# أشكال التوزيعات التكرارية

بعد معرفة درجة انتشار التكرارات حول قيمة وسطى، يتم تحديد شكل المنحنى الممثل للتوزيع للعينة المدروسة، و قد يكون هذا المنحنى متماثل أو ملتوي.

و هناك عدة مقاييس للالتواء منها ما يمكن حسابه من العلاقة بين المقاييس الموضعية أو عن طريق العزوم الاحصائية.

- يكون التوزيع متماثل إذا تحققت العلاقة $\bar{X} = M_e = M_o$ ، أي تكون نسبة 50 % من القيم على يمين و على يسار هذه المقاييس.
- عنما تتحقق المتراجحة $\bar{X} > M_e > M_o$ ، يعني التواء موجب.
- عنما تتحقق المتراجحة $\bar{X} < M_e < M_o$ ، يعني التواء سالب.

لإيجاد شكل التوزيع الاحصائي، يمكن استخدام معامل الاختلاف الربيعي كما يلي:

$$C_{yc} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

و يسمى **بمعامل يول و كاندال** $C_{yc}$، فإذا كان:

- $C_{yc} = 0$ ، يكون التوزيع متناظر ،
- $C_{yc} > 0$ ، نقول أن الالتواء موجب،
- $C_{yc} < 0$ ، نقول أن الالتواء سالب.

# العزوم
# Moments

استخدام العلماء العزوم للاستدلال على الالتواء، حيث و نميز بين نوعين العزوم البسيطة و العزوم المركزية.

**1) العزوم البسيطة:** يعرف العزم البسيط أي العزم حول الصفر على أنه مجموعة القيم المرفوعة إلى القوى r على عددها، و يعطى بالعلاقة التالية:

- حالة البيانات غير المبوبة:

$$\mathfrak{m}_r = \frac{\sum_{i=1}^{n} X_i^r}{n}$$

- حالة التوزيعات التكرارية:

$$\mathfrak{m}_r = \frac{\sum_{i=1}^{n} n_i X_i^r}{\sum_{i=1}^{n} n_i}$$

لدينا الحالات الخاصة التالية:

- أن العزم البسيط من الدرجة الأولى هو المتوسط الحسابي.
- أن العزم من الدرجة الثانية هو مربع المتوسط الحسابي.

$$r= 1 \Longrightarrow \mathfrak{m}_1 = \frac{\sum_{i=1}^{n} X_i^1}{n} = \bar{X}$$

$$r= 2 \Longrightarrow \mathfrak{m}_2 = \frac{\sum_{i=1}^{n} X_i^2}{n} = Q^2$$

**2) العزوم المركزية:** يعرف العزم المركزي أي العزم حول المتوسط الحسابي على أنه مجموع انحرافات القيم عن قيمتها المتوسطة (أي المركزية) مرفوعة إلى القوة K على عددها، حيث يعطى بالعلاقة التالية:

❀ حالة البيانات غير المبوبة:

$$\boldsymbol{u_r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^r}{n}}$$

❀ حالة التوزيعات التكرارية:

$$\boldsymbol{u_r = \frac{\sum_{i=1}^{n} n_i(X_i - \overline{X})^r}{\sum_{i=1}^{n} n_i}}$$

كما أنه لدينا الحالات الخاصة التالية:

- أن العزم الأول حول الوسط الحسابي يساوي الصفر.
- أن العزم الثاني هو التباين بطريقة الانحرافات.

$$\boldsymbol{u_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^1}{n} = \frac{0}{n} = 0}$$

$$\boldsymbol{u_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} = V(X)}$$

¥ **مثال (9):** لدينا السلسلة الاحصائية التالية:

**2 – 3 – 10 – 1 – 6**

**المطلوب:**

**1)** أوجد العزوم الأولى حول الصفر (البسيط).

$$m_1 = \frac{\sum X_i}{n} = \frac{22}{5} = 4.4$$

$$m_2 = \frac{\sum X^2}{n} = \frac{150}{5} = 30$$

$$m_3 = \frac{\sum X^3}{n} = \frac{1252}{5} = 250.4$$

$$m_4 = \frac{\sum X^4}{n} = \frac{11394}{5} = 2278.8$$

1) أوجد العزوم الأولى حول الوسط (المركزية).

$$u_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^1}{n} = \frac{0}{n} = 0$$

$$u_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = V(X) = \frac{53.2}{5} = 10.64$$

$$u_3 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n} = \frac{123.84}{5} = 24.76$$

$$u_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n} = \frac{1160.656}{5} = 232.13$$

¥ **مثال (10):** يبين الجدول التالي أوزان 100 طفل، المطلوب: إيجاد كل من:

$\boldsymbol{u_1, u_2, u_3, u_4}$

**جدول (10): التوزيع التكراري لأوزان الأطفال**

| $n_i(X_i-\bar{X})^3$ | $(X_i-\bar{X})^3$ | $n_i(X_i-\bar{X})^2$ | $(X_i-\bar{X})^2$ | $(X_i-\bar{X})^1$ | $n_iX_i$ | $n_i$ | $X_i$ |
|---|---|---|---|---|---|---|---|
| −26.62 | −1.331 | 24.2 | 1.21 | −1.1 | *120* | *20* | 6 |
| −2.16 | −0.216 | 3.6 | 0.36 | −0.6 | *65* | *10* | 6.5 |
| −0.02 | −0.001 | 0.2 | 0.01 | −0.1 | *140* | *20* | 7 |
| 1.92 | 0.064 | 4.8 | 0.16 | 0.4 | *225* | *30* | 7.5 |
| 14.58 | 0.729 | 16.2 | 0.81 | 0.9 | *160* | *20* | 8 |
| **−12.3** | ---- | **49** | ---- | ---- | ***710*** | ***100*** | ***المجموع*** |

$$u_1 = \frac{\sum_{i=1}^{n} n_i(X_i - \bar{X})^1}{\sum_{i=1}^{n} n_i} = 0$$

$$u_2 = \frac{\sum_{i=1}^{n} n_i(X_i - \bar{X})^2}{\sum_{i=1}^{n} n_i} = \frac{49}{100} = 0.49$$

$$u_3 = \frac{\sum_{i=1}^{n} n_i(X_i - \bar{X})^3}{\sum_{i=1}^{n} n_i} = 0.123$$

$$u_r = \frac{\sum_{i=1}^{n} n_i(X_i - \bar{X})^r}{\sum_{i=1}^{n} n_i} = 0.445$$

## الالتواء

## Skewness

يعرف الالتواء بأنه درجة تماثل أو بعد التماثل لتوزيع ما، و يستخدم الالتواء لمعرفة نوع التوزيع، و يعطى معامل الالتواء العزومي بالعلاقة التالية:

$$\text{معامل الالتواء العزومي} = \frac{\text{العزم الثالث}}{\text{مكعب الانحراف المعياري}}$$

و بالعلاقة الاحصائية يكتب:

$$\boldsymbol{F_1 = \frac{u_3}{\delta_x^3}}$$

فإذا كان مقياس الالتواء:

- موجباً، نقول أن التوزيع ملتو نحو اليمين (موجب الالتواء).
- سالباً، نقول أن التوزيع ملتو نحو اليسار (سالب الالتواء).
- يساوي الصفر ، فإن التوزيع متماثل.

¥ **مثال (11):** استناداً إلى المثال السابق، أوجد معامل الالتواء العزومي لفيشر؟ أو أدرس الالتواء لهذا التوزيع باستخدام معامل فيشر؟

لدينا:

$$V(X) = u_2 = 0.49 \Rightarrow \delta_x = \sqrt{0.49} = 0.7$$
$$u_3 = -0.123$$

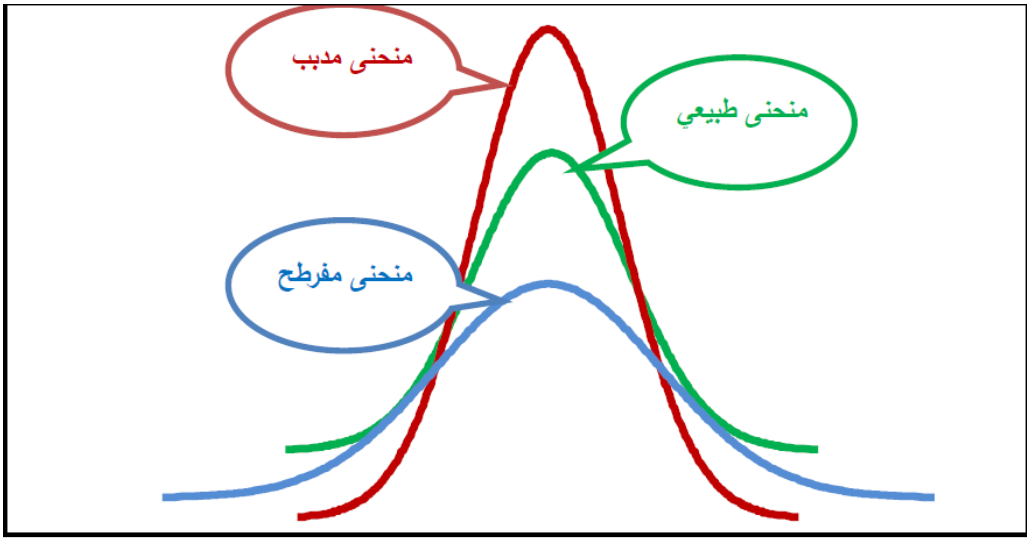
$$\boldsymbol{F_1} = \boldsymbol{\frac{u_3}{\delta_x^3}} = \frac{-0.123}{(0.7)^3} = 0.358$$

نلاحظ أن $F_1 < 0 \Leftarrow$ و يعني أن هذا التوزيع سالب الالتواء (ملتو نحو اليسار)

التفلطح
Kurtosis

التفلطح هو درجة تدبب قمة التوزيع قياساً إلى التوزيع الطبيعي، فالتوزيع ذو القمة العالية نسبياً يسمى منحنى مدبباً (متطاول)، و التوزيع ذي القمة المسطحة يسمى مفلطحاً، و التوزيع الطبيعي حيث قمته ليست مدببة (متطاولة) و لا مفلطحة.

**الكل رقم (4-1): أشكال التوزيعات الاحصائية**




و يعرف معامل التفرطح العزومي لمجموعة من البيانات على انه النسبة بين العزم الرابع حول الوسط و مربع التباين، و يعطي بالعلاقة كما يلي:

$$\text{معامل التفرطح العزومي} = \frac{\text{العزم الرابع حول الوسط}}{\text{( مربع العزم الثاني حول الوسط) مكعب أو مربع التباين}}$$

و هو معامل بيرسون الثاني:

$$P_2 = \frac{u_4}{\delta_x^4} \Rightarrow P_2 = \frac{u_4}{(u_2{}^2)}$$

- $P_2 > 3 \Longleftarrow$ نقول أن التوزيع مدبب.
- $P_2 < 3 \Longleftarrow$ نقول أن التوزيع مفرطح.
- $P_2 = 3 \Longleftarrow$ نقول أن التوزيع طبيعي ( متماثل).

¥ **مثال (12):** أحسب معامل التفرطح العزومي، و ما هو نوع التوزيع؟

$$u_4 = 0.445$$

$$\boldsymbol{P_2 = \frac{u_4}{\delta_x^4} \Rightarrow P_2 = \frac{u_4}{(u_2{}^2)}}$$

$$\boldsymbol{P_2 = \frac{0.445}{0.7^4} = \frac{0.445}{0.49^2} = 1.853}$$

نلاحظ أن $P_2 < 3 \Longleftarrow$ نقول أن التوزيع مفرطح

# الفصل الخامس

## الارتباط و الانحدار الخطي البسيط

في الفصول السابقة تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، و التشتت، و مقاييس الالتواء و التفرطح، و غيرها من المقاييس الأخرى و التي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، و ننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، و يتناول هذا الفصل[1] دراسة و تحليل العلاقة بين متغيرين، و ذلك باستخدام بعض طرق التحليل الاحصائي مثل تحليل الارتباط، الانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الانحدار، و من الأمثلة ذلك:

أ. الانفاق، و الدخل العائلي.

**ب.** سعر السلعة، و الكمية المطلوبة منها.

**ج.** الفترة الزمنية لتخزين الخبر، و عمق طراوة الخبز.

**د.** تقديرات الطلاب في مقر الاحصاء، و تقديراتهم في مقرر الرياضيات.

**هـ.** كميات السماد المستخدمة، و كمية الانتاج من محصول معين ثم تسميده بهذا النوع من السماد.

**و.** عدد مرات ممارسة نوع معين من الرياضة البدنية، و مستوى الكلسترول في الدم.

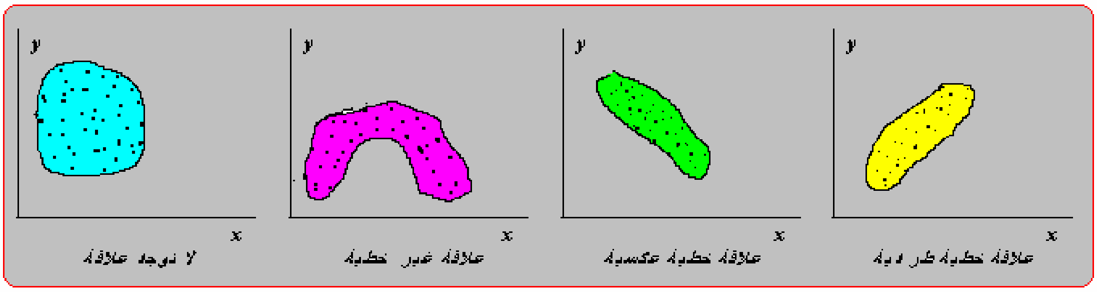
**ز.** وزن الجسم، و ضغط الدم.

---

[1] خليل، شرف الدين. بدون تاريخ. الاحصاء الوصفي، https://www.physics-pdf.com/2018/09/Book-of-descriptive-statistics-pdf.html، 2022/07/21

و الأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين (X , Y)، و تم جمع بيانات عن أزواج قيم هذين المتغيرين، و تم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالاً مختلفة على النحو التالي:

**الشكل رقم (5-1): شكل الانتشار لبيان نوع العلاقة بين X, Y**



# الارتباط الخطي البسيط
# Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع و قوة العلاقة بين متغيرين، يستخدم تحليل الارتباط، و أما إذا كان الغرض هو دراسة و تحليل أثر أحد المتغيرين على الآخر، يستخدم تحليل الانحدار، و في هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي، و سوف يجري حسابه في حالة البيانات الكمية و البيانات الوصفية المقاسة بمعيار ترتيبي.

**1) <u>الغرض من تحليل الارتباط الخطي البسيط</u>**

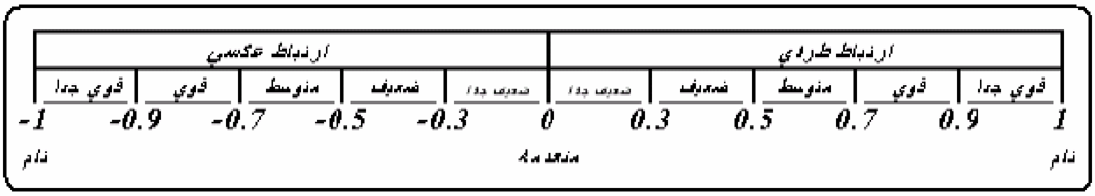
الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع و قوة العلاقة بين متغيرين، و يرمز له في حالة المجتمع بالرمز $\rho$، و في حالة العينة بالرمز $r$، و حيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف تهتم بحساب معامل الارتباط $r$، كتقدير لمعامل الارتباط في المجتمع، و من التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

❀ **نوع العلاقة:** و تأخذ ثلاثة أنواع حسب إشارة معامل الارتباط كما يلي:

- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحب انخفاض في المتغير الثاني، و العكس.
- إذا كان معامل الارتباط قيمته موجبة $(r > 0)$ توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة المتغير الثاني، و العكس.
- إذا كان معامل الارتباط قيمته صفر $(r = 0)$ دل ذلك على انعدام العلاقة بين المتغيرين.

❀ **قوة العلاقة:** و يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن $(\pm 1)$، حيث أن قيمة معامل الارتباط تقع في المدى $(-1 < \tau < +1)$ ، و قد صنف بعض الاحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

**الشكل رقم (5-2): درجات قوة معامل الارتباط**




## 2) <u>معامل الارتباط الخطي البسيط لبيرسون Pearson</u>

في حالة جمع بيانات عن متغيرين كميين (X, Y) و يمكن قياس الارتباط بينهما، باستخدام طريقة بيرسون، و من الأمثلة على ذلك: قياس العلاقة بين الوزن و الطول، و العلاقة بين الانتاج و التكلفة، و العلاقة بين الانفاق الاستهلاكي و الدخل، و العلاقة بين الدرجة التي حصل عليها الطالب و عدد ساعات الاستذكار، و هكذا الأمثلة على ذلك كثيرة.

و لحساب معامل الارتباط في العينة، تستخدم صيغة بيرسون في المعادلة (1) التالية:

$$r = \frac{S_{xy}}{S_x\, S_y} = \frac{\dfrac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y-\bar{y})^2}{(n-1)}}}$$

حيث أن:

$S_{xy} = \sum (x-\bar{x})(y-\bar{y})/(n-1)$ : هو التغاير بين $(x \ . \ y)$،

$S_x = \sqrt{\sum (x-\bar{x})^2/(n-1)}$ : هو الانحراف المعياري لقيم $(x)$،

$S_y = \sqrt{\sum (y-\bar{y})^2/(n-1)}$ : هو الانحراف المعياري لقيم $(y)$.

و يمكن اختصار الصيغة السابقة في المعادلة (2)، على النحو التالي:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}}$$

¥ **مثال (1):** فيما يلي المساحة المنزرعة بالأعلاف الخضراء بالألف هكتا، و اجمالي إنتاج اللحوم بألف طن، خلال الفترة من 1995 حتى 2002.

| السنة | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|
| المساحة | 305 | 313 | 297 | 289 | 233 | 214 | 240 | 217 |
| الكمية | 592 | 603 | 662 | 607 | 635 | 699 | 719 | 747 |

**المطلوب:** حساب معامل الارتباط بين المساحة و الكمية، و ما هو مدلوله؟

بفرض أن (X) هي المساحة المنزرعة، (y) هي الكمية، و لحساب معامل الارتباط بين (y, x) يتم تطبيق المعادلة السابقة، و ذلك على النحو التالي:

1. حساب الوسط الحسابي لكل من المساحة، و الكمية $(\bar{y},\bar{x})$:

$$\bar{x}=\frac{\sum x}{n}=\frac{2108}{8}=263.5 \ , \ \ \bar{y}=\frac{\sum y}{n}=\frac{5264}{8}=658$$

2. حساب المجاميع:

**<u>الجدول رقم (5−1): المساحة المنزرعة بالأعلاف الخضراء بالألف هكتا، و اجمالي إنتاج اللحوم بألف طن، خلال الفترة من 1995 حتى 2002.</u>**

| $x$ | $y$ | $x-\bar{x}$ | $(x-\bar{x})^2$ | $y-\bar{y}$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 305 | 592 | 41.5 | 1722.25 | -66 | 4356 | -2739 |
| 313 | 603 | 49.5 | 2450.25 | -55 | 3025 | -2722.5 |
| 297 | 662 | 33.5 | 1122.25 | 4 | 16 | 134 |
| 289 | 607 | 25.5 | 650.25 | -51 | 2601 | -1300.5 |
| 233 | 635 | -30.5 | 930.25 | -23 | 529 | 701.5 |
| 214 | 699 | -49.5 | 2450.25 | 41 | 1681 | -2029.5 |
| 240 | 719 | -23.5 | 552.25 | 61 | 3721 | -1433.5 |
| 217 | 747 | -46.5 | 2162.25 | 89 | 7921 | -4138.5 |
| 2108 | 5264 | 0 | 12040 | 0 | 23850 | -13528 |

$$\sum(x-\bar{x})^2=12040 \ , \ \sum(y-\bar{y})^2=23850 \ ,$$
$$\sum(x-\bar{x})(y-\bar{y})=-13528$$

إذن معامل الارتباط قيمته، هي:

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2}\sqrt{\sum(y-\overline{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$

$$= \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

يوجد ارتباط عكسي قوي بين المساحة المزروعة، و كمية انتاج اللحوم.

**3. تبسيط العمليات الحسابية**

في بعض الأحيان، يكون استخدام صيغة المعادلة (2) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيمة كسرية، من أجل ذلك يمكن تبسيط الصيغة (2) إلى صيغة أسهل تعتمد على مجموع القيم و ليس على انحرافات القيم عن وسطها الحسابي، و هذه الصيغة في المعادلة (3) هي:

$$\boxed{r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}}$$

و بالتطبيق على بيانات المثال السابق، يتبع الآتي:

- **حساب مجاميع:**

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 305 | 592 | 180560 | 93025 | 350464 |
| 313 | 603 | 188739 | 97969 | 363609 |
| 297 | 662 | 196614 | 88209 | 438244 |
| 289 | 607 | 175423 | 83521 | 368449 |
| 233 | 635 | 147955 | 54289 | 403225 |
| 214 | 699 | 149586 | 45796 | 488601 |
| 240 | 719 | 172560 | 57600 | 516961 |
| 217 | 747 | 162099 | 47089 | 558009 |
| 2108 | 5264 | 1373536 | 567498 | 3487562 |

| المجاميع المطلوبة |
|---|
| $\sum x = 2108$ , $\sum y = 5264$ |
| $\sum xy = 1373536$ |
| $\sum x^2 = 567498$ |
| $\sum y^2 = 3487562$ |

- **حساب معامل الارتباط:**

باستخدام المجاميع السابقة، و بالتطبيق على المعادلة (3) أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\left(\sum x^2-\frac{(\sum x)^2}{n}\right)\left(\sum y^2-\frac{(\sum y)^2}{n}\right)}}$$

$$=\frac{1373536-\frac{(2108)(5264)}{8}}{\sqrt{\left(567498-\frac{(2108)^2}{8}\right)\left(3487562-\frac{(5264)^2}{8}\right)}}$$

$$=\frac{-13528}{\sqrt{(12040)(23850)}}=\frac{-13528}{16945.619}=-0.798$$

وهي نفس النتيجة السابقة:

3) **<u>معامل الارتباط الرتبي سبيرمان</u> Pearson**

أذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين، و مثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة، و مستوى الدخل، فإنه يمكن استخدام طريقتين بيرسون السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية، و يطلق على هذا المعامل (معامل الارتباط سبيرمان) و يعبر عنه بالمعادلة (4) التالية:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

حيث أن $d$ هي الفرق بين رتب مستويات المتغير الأول $x$ ، ورتب مستويات المتغير الثاني $y$ ، أي أن : $d = R_x - R_y$ .

**¥ مثال (2):**

فيما يلي تقديرات 10 طلاب في مادتي الاحصاء، و الاقتصاد:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ب+ | ب+ | ب | أ+ | جـ+ | ب+ | د+ | د | جـ+ | أ | تقديرات إحصاء |
| ب | جـ | ب | ب+ | ب | أ | جـ | جـ | د | أ+ | تقديرات اقتصاد |

و المطلوب:

أحسب معامل الارتباط بين تقديرات الطلبة في الماديتين، و ما هو مدلوله؟

**الحل:**

**1)** بفرض أن X هي تقديرات الاحصاء، y هي تقديرات الاقتصاد، يمكن حساب معامل الارتباط بينهما باستخدام المعادلة (4)، و ذلك باتباع الآتي:

**الجدول رقم (5−2): تقديرات الطلبة في الاحصاء و الاقتصاد**

| الرتب | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| تقديرات إحصاء | أ+ | أ | ب+ | ب+ | ب+ | ب | جـ+ | جـ+ | د+ | د |
| رتب $x$ | 1 | 2 | (3+4+5)/3=4 | | | 6 | (7+8)/2=7.5 | | 9 | 10 |
| تقديرات اقتصاد | أ+ | أ | ب+ | ب | ب | ب | جـ | جـ | جـ | د |
| رتب $y$ | 1 | 2 | 3 | (4+5+6)/3=5 | | | (7+8+9)/3=8 | | | 10 |

- إذا يمكن حساب المجموع: $\sum d^2$ كما يلي:

| $x$ | $y$ | رتب $x$ | رتب $y$ | $d$ | $d^2$ |
|---|---|---|---|---|---|
| أ | أ+ | 2 | 1 | 1 | 1 |
| جـ+ | د | 7.5 | 10 | -2.5 | 6.25 |
| د | جـ | 10 | 8 | 2 | 4 |
| د+ | جـ | 9 | 8 | 1 | 1 |
| ب+ | أ | 4 | 2 | 2 | 1 |
| جـ+ | ب | 7.5 | 5 | 2.5 | 6.25 |
| أ+ | ب+ | 1 | 3 | -2 | 4 |
| ب | ب | 6 | 5 | 1 | 1 |
| ب+ | جـ | 4 | 8 | -4 | 16 |
| ب+ | ب | 4 | 5 | -1 | 1 |
| | | | | | 44.5 |

$$\sum d^2 = 44.5$$

- معامل الارتباط هو:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$

$$= 1 - 0.2697 = 0.7303$$

**2)** بما أن r = 0.703، و يدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالب في مادة الاحصاء و مادة الاقتصاد.

✓ **ملاحظة:** يمكن استخدام صيغة معامل سبيرمان في حساب الارتباط متغيرين كميين، حيث يتم استخدام رتب القيم التي يأخذها المتغير، و نترك للطالب القيام بحساب معامل ارتباط الرتب بين المساحة و الكمية في المثال السابق، و عليه أن يقوم بتفسير النتيجة

$$\sum d^2 = 148$$

# الانحدار الخطي البسيط
# Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة و تحليل أثر متغير كمي آخر، و من الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على انتاجية الدنم.
- دراسة أثر الانتاج على التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- أثر الدخل على الانفاق الاستهلاكي.

و هكذا هناك أمثلة في كثير من النواحي الاقتصادية، و الزراعية، و التجارية، و العلوم السلوكية، و غيرها.

**1) <u>نموذج الانحدار الخطي:</u>**

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين و يسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني و يسمى بالمتغير التابع أو المتنبأ به، و من ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما توضحه المعادلة (5):

$$\boxed{y = \beta_0 + \beta_1 x + e}$$

حيث أن:

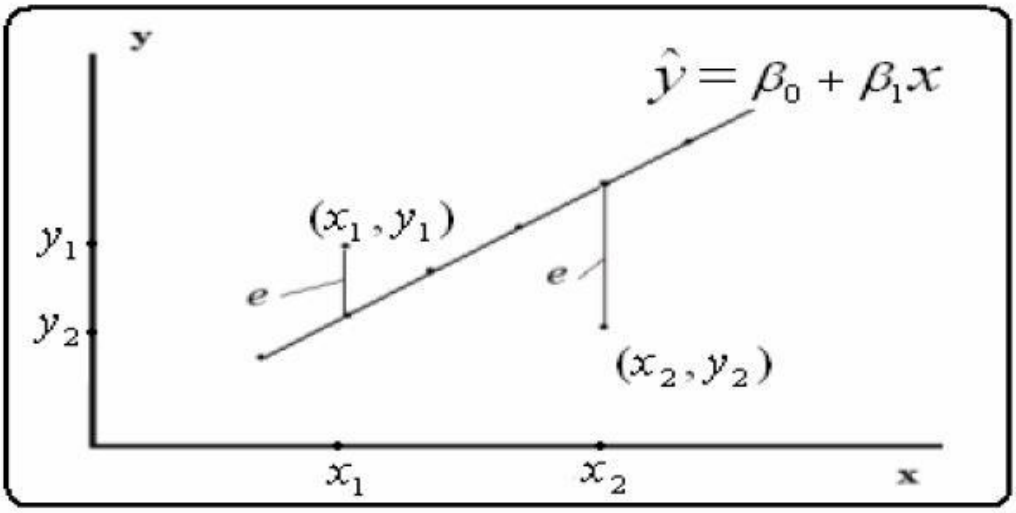
$y$ : هو المتغير التابع (الذي يتأثر)

$x$ : هو المتغير المستقل ( الذي يؤثر )

$\beta_0$ : هو الجزء المقطوع من المحور الرأسي $y$ ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل $x$ ، أي في حالة $x=0$

$\beta_1$ : ميل الخط المستقيم $(\beta_0+\beta_1 x)$ ، ويعكس مقدار التغير في $y$ إذا تغيرت $x$ بوحدة واحدة.

$e$ : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية $y$ ، والقيمة المقدرة $\hat{y}=\beta_0+\beta_1 x$ ، أي أن : $e=y-(\beta_0+\beta_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.




**<u>الشكل رقم (5-3): شكل الانحدار الخطي بين متغيرين</u>**

**2) <u>تقدير نموذج الانحدار الخطي البسيط:</u>**

يمكن تقدير معاملات الانحدار $(\beta_1, \beta_0)$ في النموذج (5) باستخدام طريقة المربعات الصغرى، و هذا التقدير هو الذي يجعل مجموع مربعات الخطاء العشوائية

$$\sum e^2 = \sum (y-(B_0+B_1 x)^2)$$

أقل ما يمكن، و يحسب هذا التقدير بالمعادلة (6) التالية:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم $\bar{x}$، $\bar{y}$ هو الوسط الحسابي لقيم $y$، و تكون القيمة المقدرة للمتغير التابع هو

$$\hat{y} = \widehat{B_0} + \widehat{B_1}x$$

و يطلق على هذا التقدير (تقدير معادلة انحدار y على x )

- **مثال (3):** فيما يلي بيانات عن كمية البروتين اليومي بالغرام التي يحتاجها العجل الرضيع، و مقدار الزيادة في وزن العجل بالكيلوم غرام، و ذلك لعينة من العجول الرضيعة حجمها 10.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| كمية البروتين | 10 | 11 | 14 | 15 | 20 | 25 | 46 | 50 | 59 | 70 |
| الزيادة في الوزن | 10 | 10 | 12 | 12 | 13 | 13 | 19 | 15 | 16 | 20 |

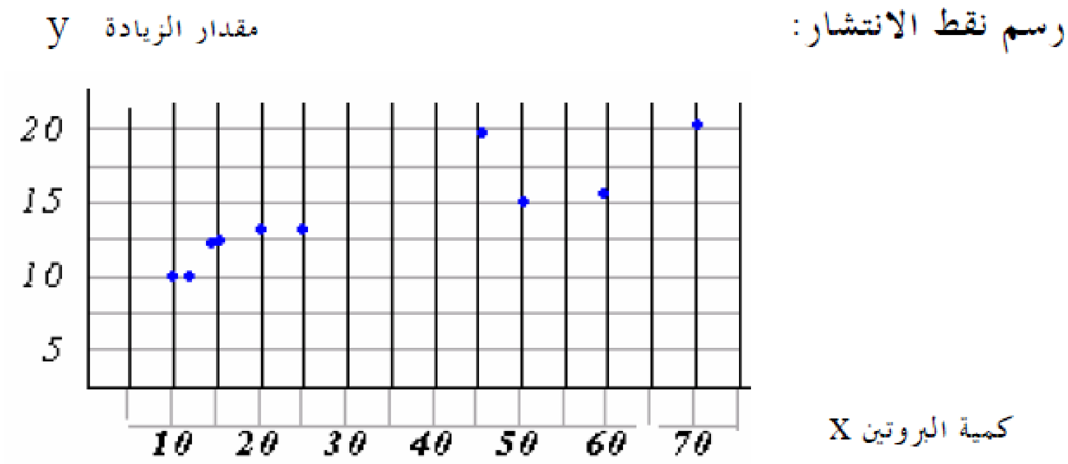
**و المطلوب:**

1. ارسم نقطة الانتشار، و ما هو توقعاتك لشكل العلاقة.
2. قدر معادلة انحدار الوزن على كمية البروتين.
3. فسر معادلة الانحدار.
4. ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 غرام من البروتين؟ و ما عو مقدار الخطأ العشوائي؟
5. ارسم معادلة الانحدار على نقط الانتشار في المطلوب الأول.

**الحل:**

**1. رسم نقطة الانتشار**

**الشكل رقم: (5-4)**




من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

**2. تقدير معادلة الانحدار.**

بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلتين في (6)، و ثم يتم حساب المجاميع التالية:

**<u>الجدول رقم (5-3): العلاقة بين البروتين و الزيادة في الوزن</u>**

| كمية البروتين $x$ | الزيادة في الوزن $y$ | $x\ y$ | $x^2$ |
|---|---|---|---|
| 10 | 10 | 100 | 100 |
| 11 | 10 | 110 | 121 |
| 14 | 12 | 168 | 196 |
| 15 | 12 | 180 | 225 |
| 20 | 13 | 260 | 400 |
| 25 | 13 | 325 | 625 |
| 46 | 19 | 874 | 2116 |
| 50 | 15 | 750 | 2500 |
| 59 | 16 | 944 | 3481 |
| 70 | 20 | 1400 | 4900 |
| 320 | 140 | 5111 | 14664 |

المجاميع المطلوبة

$$\sum x = 320$$
$$\sum y = 140$$
$$\sum xy = 5111$$
$$\sum x^2 = 14664$$

إذا الوسط الحسابي:

$$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$$

$$\bar{y} = \frac{\sum x}{n} = \frac{140}{10} = 14$$

- بتطبيق المعادلة الأولى في (6) يمكن حساب $\widehat{\beta_1}$ كما يلي:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2}$$
$$= \frac{6310}{44240} = 0.1426$$

- بتطبيق المعادلة الثانية في (6) يمكن حساب $\widehat{\beta_1}$ كما يلي:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 14 - (0.1426)(32) = 9.4368$$

- معادلة الانحدار المقدرة، هي:

$$\hat{y} = 9.44 + 0.143x$$

**3. تفسير المعادلة:**

- الثابت $\widehat{\beta_0} = 9.44$ : يدل على أنه في حالة عدم استخدام البروتين في التغذية، فإن الوزن يزيد 9.44 كغ.
- معامل الانحدار $\widehat{\beta_1} = 0.143$: يدل على أنه كلما زادت كمية البروتين غرام واحد واحد، حدث زيادة في وزن العجل بمقدار 0.143 كغ، أي زيادة مقدارها 143 غ.

**4. مقدار الزيادة في الوزن عند x = 50 هو:**

$$\hat{y} = 9.44 + 0.143(50) = 16.59$$

**5. رسم معادلة الانحدار على نقطة الانتشار.**

يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

| $x$ | 50 | 10 |
|---|---|---|
| $\hat{y}$ | 16.59 | 10.87 |

**<u>الشكل رقم (5-5): شكل معادلة الانحدار</u>**، هي: