المحور التاسع: الإرتباط والإنحدار

The ninth axis: correlation and decline

محتوى المحاضرة الحادية عشر

أولا: توزيعات المتغيرات ثنائية التغير (جداول التوافق والتكرارات المشتركة، الهامشية والشرطية)

ثانيا: الارتباط بين متغيرين مستمرين (سحابة النقاط ومعامل الارتباط الخطي، الانحدار الخطي البسيط).

تمهيد:

يعد الارتباط والانحدار الخطي البسيط من أكثر الموضوعات استخداماً في العمليات الإحصائية، والهدف من دراسة الارتباط هو الكشف عن قوة أو درجة العلاقة بين متغيرين أو أكثر، بينما الهدف الاساسي من تحليل الانحدار هو تقدير الصورة الرياضية للعلاقة بين متغير مستقل ومتغير تابع، ويستخدم تحليل الانحدار لدراسة مدى تأثير متغير مستقل واحد أو أكثر على المتغير التابع المحدد بحيث نستطيع التنبؤ بقيم المتغير التابع إذا علمنا قيم المتغير المستقل أو المتغيرات المستقلة.

أولا: توزيعات المتغيرات ثنائية التغير:

تستخدم هذه التوزيعات لإظهار العلاقة بين متغيرين أحدهما مستقل والآخر تابع وتربطهما تكرارات مشتركة.

- **المتغير المستقل:** يرمز له بالرمز x_i وهو الذي يلعب دورا في حدوث المتغير التابع.
- **المتغير التابع:** يرمز له بالرمز y_i وهو المتغير الذي نحاول تفسير وفهم أسباب حدوثه وتحديد مدى إمكانية التنبؤ به.

1- **أنواع الجداول التكرارية للمتغيرات الثنائية:**

أ- **جدول ذو مدخلين:** قبل وضع البيانات في هذا الجدول يجب ترتيبها تصاعديا.

مثال1:

يمثل التوزيع التالي أثر الأقدمية على حجم الإنتاج لمجموعة من العمال:

$(2.3)_4$ ،$(8.6)_2$ ،$(5.3)_4$ ،$(4.8)_3$ ،$(10.5)_6$

المطلوب: حدد المجتمع الإحصائي، المتغير المستقل والمتغير التابع، وضع البيانات في جدول ذو مدخلين.

الحل:

1- **تحديد المجتمع الإحصائي**: مجموعة من العمال

2- **المتغير المستقل**: حجم الإنتاج

3- **المتغير التابع**: الأقدمية

4- **وضع البيانات في جدول ذو مدخلين**:

- **ترتيب البيانات تصاعديا:** $(10.5)_6$، $(8.6)_2$، $(5.3)_4$، $(4.8)_3$، $(2.3)_4$

- جدول ذو مدخلين لتوزيع الأقدمية على حجم الإنتاج لمجموعة من العمال

x_i	y_j	f_{ij}
2	3	4
4	8	3
5	3	4
8	6	2
10	5	6
$\sum$	$\sum$	19

ب- <u>الجدول التكراري المشترك أو المزدوج</u>:

هو جدول تكراري يدرس متغيرين من نفس المجتمع في آن واحد، حيث توضع البيانات الإحصائية في هذا الجدول بالشكل التالي:

$X_i \backslash Y_j$	y_1	y_2	....	y_m	المجموع
x_1	f_{11}	f_{12}	....	f_{1m}	$\sum_{j=1}^{m} f_{1j}$
x_2	f_{21}	f_{22}	....	f_{2m}	$\sum_{j=1}^{m} f_{2j}$
. .	. .	. .	. .	. .	. .
x_n	f_{n1}	f_{n2}	....	f_{nm}	$\sum_{j=1}^{m} f_{nj}$
المجموع	$\sum_{i=1}^{n} f_{i1}$	$\sum_{i=1}^{n} f_{i2}$	....	$\sum_{i=1}^{n} f_{im}$	$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}$

حيث: تخصص الأسطر لبيانات المتغير الأول X والذي يأخذ القيم التالية: x_i : i =1,2,.........,n، وتخصص الأعمدة لبيانات المتغير الثاني Y الذي يأخذ القيم التالية: y_j : j=1,2,.........,m.

ج- <u>التكرارات الشرطية</u>(f_{ij}):

تحتوي على الصفة X و Y في نفس الوقت، حيث يمثل (i) السطر ، و (j) العمود.

<u>مثال2</u>:

في ثانوية ما ينقسم التلاميذ حسب متغيري الجنس والمستوى الدراسي كما يلي:

	أولى ثانوي	ثانية ثانوي	ثالثة ثانوي	المجموع
ذكور	295	190	128	613
إناث	269	203	174	646
المجموع	564	393	302	1259

- ✓ $\mathbf{f}_{11} = 295$: في هذه الثانوية 295 تلميذ (جنس ذكر) يدرس سنة أولى ثانوي.
- ✓ $\mathbf{f}_{12} = 190$: في هذه الثانوية 190 تلميذ (جنس ذكر) يدرس سنة ثانية ثانوي.
- ✓ $\mathbf{f}_{23} = 174$: في هذه الثانوية 174 تلميذة (جنس أنثى) تدرس سنة ثالثة ثانوي.

د- <u>التكرارات الهامشية:</u>

التكرارات الهامشية للمتغير X هي مجموع مكونات السطر(i)، وتحسب وفق الصيغة التالية:

$$\mathbf{f}_{i.} = \sum_{j=1}^{m} \mathbf{f}_{ij} \quad ; \ x_i : i = 1,2,\dots\dots\dots,n$$

التكرارات الهامشية للمتغير y هي مجموع مكونات العمود(j)، وتحسب وفق الصيغة التالية:

$$\mathbf{f}_{.j} = \sum_{i=1}^{n} \mathbf{f}_{ij} \quad ; \ y_j : j = 1,2,\dots\dots\dots,m$$

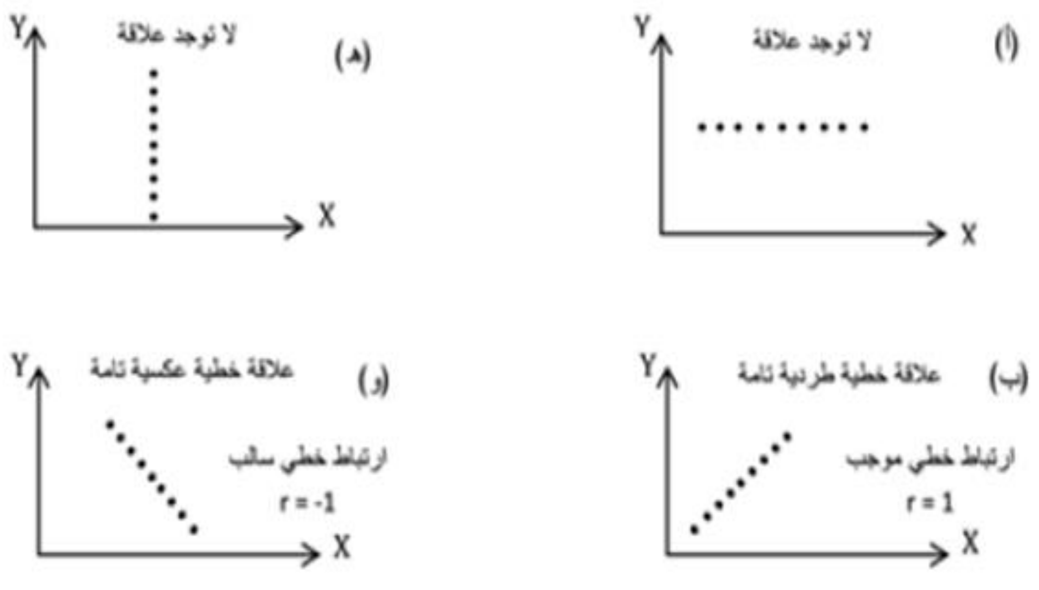
<u>من المثال السابق:</u>

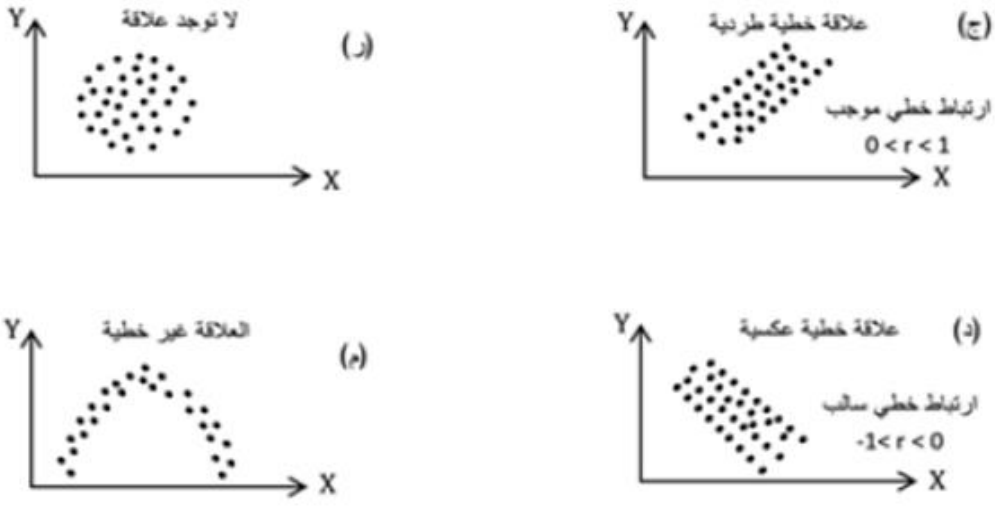
- ✓ $\mathbf{f}_{1.} = 613$: عدد التلاميذ الذكور مهما كان مستواهم التعليمي.
- ✓ $\mathbf{f}_{.2} = 393$: عدد التلاميذ سنة ثالثة ثانوي مهما كان جنسهم.
- ✓ $\mathbf{f}_{2.} = 646$: عدد التلاميذ الإناث مهما كان مستواهم التعليمي.

<u>ثانيا: الإرتباط بين متغيرين مستمرين</u>

1- <u>سحابة النقاط:</u>

لدراسة الارتباط أو العلاقة بين متغيرين(x) و (y) نضع قيم(x) على محور الفواصل و قيم (y) على محور التراتيب، ونمثل كل زوج من القيم بنقطة فنتحصل على شكل يبين كيفية انتشار هذه النقاط، و هو ما يسمى **سحابة النقاط** أو **لوحة الانتشار**، و التي من خلالها يتم التعرف بصفة مبدئية على طبيعة و قوة العلاقة بين المتغيرين. وقد تأخذ سحابة النقاط الأشكال التالية:





2- <u>معامل الإرتباط الخطي:</u>

يعتبر معامل الارتباط مؤشرا كميا على قوة العلاقة واتجاهها بين متغيرين، إذ يمكن أن يأخذ أي قيمة بين(-1، 1)، حيث تدل القيمة المحسوبة على قوة العلاقة وتدل الإشارة على اتجاهها موجبة أو سالبة. و لحساب هذه القيمة يمكن استخدام عدة مقاييس أهمها:

أ- <u>التباين المشترك:</u>

يستخدم لقياس الإرتباط الخطي الذي يمكن أن يحدث بين ثنائيات لمتغيرات إحصائية أو ثنائيات لمتغيرات عشوائية كمية. ويرمز له بالرمز:$\boldsymbol{cov(x,y)}$ ، ويمكن حسابه بطريقتين:

$$\boldsymbol{cov(x,y)}=\frac{\sum\sum \boldsymbol{f_{ij}}(\mathrm{X}_i-\bar{\mathrm{x}})(\mathrm{y}_j-\bar{\mathrm{y}})}{\boldsymbol{N}}$$

أو القانون المختصر: $$\boldsymbol{cov(x,y)}=\frac{\sum\sum \boldsymbol{f_{ij}}\mathrm{X}_i\mathrm{y}_j}{\boldsymbol{N}}-\bar{\mathrm{x}}\bar{\mathrm{y}}$$

- إذا كان $\boldsymbol{cov(x,y)=0}$: لا توجد علاقة بين المتغيرين.
- إذا كان $\boldsymbol{cov(x,y)<0}$: توجد علاقة عكسية بين المتغيرين.
- إذا كان $\boldsymbol{cov(x,y)>0}$: توجد علاقة طردية بين المتغيرين.

ومن أهم **خواص التباين المشترك** ما يلي:

الخاصية الأولى: $cov(x,y)=cov(y,x)$

الخاصية الثانية: التباين المشترك للمتغير نفسه هو التباين نفسه، $cov(x,x)=\ \boldsymbol{v_x}$

الخاصية الثالثة: $cov(ax,by)=ab\ cov(y,x)$

الخاصية الرابعة: $cov(a+x,b+y)=cov(y,x)$

والفرق بين التباين المشترك والتباين يكمن فيما يلي:

- التباين يحسب بمتغير واحد فقط، أما التباين المشترك فيحسب بمتغيرين.
- قيمة التباين دائما موجبة، أما قيمة التباين المشترك قد تكون موجبة أو سالبة.

- التباين يفسر التشابه أو التشتت بين القيم، أما التباين المشترك فيفسر درجة الارتباط بين القيم.

ب- معامل الإرتباط بيرسون:

يعتبر معامل الإرتباط بيرسون أفضل طريقة لقياس العلاقة الإحصائية أو الارتباط بين متغيرين مستمرين، فهو يعطي معلومات حول حجم الارتباط واتجاه العلاقة. يرمز له بالرمز $(\mathbf{r}_{xy})$ ويحسب بالعلاقة التالية:

$$\mathbf{r}_{xy}=\frac{cov(x,y)}{S_x S_y}$$

حيث: cov (x,y) هو التباين المشترك بين المتغيرين

$S_x S_y$ الانحراف المعياري لـ (x) و (y) على التوالي

تتراوح قيمة معامل الارتباط بين: $-1< _{xy}< 1$

- إذا كان $\mathbf{r}_{xy} = -1$:: يوجد ارتباط خطي سالب تام بين المتغيرين.
- إذا كان $\mathbf{r}_{xy} = 1$: يوجد ارتباط خطي موجب تام بين المتغيرين.
- إذا كان $-1< \mathbf{r}_{xy}< 0$: يوجد ارتباط خطي سالب غير تام بين المتغيرين.
- إذا كان $0 < \mathbf{r}_{xy}< 1$: يوجد ارتباط خطي موجب غير تام بين المتغيرين.
- إذا كان $\mathbf{r}_{xy} = 0$:: لا يوجد ارتباط خطي بين المتغيرين.

ج- معامل التحديد:

هو مربع معامل الارتباط ويقيس نسبة تفسير المتغير المستقل لتغيرات المتغير التابع، ويرمز له بالرمز ^{2} .

ويحسب كما يلي:

$$\mathrm{R}^2=\frac{(cov(x,y))^2}{{s_x}^2{s_y}^2}=\frac{(cov(x,y))^2}{v_x v_y}$$

$R^2 < 70\,\%$	علاقة ضعيفة
$70\,\% \leq R^2 < 80\,\%$	علاقة معتبرة
$80\,\% \leq R^2 < 90\,\%$	علاقة قوية
$90\,\% \leq R^2 < 1000\,\%$	علاقة قوية جدا

تمرين:

يمثل التوزيع التالي ساعات عمل لمجموعة من العمال والمتعلقة بحجم الإنتاج:

x_i	[2 4 [	[4 8 [	[8 10 [	[10 14 [	$\sum$
y_j	3	5	6	8	$\sum$
f_{ij}	4	2	6	8	20

المطلوب:

1- ضع البيانات في جدول الاقتران؟

2- أحسب وفسر، f_{11} f_{31}، f_{24}، $\mathrm{f}_{.3}$، $\mathrm{f}_{2.}$، $r\mathrm{f}_{11}$؟

3- أحسب التباين المشترك وفسره؟

4- أحسب معامل الإرتباط بيرسون وفسره؟

5- أحسب معامل التحديد وفسره؟

6- أرسم سحابة النقاط وحدد نوع العلاقة بين المتغيرين بيانيا؟

الحل:

1- **وضع البيانات في جدول الاقتران:**

x_i \ y_j	3	5	6	8	$\sum f_i$	m_i	$f_i m_i$	$f_{ij} m_i y_j$	$f_i (m_i)^2$
[2 4 [	4				4	3	12	36	36
[4 8 [		2			2	6	12	60	72
[8 10 [			6		6	9	54	324	486
[10 14 [				8	8	12	96	768	1152
$\sum f_J$	4	2	6	8	20	/	174	1188	1746
$f_j y_j$	12	10	36	64	122	/	/	/	/
$f_j (y_j)^2$	36	50	216	512	814	/	/	/	/

2- **التفسير:**

✓ $\mathbf{f_{11} = 4}$: يوجد 4 عمال يعملون ما بين 2 و 4 ساعة وحجم إنتاجهم 3 طن.

✓ $\mathbf{f_{31} = 0}$: لا يوجد عمال يعملون ما بين 8 و 10 ساعة وحجم إنتاجهم 3 طن.

✓ $\mathbf{f_{24} = 0}$: لا يوجد عمال يعملون ما بين 4 و 8 ساعة وحجم إنتاجهم 8 طن.

✓ $\mathbf{f_{.3} = 0}$: يوجد 6 عمال حجم إنتاجهم 6 طن.

✓ $\mathbf{f_{2.} = 0}$: يوجد 2 عمال يعملون ما بين 4 و 8 ساعة.

✓ $\mathbf{rf_{11} = \frac{4}{20} = 0.2}$: 20 % من العمال يعملون ما بين 2 و 4 ساعة وحجم إنتاجهم 3 طن.

✓ $\mathbf{rf_{21} = \frac{0}{20} = 0}$: 20 % من العمال يعملون ما بين 4 و 8 ساعة وحجم إنتاجهم 3 طن.

3- **حساب التباين المشترك وتفسيره:**

$$cov(x,y) = \frac{\sum\sum f_{ij} X_i y_j}{N} - \bar{X}\bar{y}$$

- **حساب الوسط الحسابي للمتغيرين:**

$$\bar{X} = \frac{\sum_{i=1}^{4} f_i m_i}{N} = \frac{174}{20} = 8.7 \quad ; \quad \bar{y} = \frac{\sum_{j=1}^{4} f_j y_j}{N} = \frac{122}{20} = 6.1$$

$$cov(x,y) = \frac{\sum\sum f_{ij} X_i y_j}{N} - \bar{X}\bar{y} = \frac{1188}{20} - 8.7(6.1) \Rightarrow \boxed{cov(x,y) = 6.33}$$

التفسير:

بما أن $cov(x,y) = 6.33 > 0$: توجد علاقة طردية بين ساعات العمل وحجم الإنتاج.

4- **حساب معامل الإرتباط بيرسون وتفسيره:**

$$r_{xy} = \frac{cov(x,y)}{S_x S_y}$$

● **حساب الإنحراف المعياري للمتغيرين:**

$$S_X = \sqrt{v_x} = \sqrt{\frac{\sum f_i (m_i)^2}{N} - \bar{x}^2} = \sqrt{\frac{1746}{20} - (8.7)^2} \Rightarrow S_X = 3.4$$

$$S_Y = \sqrt{v_y} = \sqrt{\frac{\sum f_i (y_i)^2}{N} - \bar{y}^2} = \sqrt{\frac{814}{20} - (6.1)^2} \Rightarrow S_Y = 1.87$$

$$r_{xy} = \frac{cov(x,y)}{S_x S_y} = \frac{6.33}{3.4(1.87)} \Rightarrow \boxed{r_{xy} = 0.9955}$$

التفسير:

بما أن $0 < r_{xy} < 1$: يوجد إرتباط خطي موجب قريب من التام بين ساعات العمل وحجم الإنتاج.

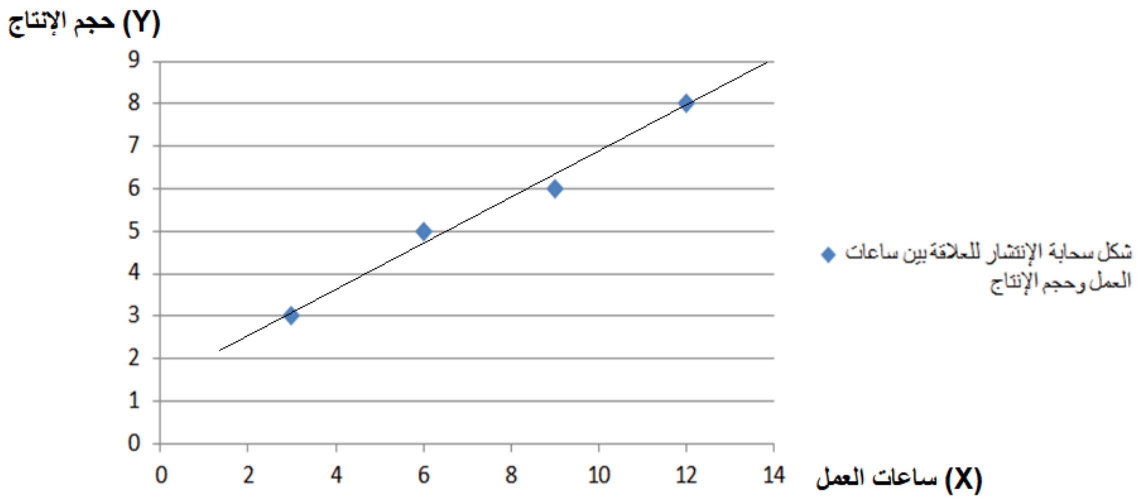
5- **حساب معامل التحديد وتفسيره:**

$$R^2 = (r_{xy})^2 .100 = (0.9955)^2 .100 \Rightarrow \boxed{R^2 = 99.12\%}$$

التفسير:

- بما أن: $90\ \% \leq R^2 < 100\ \%$، يوجد علاقة قوية جدا بين ساعات العمل وحجم الإنتاج.
- 99.12 % من التغيرات في حجم الإنتاج يفسرها التغير في ساعات العمل، و0.88 % تفسرها عوامل أخرى.

6- **رسم سحابة النقاط وتحدد نوع العلاقة بين المتغيرين بيانيا:**



3- الإنحدار الخطي البسيط:

بعد التعرف على طبيعة وقوة العلاقة بين متغيرين أحدهما متغير مستقل X والآخر متغير تابع Y باستخدام معاملات الإرتباط، يتجه الباحثون لتحليل الإنحدار الخطي البسيط والذي يهدف إلى تقييم العلاقة الخطية بين المتغيرين و دراسة أثر تغير المتغير الكمي المستقل على المتغير الكمي التابع.

مثال:

- دراسة أثر الإنتاج على التكلفة.
- د راسة أثر كمية السماد على إنتاجية المزرعة.
- دراسة أثر الدخل على الاستهلاك.

ملاحظة:

إذا كان هناك أكثر من متغير مستقل واحد نحصل على ما يعرف بالإنحدار الخطي المتعدد.

أ- معادلة الإنحدار:

إن الإنحدار الخطي البسيط التام والذي يحتوي على متغير تابع ومتغير مستقل واحد فقط يعتبر الأسهل للتقدير والتحليل والتنبؤ، وتكون جميع نقاط شكل الانتشار على استقامة واحدة، سواء كانت في الاتجاه الموجب أو في الاتجاه السالب. وتكتب معادلة الإنحدار على الشكل التالي:

$$y_i = a\ x_i + b$$

حيث:

y_i : المتغير التابع

x_i : المتغير المستقل

a : ميل خط معادلة الإنحدار ، أي ظل الزاوية α حيث: $\alpha = \frac{\Delta y}{\Delta x}$

b: ثابت المعادلة ويدل على قيمة الدالة لما x =0 ، أي نقطة تقاطع خط الإنحدار مع محور التراتيب.

ملاحظة:

إذا كانت نقاط شكل الإنتشار ليست على استقامة تامة، ولكنها تأخذ إتجاها يمكن تقريبه من معادلة خط مستقيم، يكون **الإنحدار الخطي البسيط غير تام**. وتصبح معادلة الإنحدار كما يلي:

$$y_i = a\ x_i + b + e_i$$

حيث: e_i : تمثل قيمة البعد بين النقاط الحقيقية ومعادلة خط المستقيم، وبما أنه يستحيل إيجاد هذه المعادلة لذلك يتم تقريبها إلى المعادلة الخطية التالية:

$$\widehat{y_i} = \hat{a}\ x_i + \hat{b}$$

حيث:

$\widehat{y_i}$: قيمة تقديرية لـ y_i

$\hat{a}$: قيمة تقديرية لـ a

$\hat{b}$: قيمة تقديرية لـ b

ب- **<u>إيجاد ثوابت الإنحدار$(\hat{a} \text{ ، } \hat{b})$ حسابيا وبيانيا:</u>**

- **<u>حسابيا:</u>**

يتم إيجاد ثوابت الإنحدار بما يسمى طريقة المربعات الصغرى والتي تعطي المعادلتين التقديريتين كما يلي:

$$\hat{a} = \frac{cov(x,y)}{(s_x)^2} = \frac{\sum(X_i - \overline{x})(y_j - \overline{y})}{\sum(X_i - \overline{x})^2}$$

$$\hat{b} = \overline{y} - \hat{a}\,\overline{x}$$

- **<u>بيانيا(في حالة الانحدار الخطي البسيط التام):</u>**

- رسم منحنى بياني يعبر عن الثنائيات من قيم المتغير المستقلX والتي تكون على المحور الافقي، وقيم المتغير التابع Y والتي تكون على المحور العمودي.
- نقوم بمد المستقيم حتى يقطع محور التراتيب في نقطة والتي تمثل **قيمةb والتي يكون فيها X=0**.
- نأخذ أي ثنائيتين من$(x ;y)$ ونقوم بالاسقاط على الخط المستقيم، فيتشكل لنا مثلث قائم **EFG**
- نحسب ميل الانحدار و الذي يمثل **قيمة a** وهو عبارة عن **ظل الزاوية α**$(\alpha = \frac{\Delta y}{\Delta x})$

<u>مثال:</u>

لدراسة العلاقة بين الاستهلاك والدخل لمجموعة من العائلات، تم اختيار عينة مكونة من 10 عائلات، وكانت النتائج كما يلي:

x_i	8	12	16	23	25	26	28	30	40	42
y_i	10	15	20	25	30	35	40	45	50	55

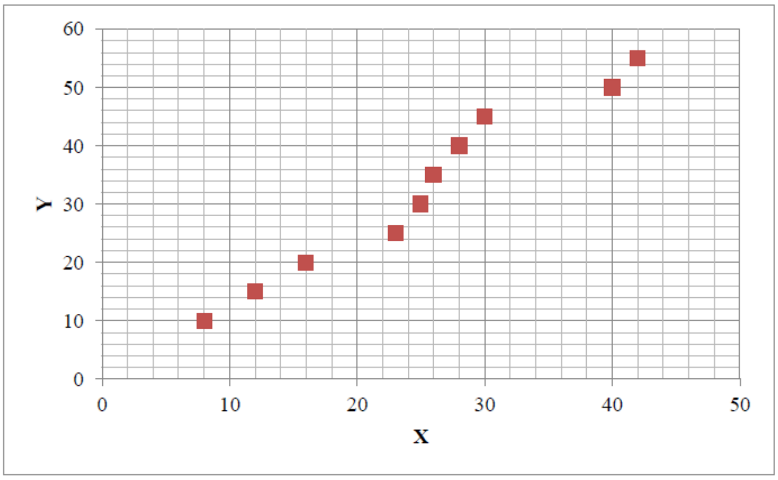
1- مثل سحابة النقاط بيانيا؟

2- هل يوجد ارتباط بين الإستهلاك والدخل؟

3- اوجد معادلة الإنحدار الخطي البسيط للدخل على الاستهلاك؟

<u>الحل:</u>

1- **<u>تمثيل سحابة النقاط بيانيا:</u>**



نلاحظ من خلال شكل انتشار النقاط أن هناك ارتباط قوي بين الدخل والاستهلاك، بحيث كلما زاد الدخل زاد الاستهلاك في نفس الاتجاه.

2- **دراسة العلاقة بين الدخل والاستهلاك (الارتباط):**

x_i (الدخل)	y_i (الاستهلاك)	x_iy_i	$(x_i)^2$	$(y_i)^2$
8	10	80	64	100
12	15	180	144	225
16	20	320	256	400
23	25	575	529	625
25	30	750	625	900
26	35	910	676	1225
28	40	1120	784	1600
30	45	1350	900	2025
40	50	2000	1600	2500
42	55	2310	1764	3025
250	325	9595	7342	12625

أ-**حساب التباين المشترك:**

$$cov(x,y)=\frac{\sum x_iy_i}{N}-\bar{x}\bar{y}$$

- **حساب الوسط الحسابي للمتغيرين:**

$$\bar{x}=\frac{\sum_{i=1}^{10}x_i}{N}=\frac{250}{10}=25 \quad ; \quad \bar{y}=\frac{\sum_{i=1}^{10}y_i}{N}=\frac{325}{10}=32.5$$

$$cov(x,y)=\frac{\sum x_iy_j}{N}-\bar{x}\bar{y}=\frac{9595}{10}-25(32.5) \Rightarrow \boxed{cov(x,y)=147}$$

التفسير:

بما أن $cov(x,y)=147>0$: توجد علاقة طردية بين الدخل والإستهلاك.

ب-**حساب معامل الإرتباط بيرسون وتفسيره:**

$$r_{xy}=\frac{cov(x,y)}{S_xS_y}$$

- **حساب الإنحراف المعياري للمتغيرين:**

$$S_x=\sqrt{v_x}=\sqrt{\frac{(m_i)^2}{N}-\bar{x}^2}=\sqrt{\frac{7342}{10}-(25)^2} \Rightarrow S_x=10.45$$

$$S_y=\sqrt{v_y}=\sqrt{\frac{(y_i)^2}{N}-\bar{y}^2}=\sqrt{\frac{12625}{10}-(32.5)^2} \Rightarrow S_y=14.36$$

- **معامل الإرتباط بين المتغيرين هو:**

$$r_{xy}=\frac{cov(x,y)}{S_xS_y}=\frac{147}{10.45(14.36)} \Rightarrow \boxed{r_{xy}=0.9796}$$

التفسير:

بما أن $0 < r_{xy} < 1$: يوجد إرتباط خطي موجب قريب من التام بين **الإستهلاك والدخل**.

ج- **حساب معامل التحديد:**

$$R^2 = (r_{xy})^2 \cdot 100 = (0.9796)^2 \cdot 100 \Rightarrow \boxed{R^2 = 95.96\,\%}$$

التفسير:

- بما أن: $100\,\% > R^2 \geq 90\,\%$، يوجد علاقة قوية جدا بين الدخل والإستهلاك.
- 95.96 % من التغيرات في الإستهلاك يفسرها التغير في الدخل، و4.04 % تفسرها عوامل أخرى.

3- **إيجاد معادلة الإنحدار الخطي البسيط للدخل على الاستهلاك:**

$$\hat{y}_i = \hat{a}\, x_i + \hat{b}$$

$$\hat{a} = \frac{cov(x,y)}{(s_x)^2} = \frac{147}{(10.45)^2} = 1.346$$

$$\hat{b} = \bar{y} - \hat{a}\,\bar{x} = 32.5 - 1.346\,(25) = -1.15$$

ومنه معادلة الإنحدار هي: $\hat{y}_i = 1.346\, x_i - 1.15$

الحل بطريقة 2:

x_i (الدخل)	y_i (الاستهلاك)	$x_i y_i$	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
8	10	80	-17	-22.5	382.5	289	506.25
12	15	180	-13	-17.5	227.5	169	306.25
16	20	320	-9	-12.5	112.5	81	156.25
23	25	575	-2	-7.5	15	4	56.25
25	30	750	0	-2.5	0	0	6.25
26	35	910	1	2.5	2.5	1	6.25
28	40	1120	3	7.5	22.5	9	56.25
30	45	1350	5	12.5	62.5	25	156.25
40	50	2000	15	17.5	262.5	225	306.25
42	55	2310	17	22.5	382.5	289	506.25
250	325	9595	/	/	1470	1092	2062.5

أ-حساب التباين المشترك:

$$cov(x,y) = \frac{\sum X_i y_i}{N} - \bar{x}\bar{y}$$

- **حساب الوسط الحسابي للمتغيرين:**

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{N} = \frac{250}{10} = 25 \quad ; \quad \bar{y} = \frac{\sum_{i=1}^{10} y_i}{N} = \frac{325}{10} = 32.5$$

$$cov(x,y) = \frac{\sum (X_i - \bar{x})(y_j - \bar{y})}{N} = \frac{1470}{10} \Rightarrow \boxed{cov(x,y) = 147}$$

التفسير:

بما أن $cov(x,y) = 147 > 0$: توجد علاقة طردية بين الدخل والإستهلاك.

- **حساب الإنحراف المعياري للمتغيرين:**

$$S_x = \sqrt{v_x} = \sqrt{\frac{\sum(X_i - \bar{x})^2}{N}} = \sqrt{\frac{1092}{10}} \Rightarrow S_x = 10.45$$

$$S_y = \sqrt{v_y} = \sqrt{\frac{\sum(y_i - \bar{y})^2}{N}} = \sqrt{\frac{2062.5}{10}} \Rightarrow S_y = 14.36$$

- **معامل الإرتباط بين المتغيرين هو:**

$$r_{xy} = \frac{cov(x,y)}{s_x s_y} = \frac{147}{10.45(14.36)} \Rightarrow \boxed{r_{xy} = 0.9796}$$

التفسير:

بما أن $0 < r_{xy} < 1$: يوجد إرتباط خطي موجب قريب من التام بين **الإستهلاك والدخل**.

- **حساب معامل التحديد:**

$$R^2 = (r_{xy})^2 .100 = (0.9796)^2 .100 \Rightarrow \boxed{R^2 = 95.96\,\%}$$

التفسير:

- بما أن: $\% 100 > R^2 \geq \% 90$، يوجد علاقة قوية جدا بين الدخل والإستهلاك.
- 95.96 % من التغيرات في الإستهلاك يفسرها التغير في الدخل، و4.04 % تفسرها عوامل أخرى.

4- **إيجاد معادلة الإنحدار الخطي البسيط للدخل على الاستهلاك:**

$$\hat{y}_i = \hat{a}\ x_i + \hat{b}$$

$$\hat{a} = \frac{cov(x,y)}{(s_x)^2} = \frac{147}{(10.45)^2} = 1.346$$

$$\hat{b} = \bar{y} - \hat{a}\,\bar{x} = 32.5 - 1.346\,(25) = -1.15$$

ومنه معادلة الإنحدار هي: $\hat{y}_i = 1.346\ x_i - 1.15$