وزارة التعليم العالي والبحث والعلمي
جامعة ديالى
كلية تربية المقداد / قسم الرياضيات

محاضرات مادة

الإحصاء والاحتمالية

للعام الدراسي (2023-2024)

المرحلة الثالثة

Chapter one

مدرس المادة: م.م هند إبراهيم محمد

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 .

وبشكل عام ان هذا العلم يعتبر جمع لفرعين رئيسيين 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 منازل 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 ، يرمز 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 ضوء 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 وتكثيف 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
أ- بيانات مبوبة Group Data
ب- بيانات غير مبوبة Ungroup Data

البيانات غير المبوبة (Ungroup Data)

أ- مقاييس النزعة المركزية (Measure of Central Tendency)
ب- مقاييس التشتت والاختلاف (Measurement of Variations)

أ- مقاييس النزعة المركزية :

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

① الوسط الحسابي (المعدل Average) : (Arithmetic Mean)

الوسط الحسابي للعينة : $\overline{X}$

لتكن n حجم العينة (عدد مفردات العينة)
$X_1 , X_2 , \ldots , X_n$ مفردات العينة

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

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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

الوسط الحسابي للمجتمع μ

لتكن $X_1, X_2, \dots, X_N$ مفردات المجتمع وان N حجم المجتمع (عدد مفردات المجتمع) ويعتبر μ ثابت (معلمة) لان عدد مفردات المجتمع ثابت لا يتغير.

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

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

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
$i = 1, 2, \dots, n$

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 :

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

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.

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

٢) الوسيط (Median)

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

امثلة

١) وسيط مجموعة الارقام ١٠، ١٠ ، ٨و٨ ، ٦ ، 5و4 ، 4و3 ، 3 هو 6 .

٢) وسيط مجموعة الارقام ١٨، ١٥، ١٢، ١١و٩و٧و5و5 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 الوسيطتين (اذا كان عدد المفردات n زوجياً)

مثال ١) اذا كانت البيانات الآتية تمثل الاطوال 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 ٢) 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 الوسيطتين (المرتبة الرابعة والخامسة)

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

---

٣) 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 واخرى 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 : 16 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 .

$$i.e.\quad M_e = M_o = \mu$$

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

الوسط الحسابي = (٣ (الوسيط) − المنوال) / ٢

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

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).

$$R = L - U$$

أكبر قيمة ↗ اقل قيمة ↗

امثلة

① جد مدى البيانات الآتية:

45 ، 50 ، 55 ، 60 ، 65

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

② مدى البيانات الآتية (الاوزان):

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 ، فنكوّن مقياس جديد يدعى التباين (Variance) الذي رمزه (S^2) ويُقرأ (سكماسكوير) 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

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

الانحراف للعينة $S = +\sqrt{S^2}$

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
$$\mu = \frac{35+34+40+38+37+32}{6} = \frac{216}{6} = 36 \text{ كغم}$$

(تباين المجتمع) ويعني تباين كل قيمة عن الوسط الحسابي μ 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 العينة)

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

$$S^2 = \sum_{i=1}^{6} (x_i - 36)^2 / 6 \quad \text{(تباين العينة)}$$
$$= \frac{42}{5} = 8.4$$

البيانات المبوبة (Group Data)

١- مقاييس النزعة المركزية ٢- مقاييس التشتت والاختلاف .

١- مقاييس النزعة المركزية :

① الوسط الحسابي :

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

x_i = يمثل مركز الفئة i

f_i = يمثل تكرار الفئة i

k = يمثل عدد الفئات .

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 الادنى للفئة الوسيطية

W_e = طول الفئة الوسيطية

G_1 = التكرار المتجمع الصاعد للفئة قبل الوسيطية

G_2 = تكرار الفئة الوسيطية.

③ المنوال: لايجاد المنوال يجب تحديد الفئة المنوالية اولاً وهي الفئة التي تقابل اكبر تكرار ونستخدم الطريقة الآتية:

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

حيث ان:

L_o : الحد الادنى 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 الاعلى للفئة الاخيرة

X_S = الحد الادنى للفئة الاولى

② الانحراف المعياري والتباين:

ليكن:

f_i = التكرار المقابل لمركز الفئة x_i

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

K = عدد الفئات

$Var(X) = S^2 =$ التباين

$S = \sqrt{S^2} =$ الانحراف المعياري

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

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 دائماً $n = \sum f_i = 10$

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

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

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

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

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

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

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

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

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

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

$\sum f_i = 10$
$\sum f_i - 1 = 9$
$\sum x_i f_i = 48$
$\sum x_i^2 f_i = 205$
$(\sum x_i f_i)^2 = (48)^2 = 2304$

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

$$-1 \le r \le +1$$

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

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

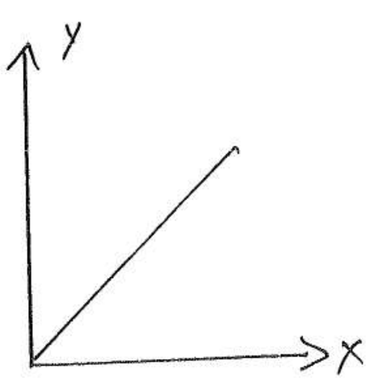
حيث ان n تمثل حجم العينة = عدد البيانات في كل عمود (عدد بيانات X = عدد بيانات Y)

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

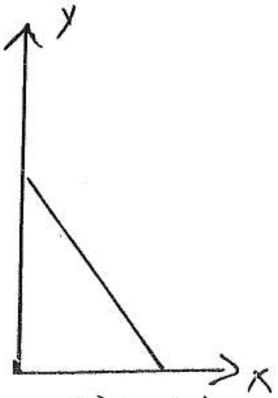
If $r_{x,y} \in [-1,0) \longrightarrow$ معامل ارتباط ضعيف

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

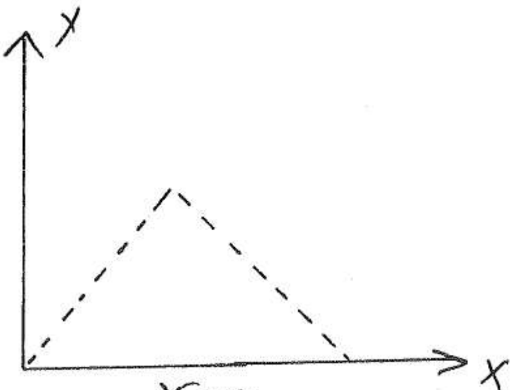
If $r_{x,y} = 0 \longrightarrow$ لا يوجد ارتباط بين X و Y



معامل ارتباط طردي



معامل ارتباط عكسي



$r=0$
لا توجد علاقة بين X و Y

Note: $r_{x,y} = r_{y,x}$

اما معامل الارتباط البسيط للمجتمع :

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

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

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

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

$\therefore$ 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 :

$$r = 1 - \frac{6\sum D_i^2}{N(N^2-1)}$$ (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)

$\sum D_i^2 = 17$

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

$$= 0.51 \in (0,1]$$ (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

$N = 5$ ، $\sum D_i^2 = 20$

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

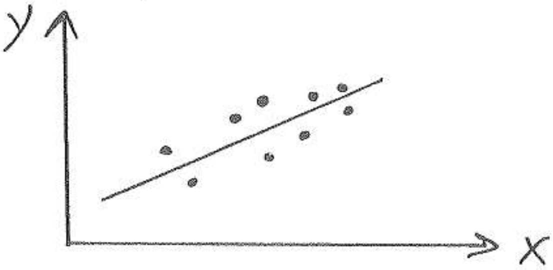
* ما نوع العلاقة بين A و B؟

الانحدار Regression

شكل الانتشار (Scatter diagram):

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 او ظل الزاوية التي يصنعها خط الانحدار مع المحور X.

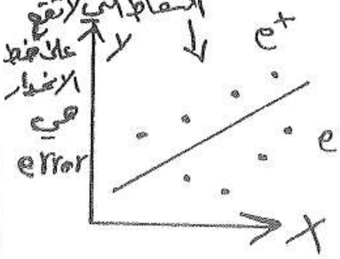
$$\text{i.e. } \beta = \tan\theta \Rightarrow \theta = \tan^{-1}\beta$$

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

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

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

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 $\bar{X} \equiv$ mean of x & $\bar{Y} \equiv$ mean of y فان:

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

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

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 عقار معين 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

٢) لايجاد معادلة خط الانحدار التقديرية :

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

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

$$\hat{\alpha} = \frac{102}{8} - 4.67\left(\frac{11}{8}\right) = 6.33$$

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

عندما $(y \mid X)$

③ لايجاد $\hat{y}(3)$ ، i.e. $X = 3$

$$\hat{y}(3) = 6.33 + 4.7(3) = 20.34$$

Chapter one
Descriptive Statistics

Ungrouped Data (البيانات غير المبوبة)

(1) جد الوسيط للارقام التالية:

3 , 4 , 4 , 5 , 6 , 8 , 8 , 8 , 10 ووسيطها هو (6) .

نرتب الاعداد تصاعدياً (تنازلياً): 10 , 9 , 8 , 11 , 12 , 16 , 9

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

الوسيط هو (10) الذي يحتل المرتبة الرابعة.

لايجاد الوسيط نرتب الاعداد تصاعدياً: 10 , 14 , 7 , 5 , 20 , 7 , 22 , 12

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

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

الوسيط $Me = (12+10)/2 = 11$

---

(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 ووسطهما 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) من مائة رقم (20) أربعة و(40) خمسة و(30) ستة والباقي كانوا سبعات.
أوجد الوسط الحسابي لهذه الارقام.

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

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

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

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 وبعدهم (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 و 70 وان عدد الساعات (ساعات المحاضرات) الاسبوعية لهذه المقررات هي 3 و 5 و 3 و 1 . أوجد متوسط الدرجات .

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

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

---

(٧) اوجد الوسط والمنوال والوسيط للارقام التالية :

$3, 5, 2, 6, 5, 9, 5, 2, 8, 6$

نرتب تصاعدياً :

$X: 2, 2, 3, 5, \underline{5, 5}, 6, 6, 8, 9 \quad (n=10)$

$\mu = Mean = \sum x_i/10 = 5.1$

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

$Mo = Mode = 5$

وللارقام التالية :

$50.3, 49.5, 48.9, 51.6, 48.7$

نرتب تصاعدياً (أو تنازلياً) :

$48.7, 48.9, \underline{49.5}, 50.3, 51.6$

$\mu = \sum x_i/5 = 49.8$

$Me = 49.5$

$Mo =$ لا يوجد

---

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

$\bar{X} = \sum x_i/5 = 86$ ، $R = \frac{n+1}{2} = \frac{5+1}{2} = 3$ (الوسيط في المرتبة الثالثة)

نرتبهم ← $Me = 85$ ، $Mo =$ لا يوجد منوال

Grouped Data (البيانات المبوبة)

(١) في تجربة لتحديد تأثير عقار معين على مستوى الكوليسترول 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 الكولسترول (الفئات)	عدد المرضى (التكرار) f_i
120 – 160	2
160 – 200 ← الفئة المنوالية	5 ← اعلى تكرار
200 – 240	3
	$n=\sum_{i=1}^{3} f_i = 10$

$D_1 = 5-2 = 3$

$D_2 = 5-3 = 2$

$W_o = 200-160 = 40$

$L_o = 160$

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

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

---

(٢) في دراسة لتحديد تأثير التدخين على الاديين تم قياس هذه المادة في البلازما (Mg/Ml) وبعد ساعتين من تناول الدواء من قبل (16) من المدخنين تم الحصول على البيانات التالية والمطلوب ايجاد معدل مستوى هذه المادة (الوسيط) .

Classes / مستوى الاديين (الفئات)	عدد المدخنين (التكرارات) f_i	c-f.d. التكرار المتجمع الصاعد
0.505 – 1.005	4	4
1.005 – 1.505	3	7
الفئة الوسيطية 1.505 – 2.000	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

$W_e = 2.000 - 1.505 = 0.495$ و $L_e = 1.505$,

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

(٩) جد الوسط والوسيط والمنوال لمجموعة الارقام والاعداد التالية :

$5 , 4 , 8 , 3 , 7 , 2 , 9$

$\mu_x = \sum x_i / 7 = 5.4$

$2 , 3 , 4 , 5 , 7 , 8 , 9$ نرتب الارقام

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

$M_e = 5$ ، $M_o =$ لايوجد منوال

---

(١٠) جد المنوال للارقام 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

$M_o = 4 , 5 , 6 , 8 , 10$

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

---

(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 .

$$\bar{X} = \frac{\sum w_i x_i}{\sum w_i} = \frac{(60)(3.0) + (20)(2.0)}{30 + 20} = \frac{220.0}{50} = \text{[illegible]}$$

---

(١٢) جد منوال الارقام التالية :

$2 , 2 , 5 , 7 , 9 , 9 , 9 , 10 , 10 , 11 , 12 , 18$

$M_o = 9$

$3 , 5 , 8 , 10 , 12 , 15 , 16 , 20$

$M_o =$ لايوجد منوال (لايوجد 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.) .

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

وزن الطفل classes	التكرار (f_i) عدد الاطفال	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

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

$$Range = R = 18 - 12 = 6$$

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

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

(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)	عدد رؤوس الغنم (التكرارات)
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

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

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

$$\delta^2 = \sum_{i=1}^{6}(x_i-36)^2/6-1 = \frac{42}{6-1} = \underline{7}(\text{كغم}^2)$$

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

نفس السؤال اعلاه لكن للعينة :

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

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

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

∴ تباين العينة هو تقدير جيد لتباين المجتمع .

---

(2) جد الوسط الوسيط والمنوال للارقام المكونة من ست ستات و سبع سبعات و ثماني ثمانيات و تسع تسعات وعشر عشرات :

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

$$R = \frac{n+1}{2}, \quad n = \sum w_i = 10+9+8+7+6 = \underset{\text{عدد زوجي}}{40}$$

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

اي المرتبة 20 و 21

والوسيط هو

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

ويمكن ايجاد الوسيط في هذا السؤال بطريقة اخرى كالاتي:

$$666666 \ 7777777 \ 888888\underset{\text{الوسيط}}{\underline{88}} \ 999999999$$

$$10\,10\,10\,10\,10\,10\,10\,10\,10\,10 .$$

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

$$\therefore M_o = 10$$

معامل الارتباط البسيط Simple Correlation Coefficient

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

مثال// جد المنوال اذا كان لديك الجدول التالي :

الفئات	التكرار
20 − 24	5
25 − 29	7
30 − 34	8
35 − 39	10
40 − 44	6
45 − 49	4

الحل

$$E = 35 + \frac{10-8}{(10-8)+(10-6)} \times 5$$

$$= 36,67$$

الارتباط والانحدار

الارتباط : تقابلنا كثيراً في الحياة العملية مواقف تتضمن متغيرين (ظاهرتين) واكثر ويكون المطلوب معرفة ما اذا كان هنالك علاقة بين هذين المتغيرين وما هو شكل هذه العلاقة وايضا كيفية التنبأ بأحد هذين المتغيرين في حالة معرفتنا في المتغير الاخر.

فكثيراً مانجد في بعض الحالات معادلة الطول مع الوزن فاذا اردتي ان تعرفي الوزن المثالي ادخلي طولكي في المعادلة ليظهر وزنك المثالي، فقد توصلوا الى هذه المعادلة او الى هذه الصيغة بدراسة العلاقة مابين المتغيرين الطول والوزن على مجموعة من الافراد اذن ومما سبق يمكن تعريف الارتباط بالتعريف التالي :

الارتباط: هو تعيين طبيعة وقوة العلاقة بين متغيرين او عدمها.

انواع الارتباط:

① الارتباط الموجب (الطردي): هو العلاقة بين المتغيرين (x,y) اذا تغير احد المتغيرين فإن الآخر يتبعه في نفس الاتجاه.

② الارتباط السالب (العكسي): وهو العلاقة بين المتغيرين (x,y) حيث اذا تغير احد المتغيرين فإن الاخر يتبعه في الاتجاه المضاد.

قياس الارتباط: تستخدم معاملات الارتباط لقياس درجة الارتباط بين المتغيرين الظاهرين.

معامل الارتباط: يعرف معامل الارتباط والذي يرمز له بالرمز ٢ لانه عبارة عن مقياس رقمي يقيس قوه الارتباط بين المتغيرين حيث تتراوح قيمة $\boxed{-1 \leqslant ٢ \leqslant +1}$.

ملاحظه// تدل اشارة المعامل الموجبة على العلاقه الطردية بينما تدل اشارة المعامل السالب على العلاقه العكسية.

ملاحظة // الجدول التالي يوضح انواع الارتباط واتجاه العلاقة وشكل الانتشار لكل نوع .

قيمة معامل الارتباط	المعنى
1 +	ارتباط طردي تام
من 0.70 الى 0.99	ارتباط طردي قوي
من 0.50 الى 0.69	ارتباط طردي متوسط
من 0.01 الى 0.49	ارتباط طردي ضعيف
0	لا يوجد ارتباط

وما قيل عن الارتباط الطردي ينطبق على الارتباط العكسي مع وضع الاشارة السالبة وتغيير كلمة الطردي الى عكسي .

انواع مقاييس الارتباط

يمكن حساب العديد من معاملات الارتباط وتعتمد ذلك على مستوى القياس (اسمي ، ترتيبي ، فترة ، نسبي) المتغيرات التي تبدو مرتبطة وهنالك عدة انواع من معاملات الارتباط وسنتناول اهم هذه الانواع :-

١- معامل بيرسون للارتباط الخطي : هو اكثر معاملات الارتباط استعمالاً خاصة في العلوم الانسانية والاجتماعية ومستوى قياس المطلوب عند تطبيق معامل بيرسون للارتباط هو ان يكون

الموضوع :   التاريخ   /   /   ٢٠١

كل المتغيرين مقياس فترة او نسبي بمعنى اخر ان تكون بيانات كلا المتغيرين (الظاهرتين) بيانات كمية

- حساب معامل بيرسون للارتباط الخطي

يمكن حساب معاملا بيرسون بدلالة القراءات للبيانات المتغيرين y, x باستخدام الصيغة التالية :

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

حيث ان :

$\sum xy$ مجموع حاصل ضرب $x\,y$

$\sum x$ مجموع قيم المتغير x

$\sum y$ مجموع قيم المتغير y

$\sum y^2$ مجموع مربعات قيم المتغير y

$\sum x^2$ مجموع مربعات قيم المتغير x

مثال // سجلت ست قراءات n كفرضية لحجم الانتاج وحجم صادرات النفط الخام ((بالمليار برميل)) خلال عدة سنوات. حدد الارتباط الخطي بين حجم الانتاج وحجم الصادرات للنفط وما مدى العلاقة بينهما؟