

الفصل العاشر

الارتباط الخطي البسيط

SIMPLE LINEAR CORRELATION

(١٠ - ١) الانحدار الخطي والارتباط الخطي :

في دراسة الانحدار الخطي لمتغير حقيقي y على متغير آخر x فرضنا أن العلاقة بينهما على الصورة $y = \alpha + \beta x$ واستخرجنا من العينة أحسن تقديرين α ، β للبارامترين المجهولين α ، β في ضوء مبدأ المربعات الصغرى ، ومن ثم أوجدنا معادلة $y = \alpha + \beta x$ يمكننا من التنبؤ بأحسن قيمة للمتغير y عند قيمة معطاة للمتغير x ، واستخدمنا تحليل التباين لتفسير الانحدار ، كما خرجنا ببعض الاستنتاجات الإحصائية في صورة اختبارات دلالة وفترات ثقة . وهذا كله يدخل تحت الموضوع المسمى بتحليل الانحدار . ويقترن بهذا الموضوع موضوع آخر يسمى تحليل الارتباط وهو يهتم بالبحث عن عدد نقيس به درجة الاعتماد المتبادل بين المتغيرين ودقة العلاقة المفروضة بينهما ، فإذا فرضنا أن العلاقة بين المتغيرين خطية يكون اهتمامنا منصّباً على إيجاد عدد أو مقياس يعبر عن درجة جودة العلاقة الخطية في وصف العلاقة الحقيقية بين المتغيرين واختبار دلالة هذا المقياس .

(١٠ - ٢) افتراضات الارتباط الخطي البسيط :

تقتضي دراسة الارتباط بين متغيرين حقيقيين x ، y وضع افتراضات يختلف بعضها عن تلك التي وضعت لدراسة الانحدار . وسنضع هنا الافتراضات الآتية :

الافتراض الأول :

« كل من المتغيرين s ، v هو متغير عشوائي » .

فمثلا قد يعبر المتغيران عن طول ذراع الإنسان وطول رجله ، أو عن عمر الزوج وعمر الزوجة عند الزواج ، أو عن وزن الدجاجة وعدد البيض الذى تنتجه أو عن درجة الرياضيات ودرجة الفيزياء لمجموعة من الطلاب ..

الافتراض الثانى :

« إن العلاقة بين المتغيرين s ، v هى علاقة خطية » بمعنى أن متوسط قيم v المناظرة لقيمة معينة s يأخذ الصورة

$$\mu_v | s = \alpha + \beta s \quad (1)$$

حيث α ، β بارامتران مجهولان .

الافتراض الثالث :

(أ) للمتغير s توزيع معتدل

(ب) عند أى قيمة ثابتة s يكون للمتغير v توزيع معتدل متوسطه $\alpha + \beta s$ وتباينه عدد ثابت مجهول σ^2 لا يتوقف على s .

وقد يكون من المفيد أن نشير إلى أن هذه الافتراضات تكافئ رياضيا القول بأن التوزيع المشترك للمتغيرين s ، v هو ذلك المسمى بالتوزيع المعتدل ذى المتغيرين

bivariate normal distribution

(١٠ - ٣) معامل الارتباط العزمى (بيرسون) :

PRODUCT MOMENT CORRELATION COEFFICIENT (PEARSON 1900)

تعريف :

إذا كان s ، v متغيرين عشوائيين ورمزنا بالرمز σ_s للانحراف المعيارى

للمتغير s وبالرمز σ_s للانحراف المعياري للمتغير s وبالرمز σ_{ss} لتغاير (s, s) فإن العدد σ_{ss} المعروف بالصيغة :

$$\frac{\sigma_s}{\sigma_{ss}} = \sigma_s$$

يسمى بمعامل الارتباط العزمي بين s ، s .

يمكن رياضيا إثبات ما يلي :

(١) لا تزيد القيمة المطلقة للعدد σ_{ss} عن الواحد الصحيح ، أى أن

$$-1 \leq \sigma_{ss} \leq 1$$

(٢) إذا كان المتغيران s ، s مستقلين فإن $\sigma_{ss} = 0$. فليس من الضروري أن يكون المتغيران مستقلين بل قد يكون بينهما علاقة غير خطية . أما إذا كان التوزيع المشترك للمتغيرين s ، s هو التوزيع المعتدل ذو المتغيرين فإن انعدام معامل الارتباط يستلزم استقلال المتغيرين .

(٣) تكون هناك علاقة دالية خطية بين المتغيرين s ، s :

$$s = a + b \cdot s \quad \text{و} \quad s = a' + b' \cdot s$$

إذا وإذا فقط كان معامل الارتباط σ_{ss} يساوى ١ أو -١ .

وهذه الخاصة تعنى أنه إذا كان هناك علاقة خطية بين s ، s فإن ذلك ينعكس على قيمة σ_{ss} فيجعلها مساوية للعدد ١ (ونقول حينئذ أن هناك ارتباط تام موجب بين المتغيرين) أو مساوية للعدد -١ (ونقول حينئذ أن هناك ارتباط تام سالب) ، والعكس صحيح .

من هذا نرى أن معامل الارتباط العزمي ليس مقياسا للاعتماد المتبادل بين

المتغيرين بشكل عام وإنما هو مقياس لدرجة الاعتماد الخطي بينهما ، وينبغي أن نتذكر ذلك دائما .

ولتقدير قيمة r من عينة عشوائية :

$\{(s_1, v_1), (s_2, v_2), \dots, (s_n, v_n)\}$ نستخدم المقياس الآتي الذي يتركب بكل بساطة من التقديرات غير المتحيزة للقيم التي يتركب منها r :

$$(2) \quad r = \frac{\sum s_i v_i}{\sum s_i \cdot \sum v_i}$$

حيث $\sum s_i v_i = \frac{1}{n-1} \sum (s_i - \bar{s})(v_i - \bar{v})$ تغاير (s, v)

$$(3) \quad \frac{1}{n-1} \sum (s_i - \bar{s})(v_i - \bar{v}) = \frac{1}{n-1} \sum (s_i v_i - \bar{s} v_i - \bar{v} s_i + \bar{s} \bar{v})$$

، $\sum s_i v_i = \frac{1}{n-1} \sum (s_i - \bar{s})(v_i - \bar{v})$ تباين (s)

$$(4) \quad \frac{1}{n-1} \sum (s_i - \bar{s})(v_i - \bar{v}) = \frac{1}{n-1} \sum (s_i v_i - \bar{s} v_i - \bar{v} s_i + \bar{s} \bar{v})$$

، $\sum s_i v_i = \frac{1}{n-1} \sum (s_i - \bar{s})(v_i - \bar{v})$ تباين (v)

ملاحظة (١)

من السهل إثبات أن (٢) يمكن أن تكتب على الصورة

$$(5) \quad r = \frac{\sum s_i v_i - \bar{s} \sum v_i - \bar{v} \sum s_i + n \bar{s} \bar{v}}{\sqrt{[\sum s_i^2 - n \bar{s}^2][\sum v_i^2 - n \bar{v}^2]}}$$

وهذه الصيغة أفضل من الصيغة (٢) من الناحية الحسابية .

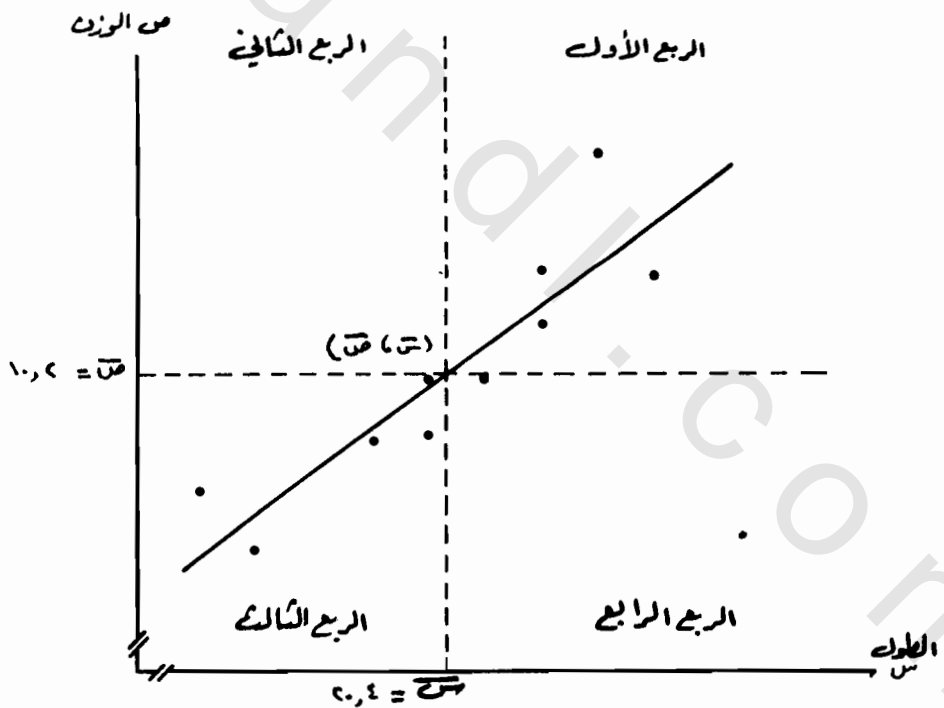
وقبل أن نوضح خصائص معامل الارتباط r والحكمة في أخذه كمقياس للارتباط الخطي نتناول المثال الآتي :

مثال (١٠ - ١) :

أخذت عينة عشوائية من نوع معين من نبات البسلة وقيست أطوالها بالمليمتر (س) وأوزانها بالمليجرام (ص) فوجد ما يلي :

س :	١٧	٢٢	٢٠	٢٣	٢٤	٢٢	٢٠	١٩	٢١	١٦
ص :	٧	١٢	٩	١٤	١٢	١١	١٠	٩	١٠	٨

ارسم شكل الانتشار لاستيضاح خطية العلاقة بين المتغيرين ثم أوجد معامل الارتباط العزمي .



الشكل (١٠ - ١) شكل الانتشار لأطوال وأوزان عينة من ١٠ نباتات بسلة

يوحى شكل الانتشار بأن النقاط تميل إلى أن تقع على خط مستقيم موجب الميل ، وهذا يشير مبدئيا إلى خطية العلاقة بين المتغيرين .

الجدول (١٠ - ١)

إيجاد معامل الارتباط العزمي من بيانات المثال (١٠ - ١)

س	ص	س ص	س ^٢	ص ^٢
١٧	٧	١١٩	٢٨٩	٤٩
٢٢	١٢	٢٦٤	٤٨٤	١٤٤
٢٠	٩	١٨٠	٤٠٠	٨١
٢٣	١٤	٣٢٢	٥٢٩	١٩٦
٢٤	١٢	٢٨٨	٥٧٦	١٤٤
٢٢	١١	٢٤٢	٤٨٤	١٢١
٢٠	١٠	٢٠٠	٤٠٠	١٠٠
١٩	٩	١٧١	٣٦١	٨١
٢١	١٠	٢١٠	٤٤١	١٠٠
١٦	٨	١٢٨	٢٥٦	٦٤
٢٠٤	١٠٢	٢١٢٤	٤٢٢٠	١٠٨٠

معامل الارتباط العزمي هو (بالتعويض في ٥)

$$= \frac{(204 \times 102) - (2124 \times 10)}{[(102)^2 - 1080 \times 10] [(204)^2 - 4220 \times 10]}$$

$$= \frac{432}{396 \times 584}$$

$$= \frac{432}{231264}$$

$$= 0.0018699$$

٤١٨

(١٠ - ٤) مميزات معامل الارتباط العزمي .

(١) تركيب معامل الارتباط :

إن قدرة معامل الارتباط المعروف في (٢) على تقدير درجة العلاقة الخطية بين المتغيرين s ، v تبين من الدراسة الرياضية للنموذج الذي وضعت له الافتراضات المذكورة آنفاً . غير أننا نستطيع أن نرى ذلك بطريقة بسيطة كالآتي .

إذا تأملنا نقط مستقيم موجب الميل ، لاحظنا أنه كلما كان الإحداثي السيني لنقطة كبيراً كلما كان إحداثيها الصادي كبيراً أيضاً . أما إذا كان المستقيم سالب الميل فإنه كلما زاد الإحداثي السيني كلما نقص الإحداثي الصادي . وعلى ذلك فإن قياس خطية العلاقة بين المتغيرين تتطلب مقياساً حساساً لدرجة اقتران القيم الكبيرة للمتغير s بالقيم الكبيرة للمتغير v إذا كانت العلاقة موجبة ، ولدرجة اقتران القيم الكبيرة للمتغير s بالقيم الصغيرة للمتغير v إذا كانت العلاقة سالبة وهذه الحساسية نجدها في التباين $s-v$ ففي المثال (١٠ - ١) السابق ، إذا رسمنا الخط الرأسي $s = \bar{s} = ٢٠,٤$ والخط الأفقي $v = \bar{v} = ١٠,٢$ في شكل الانتشار لاحظنا أنه في أغلب الحالات بل في جميع الحالات ما عدا حالة واحدة هي حالة النقطة (٢١ ، ١٠) الواقعة في الربع الرابع ما يلي :

حين تكون s كبيرة (أكبر من الوسط الحسابي $\bar{s}$) فإن v المناظرة تكون كبيرة أيضاً (أكبر من الوسط الحسابي $\bar{v}$) وحين تكون s صغيرة (أصغر من $\bar{s}$) فإن v المناظرة تكون صغيرة أيضاً (أصغر من $\bar{v}$) . أي أن :

حين ($s - \bar{s}$) موجبة تكون ($v - \bar{v}$) موجبة وحين ($s - \bar{s}$) سالبة تكون ($v - \bar{v}$) سالبة وفي كلتا الحالتين تكون ($s - \bar{s}$) ($v - \bar{v}$) موجبة ويكون المجموع $(s - \bar{s})(v - \bar{v})$ موجب .

وبالتالى يكون التغير $\bar{E}_s$ موجبا (إلا إذا كانت قيمة $(s - \bar{s})$)
(ص - $\bar{s}$) للنقطة (٢١ ، ١٠) كبيرة جداً وهذا لم يحدث) .

ويلاحظ أن قيمة $\bar{E}_s$ في هذا المثال كبيرة لأنها متوسط مجموع ٩ أعداد موجبة وعدد واحد سالب . ولو كانت النقطة التي في الربع الرابع قد وقعت في الربع الأول أو الثالث لزادت قيمة $\bar{E}_s$ وبالعكس لو كانت إحدى النقط الواقعة في الربعين الأول أو الثالث قد وقعت في الربع الثاني أو الرابع لنقصت قيمة $\bar{E}_s$

وعلى ذلك فإن التغير يعكس أمرين هما : درجة جودة العلاقة الخطية واتجاه هذه العلاقة . ولهذا يركز معامل الارتباط r المعروف في (٢) على التغير $\bar{E}_s$. هذا مع ملاحظة أن هذا المعامل يكون موجباً أو سالباً بحسب كون العلاقة الخطية موجبة أو سالبة .

غير أن قيمة التغير تعتمد على حجم العينة n كما هو واضح من التعريف (٣) ، كما أنها تعتمد على وحدات القياس . ولكي يكون مقياس الارتباط عاماً ينبغي أن يكون مستقلاً عن حجم العينة وعن وحدات القياس . ولذلك ينبغي تعديل التغير ليحقق هذين الشرطين قبل أخذه كمقياس للارتباط ، وهذا ما يحققه قسمة التغير $\bar{E}_s$ على حاصل ضرب الانحراف المعياري $\bar{E}_s$ للمتغير s والانحراف المعياري $\bar{E}_s$ للمتغير s . وهذا الإجراء يكافئ إيجاد تغير s ، s بعد وضع القيم في الصورة المعيارية $\frac{s - \bar{s}}{\bar{E}_s}$ ، $\frac{\bar{s} - \bar{s}}{\bar{E}_s}$

بهذا الإجراء يكون المقياس .

$r = \frac{\bar{E}_s}{\bar{E}_s \cdot \bar{E}_s}$ هو مقياس مطلق (لا يعتمد على وحدات القياس) ولا يعتمد على حجم العينة .

(ب) من مزايا معامل الارتباط أنه متماثل في r ، ص بمعنى أن قيمته لا تتغير بوضع r ، ص كل مكان الآخر .

(ج) قيمة معامل الارتباط r مستقلة عن اختيار نقطة الأصل لأن كلا من التغيرات والانحراف المعياري يتمتع بهذه الصفة ، وهذا يعني أن قيمة r لا تتغير بإضافة أو طرح عدد ثابت من جميع قيم r أو بإضافة أو طرح عدد ثابت من جميع قيم r . ونستفيد من هذه الخاصية أحياناً في تبسيط حساب قيمة r .

(د) يمكن إثبات أن القيمة المطلقة للمقياس r لا تزيد عن الواحد الصحيح .

أى أن .

$$(٦) \quad -1 \leq r \leq 1$$

ولما كانت r مقياساً لدرجة الارتباط الخطي بين متغيرين فإننا نقول إن المتغيرين غير مرتبطين خطياً إذا كانت $r = 0$. (وهذا لا يمنع من وجود ارتباط من نوع آخر) . وكلما اقتربت r من الواحد كلما أوحى ذلك بوجود علاقة خطية (موجبة أو سالبة) بين المتغيرين .

ملاحظة (٢)

هناك علاقة مفيدة بين الخطأ المعياري σ_r المعروف بالمعادلة (٨) بالبند (٩ - ٤) في موضوع الانحدار الخطي البسيط ومعامل الارتباط r وهى :

$$(٧) \quad \sigma_r = \frac{1 - r^2}{n - 2} \sigma_{\epsilon} \quad (١ - r^2)$$

ففي المثال (١٠ - ١) نستطيع إيجاد الخطأ المعياري كالآتي :

$$\sigma_r = \frac{1}{n - 2} [\text{م.ح. ص}^2 - \frac{(\text{م.ح. ص})^2}{n}] = \frac{1}{9} (10.2 - 10.80) = 0.44$$

$$\text{إذن } \bar{r} = \frac{9}{8} \times 4,4 (1 - 0,898) = 0,9583$$

$$\text{إذن } \bar{r} = 0,9789 \text{ تقريباً .}$$

(١٠ - ٥) دلالة معامل الارتباط العزمي :

ليكن r هو معامل الارتباط الخطي بين المتغيرين s ، r في المجتمع ، r هي معامل الارتباط الناتج من العينة . تحت الافتراضات الثلاثة المذكورة نستطيع اختبار أى فرض عن القيمة الحقيقية للمعامل r ووضع حدود ثقة لهذا المعامل .

(أولاً) اختبار الفرض $r = \text{صفر}$.

لاختبار الفرض الصفري $F : r = \text{صفر}$ (المتغيران غير مرتبطين خطياً)
ضد الفرض الآخر $F : r \neq \text{صفر}$
نستخدم الإحصاءة

$$(٨) \quad T = \frac{\sqrt{n-2} \cdot r}{\sqrt{1-r^2}}$$

التي لها توزيع T بدرجات حرية $n - 2$ بشرط صحة الفرض الصفري .

مثال (١٠ - ٢) :

في المثال (١٠ - ١) اختبر ما إذا كان المتغيران s ، r غير مرتبطين خطياً مستخدماً مستوى الدلالة ٠,٠١ .

الحل :

$$\text{لدينا } n = 10 , r = 0,898$$

ف : م = ٠ ، ف : م ≠ ٠ (اختبار ذو جانبيين)

$$\text{من (٨) : ت } = \frac{\sqrt{2-10} \cdot ٠,٨٩٨}{\sqrt{٠,٨٩٨ - 1}} = ٥,٧٧٣$$

من الجدول ت $_{[٨]٠,٠١} = ٣,٣٥٥$

بما أن $٣,٣٥٥ < ٥,٧٧٣$ نرفض الفرض الصفري عند مستوى الدلالة $٠,٠١$ ونحكم بوجود ارتباط خطي بين المتغيرين .

ملاحظة (٣)

الإحصاءة (٨) التي نختبر الفرض م = ٠ ضد الفرض م ≠ ٠ تكافئ الإحصاءة (١٠) بالبند (٩ - ٥) في موضوع الانحدار الخطي البسيط عند استخدامها لاختبار الفرض م = ٠ ضد الفرض م ≠ ٠ وذلك لأن كلاهما يقيس وجود أو عدم وجود علاقة خطية بين المتغيرين م ، م . وقد يبدو ذلك غريبا لأن الإحصاءة (٨) تتطلب أن يكون كلا المتغيرين معتدلا بينما الإحصاءة (١٠) تتطلب أن يكون المتغير م فقط معتدلا . وتزول هذه الغرابة إذا علمنا أن فيشر قد أثبت أنه في الحالة الخاصة التي يكون فيها م = ٠ فإن توزيع المعاينة لمعامل الارتباط م لا يتغير سواء كان المتغير م معتدلا أو غير معتدل طالما كان المتغير م معتدلا .

ولذلك فإن الإحصاءة (٨) تصلح فقط لاختبار الفرض الصفري م = ٠ ولا تصلح لاختبار أى فرض صفري آخر مثل م = ٥ ، ٠ أو لايجاد فترات الثقة لمعامل الارتباط م في المجتمع ، وذلك لأنه حين يكون المتغيران مرتبطين ، أى حين م ≠ ٠ يكون توزيع م ملتويا وبعيدا عن الاعتدالية . فكيف نختبر الفرض م = ك حيث ك ≠ ٠ ؟

ومن الفترة (١٢) نستطيع إيجاد فترة الثقة لمعامل الارتباط ρ للمجتمع من التحويل (٩) .

ونتجنب مشقة حساب التحويل (٩) وُضع جدول يحول r إلى E (بصرف النظر عن الإشارة) هو الجدول (١١) بملحق هذا الكتاب كما وضع جدول يحول E إلى r هو الجدول (١٢) .

مثال (١٠ - ٣) :

في عينة عشوائية من ٢٨ زوجا من مجتمع معتدل ذى متغيرين وجد أن معامل الارتباط ٠,٧ هل هذه القيمة تتمشى مع الفرض القائل أن معامل الارتباط في المجتمع هو ٠,٥ ؟ أوجد فترة ثقة بدرجة ٩٥٪ لهذا المعامل .

الحل :

الفرض الصفري : $\rho = ٠,٥$ الفرض الآخر $\rho \neq ٠,٥$

نحول كلا من معامل الارتباط في العينة ومعامل الارتباط المفروض للمجتمع إلى القيمتين المناظرتين لهما باستخدام الجدول (١١) .

$$r = ٠,٧ \text{ تنتج } E = ٠,٨٦٧$$

$$\rho = ٠,٥ \text{ تنتج } E = ٠,٥٤٩$$

$$\text{الخطأ المعياري} = \frac{1}{\sqrt{25}} = \frac{1}{\sqrt{3-1}} = ٠,٢$$

على أساس صحة الفرض الصفري نجد أن :

$$\therefore \text{ص} = \frac{٠,٥٤٩ - ٠,٨٦٧}{٠,٢} = ١,٥٩$$

وبما أن ١,٥٩ أقل من القيمة الحرجة ١,٩٦ لا يكون لدينا دليل يدعونا لرفض الفرض الصفري ، ونحكم بأن معامل الارتباط في المجتمع يمكن أن يكون ٠,٥ من الصيغة (١٢) ، نوجد فترة الثقة بدرجة ٩٥٪ للبارامتر ρ كالآتي :

الحد الأدنى للفترة $0,867 = 0,2 \times 1,96 - 0,475 = 0,48$ تقريبا

الحد الأعلى للفترة $0,867 = 0,2 \times 1,96 + 0,475 = 1,26$ تقريبا

∴ فترة الثقة للبارامتر μ هي $(0,48, 1,26)$.

من الجدول (١٢) نجد أن $(0,446, 0,851)$ هي فترة الثقة بدرجة ٩٥٪ لمعامل الارتباط.

مثال (١٠ - ٤) :

في عينة من ٩ أزواج من المشاهدات وجد أن معامل الارتباط $-0,889$ أوجد فترة ثقة بدرجة ٩٩٪ لمعامل الارتباط في المجتمع.

الحل :

من الجدول (١١) : $-0,889 = r$ وإذن $1,417 = c$ (مع إهمال إشارة r)

$$\text{الخطأ المعياري} = \frac{1}{3 - \sqrt{n}} = \frac{1}{\sqrt{6}} = 0,408$$

الحد الأدنى للفترة للبارامتر μ $0,366 = 2,58 \times 0,408 - 1,417 = c$

الحد الأعلى للفترة للبارامتر μ $2,468 = 2,58 \times 0,408 + 1,417 = c$

∴ الفترة $(0,366, 2,468)$ هي فترة ثقة بدرجة ٩٩٪ للبارامتر μ . من الجدول

(١٢) ومع استرجاع إشارة r تكون الفترة $(-0,986, -0,354)$ هي فترة ثقة

بدرجة ٩٩٪ لمعامل الارتباط ρ .

(ثالثا) اختبار دلالة الفرق بين معاملي ارتباط عيتين مستقلتين :

نفرض أن لدينا عيتين مستقلتين حجمهما n_1 ، n_2 من الأزواج أعطيتا معاملي ارتباط r_1 ، r_2 . نريد أن نختبر ما إذا كان من الممكن اعتبار هاتين العينتين مأخوذتين من نفس المجتمع (أو من مجتمعين لهما نفس معامل الارتباط) ، أى نريد أن نختبر ما إذا كان الفرق بين r_1 ، r_2 ذا دلالة .

نحول القيمتين r_1 ، r_2 إلى القيمتين المناظرتين E_1 ، E_2 باستخدام الجدول (١١) . إذا كانت العينتان مستقلتين ومأخوذتين من نفس المجتمع المعتدل فإن الفرق بين E_1 ، E_2 يكون له توزيع معتدل تباينه يساوى مجموع تباينهما ، أى

$$\sigma_{E_1 - E_2}^2 = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3} \quad (١٣)$$

ويكون الخطأ المعياري للفرق $|E_1 - E_2|$ هو الجذر التربيعي لهذا المقدار . وعلى ذلك نستطيع اختبار دلالة الفرق بين E_1 ، E_2 (أى بين r_1 ، r_2) بواسطة التوزيع المعتدل .

مثال (١٠ - ٥) :

في عينة عشوائية من ٥ أبقار وجد أن معامل الارتباط بين الزيادة في الوزن ومقدار الغذاء المأكول ٠,٨٧ ، وفي عينة عشوائية مستقلة حجمها ١٢ بقرة وجد أن معامل الارتباط ٠,٥٦ ، فهل هذان المعاملان مختلفان اختلافا جوهريا ؟ استخدم مستوى الدلالة ٥٪ .

الحل :

الفرض الصفري $r = r_0$ والفرض الآخر $r \neq r_0$

من الجدول (١١) نجد أن $r_1 = ٠,٨٧$ ومنها $E_1 = ١,٣٣٣$

$r_2 = ٠,٥٦$ ومنها $E_2 = ٠,٦٣٣$

تباين الفرق $\sigma_c - \sigma_p$ يساوى مجموع التباينين

$$0,611 = \frac{1}{9} + \frac{1}{9} =$$

$$\therefore \text{الخطأ المعياري للفرق} = \sqrt{0,611} = 0,782$$

على أساس صحة الفرض الصفري $\mu_p = \mu_c$ (أى $\sigma_c = \sigma_p$)

$$\text{ص} = \frac{(1,333 - 0,633) - \text{صفر}}{0,782} = 0,895$$

وهذه القيمة تقل عن ١,٩٦ فهي ليست ذات دلالة ولا يسعنا ألا أن نحكم بأن $\mu_p = \mu_c$.

(١٠ - ٦) التمييز بين الانحدار والارتباط في دراسة المشكلات :

كل من الانحدار والارتباط الخطى البسيط يتناول العلاقة الخطية بين متغيرين كميين μ ، σ . إلا أن هناك مواقف تستدعى استخدام أسلوب الانحدار ولا يجوز أن تُدرس بأسلوب الارتباط ، كما أن هناك مواقف تستدعى استخدام أسلوب الارتباط ولا يجوز أن تُدرس بأسلوب الانحدار . ذلك أن الافتراضات التي يبنى عليها التحليل تختلف في الانحدار عنها في الارتباط خاصة فيما يتعلق بنوعية المتغير μ ، حيث نفترض في الانحدار أن μ متغير رياضى لا يتأثر بالعوامل العشوائية بينما نفترض في الارتباط أنه عشوائى . وهذا الاختلاف في النظر إلى المتغير μ يعنى من الناحية العملية اختلافا في طريقة المعاينة . وعلى ذلك فإن اختيار الأسلوب الذى يناسب مشكلة ما - انحدار أو ارتباط - يتوقف على الطريقة التى تتبع فى عملية المعاينة أى فى عملية جمع البيانات .

فلاستخدام أسلوب الانحدار يتطلب الأمر أن يختار الباحث قيما ثابتة من المتغير μ يحددها قبل إجراء التجربة ثم يقوم بملاحظة ما يظهر من القيم المناظرة للمتغير

ص عند إجراء التجربة . فمثلا في دراسة العلاقة بين جرعات دواء مهدئ وقدرة الانسان على حل المشاكل المنطقية يبدأ الباحث بتحديد بضعة جرعات مختلفة التركيب من هذا الدواء (قيم المتغير ص) ثم يختار عينة عشوائية من الأفراد يقسمها عشوائيا إلى مجموعات عددها هو عدد الجرعات المختلفة ويعطى لأفراد كل مجموعة واحدة من تلك الجرعات ، ثم يستخدم أحد الاختبارات لقياس القدرة المنطقية لكل فرد فيحصل على قيم للمتغير ص . كذلك في دراسة العلاقة بين طول نبات ومحتوى النيتروجين في التربة يبدأ الباحث بتحديد عدة أحواض زراعية تختلف في محتوى النيتروجين (قيم ص) ويزرع النبات في كل منها ثم يقيس أطوال النباتات بعد فترة من الزمن ليحصل على قيم للمتغير ص . في مثل هاتين الحالتين تكون قيم ص فقط خاضعة للمؤثرات العشوائية ، أما قيم ص فتكون قيما ثابتة . وينبغي هنا تناول البيانات بأسلوب الانحدار حيث نقوم بإيجاد معادلة تشير إلى مدى اعتماد المتغير ص على المتغير ص ، ويحق لنا أن نصف ص بأنه المتغير المستقل وأن نصف ص بأنه المتغير التابع .

أما في أسلوب الارتباط فيتطلب الأمر أن يبدأ الباحث باختيار عينة عشوائية من المجتمع ثم يقوم بقياس كل من قيم ص ، ص لكل وحدة من وحدات العينة وبذلك تكون جميع القيم السينية والصادية خاضعة للمؤثرات العشوائية إذ يكون لكل منها حرية تامة في اتخاذ أى قيمة من القيم الممكنة في المجتمع . فمثلا في دراسة العلاقة بين طول الذراع وطول الرجل لمجتمع من الأطفال ، يختار الباحث عينة عشوائية من الأطفال ثم يقوم بقياس كل من المتغيرين . كذلك في دراسة العلاقة بين محتوى الكوليسترول في الدم ووزن الجسم في مجتمع من المرضى بمرض معين نأخذ عينة عشوائية من هؤلاء المرضى ثم نقيس كلا من هذين المتغيرين . في مثل هاتين الحالتين تكون قيم كل من المتغيرين ص ، ص عشوائية . وينبغي هنا استخدام أسلوب الارتباط حيث نقوم بإيجاد عدد نقدر به الدرجة التي يتغير بها المتغيران معا دون تمييز بين متغير مستقل وآخر تابع .

وفي الحالات التي تتطلب دراستها استخدام أسلوب الانحدار يمكن من الناحية الحسابية إيجاد معامل الارتباط ، ولكن المعامل الناتج يكون مجرد عدد لا معنى له ولا يجوز أن يؤخذ كتقدير للارتباط بين المتغيرين . (وهذا لا يتناقض مع استخدامنا لمربع معامل الارتباط وهو ما سميناه بمعامل التحديد ، في تحليل الانحدار كما رأينا في الفصل السابق) .

كذلك ، في الحالات التي تتطلب دراستها استخدام أسلوب الارتباط يمكن من الناحية الحسابية إيجاد معادلة الانحدار ، ولكن المعادلة الناتجة لا تحقق الهدف منها كمعادلة تنبؤ أو لبيان العلاقة الخطية بين المتغيرين لأن التقديرين أ ، ب يكونان في هذه الحالة تقديرين متحيزين للبارامترين α ، β . ويتضح ذلك عندما نتذكر أننا في قياس دقة المقدار ب في تقدير البارامتر β نستخدم الخطأ المعياري σ_{β} / $\sqrt{22}$ (س) الذي يعتمد على قيمة σ^2 (س) = σ^2 (س - س) - راجع البند (٩ - ٥ - أولا) فإذا لم تكن قيم σ^2 ثابتة فإن قيمة هذا الخطأ المعياري تختلف من عينة إلى أخرى وبالتالي تختلف درجة دقة التقدير من عينة إلى أخرى .

إن التمييز بين الحالات التي تدرس بأسلوب الانحدار وتلك التي تدرس بأسلوب الارتباط هو أمر على درجة كبيرة من الأهمية . وقد لوحظ أن الأمر يختلط في كثير من الحالات فتعامل مشكلات الانحدار على أنها ارتباط وتعامل مشكلات الارتباط على أنها انحدار ، بل ويصل الأمر أحيانا إلى معاملة المشكلة الواحدة على أنها انحدار وارتباط في آن واحد ، رغم أن طريقة المعاينة لا تسمح إلا باستخدام واحد فقط من هذين الأسلوبين . ويبدو أن السبب في هذا النبس يرجع إلى وجود علاقات رياضية كثيرة بين المعاملات في هذين النوعين من التحليل ، غير أن العلاقات الرياضية شيء والتحليل الإحصائي شيء آخر .

وجدير بالذكر أنه من الممكن دراسة الانحدار حين يكون كل من المتغيرين σ^2

صه عشوائيا ، ولكن ذلك يتطلب استخدام نموذج إحصائي يختلف عن النموذج الذى استخدمناه فى الفصل السابق بحيث يسمح بوجود خطأ عشوائى فى المتغير صه . وسوف لا نتعرض لدراسة هذا النموذج خاصة وأنه يفضل دراسة مثل هذه الحالة بأسلوب الارتباط .

(١٠ - ٧) معامل ارتباط الرتب (سبيرمان) :

RANK CORRELATION COEFFICIENT

بالرغم من أن معامل ارتباط بيرسون يستخدم أساسا فى حالة المتغيرات المتصلة إلا أنه يستخدم أيضا فى حالات أخرى . ومن هذه الحالات الحالة التى يتعذر فيها قياس المفردات بالمقياس العددي المعتاد وإنما يمكن قياسها بميزان الترتيب حيث تأخذ كل مفردة ترتيبين س ، ص تعبر الأولى عن ترتيب المفردة بالنسبة للمتغير الأول وتعبر الثانية عن ترتيبها بالنسبة للمتغير الثانى ويكون المطلوب قياس الارتباط بين الترتيبين . ومثال ذلك قيام اثنين من الحكام بترتيب عدد من المتقدمين لشغل وظيفة بحسب أفضليتهم لهذه الوظيفة ثم قياس مدى الاتفاق بين الحكامين . فى هذه الحالة يمكن إثبات أن معامل الارتباط العزمى يأخذ الصيغة البسيطة الآتية التى تعزى إلى سبيرمان ويرمز له بالرمز r_s حيث

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad \text{وحيث } d = s - ص \quad (١٤)$$

مثال (١٠ - ٣) :

كانت تراتيب ١٢ طالبا فى مادتي الفيزياء س والرياضيات ص كما يلى :

س	١١	٧	٩	١٢	١	٤	١٠	٣	٨	٥	٢	٦
ص	١٠	٥	٨	١١	٦	١٢	٣	٧	٤	٢	٩	١

أوجد معامل الارتباط الخاص .

الحل :

بما أن المتغيرين مقاسان بميزان الترتيب فمن الأسهل استخدام معامل ارتباط الرتب كما في الجدول الآتي ، الذي يستخدم الصيغة (١٤) .

الجدول (١٠ - ٢)

إيجاد معامل ارتباط الرتب من بيانات الخال (١٠ - ٣)

س	ص	ف	ف ^٢
١١	١٠	١	١
٧	٥	٢	٤
٩	٨	١	١
١٢	١١	١	١
١	١	٠	٠
٤	٦	٢-	٤
١٠	١٢	٢-	٤
٣	٣	٠	٠
٨	٧	١	١
٥	٤	١	١
٢	٢	٠	٠
٦	٩	٣-	٩
		صفر	٢٦

بالتعويض في (١٤) نجد أن :

$$r_{\text{س}} = 1 - \frac{26 \times 6}{143 \times 12} = 0,91$$

ونحصل على هذه النتيجة بالضبط إذا استخدمنا معامل ارتباط بيرسون المعروف في (٥) ، مع ملاحظة أنه قد يوجد فرق طفيف يعود إلى عمليات تقريب الأعداد .

(١٠ - ٨) مميزات معامل ارتباط الرتب :

كما سبق القول ، يصلح معامل ارتباط الرتب المعروف في (١٤) لقياس الارتباط بين المتغيرات التي تقاس بميزان الترتيب وهو يصلح أيضاً للمتغيرات التي تقاس بالمقياس العددي المعتاد ولكن يكون من المرغوب فيه تعيين رتب لكل مفردة بدلا من قيمها العددية . ومن أهم مميزات هذا المعامل أنه لا يشترط فرض اعتدالية المتغيرين r_s ، $r_{\text{س}}$ ، بل لا يشترط فرض وجود علاقة خطية بينهما ومن هنا فهو يفضل معامل الارتباط العزى في الحالة التي لا يتوفر فيها هذان الشرطان ، وقياس الارتباط بواسطة معامل الرتب يعتبر من فصيلة المقاييس غير البارامترية أى التي لا تستلزم وضع فروض معينة على توزيعات المجتمعات .

ومن الأمثلة التي يستخدم فيها هذا المعامل في العلوم البيولوجية قياس الارتباط بين ترتيب انبثاق يرقات مجموعة من الحشرات وترتيب حجوما ، أو بين ترتيب استنبات مجموعة من النباتات وترتيب تزهيرها ، أو لقياس الاتفاق بين اثنين من البيولوجيين في ترتيبهما لمجموعة من الكائنات العضوية من حيث أكثرها شبهاً لصيغة معينة إلى أقلها شبهاً بها .

مثال (١٠ - ٤) :

في عشرة أنواع من السجائر وجدت المقادير الآتية بالمليجرام من القار والنيكوتين . حول هذه المقادير إلى رتب ثم أوجد معامل ارتباط الرتب لقياس درجة العلاقة بين محتويات القار والنيكوتين .

النوع : (١) (٢) (٣) (٤) (٥) (٦) (٧) (٨) (٩) (١٠)

محتوى القار : ١٤ ١٧ ٢٨ ١٧ ١٦ ١٣ ٢٤ ٢٥ ١٨ ٣١

محتوى النيكتون : ٠,٩ ١,١ ١,٦ ١,٣ ١,٠ ٠,٨ ١,٥ ١,٤ ١,٢ ٢,٠

الحل :

نرتب كلا من مفردات المجموعتين بحسب الأصغر فالأكبر (أو العكس) كما في الجدول الآتي ، حيث وضعت الترتيب بدلا من القيم المعطاة .

الجدول (١٠ - ٣)

النوع	س	ص	ف	ف ^٢
(١)	٢	٢	٠	٠
(٢)	٤,٥	٤	٠,٥	٠,٢٥
(٣)	٩	٩	٠	٠
(٤)	٤,٥	٦	١,٥-	٢,٢٥
(٥)	٣	٣	٠	٠
(٦)	١	١	٠	٠
(٧)	٧	٨	١-	١
(٨)	٨	٧	١	١
(٩)	٦	٥	١	١
(١٠)	١٠	١٠	٠	٠
			صفر	٥,٥

$$\text{بالتعويض في (١٤) نجد أن } ١ - \frac{٥,٥ \times ٦}{٩٩ \times ١٠} = ٠,٩٦٦٧$$

ملاحظة (٤) :

إذا كان لمفردتين أو أكثر نفس القيمة العددية ، تعطى لكل منها رتبة تساوى الوسط الحسابي للرتب التي كانت ستأخذها هذه المفردات لو أنها كانت مختلفة القيم ، فمثلاً في محتوى القار لدينا عدداً متساويان ١٧ ، ١٧ والمفروض أن أحدهما الرابع والآخر الخامس ولذلك أعطى لكل منهما الترتيب $\frac{1}{2}$ (٤ + ٥) = ٤,٥ .

ملاحظة (٥) :

إذا كانت البيانات أصلاً على هيئة أزواج من الترتيب كما في المثال (١٠ - ٣) فإن معامل الارتباط العزمي يساوى بالضبط معامل ارتباط الرتب . أما إذا كانت البيانات أصلاً على هيئة أزواج من القيم العددية كما في المثال (١٠ - ٤) ، فإن معامل الارتباط العزمي لها لا يساوى معامل ارتباط الرتب الناتج من تحويل هذه القيم إلى رتب .

(١٠ - ٩) دلالة معامل ارتباط الرتب :

إن الإحصاءة التي يعبر عنها معامل ارتباط الرتب r_s المعروف بالصيغة (١٤) توزيعها متماثل حول الصفر . ويمكن إثبات أنه إذا كان المتغيران x ، y مستقلين فإن هذا التوزيع يقترب من توزيع معتدل وسطه الحسابي صفر وتباينه $\frac{1}{n-1}$ حين تقترب n من اللانهاية . وعلى ذلك عندما يكون حجم العينة كبيراً ($n > 30$) نستطيع أن نختبر ما إذا كان هناك ارتباط ذو دلالة بين المتغيرين وذلك بحساب القيمة .

$$t_s = \frac{r_s - 0}{\sqrt{\frac{1}{n-1}}} = \frac{r_s}{\sqrt{\frac{1}{n-1}}}$$

ومقارنتها بالقيم الحرجة للتوزيع المعتدل المعياري ، مع ملاحظة أن الفرض الصفري هو $= 0$. والفرض الآخر هو $\neq 0$.

ففيها هذا القول صحيحاً ويكون التغير في أحد المتغيرين هو فعلاً سبب أو أحد أسباب التغير في الآخر كما هو الحال مثلاً بين درجة الحرارة وعدد ضربات القلب ، وهناك حالات لا يصح فيها الاستنتاج حيث يكون الارتباط الذي ظهر بين المتغيرين ناشئاً عن وجود متغير ثالث أو أكثر يؤثر فيهما معاً فيحدث هذا الارتباط ، كما هو الحال مثلاً بين النجاح في الرياضيات والنجاح في التاريخ حيث يؤثر في كل منهما الذكاء أو المثابرة أو البيئة المنزلية وما إلى ذلك . إن تفسير وجود الارتباط لا يكون بناء على قيمة معامل الارتباط فقط بل على ما لدينا من معلومات عن المتغيرين اللذين ندرسهما .

تمارين (١٠)

في كل من المسائل الخمسة الآتية أوجد معامل الارتباط العزمي (بيرسون) من العينة المعطاة واختبر ما إذا كان المتغيران س ، ص مرتبطين خطياً .

(١) س : ٣١٠ ٣٢٠ ٢٢٠ ٨٠ ٣٢٠ ١٠٠ ٢٣٠ ٣٨٤ ٤٥ ١٠٠ ١٧٩ ١٥٩
ص : ٩,٥٢ ١٧,٣٥ ١٥,٣٩ ٤,١٩ ١٥,٨١ ١,٤١ ١٤,٩٠ ٢٢,٧٠ ٢,٥٠ ١١,٣٠ ١٥,٢٠ ١٤,٤٠

حيث س وزن الخيشوم بالمللجرام ، ص وزن الجسم بالجرامات لعينة حجمها ١٢ من نوع من سرطان البحر (أبو جلمبو) .

(٢) س : ١٦ ٢١ ١٩ ٢٠ ٢٢ ٢٤ ٢٣ ٢٠ ٢٢ ١٧
ص : ٨ ١٠ ٩ ١٠ ١١ ١٢ ١٤ ٩ ١٢ ٧

حيث س طول نوع معين من نبات البسلة بالمليمترات ، ص الوزن بالمللجرام .

(٣) س : ٥٢ ٥٠ ٤٨ ٥٣ ٥٢ ٥٢ ٥١ ٤٧ ٥١ ٥٠ ٤٨ ٥٢

ص : ٣٧ ٣٤ ٣٤ ٣٧ ٣٦ ٣٦ ٣٥ ٣٥ ٣٤ ٣٤ ٣٤ ٣٦

س : ٥٤ ٥١ ٤٨ ٥٠ ٥٠ ٤٨ ٤٨ ٤٨ ٤٩ ٥٠ ٥٠ ٥٢

ص : ٣٨ ٣٦ ٣٤ ٣٣ ٣٥ ٣٥ ٣٣ ٣٤ ٣٤ ٣٤ ٣٥ ٣٦

حيث س طول جسم طفل حديث الولادة ، ص طول محيط رأسه
بالسنتيمترات .

(٤) س : ٥٨ ٦١ ٥٥ ٦٠ ٥٥ ٦٢ ٥٢ ٥٧ ٥٦ ٥٨

ص : ٤٤ ٤٥ ٤٥ ٤٢ ٤٠ ٤٤ ٤٤ ٤٢ ٤٣ ٤١

حيث س ، ص هما أكبر وأصغر قطر لبيض الدجاج بالمليمترات .

(٥) س : ٧,٥ ٢,١ ٣,٨ ٥,٢ ٦,١ ٥,٦ ٥,٩ ٤,٥ ٤,٣

ص : ١٠,٨ ١٤,١ ١٣,٢ ١١,٨ ١١,٤ ١١,٨ ١١,٦ ١٢,١ ١٢,٦

حيث س مقدار المطر في اليوم (٠,٠١ من السنتيمتر) ، ص مقدار مازال من
تلوث الهواء نتيجة للمطر (ميكروجرام / متر مكعب) . (لتسهيل الحساب اطرح
١٠٠ من قيم ص) .

(٦) طلب من اثنين من المحكمين ترتيب ١٠ أشخاص متقدمين لوظيفة ما
فكانت النتيجة كما يلي . أوجد معامل الارتباط (سيرمان) واختبر ما إذا كان هناك
ارتباط موجب بين المحكمين .

المحكم الأول : ٩ ٤ ٦ ٥ ٣ ١ ١٠ ٨ ٢ ٧

المحكم الثاني : ٨ ٤ ٩ ٦ ١ ٢ ٧ ١٠ ٣ ٥

(٧) الجدول الآتي يبين عدد الساعات (لعينة عشوائية من ١٠ طلاب) التي استذكر فيها هؤلاء الطلاب لاختبار ما وعدد الدرجات التي حصلوا عليها في هذا الاختبار :

عدد ساعات الدراسة :	٨	٥	١١	١٣	١٠	٥	١٨	١٥	٢	٨
درجة الاختبار :	٥٦	٤٤	٧٩	٧٢	٧٠	٥٤	٩٤	٨٥	٣٣	٦٥

حول هذه القيم إلى رتب ثم أوجد معامل ارتباط الرتب (سبيرمان) .
 اختبر ما إذا كان هناك ارتباط موجب بين عدد ساعات الدراسة ودرجة الاختبار .

(٨) في عيتين مستقلتين حجمهما ٢٣ ، ٢٨ من أزواج القيم وجد أن معامل الارتباط العزمي بين المتغيرين ٠,٥ ، ٠,٨ ، على الترتيب . اختبر ما إذا كان هذان المعاملان مختلفين اختلافا ذا دلالة .