Probability and Statistics

Chapter one : Descriptive Statistics

(١) علم الاحصاء : هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهرة أو فرضية (ظواهر أو فرضيات) معينة، وتنظيم وترتيب وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

ويشكل عام ان هذا العلم يعتبر جمع لفرعين رئيسين هما :

(أ) الاحصاء الوصفي (Descriptive Statistics)

وهو العلم الذي يتضمن الطرق والاساليب المستخدمة في جمع البيانات والمعلومات عن ظاهرة معينة أو مجموعة من الظواهر، وكيفية تنظيم وتصنيف وتبويب هذه البيانات مع امكانية عرضها في جداول ورسومات بيانية، ومن ثم حساب بعض المؤشرات الاحصائية منها.

(ب) الاحصاء الاستدلالي (الاستنتاجي) (Inferential Stat.)

وهو العلم الذي يهتم بموضوعي التخمين (Estimation) واختبار الفرضيات (Testing of hypothesis).

---

(٢) المجتمع (Population) : هو جميع مفردات أو وحدات الظاهرة تحت البحث، فقد يكون المجتمع مكون من مجموعة من الناس أو مجموعة من المزارع في منطقة معينة أو وحدات سلع معينة ينتجها معمل معين وهكذا.

والمجتمعات على انواع، مجتمعات محدودة (finite) تضم عدداً محدوداً من القيم مثل مجموعة النخيل في بستان معين أو مجتمعات غير محدودة تضم عدد غير منتهي مثل عدد النجوم في السماء.

وقد يكون المجتمع متماثل اي ان كل مفردة من مفرداته تحمل نفس الصفة مثل فصيلة الدم لدى الانسان أو غير متماثل مثل الاطوال والاوزان.

وقد يكون المجتمع صغيراً أو كبيراً، معتمداً على حجم المجتمع، ويرمز لحجم المجتمع بالرمز (N).

(٣) العينة (Sample) : هي جزء من المجتمع يجري اختيارها وفق قواعد خاصة لكي تمثل المجتمع تمثيلاً صحيحاً.

العينة العشوائية Random Sample

هي مجموعة وحدات احصائية تختار من المجتمع الاحصائي على اضوء اسس خاصة. فتؤخذ من كل طبقات المجتمع وبنسب معينة بحيث تمثل المجتمع تمثيلاً صحيحاً، اي يجب ان تكون عينة متماثلة. ومن فوائد اختيار العينة انها تمثل المجتمع فتقلل بذلك الوقت والجهد والامكانيات المادية.

يرمز لحجم العينة ب (n).

البيانات (Data) هي المعلومات والارقام والاعداد التي تجمع عن ظاهرة معينة قيد الدراسة مثل اطوال الطلبة للمرحلة الثالثة في قسم الرياضيات في جامعة بغداد.

تبويب البيانات هي عملية ترتيب وتكشيف البيانات الخام التي جمعت عن ظاهرة معينة قيد الدراسة بطرق معينة، لكي تكون هذه البيانات واضحة وسهلة الفهم والوصف لاستخلاص النتائج منها.

البيانات على نوعين 1- بيانات مبوبة Group Data

2- بيانات غير مبوبة Ungroup Data

البيانات غير المبوبة (Ungroup Data)

1- مقاييس النزعة المركزية (Measure of Central Tendency)

2- مقاييس التشتت والاختلاف (Measurement of Variations)

1- مقاييس النزعة المركزية:

هناك بعض القيم تلتف حولها مفردات المجتمع تسمى هذه القيم (نقاط الوسط) أو (نقاط التمركز) .. وتدعى هذه النقاط بمقاييس النزعة المركزية ومن اهم هذه المقاييس: الوسط الحسابي، الوسط الهندسي، الوسط الموزون الوسط التوافيقي، الوسط فوق الهندسي، الوسيط، المنوال.

① الوسط الحسابي (المعدل) (Arithmetic Mean):

الوسط الحسابي للعينة: $\bar{X}$

ليكن n حجم العينة (عدد مفردات العينة)

X_1 و X_2 و ... و X_n مفردات العينة

$$\bar{X} = \frac{\text{مجموع المفردات}}{\text{عددها}}$$

$\bar{X}$ متغير من عينة الى اخرى بتغير مفردات العينة.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

الوسط الحسابي للمجتمع μ

$$\mu = \frac{\text{مجموع المفردات}}{\text{عددها}}$$

لتكن X_1 و X_2 و ... و X_N مفردات المجتمع

و ان N حجم المجتمع (عدد مفردات المجتمع)

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

ويعتبر μ ثابت (معلمة) لان عدد مفردات المجتمع ثابت لا يتغير.

ان الوسط الحسابي $\bar{X}$ أو μ عندما تكون اوزان المجتمع متساوية.
اما اذا كانت الاوزان غير متساوية فنستخدم (الوسط الحسابي الموزون)

(الوسط الموزون) : (Weighted Mean)

عندما تكون الاوزان لمفردات المجتمع غير متساوية نستخدم كل مفردة مع وزنها (w = (weight)، ويدعى هذا المقياس (بالوسط الموزون أو المرجح).

$$\bar{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

x_i مفردات المجتمع
w_i اوزان المجتمع
$i = 1, 2, \ldots, n$

مثال البيانات الآتية تمثل الدرجات التي يحصل عليها أحد الطلبة وعدد الساعات الاسبوعية لكل مادة، ويطلب ايجاد الوسط الحسابي أو معدل درجات هذا الطالب:

الدرجة	80	85	70	78	90	96	65
الساعات الاسبوعية	3	2	4	3	2	3	2

الساعات الاسبوعية تمثل الاوزان، لذلك نستخدم الوسط الحسابي الموزون:

$$\bar{X}_w = \frac{\sum_{i=1}^{7} w_i x_i}{\sum_{i=1}^{7} w_i} = \frac{3(80) + 2(85) + \cdots + 2(65)}{3 + 2 + \cdots + 2} = 79.95$$

مميزات الوسط الحسابي (Mean)

١- سهولة حسابه.
٢- يدخل في حسابه جميع مفردات المجتمع التي تمثله.
٣- يتأثر بوجود القيم المتطرفة سواء كانت صغيرة أم كبيرة، وهذا عيب فيه.
٤- مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر. (مجموع الوسط)

$$\sum_{i=1}^{n} (x_i - \bar{X}) \quad \text{عوض} \quad \sum_{i=1}^{n} (x_i - \bar{X}) = \text{Zero} \qquad \text{(Prove that).}$$

$$\sum_{i=1}^{n} x_i - n\bar{X} \quad \sum x_i = 0 \quad (\text{٣})$$

② الوسيط (Median)

هو القيمة التي تحتل المرتبة الوسطى بعد ترتيب القيم قيد الدرس (البيانات) ترتيباً تصاعدياً أو تنازلياً.

أمثلة

① وسيط مجموعة الأرقام 10، 10، 8، 8، 6، 5، 4، 4، 3 هو 6.

② وسيط مجموعة الأرقام 18، 15، 12، 11، 9، 7، 5، 5 هو $\frac{9+11}{2} = 10$

وبصورة عامة: ان مرتبة الوسيط (R) هي:

$$R = \frac{n+1}{2}$$

وان الوسيط هو القيمة المتوسطة (اذا كان عدد المفردات n فردياً)

وان الوسيط هو متوسط القيمتين الوسيطيتين (اذا كان عدد المفردات n زوجياً)

مثال ① اذا كانت البيانات الآتية تمثل الأطوال بالسنتمتر، جد الوسيط لها.

110، 106، 113، 116، 102، 114، 111

الحل: نرتب القيم تصاعدياً فنحصل على:

102، 106، 110، 111، 113، 114، 116 (عدد البيانات فردي) $n = 7$

$$R = \text{رتبة الوسيط} = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

∴ الوسيط (M) = 111 (المرتبة الرابعة)

مثال ② جد وسيط البيانات الآتية:

9، 8، 7، 12، 10، 13، 15، 17

الحل: نرتب المفردات تصاعدياً:

7، 8، 9، 10، 12، 13، 15، 17 (عدد البيانات زوجي) $n = 8$

$$R = \text{رتبة الوسيط} = \frac{n+1}{2} = \frac{8+1}{2} = 4.5$$

∴ الوسيط (M) = معدل القيمتين الوسيطيتين (المرتبة الرابعة والخامسة)

$$M = \frac{10+12}{2} = 11$$

---

③ المنوال (Mode)

هو القيمة التي تتكرر أكثر من غيرها أو هو القيمة الأكثر شيوعاً وتكراراً بين مجموعة قيم الظاهرة قيد الدرس.

وهناك منوال رئيسي ومنوال ثانوي.

وهناك مجموعة من القيم ذات منوال واحد وأخرى ذات منوالين أو قد تكون عديمة المنوال (ليس لها منوال).

امثلة

① مجموعة القيم: 18 و 12 و 11 و 10 و 10 و 9 و 9 و 9 و 7 و 2 و 2 و 5 .
لها منوال واحد هو (9)

$\therefore M_o = 9$

② مجموعة القيم: 18 و 15 و 12 و 10 و 8 و 5 و 3 .
ليس لها منوال .

③ مجموعة القيم: 9 و 7 و 7 و 7 و 5 و 5 و 5 و 4 و 4 و 3 و 2
لها منوالان هما 4 و 7 .

ما العلاقة بين الوسط والوسيط والمنوال ؟

أ- في حالة التوزيعات المتماثلة (المنتظمة) : كالتوزيع الطبيعي ، فان المقاييس الثلاثة تكون متساوية .

i.e. $Me = M_o = \mu$

ب - في حالة التوزيعات غير المتماثلة (غير المنتظمة) : فان المنوال والوسيط هما الاكثر تمثيلاً للتمركز وان :

الوسط الحسابي = (3 × الوسيط) − المنوال / 2

i.e. $\bar{X} = \frac{3Me - M_o}{2}$

ب - مقاييس التشتت أو الاختلاف

وهي مقاييس تقيس مدى تشتت القيم المختلفة في ظاهرة طبيعية عن الوسط الحسابي أو مقاييس النزعة المركزية ، ومن هذه المقاييس :
المدى ، التباين ، الانحراف المعياري ، الخطأ المعياري ، معامل الاختلاف ، الانحراف التربيعي

---

① المدى : (Range) وهو احد مقاييس التشتت و يمثل الفرق بين القيمة العليا والقيمة الدنيا لمجموعة المفردات .
ومن عيوب هذا المقياس انه لا يهتم سوى بمفردتين فقط هي القيمة العليا والدنيا ونتيجة لذلك لا يهتم بالاختلافات الموجودة بين المفردات الباقية ، فهو لا يهتم بحجم العينة فلا يمكن استخدامه كمقياس للمقارنة بين العينات عند اختلاف حجم العينة . ومن ناحية اخرى قد تكون القيم المتطرفة هي القيم العليا أو الدنيا وهذا عيب كبير وبذلك لا يعتبر هذا المقياس كفوء كلما زاد عدد مفردات العينة .

المدى (Range) :

هو الفرق بين أكبر قيمة وأصغر قيمة من قيم الظاهرة (المفردات).

$$R = L - U$$

L ← أكبر قيمة
U ← أقل قيمة

أمثلة

① جد مدى البيانات الآتية :

45 ، 50 ، 55 ، 60 ، 65

$$R = L - U = 65 - 45 = 20$$

② مدى البيانات الآتية (الأوزان) :

45 ، 100 ، 45 ، 55 ، 95 ، 50 كغم

$$R = 45 - 100 = 55 \text{ كغم}$$

---

② التباين والانحراف المعياري (Variance & Standard Deviation)

ان مقاييس التشتت (المدى) تقيس مدى اقتراب أو ابتعاد القيم للعينة (المفردات) عن نقاط التمركز (الوسط الحسابي وغيره) .. ولحل مشكلة الاختلافات الموجودة في مفردات العينة أو المجتمع ، نستخدم مقاييس الوسط الحسابي لكن المشكلة في ان (مجموع انحرافات القيم عن الوسط الحسابي = صفر) ، ولغرض التخلص من القيم السالبة في الانحرافات ، طرأت فكرة تربيع هذه الانحرافات ليدل على مدى الاختلافات ولاجل التخلص من مشكلة اخرى هي حجم المجتمع أو العينة ، تم القسمة على عدد مربعات انحرافات القيم . فتكون مقياس جديد يدعى التباين (Variance) الذي رمزه (S^2) ويقرأ (سكما سكوير) للعينة اما للمجتمع فان رمزه (σ^2) :

صيغة تباين المجتمع :

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N-1}$$

اما صيغة تباين العينة :

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

وان استخدام $(n-1)$ بدلاً من (n) الذي هو حجم العينة يعود الى اعتبارات رياضية تتعلق بالتقديرات غير المتحيزة وتسمى $(n-1)$ بدرجات الحرية (degree of freedom)

اما الانحراف المعياري (Standard deviation) : هو الجذر التربيعي للتباين

$$\sigma = \sqrt{\sigma^2}$$ الانحراف للمجتمع

$$S = \sqrt{S^2}$$ الانحراف للعينة

(مثال) جد التباين للقيم الآتية التي تمثل وزن اللحم الصافي لستة رؤوس من الغنم .. تمثل مجتمعاً قائماً بذاته :

35 و 34 و 40 و 38 و 37 و 32 (كغم)

sol.

الوسط الحسابي للمجتمع $M = \frac{35+34+40+38+37+32}{6} = \frac{216}{6} = 36$ كغم

(تباين المجتمع) ويعني تباين كل قيمة عن الوسط الحسابي M للمجتمع .

$$\sigma^2 = \sum_{i=1}^{6} (x_i - 36)^2 / 6-1 = \frac{42}{5} = 8.4 \text{ (كغم}^2\text{)}$$

(نفس المثال اعلاه لكن للعينة)

الوسط الحسابي للعينة $\bar{X} = \frac{35+34+\cdots+32}{6} = 36$ كغم

(تباين العينة) $S^2 = \sum_{i=1}^{6} (x_i - 36)^2 / 6$

$$= \frac{42}{5} = 8.4$$

البيانات المبوبة (Group Data)

١- مقاييس النزعة المركزية ٢- مقاييس التشتت والاختلاف .

١- مقاييس النزعة المركزية :

① الوسط الحسابي :

x_i = يمثل مركز الفئة ؛

f_i = يمثل تكرار الفئة ؛

k = يمثل عدد الفئات .

$$\bar{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

② الوسيط : لايجاد الوسيط نتبع الخطوات الآتية :

① ترتيب الفئات تصاعدياً ثم ايجاد التكرار المتجمع الصاعد.

② ايجاد رتبة الوسيط بالصيغة :

$$R = \frac{\sum f_{i+1}}{2} = \frac{n+1}{2}$$

③ تحديد الفئة الوسيطية من خلال مقارنة رتبة الوسيط وموقعها في أول تكرار متجمع صاعد.

④ نستخدم الصيغة التالية في ايجاد الوسيط :

$$Me = Le + \frac{R - G_1}{G_2} \times We$$

حيث ان :

L_e = الحد الأدنى للفئة الوسيطية

W_e = طول الفئة الوسيطية

G_1 = التكرار المتجمع الصاعد للفئة قبل الوسيطية

G_2 = تكرار الفئة الوسيطية .

③ المنوال : لإيجاد المنوال يجب تحديد الفئة المنوالية أولاً وهي الفئة التي تقابل أكبر تكرار ونستخدم الطريقة الآتية :

$$M_o = L_o + \frac{D_1}{D_1 + D_2} \times W_o$$

حيث ان :

L_o : الحد الأدنى للفئة المنوالية

D_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة قبل المنوالية .

D_2 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة بعد المنوالية .

W_o : طول الفئة المنوالية .

ب - مقاييس التشتت والاختلاف :

① المدى :

$$R = X_L - X_S$$

حيث ان :

X_L = الحد الأعلى للفئة الأخيرة

X_S = الحد الأدنى للفئة الأولى

② الانحراف المعياري والتباين :

ليكن :

f_i = التكرار المقابل لمركز الفئة x_i

x_i = مركز الفئة = $\left(\frac{\text{الحد الأعلى للفئة} + \text{الحد الأدنى للفئة}}{2}\right)$

k = عدد الفئات

$\text{Var}(X) = S^2 =$ التباين

$S = \sqrt{S^2} =$ الانحراف المعياري

التباين

$$S^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i - 1} - \frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{\sum f_i (\sum f_i - 1)}$$

H.W : جد الوسط والوسيط والمنوال للبيانات الآتية :

Classes	frequency (f_i)	x_i (mid point)	$x_i f_i$	C.f. (التكرار المتجمع الصاعد)
25 - 35	2	?	?	?
35 - 45	5			
45 - 55	12			
55 - 65	13			
65 - 75	20			
75 - 85	15			
85 - 95	8			
95 - 105	5			

مثال: جد الوسط (Mean) والوسيط (Median) والمنوال (Mode) والتباين (Variance) والانحراف المعياري (S.d.) والمدى (Range) لاطوال النباتات الآتية:

Classes	f_i	x_i	$x_i f_i$	c.f.
2-4	3	3	9	3
4-6	5	5	25	8
6-8	2	7	14	(10)

يساوي دائما $n = \Sigma f_i = 10$

$n = \Sigma f_i = 10$, $\Sigma x_i = 15$, $\Sigma x_i f_i = 48$

$$\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} \simeq \bar{X}_w \qquad \text{(Mean)}$$

$$Me = L_e + \frac{R - G_1}{G_2} \times w_e \qquad \text{(Median)}$$

$$= 4 + \frac{5.5 - 3}{5} \times 2 = 5 \in (4-6)$$ الفئة الوسيطية

$$Mo = L_o + \frac{D_1}{D_1 + D_2} \times w_o \qquad \text{(Mode)}$$

$$= 4 + \frac{(5-3)}{(5-3)+(5-2)} \times (2)$$

$$= 4 + \frac{(2)(2)}{2+3} = 4 + \frac{4}{5} = \frac{24}{5} = 4.8$$

$$R = X_L - X_S = 8 - 2 = 6 \qquad \text{(Range)}$$

$$\delta_x^2 = \frac{\Sigma x_i^2 f_i}{\Sigma f_i - 1} - \frac{(\Sigma x_i f_i)^2}{\Sigma f_i (\Sigma f_i - 1)} \qquad \text{(Variance)}$$

نكمل عن طريق الجدول:

x_i	x_i^2	$x_i f_i$	$x_i^2 f_i$	f_i
3	9	9	27	3
5	25	25	125	5
7	49	14	98	2

$\Sigma f_i = 10$
$\Sigma f_i - 1 = 9$
$\Sigma x_i f_i = 48$
$\Sigma x_i^2 f_i = 250$
$(\Sigma x_i f_i)^2 = (48)^2 = 2304$

الارتباط والانحدار

Correlation and Regression

الارتباط (Correlation)

هو العلاقة التي تربط بين ظاهرتين أو أكثر، وهو على انواع منها:
الارتباط البسيط، الارتباط المتعدد، الارتباط الجزئي.

الارتباط البسيط (Simple Correlation)

لمعرفة العلاقة بين ظاهرتين X و Y ونوع هذه العلاقة، نستخدم مؤشراً احصائياً يدعى (معامل الارتباط البسيط) (Simple Correlation Coefficient)
حيث ان Y يمثل المتغير المعتمد (التابع) و X هو المتغير المستقل. ويرمز لهذا المعامل بالرمز (r) حيث

$$-1 \leq r \leq +1$$

تعني الاشارة السالبة ان العلاقة بين الظاهرتين عكسية وضعيفة.
اما الاشارة الموجبة ان العلاقة بين الظاهرتين طردية وقوية.
والصيغة لهذا المعامل:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

(معامل ارتباط سبيرمان للقيم القابلة للقياس)

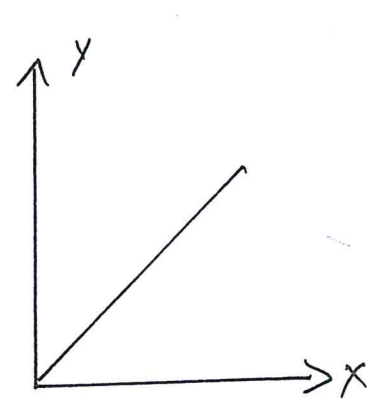
حيث ان n تمثل حجم العينة = عدد البيانات في كل عمود (عدد بيانات X = عدد بيانات Y)

If $r_{x,y} \in (0, 1]$ ⟶ معامل ارتباط قوي

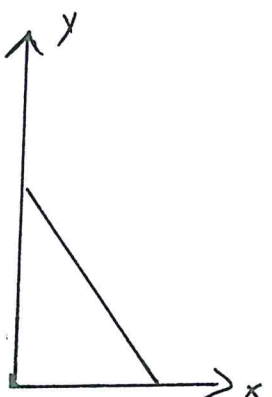
If $r_{x,y} \in [-1, 0)$ ⟶ معامل ارتباط ضعيف

If $r_{x,y} = \pm 1$ ⟶ معامل ارتباط تام

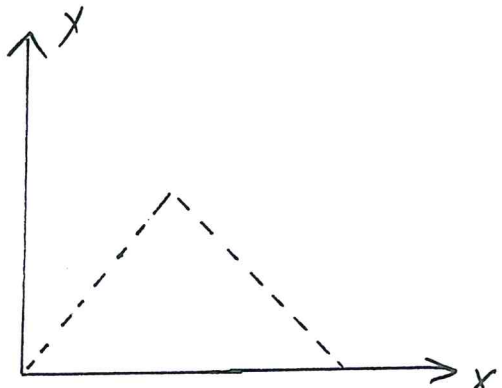
If $r_{x,y} = 0$ ⟶ لايوجد ارتباط بين X و Y



معامل ارتباط طردي



معامل ارتباط عكسي



$r = 0$
لاتوجد علاقة بين X و Y

note . ٢٠٠٠ - ٢٠٠٧

اما معامل الارتباط البسيط للمجتمع :

$$\rho_{x,y} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2}}$$

مثال : صممت تجربة لدراسة تأثير زيادة جرعة منوم معين في وقت النوم ، وتم تكوين ثلاث قراءات لكل من المستويات الثلاثة للجرعة ، والمطلوب معرفة اذا كانت هناك علاقة بين الجرعة ووقت النوم .

الجرعة (Y)	وقت النوم (X)	XY	X^2	Y^2
3	4	12	16	9
3	6	18	36	9
3	5	15	25	9
10	9	90	81	100
10	8	80	64	100
10	7	70	49	100
15	13	195	169	225
15	11	165	121	225
15	9	135	81	225
84	72	780	642	1002

$$r_{x,y} = \frac{9(780) - (72)(84)}{\sqrt{[9(642) - (72)^2][9(1002) - (84)^2]}} = \frac{972}{1079.55}$$

$$= 0.9 \in (0,1]$$

∴ العلاقة قوية وطردية بين الجرعة ووقت النوم

ملاحظة : اذا كانت كلا الظاهرتين غير قابلة للقياس ، نستطيع ايجاد العلاقة بين الظاهرتين باستخدام (معامل الارتباط لسبيرمان) ، وذلك باعطاء رتب الى القيم أو البيانات الغير قابلة للقياس ثم نستخدم الصيغة الآتية :

$$r = 1 - \frac{6\Sigma D_i^2}{N(N^2-1)}$$ (معامل سبيرمان للرتب)

D_i : تمثل فرق الرتب لكل زوج

N : تمثل عدد القيم الغير قابلة للقياس (الغير كمية) .

مثال: جد العلاقة بين درجة الطالبين A و B من البيانات الآتية:

A	B	D_i (A-B)	D_i^2
جيد (4)	مقبول (2)	2	4
مقبول (2)	جيد جداً (5)	-3	9
ضعيف (1)	مقبول (2)	-1	1
جيد (4)	متوسط (3)	1	1
جيد جداً (5)	جيد (4)	1	1
جيد (4)	متوسط (3)	1	1

* نرتب اولاً تصاعدياً ثم نضع الرتب

ضعيف (1)
مقبول (2)
متوسط (3)
جيد (4)
جيد جداً (5)

$N = 6$ (عدد البيانات)

$\sum D_i^2 = 17$

$$r_{A,B} = 1 - \frac{6.\sum D_i^2}{N(N^2-1)} = 1 - \frac{6.(17)}{6(6^2-1)}$$

$$= 0.51 \in (0,1]$$ (العلاقة قوية وطردية)

مثال آخر (H.W.)

A	B	D_i	D_i^2
متوسط (3)	جيد جداً (5)	-2	4
ضعيف (1)	متوسط (3)	-2	4
جيد (4)	مقبول (2)	2	4
متوسط (3)	جيد جداً (5)	-2	4
جيد جداً (5)	متوسط (3)	2	4

$N = 5$ ، $\sum D_i^2 = 20$

$$r_{A,B} = 1 - \frac{6(20)}{5(5^2-1)} = ?$$

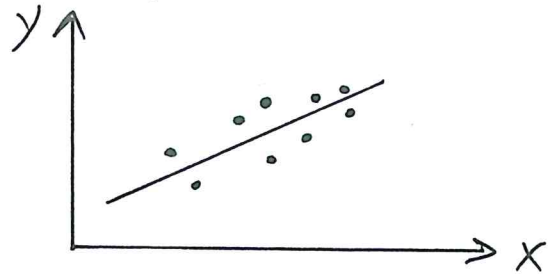
* ما نوع العلاقة بين A و B؟

الانحدار Regression

شكل الانتشار: (Scatter digram)

هو كل زوج من ازواج القيم التي تمثل متغيراتها بين محورين متعامدين وعلى شكل نقط (points).

وهذا الشكل في اغلب الاحيان يمثل علاقة دالية وذات اتجاه مستقيم (دالة خطية)



خط الانحدار (Linear Regression)

هو الخط الذي يمر بأكبر عدد ممكن من النقاط في شكل الانتشار.

اما معادلة خط الانحدار فهي:

$$Y = \alpha + \beta X + e$$

حيث ان:

α: هي المسافة بين نقطة تقاطع خط الانحدار مع المحور Y، ونقطة الاصل.

β: هي ميل خط الانحدار أو ظل الزاوية التي يصنعها خط الانحدار مع المحور X.

$$\text{i.e. } \beta = \tan\theta \Rightarrow \theta = \tan^{-1}\beta$$

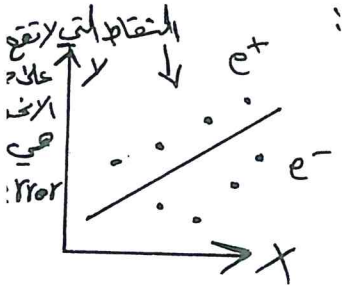
e: هو الخطأ الذي يمثل الفرق بين القيمة الحقيقية والتقديرية لـ Y

$$\text{i.e. } e_i = Y_i - \hat{Y}_i \qquad (\text{تقرأ } \hat{} \text{ بـ hat})$$

$$\text{if } e_i = 0 \Rightarrow Y_i = \hat{Y}_i$$

اما المعادلة التقديرية المستخدمة لحساب خط الانحدار هي:

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$



حيث ان:

$\hat{Y}$: القيمة التقديرية لـ Y.

$\hat{\beta}$: القيمة التقديرية لـ β.

$\hat{\alpha}$: القيمة التقديرية لـ α.

نحسب الدالة $\hat{Y}$ بحيث يكون الخطأ e اقل ما يمكن، ويمكن حساب قيمة $\hat{\alpha}$ و $\hat{\beta}$ باستخدام طريقة المربعات الصغرى في الحساب (Least Square).

$\hat{\beta}$ تسمى (معامل الانحدار) حيث ان ($\hat{\beta}=0$) يعني لا توجد علاقة بين X و Y واذا كانت $\hat{\beta}$ قيمة سالبة يعني ان العلاقة عكسية بين X و Y واذا كانت $\hat{\beta}$ قيمة موجبة يعني ان العلاقة بين X و Y طردية.

ليكن $\bar{X} \equiv \text{mean of } X$ & $\bar{Y} \equiv \text{mean of } Y$ فان:

لإيجاد $\hat{\alpha}$ و $\hat{\beta}$ باستخدام طريقة المربعات الصغرى، حيث ان $r_{x,y}$ يمثل (معامل الارتباط البسيط) وان δ هو الانحراف المعياري:

$X \mid y$	$y \mid X$
$\hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$	$\hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$
$= \frac{\delta_x}{\delta_y} \cdot r_{x,y}$	$= \frac{\delta_y}{\delta_x} \cdot r_{x,y}$
$\hat{\alpha} = \bar{X} - \hat{\beta}\bar{y}$	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{X}$

ملاحظة: ان مفهوم الانحدار الخطي على صلة وثيقة بمفهوم الارتباط الخطي، حيث ان الانحدار يقدر العلاقة الخطية بين المتغيرات اما الارتباط الخطي فهو يصف العلاقة بين المتغيرات (يصف العلاقة نفسها).

مثال (١): اجريت احدى التجارب لدراسة تأثير عقارين في تخفيض دقات القلب لغير البالغين، حيث ان المتغير المستقل (X) يمثل الجرعة بالملغم من العقار والمتغير المعتمد (y) يمثل الانخفاض في دقات القلب (دقة \ الدقيقة).

① ارسم شكل الانتشار ② جد معادلة خط الانحدار التقديرية.

③ جد مقدار الانخفاضات في دقات القلب عندما تكون الجرعة (٣) ملغم.

X	y	X^2	XY
0.5	10	0.25	5
0.75	8	0.5625	6
1.00	12	1	12
1.25	12	1.5625	15
1.5	14	2.25	21
1.75	12	3.0625	21
2.00	14	4	32
2.25	18	5.0625	40.5
11	102	17.75	152.5

② لايجاد معادلة خط الانحدار التقديرية:

$$\hat{\beta} = \frac{(8)(152.5) - 11(102)}{8(17.75) - (11)^2}$$

$$\hat{\beta} = 4.67$$

$$\hat{\alpha} = \frac{102}{8} - 4.67\left(\frac{11}{8}\right)$$

$$= 6.33$$

$$\hat{y} = 6.33 + 4.67X$$

③ لايجاد $\hat{y}(3)$ ؛ اي ان $X = 3$ عندما $(y \mid X)$

$$\hat{y}(3) = 6.33 + 4.7(3)$$

$$= 20.34$$

$$\mu_x = \frac{35+34+40+38+37+32}{6} = 36$$

$$\delta^2 = \sum_{i=1}^{6}(x_i - 36)^2/6-1 = \frac{42}{6-1} = 8.4 \text{ (كغم}^2\text{)}$$

اي ان تباين كل قيمة عن الوسط الحسابي للمجتمع μ_x هو (7) كغم^2.

نفس السؤال اعلاه لكن للعينة :

$$\bar{X} = \frac{35+34+\cdots+32}{6} = 36 \text{ كغم}$$

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{X})^2/n-1 = \sum_{i=1}^{6}(x_i-36)^2/5$$

$$= \frac{42}{5} = \underline{8.4} \text{ (كغم}^2\text{)}$$

∴ تباين العينة هو تقدير جيد لتباين المجتمع.

---

(٢) جد الوسيط والمنوال للارقام المكونة من ست ستات و سبع سبعات وثماني ثمانيات وتسع تسعات وعشر عشرات :

$$\text{Mean} = \mu = \frac{6(6)+7(7)+8(8)+9(9)+10(10)}{6+7+8+9+10} = ?$$

$$n = \sum w_i = 10+9+8+7+6 = 40 \text{ عدد القيم}$$

$$R = \frac{n+1}{2}, \quad R = \frac{40+1}{2} = 20.5$$

اي المرتبة 20 و 21

والوسيط هو

$$Me = (8+8)/2 = 8$$

ويمكن ايجاد الوسيط في هذا السؤال بطريقة اخرى كالاتي:

000666 7777777 88888888 99999999

10 10 10 10 10 10 10 10 10 10

الوسيط

المنوال (M_o) : القيمة التي تكرر اكثر من غيرها وهي اكثر قيمة تكررت هي (10)

$$\therefore M_o = 10$$

معامل الارتباط البسيط Simple Correlation Coefficient

$$r_{x,y} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

(٢) الجدول التالي يمثل أوزان خمسين طفل مصابين بمرض فقر الدم والمطلوب ايجاد الانحراف المعياري والتباين والمدى.

وزن الطفل classes	التكرار عدد الاطفال f_i	وسط الفئة x_i	x_i^2	$x_i f_i$	$x_i^2 f_i$
12-14	20	13	169	260	3380
14-16	18	15	225	270	4050
16-18	12	17	289	204	3468

$$n = \sum_{i=1}^{3} f_i = 50, \quad \sum x_i f_i = 734, \quad \sum x_i^2 f_i = 10898$$

$$\text{Range} = R = 18 - 12 = 6$$

$$\delta = \sqrt{\frac{10898}{50-1} - \frac{(743)^2}{50(50-1)}} = 1.58$$

$$\delta^2 = Var(X) = (1.58)^2 \simeq 2.5$$

(H.W) اذا كانت اوزان رؤوس الغنم في قطيع معين موزعة على النحو التالي لعينة مأخوذة من هذا القطيع. احسب قيمة كل من الوسط والوسيط والمنوال والتباين والانحراف المعياري لوزن الرأس الواحد.

الوزن (كغم)	التكرارات (عدد الرؤوس)
20-25	20
25-30	60
30-35	10
35-40	6

Ungrouped Data

(١) جد تباين القيم التالية التي تمثل وزن اللحم الصافي لستة رؤوس من الغنم وتمثل مجتمعاً قائماً بذاته:

(كغم) 32, 37, 38, 40, 34, 35

$$\delta_x^2 = \sum_{i=1}^{6} (x_i - \mu_x)^2 / N$$

(٩) جد الوسيط والوسيط والمنوال لمجموعة الأرقام والأعداد التالية :

5 ، 4 ، 8 ، 3 ، 7 ، 2 ، 9

$M_x = \Sigma x_i / 7 = 5.4$

نرتب الأرقام 2 ، 3 ، 4 ، 5 ، 7 ، 8 ، 9

الرتبة الرابعة $R = \frac{n+1}{2} = \frac{7+1}{2} = 4$ رتبة الوسيط

$M_e = 5$ ، M_o = لا يوجد منوال

---

(١٠) جد المنوال للأرقام التالية :

4 ، 7 ، 7 ، 7 ، 9 ، 9 ، 10 ، 12 ، 15

$M_o = 7$

1 ، 2 ، 3 ، 4 ، 4 ، 5 ، 5 ، 6 ، 6 ، 7 ، 8 ، 8 ، 10 ، 10 ، 11 ، 12

هناك (5) مناويل هي

$M_o = 4 ، 5 ، 6 ، 8 ، 10$

(مناويل ثانوية) $M_1 = 4$ ، $M_2 = 5$ ، $M_3 = 6$ ، $M_4 = 8$ و $M_o = 10$ منوال رئيسي

---

(١١) في شركة بها (80) عاملاً .. (60) يحصلون على 3.0 $ في الساعة وان (20) يحصلون على 2.0 $ في الساعة . أوجد متوسط دخولهم في الساعة .

$$\bar{X} = \frac{\Sigma w_i x_i}{\Sigma w_i} = \frac{(60)(3.0) + (20)(2.0)}{3+2} = \frac{220.0}{5}$$

---

(١٢) جد منوال الأرقام التالية :

2 ، 2 ، 5 ، 7 ، 9 ، 9 ، 9 ، 10 ، 10 ، 11 ، 12 ، 18

$M_o = 9$

3 ، 5 ، 8 ، 10 ، 12 ، 15 ، 16 ، 20

M_o = لا يوجد منوال (لا يوجد تكرار)

2 ، 3 ، 4 ، 4 ، 4 ، 5 ، 5 ، 7 ، 7 ، 7 ، 9 ، 10

$M_o = 7$ منوال رئيسي

$M_1 = 4$ منوال ثانوي

H.W بيّن من العينتين X و Y أفضل للدراسة وفق البيانات التالية :

$\bar{X} = 145$ ، $\bar{Y} = 80$ ، $S_x = 10$ ، $S_y = 10$ (استخدم معامل الاختلاف C.V.) .

(البيانات المبوبة) Grouped Data

① في تجربة لتحديد تأثير عقار معين لمستوى الكولسترول في الدم مقاساً بـ (Mg/100Ml) لأشخاص أعمارهم (30) سنة. سُجلت البيانات التالية لمجموعة عولجت بهذا الدواء. والمطلوب معرفة المنوال لمستوى الكولسترول في الدم.

مستوى الكوسترول (الفئات)	f_i عدد المرضى (التكرار)
120 – 160	2
160 – 200 ← الفئة المنوالية	5 ← أعلى تكرار
200 – 240	3
	$n = \sum_{i=1}^{3} f_i = 10$

$D_1 = 5 - 2 = 3$

$D_2 = 5 - 3 = 2$

$w_o = 200 - 160 = 40$

$L_o = 160$

$$M_o = L_o + \frac{D_1}{D_1 + D_2} * w_o$$

$$= 160 + \frac{3}{3+2}(40) = 160 + \frac{120}{5} = 160 + 24 = 184$$

---

② في دراسة لتحديد تأثير التوضين على الأرض تم قياس هذه المادة في البلازما (Mg/Ml) وبعد ساعتين من تناول الدواء من قبل (16) من المرضى تم الحصول على البيانات التالية والمطلوب: إيجاد مقدار مستوى هذه المادة (الوسيط).

Classes / مستوى الأرض (الفئات)	عدد المرضى (التكرارات) f_i	التكرار المتجمع الصاعد C.f.d.
0.505 – 1.005	4	4
1.005 – 1.505	3	7
1.505 – 2.000 ← الفئة الوسيطية	6	13
2.000 – 2.505	3	16
	$n = \sum f_i = 16$	

$$R = \frac{n+1}{2} = \frac{16+1}{2} = \frac{17}{2} = 8.5$$ رتبة الوسيط

$L_e = 1.505$ ، $w_e = 2.000 - 1.505 = 0.495$

$$M_e = L_e + \frac{R - G_1}{G_2}(w_e) = 1.505 + \frac{8.5 - 7}{3}(0.495) = ?$$

(٥) رتبت الارقام 85 رقماً و 150 رقماً، ما وسيطها؟

85 رقم فردي .. وهناك قيمة وسيطية واحدة هي (43)، حيث قبلها (42) رقماً وبعدها (42) رقماً .. وترتيب الوسيط الثالث والاربعين.

150 رقم زوجي .. وهناك قيمتين في الوسط حيث يتواجد قبلهما (74) رقماً وبعدهما (74) رقماً وترتيبهما (75) و(76) ووسطهما الحسابي هو الوسيط:

$$(76+75)/2 = Me$$

---

(٦) اذا كانت درجات طالب في الرياضة والطبيعية واللغة الانكليزية والصحة العامة هي على الترتيب 82، 86، 90، 70 وان عدد الساعات (ساعات المحاضرات) الاسبوعية لهذه المقررات هي 3، 5، 3، 1. أوجد متوسط الدرجات.

$$\bar{X} = \Sigma wixi / \Sigma wi$$

$$= \frac{(3)(82)+(5)(86)+(3)(90)+(1)(70)}{3+5+3+1} = 85$$

---

(٧) اوجد الوسط والمنوال والوسيط للارقام التالية:

3، 5، 2، 6، 5، 9، 5، 2، 8، 6

نرتب تصاعدياً:

X: 2، 2، 3، 5، 5، 5، 6، 6، 8، 9 $(n=10)$

$$\mu = Mean = \Sigma xi/10 = 5.1$$

$$Me = Median = (5+5)/2 = 5$$

$$Mo = Mode = 5$$

وللارقام التالية:

50.3، 49.5، 48.9، 51.6، 48.7

نرتب تصاعدياً (أو تنازلياً):

48.7، 48.9، 49.5، 50.3، 51.6

$$\mu = \Sigma xi/5 = 49.8$$

$$Me = 49.5$$

Mo = لايوجد

---

(٨) حصل طالب على درجات 96، 82، 93، 76، 85 في (5) مواد. جد الوسط والوسيط والمنوال:

$$\bar{X} = \Sigma xi/5 = 86 \quad , \quad R = \frac{n+1}{2} = \frac{5+1}{2} = 3$$

(الوسيط في المرتبة الثالثة)

لايوجد منوال Mo ، $Me = 85$ ← نرتب ثم

Chapter one
Descriptive Statistics

Ungrouped Data (البيانات غير المبوبة)

① جد الوسيط للأرقام التالية:

3, 4, 4, 5, 6, 8, 8, 8, 10

وسيطها هو (6).

نرتب الأعداد تصاعدياً (تنازلياً):

9, 16, 12, 11, 8, 9, 10

8, 9, 9, 10, 11, 12, 16

الوسيط هو (10) الذي يحتل المرتبة الرابعة.

لإيجاد الوسيط نرتب الأعداد تصاعدياً:

12, 22, 7, 5, 20, 7, 14, 10

5, 7, 7, 10, 12, 14, 20, 22

هناك مرتبتان وسيطيتان هما المرتبة الرابعة والخامسة (عدد القيم زوجي) ونحسب الوسيط بالطريقة التالية:

$Me = (12+10)/2 = 11$

---

② درجات طالب في ستة امتحانات كانت 78, 87, 68, 72, 91, 84

أوجد وسيط هذه الدرجات.

68, 72, 78, 84, 87, 91

عدد الدرجات زوجي فان هناك قيمتين في الوسط 78, 84 ووسطها الحسابي هو $(84+78)/2 = 81$ وهو الوسيط.

---

③ درجات طالب في ستة امتحانات هي 78, 87, 68, 72, 91, 84. أوجد الوسط الحسابي لهذه الدرجات.

$$\bar{X} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{84+91+72+68+87+78}{6} = \frac{480}{6} = 80$$

---

④ صف صائغ رقم (20) أربعة و(40) خمسة و (30) ستة والباقي كانوا سبعات. أوجد الوسط الحسابي لهذه الأرقام.

$$\bar{X} = \frac{\sum w_i x_i}{\sum w_i}$$

$$= \frac{(20)(4)+(40)(5)+(30)(6)+(10)(7)}{100} = \frac{530}{100} = 5.30$$

S.1.

$r_{x,y} = 0.87$, $\bar{X} = 59.9$, $\bar{Y} = 66.8$, $\sigma_x = 14.7$, $\sigma_y = 7$

$\hat{y} = \hat{\alpha} + \hat{\beta}x$ is the linear regression equation.

$$\hat{\beta} = \frac{\sigma_y}{\sigma_x} \cdot r_{x,y}$$

$$= \frac{7}{14.7}(0.87) = 0.41 \Rightarrow \hat{\beta} = 0.41$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{X}$$

$$= 66.8 - (0.41)(59.9) = 42.24$$

$$\therefore \hat{y} = 42.24 + (0.41)X$$

---

2) Based on the following data, find (estimate) the linear regression equation of $y|x$.

X: 52, 56, 60, 65, 70, 75.

Y: 44, 40, 39, 42, 38, 37.

sol. By Table:

X	Y	XY	X^2	Y^2
⋮	⋮	⋮	⋮	⋮

Then: $\Sigma X = 240$, $\Sigma Y = 378$, $\Sigma XY = 15033$, $\Sigma X^2 = 24190$, $\Sigma Y^2 = 9634$

$$\hat{\beta} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} = \frac{6(15033) - (378)(240)}{6(24190) - (378)^2} = 6.2313$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{X} = \left(\frac{240}{6}\right) - (0.2313)\left(\frac{378}{6}\right) = 25.4281$$

$$\hat{y} = \hat{\alpha} + \hat{\beta}X = 25.4281 + 6.2313X$$

If $X = 88$, find Y?

$\hat{y}(88) = 25.428 + (0.2313)(88)$.

Find Se & $S_{\hat{\beta}}$?

$$Se = \sqrt{\frac{\Sigma Y^2 - \hat{\beta}\Sigma XY}{n-2}} = \sqrt{\frac{(9634) - (2313)(15033)}{6-2}} = 39.2328$$

$$S_{\hat{\beta}} = Se \Big/ \sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{n}} = 39.2328 \Big/ \sqrt{24190 - (378)^2/6} = 2.0232$$

Q. Find the simple Correlation Coefficient between X & Y based on the following data.

X : 13 , 19 , 13 , 18 , 14 , 17 , 14 , 17 , 15 , 16

Y : 15 , 22 , 13 , 20 , 13 , 20 , 15 , 19 , 15 , 18

sol. $n = 10$

X	Y	XY	X^2	Y^2	
13	15	195	169	225	
19	22	418	361	484	
13	13	169	169	169	
18	20	360	324	400	
14	13	182	196	169	
17	20	340	289	400	
14	15	210	196	225	
17	19	323	289	361	
15	15	225	225	225	
16	18	288	256	324	
156	170	2710	2474	2982	Σ

$$r_{x,y} = \frac{(10)(2710) - (156)(170)}{\sqrt{[(10)(2474) - (156)^2][(10)(2982) - (170)^2]}} = ?$$

$\in [-1, 1]$

Simple Linear Regression

Q: If the Correlation Coefficient between X and Y is (0.87), and $\bar{X} = 59.9$, $\bar{Y} = 66.8$, $\delta_x = 14.7$ and $\delta_y = 7$. Then find the equation of linear regression.

مثال (2) اذا كان معامل الارتباط بين درجات الثانوية ودرجات الجامعة هو $r = 0.8$ والوسط الحسابي لدرجات الثانوية هو 80 والانحراف المعياري لها هو 10 ، بينما كان الوسط الحسابي لدرجات الجامعة هو 70 والانحراف المعياري هو 8 .

① كوّن معادلة خط الانحدار للتنبؤ بدرجات الجامعة اذا علمت درجات الثانوية

② ارسم خط الانحدار .

③ اذا حصل طالب على درجة (90) في الثانوية ، ماهي الدرجة المتوقعة له في الجامعة

④ اذا كانت الدرجة الحقيقية لذلك الطالب هي 80 في الجامعة كم هو مقدار الخطأ في التنبؤ ؟ (بينه على خط الانحدار) .

⑤ كوّن معادلة خط الانحدار للتنبؤ بدرجات الثانوية اذا علمت درجات الجامعة

ن = 10

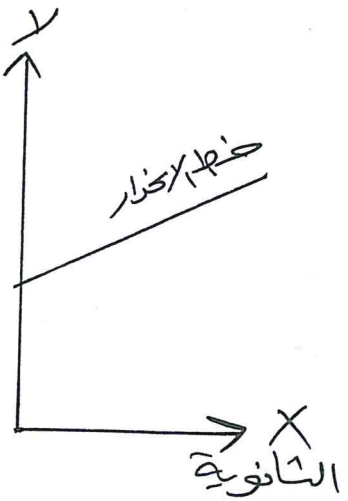
المتغير	الوسط Mean	الانحراف المعياري (S.d.)
درجات الثانوية (X)	$\bar{X} = 80$	$\sigma_x = 10$
درجات الجامعة (Y)	$\bar{Y} = 70$	$\sigma_y = 8$

② $\hat{y} = \hat{\alpha} + \hat{\beta} X$ عندما $(Y|X)$

$$\hat{\beta} = \frac{\sigma_y}{\sigma_x} \cdot r_{x,y} = \frac{8}{10}(0.8) = 0.64$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$
$$= 70 - (0.64)(80) = 18.8$$

$$\hat{y} = 18.8 + 0.64 X$$



③ $\hat{y} = 18.8 + (0.64)(90) = 76.4$

④ $e_i = Y_i - \hat{Y}_i$
$= 80 - 76.4 = 3.6$

⑤ H.W $(X|Y)$ عندما

عباس نجم سلمان
استاذ مساعد

ملخص

(Chapter one)

Descriptive Statistics (الإحصاء الوصفي)

١- البيانات غير المبوبة : وتشمل

أ- مقاييس النزعة المركزية

ب- مقاييس التشتت والاختلاف .

أ- مقاييس النزعة المركزية :

- الوسط
- الوسيط
- المنوال

① الوسط الحسابي (Mean)

هو المعدل (مجموع المفردات على عددها) .

$$\bar{X} = \sum_{i=1}^{n} x_i / n \quad \text{الوسط للعينة}$$

$$M = \sum_{i=1}^{N} x_i / N \quad \text{الوسط للمجتمع}$$

(حيث ؛ N حجم المجتمع ، n حجم العينة)

- الوسط الموزون (Weighted mean)

If w_i is weight, then; the weighted mean is;

$$\bar{X}_w = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i$$

② الوسيط : (Median)

هو القيمة التي تحتل المرتبة الوسطى عند ترتيب القيم تصاعدياً أو تنازلياً

أ- عدد القيم فردي ← القيمة الوسطى هي الوسيط

ب- عدد القيم زوجي ← معدل القيمتين الوسطيتين هي الوسيط .

ملخص

② المنوال (Mode)

هو القيمة التي تتكرر أكثر من غيرها من القيم الظاهرة، وهناك منوال رئيسي ومنوال ثانوي.

- ما العلاقة بين الوسط الوسيط والمنوال؟

الوسط − (٣) الوسيط − المنوال
<

ب - مقاييس التشتت والاختلاف

المدى
التباين
الانحراف المعياري
معامل الاختلاف

① المدى: (Range)

(هو الفرق بين أكبر قيمة وأقل قيمة)
(وهو مقياس كثير العيوب)

$$R = L - U$$

Range ← R
أكبر قيمة ← L
أقل قيمة ← U

② التباين والانحراف المعياري Variance and Standard Deviation

التباين للمجتمع

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

التباين للعينة

$$S^2 = s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

أما الانحراف المعياري للمجتمع والعينة فهو الجذر التربيعي لتباين العينة والمجتمع ويرمز له بـ σ و S : $\sigma = \sqrt{\sigma^2}$, $S = \sqrt{s^2}$

σ ← انحراف
S ← انحراف

٣) معامل الاختلاف (Coefficient of Variation)

هو مقياس للمقارنة بين عينتين من حيث الاختلاف في درجة التشتت ويرمز له بـ C.V. :

$$C.V. = \frac{S}{\bar{X}} \cdot (100)$$

حيث أن :

$\bar{X}$ ≡ الوسط الحسابي للعينة

S ≡ الانحراف المعياري للعينة

والأفضلية (لغرض المقارنة بين عينتين) تكون للعينة ذات معامل اختلاف أقل .

---

٢ - البيانات المبوبة Group Data

أ - مقاييس النزعة المركزية .

ب - مقاييس التشتت والاختلاف .

أ - مقاييس النزعة المركزية :

١) الوسط الحسابي (Mean)

$$\bar{X} = \frac{\sum_{i=1}^{k} X_i f_i}{\sum_{i=1}^{k} f_i}$$

X_i = مركز الفئة i

= (الحد الأدنى + الحد الأعلى) / ٢

f_i = تكرار الفئة i

k = عدد الفئات

والفئات متساوية بالطول

٢) الوسيط : (Median)

لإيجاد الوسيط نتبع الخطوات الآتية :

١) ترتيب الفئات تصاعدياً .

٢) إيجاد التكرار المتجمع الصاعد .

٣) إيجاد رتبة الوسيط R :

$$R = \frac{\sum_{i=1}^{k} f_i + 1}{2} = \frac{n+1}{2}$$

٤) تحديد الفئة الوسيطية من خلال مقارنة رتبة الوسيط (موقعها) مع أول تكرار متجمع صاعد .

٥) نستخدم الصيغة الآتية :

$$M = L_e + \frac{\frac{\sum f_{(i+1)}}{2} - G_1}{G_2} \times W_e$$

حيث ان:

L_e = الحد الادنى للفئة الوسيطية،

W_e = طول الفئة الوسيطية.

G_1 = التكرار المتجمع الصاعد للفئة قبل الوسيطية.

G_2 = التكرار للفئة الوسيطية.

② المنوال: (Mode) لايجاد المنوال:

① نحدد الفئة المنوالية وهي الفئة التي تقابل اكبر تكرار.

② نستخدم القانون الآتي:

$$M_o = L_o + \frac{D_1}{D_1 + D_2} \times W_o$$

حيث ان:

L_o = الحد الادنى للفئة المنوالية.

W_o = طول الفئة المنوالية

D_1 = الفرق بين تكرار الفئة المنوالية وتكرار الفئة قبل المنوالية

D_2 = الفرق بين تكرار الفئة المنوالية وتكرار الفئة بعد المنوالية.

ب- مقاييس التشتت والاختلاف:

① المدى (Range)

$$R = X_L - X_S$$

X_S = الحد الادنى للفئة الاولى

X_L = الحد الاعلى للفئة الاخيرة

② الانحراف المعياري والتباين:

التباين $= Var(X) = S^2$

الانحراف $= \sqrt{S^2} = S$

$$S^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\left(\sum_{i=1}^{k} f_i - 1\right)} - \frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{\sum_{i=1}^{k} f_i \left(\sum f_i - 1\right)}$$

Group (A)

Q1-Determine the **relationship** between degrees of two students A and B, based on the following information:-

A: good , accepted , fail , good , v.good , good .

B: accepted , v.good , accepted , intermediate , good , fail .

Q2-Based on the following data , find the ***Range, variance, Standard Deviation***

X: 2 , 3 , 2 , 4 , 3 .

Group (B)

Q1- If the correlation coefficient between X and Y is (0.87) and also $\bar{x}=59.9, \sigma_x=14.7$, $\bar{y}=66.8$ and $\sigma_y=7$. Estimate **the equation of linear regression of y | x** . Also guess the value of y when **x = 3.**

Q2-Based on the following data , find the **weighted mean** ($\bar{x}_w$)

Degree(x_i) : **62 , 80 , 90 , 75 , 82 .**

Weight (w_i): **4 , 3 , 2 , 5 , 3 .**

Group (C)

Q1-Determine the **relationship** between X and Y:

X: 3 , 2 , 4 , 11, 9

Y: 5 ,7 , 12 , 22 , 8

Q2-Based on the following data , find the **Mean, Median and Mode**:

X: 22 , 18 , 23 , 20 , 25 , 22 , 20 .

Group (D)

Q1- From the following grouped data, find the **Mean, Median and Mode**:

Classes	2-4	4-6	6-8	8-10
Frequency	2	5	4	3

***Q2-Write the general form of linear regression equation** of y/x when $\alpha=4$, and $\beta=-3$, and write this equation when $X=1$. Also, write the canonical form of α and β.*

Q1/ A: good, accepted, fail, good, V.good, good.
B: accepted, V.good, accepted, intermediate, good, fail.

fail (1)
accepted (2)
intermediate (3)
good (4)
V.good (5)

$n = 6$

A	B	D_i	D_i^2
good (4)	accepted (2)	2	4
accepted (2)	V.good (5)	-3	9
fail (1)	accepted (2)	-1	1
good (4)	intermediate (3)	1	1
V.good (5)	good (4)	1	1
good (4)	fail (1)	3	9

$\sum D_i^2 = 25$

معامل الارتباط (العلاقة بين A, B)

$$r_{x,y} = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(25)}{6(6^2-1)} \in [-1, 1]$$

$$= 1 - \frac{25}{35} = 1 - \frac{5}{7} = \frac{2}{7}$$

العلاقة طردية ضعيفة

$\simeq 0.28$

Q_2: X : 2, 3, 2, 4, 3.

Range $R = 4 - 2 = 2$

Variance $S^2 = \frac{\sum^n (x_i - \bar{x})^2}{n-1}$

or $S^2 = \frac{\sum x_i^2}{n-1} - \frac{(\sum x_i)^2}{n(n-1)}$

x_i	x_i^2
2	4
3	9
2	4
4	16
3	9
$\sum x_i = 14$	$\sum x_i^2 = 42$

$n = 5$ عدد البيانات

$S^2 = \frac{42}{4} - \frac{(14)^2}{5(4)}$

$= 10.5 - \frac{196}{20}$

$= 10.5 - 9.8$

$= 0.7$

$\therefore S = \sqrt{0.7} = \sqrt{0.7} \simeq 0.83667$

Q1: $r_{x,y} = 0.87$, $\bar{X} = 59.9$, $\sigma_x = 14.7$, $\bar{Y} = 66.8$

$\sigma_y = 7$.

$$\hat{\beta}_{y|x} = \frac{\sigma_y}{\sigma_x} * r_{x,y}$$

$$= \frac{7}{14.7}(0.87)$$

$$\simeq 0.4146$$

$$\hat{\alpha}_{x|x} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 66.8 - (0.4146)(59.9)$$

$$\simeq 66.8 - 24.8345$$

$$\simeq 41.96546$$

$$\hat{y} = \hat{\alpha} + \hat{\beta}X$$

معادلة خط الانحدار

$$\hat{y} = (41.96546) + (0.4146)X$$

$$\hat{y}(X=3) = (41.96546) + (0.4146)(3)$$

$$= (41.96546) + 1.2438$$

$$\simeq 43.20926$$

Q_2: Degree (X_i): 62 80 90 75 82

Weight (w_i): 4 3 2 5 3

$$\bar{X}_w = \frac{\sum_{i=1}^{5} X_i w_i}{\sum_{i=1}^{5} w_i}$$

$$= \frac{(62)(4)+(80)(3)+(90)(2)+(75)(5)+(82)(3)}{4+3+2+5+3}$$

$$= \frac{246+240+180+375+246}{17}$$

$$= \frac{1287}{17}$$

$$\simeq 75.7059$$

3rd Stage/Group (C) NIGHT STUDIES CHAPTER(1)	Probability Exam.	2012-2013

Group(C)

Q1/ Determine the relationship between x and y :-

	x	y	xy	x^2	y^2
	3	5	15	9	25
	2	7	21	4	49
	4	12	48	16	144
	11	22	242	121	484
n=5	9	8	72	81	64
	$\Sigma x = 29$	$\Sigma y = 54$	$\Sigma xy = 398$	$\Sigma x^2 = 231$	$\Sigma y^2 = 766$

$$r_{x,y} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}}$$

$$= \frac{(5)\times(398) - (29)(54)}{\sqrt{\left[5\times 231 - (29)^2\right]\left[5\times 766 - (54)^2\right]}}$$

$$= \frac{1990 - 1566}{\sqrt{[1155 - 841][3830 - 2916]}}$$

$$= \frac{424}{\sqrt{314 * 914}}$$

$$= \frac{424}{535.72} \simeq 0.7915$$

علاقة طردية قوية

3rd Stage/Group ( ) NIGHT STUDIES CHAPTER( )	Probability Exam.	2012-2013

Group(C)

Q_2 - $X: 22, 18, 23, 20, 25, 22, 20$

$$Mean(\bar{X}) = \frac{\sum x_i}{n} = \frac{22+18+23+20+25+22+20}{7}$$

$$= 21.43$$

Median (Me): ترتيب القيم تصاعدياً

① ② ③ [4]

18, 20, 20, 22, 22, 23, 25

$R = \frac{7+1}{2} = 4$ رتبة الوسيط

$\therefore Me = 22$

Mode (Mo):

حالات متوالين

$Mo = 22$ رئيسي وهو

$M_1 = 20$ ثانوي وهو

3rd Stage/Group (D) NIGHT STUDIES CHAPTER(1)	Probability Exam.	2012-2013

Group (D1

Q1 /

class	f_i	x_i	$x_i f_i$	Cf_i
2-4	2	3	6	2
4-6	5	5	25	7
6-8	4	7	28	11
8-10	3	9	27	(14) = n
			$\Sigma x_i f_i = 86$	

$\sum_{i=1}^{4} f_i = 14$

$n = \sum f_i = 14$

$$\bar{X} = \frac{\sum_{i=1}^{K} x_i f_i}{\sum_{i=1}^{K} f_i} \quad \text{(Mean)}$$

$$= \frac{86}{14} \simeq 6.1429$$

Median (Me)

$$Me = Le + \frac{R - Cf_{i-1}}{f_i} \times We$$

$$R = \frac{\sum f_i + 1}{2} = \frac{14+1}{2} = 7.5$$

الرتبة الوسطية = (6-8)

$$Me = 6 + \frac{7.5 - 7}{4} \times (2)$$

$$= 6 + \frac{1}{4} \simeq \frac{25}{4} \simeq 6.25 \in (6-8)$$

Mode (Mo)

الفئة الحاوية / (4-6) تقابل تكرار كبير

$$Mo = Lo + \frac{D_1}{D_1 + D_2} \times Wo$$

$$= 4 + \frac{(5-2)}{(5-2)+(5-4)} \times (2) = \frac{11}{2} = 5.5 \in (4-6)$$

3rd Stage/Group ( ) MORNING STUDIES CHAPTER( )	Probability Exam. Group (D)	2012-2013

Q2: The general form of linear regression equation of y/x :

$\hat{y} = \hat{\alpha} + \hat{\beta}X$, $\alpha = 4$, $\beta = -3$

$= 4 + (-3)X$

$= 4 - 3(1)$ if $X = 1$

$\hat{y} = 1$ when $X = 1$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{X}$

$\hat{\beta}_{y/x} = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$

or $= \dfrac{\sigma_y}{\sigma_x} * r_{x,y}$

Grouped Data / Variance (δ^2)

$$\delta^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i - 1} - \frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{\sum_{i=1}^{k} f_i \left(\sum_{i=1}^{k} f_i - 1\right)}$$

K = no. of classes عدد الفئات

f_i = frequency تكرار الفئة

x_i = midpoint مركز الفئة

$n = \sum_{i=1}^{k} f_i$ ($\sum_{i=1}^{k} f_i$ = عدد التكرارات) حجم العينة (n)

Un-Grouped Data / Variance (δ^2)

$$\delta^2 = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

← نلاحظ تطابق قانون التباين في البيانات المبوبة والغير مبوبة.

٣) معامل الاختلاف : Coefficient of Variation

يستعمل عند مقارنة عينتين من حيث الافضلية للبحث والدراسة، حيث يتم استخراج معامل الاختلاف (C.V) لكل عينة وتفضل العينة ذات معامل الاختلاف الاقل وحسب الصيغة التالية :

$$C.V = \frac{S}{\bar{X}} \times 100$$

S ← الانحراف المعياري للعينة
$\bar{X}$ ← الوسط الحسابي

مثال :

اذا كانت البيانات الاتية تمثل متوسط الدخل الشهري والانحراف المعياري للعمال وعاملات احد المصانع

العمال : $\bar{X} = 45$ دينار و $S = 4.5$ دينار
العاملات : $\bar{X} = 40$ دينار و $S = 4.3$ دينار

الحل :

معامل الاختلاف للعمال $C.V = \frac{4.5}{45} \times 100 = 10\ \%$

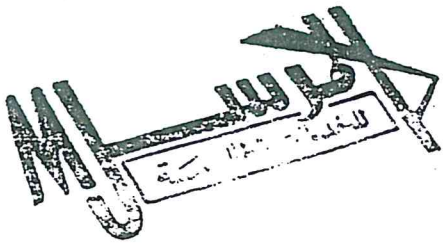
معامل الاختلاف للعاملات $C.V = \frac{4.3}{40} \times 100 = 10.75\ \%$

نجد ان معامل اختلاف الدخل الشهري للعمال اقل منه في حالة العاملات .

Theorem (1): Given a dist. x_i , $i = 1, 2, \ldots, y$

Then $\sum_{i=1}^{n} (x_i - 7) = 0$

Proof :- H.W



Theorem (2): $S^2 = \frac{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)}$

Proof :- $S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$

H.W

٢) البيانات المبوبة Group Data
مقاييس النزعة المركزية

١- الوسط الحسابي: صيغة الوسط الحسابي المستخدم في البيانات المبوبة هي كما يأتي:-

$$\bar{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \qquad \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث x_i = يمثل مركز الفئة i
f_i = " تكرار " i
k = " عدد الفئات

مثال: جد الوسط الحسابي للبيانات التالية

الفئات Classes	Frequency (f_i)	x_i (Midpoint)	$x_i f_i$
4-8	3	6	18
8-12	2	10	20
12-16	0	14	0
16-20	5	18	90

$\Sigma f_i = 10$ $\Sigma x_i f_i$ 128 ج.

$$\bar{X} = \frac{\sum_{i=1}^{4} x_i f_i}{\sum_{i=1}^{4} f_i} = \frac{128}{10} = 12.8$$

Note: اذا كانت الفئات غير متساوية بالطول نستخدم الوسيط.

٢- الوسيط: لايجاد الوسيط نتبع الخطوات التالية

① ترتيب الفئات تصاعدياً، ثم ايجاد التكرار المجمع الصاعد
② ايجاد رتبة الوسيط بالصيغة

$$R = \frac{\Sigma f_{i+1}}{2} = \frac{n+1}{2}$$

③ تحديد الفئة الوسيطية من خلال مقارنة رتبة الوسيط وموقعه في اول تكرار مجمع صاعد.

④ نستخدم الصيغة التالية في ايجاد الوسيط

$$M = Le + \frac{\frac{\Sigma f_{i+1}}{2} - G_1}{G_2} \times We$$

حيث ان:
Le هو الحد الادنى للفئة الوسيطية
G_1 هو التكرار المجمع الصاعد للفئة ماقبل الوسيطية
G_2 تكرار الفئة الوسيطية
We طول الفئة الوسيطية.

مثال:

الفئة	التكرار	التكرار المجتمع الصاعد
٢٥ - ٣٥	٢	٢
٣٥ - ٤٥	٥	٧
٤٥ - ٥٥	١٢	١٩
٥٥ - ٦٥	١٣	٣٢
٦٥ - ٧٥	٢٠	٥٢ الفئة الوسيطية
٧٥ - ٨٥	١٥	٦٧
٨٥ - ٩٥	٨	٧٥
٩٥ - ١٠٥	٥	٨٠
	٨٠	

ترتيب الوسيطة = $\frac{٨٠ + ١}{٢} = ٤٠,٥$

نلاحظ ان او تكرار مجتمع صاعد كبير $٤٠,٥$ هو ذلك التكرار الذي يقابل الفئة ٦٥-٧٥

$$\therefore M = 65 + \frac{1/2\,(80+1) - 32}{20} \times 10$$

$$= 65 + \frac{8.5}{2} = 65 + 4.25 = 69.25$$

٣- المنوال: لايجاد المنوال يجب تحديد الفئة المنوالية وهي الفئة التي تقابل أكبر تكرار ونستخدم الطريقة التالية:

$$M_0 = L_0 + \frac{D_1}{D_1 + D_2} \cdot W_0$$

حيث ان:

L_0 هو الحد الادنى للفئة المنوالية

D_1 هو الفرق بين تكرار الفئة المنوالية وتكرار الفئة قبل المنوالية

W_0 طول الفئة المنوالية.

* فضي المثال السابق:

$$M_0 = 65 + \frac{(20-13)}{(20-13)+(20-15)} \times 10$$

$$= ?$$

مقاييس التشتت:

$$R = X_L - X_S$$

① المدى: حيث X_L يمثل الحد الادنى للفئة الاخيرة

X_S " " " " " الاولى

② الانحراف المعياري والتباين عدد البيانات

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i - 1} - \frac{(\sum_{i=1}^{k} f_i x)^2}{\sum f_i (\sum f_i - 1)}}$$

$$Var(x) = S^2$$

حيث f_i تمثل التكرار في المقابل الى مركز الفئة x_i $\quad$ k عدد الفئات

مثال: للبيانات التالية، جد التباين والانحراف المعياري

الفئة	التكرار f_i	مركز الفئة x_i	$f_i x_i$	$f_i x_i^2$
15-5	3	10	30	300
25-15	5	20	100	2000
35-25	8	30	240	7200
45-35	2	40	80	3200
	18		450	12700

$$Var(x) = \left[\frac{12700}{17} - \frac{(450)^2}{18(17)}\right] = 85.3$$

$$S = \sqrt{85.3} = 92$$

اطوال الفئات — Class Intervals

هي مجموعة من الفترات المنوعة التي تغطي كل ارقام التوزيع ولكن

K: number of intervals

$K = 1 + 3.322 \log(n)$; n عدد العناصر في التوزيع

R: The range of dist.

$R = x_L - x_S$; $\quad \max(x_i) = x_L$ $\quad$ $\min(x_i) = x_S$

w: Width of an intervals

w: The number of elements in an interval

$$\boxed{w = R/K}$$

l is the lower limete of interval

u " " upper " " "

d " difference between any two Consecutive numbers.

l ($\quad$) u

$$\boxed{u = l + (w - d)}$$

Notes:

① اذا كانت x_i مجموعة من الارقام العشرية $\quad d = 1$

" " " " " " x_i " " $\quad d = .1$ (عدد عشري واحد)

" " " " " " " " $\quad d = 0.01$ (عددين عشريين)

* العينة / تقدر معالم المجتمع μ و σ^2 (وظيفتها)
العينات يتم تحديدها باختيارها بطريقة عشوائية (لها نفس فرصة الاختيار) بطريقة عشوائية اي غير متحيزة وعالية التمثيل، وغير مكلفة واعتبارية (معنى العشوائية)

* من مميزات الوسيط / تقسم البيانات الى قسمين (متساوية او غير متساوية)
* كما كانت الانحرافات قليلة كما كانت البيانات قريبة من بعضها كان العينة مناسبة ومفضلة).

* مثال الجدول الآتي يمثل اوزان خمسين طفل مصابين بمرض فقر الدم، جد الانحراف المعياري والتباين والمدى.

وزن الطفل Classes	التكرار f_i	مركز الفئة x_i	x_i^2	$x_i f_i$	$x_i^2 f_i$
12-14	20	13	169	260	3380
14-16	18	15	225	270	4050
16-18	12	17	289	204	3468
$\sum (\sum f_i)$	50			734	10898

$$S = 1.58 \quad , \quad S^2 = V(x) = (1.58)^2 \simeq 2.5$$
$$R = 18 - 12 = 6$$

* مثال / معامل اختلاف (Coefficient of Variation)
اذا كانت البيانات الآتية تمثل متوسط الدخل الشهري والانحراف المعياري لعمال وعاملات احد المصانع:

العمال	$\bar{X} = 45$ دينار	$S = 4.5$ دينار
العاملات	$\bar{X} = 40$ دينار	$S = 4.3$ دينار

$$C.V \text{ للعمال} = \frac{4.5}{45} \times 100 = 10\%$$
$$C.V \text{ للعاملات} = \frac{4.3}{40} \times 100 = 10.75\%$$

معامل اختلاف الدخل الشهري للعمال اقل منه في حالة العاملات (الافضل).