم نسقط الربع الأول ونضيف الربع الخامس ونحسب المتوسط للأرباع الأربعة ثم نسقط الربع الثاني ونضيف الربع السادس ونحسب المتوسط.. إلخ حتى يتم الانتهاء من كامل بيانات السلسلة.

٢- الخطوة الثانية هي قسمة البيانات الأصلية لكل ربع على متوسطها المتحرك المناظر لها للحصول على "النسبة إلى المتوسط المتحرك" والتي تحتوي الآن على المكونات الموسمية والعرضية لأن حساب المتوسط المتحرك ساهم في التخلص من الاتجاه الزمني والعوامل الدورية.

٣- نعيد ترتيب أرقام " النسبة إلى المتوسط المتحرك" للأرباع بحيث تكون أرقام الربع الأول معاً والربع الثاني معاً..إلخ ونأخذ المتوسط لهذه القيم وذلك لمحاولة تخليص البيانات من المكونات العرضية أو الفجائية. ونستخدم المتوسط المعدل لهذا الغرض، أي نسقط أعلى قيمة وأقل قيمة ثم نحسب المتوسط للقيم المتبقية. وسوف نحصل على أربعه متوسطات معدّلة متوسط للربع الأول، وآخر للربع الثاني، ...إلخ.

٤- يتم تعديل المتوسطات المعدلة بحيث يكون مجموعها مساوي لـ ٤٠٠ ومن ثم يكون المتوسط لكل منها ١٠٠. ويمثل المتوسط المعدل " مؤشر الموسمية " والذي نستخدمه للتعديل الموسمي للبيانات الأصلية وعليه تخليص بيانات السلسلة الزمنية من أثر التغيرات الموسمية. ويمكن إيضاح ذلك بالمثال التالي.

يوضح الجدول رقم (١١،٧) البيانات الخاصة بالصادرات للخضار ومنتجات الخضار من ١٩٩٢ إلى ١٩٩٦ بالبليون دولار. المطلوب حساب مؤشر الموسمية لهذه البيانات.

الجدول رقم (١١.٧). صادرات الخضار ومنتجات الخضار الربع سنوية بالبليون دولار.

السنة والربع		صادرات الخضار	إجمالي المتوسط المتحرك الربعي	المتوسط المتحرك	نسبة البيانات الأصلية للمتوسط المتحرك
١٩٩٢	I	١،٨٦			
	II	٢،٢٠			
	III	٢،٤٢	٨،٣٩	٢،١٠	١١٥،٢
	IV	١،٩١	٨،٤٥	٢،١١	٩٠،٥
١٩٩٣	I	١،٩٢	٨،٥٩	٢،١٥	٨٩،٣
	II	٢،٣٤	٨،٦٨	٢،١٧	١٠٧،٨
	III	٢،٥١	٨،٧٥	٢،١٩	١١٤،٦
	IV	١،٩٨	٨،٦٦	٢،١٦	٩١،٧

تابع الجدول رقم (١١.٧).

السنة والربع		صادرات الخضار	إجمالي المتوسط المتحرك الربعي	المتوسط المتحرك	نسبة البيانات الأصلية للمتوسط المتحرك
١٩٩٤	I	١,٨٣	٨,٤٧	٢,١٢	٨٦,٣
	II	٢,١٥	٨,٠٦	٢,٠٢	١٠٦,٤
	III	٢,١٠	٧,٨٨	١,٩٧	١٠٦,٦
	IV	١,٨٠	٧,٨٩	١,٩٧	٩١,٤
١٩٩٥	I	١,٨٤	٧,٧٦	١,٩٤	٩٤,٨
	II	٢,٠٢	٧,٩٦	١,٩٩	١٠١,٥
	III	٢,٣٠	٨,١٢	٢,٠٣	١١٣,٣
	IV	١,٩٦	٨,٣٥	٢,٠٩	٩٣,٨
١٩٩٦	I	٢,٠٧	٨,٧٤	٢,١٨	٩٥,٠
	II	٢,٤١	٨,٧٨	٢,٢٠	١٠٩,٥
	III	٢,٣٤	٨,٧٩	٢,٢٠	١٠٦,٤
	IV	١,٩٧			

الخطوات الأولى لحساب مؤشر الموسمية هي الحصول على الإجمالي المتحرك للأرباع الأربعة لبيانات السلسلة الزمنية. لذا فإنه لبيانات ١٩٩٢ نجمع ١,٨٦ ، ٢,٢٠ ، ٢,٤٢ ، ١,٩١ للحصول على الإجمالي الذي يساوي ٨,٣٩ ونضعه قريب من متوسط البيانات مقابل للربع الثالث. نحسب بعد ذلك الإجمالي المتحرك الثاني ونضعه مقابل الربع الرابع حيث أسقطنا الربع الأول ١,٨٦ وأضفنا الربع الخامس ١,٩٢ ونستمر

بنفس الطريقة حتى نحسب جميع القيم للإجمالي المتحرك.

الخطوة التالية هي حساب المتوسط المتحرك للأرباع الأربعة بقسمه الإجمالي المتحرك لكل ربع على الرقم (٤) وكتابه الناتج في الجدول.

في الخطوة الثالثة نحسب الموسمية المحددة بقسمة البيانات الأصلية على المتوسط المتحرك ونضربها في ١٠٠ ليكون الناتج على شكل نسبة.

والآن يمكن حساب مؤشر الموسمية. ولعمل ذلك نكتب المواسم المحددة (الجدول رقم ١١.٨) وذلك للأرباع والسنوات ويكون لدينا الآن أربع قيم لكل ربع.

الجدول رقم (١١.٨). نسب البيانات الأصلية للمتوسط المحرك وفقاً للربع.

السنة	الربع I	الربع II	الربع III	الربع IV
١٩٩٢			~~١١٥,٢~~	~~٩٠,٥~~
١٩٩٣	٨٩,٣	١٠٧,٨	١١٤,٦	٩١,٧
١٩٩٤	~~٨٦,٣~~	١٠٦,٤	~~١٠٦,٦~~	٩١,٤
١٩٩٥	٩٤,٨	~~١٠١,٥~~	١١٣,٣	~~٩٣,٨~~
١٩٩٦	~~٩٥,٠~~	~~١٠٩,٥~~		
المتوسط	٩٢,٠	١٠٧,١	١١٤,٠	٩١,٦

وحيث أننا نرغب في إيجاد متوسط واحد فإننا نأخذ المتوسط للأربعة للحصول على المتوسط المعدل وذلك بشطب القيمة الصغرى والكبرى في كل عمود ثم إيجاد المتوسط لباقي القيم. وهذه المتوسطات تقريباً هي مؤشر الموسمية ولكنها غير دقيقه لأن مجموعها لا يساوي ٤٠٠. في مثالنا الحالي مجموع المتوسطات المعدلة يساوي ٤٠٤,٧. ويمكن تحويلها ليكون مجموعها ٤٠٠ بأخذ النسبة ٤٠٠/٤٠٤,٧ = ٠,٩٨٨٤ وضربها بكل متوسط للحصول على مؤشر الموسمية (الجدول رقم ١١,٩). والآن أوجدنا

مؤشر الموسمية والموضح في العمود الثالث من الجدول رقم (١١،٩)، والتي تستخدم لتخليص البيانات من أثر الموسمية.

الجدول رقم (١١.٩). حساب مؤشر الموسمية.

الربع	المتوسط المعدل	مؤشر الموسمية
I	٩٢،٠	٩٠،٩
II	١٠٧،١	١٠٥،٩
II	١١٤،٠	١١٢،٧
IV	٩١،٦	٩٠،٥

ويتم تخليص البيانات الأصلية لصادرات الخضار من أثر الموسمية بقسمه كل قيمه بمؤشر الموسمية بعد قسمته على ١٠٠ لكل ربع. فمثلاً الربع الأول من ١٩٩٢ قيمة الصادرات ١،٨٦ بليون ومؤشر الموسمية للربع الأول ٩٠،٩ (الجدول رقم ١١،٩). لذا نقسم المؤشر ٩٠،٩ على ٠،٩٠٩ ثم نقسم الرقم ١،٨٦ عليه للحصول على القيمة الخالية من الموسمية والتي تساوي ٢،٠٥ (الجدول رقم ١١،١٠). ويتم إعادة نفس العملية حتى نكمل باقي السلسلة الزمنية.

ملاحظه ختامية Endnote

(١) هذه العبارة هي نتيجة مباشرة لنظريه جاوس ماركوف والتي تنص على أنه تحت شروط النموذج $Y_i = a + bX_i + \varepsilon_i$ حيث ε_i هي الفرق $(Y_i - Y_e)$ ، فإن مقدرات المربعات الصغرى a ، b غير متحيزة ولها أقل تباين ضمن جميع المقدرات الخطية غير المتحيزة.

الجدول رقم (١١.١٠). بيانات صادرات الخضار ومنتجات الخضار بالبليون دولار بعد استبعاد أثر الموسمية.

السنة والربع		صادرات الخضار	مؤشر الموسمية	الصادرات بعد استبعاد الموسمية
١٩٩٢	I	١,٨٦	٩٠,٩	٢,٠٥
	II	٢,٢٠	١٠٥,٩	٢,٠٨
	III	٢,٤٢	١١٢,٧	٢,١٥
	IV	١,٩١	٩٠,٥	٢,١١
١٩٩٣	I	١,٩٢	٩٠,٩	٢,١١
	II	٢,٣٤	١٠٥,٩	٢,٢١
	III	٢,٥١	١١٢,٧	٢,٢٣
	IV	١,٩٨	٩٠,٥	٢,١٩
١٩٩٤	I	١,٨٣	٩٠,٩	٢,٠١
	II	٢,١٥	١٠٥,٩	٢,٠٣
	III	٢,١٠	١١٢,٧	١,٨٦
	IV	١,٨٠	٩٠,٥	١,٩٩
١٩٩٥	I	١,٨٤	٩٠,٩	٢,٠٢
	II	٢,٠٢	١٠٥,٩	١,٩١
	III	٢,٣٠	١١٢,٧	٢,٠٤
	IV	١,٩٦	٩٠,٥	٢,١٧
١٩٩٦	I	٢,٠٧	٩٠,٩	٢,٢٨
	II	٢,٤١	١٠٥,٩	٢,٢٨
	III	٢,٣٤	١١٢,٧	٢,٠٨
	IV	١,٩٧	٩٠,٥	٢,١٨

ملحق الفصل ١١: الارتباط والانحدار باستخدام برنامج اكسل
Appendix to Chapter 11: Correlation and Regression with Excel

برنامج اكسل يحتوي على أوامر خاصة للارتباط والانحدار. وسيتم العمل أولاً باستخدام مثال لتوضيح عمل البرنامج لإيجاد الارتباط ثم نتطرق لمثالين آخرين لإيجاد الانحدار باستخدام اكسل.

الارتباط Correlation

سنقوم باختبار نتيجة ترتيب قطع لحم البقر حسب درجة الجودة X مقابل سعر البيع للرطل Y باستخدام طريقه الارتباط من قائمه الأدوات ثم تحليل البيانات في برنامج اكسل. أولاً ندخل البيانات في ورقة العمل على شكل أعمدة منفصلة X ، Y مع مراعاة وضع الرموز في الصف الأول وإدخال البيانات ثم اختيار الأمر كما سبق. يظهر لنا جدول تلقائي نضع في الصفوف الأول مدى البيانات ثم نعلّم على خيار وجود علامات للبيانات ونحدد المكان المرغوب للنتيجة ثم نوافق على ذلك. سيظهر لنا جدول النتيجة الموضح (الجدول رقم ١١.١١).

الجدول رقم (١١.١١). نتائج الارتباط من ورقة عمل برنامج اكسل.

	درجة جودة النوع	سعر البيع للرطل
درجة جودة النوع	١	
سعر البيع للرطل	٠,٩٣٨٧٣٩٨٦	١

ويجب ملاحظة أن معامل الارتباط الناتج هو نفسه معامل الارتباط الذي سبق حسابه ماعدا التقريب (٠,٩٣٨٧ أو ٠,٩٤). ويجب ملاحظة عدم توفر اختبار للمعامل في البرنامج.

الانحدار البسيط Simple Regression

يمكن استخدام المثال الخاص بمحل بيع الأدوات الزراعية للنظر في طريقة إجراء الانحدار الخطي باستخدام برنامج اكسل. أولاً ندخل البيانات في ورقه العمل ثم نختار

الأمر انحدار Regression من قائمة الأدوات ثم تحليل البيانات. يظهر لنا جدول تلقائي ندخل مدى بيانات المتغير Y في الصندوق الأول ثم بيانات مدى المتغير X في الصندوق الثاني ونعلّم المربع الخاص بالعلامات (أسماء المتغيرات) وكذلك المربع الخاص بمستوى فتره الثقة ثم نحدد المدى للنتائج ونختار المربع الخاص بالبواقي وكذلك المربع الخاص برسم البواقي وربما المربع الخاص برسم الخط ثم نختار موافق. تظهر لنا النتائج الموضحة في الجدول رقم (١١.١٢). ويجب ملاحظة أنها نفس النتائج التي تم التوصل إليها سابقاً باستخدام المعادلات والآلة الحاسبة ماعدا التقريب.

الجدول رقم (١١.١٢). نتائج الانحدار البسيط باستخدام ورقة عمل برنامج اكسل.

خلاصة النتائج	
إحصاءات الاختبار	
R المتعدد Multiple R	٠,٨٧٥٤٤٢
معامل التحديد R Square	٠,٧٦٦٣٩٨
معامل التحديد المعدلAdjusted R Square	٠,٧٣٧١٩٨
الخطاء المعياري Standard Error	٦,٨٣٧١٥
المشاهدات Observations	١٠

ANONA

	df	*SS*	*MS*	*F*	معنوية F
الانحدار	١	١٢٢٦,٩٢٧	١٢٢٦,٩٢٧	٢٦,٢٤٦٣٣	٠,٠٠٠٩٠٤
البواقي	٨	٣٧٣,٩٧٣	٤٦,٧٤٦٦٢		
الإجمالي	٩	١٦٠٠,٩			

	المعاملات	الخطاء المعياري	إحصاءة t	قيمة *P*	الحد الأدنى ٩٥٪	الحد الأعلى ٩٥٪
القاطع	٤٦,٤٨٦٤٩	٩,٨٨٤٥٦٦	٤,٧٠٢٩٣٦	٠,٠٠١٥٣٦	٢٣,٦٩٢٦٢	٦٩,٢٨٠٣٥
X المتغير ١	٥٢,٥٦٧٥٧	١٠,٢٦٠٨٦	٥,١٢٣١١٧	٠,٠٠٠٩٠٤	٢٨,٩٠٥٩٧	٧٦,٢٢٩١٦

الانحدار المتعدد A Multiple Regression Problem

المثال الخاص في الجزء المتعلق بالانحدار المتعدد للبيانات التي تم عرضها في الجدول رقم (١١,٦) يمكن حلها بسهوله باستخدام برنامج اكسل. والإجراء المستخدم هو نفسه الذي تم استخدامه للانحدار البسيط والنتيجة المتحصل عليها من ذلك موضحه بالجدول رقم (١١,١٣).

ويجب ملاحظة أن جدول تحليل التباين يحتوي على درجتي حرية للانحدار نظراً لأن لدينا متغيرين لـ X في النموذج. ماعدا ذلك فإن النتائج مشابه في الشكل لتلك المتحصل عليها لنموذج الانحدار الخطي البسيط.

الجدول رقم (١١.١٣). نتائج الحاسب للانحدار المتعدد للنموذج الخطي لإنتاجية القطن.

خلاصة النتائج	
إحصاءات الاختبار	
Multiple R المتعدد R	٠,٩٨٦٥٨٨
معامل التحديد R Square	٠,٩٧٣٣٥٥
معامل التحديد المعدل Adjusted R Square	٠,٩٦٧٤٣٤
الخطاء المعياري Standard Error	١٢,٠٥٧٩٤
المشاهدات Observations	١٢

ANONA

	df	SS	MS	F	معنوية F
الانحدار	٢	٤٧٨٠٢,٣٧	٢٣٩٠١,١٩	١٦٤,٣٨٩٢	$٨,٢٣ \times ١٠^{-٨}$
البواقي	٩	١٣٠٨,٥٤٥	١٤٥,٣٩٣٩		
الإجمالي	١١	٤٩١١٠,٩٢			

تابع الجدول رقم (١١.١٣).

	المعاملات	الخطـــــــاء المعياري	إحصاءة t	قيمة *P*	الحد الأدنى ٩٥٪	الحد الأعلى ٩٥٪
القاطع	٢٤٨,٠٦٣	١٥,٢٩٥٣٤	١٦,٢١٨٢	$٥,٧٢ \times ١٠^{-٨}$	٢١٣,٤٦٢٥	٢٨٢,٦٦٣٥
X المتغير ١	١٠,٣١١٦٤	٢,٦٥٥٥٥٥	٣,٨٨٣٠٤٥	٠,٠٠٣٧١٤	٤,٣٠٤٣٥٢	١٦,٣١٨٩٣
X المتغير ٢	٤,٢٨٥٤٧٧	٠,٦٨٣١٢٧	٦,٢٧٣٣٢٣	٠,٠٠٠١٤٦	٢,٧٤٠١٣٥	٥,٨٣٠٨٢

ويلاحظ من بيانات البواقي بأنها لا تحتوي على أي اتجاه معين حيث أنها تتوزع حول خط الصفر.

تابع الجدول رقم (١١.١٣).

نتائج البواقي		
المشاهدة	**Y المقدرة**	**البواقي**
١	٤٢٨,١٨٥٥	-٨,١٨٥٥٣
٢	٤٦٩,٤٣٢١	-١١,٤٣٢١
٣	٣٩٦,٤٤٦٥	٣,٥٥٣٤٩٤
٤	٥١٢,٢٨٦٩	-٢,٢٨٦٨٧
٥	٥٢٢,٥٩٨٥	-٧,٤٠١٤٩
٦	٤٧٠,٢٣٦٢	٩,٧٦٣٧٩٧
٧	٥٨٦,٠٧٦٦	-١١,٠٧٦٦
٨	٤٥٩,٩٢٤٦	٠,٠٧٥٤٣٨
٩	٥٣٣,٧١٤٣	-٣,٧١٤٢٦
١٠	٦٠٧,٥٠٤	٢,٤٩٦٠٤٤
١١	٥٨٦,٨٨٠٧	٣,١١٩٣٢٦
١٢	٥٣٣,٧١٤٣	-٩,٧١٤٢٦

تمارين Exercises

١- ما الفرق بين تحليل الانحدار والارتباط في اختبار قيم X ؟ هل يمكن تحويل مسألة الارتباط إلى انحدار أو العكس؟ ناقش ذلك.

٢- ما هو " الخطأ المعياري للتقدير" ؟ هل هو معلمة أم إحصاءة ؟ ناقش ذلك.

٣- ماذا تمثل النقاط على شكل الانتشار ؟ ماذا تعني النقطة على خط الانحدار باستخدام المربعات الصغرى للعينة ؟ ناقش ذلك.

٤- البيانات التالية توضح صافي الدخل المزرعي القومي، X ، وإجمالي الأصول غير العقارية للمزرعة Y بالبليون دولار للفترة من ١٩٨٨ حتى ١٩٩٦. أحسب معامل الارتباط r ثم اختبر مدى اختلافه عن الصفر باستخدام مستوى معنوية ٥٪. ثم فسر ماذا يعني هذا المعامل في حالة هذا المثال.

السنة	X صافي الدخل المزرعي (بالبليون دولار)	Y إجمالي الأصول غير العقارية للمزرعة (بالبليون دولار)
١٩٨٨	٣٨	٢٠٤
١٩٨٩	٤٥	٢١٢
١٩٩٠	٤٤	٢٢٠
١٩٩١	٣٨	٢١٥
١٩٩٢	٤٧	٢٢٦
١٩٩٣	٤٣	٢٣١
١٩٩٤	٤٨	٢٣٧
١٩٩٥	٣٦	٢٢١
١٩٩٦	٥٢	٢٢٧

٥- تستخدم المساحة المزروعة بالايكر من الأرز X (بالمليون) للتنبؤ بالإنتاج الكلي Y (بالمليون وزن مئوي) في تحليل الانحدار. احسب خط الانحدار بالمربعات الصغرى لهذه البيانات.

X المساحة المزروعة (بالمليون)	Y الإنتاج (بالمليون وزن متوي)
٣,٢	١٨٠
٢,٩	١٥٦
٣,٤	١٩٨
٣,١	١٧٤
٢,٨	١٧١
٣,١	١٨٠

أ) اختبر ما إذا كان معامل الميل يختلف عن الصفر باستخدام تحليل التباين ومستوى معنوية ٥ ٪. ثم فسر معنى الميل.

ب) احسب معامل التحديد ثم فسر معناه.

ج) تنبأ بالإنتاج عندما تكون المساحة المزروعة ٣ مليون أيكر باستخدام المعادلة لخط الانحدار. كوّن فتره ٩٥٪ لمتوسط الإنتاج المقدّر. ثم فسر معنى فتره الثقة المتحصل عليها.

د) ارسم شكل الانتشار لهذه البيانات ثم ارسم عليه خط الانحدار للمربعات الصغرى.

٦- البيانات التالية توضح العلاقة بين الدخل الشخصي للفرد X (بالألف دولار لعام ١٩٩٢) ومتوسط الاستهلاك الفردي من لحوم البقر Y (بالرطل).

X الدخل (بالألف)	Yاستهلاك لحوم البقر (بالرطل)
١٧,٦	٦٩
١٧,٨	٦٥
١٧,٩	٦٤
١٧,٨	٦٣
١٨,١	٦٢
١٨,١	٦١

أ) احسب خط الانحدار بالمربعات الصغرى للمتغير Y على X .

ب) استخدم اختبار t لتحديد ما إذا كان ميل الخط يختلف عن الصفر عند مستوى معنوية ٥٪.

ج) قدر استهلاك لحوم البقر عندما يكون الدخل الفردي ١٨ ألف دولار. أوجد فتره ثقه ٩٥٪ لهذا التقدير ثم فسر معناه.

د) احسب معامل التحديد r^2 ثم فسر معناه.

هـ) ارسم شكل الانتشار لهذه البيانات ثم ارسم خط الانحدار بالمربعات الصغرى عليه.

٧- استخدم برنامج اكسل لحل التمرين ٤ .

٨- استخدم برنامج اكسل لحل التمرين ٥ .

٩- استخدم برنامج اكسل لحل التمرين ٦ .

١٠- تحصل مهندس زراعي على بيانات تجربة لإنتاجية القطن للايكر Y (بالرطل من الوبر) وكمية سماد النتروجين X_1 (بالرطل) وكمية مياه الري المستخدمة X_2 (انش/ايكر) . قدر خط الانحدار بالمربعات الصغرى باستخدام برنامج اكسل.

Y إنتاجية القطن (بالرطل)	X_1 النتروجين (بالرطل)	X_3 الري (انش– ايكر)
٢٥٠	٠	٠
٣٢٠	١٠	٠
٣٧٥	١٥	٣
٤٠٠	٢٠	٣
٤١٠	٢٥	٣
٤٤٠	٢٥	٦

تابع

Y إنتاجية القطن (بالرطل)	X_1 النتروجين (بالرطل)	X_3 الري (انش- ايكر)
٤٥٠	٣٠	٦
٤٧٠	٣٥	٦
٥١٠	٣٥	٩
٥٣٠	٤٠	٩
٥٣٢	٤٥	٩

أ) أوجد معادلة الانحدار ثم تحقق من جودة النموذج ككل باستخدام مستوى معنوية ٥٪ $\alpha = 0.05$.

ب) اختبر معنوية معامل الميل باستخدام مستوى معنوية ٥٪.

ج) أوجد R^2 ثم فسر معناه. ثم اوجد تقدير الخطأ المعياري.

١١- ترغب مؤسسة تصنيع زراعية صغيرة في تقدير دالة التكاليف الكلية الأسبوعية. تحصل المدير على البيانات المحاسبية من المؤسسة واكتشف العلاقة التالية:

Y إجمالي التكاليف (دولار)	X الإنتاج الأسبوعي
٤٨٠٠	٠
٤٨١٥	٥
٤٨٣٠	١٠
٤٨٥٠	١٥
٤٨٧٥	٢٠
٤٩٠٠	٢٥
٤٩٢٥	٣٠
٤٩٥٠	٣٥
٥٠٠٠	٤٠
٥٠٦٠	٤٥

قدر معادلة التكاليف الكلية باستخدام معادلة الانحدار المتعدد التي تحتوي على X ، X^2 ثم أجب على الأجزاء من أ إلى ج الواردة في التمرين ١٠ آنفاً لهذه المســألة.

١٢- ترتبط مبيعات بذور الذرة بعدد أيام الإعلان في الراديو والتلفزيون في المطاعم للمزارعين وكذلك القيمة المنفقة على الإعلان في المجتمعات، كما في الجدول التالي:

Y مبيعات بذور الذرة (١٠٠٠ دولار)	X_1 عدد الأيام	X_2 القيمة المنفقة (١٠٠٠ دولار)
٢٣٥	٨	١٦
٢٤٥	١٠	٢٠
٢٦٠	١٢	٢٥
٢٧٠	١٤	٣٠
٢٥٥	١١	٢٢
٢٦٥	١٣	٢٦
٢٧٥	١٥	٣٢
٢٤٠	٩	١٨
٢٣٣	٧	١٥

أ) اختبر ما إذا كانت معاملات الميل تختلف عن الصفر باستخدام مستوى معنوية ٥٪.

ب) حدد معامل التحديد ثم فسره.

ج) تنبأ بالمبيعات إذا استمر الإعلان في المجتمعات ١٠ أيام وكانت تكلفته ٢٤٠٠٠ دولار.

د) ما هي المشكلة التي تتوقع وجودها في هذه البيانات؟ وهل هناك أي دليل لإثبات ذلك.

١٣- مجموعه عجول أوزانها في الفئة ٤٠٠-٦٠٠رطل (x_3) تم رعيها في أماكن العشب والرعي خلال الربيع والخريف والصيف (X_1 : ١، ٢، ٣ على التوالي)، مع إعطائها حبوب حسب وزن الجسم (x_2 = نسبة وزن الجسم) وتم قياس الوزن المكتسب بالرطل Y . احسب العلاقة بين الوزن المكتسب Y والمتغيرات الأخرى مستخدماً نموذج الانحدار المتعدد.

Y الوزن المكتسب رطل/يوم	X_1 الموسم	X_2 الحبوب	X_3 وزن الجسم
٢,٥	١	١,٠	٥٩٠
٢,٠	١	١.٠	٤١٥
٢,١	١	٠,٧٥	٥٣٠
١,٠	٣	١,٠	٥٧٠
٠,٨	٣	٠,٧٥	٥٢٠
٠,٥	٣	٠,٥	٤٢٥
١.٥	٢	٠.٧٥	٥٤٠
١,٨	٢	١,٠	٤٠٥
٢,٠	٢	١,٠	٥٦٢

أ) اختبر ما إذا كانت معاملات الميل تختلف عن الصفر باستخدام مستوى معنوية ٥٪ ثم فسر تلك النتائج بناء على المسألة.

ب) احسب R^2 ثم فسر معناها.

ج) تنبأ بالوزن بالرطل لليوم لعجل وزنه ٥١٠ تمت تغذيته بحبوب مساوية لـ ١٪ من وزن جسمه وتم رعيه في فصل الصيف.

١٤- استخدم طريقه المربعات الصغرى لتقدير الاتجاه الخطي للأسعار القياسية لمنتجات المزرعة (١٩٨٢- ١٠٠) لفترة من ١٩٨٠ إلى ١٩٩٥ .

السنة	مؤشر منتجات المزرعة
١٩٨٠	١٠٢,٩
١٩٨١	١٠٥,٢
١٩٨٢	١٠٠,٠
١٩٨٣	١٠٢,٤
١٩٨٤	١٠٥,٥
١٩٨٥	١٠٦,١
١٩٨٦	١٠٥,٢
١٩٨٧	١٠٨,٣
١٩٨٨	١٠٩,٩
١٩٨٩	١٠٨,٩
١٩٩٠	١١٢,٢
١٩٩١	١١١,٧
١٩٩٢	١١٤,٦
١٩٩٣	١١٢,١
١٩٩٤	١١٦,٣
١٩٩٥	١١٧,٤

١٥- استخدم الانحدار الخطي في برنامج آكسل لتقدير خط الاتجاه للبيانات في التمرين رقم (١٤). ثم قارن كلا التقديرين. ما هو انطباعك حول هذا الاتجاه. ثم

استخدم خط الانحدار لتقدير مؤشر سعر منتجات المزرعة لعام ١٩٩٧.

١٦- يوضح الجدول التالي ديون الرهن العقاري المستحقة على الملكية الزراعية بالبليون للفترة من ١٩٨٠-١٩٩٣. ارسم خط الاتجاه للبيانات باستخدام المنحنى الاسي والخط المستقيم. حدد ما هو الأفضل لتمثيل البيانات ؟ ثم تنبأ بالديون المستحقة في ١٩٩٥.

السنة	ديون الرهن العقاري (بالبليون دولار)
١٩٨٠	٣٠,٥
١٩٨١	٣٢,٤
١٩٨٢	٣٥,٤
١٩٨٣	٣٩,٨
١٩٨٤	٤٤,٩
١٩٨٥	٤٩,٩
١٩٨٦	٥٥,٤
١٩٨٧	٦٣,٩
١٩٨٨	٧٢,٨
١٩٨٩	٨٦,٨
١٩٩٠	٩٧,٥
١٩٩١	١٠٧,٢
١٩٩٢	١١٨,٣
١٩٩٣	١٣٣,٧

١٧- معادلة الاتجاه التالية ناتجة من رسم القطع المكافئ بتحليل الانحدار لحجم قوة العمل الزراعي في محافظه معينه.

$$Y = 49.17 + 4.23X - 0.19X^2$$
$$X = 0 \text{ in } 1975$$

X تمثل فتره ٢,٥ سنة.

Y حجم قوة العمل الزراعية بالألف.

أ) إذا افترضنا أن خط الاتجاه مثل البيانات بكفاءة، ما هو التعميم الممكن قوله حيال الطريقة التي تنمو بها قوة العمل في هذه المحافظة بالنسبة المئوية ؟ بالقيمة المطلقة؟.

ب) في الجزء (أ) افترضنا أن الخط مثّل البيانات بكفاءة. ولكن الحجم الفعلي لقوة العمل الزراعية كانت ٩٢٠٧٣ في عام ٢٠٠٠م. وهذا الرقم يشير إلى أن الخط لم يمثل البيانات جيداً. هل توافق على ذلك ؟ ناقش.

١٨- تنبأ أحد المديرين بشركتك بأن مبيعات الشركة للسنة القادمة تساوي ١٢ مليون دولار باستخدام خط الاتجاه. ونتوقع عدم وجود تغيرات دوريه حادة السنة القادمة ونعتقد أن أثر الاتجاه خلال السنة سوف يكون بسيط ونتوقع استمرار النمط الموسمي في المبيعات. وذلك النمط كالتالي :

الربع	I	II	III	IV
مؤشر الموسمية	٩٠	٧٥	١٣٠	١٠٥

المطلوب التنبؤ بالمبيعات للربع الثاني والثالث للسنة القادمة.

١٩- مؤشر التغيرات الموسمية في إنتاج لحوم البقر لبعض الأشهر المختارة هو ٦٠ ليناير، ٧٠ لمارس ، ١٢٢لأغسطس ، ١٠٠لسبتمبر.

أ) في أي هذه الأشهر كان إنتاج لحوم البقر الأعلى.

ب) في أي هذه الأشهر كان الإنتاج تقريباً هو المتوسط الشهري المعتاد.

ج) زاد إنتاج لحوم الأبقار في الجنوب الغربي من ٦,٠٦ مليون رطل في يناير إلى ١١,٥٩ مليون رطل في أغسطس. ما هي نسبة التغير في الإنتاج بعد استبعاد المتغيرات الموسمية.

٢٠- مسلخ تجاري لصغار الدواجن عرض البيانات الشهرية التالية. (بالمليون رأس).

الشهر	السنة			
	١٩٩٧	١٩٩٨	١٩٩٩	٢٠٠٠
يناير	٣٣	٣٦	٤٠	٤٢
فبراير	٣٠	٣١	٣٧	٣٦
مارس	٣٥	٣٧	٣٧	٣٨
ابريل	٣٢	٣١	٣٨	٤١
مايو	٣٣	٣٩	٣٩	٤٢
يونيو	٣٥	٤١	٣٥	٣٩
يوليو	٣٤	٣٦	٤١	٤٢
أغسطس	٣٨	٣٨	٣٩	٤١
سبتمبر	٣٧	٣٦	٣٦	٤١
أكتوبر	٣٥	٣٨	٤١	٤٥
نوفمبر	٣٤	٣٥	٣٩	٣٧
ديسمبر	٣٢	٣٤	٣٧	٤٢

أ) ارسم هذه البيانات مع الزمن.

ب) احسب الاتجاه الزمني ومؤشر الموسمية الشهري باستخدام ١٢ شهر متوسط متحرك مركزه الشهر السادس.