الفصل التاسع

الانحدار الخطى البسيط

SIMPLE LINEAR REGRESSION

(٩ – ١) المجتمع ذو المتغيرين : BIVARIATE POPULATION

المجتمع ذو المتغيرين هو مجتمع ننظر إليه من حيث متغيرين س ، ص يكونان موضع اهتمامنا في وقت ما ، فمثلا في مجتمع من الطلاب قد نهتم بالمتغير س الذى يعبر عن نسبة الذكاء والمتغير ص الذى يعبر عن درجة الطالب في مادة الإحصاء وفي مجتمع من القمح قد تعبر س عن تاريخ الزراعة وتعبر ص عن مقدار المحصول الناتج ، وفي مجتمع من الأبقار قد تعبر س عن مقدار الغذاء اليومى وتعبر ص عن الزيادة في الوزن بعد مدة من الزمن .

كما سبق الذكر ، يميز الإحصاء بين نوعين من المتغيرات : المتغيرات العشوائية والمتغيرات غير العشوائية ، والمتغير العشوائي هو متغير حقيقي يخضع لمؤثرات عشوائية غير منظورة ولذلك لا نستطيع التحكم فيه تجريبياً . وهذا النوع من المتغير يكون له توزيع احتمال بمعني أنه لو كان من النوع الوثاب مثلا فإن ظهور أى عنصر من عناصره يكون مصحوباً باحتمال ما ، أما المتغير غير العشوائي (أو الرياضي) فليس له توزيع احتمال ويمكن التحكم فيه تجريباً أو تحديد قيمه مقدماً أو قياسها بدقة أو بخطأ يمكن إهماله .

وفي دراستنا لمجتمع ذى متغيرين س ، ص كثيراً ما يكون اهتمامنا منصباً على إيجاد العلاقة بينهما إن كان هناك ثمة علاقة ، وعلى محاولة وضع هذه العلاقة على

هيئة دالة ص = د (س) . وتختلف هذه الدالة بطبيعة الحال باختلاف العلاقة بين المتغيرين فقد تكون على صورة خطية ص = $\alpha + \beta$ س أو على صورة تربيعية ص = $\alpha + \beta$ س + لا $س^2$ أو على صورة أسية صـ = α هـ$^{\beta س}$ وهكذا ... إلا أننا سوف نقصر اهتمامنا هنا على الحالات التي تكون فيها الدالة على الصورة الخطية .

سنبني هذه الدراسة على الافتراضين الآتيين وافتراض ثالث نقدمه في البند (٩ - ٥) ، والنموذج الذى سنستخدمه فى هذه الدراسة يسمى بالنموذج I لتحليل الانحدار .

الافتراض الأول :

« المتغير س هو متغير رياضي والمتغير صـ هو متغير عشوائي » .

وهذاالافتراض يعنى من الناحية التطبيقية أننا نبدأ بتحديد قيم ثابتة $س_1$ ، $س_2$ ، ... ، $س_ن$ من المتغير س ثم نقوم بملاحظة أو قياس القيم (العشوائية) المناظرة $ص_1$ ، $ص_2$ ، ... ، $ص_ن$ من المتغير صـ . فمثلا قد تكون القيم السينية هى درجات حرارة معينة ١٠ ، ١٥ ، ٢٠ ، ... ، ٥٠ وتكون القيم الصادية هى مقادير ما نشاهده من تمدد معدن عند هذه الدرجات . أو تكون قيم س هى أطوال أطفال عند الولادة بالسنتمترات وتكون قيم صـ هى مدد الحمل بالأيام . أو تكون قيم س هى أعماق محددة تحت سطح البحر وتكون القيم الصادية هى نسب الملوحة عند هذه الأعماق .

الافتراض الثانى :

« العلاقة بين المتغيرين س ، صـ (في المجتمع) هى علاقة خطية » .

ونعني بهذا الفرض رياضياً أن يكون متوسط قيمة ص عند قيمة معينة س يأخذ الصيغة الخطية الآتية :

$$\mu_{ص/س} = \alpha + \beta س \qquad (١)$$

حيث α ، β بارامتران مجهولان ينبغى تقديرهما من العينة . هذا مع ملاحظة أنه عند أى قيمة س يمكن أن تأخذ ص قيماً كثيرة لأنها كما ذكرنا تتأثر بعوامل عشوائية لا نستطيع التحكم فيها ولهذا يعبر الطرف الأيمن من الصيغة (١) عن متوسط هذه القيم وهو بالطبع أحسن قيمة تناظر القيمة س . ويسمى المتغير س بالمتغير المستقل كما يسمى المتغير ص بالمتغير التابع .

إن هذا الافتراض لا يوضع إلا في الحالات التي نعلم بخبرتنا السابقة أو من معلومات خاصة أنه صحيح . ومع ذلك إذا كنا نشك في صحته فهناك طريقة إحصائية سنذكرها بعد للتحقق من سلامته .

إن مهمتنا الابتدائية في دراسة العلاقة الخطية بين متغيرين س ، ص هى إيجاد أحسن تقديرين للبارامترين المجهولين α ، β الموجودين بالعلاقة (١) ، ونعتمد في ذلك كالمعتاد على تصميم تجربة على عينة من المجتمع الذى ندرسه ومنها نحصل على قيم تجريبية للمتغيرين س ، ص تبدو كما يلى :

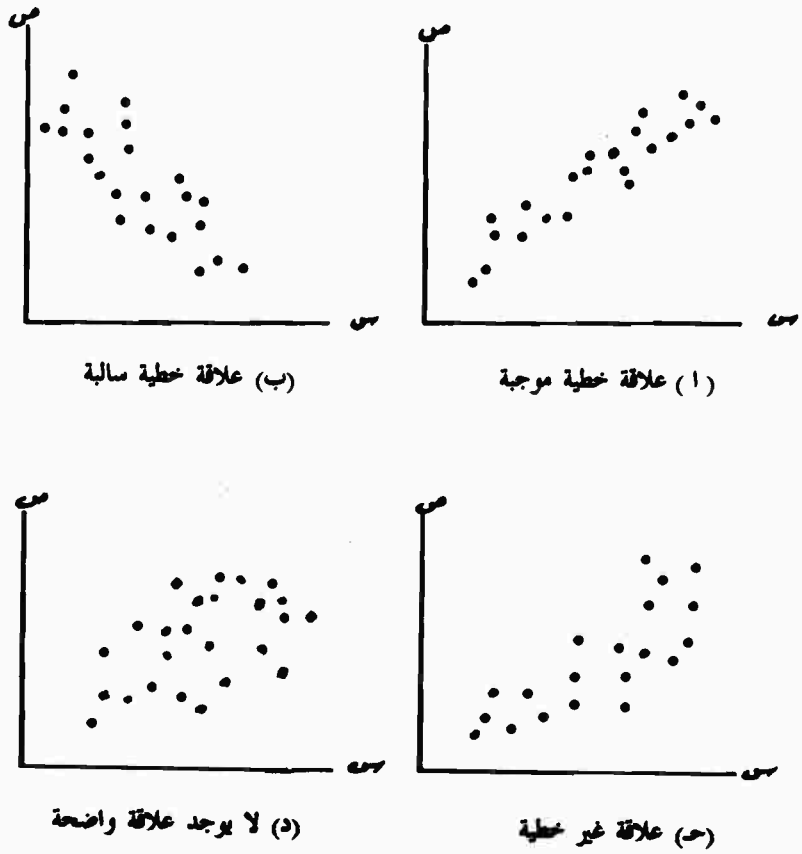
$س_1$	$س_2$	$س_3$	...	$س_ن$
$ص_1$	$ص_2$	$ص_3$	...	$ص_ن$

ومن هذه الأزواج من القيم نوجد التقديرين المطلوبين كما سنرى بعد ، ويساعدنا التمثيل البياني لهذه الأزواج على تصور المشكلة التي نتناولها والخط الذى ننشده .

(٩ - ٢) شكل الانتشار : SCATTERGRAM

على أساس أن المتغيرين س ، ص حقيقيان نستطيع تمثيل أزواج قيم $(س_ر ، ص_ر)$ التي حصلنا عليها من العينة على نظام إحداثيات ذى محورين متعامدين . فيمثل كل زوج منها بنقطة معينة في المستوى ، ومجموعة النقط الناتجة تؤلف شكلا يسمى بشكل الانتشار . وإذا كان هناك ثمة علاقة بين المتغيرين فإن هذه النقط لا تكون مبعثرة كيفما اتفق بل تتخذ نمطاً معيناً يوحى بوجود وطبيعة هذه

العلاقة . فإذا بدا شكل الانتشار كما في أى من الشكلين (٩ – ١ – ا) أو (٩ – ١ – ب) فإنه يشير بصفة مبدئية بأن العلاقة بين المتغيرين هى علاقة خطية ، لأننا نستطيع أن نتصور وجود خط مستقيم تقع النقط من حوله وقريبة منه وإن كانت لا تمر جميعها به .



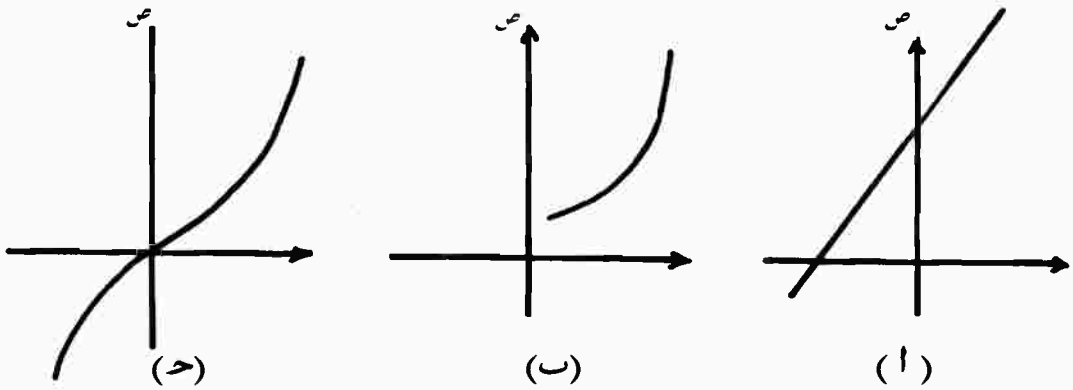
(ا) علاقة خطية موجبة (ب) علاقة خطية سالبة

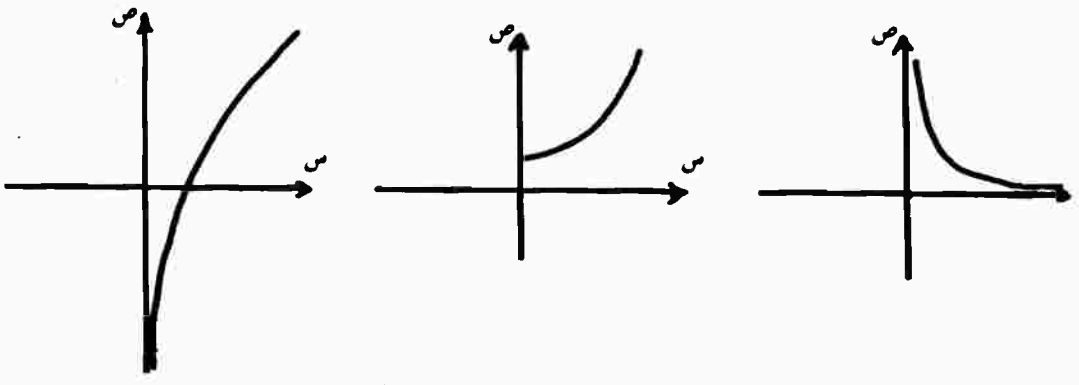
(حـ) علاقة غير خطية (د) لا يوجد علاقة واضحة

الشكل (٩ – ١) بعض صور شكل الانتشار

أما إذا بدا شكل الانتشار كما فى أى من الشكلين (حـ) ، (د) فإننا نشك فى خطية العلاقة بين المتغيرين وعلينا أن نبحث عن افتراض آخر غير افتراض الخطية نتعامل به مع المتغيرين . إن معرفة الباحث بخواص المنحنيات تساعده على كشف طبيعة

العلاقة بين المتغيرين واقتراح المعادلة التى تناسبها . وفيما يلى بعض المنحنيات التى تعاون على كشف الأنماط التى قد يشير إليها شكل الانتشار والدوال التى تعبر عنها جبريا . وسنرى فى البند (٩ – ٧) أنه يمكن تناول بعض هذه الدوال كدوال خطية بعد إجراء تحويلات مناسبة عليها .





الشكل (٩ – ٢) : بعض أنماط شكل الانتشار

(٩ - ٣) تقدير البارامترين α ، β .

نفرض أن ($س_ر$ ، $ص_ر$) هى إحدى القيم المشاهدة في العينة . حسب الافتراض الثاني ينبغى أن تكون العلاقة بين $س_ر$ ، $ص_ر$ على الصورة :

$$ص_ر = \alpha + \beta\ س_ر + خ_ر \qquad (٢)$$

حيث $خ_ر$ هو الفرق بين القيمة المشاهدة $ص_ر$ المناظرة للقيمة $س_ر$ وبين القيمة المتوقعة لها وهى $\alpha + \beta\ س_ر$ ولذلك تسمى القيم $خ_ر$ بالبواقي residuals وهى تعبر عن الأخطاء العشوائية في قياس المتغير العشوائي ص . ونظراً لأن هذه الأخطاء تكون بالزيادة لبعض قيم ص وبالنقصان للبعض الآخر ، فإننا نفترض أن متوسط هذه الأخطاء يؤول إلى الصفر على المدى البعيد .

طريقة المربعات الصغرى : LEAST SQUARES METHOD

كما أشرنا من قبل ، إن مهمتنا الأساسية هى استخدام أزواج القيم ($س_ر$ ، $ص_ر$) المشاهدة في العينة لإيجاد أحسن تقديرين للبارامترين α ، β . فإذا كان ا ، ب هما هذان التقديران فإن المعادلة :

$$\hat{ص} = ا + ب\ س \qquad (٣)$$

تكون معبرة أحسن تعبير عن المعادلة (١) ويكون الخط الممثل للمعادلة (٣) هو أحسن خط يلائم مجموعة نقط العينة . فإذا وفقنا في إيجاد ا ، ب سمى الخط (٣) بخط أحسن مطابقة line of best fit أو خط انحدار ص على س regression line of Y on X وأحياناً يسمى بصيغة التنبؤ prediction formula .

والطريقة الأكثر شيوعاً لإيجاد ا ، ب تبني على مبدأ يعرف بمبدأ المربعات الصغرى يمكن صياغته كالآتي ، مع ملاحظة الافتراض الأول عن أن قيم س ثابتة وقيم ص متغيرة .

« أحسن خط يلائم مجموعة النقط ($س_ر$ ، $ص_ر$) هو ذلك الذى نحدده بحيث يكون مجموع مربعات البواقي $خ_ر$ أصغر ما يمكن » .

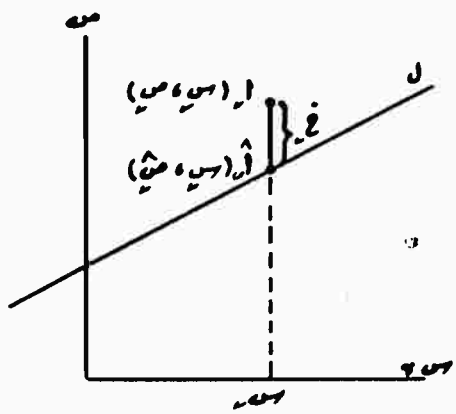
من المعادلة (٢) نرى أن $خ_ر = ص_ر - \alpha - \beta\ س_ر$

ليكن د (α ، β) = مـ $خ_ر^2$ = مـ ($ص_ر - \alpha - \beta\ س_ر$)^2 (٤)

فإذا نظرنا للدالة (٤) على أنها دالة في المجهولين α ، β فإن مبدأ المربعات الصغرى يقول إن أحسن خط هو ذلك الذى تحدد فيه قيمتاً α ، β بحيث تكون قيمة هذه الدالة نهاية صغرى .

ويوضح هذا المبدأ هندسياً كالآتي :

لتكن $ا_ر$ ($س_ر$ ، $ص_ر$) إحدى النقط المشاهدة في العينة أى إحدى نقط شكل الانتشار ، ولتكن $\hat{ا}_ر$ ($س_ر$ ، $\hat{ص}_ر$) هى نقطة واقعة على خط مستقيم ل وتشترك مع النقطة $ا_ر$ في الإحداثي السيني $س_ر$.



الشكل (٩ - ٣)

إن مبدأ المربعات الصغرى يقول إن الخط ل يكون هو خط أحسن مطابقة إذا كان مجموع مربعات المسافات الرأسية بين النقط $ا_ر$ ، $\hat{ا}_ر$ نهاية صغرى ، أى إذا كانت الدالة

د (α ، β) = مـ ($ص_ر - \hat{ص}_ر$)^2 = مـ ($ص_ر - \alpha - \beta\ س_ر$)^2

نهاية صغرى . بطرق التفاضل المعتادة نفاضل الدالة (٤) جزئياً بالنسبة إلى كل من α ، β . وبمساواة كل من الناتجين بالصفر نحصل على معادلتين آنيتين يكون

حلهما معاً هما القيمتان ا ، ب المطلوبتان. وهاتان المعادلتان تسميان بالمعادلتين المعتادتين وتأخذان الصورتين الآتيتين :

$$\text{مجـ ص} = \text{ن } \alpha + \beta \text{ مجـ س} \quad ، \quad \text{مجـ س ص} = \alpha \text{ مجـ س} + \beta \text{ مجـ س}^{٢}$$

وبحلهما معا ينتج ما يلى :

أحسن تقديرين للبارامترين α ، β من العينة هما ا ، ب حيث :

$$\text{ب} = \frac{\text{ن مجـ س ص} - (\text{مجـ س})(\text{مجـ ص})}{\text{ن مجـ س}^{٢} - (\text{مجـ س})^{٢}} \qquad (٥)$$

وحيث ن هى عدد أزواج القيم $(\text{س}_{\text{ر}} ، \text{ص}_{\text{ر}})$

$$\text{ا} = \bar{\text{ص}} - \text{ب } \bar{\text{س}} \quad ، \qquad (٦)$$

وبالتالى تكون معادلة انحدار ص على س هى

$$\hat{\text{ص}} = (\bar{\text{ص}} - \text{ب } \bar{\text{س}}) + \text{ب س}$$

$$\text{أو } \hat{\text{ص}} = \bar{\text{ص}} + \text{ب } (\text{س} - \bar{\text{س}}) \qquad (٧)$$

حيث ب معطاة بالصيغة (٥) وتسمى بمعامل انحدار ص على س وهى تعبر عن ميل خط الانحدار (٧) عن الاتجاه الموجب لمحور السينات .

ملاحظة (١) :

من الواضح أن خط الانحدار (٧) يمر بالنقطة $(\bar{\text{س}} ، \bar{\text{ص}})$.

ملاحظة (٢) :

بقسمة كل من بسط ومقام الطرف الأيسر من (٥) على (ن − ١) وبعد عمليات جبرية بسيطة نجد أن :

$$ب = \frac{\frac{1}{ن-1} \text{مح} (س - \bar{س})(ص - \bar{ص})}{\frac{1}{ن-1} \text{مح} (س - \bar{س})^2} = \frac{ع_{سص}}{ع^2_{س}} = \frac{\text{تغاير} (س ، ص)}{\text{تباين } س} \quad (٨)$$

وفى إيجاد ب من (٥) أو (٨) يمكن أن نجمع أو نطرح أى عدد من جميع قيم س وأى عدد من جميع قيم ص دون أن تتأثر قيمة ب . وفى حالة التوزيعات التكرارية حيث يكون لكل زوج ($س_ر$ ، $ص_ر$) تكرار $ك_ر$ نحصل على ب من أى من الصيغتين بوضع $ك_ر$ بعد كل رمز مح .

مثال (٩ - ١) :

أوجد معادلة انحدار ص على س من البيانات الآتية :

س :	١,٥	١,٨	٢,٤	٣,٠	٣,٥	٣,٩	٤,٤	٤,٨	٥,٠
ص :	٤,٨	٥,٧	٧,٠	٨,٣	١٠,٩	١٢,٤	١٣,١	١٣,٦	١٥,٣

ومنها أوجد أحسن تقدير لقيمة ص عندما س = ٣,٧ .

الحل :

(يلاحظ أننا لو رسمنا شكل الانتشار لوجدنا أن افتراض الخطية هو افتراض معقول) . يتطلب الحل إيجاد كل من مح س ، مح ص ، مح $س^2$ ، مح س ص وهذه جميعاً نحصل عليها من الجدول (٩ - ١) الآتى .

لاحظ أن العمود الأخير لا لزوم له في إيجاد معادلة انحدار ص على س ولكنه سيلزمنا فيما بعد في تحليل الانحدار . ويمكن استخدام حاسبات الجيب للحصول على هذه المجاميع دون ضرورة لكتابة التفاصيل المبينة بالجدول وذلك توفيرا للوقت والجهد .

جدول (٩ – ١)

إيجاد معادلة خط الانحدار من بيانات المثال (٩ – ١)

س	ص	س ص	$س^{٢}$	$ص^{٢}$
١,٥	٤,٨	٧,٢٠	٢,٢٥	٢٣,٠٤
١,٨	٥,٧	١٠,٢٦	٣,٢٤	٣٢,٤٩
٢,٤	٧,٠	١٦,٨٠	٥,٧٦	٤٩,٠٠
٣,٠	٨,٣	٢٤,٩٠	٩,٠٠	٦٨,٨٩
٣,٥	١٠,٩	٣٨,١٥	١٢,٢٥	١١٨,٨١
٣,٩	١٢,٤	٤٨,٣٦	١٥,٢١	١٥٣,٧٦
٤,٤	١٣,١	٥٧,٦٤	١٩,٣٦	١٧١,٦١
٤,٨	١٣,٦	٦٥,٢٨	٢٣,٠٤	١٨٤,٩٦
٥,٠	١٥,٣	٧٦,٥٠	٢٥,٠٠	٢٣٤,٠٩
٣٠,٣	٩١,١	٣٤٥,٠٩	١١٥,١١	١٠٣٦,٦٥

من الصيغة (٥) نجد أن ب $= \frac{٩١,١ \times ٣٠,٣ - ٣٤٥,٠٩ \times ٩}{(٣٠,٣)^{٢} - ١١٥,١١ \times ٩} = ٢,٩٣٠٣$

كما أن $\bar{س} = \frac{٣٠,٣}{٩} = ٣,٣٦٦٧$ ، $\bar{ص} = \frac{٩١,١}{٩} = ١٠,١٢٢٢$

∴ معادلة انحدار ص على س (من الصيغة ٧) هى

$\hat{ص} = ١٠,١٢٢٢ + ٢,٩٣٠٣ (س - ٣,٣٦٦٧)$

ومنها ص̂ = ٠,٢٥٦٨ + ٢,٩٣٠٣ س

وحينما س = ٣,٧ فإن أحسن تقدير لقيمة ص هى :

ص̂ = ٠,٢٥٦٨ + ٢,٩٣٠٣ × ٣,٧ = ١١,٠٩٨٩

(٩ - ٤) الخطأ المعيارى للتقدير Standard Error of Estimate

كما هو الحال عند دراسة بيانات عن متغير واحد حيث نصف هذه البيانات بواسطة الوسط الحسابي الذى يعطى تقديراً لمتوسط هذه البيانات ثم ندعم هذا الوصف بتقدير التشتت بواسطة الانحراف المعيارى ، فاننا نصف البيانات ذوات المتغيرين بواسطة خط الانحدار الذى يعطى تقديراً لمتوسطات قيم ص عند قيم س ونستكمل هذا الوصف بتقدير مدى تشتت نقط شكل الانتشار حول هذا الخط . ويقاس هذا التشتت بما يسمى بالخطأ المعيارى لتقدير ص من معادلة الانحدار (٧) ، ويرمز له بالرمز ع$_{ص.س}$ ويعرف كالآتي :

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢} \text{ مح } (ص_{س} - ص̂_{س})^{٢} \qquad (٩)$$

وهذا المقياس له أهمية كبرى في عمليات الاستنتاج الإحصائي كما سنرى بعد .

ملاحظة :

يمكن إثبات أن .

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢} (\text{مح } ص^{٢} - ا \text{ مح } ص - ب \text{ مح } س ص) \qquad (١٠)$$

وهذه الصيغة أسهل في حساب الخطأ المعيارى من الصيغة (٩) وتستخرج من نفس المجاميع التى نوجدها لحساب معادلة الانحدار .

مثال (٩ - ٢) :

أوجد الخطأ المعيارى لخط الانحدار الناتج في المثال (٩ - ١) السابق .

الحل :

نعوض في الصيغة (١٠) من مجاميع الجدول (٩ - ١) بالقيمتين

ا = ٠,٢٥٦٨ ، ب = ٢,٩٣٠٣

$$ع^{٢}_{ص.س} = \frac{١}{٧}(٣٤٥,٠٩ - ٢,٩٣٠٣ \times ٩١,١ - ٠,٢٥٦٨ \times ١٠٣٦,٥٦)$$

$$= ٠,٢٩١٢$$

$$\therefore ع_{ص.س} = ٠,٥٣٩٦$$ (الخطأ المعيارى مقدراً من العينة)

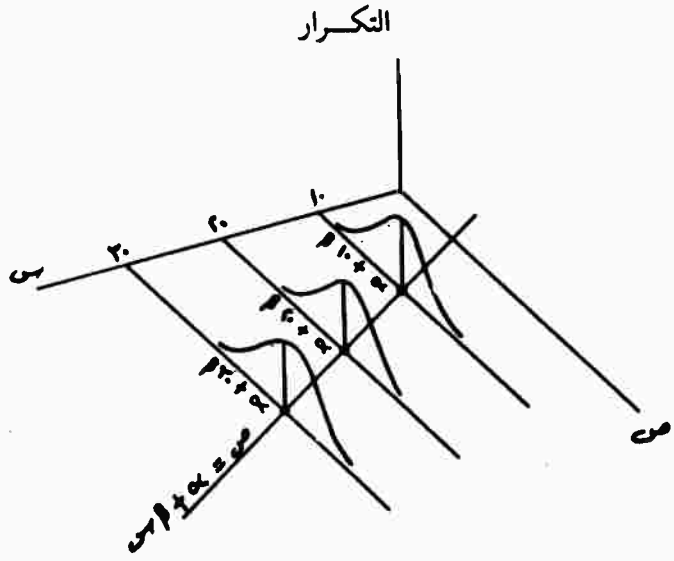
(٩ - ٥) استنتاجات إحصائية :

إن إيجاد معادلة الانحدار الخطى لم يستلزم إلا وضع الافتراضين الأول والثاني السابق ذكرهما ، أما إذا أردنا القيام باستنتاجات إحصائية أى إصدار قرارات عن المجتمع عن طريق العينة فإن ذلك يتطلب وضع افتراض خاص عن توزيع احتمال المتغير العشوائي ص- . وفى المعتاد نضع الافتراض الآتي (الذى يمكن هو أيضاً اختبار صحته) .

الافتراض الثالث :

« عند أى قيمة ثابتة س يكون للمتغير العشوائي ص- توزيع معتدل متوسطه $\alpha + \beta$ س وتباينه عدد ثابت مجهول $\sigma^{٢}$ مستقل عن س » .

ويمكن تصوير هذا الافتراض كالآتى :



الشكل (٩ – ٤) توزيعات معدلة للمتغير ص عند س = ١٠ ، ٢٠ ، ٣٠

ويلاحظ ما يلى :

(أ) متوسطات التوزيعات الصادية تختلف باختلاف قيم س وهى تقع جميعها على خط الانحدار .

(ب) جميع التوزيعات الصاديةمتوازية ولها نفس الشكل لأن لها نفس التباين وتختلف فقط فى مواقعها كما جاء فى (أ) .

من الافتراضات الثلاثة المذكورة ، بالاضافة إلى افتراض العشوائية الذى يضمن استقلال وحدات التجريب عن بعضها البعض ، يمكن أن نخرج بعدة استنتاجات نختار منها ما يلى :

(أولا) اختبار الفرض β = ك حيث ك عدد معين .

للوصول إلى خط الانحدار $\hat{ص}$ = ا + ب س نوجد معامل الانحدار ب من واقع بيانات مأخوذة من عينة ما ، وإذا اخترنا عينة أخرى نحصل على قيمة مختلفة

لهذا المعامل . أى أن قيمة ب تختلف من عينة إلى أخرى ، ولذلك نعتبر أن أى قيمة للمعامل ب هى إحدى القيم المشاهدة من متغير عشوائى B ذى عدد غير منتهى من القيم . وتحت الافتراضات الثلاثة للانحدار يمكن إثبات أن لهذا المتغير توزيعا معتدلا وسطه الحسابى β وانحرافه المعيارى يقدر بالمقدار $ع_{ص.س} / \sqrt{م_{س}}$ حيث $ع_{ص.س}$ هو الخطأ المعيارى للتقدير المعطى بالصيغة (٩) وحيث

$$م_{س} = مح (س - \bar{س})^{٢} = مح س^{٢} - \frac{(مح س)^{٢}}{ن}$$

وينتج من البند (٦ - ٦) أن الإحصاءة

$$ت = \frac{\beta - B}{ع_{ص.س} / \sqrt{م_{س}}} \qquad (١١)$$

يكون لها توزيع ت بدرجات حرية (ن - ٢) . وبهذه الإحصاءة نستطيع كالمعتاد اختبار الفرض الصفرى ف$_{٠}$: β = ك ضد أى فرض آخر وذلك بإيجاد ت$_{ح}$ من بيانات العينة أى بوضع B = ب وعلى أساس صحة الفرض الصفرى أى بوضع β = ك ثم مقارنة هذه القيمة بالقيمة الحرجة $ت_{\alpha[ن-٢]}$ التى نستخرجها من جدول ت .

نتيجة : اختبار خطية العلاقة بين س ، ص .

بصفة خاصة نستطيع اختبار الفرض الصفرى ف : β = ٠ ضد الفرض $\beta \neq$ ٠ . فإذا أشار الاختبار بقبول الفرض الصفرى حكمنا بعدم وجود علاقة خطية بين س ، ص لأن معني β = ٠ أن تكون ص = α أى ص تكون ثابتة ولا يكون لقيم س أى أثر خطى على ص . أما إذا رفضنا الفرض الصفرى لصالح الفرض

الآخر $\beta \neq$ ٠ . نحكم بأن هناك علاقة خطية بين المتغيرين وإن كان هذا لا يمنع من وجود علاقة أخرى قد تكون أفضل من العلاقة الخطية مثل ص = α + β س + لا س^{٢} وهناك اختبار آخر نعرف منه مدى انحراف العلاقة الحقيقية بين س ، ص عن العلاقة الخطية ، وسنقدم هذا الاختبار فى البند (٩ - ٩) .

مثال (٩ - ٣) :

ابحث وجود علاقة خطية بين المتغيرين س ، ص من بيانات المثال (٩ - ١) مستخدماً مستوى الدلالة ٠,٠١ .

الحل :

الفرض الصفرى ف$_{٠}$: β = ٠ (لا يوجد علاقة خطية)

الفرض الآخر ف$_{١}$: $\beta \neq$ ٠ (اختبار ذو جانبين)

$$م م_{س} = مج س_{ر}^{٢} - \frac{(مج س_{ر})^{٢}}{ن} = ١١٥,١١ - \frac{(٣٠,٣)^{٢}}{٩}$$

$$= ١٣,١$$

$$\therefore \sqrt{م م_{س}} = ٣,٦١٩٤$$

كما أن ع$_{ص.س}$ = ٠,٥٣٩٦ وقد سبق إيجادها

من (١١) وعلى أساس أن β = ٠ نجد أن :

$$ت_{ى} = \frac{٣,٦١٩٤ \times ٢,٩٣٠٣}{٠,٥٣٩٦} = ١٩,٦٥٥٢$$

من الجدول ت$_{٠,٠١}$ [٧] = ٣,٤٩٩

نرفض ف$_{٠}$ ونحكم بأن هناك علاقة خطية بين المتغيرين .

مثال (٩ – ٤) :

في بيانات المثال (٩ – ١) ابحث ما إذا كانت $\beta > ٢,٥$ عند مستوى الدلالة ٠,٠١

الحل :

ف$_{٠}$: $\beta = ٢,٥$

ف$_{١}$: $\beta > ٢,٥$ (اختبار ذو جانب واحد)

من (١١) وعلى أساس صحة الفرض الصفرى $\beta = ٢,٥$ نجد أن

$$ت_{ى} = \frac{(٢,٥ - ٢,٩٣٠٣) \times ٣,٦١٩٤}{٠,٥٣٩٦} = ٢,٨٨٦٣$$

من الجدول : $ت_{٠,٠٢}$ [٧] = ٢,٩٩٨

نقبل الفرض الصفرى أن $\beta = ٢,٥$ عند المستوى ٠,٠١ ونحكم بأن β لا تزيد عن ٢,٥ .

(ثانياً) فترات الثقة للبارامتر β .

من الإحصاءة (١١) نستطيع أن نثبت أن العددين

$$ب \pm \frac{ع_{ص.س}}{\sqrt{م_{٢}\, م_{س}}} \times ت_{\alpha\,[ن - ٢]} \qquad (١٢)$$

هما حدا الثقة بدرجة $(١ - \alpha)$ للبارامتر β .

(ثالثاً) فترات الثقة للقيمة الحقيقية للمتغير صـ عند قيمة معينة سـ. :

يمكن إثبات أن العددين

$$\hat{ص} \pm ع_{ص.س} \sqrt{١ + \frac{١}{ن} + \frac{(س. - \bar{س})^{٢}}{م_{س}}} \times ت_{\alpha[ن-٢]} \qquad (١٣)$$

هما حدا الثقة بدرجة $(١ - \alpha)$ لقيمة صـ عندما تأخذ سـ القيمة سـ.

مثال (٩ – ٥) :

للمثال (٩ – ١) أوجد فترة ثقة بدرجة ٩٥٪ لكل من :

(أ) البارامتر β (ب) القيمة الحقيقية للمتغير صـ عند سـ = ٢

الحل :

(أ) نعوض في (١٢) مع ملاحظة أن $ت_{٠,٠٥[٧]} = ٢,٣٦٥$

$$\text{الحد الأدنى للفترة} = ٢,٩٣٠٣ - \frac{٠,٥٣٩٦}{٣,٦١٩٤} \times ٢,٣٦٥ = ٢,٥٧٨$$

$$\text{الحد الأعلى للفترة} = ٢,٩٣٠٣ + \frac{٠,٥٣٩٦}{٣,٦١٩٤} \times ٢,٣٦٥ = ٣,٢٨٣$$

إذن الفترة (٢,٥٧٨ ، ٣,٢٨٣) هي فترة ثقة بدرجة ٩٥٪ للبارامتر β .

(ب) نعوض في (١٣) ، مع ملاحظة أنه عندما سـ = ٢ فإن

$$\hat{ص} = ٠,٢٥٦٨ + ٢,٩٣٠٣ \times ٢ = ٦,١١٧٤$$

$$\text{الحد الأدنى للفترة} = ٦,١١٧٤ - ٠,٥٣٩٦ \sqrt{١ + \frac{١}{٩} + \frac{(٢ - ٣,٣٦٦٧)^{٢}}{١٣,١}} \times ٢,٣٦٥$$

$$= ٦,١١٧٤ - ٠,٥٣٩٦ \times ١,١١٩٧ \times ٢,٣٦٥ = ٤,٦٨٩$$

الحد الأعلى للفترة = ٧,٥٤٦

إذن الفترة (٤,٦٨٩ ، ٧,٥٤٦) هى فترة ثقة بدرجة ٩٥٪ لقيمة ص عند س = ٢

ملاحظة :

هناك برامج جاهزة للحاسب الالكترونى تعطى جميع القيم المطلوبة في تحليل الانحدار بدءا من قيم الصيغة (٥) إلى قيم الصيغة (١٣) بمجرد تغذيته بالبيانات الخام .

(٩ - ٦) التوسع في استخدام الانحدار الخطى البسيط :

يمكن تناول بعض العلاقات غير الخطية بنفس الطريقة التي نتناول بها العلاقات الخطية وذلك باختيار تحويلات مناسبة للمتغيرات بحيث تأخذ العلاقة المعطاة الصورة الخطية . وكمثال لذلك نفرض أن العلاقة بين المتغيرين س ، ص على الصورة :

$$ص = \alpha \, هـ^{\beta س} \quad \text{حيث } \alpha > ٠ \text{ ، } ص > ٠$$

ونريد تقدير α ، β من العينة . بأخذ لوغاريتم كل من الطرفين للأساس هـ نجد أن :

$$\text{لو } ص = \text{لو } \alpha + \beta س$$

$$\text{أو } ص' = \alpha' + \beta س$$

$$\text{حيث } \alpha' = \text{لو } \alpha \text{ ، } ص' = \text{لو } ص$$

وهذه العلاقة الأخيرة على صورة خطية . لإيجاد تقديرين ا ، ب للمجهولين α ، β نبدأ بتحويل كل قيمة ص إلى لو ص فيصبح لدينا الأزواج (س ، صَ) بدلا من الأزواج (س ، ص) . نوجد التقديرين ب ، اَ كالمعتاد من الصيغتين (٥ ، ٦) فتكون ب تقديراً للبارامتر β وتكون ا = هَ$^{\text{اَ}}$ هى تقدير للبارامتر α وتكون معادلة الانحدار هى :

$$\text{ص} = \text{ا هـ}^{\text{ب س}}$$

وبنفس الطريقة يمكن تناول المعادلات الآتية :

ص = $\alpha + \frac{\beta}{\text{س}}$ حيث س $\neq$ ٠ (ضع سَ = $\frac{١}{\text{س}}$ التحويل القالب)

ص = α س$^{\beta}$ حيث س > ٠ ، α > ٠ ، ص > ٠ (تحويل لوغاريتمي)

ص = $\alpha + \beta$ حتا ٣ س (ضع سَ = حتا ٣ س)

ص س$^{\alpha}$ = β حيث س > ٠ ، ص > ٠ ، β > ٠ (تحويل لوغاريتمي)

مثال (٩ - ٦) :

المعروف أن العلاقة بين ضغط الغاز صه وحجمه ح تأخذ الصورة صه ح$^{\beta}$ = α أوجد تقديراً للبارامترين المجهولين α ، β من البيانات التجريبية الآتية ثم أوجد أحسن تقدير لضغط الغاز حين يكون حجمه ح = ١٠٠

ح :	١٩٤,٠	١١٨,٦	٨٨,٧	٧٢,٤	٦١,٨	٥٤,٣
ض :	١٠,١	١٩,٢	٢٨,٤	٣٧,٦	٤٩,٥	٦١,٢

الحل :

نحول المعادلة صـ $= \alpha$ ح$^{\beta}$ إلى صورة خطية بأخذ لوغاريتم كل من الطرفين للأساس ١٠ .

$\therefore$ لو ض $= \beta$ لو ح + لو α أو لو ض = لو α + β لو ح

أو ص = $\bar{\alpha}$ + $\bar{\beta}$ س

حيث ص = لو ض ، $\bar{\alpha}$ = لو α ، $\bar{\beta}$ = $-\beta$ ، س = لو ح

لتقدير بارامترات هذه المعادلة الخطية ينبغى أن نوجد لوغاريتم كل قيمة معطاة من قيم الحجم ح ولوغاريتم كل قيمة مناظرة من قيم الضغط صـ . وباستخدام جداول اللوغاريتمات أو الحاسبات نحصل على العمودين الأول والثاني من الجدول الآتي ، ثم نستكمل الحساب كما في المثال (٩ - ١) .

الجدول (٩ - ٢)

س = لو ح	ص = لو ض	س^{٢}	س ص
٢,٢٨٧٨	١,٠٠٤٣	٥,٢٣٤٠	٢,٢٩٧٦
٢,٠٧٤١	١,٢٨٣٣	٤,٣٠١٩	٢,٦٦١٧
١,٩٤٧٩	١,٤٥٣٣	٣,٧٩٤٣	٢,٨٣٠٩
١,٨٥٩٧	١,٥٧٥٢	٣,٤٥٨٥	٢,٩٢٩٤
١,٧٩١٠	١,٦٩٤٦	٣,٢٠٧٧	٣,٠٣٥٠
١,٧٣٤٨	١,٧٨٦٨	٣,٠٠٩٥	٣,٠٩٩٧
١١,٦٩٥٣	٨,٧٩٧٥	٢٣,٠٠٥٩	١٦,٨٥٤٣

سَ = ١,٩٤٩٢ ، صَ = ١,٤٦٦٣

على فرض أن أَ ، بَ هما التقديران المطلوبان للثابتين α ، β نجد ما يلى :

من الصيغة (٥) : بَ = $\frac{٦ \times ١٦,٨٥٤٣ - ٨,٧٩٧٥ \times ١١,٦٩٥٣}{٦ \times ٢٣,٠٠٥٩ - (١١,٦٩٥٣)^{٢}}$

= -١,٤٠

من الصيغة (٦) : أَ = صَ - بَ سَ = ١,٤٦٦٣ + ١,٤٠ × ١,٩٤٩٢

= ٤,١٩٥٢ = ٤,٢ تقريبا

إذن ب = - ب = ١,٤٠

، أ = ١٠^{٤,٢} = ١٥٨٤٨,٩ = ١٥٨٤٩ تقريباً

وهذان هما التقديرات المطلوبان للثابتين β ، α وعلى ذلك فإن المعادلة التي تربط الحجم والضغط الناتجة من العينة هى :

ص ح^{١,٤٠} = ١٥٨٤٩

وحين ح = ١٠٠ فإن أحسن تقدير للضغط ض ينتج كالآتى :

ض × ١٠٠^{١,٤٠} = ١٥٨٤٩ ومنها ض × ١٠^{٢,٨٠} = ١٥٨٤٩

إذن ض = ٢٥,١٢ تقريباً .

تمارين (٩ - ١)

من كل من البيانات المبينة في المسائل الخمسة الآتية :

(ا) أوجد معادلة انحدار ص على س .

(ب) أوجد الخطأ المعيارى $ع_{ص.س}$ لخط الانحدار .

(حـ) ابحث ما إذا كان هناك علاقة خطية بين المتغيرين س ، ص .

(ء) إذا ثبت وجود علاقة خطية فأوجد فترة ثقة بدرجة ٩٥٪ للقيمة الحقيقية للمتغير ص عند القيمة $س_{.}$ المعطاة .

(١) س : ٢,٨ ٢,٩ ٣,٠ ٣,١ ٣,٢ ٣,٢ ٣,٢ ٣,٣ ٣,٤

ص : ٢٧ ٢٣ ٣٠ ٢٨ ٣٠ ٣٢ ٣٤ ٣٣ ٣٠

حيث س ترمز إلى كثافة الحديد الخام (جرام / سم^{٣})

، ص ترمز إلى النسبة المئوية لمحتوى الحديد .

، $س_{.}$ = ٣

(٢) س : ٤٨ ٤٩ ٥٠ ٥١ ٥٢

ص : ٢٧٧,١ ٢٧٩,٣ ٢٨١,٤ ٢٨٣,٢ ٢٨٤,٨

حيث س ترمز إلى طول الطفل عند الولادة بالسنتيمترات

، ص ترمز إلى مدة الحمل بالأيام .

، $س_{.}$ = ٥٠

(٣) س : ٨ ٩ ٩ ١٢ ١٣ ١٤ ١٤ ١٦ ١٦ ١٦

ص : -٥٥ -٥٥ -٥٣ -٥٠ -٥٠ -٣٥ -٣٧ -٣٥ -٢٧ -٢٤

حيث س ترمز إلى درجات الحرارة بالسنتيجراد

، ص ترمز إلى الانحراف (بمضاعفات الـ $\frac{1}{1000}$ زاوية دائرية) لنوع معين من الرؤية التلسكوبية .

، $س_{.}$ = ١٣,٥

(٤) س :	١٥	٣٠	٦٠	١٢٠	٣٠٠	٦٠٠
ص :	٤,٠	٦,٣	٥,٢	٦,٤	٥,٥	٦,٤

حيث س ترمز إلى المدة بالثواني

، ص ترمز إلى مقدار المادة التي تبددت (بالجرامات في اللتر) من استحلاب الزئبق في محلول زيتريت الصديون بتأثير الاهتزاز الناتج من الصوت عن ذبذبات فوق الصوتية في المدة س .

، $س_{.}$ = ٧٠

(٥) س :	٠,٠١	٠,٠٢	٠,٠٤	٠,٠٨	٠,١٥	٠,٣٠	٠,٥٠	١,٠٠	٢,٠٠
ص :	١٢	١٨	٣	٢٧	١٦	١٣	١٠	٧٠	٩٠

حيث س ترمز إلى النسبة المئوية لدرجة تركيز الكلورونافتالين

، ص ترمز إلى النسبة المئوية لموت النمل الأبيض .

، $س_{.}$ = ٠,٢٥

(٦) في عينة عشوائية من ٨ أزواج من القيم ($س_ر$ ، $ص_ر$) وجد أن معامل الانحدار

ب = ٠,٣١ وأن $ع^{٢}_{ص.س}$ = ٠,٤ ، $\overline{م_{س}^{٢}}$ = ٠,٣

اختبر الفرض أن β = ٠ ضد الفرض أن β > .

(٩ - ٧) معنى آخر للانحدار - تحليل التباين :

فى البنود السابقة من هذا الفصل كنا نأخذ الانحدار على أنه وسيلة لإيجاد معادلة تمكننا من معرفة أحسن تقدير لقيمة متغير عشوائى صـ عن طريق قيمة معطاة لمتغير سـ . ولذلك وصفنا معادلة الانحدار بأنها معادلة تنبؤ . إلا أن الانحدار يؤخذ أيضا على أنه وسيلة لتفسير الاختلاف المشاهد فى قيم المتغير صـ ، وذلك استجابة لتساؤل هام عن العوامل المؤثرة فى هذا الاختلاف ، وبصفة خاصة عما إذا كان هذا الاختلاف يرجع إلى التغير فى قيم المتغير سـ أم يرجع إلى عوامل عشوائية أو عوامل أخرى لا تتعلق بالمتغير سـ .

إن الإجابة عن هذا التساؤل تتطلب تحليل الاختلاف فى قيم صـ وهو م م (ص) = مـ (صـ ر - صَ)^{٢} إلى مركبتين مستقلتين تعتمد أحدهما على قيم المتغير سـ ولا تعتمد الأخرى عليها ثم تقييم كل من هاتين المركبتين .

ولبيان كيفية هذا التحليل نعتبر المتساوية الآتية :

$$ص_ر - \bar{ص} = ص_ر - \hat{ص}_ر + \hat{ص}_ر - \bar{ص}$$

$$= (\hat{ص}_ر - \bar{ص}) + (ص_ر - \hat{ص}_ر) \text{ حيث } ر = ١، ٢ ، \ldots ، ن$$

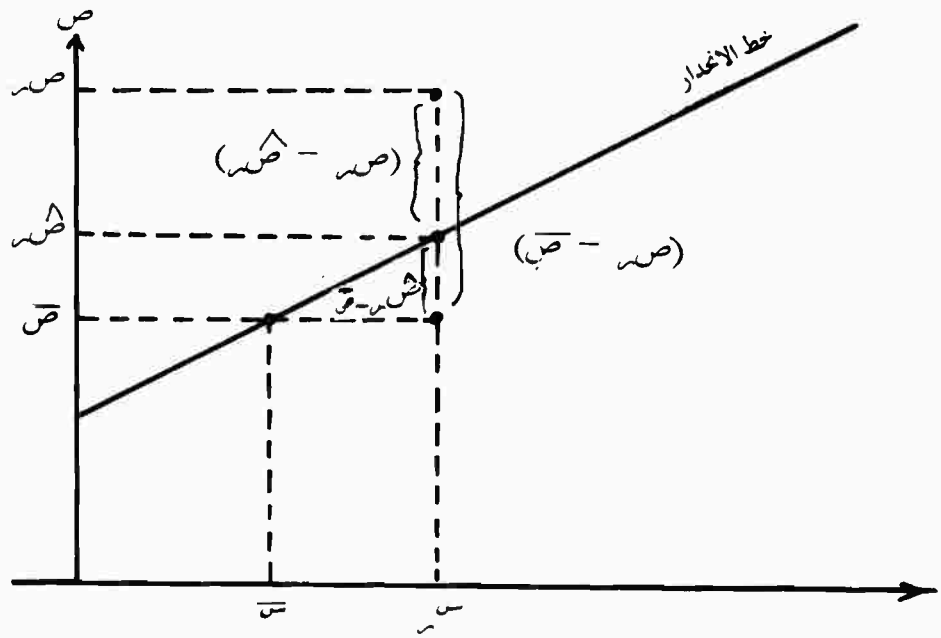
إن هذه المتساوية واضحة جبريا ، كما أنها تتضح هندسيا من الشكل (٩ - ٥) مع ملاحظة أن النقطة ($\bar{س}$ ، $\bar{ص}$) تقع دائما على خط الانحدار . وبتربيع طرفى المتساوية ثم الجمع تنتج المتساوية الآتية :

$$\sum_{ر=١}^{ن} (ص_ر - \bar{ص})^{٢} = \sum_{ر=١}^{ن} (\hat{ص}_ر - \bar{ص})^{٢} + \sum_{ر=١}^{ن} (ص_ر - \hat{ص}_ر)^{٢} \quad (١٤)$$

مع ملاحظة أن الحد الأوسط فى عملية تربيع الطرف الأيسر ينعدم عند عملية

الجمع . وبذلك نكون قد جزأنا الاختلاف الكلى فى ص إلى مركبتين بطريقة مشابهة لتقسيم الاختلاف الكلى فى تحليل التباين .

لنبحث الآن فى المعنى الذى تتضمنه كل من هاتين المركبتين .



الشكل (٩ – ٥) : تجزىء الاختلاف فى المتغير ص

الاختلاف المفسر والاختلاف غير المفسر

EXPLAINED AND UNEXPLAINED VARIATION

من الشكل (٩ – ٥) نرى أننا إذا رسمنا خط الانحدار وحددنا قيمة معينة $س_ر$ فإن قيمة $\hat{ص}_ر$ المناظرة تكون قد تحددت تماما لأنها تقع على خط الانحدار ويكون الفرق $(\hat{ص}_ر - \bar{ص})$ بين $\hat{ص}_ر$ والقيمة الثابتة $\bar{ص}$ هو فرق يعتمد كلية على قيمة $س_ر$ وبالتالى فإن مجموع مربعات الفروق $مح (\hat{ص}_ر - \bar{ص})^2$ يعتمد كلية على قيم المتغير س . ولما كان هذا المقدار هو جزء من الاختلاف الكلى فى ص كما يظهر من المتساوية (١٤) فإن $مح (\hat{ص}_ر - \bar{ص})^2$ يكون هو الجزء من الاختلاف فى ص

الذى يُعزّى إلى التغير فى س ، أى إلى التغير الذى حدث فى ص نتيجة للتغير فى س ، وهذا ما نسميه بانحدار ص على س ولذلك نرمز لهذا الاختلاف بالرمز م م (الانحدار الخطى) وله درجة واحدة من درجات الحرية . ولما كان هذا الاختلاف مصدره معروف (وهو التغير فى س) فقد اصطلح على تسميته أيضا بالاختلاف المفسر ونكتبه رمزيا كالآتى :

الاختلاف المفسر = م م (الانحدار) = مح $(\hat{ص}_{ر} - \bar{ص})^{٢}$ بدرجة حرية واحدة (١٥)

ويمكن إثبات أن هذا الاختلاف يمكن أن يحسب كالآتى :

الاختلاف المفسر = $ب^{٢}$ م م (س) حيث ب معامل الانحدار (١٦)

أو $$= \frac{[م صـ (س ، ص)]^{٢}}{م م (س)}$$ (١٧)

حيث م م (س) $= مح (س - \bar{س})^{٢} = مح س^{٢} - \frac{(مح س)^{٢}}{ن}$

، م صـ (س ، ص) $= مح (س - \bar{س})(ص - \bar{ص}) = مح س ص - \frac{(مح س)(مح ص)}{ن}$

أما المركبة الثانية وهى مح $(ص_{ر} - \hat{ص}_{ر})^{٢}$ فهى الاختلاف الذى يتبقى بعد طرح الاختلاف المفسر من الاختلاف فى ص ، وهو يعتمد على عوامل مجهولة لا تتعلق بالمتغير س . وقد اصطلح على تسمية هذا الاختلاف بالاختلاف المتبقى residual variation أو بالاختلاف غير المفسر . ومع تذكر أن $ص_{ر}$ هى قيم عشوائية فإن هذا الاختلاف يعبر عن التشتت غير المنتظم لنقط شكل الانتشار حول خط الانحدار أى عن الانحرافات الرأسية حول خط الانحدار وسنرمز له بالرمز م م (الانحراف عن خط الانحدار) وله ن - ٢ من درجات الحرية . أى أن

الاختلاف غير المفسر = م م (الانحراف عن خط الانحدار)

$$= \text{مح } (ص_{ر} - \hat{ص}_{ر})^{٢}$$

$$= \text{م م } (ص) - \text{الاختلاف المفسر} \qquad (١٨)$$

والمعنى الذى يتضمنه هذا الاختلاف يبرر استخدامه كأساس لحساب الخطأ المعيارى $ع_{ص.س}$ الوارد فى الصيغة (٩) بالبند (٩ - ٤) ، حيث

$$ع^{٢}_{ص.س} = \frac{\text{مح } (ص_{ر} - \hat{ص}_{ر})^{٢}}{ن - ٢} = \frac{\text{الاختلاف غير المفسر}}{\text{درجات الحرية}} \qquad (١٩)$$

نستطيع حينئذ أن نكتب المتساوية (١٤) كالآتى :

م م (ص) = م م (الانحدار الخطى) + م م (الانحراف عن خط الانحدار) (٢٠)

بدرجات حرية ن - ١ ، ١ ، ن - ٢ على الترتيب .

ويفضل تسجيل قيم هذه المتساوية فى جدول التباين الآتى .

الجدول (٩ - ٣)

جدول التباين للانحدار

مصدر التباين	م م	د ح	ط م	ف ى
الاختلاف المفسر (الانحدار الخطى)	$\text{مح } (\hat{ص}_{ر} - \bar{ص})^{٢}$	١	$ع^{٢}_{١}$	$\frac{ع^{٢}_{١}}{ع^{٢}_{٢}}$
الاختلاف غير المفسر (الانحراف عن الخطية)	$\text{مح}(ص_{ر} - \hat{ص}_{ر})^{٢}$	ن - ٢	$ع^{٢}_{٢}$	
الاختلاف الكلى فى ص	$\text{مح}(ص_{ر} - \bar{ص})^{٢}$	ن - ١		

بعد تجزىء الاختلاف الكلى فى ص بهذه الصورة يبقى أن نختبر ما إذا كان الانحدار الخطى قد فسر جزءا ذا بال من هذا الاختلاف ، أى أن نختبر ما إذا كان التباين المفسر أكبر كبرا جوهريا من تباين الاختلاف غير المفسر . وهذا ما نقيسه باختبار ف بالصيغة الآتية بشرط توفرافتراضات الانحدار ، ومع ملاحظة أن هذين التباينين هما تقديران مستقلان للتباين σ^2 لتوزيع المتغير صـ .

$$\text{ف} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \qquad \text{بدرجتى حرية ١، ن} - ٢ \qquad (٢١)$$

وهذا الاختبار يكافىء اختبار ت الوارد بالصيغة (١١) بالبند (٩ – ٥) لاختبار وجود علاقة خطية بين المتغيرين سـ ، صـ أى لاختبار الفرض الصفرى $\beta = ٠$ ضد الفرض $\beta \neq ٠$. ويمكن اشتقاق أى منهما من الآخر ، إذ أن قيمة ف بالصيغة (٢١) تساوى مربع قيمة ت بالصيغة (١١) . وبذلك نكون قد توصلنا إلى طريقة أخرى لاختبار وجود علاقة خطية بين المتغيرين .

وجدير بالذكر أن الصيغة (٢١) هى صيغة عامة لاختبار مدى دقة التنبؤ من خط الانحدار مهما كان عدد المتغيرات المستقلة المستخدمة فى عملية التنبؤ، وتختلف طريقة حساب البسط والمقام فى هذه الصيغة باختلاف عدد المتغيرات التنبؤية .

مثال (٩ – ٧) :

ابحث وجود علاقة خطية بين المتغيرين سـ ، صـ من بيانات المثال (٩ – ١) مستخدما طريقة تحليل التباين .

الحل :

الفرض الصفرى : $\beta = ٠$ الفرض الآخر : $\beta \neq ٠$.

من الجدول (٩ – ١) نستطيع حساب القيم اللازمة لاستخدام الصيغة (٢١) كالآتى :

$$م_{ص} (ص) = ١٠٣٦,٦٥ - \frac{٩١,١^{٢}}{٩} = ١١٤,٥١٥٦$$

$$م_{س} (س) = ١١٥,١١ - \frac{(٣٠,٣)^{٢}}{٩} = ١٣,١$$

$$م_{صـ} (س ، ص) = ٣٤٥,٠٩ - \frac{٩١,١ \times ٣٠,٣}{٩} = ٣٨,٣٨٦٧$$

من (١٧) : الاختلاف المفسر $= \frac{(٣٨,٣٨٦٧)^{٢}}{١٣,١} = ١١٢,٤٨٣٩$ $\quad \nu_{١} = ١$

∴ الاختلاف غير المفسر = ١١٤,٥١٥٦ – ١١٢,٤٨٣٩

$= ٢,٠٣١٧$ $\quad \nu_{٢} = ٧$

وينشأ جدول التباين الآتى :

الجدول (٩ – ٤)

مصدر التباين	م م	د ح	ط م	ف
الاختلاف المفسر	١١٢,٤٨٣٩	١	١١٢,٤٨٣٩	٣٨٧,٦٠٨**
الاختلاف غير المفسر	٢,٠٣١٧	٧	٠,٢٩٠٢	
الاختلاف الكلى	١١٤,٥١٥٦	٨		

إن الفيمة ف$_{ب}$ = ٣٨٧,٦٠٨ أكبر بكثير من القيمة الحرجة ف ٠,٠١[٧،١] = ١٢,٢ مما يجعلنا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونحكم بوجود علاقة خطية بين المتغيرين .

نلاحظ أن الجذر التربيعى لقيمة ف$_{ب}$ وهو $\sqrt{٣٨٧,٦٠٨}$ = ١٩,٦٨٨ يساوى قيمة ت$_{ب}$ = ١٩,٦٥٥٢ السابق إيجادها بالمثال (٩ – ٣) ويرجع الفرق بين القيمتين إلى أخطاء التقريب .

معامل التحديد COEFFICIENT OF DETERMINATION

استكمالا للانحدار كوسيلة لتفسير الاختلاف فى قيم المتغير التابع ص يهمنا أن نقدر نسبة الاختلاف الذى فسره الانحدار الخطى إلى الاختلاف الكلى فى المتغير ص . وهذه النسبة تسمى بمعامل التحديد ويرمز لها بالرمز ر^{٢} . أى أن :

$$ر^{٢} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (٢٢)$$

ففى المثال السابق نجد أن :

$$ر^{٢} = \frac{١١٢,٤٨٣٩}{١١٤,٥١٥٦} = ٠,٩٨٢ \text{ تقريبا}$$

وهذا يعنى أن الانحدار الخطى قد فسر حوالى ٩٨,٢٪ من الاختلاف الكلى أى جزءا جوهريا منه ، مما يؤكد صحة العلاقة الخطية بين المتغيرين .
ويرمز لمعامل التحديد بالرمز ر^{٢} لأنه يساوى مربع معامل الارتباط ر الذى سنتناوله فى الفصل التالى . ويلاحظ أن

$$٠ \leqslant ر^{٢} \leqslant ١$$

لأن كلا من البسط والمقام فى التعريف (٢٢) هو مجموع مربعات لا يمكن أن يكون سالبا وإذن ر^{٢} ⩾ ٠ كما أن البسط جزء من المقام كما يتضح من المتساوية

(١٤) واذن $ر^{٢} \leq ١$. وإذا كانت $ر^{٢} = ٠$ فإن هذا يعنى أن الانحدار الخطى لا يفسر شيئا من الاختلاف فى ص ولا يكون للمتغير فى س أى أثر فى التغير فى ص . أما إذا كانت $ر^{٢} = ١$ فإن هذا يعنى أن الانحدار قد فسر الاختلاف فى ص بأكمله ، وهذا يحدث حين تكون القيم المشاهدة $ص_{ر}$ مساوية للقيم المناظرة $ص_{ر}$ المقدرة من معادلة الانحدار ، أى تكون النقط ($س_{ر}$ ، $ص_{ر}$) المشاهدة واقعة جميعها على خط الانحدار . وحين تكون قيمة $ر^{٢}$ صغيرة فإن هذا يعنى أن الجزء الأكبر من الاختلاف فى ص يرجع إلى عوامل ومتغيرات لا تدخل فى الانحدار أى لا علاقة لها بالتغير فى المتغير المستقل س .

من الصيغة (١٧) يمكن أن نكتب معامل التحديد بدلالة مجاميع المربعات ومجاميع حواصل الضرب كالآتى :

$$ر^{٢} = \frac{[م\ صـ\ (س\ ،\ ص)]^{٢}}{م\ م\ (س)\ .\ م\ م\ (ص)} \qquad (٢٣)$$

هذا ويمكن كتابة اختبار ف بالصيغة (٢١) بدلالة معامل التحديد كالآتى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{\text{الاختلاف المفسر} \div ١}{\text{الاختلاف غير المفسر} \div (ن - ٢)}$$

$$= \frac{\text{الاختلاف المفسر}}{(\text{الاختلاف الكلى} - \text{الاختلاف المفسر}) \div (ن - ٢)}$$

وبقسمة كل من البسط والمقام على الاختلاف الكلى م م (ص) ينتج أن

$$ف = \frac{ر^{٢}}{(١ - ر^{٢})/(ن - ٢)} \qquad (٢٤)$$

وهذه صيغة ثالثة لاختبار وجود علاقة خطية بين متغيرين .

(٩ – ٨) تحليل الانحدار حين يكون هناك أكثر من قيمة ص لكل قيمة س .

فى الأمثلة والتمارين السابقة كانت التجربة تحدد قيمة عشوائية واحدة ص لكل قيمة ثابتة س . إلا أن بعض الأبحاث تفضل تصميم التجربة بحيث نحصل منها على أكثر من قيمة عشوائية من المتغير صـ لكل قيمة من قيم المتغير سـ ، خاصة وأن هذا التصميم يوفر لنا الفرصة للحكم على جودة العلاقة المفروضة بين المتغيرين كما سنرى فى البند (٩ – ٩) .

فإذا كان هناك ك من القيم السينية $س_1$ ، $س_2$ ، ... ، $س_و$ يكون لدينا ك من المجموعات الصادية المناظرة لها ، وتتخذ البيانات فى هذه الحالة الصورة المبينة بالجدول (٩ – ٥) والتى تشبه الصورة التى يتخذها تحليل التباين للتجارب ذوات العامل الواحد .

الجدول (٩ – ٥)

$س_1$	$س_2$	...	$س_و$	...	$س_ك$
$ص_{11}$	$ص_{21}$	...	$ص_{1و}$	...	$ص_{1ك}$
$ص_{12}$	$ص_{22}$	...	$ص_{2و}$	...	$ص_{2ك}$
...	...	...	...	...	...
...	...	...	...	...	...
$ص_{ر1}$	$ص_{ر2}$	...	$ص_{رو}$	...	$ص_{رك}$
...	...	...	...	...	...
...	...	...	...	...	...
$ص_{ن1}$	$ص_{ن2}$	...	$ص_{نو}$	...	$ص_{نك}$

وإذا تبنينا نفس الافتراضات الثلاثة للانحدار الخطى فإن أسلوب التحليل يسير على نفس النمط السابق تقديمه مع بعض التعديلات التى يقتضيها الوضع الجديد للبيانات كما يتبين من المثال الآتى .

مثال (٩ - ٨) :

فى دراسة عن توزيع الأوزان لمجتمع من الرجال وعلاقة هذا التوزيع بالأطوال ، قسم مجتمع الرجال من حيث الطول إلى ٤ أقسام تتساوى فيها الأطوال بالتقريب وتمثلها الأطوال ٦٠ ، ٦٢ ، ٦٤ ، ٧٠ . وفى كل من هذه الأطوال اختير عدد من الرجال عشوائيا وقيست أوزانهم وسجلت المقاييس فى الجزء العلوى من الجدول (٩ - ٦) الآتى :

الجدول (٩ - ٦)

الأوزان	الأطوال			
	$س_1$	$س_2$	$س_3$	$س_4$
	٦٠	٦٢	٦٤	٧٠
	١١٠	١٢٠	١٥٠	١٧٠
	١٣٥	١٤٠	١٤٥	١٨٥
	١٢٠	١٣٠		١٦٠
		١٣٥		

$ن_و$	٣	٤	٢	٣
$مح_ر ص_{ر و}$	٣٦٥	٥٢٥	٢٩٥	٥١٥
$مح_ر ص^2_{ر و}$	٤٤٧٢٥	٦٩١٢٥	٤٣٥٢٥	٨٨٧٢٥
$ن_و س_و$	١٨٠	٢٤٨	١٢٨	٢١٠
$ن_و س^2_و$	١٠٨٠٠	١٥٣٧٦	٨١٩٢	١٤٧٠٠
$س_و مح_ر ص_{ر و}$	٢١٩٠٠	٣٢٥٥٠	١٨٨٨٠	٣٦٠٥٠

ن = ١٢

مح مح ص = ١٧٠٠

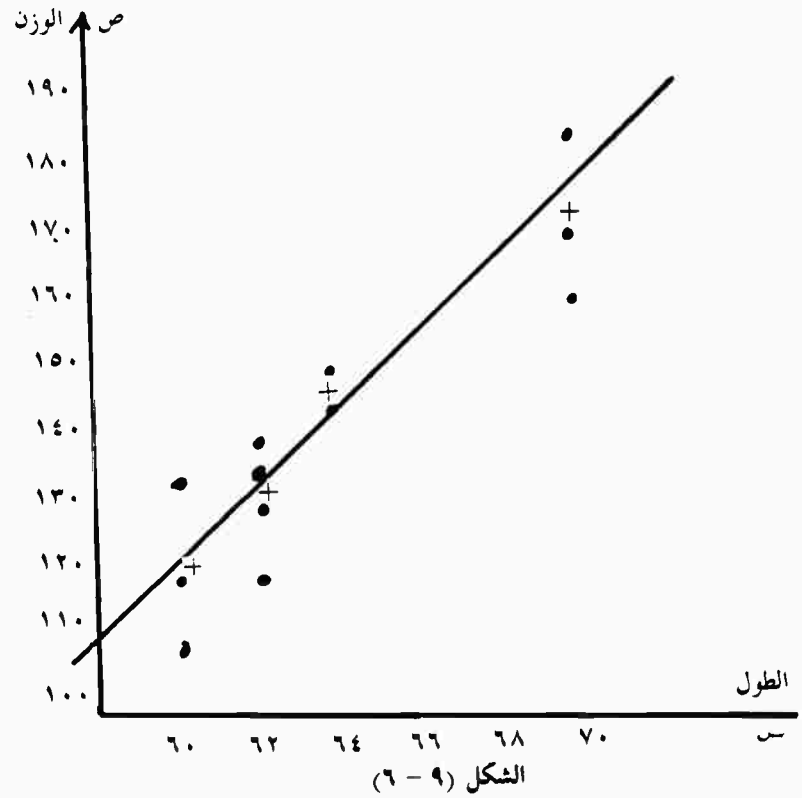
مح مح $ص^2$ = ٢٤٦١٠٠

مح مح س = ٧٦٦

مح مح $س^2$ = ٤٩٠٦٨

مح مح س ص = ١٠٩٣٨٠

نعبر عن هذه البيانات هندسيا كما فى الشكل (٩ – ٦) الذى يعرض شكل الانتشار ويشتمل على ١٢ نقطة تشترك بعضها فى الإحداثيات السينية وكل نقطة تعبر عن طول ووزن أحد الرجال . أما النقط المشار إليها بالعلامة + فتمثل متوسطات المجموعات الصادية عند القيم السينية المناظرة أى تمثل النقط ($س_و$ ، $\bar{ص}_و$) حيث و = ١ ، ٢ ، ٣ ، ٤ .



الشكل (٩ – ٦)

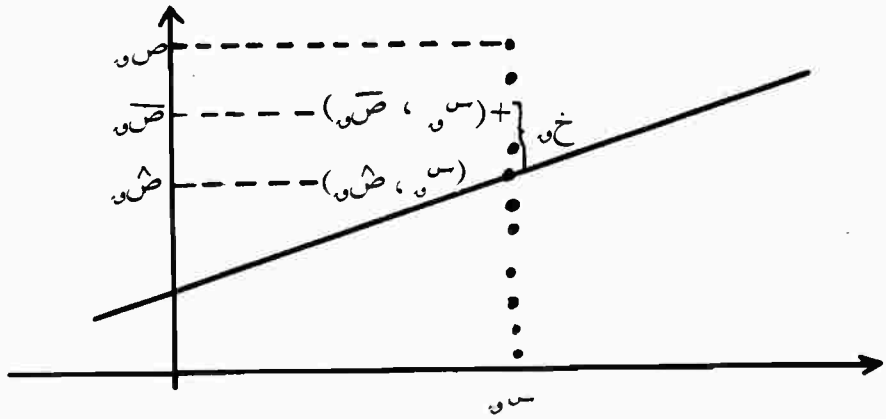
شكل الانتشار وخط الانحدار لبيانات المثال (٩ – ٨)

المطلوب فى هذا المثال إيجاد معادلة انحدار ص على س واختبار دلالة هذا الانحدار .

إيجاد معادلة الانحدار :

فى المثال (٩ – ٧) حيث كان لدينا قيمة واحدة ص لكل قيمة س كان بحثنا يهدف إلى معرفة ماإذا كانت القيم الصادية ص$_١$ ، ص$_٢$ ،... نقع على خط مستقيم هو خط الانحدار،

أما فى المثال (٩ – ٨) حيث لدينا مجموعة من القيم الصادية لكل قيمة س فان بحثنا يهدف فى هذه الحالة إلى معرفة ما إذا كانت المتوسطات $\bar{ص}_1$ ، $\bar{ص}_2$ ، ... للمجموعات الصادية تقع على خط مستقيم . وعلى ذلك فإنه حسب مبدأ المربعات الصغرى – راجع البند (٩ – ٣) – يكون خط الانحدار هو ذلك الذى تختار ثوابته بحيث تجعل الدالة مـ $(\bar{ص}_و - \hat{ص}_و)^2$ نهاية صغرى ، حيث $\hat{ص}_و$ هو الإحداثى الصادى للنقطة الواقعة على خط الانحدار والتى إحداثيها السينى $س_و$ – انظر الشكل (٩ – ٧) .



الشكل (٩ – ٧)

ويهمنا أن نشير إلى أننا إذا وضعنا بدلا من كل قيمة $ص_{ر و}$ فى أى عمود هـ الوسط الحسابى $\bar{ص}_و$ للصادات فى هذا العمود فإن معادلة الانحدار التى تنتج تكون هى بذاتها معادلة انحدار ص على س لأن هذا التغيير لا يؤثر فى المجاميع أو مجاميع المربعات أو مجاميع حواصل الضرب ، وهذا ما يمكن بيانه جبريا . وهذه الحقيقة تعنى أنه إذا وقعت متوسطات الأعمدة على خط مستقيم فإن هذا الخط ينطبق على خط انحدار ص على س .

لايجاد معامل الانحدار ب نستخدم الصيغة (٥) التى تأخذ فى هذه الحالة الشكل الآتى :

$$ب = \frac{\text{مح مح س ص} - \frac{(\text{مح مح س})(\text{مح مح ص})}{ن}}{\text{مح مح س}^{٢} - \frac{(\text{مح مح س})^{٢}}{ن}}$$

وفى هذا المثال نجد أن :

$$ب = \frac{١٠٩٣٨٠ - \frac{٧٦٦ \times ١٧٠٠}{١٢}}{٤٩٠٦٨ - \frac{٧٦٦^{٢}}{١٢}} = \frac{٨٦٣,٣٤}{١٧١,٦٦٧} = ٥,٠٢٩٢$$

$$\bar{س} = \frac{١}{ن} \text{مح مح س} = \frac{٧٦٦}{١٢} = ٦٣,٨٣٣٣$$

$$\bar{ص} = \frac{١}{ن} \text{مح مح ص} = \frac{١٧٠٠}{١٢} = ١٤١,٦٦٦٧$$

∴ معادلة انحدار ص على س مقربة إلى ٣ خانات عشرية هى :

$$\hat{ص} = ١٤١,٦٦٧ + ٥,٠٢٩ \ (س - ٦٣,٨٣٣)$$

أى $\hat{ص} = - ١٧٩,٣٤٩ + ٥,٠٢٩$ س

اختبار وجود علاقة خطية :

على فرض توفر شروط الانحدار يمكننا اختبار وجود علاقة خطية بين المتغيرين أى اختبار الفرض الصفرى $\beta = ٠$ ضد الفرض $\beta \neq ٠$ باستخدام اختبار ت بالصيغة (١١) أو اختبار ف بالصيغة العامة (٢١) أو بالصيغة المكافئة (٢٤) . وسنستخدم هنا الصيغة العامة وهى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$ بدرجتى حرية ١ ، ن − ٢

من بيانات المثال نجد مايلى :

الاختلاف الكلى $= م م (ص) = مجـ مجـ ص^{٢} - \frac{(مجـ مجـ ص)^{٢}}{ن}$

$= ٢٤٦١٠٠ - \frac{١٧٠٠^{٢}}{١٢} = ٥٢٦٦,٦٧$

الاختلاف المفسر $= \frac{[م صـ (س ، ص)]^{٢}}{م م (س)}$

$= \frac{[\frac{١٧٠٠ \times ٧٦٦}{١٢} - ١٠٩٣٨٠]^{٢}}{\frac{٧٦٦^{٢}}{١٢} - ٤٩٠٦٨} = \frac{(٨٦٣,٣٤)^{٢}}{١٧١,٦٦٧} = ٤٣٤١,٨٧٠٨$

∴ الاختلاف غير المفسر $= ٥٢٦٦,٦٧ - ٤٣٤١,٨٧٠٨ = ٩٢٤,٧٩٩٢$

وبذلك نحصل على جدول التباين الآتى :

الجدول (٩ - ٧)

مصدر التباين	م م	د ح	ط م	ف
الاختلاف المفسر	٤٣٤١,٨٧٠٨	١	٤٣٤١,٨٧٠٨	٤٦,٩٥**
الاختلاف غير المفسر	٩٢٤,٧٩٩٢	١٠	٩٢,٤٧٩٩	
الاختلاف الكلى	٥٢٦٦,٦٧٠٠	١١		

بما أن ٤٦,٩٥ أكبر بكثير من القيمة الحرجة ف $_{٠,٠١}$ $_{[١٠ ، ١]}$ $= ١٠,٠٠$

نرفض الفرض الصفرى $\beta = ٠$ عند مستوى عالى من الدلالة ونحكم بوجود علاقة خطية بين طول الرجل ووزنه .

(٩ – ٩) اختبار جودة العلاقة الخطية :

نعلم أننا إذا قبلنا الفرض الصفرى $\beta = 0$ ، فإننا نحكم بعدم وجود علاقة خطية بين المتغيرين ولا نحتاج حينئذ إلى إجراء أى اختبارات أخرى تتعلق بخطية هذه العلاقة . أما إذا رفضنا هذا الفرض فإننا نحكم بوجود علاقة خطية بين المتغيرين لأن الانحدار الخطى يكون قد فسر جزءا ذا دلالة من الاختلاف الكلى . غير أن هذا لا يعنى أن العلاقة الخطية هى أحسن علاقة تصف العلاقة الحقيقية بين المتغيرين ، فقد تكون هناك علاقة أخرى مثل $ص = \alpha + \beta_1 س + \beta_2 س^2$ تفضل العلاقة الخطية فى ذلك . ويستلزم الأمر هنا الاستمرار فى تحليل البيانات لتقدير درجة جودة العلاقة الخطية ، وذلك بتقييم الاختلاف غير المفسر الذى يعبر عن الانحراف عن الخطية ، فإذا كان هذا الاختلاف غير ذى دلالة أى يرجع إلى العوامل العشوائية فإن العلاقة الخطية تكون هى أحسن العلاقات ، أما إذا كان هذا الاختلاف ذا دلالة فإن العلاقة الخطية لا تكون هى العلاقة المثلى بين المتغيرين .

ولكن كيف نقيّم الاختلاف المعبر عن الانحراف عن الخطية ؟ إننا نحتاج هنا إلى مقياس نقيس به دلالة هذا الانحراف . ولتحقيق هذا الغرض ينبغى تصميم تجربة نحصل منها على بيانات بحيث يناظر كل قيمة من القيم السينية مجموعة من القيم الصادية كما فى المثال (٩ – ٨) السابق ، لأن وجود هذه المجموعات الصادية يتيح لنا فرصة إيجاد الاختلاف **داخل** هذه المجموعات وهو الذى يعبر عن خطأ التجريب ، وبالتالى يمكن اتخاذه معيارا لمدى دلالة الانحراف عن الخطية .

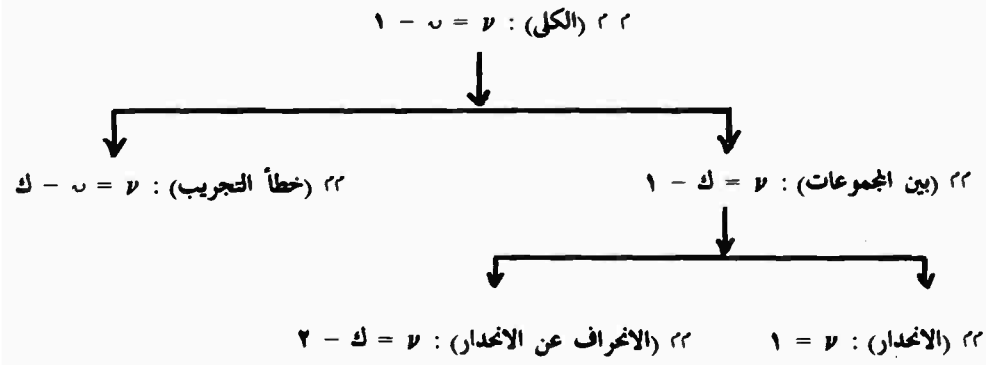
وعلى ذلك ، وعلى فرض أن لدينا بيانات من مثل هذه التجربة نبدأ بخطوة هامة هى تحليل التباين للمجموعات الصادية ، أى فصل الاختلاف الكلى فى القيم الصادية (كالمعتاد) إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف بين المجموعات وتعبر الثانية عن الاختلاف داخل المجموعات (خطأ التجريب) ، وهذه الخطوة تتخذ الشكل المعتاد الآتى :

م م (ص) = م م (بين المجموعات) + م م (داخل المجموعات)
بدرجات حرية ن − ١ ، ك − ١ ، ن − ك على الترتيب .

وجدير بالإشارة هنا إلى أن المتغير المستقل س فى الانحدار هو متغير كمى بينما المتغير المستقل (عامل التجريب) فى تحليل التباين هو متغير نوعى . وحين نجرى تحليل التباين للانحدار نعتبر أن القيم العددية للمتغير س هى مجرد إشارات تعبر عن مستويات عامل التجريب .

ولما كان الهدف من تحليل الانحدار اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط مستقيم فإن الاختلاف الذى ينبغى تحليله هو الاختلاف **بين المجموعات** . وعلى ذلك فإن الخطوة الثانية هى تحليل الاختلاف بين المجموعات إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف المفسر : م م (الانحدار الخطى) وتعبر الثانية عن الاختلاف غير المفسر : م م (الانحراف عن الانحدار الخطى) وهذه الخطوة تتخذ الشكل الآتى :
م م (بين المجموعات) = م م (الانحدار الخطى) + م م (الانحراف عن خط الانحدار)
بدرجات حرية ك − ١ ، ١ ، ك − ٢ على الترتيب .

الشكل (٩ − ٨) الآتى يلخص خطوتى التحليل السابق ذكرهما .



الشكل (٩ − ٨) : خطوتا اختبار جودة الانحدار الخطى

كما أن الجدول (٩ – ٨) الآتى هو جدول التباين لعمليتى التحليل .

الجدول (٩ – ٨)

مصدر التباين	مجموع المربعات	درجات الحرية
بين المجموعات	ح ن و (صَ و – صَ)٢	ك – ١
الانحراف الخطى	ح ن و (صُ و – صَ)٢	١
الانحراف عن خط الانحدار	ح ن و (صَ و – صُ و)٢	ك – ٢
خطأ التجريب (داخل المجموعات)	ح ح (صـ ر و – صَ و)٢	ن – ك
الكلى	ح ح (صـ ر و – صَ)٢	ن – ١

من هذا التحليل نستطيع أن نختبر ثلاثة أمور هى :

(أولا) اختبار دلالة الاختلاف المشاهد بين المجموعات :

أى اختبار ما إذا كانت متوسطات المجموعات الصادية تختلف باختلاف القيم السينية . والاختبار الذى يصلح لذلك هو اختبار ف المعتاد حيث

$$ف = \frac{\text{التباين بين المجموعات}}{\text{التباين داخل المجموعات}}$$ بدرجتى حرية ك – ١ ، ن – ك (٢٥)

والفرض الصفرى هنا هو أن المتوسطات متساوية جميعها . فإذا قبلنا هذا الفرض نحكم بعدم وجود أى علاقة بين التغير فى سـ والتغير فى صـ ويتوقف البحث عند هذا الحد . أما إذا رفضنا الفرض الصفرى فنحكم بأن التغير فى سـ يؤثر فى التغير فى صـ وينبغى حينئذ أن نستمر فى البحث بحسب الخطوة الثانية .

(ثانيا) بحث العلاقة الخطية بين المتغيرين :

(ا) اختبار وجود علاقة خطية :

أى اختبار الفرض الصفرى $\beta = 0$ ضد الفرض $\beta \neq 0$. وكما فعلنا فى المثال (٩ – ٨) نستخدم الصيغة (٢١) وهى :

$$\text{ف} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

مع ملاحظة أن الاختلاف غير المفسر هو ذلك الاختلاف الذى لا يعتمد على المتغير س وهو يشمل الاختلاف الناشىء عن الانحراف عن خط الانحدار كما يشمل الاختلاف الناشىء عن خطأ التجريب ، وعلى ذلك فإن التعويض فى هذه الصيغة من بيانات الجدول (٩ – ٨) يكون كالآتى :

$$\text{ف} = \frac{\text{م م (الانحدار الخطى)} \div 1}{[\text{م م (الانحراف عن الخطية)} + \text{م م (خطأ التجريب)}] \div (\text{ن}-2)} \quad (٢٦)$$

بدرجتى حرية ١ ، ن – ٢ .

وكما سبق القول فى (أولا) ليس هناك ما يدعو للقيام بهذا الاختبار أو بالاختبار الذى سيرد فى (ب) إلا إذا ثبت من الخطوة السابقة دلالة الاختلاف بين المجموعات الصادية .

(ب) اختبار جودة العلاقة الخطية :

أى اختبار دلالة الانحراف عن خط الانحدار مقدرا بمجموع مربعات الانحرافات $\text{ص}_{\text{و}} - \hat{\text{ص}}_{\text{و}}$ ، أو بمعنى آخر اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط الانحدار أو قريبة منه أم تنتشر بعيدة عنه أبعادا جوهرية . والاختبار الذى يصلح لذلك هو اختبار ف بالصورة

$$ف = \frac{\text{تباين الانحراف عن الخطية}}{\text{تباين خطأ التجريب}} \qquad (٢٧)$$

بدرجتى حرية ك – ٢ ، ن – ك .

فى هذا الاختبار يلعب التباين المقدر لخطأ التجريب دورا رئيسيا كمعيار يقاس بالنسبة إليه الانحراف عن الخطية ، وهذا هو السبب فى ضرورة أن تصمم التجارب التى تهدف إلى قياس جودة العلاقات الخطية بحيث يكون لكل قيمة س قيمتان أو أكثر من قيم ص وإلا ما استطعنا الحصول على هذا المعيار .

ملاحظة :

إذا وجدنا من الخطوة (أ) أن الانحدار الخطى ذو دلالة ووجدنا من الخطوة (ب) أن الانحراف عن الخطية غير ذى دلالة ، نحكم بأن العلاقة الخطية هى علاقة جيدة وتصف بجدارة العلاقة الحقيقية بين المتغيرين . أما إذا وجدنا أن كلا من الانحدار الخطى والانحراف عن الخطية ذو دلالة فنحكم بأنه بالرغم من وجود علاقة خطية بين المتغيرين إلا أنها ليست أحسن العلاقات التى تعبر عن حقيقة العلاقة بينهما وعلينا إذا أردنا أن نبحث عن علاقة أفضل .

مثال (٩ – ٩) :

ابحث خطية العلاقة بين الطول والوزن مستخدما بيانات المثال (٩ – ٨) .

الحل :

(أولا) نبدأ بتحليل التباين للمتغير ص للحكم على دلالة الاختلاف بين متوسطات المجموعات الصادية .

م م (الكلى) = ٥٢٦٦,٦٧ سبق إيجاده بالمثال (٩–٨)

$$م\ م\ (\text{بين المجموعات}) = \frac{٣٦٥^{٢}}{٣} + \frac{٥٢٥^{٢}}{٤} + \frac{٢٩٥^{٢}}{٢} + \frac{٥١٥^{٢}}{٣} - \frac{١٧٠٠^{٢}}{١٢}$$

$$= ٤٤٠٢,٠٨ \qquad \nu = ٣$$

$$\therefore م\ م\ (\text{خطأ التجريب}) = ٥٢٦٦,٦٧ - ٤٤٠٢,٠٨ = ٨٦٤,٥٩ \qquad \nu = ٨$$

$$\text{من (٢٥) : ف} = \frac{٤٤٠٢,٠٨ \div ٣}{٨٦٤,٥٩ \div ٨} = ١٣,٥٧٧^{**}$$

وبما أن ف $_{٠,٠١\ [٣،٨]}$ = ٧,٥٩ نرفض الفرض الصفرى عن تساوى المتوسطات عند مستوى الدلالة ٠,٠١ ونحكم بأن أوزان الرجال ليست مستقلة عن أطوالهم . ومادام الأمر كذلك نستمر فى التحليل .

(ثانيا) نقوم بتحليل الانحدار للاختلاف بين المجموعات كما يلى :

$$م\ م\ (\text{الانحدار الخطى}) = ٤٣٤١,٨٧٠٨ \text{ سبق إيجاده بالمثال (٩-٨)} \qquad \nu = ١$$

$$\therefore م\ م\ (\text{الانحراف عن الخطية}) = م\ م\ (\text{بين المجموعات}) - م\ م\ (\text{الانحدار الخطى})$$

$$= ٤٤٠٢,٠٨ - ٤٣٤١,٨٧٠٨$$

$$= ٦٠,٢٠٩٢ \qquad \nu = ٢$$

بضم هذه النتائج إلى نتائج الخطوة السابقة ينتج جدول التباين الآتى :

الجدول (٩ - ٩)

مصدر التباين	مجـ ٢	د ح	ط ٢
بين المجموعات	٤٤٠٢,٠٨	٣	١٤٦٧,٣٦
الانحدار الخطى	٤٣٤١,٨٧٠٨	١	٤٣٤١,٨٧٠٨
الانحراف عن الخطية	٦٠,٢٠٩٢	٢	٣٠,١٠٤٦
خطأ التجريب	٨٦٤,٥٩	٨	١٠٨,٠٧٣٨
الكل	٥٢٦٦,٦٧	١١	

(أ) اختبار وجود علاقة خطية

$$\text{من (٢٦) : } ف_{ى} = \frac{١ \div ٤٣٤١,٨٧٠٨}{١٠ \div (٨٦٤,٥٩ + ٦٠,٢٠٩٢)} = ٤٦,٩٥^{**}$$

بما أن $ف_{٠,٠١ [١ ، ١٠]} = ١٠,٠٠$ نرفض الفرض الصفرى أن $\beta = ٠$ ونحكم بوجود علاقة خطية بين أطوال الرجال وأوزانهم .

(ب) اختبار دلالة الانحراف عن خط الانحدار

$$\text{من (٢٧) : } ف_{ى} = \frac{٣٠,١٠٤٦}{١٠٨,٠٧٣٨} < ١$$

وإذن نقبل الفرض الصفرى بعدم وجود انحراف ذى دلالة عن خط الانحدار .

الخلاصة :

من هذه التجربة نستخلص أن العلاقة الخطية التى تمثلها معادلة الانحدار ص = - ١٧٩,٣٥ + ٥,٠٢٩ س تعبر بجدارة عن العلاقة الحقيقية بين أطوال وأوزان الرجال . ومع ملاحظة أن معامل التحديد

$$ر^{٢} = \frac{[م_{صـ}(س ، ص)]^{٢}}{م_{م}(س) . م_{م}(ص)}$$

$$= \frac{(٨٦٣,٣٤)^{٢}}{١٧١,٦٦٧ \times ٥٢٦٦,٦٧} = ٠,٨٢٤٤$$

نرى أن العلاقة الخطية تفسر حوالى ٨٢,٤٪ من الاختلاف الكلى فى قيم المتغير ص ، وهذا يدعم القول بخطية العلاقة بين المتغيرين .

(٩ - ١٠) ملاحظات عن افتراضات الانحدار :

إن صحة الاستنتاجات الإحصائية تتوقف على مدى انطباق الافتراضات التى وضعت فى تعريف النموذج الذى بنيت عليه الدراسة على الموقف التجريبى الذى تتناوله هذه الدراسة . ولذلك ينبغى أن نتفهم جيدا ما تعنيه هذه الافتراضات وما تتطلبه من شروط إجرائية تحسبا من العواقب التى قد تنجم عن وجود تناقضات ذات بال بين الافتراضات الموضوعة والظروف الفعلية لعملية التجريب . كما ينبغى أن نكون على استعداد لتغيير افتراض أو تعديله إذا اتضح لنا عدم إمكانية تحققه عمليا ولو بشىء من التقريب ، على أن نراعى ما يستلزمه هذا التغيير أو التعديل بالنسبة لما نقدمه من استنتاجات .

لنعتبر الافتراضات الثلاثة التى استخدمناها فى الانحدار الخطى البسيط ولنبدأ بالافتراض الأول . إن هذا الافتراض يتضمن أن يكون المتغير س متغيرا غير عشوائى

وهذا يعنى أن الباحث يتحكم تجريبيا فى هذا المتغير ويستطيع تسجيل القيم التى يدخلها فى بياناته بدقة تامة . غير أنه فى كثير من المواقف التجريبية لا يتحقق هذا الافتراض فقد يكون الباحث مهتما بالحصول على بيانات عن المتغير س- دون أن يكون متحكما فيه أى بصرف النظر عن كون هذا المتغير عشوائيا أو غير عشوائى مادامت هذه البيانات تعبر فى نظر الباحث عن قيم نموذجية لهذا المتغير . وفى بعض الدراسات يضطر الباحث إلى اختيار مشاهدات تقع فى مدى معين محدد من قبل أو يأخذ منها قيما معينة محددة مسبقا . فى مثل هذه الحال ينبغى للباحث أن يستبدل بهذا الافتراض افتراضا آخر يتلاءم مع الموقف الذى يتناوله ، كأن يضع الافتراض «المتغير س- قيمه مشروطة » وهذا الافتراض الجديد لا يؤثر فى سلامة استخدام طريقة المربعات الصغرى إلا أن الاستنتاجات الإحصائية التى نخرج بها عن المجتمع الذى ندرسه ينبغى أن تكون مشروطة بالنسبة للمتغير س- أى لا يجوز تعميمها لقيم غير ممثلة فى البيانات التى استخدمت فى الدراسة أو ليس لها نفس خصائصها .

أما الافتراض الثانى فيعنى أن الصيغة الحقيقية للعلاقة بين المتغيرين هى الصيغة الخطية ، وعلى هذا لا تكون استنتاجاتنا صحيحة إلا إذا كان لدينا ما يضمن سلامة هذه الصيغة ، ويساعدنا فى ذلك الأسلوب المذكور فى نتيجة البند (٩ – ٥) لاختبار خطية هذه العلاقة كما يساعدنا الأسلوب المبين بالبند (٩ – ٩) لاختبار دلالة الانحراف عن خط الانحدار . فإذا تبين لنا عدم انطباق هذه الصيغة فى موقف ما ينبغى أن نبحث عن صيغة أخرى تناسبه .

وبالنسبة للافتراض الثالث فهو يعنى أنه عند أى قيمة ثابتة من المتغير س- تتصف القيم التى يأخذها المتغير ص- بما يلى :

(١) قيم ص- مستقلة ، أى أن صغر أو كبر الخطأ العشوائى فى أحدها لا يؤثر فى مقدار الخطأ فى القيم الأخرى . وهذه الصفة يمكن تحقيقها عمليا بإحكام عملية التجريب وخاصة فيما يتعلق بعشوائية العينة .

(٢) قيم صـ لها تباين ثابت σ^2 ، وهذه الصفة تتحقق تلقائيا إذا كانت هذه القيم هى مشاهدات مستقلة من نفس المجتمع .

(٣) قيم صـ تتوزع توزيعا معتدلا .

وجدير بالإشارة هنا إلى أنه بصفة عامة تكون النماذج المستخدمة بما يصاحبها من فروض هى نماذج نظرية قلما تتحقق فى الواقع العملى إلا على وجه التقريب . ونحن إذ نستخدم هذه النماذج نتعلق بالأمل فى ألا تؤدى بنا هذه الحقيقة إلى وجود فروق كبيرة بين التقديرات التى نحصل عليها منها وبين القيم الحقيقية للبارامترات التى نبحث عنها ، كما نتعلق بالأمل بأن تكون الإجراءات التى اتخذناها فى عملية التقدير هى إجراءات مناسبة حتى إذا كان الموقف الذى نتناوله لا يحقق الفروض تحقيقا تاما .

(٩ - ١١) استخدامات الانحدار :

لعله من المناسب الآن أن نبرز الأغراض التى يستخدم الانحدار الخطى البسيط من أجلها . ومن الدراسة التى مرت بنا فى هذا الفصل يمكننا تلخيص هذه الأغراض فيما يلى :

(١) التنبؤ بقيم متغير عشوائى صـ بمعلومية متغير رياضى سـ مع تقدير درجة دقة هذا التنبؤ . هذا مع ملاحظة أن خط الانحدار هو خط مستقيم يمتد بغير نهاية من الطرفين . ولكننا فى عملية التنبؤ لا يجوز التنبؤ بقيم صادية مناظرة لقيم سينية تخرج كثيرا عن مدى القيم التى استخدمت فى إنشاء هذا الخط إلا إذا كان لدينا ما يبرر ذلك . فمثلا إذا أوجدنا خط انحدار لأطوال الذكور بين العمر ١٠ والعمر ١٥ فمن الخطأ استخدام هذا الخط للتنبؤ بأطوال الرجال فى أعمار فوق العشرين لأن الطول يتوقف عند بلوغ سن النضج .

(٢) دراسة طبيعة العلاقة بين متغيرين سـ ، صـ وشكل المنحنى أو الدالة التى تعبر عن هذه العلاقة .

(٣) تفسير بعض الاختلاف فى قيم المتغير صـ بدلالة الاختلاف فى قيم سـ على أساس اتخاذ المتغير سـ كضابط إحصائى ststistical control .

(٤) دراسة ما إذا كان التغير فى سـ **يسبب** ولو جزئيا التغير فى صـ ، مع ملاحظة أن دراسة السببية تحتاج إلى أكثر من إثبات وجود علاقة ذات دلالة بين المتغيرين ، إذ أن هذه العلاقة قد تكون ناشئة عن وجود متغيرات أو عوامل أخرى تؤثر فى المتغيرين سـ ، صـ معا . فمثلا فى الأحياء المزدحمة من بعض المدن الكبيرة نجد علاقة ذات دلالة بين ازدحام السكان والتدرن الرئوى (السل) فهل هذا يعنى أن الازدحام سبب فى الإصابة بالتدرن الرئوى ؟ لا نستطيع الإجابة بالإيجاب أو النفى عن هذا السؤال إذا لاحظنا أنه فى الأحياء المزدحمة بصفة عامة يكون مستوى الحياة منخفضا وبالتالى يكون هناك سوء تغذية للسكان ، وقد يكون سوء التغذية هو السبب الحقيقى لهذا المرض . إن تقرير السببية أمر متروك للباحث يفتى فيه بما لديه من خبرة ومعلومات عن المتغيرات التى يتناولها مستعينا بما يستخلصه من التحليل الاحصائى .

(٥) يستخدم الانحدار أيضا فى أنواع أخرى من التحليل الإحصائى منها تحليل التغاير حيث يكون الهدف معرفة مدى تأثير عامل نوعى على متغير عددى صـ بعد استبعاد أثر متغير عددى سـ مرتبط بالمتغير صـ ، وهذا ما سنتناوله فى فصل لاحق .

تمارين (٩ - ٢)

(اعتبر أن افتراضات الانحدار متوفرة) .

(١) فى تجربة عن تأثر طاقة التمثيل الغذائى بدرجة الحرارة على نوع من الطيور فى فترة ضوئية ثابتة مدتها ١٠ ساعات ، اختيرت أربع درجات حرارة هى ٥ ، ١٠ ، ١٥ ، ٢٠ وعرضت خمسة طيور لكل من هذه الدرجات ثم حسبت مقادير طاقة التمثيل الغذائى لكل من هذه الطيور وسجلت بالجدول الآتى . أوجد معادلة

الانحدار الخطى لطاقة التمثيل الغذائى على درجة الحرارة واختبر جودة العلاقة الخطية بين هذين المتغيرين .

	درجات الحرارة (س)			
	٥	١٠	١٥	٢٠
طاقة التمثيل الغذائى (ص)	٢٣,٢	٢٤,٢	١٨,٣	١٥,٨
	٢٣,٥	٢٤,٢	١٨,٨	١٥,٢
	٢٣,٤	٢٤,١	١٨,٦	١٥,٦
	٢٣,١	٢٤,٣	١٨,٧	١٥,٣
	٢٣,٦	٢٤,٢	١٨,١	١٥,١

لكل من العينات الثلاث الآتية أوجد معادلة الانحدار الخطى واختبر جودة العلاقة الخطية :

(٢)

	س			
	٠	١	٢	٤	٦
ص	١-	٠	١	٨	١١
	١	٢	٥	١٢	١٣

(٣)

	س			
	١	٢	٣	٤
ص	٠	٥	٩	١٠
	٢	٧	١٣	١٤