الفصل الثانى عشر

الانحدار والارتباط الخطى المتعدد

MULTIPLE LINEAR REGRESSION AND CORRELATION

سنعتمد فى هذا الفصل على الحاسب الإلكترونى فى إجراء الحسابات اللازمة لحل مشكلات الانحدار والارتباط تخففا من الأعباء الحسابية الضخمة التى يتطلبها التحليل . إلا أن هذا لا يعفينا بل يوجب علينا الإحاطة بالأسس والافتراضات والمفاهيم التى يبنى عليها التحليل تحسبا من أخطاء ومزالق التطبيق .

أولا : الانحدار الخطى المتعدد

(١٢ – ١) الانحدار الخطى المتعدد كامتداد للانحدار الخطى البسيط :

فى تناولنا للانحدار الخطى البسيط فى الفصل التاسع من هذا الكتاب افترضنا أن لدينا متغيرا عشوائيا ص نرغب فى التنبؤ بقيمه عن طريق قيم متغير غير عشوائى س على أساس أن العلاقة بين هذين المتغيرين هى علاقة خطية :

ص = $\alpha + \beta$ س   حيث α ، β بارامتران مجهولان ،

وعلى أساس أنه عند أى قيمة ثابتة س يكون للمتغير ص توزيع معتدل متوسطه $\alpha + \beta$ س يتوقف على قيمة س ، وتباينه عدد ثابت σ^2 لا يتوقف على س .

إلا أنه فى كثير من التطبيقات يرى الباحث أن استخدام متغير واحد س لا يصلح للتنبؤ بقيم المتغير ص بدقة كافية ، ويأمل أن يحصل على تنبؤات أكثر دقة إذا استخدم عدة متغيرات $\text{س}_{١}$ ، $\text{س}_{٢}$ ، ... ، $\text{س}_{\text{ك}}$ يشعر بخبرته فى ميدان عمله أن لها دورا فى عملية التنبؤ . وإذا تبنينا افتراضات مماثلة لتلك التى تبنيناها فى الانحدار الخطى البسيط من أن هذه المتغيرات غير عشوائية وترتبط بالمتغير العشوائى ص بعلاقة خطية :

$$\text{ص} = \alpha + \beta_{١}\,\text{س}_{١} + \beta_{٢}\,\text{س}_{٢} + \ldots + \beta_{\text{ك}}\,\text{س}_{\text{ك}} \qquad (١)$$

حيث α ، $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{\text{ك}}$ بارامترات مجهولة ، ومن أنه عند أى مجموعة ثابتة $\text{س}_{١}$ ، $\text{س}_{٢}$ ، ... ، $\text{س}_{\text{ك}}$ يكون للمتغير ص توزيع معتدل ذو تباين ثابت $\sigma^{٢}$ ، فإن الانحدار يسمى فى هذه الحال بالانحدار الخطى المتعدد أو اختصارا بالانحدار الخطى كما تسمى البارامترات $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{\text{ك}}$ بمعاملات الانحدار الجزئية **partial regression coefficients** ، وتكون دراسة هذا الانحدار مجرد تعميم للمفاهيم والأساليب والصيغ والاختبارات المستخدمة فى الانحدار الخطى البسيط ولا تختلف عنها إلا فى درجة المشقة فى تناول البيانات ، كما تختلف بطبيعة الحال فى درجات الحرية التى تعتمد على عدد الثوابت فى معادلة الانحدار التى تقدر من العينة . وإذا كان القارىء قد اطمأن لمعرفته بأسس وأساليب الانحدار الخطى البسيط فسوف لا يجد صعوبة تذكر فى تناول الانحدار الخطى المتعدد كما يتبين من الفقرات الآتية .

(ا) ايجاد معادلة الانحدار :

فى الانحدار الخطى البسيط نقوم بتجربة نحصل منها على عينة من ن من المشاهدات ($\text{س}_{\text{ر}}$ ، $\text{ص}_{\text{ر}}$) تأخذ الشكل الآتى :

س١	س٢	...	سن
ص١	ص٢	...	صن

وتستخدم هذه البيانات فى تقدير البارامترين المجهولين α ، β بقيمتين ا ، ب توطئة لكتابة معادلة الانحدار :

ص̂ = ا + ب س

التى تمثل « أحسن خط » يلائم تلك المشاهدات من وجهة نظر مبدأ المربعات الصغرى ، الذى يستلزم كما نعلم تحديد القيمتين أ،ب اللتين تجعلان الدالة :

مج (صر − ص̂ر)^2 = مج (صر − α − β سر)^2

نهاية صغرى . وتحقيق هذا الفرض يستلزم بدوره إجراء عملية التفاضل الجزئى بالنسبة إلى كل من α ، β ومساواة كل من الناتجين بالصفر للحصول على معادلتين خطيتين فى هذين المجهولين تسميان بالمعادلتين المعتادتين وهما :

مج ص = ن α + β مج س

مج س ص = α مج س + β مج س^2

وحل هاتين المعادلتين معا يعطينا القيمتين ا ، ب المطلوبتين .

وهذه الخطوات هى بذاتها الخطوات التى تتخذ عند تناول الانحدار الخطى المتعدد . إذ نقوم بتجربة نحصل منها على عينة من ن من المشاهدات (س١ر ، س٢ر ، .... ، سكر ، صر) حيث ر = ١ ، ٢ ، ... ، ن تأخذ الشكل المبين بالجدول (١٢ − ١) الآتى :

الجدول (١٢ - ١)

بيانات الانحدار الخطى المتعدد

المشاهدات	ص	$س_{١}$	$س_{٢}$	....	$س_{ك}$
(١)	$ص_{١}$	$س_{١١}$	$س_{٢١}$	...	$س_{١ك}$
(٢)	$ص_{٢}$	$س_{١٢}$	$س_{٢٢}$	...	$س_{٢ك}$
...	...	...	...	...	
(ر)	$ص_{ر}$	$س_{١ر}$	$س_{٢ر}$	...	$س_{رك}$
...	...	...	...	...	...
(ن)	$ص_{ن}$	$س_{١ن}$	س	...	$س_{نك}$

ولتقدير البارامترات المجهولة α ، $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{ك}$ من هذه البيانات نستخدم مبدأ المربعات الصغرى لايجاد القيم ا ، $ب_{١}$ ، $ب_{٢}$ ، .... ، $ب_{ك}$ توطئة لكتابة معادلة الانحدار :

$$\hat{ص} = ا + ب_{١}\, س_{١} + ب_{٢}\, س_{٢} + \dots + ب_{ك}\, س_{ك} \qquad (٢)$$

وتحقيق هذا المبدأ يستلزم أن تؤخذ القيم أ،$ب_{.}$ ، . . . ، $ب_{ن}$ بحيث تكون الدالة

$$مج\,(ص_{ر} - \hat{ص}_{ر})^{٢} = مج\,(ص_{ر} - \alpha - \beta_{١}\, س_{١ر} - \beta_{٢}\, س_{٢ر} - \dots - \beta_{ك}\, س_{رك})^{٢}$$

نهاية صغرى ، وهذا يستلزم بدوره إجراء عملية التفاضل الجزئى بالنسبة إلى كل من α ، $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{ك}$ ومساواة النواتج بالصفر للحصول على ك + ١ من المعادلات الخطية فى هذه المجاهيل تسمى بالمعادلات المعتادة . وحل هذه المعادلات معا يعطينا التقديرات المطلوبة أ ، $ب_{١}$ ، $ب_{٢}$ ، ... ، $ب_{ك}$ وعددها ك + ١ .

ولنا أن نتصور هنا مدى المشقة التى نلقاها فى حساب المجاميع ومجاميع المربعات ومجاميع حواصل الضرب المطلوبة لوضعها فى هذه المعادلات ثم حل المعادلات ذاتها ، خاصة إذا كان عدد المتغيرات السينية أكبر من اثنين . وحتى فى حالة وجود متغيرين $س_{١}$ ، $س_{٢}$ حيث نستهدف الوصول إلى معادلة الانحدار

$$\hat{ص} = ا + ب_{١}\, س_{١} + ب_{٢}\, س_{٢}$$

يكون علينا حل المعادلات المعتادة الآتية :

$$\text{مج}\, ص = ن\, \alpha + \beta_{١}\, \text{مج}\, س_{١} + \beta_{٢}\, \text{مج}\, س_{٢}$$

$$\text{مج}\, س_{١}\, ص = \alpha\, \text{مج}\, س_{١} + \beta_{١}\, \text{مج}\, س_{١}^{٢} + \beta_{٢}\, \text{مج}\, س_{١}\, س_{٢} \qquad (٣)$$

$$\text{مج}\, س_{٢}\, ص = \alpha\, \text{مج}\, س_{٢} + \beta_{١}\, \text{مج}\, س_{١}\, س_{٢} + \beta_{٢}\, \text{مج}\, س_{٢}^{٢}$$

وهذا الحل يحتاج أولا إلى حساب ثمانية مجاميع هى مج ص ، مج $س_{١}$ ، مج $س_{٢}$ ، مج $س_{١}^{٢}$ ، مج $س_{٢}^{٢}$ ، مج $س_{١}\, س_{٢}$ ، مج $س_{١}$ ص ، مج $س_{٢}$ ص مما يحتاج إلى كثير من الجهد والوقت . ولذلك نلجأ إلى الحاسب الالكترونى ليقوم عنا بكل هذه العمليات ويعطينا التقديرات المطلوبة بكل سرعة ودقة .

(ب) إيجاد الخطأ المعيارى لتقدير ص من خط الانحدار :

فى الانحدار الخطى البسيط استخدمنا مقياسا رأينا أهميته البالغة فى عمليات الاستنتاج الإحصائى هو الخطأ المعيارى للتقدير الذى رمزنا له بالرمز $ع_{ص.س}$ وعرفناه فى البند (٩ — ٤) بأنه الانحراف المعيارى للقيم المعيارية المشاهدة $ص_{ر}$ حول القيم $\hat{ص}_{ر}$ المقدرة من معادلة الانحدار أى عرفناه كالآتى :

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢}\, \text{مج}\, (ص_{ر} - \hat{ص}_{ر})^{٢} \qquad \text{بدرجات حرية } ن - ٢$$

وحين حللنا الاختلاف الكلى فى القيم الصادية وهو مج $(ص_ر - \bar{ص})^٢$ إلى مركبتين تعبر الأولى وهى مج $(\hat{ص}_ر - \bar{ص})^٢$ عن الاختلاف الناشىء عن الانحدار (الاختلاف المفسر) وتعبر الثانية وهى مج $(ص_ر - \hat{ص}_ر)^٢$ عن الاختلاف المتبقى الناشىء عن الانحراف عن خط الانحدار (الاختلاف غير المفسر) أمكننا كتابة مربع الخطأ المعيارى للتقدير كالآتى :

$$ع^٢_{ص.س} = \frac{١}{ن - ٢} \text{ مج } (ص_ر - \hat{ص}_ر)^٢ = \frac{١}{ن - ٢} \times \text{الاختلاف غير المفسر}$$

بنفس المنطق نستخدم نفس التعريف فى حالة الانحدار الخطى المتعدد مع مراعاة درجات الحرية ، فنرمز للخطأ المعيارى لتقدير ص من خط الانحدار بالرمز

$ع_{ص. س_١ س_٢ ... س_ك}$ حيث :

$$ع^٢_{ص.س_١ س_٢ ... س_ك} = \frac{١}{ن - (ك + ١)} \times \text{الاختلاف غير المفسر} \qquad (٤)$$

حيث ن حجم العينة ، ك عدد المتغيرات السينية ومع ملاحظة أن ك + ١ هو عدد الثوابت فى معادلة الانحدار التى تقدر من العينة .

ملاحظة (١) :

ليس من المناسب هنا تقديم الصيغة العامة للخطأ المعيارى للتقدير ويكفينا أن نفهم المعنى الذى يتضمنه ، أما حساب قيمته فنتركه للحاسب الالكترونى . إلا أنه حين يكون هناك متغيران تنبؤيان اثنان فقط (ك = ٢) فيمكن إثبات أن

$$ع^٢_{ص. س_١ س_٢} = \frac{١}{ن - ٣} (\text{مج ص}^٢ - \text{أ مج ص} - ب_١ \text{ مج س}_١ \text{ص} - ب_٢ \text{ مج س}_٢ \text{ص}) \quad (٥)$$

وهذه الصيغة هى امتداد للصيغة (١٠) بالبند (٩ – ٤) التى تتناول الحالة التى يكون لدينا فيها متغير تنبؤى واحد .

هذا مع ملاحظة أن الصيغة العامة للخطأ المعيارى تكتب بأسلوب رياضى يتطلب معرفته دراسة مسبقة لموضوع المصفوفات ، والواقع أن الدراسة النظامية للانحدار المتعدد تعتمد برمتها على هذا الأسلوب غير أننا فى التطبيق العملى لا نحتاج إليه .

(ح) اختبار دلالة الانحدار ككل :

فى الانحدار الخطى البسيط اهتممنا باختبار دلالة الانحدار أى باختبار وجود علاقة خطية بين المتغير التابع ص والمتغير المستقل س وهذا يكافىء اختبار الفرض الصفرى $\beta = ٠$ ضد الفرض $\beta \neq ٠$ واستخدمنا لهذا الغرض اختبار ت بالصيغة (١١) بالبند (٩ – ٥ – أولا) مع وضع $\beta =$. أو اختبار ف بالصيغة (٢١) بالبند (٩ – ٧) أو بالصيغة المكافئة (٢٤) ، بشرط توفر شروط الانحدار .

كذلك يهمنا فى حالة الانحدار الخطى المتعدد اختبار وجود علاقة خطية بين المتغير التابع ص والمتغيرات المستقلة $س_{١}$ ، $س_{٢}$ ، . . . ، $س_{ك}$ وهذا الاختبار يكافىء اختبار الفرض الصفرى $\beta_{١} = \beta_{٢} = \cdots = \beta_{ك} = ٠$ وفى هذه الحالة نستخدم أى من الصيغتين (٢١) أو (٢٤) المذكورتين مع مراعاة استخدام درجات الحرية المناسبة . وهاتان الصيغتان هما :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بدرجتى حرية ك ، ن – ك – ١} \qquad (٦)$$

$$\text{أو } ف = \frac{ر^{٢} \div ك}{(١ - ر^{٢}) \div (ن - ك - ١)} \quad \text{بدرجتى حرية ك ، ن – ك – ١} \qquad (٧)$$

حيث ن حجم العينة ، ك عدد المتغيرات السينية ، $ر^{٢}$ هو معامل التحديد الذى يعبر كالمعتاد عن نسبة الاختلاف فى المتغير ص التى يفسرها خط الانحدار وسنرمز له هنا بالرمز $ر^{٢}_{ص(٢١...ك)}$ حيث

$$ر^{٢}_{ص(٢١...ك)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (٨)$$

(د) اختبار دلالة معاملات الانحدار الجزئية :

فى الانحدار الخطى البسيط يكون لدينا معامل انحدار واحد β ، ويمكن أن نختبر ما إذا كان هذا المعامل يأخذ قيمة ما نفترضها بواسطة الإحصاءة (١١) المقدمة بالبند (٩ – ٥ – أولا) وهى

$$ت = \frac{B - \beta}{ع_{ص.س} / \sqrt{م م (س)}} = \frac{B - \beta}{ع\,(ب)} \qquad (٩)$$

التى يكون لها توزيع ت بدرجات حرية ن – ٢ بشرط تحقق افتراضات الانحدار . ويلاحظ أن مقام هذا الكسر وهو ع (ب) = $ع_{ص.س} / \sqrt{م م(س)}$ هو تقدير للانحراف المعيارى للمتغير العشوائى B الذى تعتبر ب إحدى قيمه المشاهدة – وحيث ب تقدير معامل الانحدار β الذى نجده من العينة . والمتغير B هو متغير معتدل وسطه الحسابى β وانحرافه المعيارى σ (ب) يقدر من العينة بالمقدار ع (ب) .

أما فى الانحدار الخطى المتعدد فلدينا ك من معاملات الانحدار الجزئية $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{ك}$ (بالإضافة إلى البارامتر α) ويهمنا أن نختبر ما إذا كانت هذه المعاملات تأخذ قيما معينة . وبنفس منطق البند (٩ – ٥) نعتبر أن القيمة $ب_{و}$

(و = ١ ، ٢ ، ... ، ك) التى نحصل عليها من عينة لتقدير البارامتر $\beta_{و}$ ، هى إحدى القيم المشاهدة من متغير عشوائى $B_{و}$ ذى توزيع معتدل متوسطه $\beta_{و}$ وانحرافه المعيارى σ (ب$_{و}$) نقدره بقيمة سنرمز لها بالرمز ع (ب$_{و}$) وبالتالى يكون للإحصاءة

$$ت_{و} = \frac{B_{و} - \beta_{و}}{ع\ (ب_{و})} \qquad و = ١ ، ٢ ، ... ، ك \qquad (١٠)$$

توزيع ت بدرجات حرية ن − (ك + ١) بشرط تحقق افتراضات الانحدار . وهذه الإحصاءة تصلح لاختبار أن يأخذ أى معامل $\beta_{و}$ أى قيمة نفترضها ، وبصفة خاصة لاختبار الفرض $\beta_{و}$ = ٠ لأهمية ذلك فى بحث مدى مساهمة المتغير المناظر س$_{و}$ فى التنبؤ من معادلة الانحدار .

ملاحظة (٢) :

ليس من المناسب هنا أيضا تقديم الصيغة العامة التى نوجد منها التقدير ع (ب$_{و}$) للانحراف المعيارى للمتغير $B_{و}$ وسنعتمد فى ذلك أيضا على الحاسب الالكترونى ويكفينا أن نحيط بالدور الذى يقوم به . إلا أنه حين يكون هناك متغيران تنبؤيان اثنان فقط (ك = ٢) فيمكن إثبات أن هذا التقدير يحسب من الصيغة الآتية :

$$ع\ (ب_{و}) = \frac{ع_{ص.س_{١}س_{٢}}}{\sqrt{(١ - ر^{٢}_{٢١})\ .\ م\ م\ (س_{و})}} \qquad و = ١ ، ٢ \qquad (١١)$$

حيث $ع^{٢}_{ص.س_{١}س_{٢}} = \frac{١}{ن - ٣} \times$ الاختلاف غير المفسر

هو مربع الخطأ المعيارى للتقدير ويمكن حسابه من الصيغة (٥) ، وحيث $ر^{٢}_{٢١}$ هو معامل التحديد للمتغيرين س$_{١}$ ، س$_{٢}$ ويمكن حسابه من الصيغة :

$$ر^{2}_{21} = \frac{[\text{م ص}\ (\text{س}_{1}\ ،\ \text{س}_{2})]^{2}}{\text{م م}\ (\text{س}_{1})\ .\ \text{م م}\ (\text{س}_{2})} \qquad (12)$$

كما أن م م (س_{1}) هو مجموع المربعات للمتغير س ، م م (س_{2}) هو مجموع المربعات للمتغير س_{2} .

(١٢ - ٢) استخدام الحاسب الالكترونى :

يتضح من البند السابق أن حل المشكلات العملية فى تحليل الانحدار يتطلب القيام بعمليات حسابية طويلة سواء فى إيجاد المجاميع أو حل المعادلات أو فى استخراج تقديرات للأخطاء المعيارية أو فى استخراج قيم ت أو قيم ف اللازمة للاستنتاج الإحصائى ، مما يستغرق جهدا شاقا ووقتا طويلا ، إضافة إلى صعوبة تلافى أخطاء هذه الحسابات إذا أجريت يدويا . ولذلك فإن معظم المشكلات التطبيقية تستعين بالحاسبات الالكترونية التى تنوب عنا فى استخراج ما نريد بكل سرعة ودقة وتترك لنا فقط مهمة تفسير النتائج واتخاذ القرارات بناء على ما قدمته من معلومات .

واستخدام هذه الحاسبات لا يستلزم أن يكون الباحث قادرا على تشغيلها بنفسه بل يكفيه أن يتصل بأحد مراكز الحاسبات وتقديم مصفوفة البيانات التى حصل عليها من التجربة مع تحديد مجموعة الأعداد التى تخص كلا من المتغير الصادى والمتغيرات السينية ، وتحديد ما يريد إيجاده من معلومات .

وهناك كثير من الحاسبات تختزن برامج إحصائية تقوم بكافة أنواع التحليل ، ومن هذه البرامج ما يلى :

STATPACK, MINITAB, COSAP AND SPSS

ومثل هذه البرامج من شأنها توفير الجهد والوقت مع ضمان الدقة والأمان ولذلك يعتمد هذا الفصل فى حساباته وتحليلاته على الحاسبات الالكترونية التى أصبحت

اليوم فى متناول الجميع . وسنوضح ذلك بالمثال الآتى الذى يتناول حالة متغيرين تنبؤيين $س_١$ ، $س_٢$. على أن الأسلوب المستخدم يمتد إلى الحالات التى تتناول أكثر من متغيرين دون أى تعديل .

مثال (١٢ - ١) :

أرادت إحدى الشركات التجارية الكبيرة أن تعد طريقة للتنبؤ بعدد الوحدات التى تباع فى الشهر فى أى فرع من فروعها العشرة ، على افتراض أن الوحدات المباعة فى أى فرع تتأثر بعاملين هما (١) عدد البائعين فى الفرع ، (٢) مقدار ما يصرفه الفرع شهريا على الإعلان . وتحقيقا لهذا الغرض جمعت بيانات من الفروع ودونت فى الجدول (١٢ - ٣) الآتى .

الجدول (١٢ - ٣)

الفرع	عدد الوحدات المباعة ص	عدد العاملين $س_١$	تكاليف الاعلان $س_٢$
(١)	٢٥	٥	١٠
(٢)	٢٠	٢	١١
(٣)	٣٠	٦	١٢
(٤)	٢٥	٤	١٣
(٥)	٢٥	٣	١٤
(٦)	٣٢	٦	١٥
(٧)	٢٥	٤	١٢
(٨)	٢١	٣	١١
(٩)	٢٠	٢	١٠
(١٠)	٢٧	٥	١٢

المطلوب إيجاد معادلة الانحدار الخطى للمتغير ص على المتغيرين س١ ، س٢ واختبار دلالة الانحدار ككل ودلالة كل من معاملى الانحدار .

الحل :

للتوضيح سنقوم بحل هذه المسألة مرتين : باستخدام الحاسب ثم بدونه .

(أولا) الحل باستخدام الحاسب :

أدخلت البيانات المدونة بالجدول (١٢ – ٣) فى حاسب الكترونى فأخرج المعلومات المدونة بالجدول (١٢ – ٤) الآتى ، مع ملاحظة أن عدد المتغيرات التنبؤية ك = ٢ وأن حجم العينة ن = ١٠ .

الجدول (١٢ – ٤)

مخرجات الحاسب الالكترونى لبيانات المثال (١٢ – ١)

(1)	Regress.of	Y	Number of Units Sold		
(2)	on	X_1	Number of Salespeople		
(3)		X2	Amount of Advertising Expenditure		
(4)	Variable Name	Regress. Coeff	S.E. of Coeff.	t	D.F.
(5)	Constant	6.05263			
(6)	X1	2.10526	.17708	11.88891	7
(7)	X2	.87719	.16125	5.42652	7

(8) Coefficient of Determination (R ^ 2) = .974659

(9) Estimated Standard Error of Estimate = .722013

Analyis of Variance for Regression

(10) Source of Variation	SS	D.F.	MS	F
(11) Regression	140.351	2	70.1754	134.615
(12) Residual	3.64912	7	.521303	

(ا) معادلة الانحدار :

بالتأمل فى الجدول (١٢ – ٤) نرى أن الحاسب قد قام بإيجاد التقديرات المطلوبة للبارامترات α ، β_1 ، β_2 وهى:

ا = ٦,٠٥٢٦٣ ، ب$_{١}$ = ٢,١٠٥٢٦ ، ب$_{٢}$ = ٠,٨٧٧١٩

وهذه القيم الثلاث هى تلك المدونة بالصفوف ٥ ، ٦ ، ٧ من العمود الثانى بالجدول . وإذن معادلة الانحدار هى – التقريب إلى ٣ خانات عشرية :

$\hat{ص}$ = ٦,٠٥٣ + ٢,١٠٥ س$_{١}$ + ٠,٨٧٧ س$_{٢}$

فمثلا إذا كان عدد البائعين فى الفرع س$_{١}$ = ٤ والتكاليف الشهرية س$_{٢}$ = ١٢ فإننا نتوقع أن يكون عدد الوحدات المباعة حوالى ٢٤ وحدة .

(ب) اختبار دلالة الانحدار الخطى :

من الجدول (١٢ – ٤) نجد أن الحاسب قد حسب لنا جدول التباين فى السطور الثلاثة الأخيرة ، فقد أعطانا كلا من :

الاختلاف المفسر (الانحدار) = ١٤٠,٣٥١ بدرجات حرية ك = ٢

الاختلاف غير المفسر (الانحراف عن خط الانحدار) = ٣,٦٤٩١٢ بدرجات حرية ﻫ – ك – ١ = ٧ بل أعطانا أيضا نسبة التباين ف$_{ى}$ مستخدما الصيغة (٦) وهى

$$ف_{ى} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{٧٠,١٧٥٤}{٠,٥٢١٣٠٣} = ١٣٤,٦١٥$$

الاستنتاج :

لما كانت القيمة ١٣٤,٦١٥ تزيد كثيرا عن القيمة الحرجة ف$_{٠,٠١ [٢ ، ٧]}$ = ٩,٥٥

نرفض الفرض الصفرى أن $\beta_{١} = \beta_{٢} =$ ٠ عند مستوى عالى من الدلالة بمعنى أن هناك علاقة خطية بين عدد الوحدات المباعة من ناحية وعدد البائعين وتكاليف الإعلان من ناحية أخرى .

كما أن الحاسب قد أعطانا قيمة معامل التحديد بين المتغير ص والمتغيرين $س_{١}$ ، $س_{٢}$ وهو $ر^{٢}_{ص (١ ، ٢)}$ = ٠,٩٧٤٦٥٩ وهذا يعنى أن المتغيرين التنبؤيين قد فسرا حوالى ٩٧,٥٪ من التغير فى ص وهذا جزء كبير يدعم الحكم بوجود العلاقة الخطية .

(حـ) اختبار دلالة كل من معاملى الانحدار :

أعطانا الحاسب فى السطرين السادس والسابع من العمود الثالث التقديرين الخاصين بالخطأ المعيارى لمعاملى الانحدار الجزئيين وهما ع ($ب_{١}$) = ٠,١٧٧٠٨ ، ع ($ب_{٢}$) = ٠,١٦١٢٥ ، كما أعطانا فى العمودين الأخيرين من هذين السطرين قيمتى $ت_{١}$ ، $ت_{٢}$ مستخدما الصيغة (١٠) بعد وضع $\beta_{١}$ = ٠ ، $\beta_{٢}$ = ٠ وهما $ت_{١}$ = ١١.٨٨٨٩١ ، $ت_{٢}$ = ٥,٤٢٦٥٢ ولكل منهما سبع درجات حرية .

الاستنتاج :

نظرا لأن $ت_{(٧ ، ٠,٠١)}$ = ٣,٤٩٩ نرفض كلا من الفرضين الصفرين $\beta_{١}$ = ٠ ، $\beta_{٢}$ = ٠ عند مستوى الدلالة ٠,٠١ وهذا يعنى أن كلا من المتغيرين التنبؤيين يسهم إسهاما جوهريا فى التنبؤ بعدد الوحدات المباعة .

نلخص ما وجدناه فى هذه التجربة كما يلى : بناء على البيانات المشاهدة فى العينة على أساس أن افتراضات الانحدار متحققة ، تأخذ معادلة الانحدار الشكل الآتى :

$$\hat{ص} = ٦,٠٥٣ + ٢,١٠٥\ س_{١} + ٠,٨٧٧\ س_{٢}$$

وهذه المعادلة تعبر تعبيرا مناسبا عن العلاقة الحقيقية بين المتغير ص (عدد الوحدات المباعة) والمتغيرين $س_{١}$ ، $س_{٢}$ (عدد البائعين وقيمة تكاليف الإعلان) ، وتفسر حوالى ٩٧,٥٪ من التغير فى قيم ص . كما أن كلا من هذين المتغيرين يسهم إسهاما جادا فى التنبؤ بقيم هذا المتغير .

(ثانيا) الحل بغير استخدام الحاسب :

إذا لم يكن الحاسب الآلى متوفرا وكان الانحدار ذا متغيرين تنبؤيين فقط ، يمكن إجراء العمليات الحسابية المطلوبة باستخدام حاسب الجيب دون تحمل مشقة كبيرة . وفى هذا المثال يتخذ الحل الخطوات الآتية :

(ا) ايجاد معادلة الانحدار :

نبدأ بحساب المجاميع الآتية

مجـ ص = ٢٥٠ ، مجـ $ص^{٢}$ = ٦٣٩٤ ، مجـ $س_{١}$ = ٤٠ ، مجـ $س_{١}^{٢}$ = ١٨٠ ، مجـ $س_{٢}$ = ١٢٠ ، مجـ $س_{٢}^{٢}$ = ١٤٦٤ ، مجـ $س_{١}$ $س_{٢}$ = ٤٨٩ ، مجـ $س_{١}$ ص = ١٠٥٠ ، مجـ $س_{٢}$ ص = ٣٠٤٠ ثم نعوض بهذه القيم فى المعادلات المعتادة (٣) كالآتى :

$$١٠\,\alpha + ٤٠\,\beta_{١} + ١٢٠\,\beta_{٢} = ٢٥٠$$

$$٤٠\,\alpha + ١٨٠\,\beta_{١} + ٤٨٩\,\beta_{٢} = ١٠٥٠$$

$$١٢٠\,\alpha + ٤٨٩\,\beta_{١} + ١٤٦٤\,\beta_{٢} = ٣٠٤٠$$

وهناك عدة طرق لحل مثل هذه المعادلات نختار منها هنا طريقة دوليتل Doolittle التى تحول مصفوفة المعاملات والثوابت فى المعادلات المعتادة إلى مصفوفة مثلثية تكون جميع عناصر القطر الرئيسى فيها مساوية للواحد . وهذه الطريقة روتين خاص من التعليمات تتبين من الجدول (١٢ - ٥) الآتى .

الجدول (١٢ - ٥)

طريقة دوليتل لحل المعادلات الخطية

التعليمات	الصف	α	β_1	β_2	الثوابت
معاملات وثابت المعادلة الأولى	$م_{1}$	١٠	٤٠	١٢٠	٢٥٠
معاملات وثابت المعادلة الثانية	$م_{2}$	٤٠	١٨٠	٤٨٩	١٠٥٠
معاملات وثابت المعادلة الثالثة	$م_{3}$	١٢٠	٤٨٩	١٤٦٤	٣٠٤٠
$م_{1}$	$م'_{1}$	١٠	$\underline{٤٠}$	$\underline{١٢٠}$	٢٥٠
$م'_{1} \div ١٠$	$ص_{1}$	١	٤	١٢	٢٥
$م_{2ر} - م'_{21}\, ص_{1ر}$	$م'_{2}$	٠	٢٠	$\underline{٩}$	٥٠
$م'_{2} \div ٢٠$	$ص_{2}$	٠	١	٠,٤٥	٢,٥
$م_{3ر} - م'_{31}\, ص_{1ر} - م'_{32}\, ص_{2ر}$	$م'_{3}$	٠	٠	١٩,٩٥	١٧,٥
$م'_{3} \div ١٩,٩٥$	$ص_{3}$	٠	٠	١	٠,٨٧٧١٩

يلاحظ فى هذه التعليمات أن هناك ثلاثة مقادير ثابتة هى $م'_{21}$ = ٤٠ ، $م'_{31}$=١٢٠ ، $م'_{32}$ = ٩ وهذة تسمى **محاور الارتكاز** . أما بقية القيم التى تشتمل على الرمز ر فتتغير حسب قيمة ر التى تعبر عن رقم العمود (ر = ١ ، ٢ ، ٣ ، ٤) فمثلا $م_{2ر}$ تعبر عن العنصر الذى بالصف $م_{2}$ والعمود ر وبهذا يكون

$م_{12}$ = ٤٠ ، $م_{22}$ = ١٨٠ ، $م_{32}$ = ٤٨٩ ، $م_{42}$ = ١٠٥٠

تبدأ طريقة دولتيل بتدوين معاملات وثوابت المعادلات كما هو مبين بالصفوف الثلاثة الأولى المشار إليها بالرموز م$_1$ ، م$_2$ ، م$_3$. ثم تحسب عناصر الصفوف التالية الواحد بعد الآخر باستخدام التعليمات المبينة أمامه .

فالصف الرابع م$'_1$ يتكون من نفس عناصر الصف م$_1$. والصف التالى ص$_1$ يتكون بقسمة عناصر الصف السابق م$'_1$ على معامل α وهو ١٠ ، والصف التالى م$'_2$ تتكون عناصره باستخدام التعليمات المبينة أمامه كالآتى :

بوضع ر = ١ نجد أن العنصر الأول فى هذا الصف هو ٤٠ − ٤٠ × ١ = ٠

وبوضع ر = ٢ نجد أن العنصر الثانى فى هذا الصف هو ١٨٠ − ٤٠ × ٤ = ٢٠

وبوضع ر = ٣ نجد أن العنصر الثالث فى هذا الصف هو

٤٨٩ − ٤٠ × ١٢ = ٩

وبوضع ر = ٤ نجد أن العنصر الرابع فى هذا الصف هو

١٠٥٠ − ٤٠ × ٢٥ = ٥٠

ويلاحظ أننا استخدمنا هنا محور الارتكاز م$'_{21}$ = ٤٠ .

بنفس الطريقة نوجد بقية الصفوف الثلاثة . وبذلك تتحول المصفوفة التى رمز لصفوفها بالرموز م$_1$ ، م$_2$ ، م$_3$ وهى المصفوفة الأصلية إلى المصفوفة المثلثية التى رمز لصفوفها بالرموز ص$_1$ ، ص$_2$ ، ص$_3$ وبالتالى تتحول مجموعة المعادلات المعتادة إلى المجموعة المكافئة الآتية :

$\alpha + ٤\beta_1 + ١٢\beta_2 = ٢٥$ (من الصف ص$_1$)

، $\beta_1 + ٠{,}٤٥\beta_2 = ٢{,}٥$ (من الصف ص$_2$)

، $\beta_2 = ٠{,}٨٧٧١٩$ (من الصف ص$_3$)

من هذه المعادلات ينتج أن القيم ا ، ب$_1$ ، ب$_2$ التى تحقق هذه المعادلات معا هى :

$ب_٢ = ٠,٨٧٧١٩$

$\therefore ب_١ = ٢,٥ - ٠,٤٥ \; ب_٢ = ٢,٥ - ٠,٤٥ \times ٠,٨٧٧١٩$

$= ٢,١٠٥٢٦$

، $ا = ٢٥ - ١٢ \; ب_٢ - ٤ \; ب_١$

$= ٢٥ - ١٢ \times ٠,٨٧٧١٩ - ٤ \times ٢,١٠٥٢٦$

$= ٦,٠٥٢٦٨$

فتكون معادلة الانحدار مقربة إلى ثلاث خانات عشرية هى :

$$\hat{ص} = ٦,٠٥٣ + ٢,١٠٥ \; س_١ + ٠,٨٧٧ \; س_٢$$

وهى نفس المعادلة التى أوجدها لنا الحاسب الآلى .

(ب) اختبار دلالة الانحدار :

نحتاج هنا إلى تحليل الاختلاف فى ص إلى مركبتين كالآتى :

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

$$= ٦٣٩٤ - \frac{(٢٥٠)^٢}{١٠} = ١٤٤ \quad \text{بدرجات حرية} \; ن - ١ = ٩$$

من الصيغة (٥) نجد أن :

$$\text{الاختلاف غير المفسر} = مجـ \; ص^٢ - ا \; مجـ \; ص - ب_١ \; مجـ \; س_١ص - ب_٢ \; مجـ \; س_٢ص$$

$$= ٦٣٩٤ - ٦,٠٥٢٦٨ \times ٢٥٠ - ٢,١٠٥٢٦ \times ١٠٥٠ - ٠,٨٧٧١٩ \times ٣٠٤٠$$

$$= ٣,٦٤٩٤ \qquad \text{بدرجات حرية} \; ن - ك - ١ = ٧$$

∴ الاختلاف المفسر = ١٤٤ − ٣,٦٤٩٤

= ١٤٠,٣٥٠٦ بدرجات حرية ك = ٢

لاختبار دلالة الانحدار نستخدم الصيغة (٦) كالآتى :

$$ف_{ى} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{١٤٠,٣٥٠٦ \div ٢}{٣,٦٤٩٤ \div ٧}$$

$$= ١٣٤,٦٠٤٩$$

وهذه هى نفس القيمة التى أوجدها الحاسب (مع فارق التقريب) وهى كما ذكرنا ذات دلالة عالية وتدعونا لرفض الفرض الصفرى $\beta_{١} = \beta_{٢} = ٠$.

$$\text{معامل التحديد } ر^{٢}_{ص\,(١\,،\,٢)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} = \frac{١٤٠,٣٥٠٦}{١٤٤}$$

$$= ٠,٩٧٤٦٦$$

وهى نفس القيمة التى أوجدها الحاسب .

(حـ) اختبار دلالة كل من معاملى الانحدار :

نحتاج هنا أولا إلى تقدير الخطأ المعيارى لكل من المعاملين بالصيغة (١١) وهى

$$ع(ب_{و}) = \frac{ع_{ص.س_{١}س_{٢}}}{\sqrt{(١ - ر^{٢}_{١٢}).\,م\,م\,(س_{و})}} \qquad و = ١ ، ٢$$

ثم استخدام اختبار ت لكل منهما بالصيغة (١٠) وهى

$$ت_{و} = \frac{B_{و} - ٠}{ع\,(ب_{و})} \qquad و = ١ ، ٢$$

بدرجات حرية عددها ن − ك − ١ = ٧ .

لدينا : $\text{ع}^{\text{٢}}_{\text{ص.س}_{\text{١}}\text{س}_{\text{٢}}} = \frac{\text{١}}{\text{ن} - \text{ك} - \text{١}} \times$ الاختلاف غير المفسر

$$= \frac{\text{١}}{\text{٧}} \times \text{٣,٦٤٩٤} = \text{٠,٥٢١٢٤٢٧}$$

$$\therefore \text{ع}_{\text{ص.س}_{\text{١}}\text{س}_{\text{٢}}} = \text{٠,٧٢٢٠}$$

$$\text{، م م (س}_{\text{١}}\text{)} = \text{١٨٠} - \frac{\text{٤٠}^{\text{٢}}}{\text{١٠}} = \text{٢٠}$$

$$\text{، م م (س}_{\text{٢}}\text{)} = \text{١٤٦٤} - \frac{\text{١٢٠}^{\text{٢}}}{\text{١٠}} = \text{٢٤}$$

$$\text{، م ص (س}_{\text{١}}\text{ ، س}_{\text{٢}}\text{)} = \text{٤٨٩} - \frac{\text{٤٠} \times \text{١٢٠}}{\text{١٠}} = \text{٩}$$

$$\therefore \text{ر}^{\text{٢}}_{\text{٢١}} = \frac{\text{٨١}}{\text{٢٠} \times \text{٢٤}} = \text{٠,١٦٨٧٥}$$

$$\therefore \text{١} - \text{ر}^{\text{٢}}_{\text{٢١}} = \text{٠,٨٣١٢٥}$$

$$\text{ومنها ع(ب}_{\text{١}}\text{)} = \frac{\text{٠,٧٢١٩٨}}{\sqrt{\text{٠,٨٣١٢٥} \times \text{٢٠}}} = \text{٠,١٧٧٠٧}$$

$$\text{، ع(ب}_{\text{٢}}\text{)} = \frac{\text{٠,٧٢١٩٨}}{\sqrt{\text{٠,٨٣١٢٥} \times \text{٢٤}}} = \text{٠,١٦١٦٤}$$

$\therefore \text{ت}_{\text{١}} = \frac{\text{٢,١٠٥}}{\text{٠,١٧٧٠٧}} = \text{١١,٨٨٨}$ بدرجات حرية ٧

، $\text{ت}_{\text{٢}} = \frac{\text{٠,٨٧٧}}{\text{٠,١٦١٦٤}} = \text{٥,٤٢٥٦}$ بدرجات حرية ٧

وهاتان القيمتان أعطاهما الحاسب وقد رأينا أن كليهما ذو دلالة عالية وتدعوان إلى رفض كل من الفرضين $\beta_1 = 0$ ، $\beta_2 = 0$ عند مستوى الدلالة ٠,٠١ .

(١٢ - ٣) أسلوب آخر لاختبار دلالة معاملات الانحدار الجزئية :

لنبدأ بالحالة التى يكون لدينا فيها متغيران سينيان $س_1$ ، $س_2$. إذا أوجدنا الاختلاف المفسر الناشىء عن انحدار هذين المتغيرين على المتغير ص حين يستخدمان معا (بدرجتين من درجات الحرية) ثم أوجدنا الاختلاف الناشىء عن انحدار المتغير $س_2$ على المتغير ص حين يستخدم وحده (بدرجة واحدة من الحرية) فإن الفرق بين هذين الاختلافين هو مقدار الاختلاف المفسر الذى ساهم به المتغير $س_1$ عند ضمه إلى المتغير $س_2$. أي هو المساهمة الخاصة بالمتغير $س_1$ بعد استبعاد مساهمة $س_2$ ويمكن اختبار الفرض الصفرى $\beta_1 = 0$ ضد الفرض $\beta_1 \neq 0$ بواسطة اختبار ف بالصيغة (٦) كالآتى :

$$\text{ف} = \frac{\text{التباين المفسر (للمتغير } س_1 \text{ بعد استبعاد } س_2\text{)}}{\text{التباين غير المفسر}}$$

$$= \frac{[\text{م م (انحدار } س_1 \text{ ، } س_2\text{)} - \text{م م (انحدار } س_2\text{)}] / 1}{[\text{م م(ص)} - \text{م م (انحدار } س_1\text{،}س_2\text{)}] / (\text{ن} - 3)} \qquad (13)$$

مع ملاحظة أن م م (ص) - م م (انحدار $س_1$ ، $س_2$) هو الاختلاف غير المفسر . وبالمثل ، إذا أوجدنا الاختلاف الذى يفسره $س_1$ وحده وطرحناه من الاختلاف الذى يفسره المتغيران معا نحصل على المساهمة الخاصة بالمتغير $س_2$ بعد استبعاد أثر $س_1$ ويمكن اختبار هذه المساهمة بنفس الطريقة .

ففى المثال – انظر الحل بغير الحاسب – وجدنا ما يلى :

م م (ص) = الاختلاف الكلى = ١٤٤

م م (انحدار $س_{١}$ ، $س_{٢}$) = الاختلاف المفسر للمتغيرين معا = ١٤٠,٣٥١٢

الاختلاف غير المفسر = ٣,٦٥ بدرجات حرية ن – ٣

نحسب الآن الاختلاف الذى يفسره كل من $س_{١}$ ، $س_{٢}$ عندما يستخدم كل منهما على حدة .

$$\text{م م(انحدار } س_{١}) = \frac{[\text{م صـ } (س_{١} \text{ ، ص})]^{٢}}{\text{م م } (س_{١})}$$

$$= \frac{\left(١٠٥٠ - \frac{٢٥٠ \times ٤٠}{١٠}\right)^{٢}}{٢٠} = ١٢٥ \text{ بدرجة واحدة من الحرية}$$

$$\text{م م (انحدار } س_{٢}) = \frac{\left(٣٠٤٠ - \frac{٢٥٠ \times ١٢٠}{١٠}\right)^{٢}}{٢٤} = ٦٦,٦٧ \text{ بدرجة واحدة من الحرية}$$

من الصيغة (١٣) ، لاختبار الفرض $\beta_{١} = ٠$

$$ف_{١} = \frac{١٤٠,٣٥١٢ - ٦٦,٦٧}{٣,٦٥ \div ٧} = \frac{٧٣,٦٨١٢}{٠,٥٢١٤}$$

$$= ١٤١,٣١٤ \qquad \text{بدرجتى حرية ١ ، ٧}$$

ولاختبار الفرض $\beta_{٢} = ٠$

$$ف_{٢} = \frac{١٤٠,٣٥١٢ - ١٢٥}{٣,٦٥ \div ٧} = \frac{١٥,٣٥١٢}{٠,٥٢١٤}$$

$$= ٢٩,٤٤٢ \qquad \text{بدرجتى حرية١ ، ٧}$$

ونظرا لأن ف$_{٠,٠١[١ ، ٧]}$ = ١٢,٢ نرفض كلا من $\beta_١$ = ٠ ، $\beta_٢$ = ٠ عند مستوى الدلالة ٠,٠١ .

يلاحظ أن اختبار ف المستخدم هنا يكافىء اختبار ت الذى نتج عن الحل السابق ، ويتأكد هذا من ملاحظة أن $\sqrt{١٤١,٣١٤}$ = ١١,٨٨٨ = ت$_١$ ، $\sqrt{٢٩,٤٤٢}$ = ٥,٤٢٦ = ت$_٢$. ويمكن إذا أردنا أن ندون هذه النتائج فى جدول كالآتى .

الجدول (١٢ - ٦)

اختبار كل من المتغيرين بعد استبعاد أثر المتغير الآخر

مصدر الاختلاف	م م	د ح	ط م	ف
انحدار س$_١$ ، س$_٢$ معا	١٤٠,٣٥	٢		
انحدار س$_١$ وحده	١٢٥	١		
انحدار س$_٢$ وحده	٦٦,٦٧	١		
انحدار س$_١$ بعد س$_٢$	٧٣,٦٨	١	٧٣,٦٨	١٤١,٣١٤**
انحدار س$_٢$ بعد س$_١$	١٥,٣٥	١	١٥,٣٥	٢٩,٢٤٢**
الاختلاف غير المفسر	٣,٦٥	٧	٠,٥٢١٤	
الكلى	١٤٤,٠٠	٩		

ملاحظات :

من التحليل الملخص بالجدول (١٢ – ٦) نخرج بالنتائج والملاحظات الآتية :

(١) المتغير $س_{١}$ حين يستخدم وحده للتنبؤ بقيم ص أفضل من المتغير $س_{٢}$ حين يستخدم وحده لنفس الغرض ، وذلك لأن الاختلاف الذى يفسره $س_{١}$ منفردا وهو ١٢٥ أكبر من الاختلاف الذى يفسره $س_{٢}$ منفردا وهو ٦٦,٦٧ . ونصل إلى نفس النتيجة إذا قارنا معاملى التحديد ، فمعامل التحديد للمتغير $س_{١}$ هو :

$$ر^{٢}_{ص\,س_{١}} = \frac{١٢٥}{١٤٤} = ٠,٨٦٨٠٦ \quad \text{بينما} \quad ر^{٢}_{ص\,س_{٢}} = \frac{٦٦,٦٧}{١٤٤} = ٠,٤٦٢٩٩$$

وبهذا الأسلوب نستطيع مقارنة أى عدد من المتغيرات تستخدم فرادا .

(٢) فى تفسير الاختلاف الكلى فى ص تكون مساهمة أى متغير أكبر فى حالة استخدامه منفردا عنها فى حالة استخدامه بعد متغير آخر فبالنسبة للمتغير $س_{١}$ نجد أن ١٢٥ > ٧٣,٦٨ وبالنسبة للمتغير $س_{٢}$ نجد أن ٦٦,٦٧ > ١٥,٣٥ وهذه نتيجة عامة مهما كان عدد المتغيرات التنبؤية ، فالمساهمة المنفردة لأى متغير تكون أكبر دائما من مساهمته مع متغير أو عدة متغيرات أخرى .

(٣) معامل التحديد المحسوب من معادلة الانحدار للمتغيرين $س_{١}$ ، $س_{٢}$ معا وهو $ر^{٢}_{ص(١ ، ٢)} = \frac{١٤٠,٣٥}{١٤٤} = ٠,٩٧٥$ أكبر من معامل التحديد المحسوب من معادلة الانحدار لأى من المتغيرين وهما ٠,٨٦٨ ، ٠,٤٦٣ . وهذه نتيجة عامة ، فمعامل التحديد يقل دائما حين يستبعد واحد أو أكثر من المتغيرات التنبؤية .

(٤) فى معادلة الانحدار للمتغيرين $س_{١}$ ، $س_{٢}$ معا يصعب تقدير الأهمية النسبية لهذين المتغيرين من حيث مقدار المساهمة فى التنبؤ بقيم المتغير التابع ص ،

لأننا إذا قدرنا هذه الأهمية النسبية بواسطة الاختلافين المفسرين عند استخدام كل متغير على حدة وهما ١٢٥ ، ٦٦,٦٧ نجد أن هذا التقدير غير مناسب لأن مجموع هذين الاختلافين وهو ١٩١,٦٧ يزيد عن الاختلاف الكلى فى ص وهو ١٤٤ . ومن ناحية أخرى ، إذا اعتبرنا الاختلاف الناشيء عن المتغير $س_{١}$ بعد استبعاد أثر المتغير $س_{٢}$ وهو ٧٣,٦٨ والاختلاف الناشيء عن المتغير $س_{٢}$ بعد استبعاد أثر المتغير $س_{١}$ وهو ١٥,٣٥ نجد أن مجموعهما وهو ٨٩,٠٣ يقل عن الاختلاف الذى يفسره المتغيران معا وهو ١٤٠,٣٥ . وهذه الصعوبة فى تقدير الأهمية النسبية للمتغيرات هى إحدى الصعوبات التى يلقاها الباحث عند التصدى لمقارنة المتغيرات فى الانحدار الخطى المتعدد وعند اختيار أفضل المتغيرات التى تدخل فى معادلة الانحدار كما سيأتى بالبند التالى .

(٥) قمية أى معامل انحدار جزئى $ب_{ى}$ لمتغير $س_{ى}$ هى قيمة **شرطية** تختلف باختلاف المتغيرات التى تدخل معه فى معادلة الانحدار .

إن أسلوب اختبار دلالة معاملات الانحدار الجزئية الذى أدى إلى الصيغة (١٣) هو أسلوب عام مهما كان عدد المتغيرات التنبؤية ، ويستخدم فى التعرف على دلالة ما يحدث من نقص فى دقة التنبؤ حين يستبعد واحد أو أكثر من المتغيرات من معادلة الانحدار . وبصفة عامة إذا كنا قد وفقنا معادلة انحدار فى ك من المتغيرات $س_{١}$ ، $س_{٢}$ ، . . . ، $س_{ك}$ ووفقنا معادلة انحدار فى $ل < ك$ من هذه المتغيرات أى بعد استبعاد ك − ل منها ، وأردنا اختبار ما إذا كانت دقة التنبؤ قد نقصت نقصا ذا دلالة نتيجة لهذا الاستبعاد فإننا نستخدم الصيغة العامة الآتية :

$$ف = \frac{[ر^{٢}(ك) - ر^{٢}(ل)]/(ك - ل)}{[ر^{٢}(ص) - ر^{٢}(ك)] \;/\; (ن - ك - ١)} \qquad (١٤)$$

بدرجتى حرية ك − ل ، ن − ك − ١ ،

حيث م م (ك) هو الاختلاف الناشيء عن الانحدار على ك من المتغيرات

، م م (ل) هو الاختلاف الناشيء عن الانحدار على ل من هذه المتغيرات

، م م (ص) هو الاختلاف فى قيم المتغير التابع ص

إذ يمكن إثبات أن كلا من بسط ومقام هذه النسبة هو تقدير مستقل للتباين σ^2 .

إن الصيغة (١٤) يمكن كتابتها بدلالة معاملات التحديد كالآتى :

$$\text{ف} = \frac{[\text{ر}^2(\text{ك}) - \text{ر}^2(\text{ل})] \,/\, (\text{ك} - \text{ل})}{[1 - \text{ر}^2(\text{ك})] \,/\, (\text{ن} - \text{ك} - 1)} \qquad (15)$$

بدرجتى حرية (ك − ل) ، (ن − ك − ١)

مثال (١٢ − ٢) :

أجرى انحدار خطى لمتغير عشوائى صـ على خمسة متغيرات تنبؤية سـ_1 ، سـ_2 ، سـ_3 ، سـ_4 ، سـ_5 ووجد أن معامل التحديد فى عينة حجمها ٤٦ هو ٠,٦٢ وعندما أجرى انحدار خطى لنفس المتغير على ثلاثة من هذه المتغيرات سـ_1 ، سـ_3 ، سـ_4 وجد أن معامل التحديد ٠,٥٧ هل يمكن الاستغناء عن المتغيرين سـ_2 ، سـ_5 والاكتفاء بالمتغيرات الثلاثة الأخرى دون أن نفقد شيئا من دقة التنبؤ بقيم المتغير صـ ؟

الحل :

من الصيغة (١٥) :

$$\text{ف}_{\text{ى}} = \frac{(0.62 - 0.57) \div 2}{(1 - 0.62) \div 40} = 2.632$$

وهذه القيمة تقل عن القيمة الحرجة ف$_{٠,٠٥[٢ ، ٤٠]}$ = ٣,٢٣ مما يدعونا إلى قبول الفرض الصفرى أن دقة التنبؤ لم تتأثر ، وبالتالى نستطيع الاكتفاء بالمتغيرات سـ$_١$ ، سـ$_٣$ ، سـ$_٤$ للتنبؤ بالمتغير صـ دون أن نفقد شيئا من دقة التنبؤ .

مثال (١٢ – ٣) :

فى المثال (١٢ – ١) هل نستطيع الاكتفاء بأحد المتغيرين للتنبؤ بقيم المتغير التابع صـ دون أن نفقد شيئا من دقة التنبؤ ؟

الحل :

سبق أن أجبنا عن هذا السؤال حين استخدمنا الصيغة (١٣) – التى هى حالة خاصة من الصيغة (١٤) أو (١٥) – فى رفض كل من الفرضين $\beta_١ = ٠$ ، $\beta_٢ = ٠$ لأن هذا يعنى أن وجود أى من المتغيرين له أثر ذو دلالة فى عملية التنبؤ . باستخدام الصيغة (١٥) لاختبار إمكانية الاستغناء عن المتغير سـ$_١$ لدينا :

$$ر^٢_{ص(س_١ س_٢)} = ٠,٩٧٤٦٦ \quad ، \quad ر^٢_{ص س_٢} = ٠,٤٦٢٩٩$$

$$ف_ى = \frac{(٠,٩٧٤٦٦ - ٠,٤٦٢٩٩) \div ١}{(١ - ٠,٩٧٤٦٦) \div ٧} = \frac{٠,٥١١٦٧}{٠,٠٠٣٦٢}$$

$$= ١٤١,٣٥$$

وهذه القيمة ذات دلالة عالية وتعنى أن الاستغناء عن سـ$_١$ يقلل من دقة التنبؤ . وبالمثل ، لاختبار إمكانية الاستغناء عن سـ$_٢$

$$ف = \frac{٠,٩٧٤٦٦ - ٠,٨٦٨٠٦}{٠,٠٠٣٦٢} = ٢٩,٤٥$$ وهذه أيضا ذات دلالة عالية وتعنى أن الاستغناء عن سـ$_٢$ يقلل من دقة التنبؤ .

(١٢ - ٤) اختيار متغيرات التنبؤ :

من الصعوبات التى يلقاها الباحثون فى الانحدار المتعدد كيفية اختيار المتغيرات التى تدخل فى معادلة الانحدار لتعطى أعلى درجة من الدقة فى التنبؤ بالمتغير التابع ص . وفى المعتاد يكون لدى الباحث مجموعة كبيرة من المتغيرات المرشحة لذلك ويكون من المفيد عمليا اختصار هذه المجموعة إلى مجموعة تتكون من أقل عدد ممكن من هذه المتغيرات بشرط أن تعطى هذه المجموعة الجزئية معادلة تنبؤ لا تقل كفاءة عن المعادلة التى تستخدم جميع المتغيرات المتاحة . إن عملية الاختصار هذه ممكنة فى الغالب لأن بعض المتغيرات المتاحة لا تسهم بدرجة كافية فى عملية التنبؤ كما قد يبدو لأول وهلة ، كما أنه قد توجد مجموعة (أو أكثر) من المتغيرات المتاحة ترتبط ببعضها ارتباطا عاليا وبالتالى تعطى معلومات مماثلة ويمكن حينئذ الاكتفاء بمتغير واحد فقط من هذه المجموعة ليمثلها جميعا .

ويتلخص السؤال المطروح هنا فيما يلى : إذا كان لدينا ك من المتغيرات التنبؤية المرشحة للدخول فى معادلة الانحدار فكيف نختار أصغر مجموعة جزئية من هذه المجموعة بحيث تعطى نفس الدرجة من الدقة فى التنبؤ بالمتغير التابع ؟ فمثلا إذا كان لدينا مجموعة من عشرة متغيرات تنبئوية فهل يمكن أن نكتفى بمجموعة من اثنين أو ثلاثة منها لتعبر عن المجموعة بكاملها (دون أن نفقد شيئا من دقة التنبؤ) ؟

الواقع أنه لا توجد إجابة واحدة شافية لهذا السؤال ولكن هناك عدة اقتراحات بمحاولات نسترشد بها فى تحديد المجموعة المطلوبة ، وتتطلب هذه المحاولات النظر إلى البيانات المشاهدة عدة مرات من جوانب مختلفة ، وهنا يلعب الحاسب دورا رئيسيا لأن تنفيذ هذه المحاولات يحتاج إلى جهد حسابى شاق .

نفرض أننا حصلنا على بيانات مشاهدة فى عينة حجمها ن كما فى الجدول (١٢ - ١) السابق .

(ا) إن أول ما يخطر بالبال هو اختيار المتغيرات الأكثر علاقة بالمتغير التابع صـ ، فنقوم بالمقارنة بينها كمتغيرات فرادا إما عن طريق مقادير الاختلافات المفسرة $م^{٢}$ (الانحدار) أو عن طريق معاملات التحديد $ر^{٢}$ كما جاء بالملاحظة (١) بالبند (١٢ - ٣) السابق ، فتكون المتغيرات ذوات القيم الأكبر مرشحة مبدئيا للدخول فى معادلة انحدار متعدد .

(ب) نبحث عما إذا كان هناك مجموعة (أو أكثر) من المتغيرات التنبؤية التى ترتبط ببعضها ارتباطا عاليا ، فإذا وجدت مثل هذه المجموعة فإن أحدها يمكن أن يمثلها جميعا . ويحتاج الأمر هنا إلى إيجاد معامل الارتباط بين كل زوج من هذه المتغيرات ، ويفضل وضع هذه المعاملات وأيضا معاملات الارتباط بين كل من المتغيرات والمتغير التابع فى جدول لتسهيل الرجوع إليها . وعلى فرض أن هناك خمس متغيرات تنبؤية يكون الجدول كالآتى :

	$س_{١}$	$س_{٢}$	$س_{٣}$	$س_{٤}$	$س_{٥}$
$س_{١}$	١	$ر_{٢١}$	$ر_{٣١}$	$ر_{٤١}$	$ر_{٥١}$
$س_{٢}$		١	$ر_{٣٢}$	$ر_{٤٢}$	$ر_{٥٢}$
$س_{٣}$			١	$ر_{٤٣}$	$ر_{٥٣}$
$س_{٤}$				١	$ر_{٥٤}$
ص	$ر_{ص١}$	$ر_{ص٢}$	$ر_{ص٣}$	$ر_{ص٤}$	$ر_{ص٥}$

(حـ) نوجد معادلة انحدار ص على جميع ما لدينا من متغيرات ثم نفحص دلالة كل من معاملات الانحدار الجزئية $ب_{١}$ ، $ب_{٢}$ ، . . . ، $ب_{ن}$ عن طريق اختبار

ت بالصيغة (١٠) فتكون المتغيرات التى ترشح للدخول فى معادلة الانحدار هى تلك التى تحظى بالقيم الأكبر من قيم ت ، أما المتغيرات التى تناظر القيم الأصغر من ت فتكون معرضة للاستبعاد . على أن ذلك لا ينبغى أن يتم دون تدبر فإن استخدام أو استبعاد متغير مالا يقرر لمجرد أن قيمة ت المناظرة كبيرة أو صغيرة ، لأن قيم معاملات الانحدار الجزئية وبالتالى قيم ت هى قيم شرطية وتتغير بتغير المتغيرات التى تدخل معها فى الانحدار كما سبق القول بالملاحظة (٥) ، ومن ناحية أخرى قد توجد أسباب نظرية أو خبرات تحتم الاحتفاظ ببعض المتغيرات حتى ولو كان معامل الانحدار الجزئى المناظر غير ذى دلالة . فمثلا إذا كان لدينا عدة متغيرات نرغب استخدامها فى التنبؤ بالمسافة التى تقطعها سيارة ما وكان من بين هذه المتغيرات متغير « مقدار الوقود المستهلك » فلابد من الاحتفاظ بهذا المتغير حتى ولو كان معامل الانحدار المناظر له غير ذى دلالة .

(ى) على ضوء ما نجده فى (أ) و(ب) و(حـ) نختار مجموعة أو أكثر من المتغيرات ونحسب معادلات انحدار كل منها مع إيجاد معامل التحديد ومقارنته بمعامل التحديد الذى ينجم من الانحدار على جميع المتغيرات المتاحة واختبار دلالة الفرق بينهما فى دقة التنبؤ باستخدام الصيغة (١٥) . وإذا وجدنا أن إحدى المجموعات المختارة بنفس كفاءة المجموعة الكاملة نحاول اختزالها بنفس الطريقة إلى مجموعة ذات عدد أقل من المتغيرات .

ويبنى اختيارنا النهائى لمجموعة المتغيرات التى تدخل فى معادلة التنبؤ التى ننشدها على أساس أنها (أولا) تشتمل على أقل عدد من المتغيرات و(ثانيا) يمكنها أن تعبر عن مجموعة المتغيرات المتاحة بكاملها ، أى بحيث لا تقل دقة التنبؤ منها عن دقة التنبؤ من استخدام جميع المتغيرات المتاحة . هذا ويجدر الإشارة أن المجموعة التى تختار على هذا الأساس ليست فريدة ، بل يمكن أن نعثر على أكثر من مجموعة تستوفى الشرطين المطلوبين .

تمارين (١٢ – ١)

أجريت دراسة لمعرفة أثر العوامل الجغرافية على حجم نوع من الضفادع واستهدف جزء من هذه الدراسة التنبؤ بطول جسم الضفدعة عن طريق الارتفاع عن سطح البحر والمتوسط السنوى لدرجة الحرارة . وقد جمعت بيانات من ١٣ موقعا ودونت بالجدول الآتى :

الجدول (١٢ – ٧)

الموقع	طول الضفدعة ص	الارتفاع عن سطح البحر $س_١$	درجة الحرارة $س_٢$
(١)	٢٣,٧	٦٠٠	٦٠
(٢)	٢٣,٩	٨٠٠	٥٥
(٣)	٢٣,٤	١٤٥٠	٥٠
(٤)	٢٣,٤	١٤٠٠	٥٠
(٥)	٢٢,٢	١٠٠	٥٥
(٦)	١٩,٥	٤٥٠	٥٠
(٧)	٢٣,٥	٤٠٠	٥٠
(٨)	٢٤,٧	٨٠٠	٦٠
(٩)	٢٣,٣	٨٠٠	٦٠
(١٠)	٢١,٧	١١٠٠	٥٠
(١١)	٢١,٣	٥٠٠	٥٠
(١٢)	٢٠,٢	٤٠٠	٥٠
(١٣)	٢٢,١	٤٠٠	٥٠

مجـ ص = ٢٩٢,٩   مجـ ص^٢ = ٦٦٢٧,٣٧   مجـ س١ = ٩٢٠٠   مجـ س١^٢ = ٨٤٩٥٠٠٠
مجـ س٢ = ٦٩٠   مجـ س٢^٢ = ٣٦٨٥٠   مجـ س١ س٢ = ٤٨٦٥٠٠
مجـ ص س١ = ٢١٠٢٦٥   مجـ ص س٢ = ١٥٥٩٢,٥

(أولا) استخدم مخرجات الحاسب الالكترونى المبينة بالجدول (١٢ – ٨) الآتى لإيجاد معادلة الانحدار الخطي للمتغير ص على المتغيرين س١ ، س٢ .
(ثانيا) اختبر دلالة الانحدار ككل ودلالة كل من معاملى الانحدار . هل من الممكن الاستغناء عن أحد المتغيرين التنبؤيين دون أن نفقد شيئا من دقة التنبؤ ؟

(ثالثا) أجب عن الجزئين (أولا) و(ثانيا) دون الاستعانة بمخرجات الحاسب .

الجدول (١٢ – ٨)

Regress. of	Y	Length of Frog		
On	X_1	Altitude		
	X_2	Temperature		
Variable Name	Regress. Coeff.	S.E. of Coeff.	t	D.F.
Constant	9.785579	4.129044	2.37	
X_1	.1700912 E-02.	.81158094—03	2.10	10
X_2	.2174479	.7589070 E—1	2.87	10

Multiple Corr. Coeff. (R) = .73379

Coeff. of Determination (R ^ 2) = .5384

Estimated Standard Error of Estimate = 1.1390588

Analyis of Variance for Regressioon

Source of Variation	SS	D.F.	M S	F
Regression	15.13315	2	7.566575	5.8319
Residual	12.97455	10	1.297455	

ملاحظة :

الرمز E - 01 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{-١}$

الرمز E - 02 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{-٢}$
وهكذا ...

الرمز E + 02 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{٢}$
وهكذا ...

فمثلا العدد 1700912 E - 02 . يعنى العدد 001700912 .

هذا ما يسمى بالرمزية العلمية scientific notation

ثانيا – الارتباط الخطى المتعدد

سنفترض هنا أن لدينا ك + ١ من المتغيرات العشوائية $ص_١$ ، $ص_٢$ ، . . . ، $ص_{ك+١}$ كما سنفترض أن توزيع الاحتمال المشترك لهذه المتغيرات هو توزيع معتدل متعدد المتغيرات multivariate normal distribution إن هذا الفرض يتضمن أمرين : الأول أن توزيع أى من هذه المتغيرات (على حدة) هو توزيع معتدل ، والثانى أن العلاقة بين أى من هذه المتغيرات وأى متغير آخر أو مجموعة من المتغيرات الأخرى هى علاقة خطية .

وامتدادا لدراساتنا فى الارتباط الخطى البسيط فى الفصل العاشر نتناول فى البندين الآتيين معاملين رئيسيين هما معامل الارتباط الخطى المتعدد ومعامل الارتباط الجزئى ونرى كيف نختبر دلالة كل منهما .

(١٢ – ٥) معامل الارتباط المتعدد

MULTIPLE CORRELATION COEFFICIENT

إذا أردنا تقدير درجة العلاقة الخطية بين أحد هذه المتغيرات وليكن $ص_١$ وبقية المتغيرات ، فإننا نعرّف معامل الارتباط المتعدد بين $ص_١$ والمتغيرات الأخرى بأنه

معامل الارتباط البسيط بين المتغير $ص_١$ والمتغير العشوائى الذى يتكون من التركيب الخطى $\alpha + \beta_٢ ص_٢ + \beta_٣ ص_٣ + \ldots + \beta_{ك+١} ص_{ك+١}$ وسنرمز لهذا المعامل بالرمز $ص_{١(٢٣ \ldots ك+١)}$.

ولتقدير هذا المعامل نأخذ عينة عشوائية من ن من وحدات المجتمع ونقوم بقياس قيمة كل من هذه المتغيرات لكل وحدة فنحصل على ن من المشاهدات كل منها على الصورة $(ص_{ر١} ، ص_{ر٢} ، \ldots ، ص_{ر، ك+١})$ حيث $ر = ١ ، ٢ ، \ldots ، ن$ وتوضع هذه القياسات عادة كما فى الجدول الآتى .

جدول (١٢ - ٨)

المشاهدات	$ص_١$	$ص_٢$	$ص_٣$	...	$ص_{ك+١}$
(١)	$ص_{١١}$	$ص_{٢١}$	$ص_{٣١}$	...	$ص_{١، ك+١}$
(٢)	$ص_{١٢}$	$ص_{٢٢}$	$ص_{٣٢}$	...	$ص_{٢، ك+١}$
...	...	...	...	...	...
(ر)	$ص_{ر١}$	$ص_{ر٢}$	$ص_{ر٣}$	...	$ص_{ر، ك+١}$
...	...	...	...	...	...
(ن)	$ص_{ن١}$	$ص_{ن٢}$	$ص_{ن٣}$	...	$ص_{ن، ك+١}$

من هذه المشاهدات نحصل على التقدير المطلوب بعدد سنرمز له بالرمز $ر_{١(٢٣ \ldots ك+١)}$ ونحسبه كما فى حالة الارتباط الخطى البسيط من الصيغة

$$ر^٢_{١(٢٣ \ldots ك)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (١٦)$$

أى أن التقدير المطلوب هو الجذر التربيعى لنسبة الاختلاف الذى يفسره الانحدار الخطى المتعدد إلى الاختلاف الكلى فى المتغير $\text{ص}_{\text{١}}$. ويمكن إثبات أنه فى هذه الحالة .

$$\text{١} \geq \text{ر}_{\text{١}(\text{٢٣} \ldots \text{ك} + \text{١})} \geq \text{٠} \qquad (\text{١٧})$$

أى أن هذا المعامل لا يمكن أن يكون سالبا وهو يختلف فى هذه الصفة عن معامل الارتباط البسيط الذى نعلم أنه يمكن أن يأخذ قيما سالبة .

وكما فى الانحدار الخطى المتعدد نعتمد فى حساب هذا المعامل على الحاسب الالكترونى توفيرا للجهد والوقت . وفى الحالة البسيطة التى يكون لدينا فيها ثلاثة متغيرات عشوائية $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$ يمكننا حساب معامل الارتباط المتعدد $\text{ر}_{\text{١}(\text{٢} ، \text{٣})}$ بين المتغير $\text{ص}_{\text{١}}$ والمتغيرين $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$ من الصيغة الآتية :

$$\text{ر}^{\text{٢}}_{\text{١}(\text{٣٢})} = \frac{\text{ر}^{\text{٢}}_{\text{٢١}} + \text{ر}^{\text{٢}}_{\text{٣١}} - \text{٢}\, \text{ر}_{\text{٢١}} \,.\, \text{ر}_{\text{٣١}} \,.\, \text{ر}_{\text{٣٢}}}{\text{١} - \text{ر}^{\text{٢}}_{\text{٣٢}}} \qquad (\text{١٨})$$

ويتطلب حساب هذه القيمة إيجاد كل من :

$\text{ر}_{\text{٢١}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٢}}$

، $\text{ر}_{\text{٣١}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٣}}$

، $\text{ر}_{\text{٣٢}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$

ومن الواضح أن قيمة معامل الارتباط المتعدد تعتمد على قيم معاملات الارتباط البسيط بين أزواج المتغيرات المستخدمة ، وهذه الملاحظة صحيحة مهما كان عدد المتغيرات .

اختبار دلالة معامل الارتباط المتعدد :

لاختبار الفرض الصفرى $ر_{1(32...ك+1)} = 0$ ضد الفرض $ر_{1(32...ك+1)} \neq 0$ نستخدم نفس الصيغة (٦) التى وردت بالبند (١٢ - ١) وهى :

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

$$= \frac{ر^2_{1(32...ك+1)} / ك}{(1 - ر^2_{1(32...ك+1)}) / (ن - ك - 1)} \qquad (19)$$

وإذا توفرت الشروط المذكور فى مستهل الجزء الثانى من هذا الفصل فإن هذه الإحصاءة يكون لها توزيع ف بدرجتى حرية ك ، ن - ك - ١ حيث ك + ١ عدد جميع المتغيرات المستخدمة . يلاحظ أن كلا من الصيغتين (٦) ، (١٩) تختبر وجود أو عدم وجود العلاقة الخطية .

مثال (١٢ - ٤) :

فى عينة عشوائية من ٢٣ مجموعة من القيم من مجتمع معتدل ذى ثلاثة متغيرات عشوائية وجدت معاملات الارتباط البسيطة الآتية : $ر_{21} = 0,38$ ، $ر_{31} = 0,46$ ، $ر_{32} = 0,51$. أوجد معامل الارتباط المتعدد $ر_{1(32)}$ واختبر دلالته .

الحل :

من الصيغة (١٨) :

$$ر^2_{1(32)} = \frac{0,1444 + 0,2116 - 2 \times 0,38 \times 0,46 \times 0,51}{1 - 0,2601}$$

$$= 0,240$$

∴ معامل الارتباط المتعدد المطلوب = $\sqrt{0,240}$ = ٠,٤٩ تقريبا .

من الصيغة (١٩) :

$$ف_{ى} = \frac{٠,٢٤٠ \div ٢}{(١ - ٠,٢٤) \div ٢٠} = ٣,١٥٧٩ \qquad \text{بدرجتي حرية ٢، ٢٠}$$

وهذه القيمة ليست بذات دلالة عند المستوى ٠,٠٥ وتشير إلى عدم وجود ارتباط خطى بين المتغير $ص_{١}$ والمتغيرين $ص_{٢}$ ، $ص_{٣}$.

(١٢ - ٦) معامل الارتباط الجزئى

PARTIAL CORRELATION COEFFICIENT

إن معامل الارتباط البسيط $ص_{٢١}$ بين متغيرين عشوائيين $ص_{١}$ ، $ص_{٢}$ يفترض أن هذين المتغيرين يتأثران ببعضهما فقط ولا يتأثران بأى متغير آخر . إلا أنه فى كثير من الحالات يكون كل من هذين المتغيرين متأثرا تأثرا خطيا بمتغير عشوائى ثالث $ص_{٣}$ ، وفى هذه الحالة يقتضى القياس الصحيح للارتباط بين $ص_{١}$ ، $ص_{٢}$ استبعاد أثر هذا المتغير الثالث من كل منهما . إن المعامل الذى يقيس مثل هذا الارتباط يسمى بمعامل الارتباط الجزئى بين $ص_{١}$ ، $ص_{٢}$ وسنرمز له بالرمز $ص_{٢١ - ٣}$. ويعرّف هذا المعامل أيضا بأنه معامل الارتباط البسيط بين $ص_{١}$ ، $ص_{٢}$ عند قيمة ثابتة للمتغير $ص_{٣}$ أى فى قطاع الوحدات التى تأخذ نفس القيمة للمتغير $ص_{٣}$. ويتفق هذان التعريفان طالما كان توزيع الاحتمال المشترك للمتغيرات الثلاثة معتدلا trivariate normal distribution إذ أنه فى هذه الحالة تظل قيمة $ص_{٢١-٣}$ كما هى مهما كانت القيمة الثابتة للمتغير $ص_{٣}$ ، وهذه هى الحالة التى نتناولها هنا والتى تتمشى مع ما افترضناه فى مستهل هذا الجزء .

ويقدر هذا المعامل من العينة بمقدار سنرمز له بالرمز $ر_{٢١-٣}$ حيث

$$ر_{٢١-٣} = \frac{ر_{٢١} - ر_{٣١} \cdot ر_{٣٢}}{\sqrt{(١ - ر^{٢}_{٣١})(١ - ر^{٢}_{٣٢})}} \qquad (٢٠)$$

وحيث $ر_{٢١}$ ، $ر_{٣١}$ ، $ر_{٣٢}$ هى معاملات الارتباط البسيط بين أزواج المتغيرات . وبنفس الطريقة نحسب كلا من $ر_{٣١-٢}$ ، $ر_{٣٢-١}$ مع تذكر أننا هنا لا نميز بين متغير مستقل ومتغير تابع .

أما دلالة هذا المعامل فنختبره عن طريق الإحصاءة :

$$ت = \frac{ر_{٢١-٣}\sqrt{ن - ٣}}{\sqrt{١ - ر^{٢}_{٢١-٣}}} \qquad \text{بدرجات حرية } ن - ٣ \qquad (٢١)$$

أو عن طريق الإحصاءة المكافئة:

$$ف = \frac{ر^{٢}_{٢١-٣}}{(١ - ر^{٢}_{٢١-٣}) / (ن - ٣)} \qquad \text{بدرجات حرية } ١ \text{ ، } ن - ٣ \qquad (٢٢)$$

مثال (١٢ - ٥) :

فى دراسة للعلاقة بين ضغط الدم وكلوسترول الدم أخذت عينة عشوائية من ١٤٢ من السيدات كبار السن وقيس كل من ضغط الدم $ص_{١}$ ودرجة تركيز الكلوسترول $ص_{٢}$ والعمر $ص_{٣}$ ، ثم حسبت معاملات الارتباط البسيط لكل زوج من هذه المتغيرات فوجدت كما يلى : $ر_{٢١} = ٠,٢٤٩٥$ ، $ر_{٣١} = ٠,٣٣٣٢$ ، $ر_{٣٢} = ٠,٥٠٢٩$ ابحث دلالة معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول (أولا) دون استبعاد متغير العمر ، (ثانيا) بعد استبعاد متغير العمر .

الحل :

(أولا) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول دون

استبعاد متغير العمر هو $ر_{٢١}$ = ٠,٢٤٩٥ ونختبر دلالة هذا المعامل باستخدام الصيغة (٨) بالبند (١٠ – ٥) وهى :

$$ت = \frac{ر_{٢١}\sqrt{ن - ٢}}{\sqrt{١ - ر^{٢}_{٢١}}}$$ بدرجات حرية ن – ٢

أو الصيغة المكافئة ف = $\frac{ر^{٢}_{٢١}}{(١ - ر^{٢}_{٢١}) / (ن - ٢)}$ بدرجتى حرية ١ ، ن – ٢

لدينا : $ت_{ى} = \frac{٠,٢٤٩٥\sqrt{١٤٠}}{\sqrt{١ - (٠,٢٤٩٥)^{٢}}} = ٣,٠٤٨٦$ بدرجات حرية ١٤٠

وهذه القيمة تزيد عن القيمة الحرجة $ت_{٠,٠١[١٤٠]}$ = ٢,٥٧٦ فهى ذات دلالة عالية وتشير إلى وجود ارتباط خطى بين المتغيرين .

(ثانيا) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول بعد استبعاد متغير العمر هو معامل الارتباط الجزئى $ر_{٢١-٣}$. من الصيغة (٢٠) نجد أن

$$ر_{٢١-٣} = \frac{٠,٢٤٩٥ - ٠,٣٣٣٢ \times ٠,٥٠٢٩}{\sqrt{(١ - ٠,٣٣٣٢^{٢})(١ - ٠,٥٠٢٩^{٢})}} = ٠,١٢٣٣$$

ونختبر دلالة هذا المعامل من الصيغة (٢١) كالآتى :

$$ت_{ى} = \frac{٠,١٢٣٣\sqrt{١٣٩}}{\sqrt{١ - ٠,١٢٣٣^{٢}}} = ١,٤٦٥$$ بدرجات حرية ١٣٩

وهذه القيمة تقل عن القيمة الحرجة $ت_{٠,٠٥\ [١٣٦]}$ = ١,١٩٨٠ فهى ليست بذات دلالة عند المستوى ٠٥,٠ ، وتشير إلى أنه حين نستبعد عامل العمر لا نستطيع الاستدلال على وجود ارتباط خطى بين ضغط الدم وكلوسترول الدم . هذا مع ملاحظة أن كلا من ضغط الدم ومقدار الكلوسترول فى الدم يزداد بزيادة العمر ، ومن هنا نرى أن النتيجة التى حصلنا عليها فى (أولا) كانت نتيجة مضللة .

معاملات الارتباط الجزئى من مراتب أعلى :

إن معامل الارتباط الجزئى من النوع $ص_{رو-ل}$ يوصف بأنه معامل ارتباط جزئى من المرتبة الأولى of the first order لأنه يعطى معامل الارتباط بين متغيرين بعد استبعاد أثر متغير واحد من كل منهما . وبالمثل إذا كان لدينا أربعة متغيرات عشوائية $ص_١$ ، $ص_٢$ ، $ص_٣$ ، $ص_٤$ ، فإننا نرمز بالرمز $ص_{٤٣-٢١}$ لمعامل الارتباط الجزئى بين المتغيرين $ص_١$ ، $ص_٢$ بعد استبعاد أثر المتغيرين $ص_٣$ ، $ص_٤$ من كل منهما ، ونصف هذا المعامل حينئذ بأنه معامل ارتباط جزئى من المرتبة الثانية . ويحسب هذا المعامل من معاملات الارتباط الجزئية من المرتبة الأولى كالآتى :

$$ر_{٤٣-٢١} = \frac{ر_{٤-٢١} - ر_{٤-٣١} \,.\, ر_{٤-٣٢}}{\sqrt{(١ - ر^{٢}_{٤-٣١})\,(١ - ر^{٢}_{٤-٣٢})}} \qquad (٢٣)$$

أو كالآتى :

$$ر_{٤٣-٢١} = \frac{ر_{٤-٢١} - ر_{٣-٤١} \,.\, ر_{٣-٤٢}}{\sqrt{(١ - ر^{٢}_{٣-٤١})\,(١ - ر^{٢}_{٣-٤٢})}} \qquad (٢٤)$$

وتختبر دلالة هذا المعامل من الصيغة :

$$\text{ت} = \frac{\text{ر}_{\text{٢١-٤٣}}\sqrt{\text{ن} - ٤}}{\sqrt{١ - \text{ر}^{٢}_{\text{٢١-٤٣}}}} \qquad \text{بدرجات حرية ن} - ٤ \qquad (٢٥)$$

ويمتد مفهوم معاملات الارتباط الجزئية لأى عدد منتهى ك من المتغيرات العشوائية التى تشترك فى توزيع معتدل متعدد المتغيرات . ومعامل الارتباط الجزئى من المرتبة م (حيث م < ك − ٢) هو معامل الارتباط بين اثنين من المتغيرات بعد استبعاد أثر م من المتغيرات العشوائية الأخرى . ويعتمد حساب معامل الارتباط الجزئى من المرتبة م على معاملات الارتباط الجزئية من المرتبة السابقة عليها أى من المرتبة م − ١ . هذا ويجدر بنا ملاحظة التماثل فى صيغ هذه المعاملات . أما اختبار دلالة معامل الارتباط الجزئى من المرتبة م فهو تعميم للصيغتين (٢١) ، (٢٥) ويأخذ الصيغة الآتية :

$$\text{ت} = \frac{\text{ر}\sqrt{\text{ن} - ٢ - \text{م}}}{\sqrt{١ - \text{ر}^{٢}}} \qquad \text{بدرجات حرية ن} - ٢ - \text{م} \qquad (٢٦)$$

حيث ر هو معامل الارتباط الجزئى من المرتبة م .

تمارين (١٢ − ٢)

(١) فى عينة عشوائية من ٣٠ وحدة من مجتمع معتدل ذى ثلاثة متغيرات وجد أن معامل الارتباط المتعدد $\text{ر}_{\text{١(٣٢)}}$ = ٠,٥ . بيّن أن هذه القيمة تدل على وجود ارتباط فى المجتمع بين المتغير $\text{ص}_{١}$ والمتغيرين $\text{ص}_{٢}$ ، $\text{ص}_{٣}$ مستخدما مستوى الدلالة ٠,٠٥ .

(٢) فى عينة عشوائية حجمها ٣٩ من مجتمع معتدل ذى ثلاثة متغيرات وجد أن معاملات الارتباط البسيط بين أزواج هذه المتغيرات هى $ر_{٢١}$ = ٠,٢٨ ، $ر_{٣٢}$ = ٠,٤٩ ، $ر_{٣١}$ = ٠,٥١ . اثبت أن $ر_{٢-٣١}$ = ٠,٤٤٥ ، $ر_{١-٢٣}$ = ٠,٤٢ واختبر دلالة كل منهما .

(٣) بين أن معامل الارتباط الجزئى من المرتبة الثانية والذى قيمته $ر_{٢١-٤٣}$ = ٠,٥ المحسوب من عينة عشوائية حجمها ٢٠ من مجتمع معتدل ذى أربعة متغيرات يكون ذا دلالة عند المتسوى ٠,٠٥ .

(٤) فى عينة عشوائية من ٥٤ من السيدات حسبت المقادير المستهلكة فى فترة ما من نوعين من الغذاء هما البروتين ($\tilde{ص}_{١}$) والدهون ($\tilde{ص}_{٢}$) كما حسب العمر $\tilde{ص}_{٣}$ وقد وجدت معاملات الارتباط البسيط الآتية :

$ر_{٢١}$ = ٠,٥٧٨٤ ، $ر_{٣١}$ = −٠,٤٨٦٥ ، $ر_{٣٢}$ = −٠,٥٢٩٦ . أوجد معامل الارتباط الجزئى بين استهلاك البروتين واستهلاك الدهون مستقلا عن أثر العمر ، واختبر دلالته عند المستوى ٠,٠١ .