الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري يتم إتباع الخطوات التالية

. تكوين الجدول التكراري المتجمع الصاعد

. تحديد رتبة الوسيط $(\frac{n}{2})=(\frac{\sum f}{2})$

. تحديد فئة الوسيط كما الشكل التالي :

تكرار متجمع صاعد سابق f_1 الحد الأدنى لفئة الوسيط (A)

رتبة الوسيط $(n/2)$ ⟵ الوسيط Med

تكرار متجمع صاعد لاحق f_2 الحد الأعلى لفئة الوسيط

. ويحسب الوسيط بتطبيق المعادلة

$$\boldsymbol{Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L}$$

حيث أن

L هي طول فئة الوسيط وتحسب بالمعادلة التالية

طول الفئة = الحد الأعلى – الحد الأدنى

L=Upper - Lower

مثال

فيما يلي توزيع عجل متوسط الحجم حسب احتياجاته اليومية من الغذاء الجاف بالكيلوجرام

فئات الاحتياجات اليومية	1.5-	4.5	7.5	10.5	13.5 – 16.5
عدد العجول f	4	12	19	10	5

خطوات الحل

رتبة الوسيط $25 = \frac{50}{2} = \frac{\sum f}{2} = \frac{n}{2}$

الجدول التكراري المتجمع الصاعد

تكرار متجمع صاعد	أقل من
0	1.5
4	4.5
f_1 16	A 7.5الحد الأدنى
25 رتبة الوسيط	Med
f_2 35	10.5 الحد الأعلى
45	13.5
50	16.5

بتطبيق معادلة الوسيط على المثال

$$\boldsymbol{Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L}$$

$$f_1 = 16 \qquad f_2 = 35 \qquad L = 10.5 - 7.5 = 3 \qquad A = 7.5$$

$$\boldsymbol{Med = 7.5 + \frac{\frac{50}{2} - 16}{35 - 16} \times 3 = 8.921\ k.g}$$

مزايا وعيوب الوسيط

مزايا

1- لا يتأثر بالقيم الشاذة أو المتطرفة

2- سهل في الحساب

3- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى أي أن

$$\sum |x - Med| \leq \sum |x - a| \, , a \neq Med$$

العيوب

1- لا يأخذ عند حسابة كل القيم في الاعتبار فهو يعتمد على قيمة أو قيمتين فقط

٢- يصعب حسابة في حالة البيانات الوصفية المقاسة بمعيار اسمي nominal

المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعاً أو تكراراً ويكثر استخدامه في حالة البيانات الوصفية لمعرفة النمط (المستوى)

- حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mod) = القيمة الأكثر تكراراً

- حساب المنوال في حالة البيانات المبوبة (طريقة الفروق)

$$Mod = A + \frac{d_1}{d_1 + d_2} * L$$

A الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار)

d_1 الفرق الأول =(تكرار فئة المنوال – تكرار سابق) تكرار سابق

تكرار فئة المنوال

d_2 الفرق الثاني =(تكرار فئة المنوال – تكرار لاحق) تكرار لاحق

L طول فئة المنوال

فئة المنوال= الفئة المناظرة لأكبر تكرار

مثال

اختبرت عينات عشوائية من طلاب بعض أقسام كلية علوم الأغذية والزراعة وتم رصد درجات هؤلاء الطلاب في مقرر إحصاء وكانت النتائج كالتالي

قسم وقاية النباتات	80	77	75	77	77	77	65	70	58	67
قسم علوم الأغذية	88	68	60	75	93	65	77	85	95	90
قسم الاقتصاد	80	65	69	80	65	88	76	65	86	80
قسم الإنتاج الحيواني	85	73	69	85	73	69	69	73	72	85

الحل

هذه البيانات غير مبوبة لذا فإن المنوال = القيمة الأكثر تكراراً

والجدول التالي يبين منوال الدرجة لكل قسم من الأقسام

القيمة المنوالية	القيمة الأكثر تكرار	القسم
المنوال= 77 درجة	الدرجة 77 تكررت 4 مرات	قسم وقاية النباتات
لا يوجد منوال	جميع القيم ليس لها تكرار	قسم علوم الأغذية
يوجد منوالان هما المنوال الأول = 65 المنوال الثاني = 80	الدرجة 65تكررت 3 مرات الدرجة 80 تكررت 3 مرات	قسم الاقتصاد
يوجد ثلاث منوال هما المنوال الأول = 69 المنوال الثاني = 73 المنوال الثالث = 85	الدرجة69 تكررت 3 مرات الدرجة 73 تكررت 3 مرات الدرجة 85 تكررت 3 مرات	قسم الإنتاج الحيواني

مثال

في الجدول التالي توزيع أسرة حسب الإنفاق الاستهلاكي الشهري لها بالألف ريال والمطلوب حساب منوال الإنفاق الشهري

فئات الإنفاق	2-	5-	8-	11-	14-17
عدد الأسر f	4	7	10	5	4

الحل

تحديد الفئة المنوالية

الفئة المنوالية هي الفئة المناظرة لأكبر تكرار (8-11)

التكرارات	الفئات
4	2-
7	5-
10	8- فئة المنوال A
5	11-
4	14-17

أكبرتكرار

d1= 10 – 3=3

d2= 10 – 5=5

L=3 طول فئة المنوال

الحد الأدنى لفئة المنوالية (A=8)

$$Mod = A + \frac{d_1}{d_1 + d_2} * L$$

$$Mod = 8 + \frac{3}{3+5} * 3$$

8+1.125=9.125

مقاييس التشتت

هي مقاييس إحصائية تستخدم لقياس مدى تباين أو تشتت البيانات عن بعضها البعض وهي تشمل

١- المدى

الفرق بين أكبر قيمة وأصغر قيمة في البيانات .

٢- التباين

متوسط مربع الانحرافات عن المتوسط الحسابي.

٣- الانحراف المعياري

الجذر التربيعي للتباين .

تستخدم مقاييس التشتت لقياس مدى تجانس البيانات أو مدى انتشار البيانات حول مقاييس النزعة المركزية سوف نذكر مثال يبين ذلك

إذا كان لدينا مجموعتين من الطلاب وكان درجات المجموعتين كالتالي

المجموعة الأولى	63	70	78	81	85	67	88
المجموعة الثانية	73	78	77	78	75	74	77

لو حسبنا الوسط الحسابي لكل مجموعة نجد أن الوسط الحسابي لكل منها يساوي 76 درجة ومع ذلك درجات المجموعة الثانية أكثر تجانساً من درجات المجموعة الأولى

المدى Rang

أبسط مقاييس التشتت يعطي فكرة عن مدى تغير الظاهرة

البيانات غير المبوبة

المدى = أكبر قيمة – أقل قيمة

Rang=Max – Min

البيانات المبوبة

توجد طريقتين

المدى في حالة البيانات المبوبة =مركز الفئة الأخيرة – مركز الفئة الأولى

المدى في حالة البيانات المبوبة =الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

مثال

لدينا مجموعة درجات طلاب مستوى ثاني تقنيات معلومات في مقرر اساسيات إحصاء

درجات الطلاب	12	15	18	20	22	25	30

الحل

أكبر قيمة =30

أصغر قيمة = 12

المدى = أكبر قيمة – أصغر قيمة

Rang=30 – 12 = 18

مثال

لدينا مجموعة درجات طلاب مستوى ثاني تقنيات معلومات في مقرر اساسيات إحصاء

درجات الطالب	60 - 69	70 - 79	80 - 89	90 - 99
التكرار	2	5	8	5

الحل

حساب مركز الفئة الأولى

مركز الفئة الأولى =(2)\(60+69)= 64.5

حساب مركز الفئة الثانية

مركز الفئة الثانية =(2)\(90+99)= 94.5

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

Rang= 94.5 – 64.5 =30

التباين Variance

هو أحد مقاييس التشتت وأكثرها استخداماً في النواحي التطبيقية ويعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي

التباين في المجتمع (σ^2)

إذا توافر لدينا قراءات عن كل مفردات المجتمع ولتكن $x_1, x_2, \ldots . x_N$ فأن التباين في المجتمع ويرمز له بالرمز σ^2 سيجما يحسب باستخدام المعادلة التالية

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

حيث μ هو الوسط الحسابي في المجتمع أي أن $\mu = {}^{\Sigma X}/_{N}$

مثال

مصنع لتعبئة المواد الغذائية يعمل به 15 عامل وكانت عدد سنوات الخبرة لهؤلاء العمال كما يلي بفرض أن هذه البيانات تم جمعها عن كل مفردات المجتمع فأوجد التباين لعدد سنوات الخبرة

5	13	7	14	12	9	6	8	10	13	14	6	11	12	10

$(x - \mu)^2$	$(x - \mu)$	سنوات الخبرة X
25	5 – 10 = -5	5
9	3	13
9	-3	7
16	4	14
4	2	12
1	-1	9
16	-4	6
4	-2	8
0	0	10
9	3	13
16	4	14
16	-4	6
1	1	11
4	2	12
0	0	10
130	0	150

الحل

الوسط الحسابي في المجتمع μ

$$\mu = \frac{1}{N}\sum X$$

$$\mu = \frac{1}{15}(5 + 13 + 7 + 14 + 12 + 9 + 6 + 8 + 10 + 13 + 14 + 6 + 11 + 12 + 10)$$

$$= \frac{1}{15}(150) = 10$$

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N} = \frac{130}{15} = 8.67$$

التباين في العينة (S^2)

في كثير من الحالات يكون تباين المجتمع S^2 غير معلوم وعندئذ يتم سحب عينة من هذا المجتمع ويحسب من بيانات العينة كتقدير لتباين المجتمع فإذا كانت قراءات عينة عشوائية حجمها n هي $x_1, x_2, \ldots . x_n$ فإن تباين العينة ويرمز له بالرمز S^2

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

حيث أن $\bar{x}$ هو الوسط الحسابي لقراءات العينة أي أن وتباين العينة هو التقدير غير المتحيز لتباين المجتمع

مثال

إذا تم سحب عينة من عمال المصنع حجمها 5 عمال وسجل عدد سنوات الخبرة وكانت كالتالي أحسب تباين سنوات الخبرة في العينة

9	5	10	13	8

الحل

الوسط الحسابي في العينة

$$\bar{x} = \frac{1}{n}\sum x = \frac{1}{5}(8 + 13 + 10 + 5 + 9) = \frac{1}{5}(45) = 9$$

سنوات الخبرة X	$(x - \bar{x})$	$(x - \bar{x}^2)$
8	-1	1
13	4	16
10	1	1
5	-4	16
9	0	0
45	0	34

إذا تباين سنوات الخبرة في العينة قيمته هو

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

$$S^2 = \frac{34}{5 - 1} = \frac{34}{4} = 8.5$$

التباين للبيانات المبوبة

$$\sigma^2 = \frac{\sum(x - \bar{x})^2 f}{n - 1}$$

حيث أن f هو تكرار الفئة ، x هو مركز الفئة ، $\bar{x}$ هو الوسط الحسابي ، n مجموع التكرارات

مثال

اذا كانت درجات مجموعة من طلاب مستوى ثاني تقنيات معلومات في اختبار النصفي لمقرر إحصاء

الدرجات	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30
التكرار	2	0	4	3	1

$(x-\bar{x})^2 \times f$	$(x-\bar{x})^2$	$(x-\bar{x})$	$x \times f$	مركز الفئة x	التكرار f	الفئة
220.5	110.25	-10.5	15	7.5	2	5-
0	30.5	-5.5	0	12.5	0	10-
1	0.25	-0.5	70	17.5	4	15-
60.75	20.25	4.5	67.5	22.5	3	20-
90.25	90.25	9.5	27.5	27.5	1	25-30
372.5			180		10	

$$\bar{x} = \frac{\sum x \times f}{\sum f} = \frac{180}{10} = 18$$

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{n - 1}$$

$$\sigma^2 = \frac{372.5}{9} = 41.38$$

الانحراف المعياري

هو الجذر التربيعي للتباين

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

الارتباط والانحدار الخطي البسيط

تعرفنا فيما سبق على بعض أساليب إحصائية والتي يمكن استخدامها لتحليل متغير واحد مثل الدخل الشهرية وغيرها من امثلة فرأينا كيف نجمع البيانات ونلخصها ونحسب لها مقاييس النزعة المركزية والتشتت وهنا سوف نتعرف على أساليب تحليل جديدة نستطيع باستخدامها تحليل متغير من خلال علاقته بمتغير آخر أو بعدة متغيرات أخرى والأمثلة على هذه الظواهر أو المتغيرات التي يوجد بينها علاقة كثيرة مثل عدد ساعات الدراسة ودرجات الاختبار وأيضا بين استخدام التكنولوجيا في الفصول الدراسية ومستوى التحصيل الأكاديمي وغيرها من أمثلة

لدراسة العلاقة بين متغيرين نستخدم أسلوب تحليل الارتباط وعندما نريد دراسة أثر أحد المتغيرين على الاخر نستخدم أسلوب تحليل الانحدار.

تقسم العلاقات بين الظواهر الى قسمين

1- علاقات تابعية

عندما يكون مقابل كل قيمة للمتغير المستقل قيمة للمتغير التابع مثل العلاقة بين مساحة الدائرة بنصف القطر

2- علاقات ارتباطية

عندما يكون مقابل كل قيمة للمتغير المستقل قيمة تقريبية أو احتمالية للمتغير التابع ومن امثلته العلاقة بين جودة البيانات ودقة نماذج التعلم الآلي العلاقة بين استخدام تقنيات الأمان وتحسين أمان الأنظمة العلاقة بين عدد المستخدمين وسرعة الشبكة

الارتباط

ظاهرة إحصائية تطلق على العلاقة بين متغيرين .

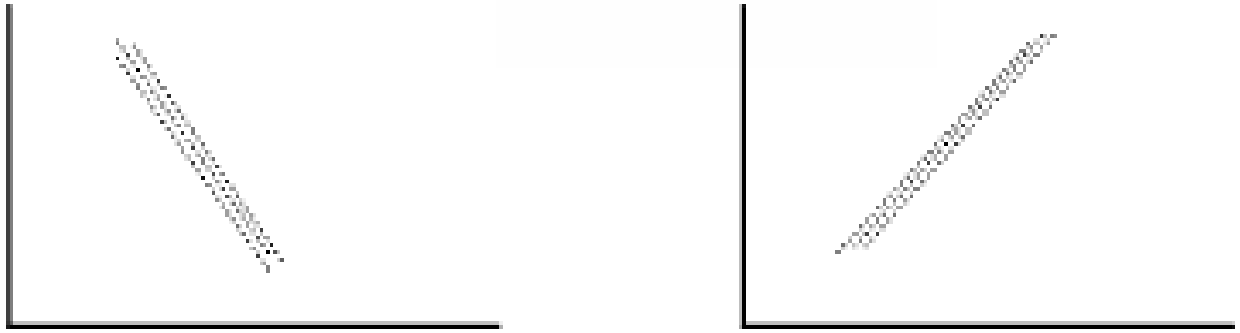
معامل الارتباط (Coefficient of Correlation)

هو مقياس لدرجة قوة العلاقة الخطية واتجاهها بين متغيرين .

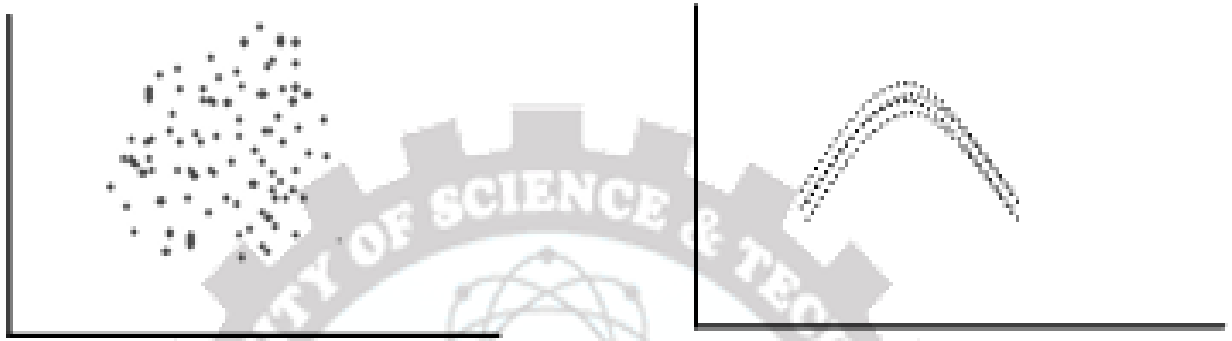
مقياس إحصائي يستخدم لبيان نوع العلاقة بين المتغيرات سواء كانت هذه العلاقة طردية أو عكسية ويأخذ معامل الارتباط قيمة من القيم المحصورة بين $(1 - \text{و} 1+)$ فمعامل ارتباط () بين متغيرين يعبر عن علاقة موجبة تامة (طردية) بين المتغيرين أما معامل ارتباط () فيعبر عن علاقة تامة ولكنها سالبة (عكسية) بين المتغيرين ومعاملات الارتباط بين أزواج من المتغيرات والتي تتراوح قيمها بين $(1 - ,1+)$ تعبر عن علاقة غير تامة فمعامل ارتباط (± 0.90) يعبر عن علاقة قوية بين المتغيرات ومعامل ارتباط (± 0.50) يعبر عن علاقة متوسطة أما معامل ارتباط (± 0.20) فيعبر عن علاقة ضعيفة بين المتغيرين

شكل الانتشار (Scatter Diagram)

إذا اردنا تحديد قوة واتجاه العلاقة بين متغير وآخر فإننا نحتاج لعرض شكل الانتشار لتمثيل العلاقة بين ظاهرتين باستخدام الرسم البياني (للمتغير المستقل X والذي يمثل المحور الأفقي والمتغير التابع Y والذي يمثل المحور العمودي) وتمثل كل ثنائية (زوج) من القيم (x,y) بنقطة في المستوى الإحداثي فنحصل على شكل الانتشار والذي يتوقف على نوع العلاقة بين المتغيرين



(a) ارتباط خطي موجب قوي



(b)ارتباط خطي سالب قوي

(c) ارتباط غير خطي موجب قوي (علاقة غير خطية)

(d)لا يوجد ارتباط

من خلال الأشكال السابقة ومن قراءة هذه الأشكال يمكن تحديد طبيعة ونوع ومتانة العلاقة بين المتغيرات فإذا كانت نقاط الانتشار تقع في مجال متقارب وعلى امتداد خط مستقيم واحد من أسفل إلى أعلى فأننا نستنتج وجود علاقة خطية طردية قوية بين المتغير x والمتغير yكما في الشكل (a) أما إذا كان شكل الانتشار في حدود خط مستقيم يتجه من أعلى إلى أسفل وبدرجة تركيز عالية حول الخط فإن هذا يدل على وجود علاقة خطية عكسية قوية بين المتغيرين x,y كما في الشكل (b)أما إذا ظهر شكل الانتشار تركيزاً للنقاط ولكن حول منحنى وليس خط مستقيماً فإننا نستنتج من ذلك عدم وجود علاقة خطية بين المتغيرين x ,y كما في الشكل (c)أما إذا كان شكل الانتشار أكثر تشتتاً للنقط بحيث يتعذر معه أن تقع في شكل خط مستقيم أو منحنى فإن ذلك يدل على عدم وجود علاقة ارتباطية بين المتغيرين x ,y كما في الشكل (d)

الأشكال السابقة هي عبارة عن أمثلة لأشكال الانتشار التي تبين مدى قوة أو ضعف الارتباط بين المتغيرين x ,y أي أنها تعطي فكرة واضحة فيما إذا كانت yتزداد بزيادة قيمة x أو تنقص بزيادة x أو أنها لا تتأثر

حساب معامل الارتباط بين متغيرين

يمكن حساب معامل الارتباط بعدة طرق ونذكر منها أكثر شيوعاً

١- معامل ارتباط بيرسون(Pearson)

يستخدم عندما يكون كلا المتغيرين كمي ويقاس بمقياس فئوي مثل قياس العلاقة بين الوزن والطول والعلاقة بين الدرجة التي حصل عليها الطالب وعدد ساعات المحاضرة وغيرها من الأمثلة .

٢- معامل ارتباط سبيرمان (Spearman)

يسمى أيضاً معامل ارتباط الرتب ويستخدم عندما يكون المتغيرين مقاسين بمقياس ترتيبي مثل إيجاد الارتباط بين متغير مستوى الدخل (عالي ، متوسط ، منخفض) ومتغير تأييد رأي معين (أوافق ، متردد ، غير موافق) .

إشارة معامل الارتباط

1- إذا كان$r > 0$ فإن الارتباط يكون إيجابي (أي توجد علاقة طردي بين المتغيرين) بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني

2- إذا كان $r < 0$ فإن الارتباط يكون سلبي (أي توجد علاقة عكسية بين المتغيرين) بمعنى أن زيادة أحد المتغير يصاحبه انخفاض في المتغير الثاني

3- إذا كان $r = 0$ فإنه لا توجد ارتباط بمعنى انعدام العلاقة بين المتغيرين

قوة العلاقة

يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (∓ 1) حيث أن قيمة معامل الارتباط تقع في المدى $(-1 < r < 1)$

0.9-1	علاقة قوية جداً
0.7 – 0.89	علاقة قوية
0.5 – 0.69	علاقة متوسطة
0.3 – 0.49	علاقة ضعيفة
0 – 0.29	علاقة غير موجودة أو ضعيفة جداً
(-0.3) –(-0.49)	علاقة ضعيفة سلبية
(-0.5) –(-0.69)	علاقة متوسطة سلبية
(-0.7) –(-0.89)	علاقة قوية سلبية
(-0.9) –(-1)	علاقة قوية جداً سلبية

لحساب معامل الارتباط باستخدام صيغة بيرسون

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

S_{xy} التغاير بين المتغيرين

S_x ، S?y الانحراف المعياري

الجدول التالي يوضح درجات 10 طلاب في الاختبار الأول والاختبار النهائي

الاختبار الأول	80	70	90	85	75	95	80	70	85	90
الاختبار النهائي	85	75	95	90	80	98	85	75	90	95

هل يوجد علاقة بين درجات الطلاب في الاختبار الأول والاختبار النهائي

الحل

أولاً نوجد المتوسط الحسابي للمتغيرين

$$\bar{x} = 83 \quad \text{و} \quad \bar{y} = 88$$

$(y-\bar{y})^2$	$(x-\bar{x})^2$	$(y-\bar{y})(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})$	y	x
9	9	9	-3	-3	85	80
169	169	169	-13	-13	75	70
2401	49	49	49	7	95	90
16	4	8	4	2	90	85
64	64	64	-8	-8	80	75
100	144	120	10	12	98	95
9	9	9	-3	-3	85	80
169	169	169	-13	-13	75	70
16	4	4	4	2	90	85
49	49	49	7	7	95	90
610	610	646				sum

$r = \frac{646}{\sqrt{610}\sqrt{610}} = 1.05$ يوجد علاقة إيجابية قوية جداً بين درجات الطلاب في الاختبار الأول والاختبار النهائي

لحساب معامل الارتباط باستخدام صيغة اسبيرمان

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

d الفرق بين رتب مستويات المتغير الأول ورتب مستويات المتغير الثاني

مثال

لنفترض أنه لدينا المعلومات التالية عن متوسط عدد الأطفال في الأسرة والمستوى التعليمي للأم المطلوب تقدير متانة ونوع العلاقة الارتباطية بين المستوى التعليمي للأم ومتوسط عدد الأطفال في الأسرة

المستوى التعليمي للأم	أمية	تقرأ وتكتب	ابتدائية	إعدادية	ثانوية	جامعية	دكتورا
متوسط عدد الأطفال	10	8	9	7	5	4	2

المستوى التعليمي للأم	متوسط عدد الأطفال	رتب المستوى التعليمي للأم	رتب متوسط عدد الأطفال	D	D^2
إعدادية	7	4	4	0	0
ابتدائية	9	3	6	-3	9
ثانوية	5	5	3	2	4
أميه	10	1	7	-6	36
تقرأ وتكتب	8	2	5	-3	9
جامعية	4	6	2	4	16
دكتوراه	2	7	1	6	36
Sum					110

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(110)}{7(49 - 1)} = 1 - 1.964 = -0.96$$

هذا يدل على وجود علاقة ارتباط متينة بين المستوى التعليمي للأم ومتوسط عدد الأطفال في الأسرة وهي علاقة عكسية قوية جداً وهذا يدل على أنه كلما زاد المستوى التعليمي للأم كلما قل متوسط عدد الأطفال في الأسرة

ارتباط الصفات (Correlation of Attributes)

أوضحاً سابقاً بأن معامل ارتباط (بيرسون) يقيس قوة الارتباط في حالة البيانات الكمية سواء كانت مبوبة أو غير مبوبة وكذلك معامل ارتباط الرتب(لسبيرمان)

والذي يستخدم لحساب قوة الارتباط في حالة البيانات الكمية والوصفية التي لها صفة الترتيب إلا أنه قد توجد بيانات وصفية لها صفات مميزة ولا نستطيع ترتيبها مثل الحالة الاجتماعية (متزوج ، أعزب ، مطلق ، أرمل) وكذلك لون العينين (زرقاء ، سوداء ، بنية) ولقياس قوة الارتباط لهذه البيانات نشأت الحاجة إلى إيجاد مقاييس تقيس الارتباط بين هذه الصفات وسنذكر منها

معامل الاقتران (Coefficient of Association)

يستخدم لقياس قوة الارتباط عندما يكون لدينا متغيرين نوعيين ولكل متغير صفتين فقط أي يكون لدينا جدول اقتران 2 * 2 كأن تدرس العلاقة بين الإصابة بالسرطان والتدخين لمجموعة أشخاص مدخنين وغير مدخنين فالمتغيرين هنا الأول المرض (مصاب ، غير مصاب) والثاني التدخين (مدخن ، غير مدخن) أو ندرس قياس قوة العلاقة بين التدخين والتعليم فإننا نجد أن التعليم ينقسم إلى (متعلم ، غير متعلم) والتدخين (مدخن ، غير مدخن) ولا يستخدم معامل الاقتران عندما يكون كل من المتغيرين أو أحدهما له أكثر من صفتين ومعامل الاقتران كمعامل الارتباط قيمته تكون محصورة بين الصفر والواحد الصحيح

ولحساب معامل الاقتران تقسم المفردات حسب المتغيرات والصفات كما في الجدول

المتغير A / المتغير B	القيمة (1)	القيمة (2)	المجموع
القيمة (1)	n_{11}	n_{12}	n_1
القيمة (2)	n_{21}	n_{22}	n_2
المجموع	n_1	n_2	N

حيث $N = n_{11} + n_{12} + n_{21} + n_{22}$

ويحسب معامل الاقتران من العلاقة

$$r = \frac{n_{11} * n_{22} - n_{21} * n_{12}}{n_{11} * n_{22} + n_{21} * n_{12}}$$

ويكون هناك علاقة قوية بين المتغيرات كلما اقتربت قيمة معامل الاقتران من الواحد وبشكل عام نقول أن هناك علاقة عندما تكون قيمة rمن اكبر

$r > 0.5$

مثال

قام احد الباحثين بدراسة العلاقة بين التدخين والتهاب اللثة فجمع بيانات عن 200 شخص فكان توزيعهم حسب الصفات كما يلي:

A / B	مدخن	غيرمدخن	المجموع
مصاب	70	10	80
غير مصاب	20	100	120
المجموع	90	110	200

$$r = \frac{n_{11} * n_{22} - n_{21} * n_{12}}{n_{11} * n_{22} + n_{21} * n_{12}}$$

$$r = \frac{70 * 100 - 20 * 10}{70 * 100 + 20 * 10} = \frac{6800}{7200} = 0.944$$

هناك علاقة قوية جداً بين التدخين ومرض التهاب اللثة وتجدر الإشارة إلى أنه ليس هناك أي مدلول لإشارة معامل الاقتران وذلك لأن إشارة تختلف حسب ترتيب تكرار الصفات في الجدول لذلك تؤخذ قيمته المطلقة للدلالة على وجود العلاقة من عدمها وقيمة هذا المعامل تكون دائماً أقل من الواحد الصحيح مادامت $n_{21} * n_{12} > 0$ ولا تساوي الواحد إلا إذا كانت $n_{21} * n_{12} = 0$

2.7.2 معامل التوافق (Coefficient of Contingency) :

نعلم من دراستنا السابقة أن معامل الاقتران يستخدم لقياس متانة العلاقة بين متغيرين نوعيين لكل منهما قيمتين (أن الجدول يحتوي على أربع خانات) ولا نستطيع استخدامه إذا كان لأحد المتغيرين على الأقل أكثر من صفتين. وفي هذه الحالة يكون الجدول مكون من أكثر من أربع خلايا فنلجأ إلى استخدام مقياس آخر هو معامل التوافق ويرمز له بالرمز r_c ولحساب معامل التوافق نفرض أن لدينا ظاهرة X لها r من الصفات والظاهرة Y لها s من الصفات ونوضح ذلك في الجدول (8.3) :

جدول(8.3)

Y \ X	x_1	x_2	.....	x_r	المجموع
y_1	f_{11}	f_{12}	.....	f_{1r}	$f_{1.}$
y_2	f_{21}	f_{22}	.....	f_{2r}	$f_{2.}$
. .	. .	. .	. .	. .	. .
y_s	f_{s1}	f_{s2}	.....	f_{rs}	$f_{s.}$
المجموع	$f_{.1}$.	$f_{.2}$	.....	$f_{.r}$	$f_{..}$

ويحسب معامل التوافق r_c بالعلاقة الآتية:

$$r_c = \sqrt{\frac{G-1}{G}}$$

حيث :

$$G = \frac{(f_{11})^2}{f_{.1} \times f_{1.}} + \frac{(f_{12})^2}{f_{.2} \times f_{1.}} + \dots + \frac{(f_{rs})^2}{f_{s} \times f_{r.}}$$

مثال (8.3):

من بيانات الجدول التالي أوجد معامل التوافق بين متغيري الحالة الاجتماعية والتدخين .

الحالة الاجتماعية / التدخين	a	b	c	المجموع
يدخن	30	80	20	130
لا يدخن	25	15	30	70
المجموع	55	95	50	200

الحل : لدينا :

$$G = \frac{(30)^2}{(55)(130)} + \frac{(80)^2}{(95)(130)} + \frac{(20)^2}{(50)(130)} + \frac{(25)^2}{(55)(70)} + \frac{(15)^2}{(95)(70)} + \frac{(30)^2}{(50)\ (70)}$$

$$= 1.15946$$

وبتطبيق الصيغة

$$r_c = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{0.15946}{1.15946}} = 0.37$$

التي تشير إلى علاقة متوسطة القوة بين التدخين والحالة الاجتماعية للشخص مع التنويه إلى أن زيادة عدد الأعمدة والصفوف من شأنها أن تزيد في ارتفاع معامل التوافق عموماً ، لذلك فإن أصغر قيمة لـ r_c هي الصفر وأكبر قيمة تعتمد على عدد الصفوف و عدد الأعمدة ولكنها في كل الأحوال لا تتجاوز الواحد .

2.3 الانحدار الخطي البسيط

تحليل الانحدار عبارة عن وسيلة إحصائية تستخدم لتحديد نوع العلاقة بين متغير مستقل واحد أو أكثر ومتغير تابع واحد ، وعند وجود علاقة خطية بسيطة بين متغير تابع Y ومتغير مستقل X يستخدم عادة نموذج الانحدار الخطي الآتي:

$$Y_i = \propto + \beta X + e_i$$

حيث

Y : المتغير التابع أو الاستجابة في التجربة .

$\propto$: معامل ثابت أو الجزء المقطوع من محور Y ويصبح مساوياً لقيمة Y عندما $x = 0$

β : يدعى بميل الانحدار ويمثل مقدار التغير في Y عند زيادة المتغير X بمقدار وحده واحدة ، كما يطلق عليه بمعامل الانحدار.

X_i : المتغير المستقل في التجربة .

e : الخطأ العشوائي : وهو متغير عشوائي يفترض أنه يتبع التوزيع الطبيعي بمتوسط $E(\varepsilon_i) = 0$ وتباين $\mathrm{var}(\varepsilon_i) = \sigma^2$ أي أن: $\varepsilon_i \sim N(0, \sigma^2)$ والتباين المشترك بين $\varepsilon_i , \varepsilon_j$ يساوي صفر أي أن: $Cov(\varepsilon_i , \varepsilon_j) = 0$ لكل $i, j : i \neq j$.

وقد تم إدراج حد الخطأ العشوائي e في نموذج الانحدار لأسباب منها:

1. يوجد عنصر أساسي لا يمكن التنبؤ به هو العشوائية في الاستجابات البشرية التي يمكن التعبير عنها ضمن حد الخطأ.
2. الصياغة الناقصة حال بناء النموذج الرياضي.
3. لتلافي وجود أخطاء المشاهدات أو القياس.

ويسمى النموذج السابق بنموذج الانحدار الخطي البسيط، كما أن معادلة الانحدار تدلنا على القيمة المتوقعة لـ Y لأنها الأداة التي يمكن من خلالها الحصول على تقدير للمتغير Y ، ومن الناحية العملية لن نحصل على معادلة الانحدار وإنما تقديراً لها من عينة من البيانات الإحصائية باستخدام طريقة المربعات الصغرى.

ولتقدير معلمات الانحدار في النموذج نستخدم طريقة المربعات الصغرى وهي من أكثر الطرق استخداماً في تقدير معلمات الانحدار للمجتمع وتتلخص في جعل

مجموع مربعات الأخطاء العشوائية أقل ما يمكن وتقدر β ، $\propto$ بالمقدرات $\hat{\beta}$ ، $\hat{\propto}$ وبعد حسابهما والتعويض في النموذج السابق نحصل على المعادلة التقديرية $\hat{Y} = \hat{\propto} + \hat{\beta} X + e$ حيث $\hat{Y}$ (تقرأ Y هات) تمثل القيمة المقدرة المتنبئة للقيمة Y التي تقابل القيمة X .

للحصول على خط انحدار المتغير التابع Y على المتغير المستقل X بطريقة المربعات الصغرى، فإنه يمكن اعتبار هذا الخط تقديراً لخط انحدار المجتمع $Y = \propto + \beta X$. وهي إحدى الطرق الرياضية لحساب الميل والحد الثابت نفترض أن تكون معادلة الخط المستقيم هي:

$$Y = \propto + \beta X$$

وبأخذ المجموع للطرفين نحصل على

$$\sum Y_i = \sum (\propto + \beta X)$$

$$\sum Y_i = n \propto + \beta \sum X_i$$

وبضرب معادلة الخط المستقيم بالمتغير المستقل X وأخذ المجموع للطرفين نحصل على:

$$\sum X_i Y_i = \propto \sum X_i + \beta \sum X_i^{\,2}$$

وبحل هاتين المعادلتين نحصل على:

$$\hat{\beta} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^{\,2} - \left(\sum X_i\right)^2}$$

وهذه المعادلة يمكن أن تصاغ بصورة أخرى :

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\,\overline{X}^2}$$

وبعد حصولنا على قيمة الميل نستخرج قيمة الحد الثابت وفق العلاقة الآتي:

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

حيث:

$\overline{X}$: تمثل الوسط الحسابي لمشاهدات المتغير المستقل X_i

$\overline{Y}$: تمثل الوسط الحسابي لمشاهدات المتغير التابع Y_i

وبعد تقدير معالم النموذج β , α (قيمة الميل والمقدار الثابت) نعوض بقيمة كل منهما في معادلة الانحدار التقديرية أو دالة الانحدار $\hat{Y} = \hat{\alpha} + \hat{\beta} X + e$ فنحصل على معادلة خط انحدار المتغير التابع Y على المتغير المستقل X

مثال(9.3):

احسب معادلة خط انحدار Y على X ثم احسب قيمة y عندما X =10 إذا كانت

X	2	8	4	6	5
Y	3	7	10	4	6

الحل: نوجد الوسط الحسابي لكل من y و X

$$\overline{X} = \frac{\sum X_i}{n} = \frac{25}{5} = 5$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

وأيضاً نكون الجدول الآتي:

i	X_i	Y_i	X_iY_i	X_i^2
1	2	3	6	4
2	8	7	56	64
3	4	10	40	16
4	6	4	24	36
5	5	6	30	25
$\sum$	25	30	156	145

نوجد قيمة المقدر $\hat{\beta}$ من العلاقة الآتية:

$$\because \hat{\beta} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^{\ 2} - \left(\sum X_i\right)^2}$$

وبالتعويض من الجدول في المعادلة السابقة نحصل على ميل الانحدار $\hat{\beta}$

$$\hat{\beta}_i = \frac{5(156) - (25)(30)}{5(145) - \ (25)^2}$$

$$\hat{\beta} = \frac{780 - 750}{100} = 0.3$$

وبالتعويض أيضاً بقيمة $\hat{\beta}$ في معادلة الحد الثابت

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\ \overline{X}$$

$$\hat{\alpha} = 6 - (0.3) * 5$$

$$\hat{\alpha} = 6 - 1.5 = 4.5$$

لاحظ أن إشارة الحد الثابت موجبة وهذا يطابق ما استنتجناه سابقاً عند تناولنا شكل الانتشار وبذلك تكون معادلة خط انحدار y على x بشكلها النهائي على الصورة:

$$\hat{Y} = 0.3x + 4.5$$

وبالتعويض عندما $X = 10$ في معادلة خط انحدار Y على X نحصل على قيمة $\hat{Y}$ أي:

$$\hat{Y} = 4.5 + (0.3)(10) = 7.5$$

الانحراف المتوسط

اعداد الطالبة بثينة النجادي
استاذ المقرر
الاستاذة / هيفاء الشاوش

ما هو الانحراف المتوسط

الانحراف المتوسط هو أحد مقاييس التشتت (Measures of Dispersion) في الإحصاء. يقيس متوسط المسافة المطلقة بين كل قيمة في مجموعة البيانات ومتوسط هذه البيانات (أو وسيطها).

* إذا كان الانحراف المتوسط صغيراً، فهذا يعني أن القيم متقاربة جداً من المتوسط.

* إذا كان الانحراف المتوسط كبيراً، فهذا يعني أن القيم متباعدة ومنتشرة بشكل أكبر عن المتوسط.

لماذا نستخدمه؟

نستخدمه لفهم مدى انتشار أو تشتت البيانات. يعطينا فكرة عن مدى تجانس البيانات.

خطوات حساب الانحراف المتوسط حول المتوسط الحسابي:

1. احسب المتوسط الحسابي (Mean): اجمع كل القيم في مجموعة البيانات واقسمها على عدد القيم.

   المتوسط (μ) = (مجموع القيم) / (عدد القيم)

2. احسب الانحراف المطلق لكل قيمة عن المتوسط: اطرح المتوسط من كل قيمة في مجموعة البيانات، ثم خذ القيمة المطلقة للنتيجة (تجاهل الإشارة السالبة إذا ظهرت).

   الانحراف المطلق للقيمة (x_i) = $|x_i - \mu|$

3. اجمع كل الانحرافات المطلقة: اجمع كل القيم المطلقة التي حصلت عليها في الخطوة 2.

4. اقسم مجموع الانحرافات المطلقة على عدد القيم: النتيجة هي الانحراف المتوسط.

الانحراف المتوسط (MD) = (مجموع الانحرافات المطلقة) / (عدد القيم)

مثال عملي:

لدينا مجموعة البيانات التالية تمثل درجات 5 طلاب في اختبار:
88 ,92 ,78 ,90 ,85

لنحسب الانحراف المتوسط حول المتوسط الحسابي:

1. حساب المتوسط الحسابي (μ):

$$\mu = (85 + 90 + 78 + 92 + 88) / 5$$

$$\mu = 433 / 5$$

$$\mu = 86.6$$

2. حساب الانحراف المطلق لكل قيمة عن المتوسط:

* $|86.6 - 85| = |-1.6| = 1.6$
* $|86.6 - 90| = |3.4| = 3.4$
* $|86.6 - 78| = |-8.6| = 8.6$
* $|86.6 - 92| = |5.4| = 5.4$
* $|86.6 - 88| = |1.4| = 1.4$

3. جمع الانحرافات المطلقة:

مجموع الانحرافات المطلقة = 1.6 + 3.4 + 8.6 + 5.4 + 1.4 = 20.4

4. حساب الانحراف المتوسط:

MD = 20.4 / 5

MD = 4.08

النتيجة: الانحراف المتوسط لدرجات الطلاب هو 4.08. هذا يعني أن درجات الطلاب تبتعد عن المتوسط (86.6) بمتوسط قدره 4.08 درجة.

COEFFICIENT OF VARIATION

(معامل الاختلاف)

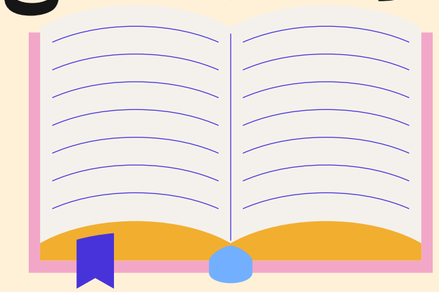
شرح الطالب

(عبد الرحمن الشجاع)

الأهداف المحاضرة :

في نهاية هذه المحاضرة يكون الطالب قادر على:

- ✓ ان يعرف استخراج قيمه الانحراف المعياري و التباين ومعامل الاختلاف من البيانات غير المبوبة بطريقة صحيحة.
- ✓ ان يعرف استخراج قيمه الانحراف المعياري و التباين و معامل الاختلاف من البيانات المبوبة بطريقة صحيحة.



COEFFICIENT OF VARIATION

(معامل الاختلاف)

- معامل الاختلاف أو ما يسمى الانحراف النسبي هو مقياس لقياس الاختلافات ولكن في صورة نسبية أي لا يأخذ وحدات قياس بينما مقاييس التشتت الأخرى وهي المدى والتباين والانحراف القياسي فهي تقيس الاختلاف بوحدات القياس أو مربعها كما في حالة التباين ، ويرمز للانحراف النسبي أو معامل الاختلاف بالرمز

(CV)

.

مقاييس التشتت (Measures Dispersion or Variation)

يقصد بالتشتت او الاختلاف بانه التباعد او التقارب الموجود بين قيم المشاهدات حول نقطه التركز (الوسط الحسابي) او المقياس الذي يقيس مدى تشتت القيم عن وسطها بمعنى كلما كبرت قيم مقايس التشتت دل ذلك على درجة كبيرة من الاختلاف بين البيانات وكلما صغرت قيم مقايس التشتت دل ذلك على درجة قليله من الاختلاف بين البيانات وهي:

- التباين
- الانحراف المعياري
- المدى
- (معامل الاختلاف)

معامل الاختلاف (coefficient of variation):

(ان معامل الاختلاف يفضل على الانحراف المعياري وغيرة من مقايس التشتت لمقارنة تشتت البيانات بين مجموعات من البيانات.)

ف معامل الاختلاف يعطي نسبة الانحراف المعياري الى الوسط الحسابي وبما ان معامل الاختلاف هو مقياس لقياس التغير النسبي على شكل نسبة مئوية لذلك معامل التغير يمكن استخدامه لمقارنة التشتت داخل عدة مجموعات من البيانات حتى وان كانت وحدات القياس لهذه المجموعة مختلفة. ويحسب من خلال القانون التالي:

معامل الاختلاف (C.V)= $\frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$ ------القانون

مثال/أجريت تجربة لدراسة طول النبات وكمية المحصول لـ (150) نباتا من محصول الذرة فكانت النتائج كالاتي:

	كمية المحصول	الطول
الوسط الحسابي	800	200
الانحراف المعياري	36	16

الحل/

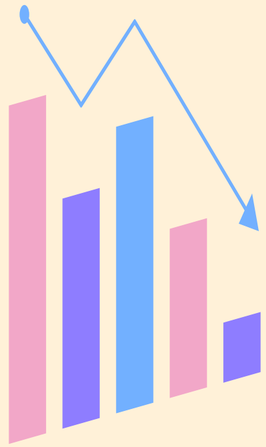
بما ان هناك مجموعتين من الصفات فان المقياس المناسب للمقارنة التشتت هو مقياس معامل الاختلاف وكما يلي:

بالنسبة لمجموعه المحصول :

معامل الاختلاف (CV)= الانحراف المعياري / الوسط الحسابي × 100

$$= \frac{36}{800} \times 100$$

$$= 4.5\ \%$$



بالنسبة لمجموعه الطول :

معامل الاختلاف (CV)= الانحراف المعياري / الوسط الحسابي × 100

$$= \frac{16}{200} \times 100$$

$$= 8\ \%$$

من خلال النتائج أعلاه ومن خلال المقارنة ما بين المجموعتين تبين بان مجموعه لها تشتت اكبر.