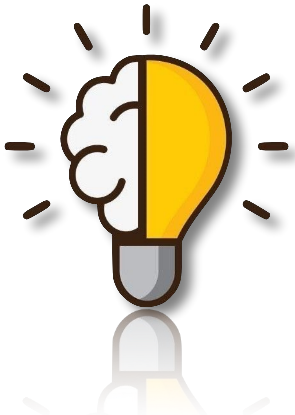
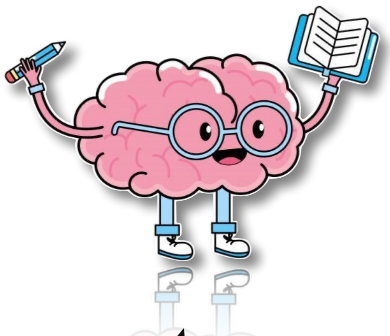
جامعة شط العرب
كلية الادارة والاقتصاد
قسم إدارة وتسويق النفط والغاز





مبادئ الاحصاء

Principles of Statistics

أعداد
م.م علياء ماجد دخيل

2025-2026

المرحلة الاولى

علم الاحصاء

علـم الاحـصاء Statistics science

الإحصاء هو فرع من فروع الرياضيات يهتم بجمع البيانات، وتحليلها، وتفسيرها، وعرضها. يُستخدم الإحصاء في مجموعة واسعة من المجالات، مثل الاقتصاد، والعلوم الاجتماعية، والطب، والهندسة، والعلوم الطبيعية، حيث يساعد في اتخاذ القرارات المستندة إلى بيانات موثوقة. تتضمن العمليات الإحصائية عدة مراحل، بدءًا من تصميم الدراسة وجمع البيانات، وصولاً إلى تحليل النتائج وتقديم التوصيات. يمكن تقسيم الإحصاء إلى نوعين رئيسيين: الإحصاء الوصفي، الذي يركز على تلخيص البيانات وتقديمها بطريقة مفهومة، والإحصاء الاستنتاجي، الذي يستخدم لتعميم النتائج من عينة إلى مجتمع أكبر.
تعتبر الإحصاءات أداة قوية لفهم الظواهر المختلفة وتحديد الأنماط والاتجاهات، مما يجعلها ضرورية في البحث العلمي وصنع السياسات.

اولاً / مفهوم علم الاحصاء

علم الإحصاء هو فرع من فروع الرياضيات الذي يركز على جمع، وتحليل، وتفسير، وعرض البيانات.
يُعتبر الإحصاء أداة مهمة في العديد من المجالات، حيث يساهم في فهم الظواهر وتقديم رؤى قيمة مبنية على أدلة.

ثانياً / مكونات علم الإحصاء

1. جمع البيانات : يتضمن تصميم الدراسات وجمع المعلومات من مصادر مختلفة، مثل الاستبيانات، والمقابلات، والسجلات.
2. تحليل البيانات : يشمل استخدام تقنيات رياضية وإحصائية لفهم البيانات واستخراج الأنماط والاتجاهات.
3. تفسير النتائج: يتم تحليل النتائج لتقديم استنتاجات واضحة ومفيدة، مما يساعد في اتخاذ القرارات المستندة إلى البيانات.
4. عرض البيانات: يتم تقديم البيانات بطريقة مرئية، مثل الرسوم البيانية والجداول، لتسهيل الفهم والتواصل.

ثالثاً / أنواع الإحصاء

1 / الإحصاء الوصفي

يركز على تلخيص البيانات باستخدام مقاييس مثل المتوسط والانحراف المعياري.

ومن الأمثلة على الإحصاء الوصفي

-متوسط الدرجات

في مدرسة، يمكن حساب متوسط درجات الطلاب في مادة معينة لتقديم نظرة عامة عن الأداء الأكاديمي.

-الانحراف المعياري

يُستخدم لتحديد مدى تشتت درجات الطلاب حول المتوسط. إذا كان الانحراف المعياري منخفضًا، فهذا يعني أن الدرجات متقاربة.

- التوزيع التكراري

في دراسة حول أعمار المشاركين في حدث معين، يمكن إنشاء جدول يوضح عدد المشاركين في كل فئة عمرية (مثل: 18-25، 26-35، إلخ).

- الرسوم البيانية

يمكن استخدام الرسوم البيانية، مثل المدرجات التكرارية أو المخططات الدائرية، لتقديم البيانات بطريقة مرئية. على سبيل المثال، يمكن عرض النسب المئوية لمختلف الفئات في دراسة سكانية.

- المقاييس الوصفية

يمكن حساب المقاييس الأساسية مثل الوسيط (العدد الذي يتوسط مجموعة من البيانات) والنمط (القيمة الأكثر تكرارًا) لفهم توزيع البيانات.

- **تحليل الاستبيانات**
عند إجراء استبيان حول رضا العملاء، يمكن تلخيص النتائج باستخدام النسب المئوية (مثل نسبة العملاء الذين كانوا راضين أو غير راضين).
- **التحليل الوصفي في البحث العلمي**
في دراسة طبية، يمكن تقديم معلومات عن عدد المشاركين وخصائصهم (مثل الجنس، والعمر، والتاريخ الطبي) دون محاولة تعميم النتائج.

2/ الإحصاء الاستنتاجي او الاستدلالي
يستخدم لاستخلاص استنتاجات من عينة معينة وتعميمها على مجتمع أكبر.

ومن الامثلة على استخدام الإحصاء الاستدلالي

- **استطلاعات الرأي**
تستخدم لاستنتاج آراء أو تفضيلات الناخبين حول قضايا معينة أو مرشحين سياسيين. على سبيل المثال، يمكن إجراء استطلاع على عينة من الناخبين لتوقع نتائج الانتخابات.
- **البحوث الطبية**
يُستخدم لتقييم فعالية علاج أو دواء جديد. يمكن إجراء تجربة سريرية على مجموعة صغيرة من المرضى وتحليل النتائج لاستنتاج فعالية العلاج على المجتمع ككل.
- **دراسات السوق**
تساعد الشركات في فهم سلوك المستهلكين. يمكن إجراء مسح على عينة من العملاء لفهم تفضيلاتهم، ثم تعميم النتائج على قاعدة العملاء الأكبر.
- **التعليم**
تُستخدم لتقييم أداء الطلاب. يمكن جمع بيانات من مجموعة من الطلاب في اختبار معين لاستنتاج مستوى الأداء الأكاديمي للطلاب في المدرسة ككل.
- **العلوم الاجتماعية**
تُستخدم لإجراء الدراسات السكانية. يمكن أخذ عينة من السكان في منطقة معينة لاستنتاج خصائص اجتماعية واقتصادية للسكان بشكل عام.
- **جودة المنتجات**
تُستخدم للتحقق من جودة الإنتاج في المصانع. يمكن أخذ عينات من المنتجات لاختبار الجودة واستنتاج مستوى الجودة في الإنتاج الكلي.

رابعاً / الفرق بين الإحصاء الوصفي والاستدلالي

1/ التعريف

-الإحصاء الوصفي
يُستخدم لتلخيص وعرض البيانات بطريقة واضحة ومباشرة. يهدف إلى تقديم معلومات وصفية حول مجموعة بيانات معينة دون محاولة تعميم النتائج على مجتمع أكبر.
- الإحصاء الاستدلالي
يُستخدم لاستخلاص استنتاجات أو تعميمات عن مجموعة سكانية أكبر بناءً على بيانات من عينة. يعتمد على أساليب رياضية لتحليل البيانات واستنتاج النتائج.

2/ الهدف

-الإحصاء الوصفي
يهدف إلى تقديم وصف شامل للبيانات، مثل المتوسطات، والانحرافات المعيارية، والرسوم البيانية.

- الإحصاء الاستدلالي

يهدف إلى استخدام البيانات من العينة لاستنتاج خصائص أو سلوك المجتمع الأكبر، مثل اختبار الفرضيات وتقدير المعلمات.

3/ البيانات المستخدمة

-الإحصاء الوصفي

يعتمد على البيانات المجمعة من العينة أو المجتمع المحدد فقط.

- الإحصاء الاستدلالي

يعتمد على البيانات المجمعة من عينة تمثل المجتمع ككل، وغالبًا ما يتطلب استخدام أساليب عشوائية لجمع البيانات.

4/ الأدوات والأساليب

- الإحصاء الوصفي

يستخدم مقاييس مثل المتوسط، الوسيط، النمط، والانحراف المعياري، بالإضافة إلى الرسوم البيانية مثل المدرجات التكرارية والمخططات الدائرية.

- الإحصاء الاستدلالي

يستخدم تقنيات مثل اختبارات الفرضيات، فترات الثقة، وتحليل الانحدار.

5/ النتائج

- الإحصاء الوصفي

يوفر معلومات دقيقة حول مجموعة البيانات المحددة، مثل مدى الانتشار أو الاتجاهات.

- الإحصاء الاستدلالي

يوفر استنتاجات وتوقعات حول المجتمع الأكبر، مع وجود مستوى معين من الثقة أو الخطأ.

خامساً / أهمية علم الإحصاء

يُعتبر علم الإحصاء أساسيًا في البحث العلمي وصنع القرار، حيث يساعد في:

- فهم العلاقات بين المتغيرات.
- تقييم فعالية البرامج والخدمات.
- توقع الاتجاهات المستقبلية.

بالتالي، يلعب الإحصاء دورًا حيويًا في تحسين المعرفة والقرارات في مختلف المجالات.

الخلاصة

بشكل عام، يعد الإحصاء الوصفي أداة قوية لفهم البيانات الحالية، بينما يساعد الإحصاء الاستدلالي في اتخاذ القرارات المستندة إلى بيانات تقديرية.

الــمـتغيـرات Variables

اولاً / مفهوم المتغيرات في الإحصاء

المتغيرات هي عناصر أساسية في علم الإحصاء، حيث تمثل الخصائص أو السمات التي يمكن قياسها أو تسجيلها. تُستخدم المتغيرات لجمع البيانات وتحليلها. إليك تفاصيل حول أنواع المتغيرات:

ثانياً /أنواع المتغيرات

1/ المتغيرات الكمية (Numerical Variables)

- تعريف
- تمثل قيمًا عددية يمكن قياسها.
- أنواع
- المتغيرات المستمرة (Continuous Variables) : يمكن أن تأخذ أي قيمة ضمن نطاق معين.
- مثال
- الوزن، الطول، والوقت.
- المتغيرات المنفصلة (Discrete Variables) : تأخذ قيمًا عددية محددة، غالبًا ما تكون أعدادًا صحيحة.
- مثال

عدد الطلاب في فصل دراسي.

2/ المتغيرات النوعية (Qualitative Variables)

- تعريف
- تمثل خصائص أو سمات غير رقمية.
- أنواع
- المتغيرات الاسمية (Nominal Variables): تُستخدم لتصنيف البيانات دون ترتيب.
- مثال
- الجنس، لون العين.
- المتغيرات الترتيبية (Ordinal Variables) : تُستخدم لتصنيف البيانات مع وجود ترتيب.
- مثال

مستويات التعليم (ثانوي، جامعي، دراسات عليا).

ثالثاً / المتغيرات المستقلة والتابعة

- المتغير المستقل (Independent Variable):

هو المتغير الذي يتم تغييره أو التحكم فيه في الدراسة. يُعتبر السبب في التغيرات التي تحدث في المتغير التابع.

- المتغير التابع (Dependent Variable):

هو المتغير الذي يتأثر بالتغيرات في المتغير المستقل. يُعتبر النتيجة أو التأثير.

رابعاً /أهمية المتغيرات في الإحصاء

- تحليل البيانات: تساعد المتغيرات في تنظيم وتحليل البيانات لفهم العلاقات والأنماط.
- تفسير النتائج: تساهم في استنتاجات دقيقة حول الظواهر المدروسة.
- تصميم الدراسات: تُستخدم لتحديد كيفية جمع البيانات وتحليلها، مما يساعد في تصميم التجارب والدراسات بشكل فعال.

خلاصة

تعتبر المتغيرات أساسًا مهمًا في الإحصاء، حيث تُستخدم لجمع البيانات وتحليلها. فهم أنواع المتغيرات واستخداماتها يمكن أن يكون له تأثير كبير على جودة البحث ونتائجه.

المجـــتــمـــع population

اولاً / مفهوم المجتمع في الإحصاء

في الإحصاء، يُشير "المجتمع" إلى مجموعة كاملة من الأفراد أو العناصر التي يتم دراستها أو تحليلها. يمكن أن يتكون المجتمع من أي عدد من العناصر، مثل الأشخاص، أو الأشياء، أو الأحداث، حسب موضوع الدراسة.

ثانياً / أهمية تعريف المجتمع

- تحديد العينة: يساعد تعريف المجتمع في تحديد كيفية اختيار العينات المناسبة للدراسة.
- تحليل البيانات: يتيح فهم النتائج وتحليلها بشكل صحيح بناءً على خصائص المجتمع.
- تعميم النتائج: يساعد في تعميم النتائج المستخلصة من العينة إلى المجتمع الأوسع.

ثالثاً / خصائص المجتمع

- الشمولية

يجب أن يتضمن المجتمع جميع العناصر ذات الصلة بالسؤال البحثي أو الهدف من الدراسة.

- التنوع

يمكن أن يكون المجتمع متنوعًا من حيث الخصائص والسمات، مما يعكس التنوع الموجود في العالم الحقيقي.

- التمثيل

يُعتبر المجتمع نقطة انطلاق لفهم الظواهر أو الاتجاهات، حيث يتم استخدام عينات تمثل المجتمع لدراسة الخصائص.

خلاصة
يُعتبر تعريف المجتمع جزءًا أساسيًا من البحث الإحصائي، حيث يشكل الأساس الذي تُبنى عليه الدراسات والأبحاث لفهم الظواهر بشكل أفضل.

الــعـيـنــة Sample

اولاً / مفهوم العينة في الاحصاء

العينة هي مجموعة فرعية تم اختيارها من المجتمع الإحصائي. تُستخدم العينة لجمع البيانات وتحليلها بهدف استنتاج خصائص أو سلوك المجتمع الأكبر دون الحاجة إلى دراسة جميع أفراده.

ثانياً / أهمية العينة

- توفير الوقت والتكاليف: دراسة عينة أصغر أقل تكلفة وأكثر فاعلية من دراسة المجتمع بأكمله.
- تسهيل التحليل: يسمح بتحليل البيانات بشكل أسهل وأسرع.
- توقع النتائج: يمكن أن تعطي نتائج العينة صورة واضحة عن المجتمع ككل.

ثالثاً / أنواع العينات

- العينة العشوائية (Random Sample)

- الوصف: يتم اختيار الأفراد بشكل عشوائي، مما يضمن أن كل فرد في المجتمع لديه فرصة متساوية للاختيار.
- الفائدة: يقلل من التحيز ويزيد من إمكانية تعميم النتائج.

- العينة الطبقية (Stratified Sample)

- الوصف : يتم تقسيم المجتمع إلى طبقات (مثل العمر أو الجنس) ثم تُختار عينة من كل طبقة.
- الفائدة : يضمن تمثيل جميع الفئات داخل المجتمع.

- العينة المريحة (Convenience Sample)

- الوصف: يتم اختيار الأفراد بناءً على سهولة الوصول إليهم.
- الفائدة : سريع وسهل، لكنه قد يكون غير ممثل للمجتمع.

- العينة النظامية (Systematic Sample)

- الوصف : يتم اختيار الأفراد بناءً على نظام معين، مثل اختيار كل فرد خامس.
- الفائدة: يُستخدم في الدراسات التي تتطلب تنظيمًا معينًا.

- العينة العنقودية (Cluster Sample)

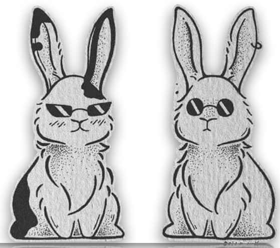
- الوصف: يتم تقسيم المجتمع إلى مجموعات (عناقيد) ثم اختيار مجموعة واحدة أو أكثر للدراسة.
- الفائدة: يُستخدم في الدراسات التي تشمل مناطق جغرافية واسعة.

رابعاً / خطوات اختيار العينة

1. تحديد المجتمع: معرفة من هم الأفراد الذين سيتم دراستهم.
2. اختيار طريقة العينة: تحديد نوع العينة المناسبة.
3. تحديد حجم العينة: حساب عدد الأفراد المطلوبين للحصول على نتائج دقيقة.
4. جمع البيانات: إجراء الدراسة على الأفراد المختارين.

خلاصة

العينة هي جزء مهم من البحث الإحصائي، حيث تتيح للباحثين جمع البيانات وتحليلها بطريقة فعالة. اختيار نوع العينة المناسب يؤثر بشكل كبير على دقة النتائج وقابليتها للتعميم.



جدول التوزيع التكراري

الـــعرض الجدولي والتمثيل البياني

إنشاء جدول التوزيع التكراري

جدول التوزيع التكراري هو وسيلة منظمة لعرض البيانات، حيث يتم تقسيم القيم إلى فئات وتسجيل عدد المرات التي تظهر فيها كل فئة. إليك خطوات إنشاء جدول التوزيع التكراري:

1- جمع البيانات
اجمع البيانات التي تريد تحليلها. على سبيل المثال، لنفترض أن لدينا مجموعة من الدرجات
88, 90, 92, 95

2- تحديد الفئات
حدد عدد الفئات التي تريد تقسيم البيانات إليها. يمكن استخدام (Sturges' Rule) لتحديد عدد الفئات:

$$k = 1 + 3.3 \cdot \log(n)$$ ستور جس

حيث(n) هو عدد القيم في البيانات.

3- حساب نطاق الفئات او المدى الكلي
{النطاق}= {القيمة القصوى} - {القيمة الدنيا}

4- ثم قسم النطاق على عدد الفئات لتحديد عرض كل فئة.

5- إنشاء الجدول

مثال / لنفترض أن لدينا البيانات التالية

55, 67, 70, 72, 73, 75, 80, 82, 85, 88, 90, 92, 95

1- ماهو جدول التوزيع التكراري للبيانات اعلاه
2- جد الحدود الحقيقية والتكرار النسبي والمئوي للفئة
3- التوزيعات المتجمعة (UCF , LCF)
3- مثل ذلك بيانياً

الحل / 1- جدول التوزيع التكراري

الفئة C	التكرار fi
55 - 62	1
63 - 70	2
71 - 78	3
79 - 86	4
87 - 94	3
المجموع	**13**

- تحديد عدد الفئات

لدينا 13 قيمة، لذا باستخدام (Sturges' Rule) :

$$k \approx 1 + 3.322 \cdot \log_{10}(13) \approx 5$$

- حساب النطاق (المدى الكلي)

القيمة القصوى: 95

القيمة الدنيا: 55

المدى = (40 = 55 - 95)

- تحديد طول الفئة

طول كل فئة

$$\frac{40}{5} = 8$$

2- الحدود الحقيقية والتكرار النسبي والمئوي ، 3- التوزيعات المتجمعة (UCF , LCF)

الفئة C	التكرار fi	مركز الفئة x_i	الحد الادنى الحقيقي	الحد الاعلى الحقيقي	التكرار النسبي	التكرار المئوي %	التكرار التجميعي التصاعدي UCF	التكرار التجميعي التنازلي LCF
55 - 62	1	58.5	54.5	62.5	0.07692	7.69	1	13
63 - 70	2	66.5	62.5	70.5	0.15384	15.38	3	12
71 - 78	3	74.5	70.5	78.5	0.23076	23.08	6	10
79 - 86	4	82.5	78.5	86.5	0.30769	30.77	10	6
87 - 94	3	90.5	86.5	94.5	0.23076	23.08	13	3
المجموع	**13**							

- مركز الفئة = الحد الادنى + الحد الاعلى / 2
- الحد الادنى الحقيقي = مركز الفئة - نصف طول الفئة
- الحد الاعلى الحقيقي = مركز الفئة + نصف طول الفئة

- التكرار النسبي ≈ التكرار / مجموع التكرار
- التكرار المئوي ≈ التكرار / مجموع التكرار * 100

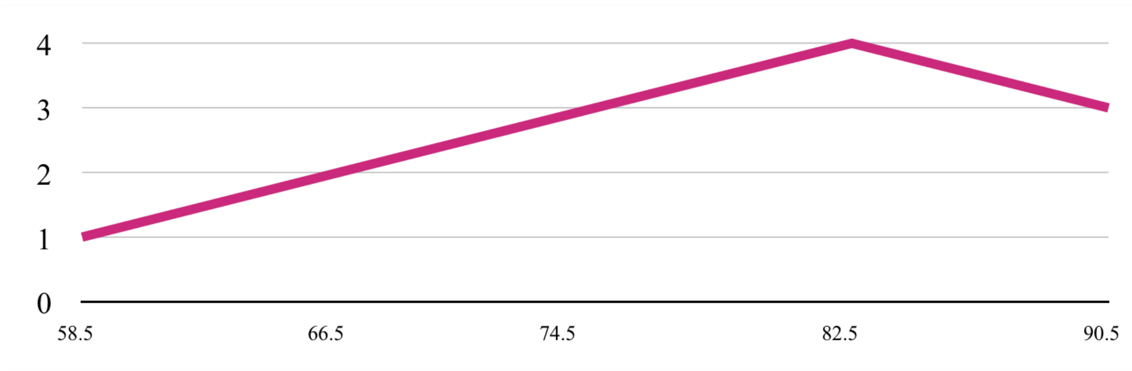
- التكرار التجميعي التنازلي (LCF) للفئة الاولى = مجموع التكرارات
- التكرار التجميعي التنازلي (LCF) للفئة الثانيه = مجموع التكرارات - تكرار الفئة الاولى
- التكرار التجميعي التنازلي (LCF) للفئة الاولى = مجموع التكرارات - تكرار الفئة الثانيه

:

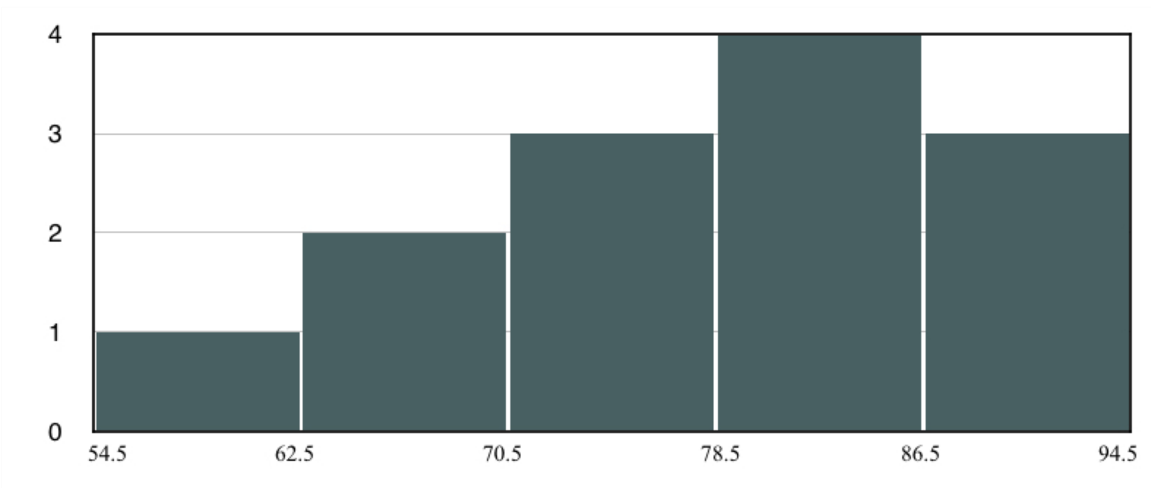
- التكرار التجميعي التصاعدي (UCF) للفئة الاولى = تكرار الفئة الاولى
- التكرار التجميعي التصاعدي (UCF) للفئة الثانيه = تكرار الفئة الثانية + تكرار الفئة الاولى
- التكرار التجميعي التصاعدي (UCF) للفئة الثالثة = تكرار الفئة الثالثة + تكرار الفئة الثانيه

:

4- الرسم البياني بشكل * المضلع التكراري



*** المدرج التكراري**



نشاط 1/

لنفترض أن لدينا مجموعة بيانات تمثل درجات الطلاب في اختبار معين، كما يلي:

45, 55, 67, 70, 72, 75, 75, 80, 82, 85, 85, 88, 90, 92, 95, 95, 100, 100, 100, 100

الخطوات المطلوبة

1. إنشاء جدول التوزيع التكراري .
2. حساب الحدود الحقيقية للفئات .
3. حساب التكرار النسبي والمئوي .
4. التوزيعات المتجمعة (UCF , LCF)
5. التمثيل البياني لجدول التوزيع التكراري

نشاط 2 /

إليك مجموعة بيانات تمثل درجات 50 طالبًا في اختبار:

45, 56, 67, 70, 72, 75, 78, 80, 82, 85,55, 60, 65, 68, 75, 77, 80, 83, 88, 90,92, 94, 96, 97, 85, 88, 91, 93, 95, 99,100, 100, 98, 97, 96, 95, 94, 93, 92, 91, 89, 88, 87, 86, 85, 84, 82, 81, 80, 79

1. تحديد الفئات.
2. حساب التكرار لكل فئة.
3. حساب التكرار النسبي والتكرار المئوي .
4. التوزيعات المتجمعة (UCF , LCF)
5. التمثيل البياني لجدول التوزيع التكراري

مقاييس النزعة المركزية

مـقـايـيـس الـنـزعـة الـمـركـزيـة Measures of central tendency

تعد مقاييس النزعة المركزية أدوات إحصائية تُستخدم **لتلخيص** و**توصيف** مجموعة من البيانات من خلال تحديد نقطة مركزية تمثل تلك البيانات. تعتبر هذه المقاييس أساسية في تحليل البيانات لأنها تساعد في فهم خصائص مجموعة البيانات وتوجهاتها.

تُعتبر مقاييس النزعة المركزية أدوات حيوية في علم الإحصاء، حيث توفر نظرة شاملة حول البيانات وتساعد الباحثين وصانعي القرار في استخلاص النتائج والتوصيات بناءً على المعلومات المتاحة.

أهمية مقاييس النزعة المركزية:

1. تساعد في تقديم فكرة عامة عن البيانات بطريقة سهلة الفهم.
2. تُستخدم للمقارنة بين مجموعات بيانات مختلفة.
3. تساعد في تحديد الاتجاهات العامة في البيانات.

الأنواع الرئيسية لمقاييس النزعة المركزية:

1. الوسط : هو القيمة التي يتم حسابها بجمع جميع القيم وقسمتها على عددها. يُعتبر مقياسًا شائعًا، لكن قد يتأثر بالقيم الشاذة.
2. الوسيط: هو القيمة التي تقسم مجموعة البيانات إلى نصفين متساويين، مما يجعله أقل تأثرًا بالقيم الشاذة.
3. المنوال: هو القيمة الأكثر تكرارًا في مجموعة البيانات، ويُستخدم بشكل خاص في البيانات الفئوية.

اولاً / الوسط الحسابي $(\bar{x})$ the arithmetic mean

يعتبر الوسط الحسابي مقياس للنزعة المركزية يُستخدم لتحديد القيمة المتوسطة لمجموعة من البيانات. يُحسب بجمع جميع القيم ثم قسمتها على عددها.

-خصائص الوسط الحسابي

1. التأثر بالقيم الشاذة : قد يتأثر الوسط الحسابي بشكل كبير بالقيم العالية أو المنخفضة بشكل غير عادي.
2. سهولة الفهم: يُعتبر سهل الفهم والحساب، مما يجعله شائعًا في التحليل الإحصائي.
3. استخدامات متعددة : يُستخدم في مختلف المجالات مثل الاقتصاد، علم النفس، والبحوث العلمية.

○ الوسط الحسابي للبيانات غير المبوبه

إذا كانت لديك مجموعة من القيم $x_1, x_2, x_3, \ldots, x_n$ ، فإن الوسط الحسابي يُحسب على النحو التالي:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Or

$$\bar{x} = \frac{\sum xi}{n}$$

إذ أن :

$\sum xi$ هو مجموع القيم

- $\bar{x}$ هو الوسط الحسابي.
- (n) هو عدد القيم في المجموعة.

مثال/

لنفترض أن لدينا القيم التالية:

5 ,6 ,8 ,4

ماهوالوسط الحسابي للقيم اعلاه

الحل /

$$4 + 8 + 6 + 5 = 23$$

$$\bar{x} = \frac{23}{4} = 5.75$$

مثال/

لنفترض أن لدينا القيم التالية:

M,10, 8 ,2

أذ علمت بأن الوسط الحسابي للقيم اعلاه $(\bar{x} = 25)$ ، جد قيمة M

الحل /

$$\bar{x} = \frac{\sum xi}{n}$$

$$25 = \frac{\sum xi}{4}$$

مجموع القيم يساوي

$$\sum xi = 100$$

$$100 - 20 = 80$$

قيمة M =80

للتأكد من صحة الحل

مجموع القيم

$$\sum xi = 80+2+8+10$$

$$=100$$

الوسط الحسابي

$$\bar{x} = \frac{\sum xi}{n}$$

$$\bar{x} = \frac{100}{4}$$

$$\bar{x} = 25$$

ملاحظه مهمة /

الوسط الحسابي هو أداة مهمة لفهم البيانات، لكنه ليس دائمًا المقياس الأفضل للنزعة المركزية، خاصة في وجود قيم شاذة مثل قيمة M التي تساوي (80) في المثال السابق الذي كان تأثيرها على قيمة الوسط الحسابي مما جعلها تساوي 25 على الرغم من أن جميع القيم كانت اقل او تساوي 10 الا أنه كان متأثراً بالقيمة الشاذة (M) . لذا يُفضل استخدامه جنبًا إلى جنب مع مقاييس أخرى مثل الوسيط والوضع للحصول على صورة شاملة عن البيانات .

o الوسط الحسابي للبيانات المبوبه

لحساب الوسط الحسابي للبيانات المبوبة، يمكنك اتباع الخطوات التالية:

1. تحديد الفئات: قم بتقسيم البيانات إلى فئات (intervals) إذا لم تكن مقسمة بالفعل.

2. حساب مركز الفئة : يمكن حساب نقطة المنتصف او مركز الفئة (midpoint) باستخدام الصيغة:

مركز الفئة =حدود الفئة السفلى +حدود الفئة العليا / 2

3. حساب التكرارات وجمعها

4. حساب الوسط الحسابي: استخدم الصيغة التالية:

$$\bar{x} = \frac{\sum (f_i \cdot x_i)}{\sum f_i}$$

إذ أن :-

f_i هو تكرار الفئة

x_i هو نقطة المنتصف للفئة.

مثال/

من البيانات التالية ، جد $\bar{x}$

الفئة C	التكرار fi
0 - 10	3
10 - 20	5
20 - 30	2
المجموع	**10**

علماً بأن هذه البيانات كانت لدرجات مجموعة من الطلاب

الحل /

الفئة C	التكرار fi	مركز الفئة x_i	$f_i\, x_i$
0 - 10	3	5	15
10 - 20	5	15	75
20 - 30	2	25	50
المجموع	**10**		**140**

$$\bar{x} = \frac{\sum (f_i \cdot x_i)}{\sum f_i}$$

$$\bar{x} = \frac{(3 \cdot 5) + (5 \cdot 15) + (2 \cdot 25)}{3 + 5 + 2} = \frac{15 + 75 + 50}{10} = \frac{140}{10} = 14$$

الخلاصه

- **البيانات المبوبة**

هي البيانات التي تُقسم إلى فئات أو مجموعات. يُستخدم هذا الأسلوب لتسهيل التحليل وفهم التوزيع.

مثال

إذا كان لديك مجموعة من الدرجات، يمكنك تقسيمها إلى فئات مثل:

- 0-10
- 11-20
- 21-30

في هذه الحالة، تم تنظيم البيانات في فئات محددة، مما يسهل حساب الإحصائيات مثل الوسط الحسابي والانحراف المعياري.

- **البيانات غير المبوبة**

هي البيانات التي لا تُقسم إلى فئات، بل تُسجل بشكل فردي.

مثال

مجموعة من الدرجات مثل: 8، 15، 22، 30.

هنا، يتم التعامل مع كل درجة كقيمة فردية بدون تجميعها في فئات.

- **الفرق بينهما**

- التنظيم : البيانات المبوبة تُنظم في فئات، بينما البيانات غير المبوبة تُترك دون تنظيم.
- التحليل : البيانات المبوبة تسهل التحليل الإحصائي، بينما البيانات غير المبوبة قد تتطلب معالجة إضافية للحصول على إحصائيات مفيدة.

مثال/ من الجدول جد الوسط الحسابي للبيانات التالية :-

التكرار fi	الفئة C
3	40-49
6	50-59
7	60-69
6	70-79
6	80-89
4	90-100

علماً بأن هذه البيانات كانت لدرجات مجموعة من الطلاب

الحل /

$f_i x_i$	مركز الفئة x_i	التكرار fi	الفئة C
133.5	44.5	3	40-49
327	54.5	6	50-59
451.5	64.5	7	60-69
447	74.5	6	70-79
507	84.5	6	80-89
378	94.5	4	90-99
2244		**32**	المجموع

$$\bar{x} = \frac{\sum (f_i \cdot x_i)}{\sum f_i}$$

$$= \frac{2244}{32} = 70.13$$

التفسير

يُظهر من خلال الوسط الحسابي (70.19) أن الأداء العام للطلاب كان جيدًا

ثانياً / الوسط الترجيحي The weighted average (X_w)

يعد الوسط الترجيحي هو مقياس مركزي يُستخدم عندما تكون بعض القيم لها أهمية أو وزن أكبر من غيرها.
يُحسب الوسط الترجيحي باستخدام الأوزان المعطاة لكل قيمة.
تُحسب قيمة الوسط الترجيحي باستخدام الصيغة التالية:

$$X_w = \frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

إذ ان

- x_i : القيم .
- w_i : الأوزان المرتبطة بكل قيمة.

مثال / لنفترض أن لدينا البيانات التالية حول درجات الطلاب في اختبار معين، مع الأوزان المعطاة لكل درجة بناءً على أهمية كل اختبار

الدرجة x_i	الاوزن w_i
85	3
90	2
78	1
92	4
88	2

الحل /

الدرجة x_i	الاوزن w_i	$x_i \cdot w_i$
85	3	255
90	2	180
78	1	78
92	4	368
88	2	176
المجموع	12	1057

$$X_w = \frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

$$= \frac{1057}{12} \approx 88.08$$

التفسير
أن متوسط درجات الطلاب، مع الأخذ في الاعتبار أوزان الاختبارات، هو 88.08. يُظهر ذلك أن الأداء العام للطلاب كان جيدًا، مع مراعاة أهمية كل اختبار في الحساب

مثال / لنفترض أن شركة لديها الأرباح التالية من ثلاثة منتجات جد Xw:

المنتج	الارباح x_i	الاوزن w_i (حجم المبيعات)
A	100,000	5
B	150,000	3
C	50,000	2

الحل /

المنتج	الارباح x_i	الاوزن w_i (حجم المبيعات)	$x_i \cdot w_i$
A	100,000	5	500,000
B	150,000	3	450,000
C	50,000	2	100,000
المجموع		**10**	**1,050,000**

$$X_w = \frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

$$= \frac{1,050,000}{10} = 105,000$$

تفسير
هذا يعني أن متوسط الأرباح مع الأخذ في الاعتبار حجم المبيعات لكل منتج هو 105,000، مما يساعد الشركة على فهم الأداء العام للأرباح

متى يُستخدم الوسط الترجيحي في حساب متوسط الأرباح؟

1. **أرباح من مصادر مختلفة :** إذا كانت الشركة تحقق أرباحًا من عدة منتجات أو خدمات، وقد تكون بعض المنتجات أكثر أهمية أو مبيعًا من الأخرى. في هذه الحالة، يمكن استخدام الأوزان لتمثيل حجم المبيعات أو إيرادات كل منتج.
2. **فترات زمنية مختلفة:** إذا كانت هناك أرباح من فترات زمنية مختلفة (مثل الربع الأول والثاني والثالث والرابع) وتختلف الأهمية أو الأثر الاقتصادي لكل فترة، يمكن استخدام الأوزان لحساب متوسط الأرباح.
3. **فوائد مالية :** إذا كانت هناك فوائد أو تكاليف مختلفة مرتبطة بأرباح معينة، يمكن استخدام الأوزان لتمثيل هذه الفوائد أو التكاليف.

ثالثاً / الوسط التوافقي (Harmonic Mean) (H)

الوسط التوافقي هو مقياس مركزي يُستخدم بشكل خاص عندما نريد حساب متوسط مجموعة من النسب أو المعدلات. يُعتبر الوسط التوافقي مفيدًا في حالات مثل حساب متوسط السرعات أو المعدلات التي تتعلق بالزمن.

o الوسط التوافقي للبيانات غير المبوبه

يُحسب الوسط التوافقي للبيانات غير المبوبة باستخدام الصيغة التالية :

$$H=\frac{n}{\sum \frac{1}{x_i}}$$

حيث:

- (n) هو عدد القيم.
- (x_i) هي القيم.

مثال/

لنفترض أن لدينا مجموعة من السرعات (km/h)

60 , 90 ,120

الحل /

السرعه (km/h)	$\frac{1}{x_i}$
60	$\frac{1}{60} \approx 0.01667$
90	$\frac{1}{90} \approx 0.01111$
120	$\frac{1}{120} \approx 0.00833$
المجموع	0.03611

$$H=\frac{n}{\sum \frac{1}{x_i}}$$

$$H=\frac{3}{0.03611} \approx 83.05 \text{ km/h}$$

o الوسط التوافقي للبيانات المبوبه

$$H = \frac{\sum f_i}{\sum \left(fi \frac{1}{xi}\right)}$$

<u>مثال/ لنفترض أن لدينا البيانات المبوبه التالية ، احسب قيمة الوسط التوافقي (H)</u>

الفئة C	التكرار fi	مركز الفئة x_i	$\frac{1}{x_i}$	$fi\frac{1}{x_i}$
0-10	4	5	$\frac{1}{5} = 0.2$	$4 \cdot 0.2 = 0.8$
11-20	6	15.5	$\frac{1}{15.5} \approx 0.0645$	$6 \cdot 0.0645 \approx 0.387$
21-30	8	25.5	$\frac{1}{25.5} \approx 0.0392$	$8 \cdot 0.0392 \approx 0.314$
31-40	2	35.5	$\frac{1}{35.5} \approx 0.0282$	$2 \cdot 0.0282 \approx 0.0564$
المجموع	**20**			**1.5574**

الحل /

$$H = \frac{\sum f_i}{\sum \left(fi \frac{1}{xi}\right)}$$

$$= \frac{20}{1.5574} \approx 12.85$$

<u>الوسيط The median (Me)</u>

الوسيط هو مقياس مركزي يُستخدم لتحديد القيمة المتوسطة لمجموعة من البيانات. يُعتبر الوسيط مفيدًا بشكل خاص عندما تكون البيانات تحتوي على قيم متطرفة (outliers) أو عندما لا تكون موزعة بشكل متساوٍ.

o كيفية حساب الوسيط للبيانات غير المبوبة

1. ترتيب البيانات :

قم بترتيب القيم من الأصغر إلى الأكبر.

2. تحديد عدد القيم (n) :

- إذا كان عدد القيم فرديًا، يكون الوسيط هو القيمة التي تقع في المنتصف.
- إذا كان عدد القيم زوجيًا، يُحسب الوسيط كمتوسط القيمتين في المنتصف.

- خطوات حساب الوسيط

حالة عدد القيم فرديًا

- لنفترض أن لدينا القيم التالية:

3, 1, 4, 2, 5

1. ترتيب البيانات :

1, 2, 3, 4, 5

2. تحديد الوسيط :

- العدد الإجمالي للقيم هو 5 (عدد فردي).

الوسيط هو القيمة الثالثة 3.

حالة عدد القيم زوجيًا

- لنفترض أن لدينا القيم التالية:

7, 1, 3, 2

1. ترتيب البيانات:

1, 2, 3, 7

2. تحديد الوسيط:

- العدد الإجمالي للقيم هو 4 (عدد زوجي).
- الوسيط هو متوسط القيمتين في المنتصف:

$$Me = \frac{2+3}{2} = 2.5$$

- في الحالة الفردية: الوسيط هو 3
- في الحالة الزوجية : الوسيط هو 2.5.

o كيفية حساب الوسيط للبيانات المبوبة

لحساب الوسيط للبيانات المبوبة، يمكنك اتباع الخطوات التالية:

1. جمع البيانات:

- اجمع البيانات وقم بتنظيمها في جدول مبوب، حيث يحتوي على الفئات أو الفترات وعدد الملاحظات في كل فئة.

2. حساب التكرارات التراكمية:

- أضف عمودًا للتكرارات التراكمية. قم بجمع التكرارات من أعلى إلى أسفل للحصول على التكرار التراكمي لكل فئة.

3. تحديد موقع الوسيط:

- احسب موقع الوسيط باستخدام الصيغة:

$$N = \sum f \quad ، \quad = \frac{N}{2}$$

حيث N هو إجمالي عدد الملاحظات.

4. تحديد الفئة الوسيطة:
- **<u>إبحث عن الفئة التي تحتوي على القيمة التي تعادل أو تتجاوز موقع الوسيط.</u>** هذه الفئة تعرف بالفئة الوسيطة.

5. حساب الوسيط:
- استخدم الصيغة التالية لحساب الوسيط:

$$\text{Me} = L + \left(\frac{\left(\frac{N}{2} - CF \right)}{f} \right) \times c$$

إذ أن :
- (L) = الحد الأدنى للفئة الوسيطة.
- (CF) = التكرار التراكمي للفئة السابقة للفئة الوسيطة.
- (f) = تكرار الفئة الوسيطة.
- (c) = طول الفئة.

<u>مثال / من البيانات المبوبة التالية ، جد الوسيط .</u>

الفئة C	التكرار f	التكرار التراكمي CF
10-20	5	5
20-30	10	15
30-40	15	30
40-50	5	35

N= 35

موقع الوسيط:

$$\frac{35}{2} = 17.5$$

الفئة الوسيطة:

30-40

(لأن التكرار التراكمي يصل إلى 30).

5. حساب الوسيط:
- (L = 30)
- (CF = 15) (التكرار التراكمي للفئة 20-30)
- (f = 15) (تكرار الفئة 30-40)
- (c = 10) (طول الفئة)

$$\text{Me} = 30 + \left(\frac{(17.5 - 15)}{15} \right) \times 10$$

$$= 30 + \left(\frac{2.5}{15} \right) \times 10$$

$$= 30 + \frac{25}{15} = 30 + 1.67 \approx 31.67$$

لذا، الوسيط للبيانات المبوبة هو حوالي 31.67 .

<u>مثال / من البيانات المبوبة التالية ، جد الوسيط .</u>

الفئة C	التكرار f	التكرار التراكمي CF
0-10	**8**	**8**
10-20	12	**20**
20-30	**20**	**40**
30-40	15	55
40-50	10	65
50-60	5	**70**

N=70

موقع الوسيط

$$\frac{70}{2} = 35$$

تحديد الفئة الوسيطة

- ابحث عن الفئة التي تحتوي على القيمة تتجاوز او تعادل القيمة 35 تعتمد على التكرار التراكمي 40 20-30

حساب الوسيط

- (الحد الأدنى للفئة الوسيطة) $(L = 20)$
- (التكرار التراكمي للفئة السابقة 20-10) $(CF = 20)$
- (تكرار الفئة 30-20) $(f = 20)$
- (طول الفئة) $(c = 10)$

$$Me = 20 + \left(\frac{(35 - 20)}{20}\right) \times 10$$

$$= 20 + \left(\frac{15}{20}\right) \times 10$$

ح

- الوسيط للبيانات المبوبة هو 27.5.

المنوال The mode (MO)

○ كيفية حساب المنوال للبيانات غير المبوبة

لحساب المنوال للبيانات غير المبوبة، يمكنك اتباع الخطوات التالية:
- اجمع جميع القيم في مجموعة البيانات.
- احسب عدد مرات تكرار كل قيمة في البيانات.
- ابحث عن القيمة التي تتكرر أكثر من غيرها. هذه القيمة تُعرف بالمنوال.

مثال / البيانات غير المبوبة جد المنوال :

5, 7, 8, 5, 10, 7, 5, 12, 15, 7

الحل /
- تحديد تكرار كل قيمة:
  -(5): 3 مرات
  -(7): 3 مرات
  -(8): 1 مرة
  -(10): 1 مرة
  -(12): 1 مرة
  -(15): 1 مرة
- القيم الأكثر تكرارًا هي(5) و(7) (كل منهما 3 مرات).

النتيجة:
- المنوال: البيانات غير المبوبة تحتوي على منوال مزدوج (بسبب تساوي التكرار)، لذا المنوال هنا هو 5 و7.

○ كيفية حساب المنوال للبيانات المبوبة

لحساب المنوال للبيانات المبوبة، يمكنك اتباع الخطوات التالية:
الصيغة التالية لحساب المنوال:

$$\text{Mo} = L + \left(\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \right) \times c$$

إذ أن :
- L = الحد الأدنى للفئة المنوالية.
- f_1 = تكرار الفئة المنوالية. **(الفئة التي تحتوي على أعلى تكرار، تُعرف بالفئة المنوالية)**
- f_0 = تكرار الفئة السابقة للفئة المنوالية.
- f_2 = تكرار الفئة التالية للفئة المنوالية.
- c = طول الفئة.

مثال / من البيانات المبوبة ، جد المنوال

الفئة C	التكرار f
0-10	8
10-20	**12**
20-30	**20**
30-40	**15**
40-50	10
50-60	5

الحل /

1. تحديد الفئة الأكثر تكرارًا او الاعلى تكراراً
- الفئة 20-30 هي الفئة المنوالية، حيث تكرارها هو 20 (الأعلى).

2. حساب المنوال:
- ($L = 20$) (الحد الأدنى للفئة المنوالية).
- $f_1 = 20$ (تكرار الفئة المنوالية).
- $f_0 = 12$ (تكرار الفئة السابقة 10-20).
- $f_2 = 15$ (تكرار الفئة التالية 30-40).
- $c = 10$ (طول الفئة).

$$Mo = 20 + \left(\frac{20 - 12}{(20 - 12) + (20 - 15)} \right) \times 10$$

$$= 20 + \left(\frac{8}{8 + 5} \right) \times 10$$

$$= 20 + \left(\frac{8}{13} \right) \times 10 \approx 20 + 6.15 \approx 26.15$$

- المنوال للبيانات المبوبة هو **حوالي** 26.15

مقاييس

التشتت اوالاختلاف

مقاييس التشتت او الاختلاف

Measures of dispersion or variation

تُعد مقاييس التشتت او الاختلاف من العناصر الأساسية في علم الإحصاء، حيث تلعب دورًا حيويًا في تحليل البيانات وفهم توزيعها. فعلى الرغم من أن المقاييس المركزية مثل المتوسط والوسيط تعكس القيم المركزية لمجموعة بيانات معينة، إلا أن هذه المقاييس وحدها لا تكفي لتقديم الصورة الكاملة عن البيانات.

تساعد مقاييس التشتت في توضيح مدى تباين أو تشتت القيم عن المتوسط، مما يمنح الباحثين والمحللين القدرة على تقييم مدى اتساق البيانات ووجود القيم المتطرفة. من خلال استخدام مقاييس مثل الانحراف المعياري، والتباين، والمدى، يمكن فهم طبيعة البيانات بشكل أعمق، وتوفير معلومات قيمة حول استقرار النتائج.

تُستخدم مقاييس التشتت في مجالات متعددة، مثل الاقتصاد، والعلوم الاجتماعية، والطب، والهندسة، حيث تساعد في اتخاذ القرارات المستندة إلى بيانات موثوقة. في هذه المقدمة، سنستعرض أبرز مقاييس التشتت والاختلاف، وكيفية استخدامها في تحليل البيانات لفهم الظواهر المختلفة بشكل أفضل. ومن المقاييس الشائعة في الاحصاء هي :-

1. **المدى (Range)**
   - يُحسب بطرح أقل قيمة من أعلى قيمة في مجموعة البيانات.
   - يوضح مدى انتشار البيانات.
2. **الانحراف المتوسط (Average deviation)**
3. **الانحراف المعياري (Standard Deviation)**
   - يقيس متوسط مدى تشتت القيم عن المتوسط.
   - كلما زاد الانحراف المعياري، زادت درجة تشتت البيانات.
4. **التباين (Variance)**
   - هو مربع الانحراف المعياري ويعبر عن مدى انتشار البيانات.
   - يُستخدم في التحليلات الإحصائية المختلفة.

اولاً / المدى (Range)

المدى (Range) هو أحد أبسط مقاييس التشتت، ويستخدم لقياس مدى انتشار البيانات. يُحسب المدى عن طريق طرح أقل قيمة في مجموعة البيانات من أعلى قيمة، مما يعطي فكرة عن الفجوة بين القيم القصوى.

- **المدى للبيانات غير المبوبة**

- كيفية حساب المدى

1. تحديد القيم القصوى
   - ابحث عن أقل قيمة في البيانات (الحد الأدنى).
   - ابحث عن أعلى قيمة في البيانات (الحد الأقصى).
2. حساب المدى من الصيغة التالية

المدى = الحد الأقصى - الحد الأدنى

مثال/

لنفترض أن لدينا مجموعة البيانات التالية: 5, 12, 7, 3, 9 جد المدى

الحل /
- الحد الأدنى: 3
- الحد الأقصى: 12

$$= 12 - 3$$
$$= 9$$

- **أهمية المدى:**

- بسيط وسريع: المدى سهل الحساب ويوفر نظرة عامة سريعة عن تشتت البيانات.
- مؤشر أولي: يساعد في فهم الفجوة بين القيم القصوى، **لكنه لا يعكس توزيع البيانات بشكل كامل.**

ومع ذلك، يجب أن نلاحظ أن المدى قد يتأثر بشدة بالقيم المتطرفة، مما يحد من فعاليته في بعض الحالات. لذا، يُفضل استخدامه مع مقاييس تشتت أخرى للحصول على صورة أوضح عن البيانات.

مثال /

لنفترض أن لدينا مجموعة البيانات التالية : 2, 4, 5, 6, 100 جد المدى

الحل /
- الحد الأدنى: 2
- الحد الأقصى: 100

نحسب المدى كالتالي:

$$= 100 - 2$$
$$= 98$$

تحليل المدى:
- المدى: 98
- القيمة المتطرفة: 100

- **الملاحظات:**

- في هذا المثال، يكون المدى مرتفعًا جدًا بسبب وجود قيمة متطرفة (100) مقارنة بالقيم الأخرى.
- هذا يوضح أن المدى قد لا يعكس بشكل دقيق تشتت البيانات في حالة وجود قيم متطرفة، حيث أن معظم القيم قريبة من الحد الأدنى (2) ولا تعكس التشتت الحقيقي للبيانات.

لذا، يُفضل استخدام مقاييس تشتت أخرى مثل الانحراف المعياري أو الربعية للحصول على صورة شاملة عن توزيع البيانات، خاصة في وجود قيم متطرفة.

المدى للبيانات المبوبة

عند حساب المدى للبيانات المبوبة، يمكن استخدام مركز كل فئة بدلاً من الحدود القصوى. مركز الفئة هو القيمة المتوسطة التي تمثل تلك الفئة، وغالبًا ما يُحسب باستخدام الصيغة:

مركز الفئة = الحد الأدنى للفئة + الحد الأقصى للفئة / 2
المدى = الحد الأقصى لمركز الفئة - الحد الأدنى لمركز الفئة

خطوات حساب المدى للبيانات المبوبة:

1. تحديد الفئات: قم بإنشاء جدول يوضح الفئات وتكرار كل فئة.
2. حساب مركز الفئة: احسب مركز الفئة لكل فئة.
3. تحديد المدى: استخدم مركز الفئة للحد الأدنى والحد الأقصى.

مثال / لنفترض أن لدينا جدول تكراري يوضح توزيع درجات الطلاب كما يلي ، فما هو مقياس المدى لتلك البيانات

الفئة C	التكرار fi
0-10	2
11-20	5
21-30	8
31-40	4
41-50	1

الحل /

الفئة C	التكرار fi	مركز الفئة x_i
0-10	2	5
11-20	5	15.5
21-30	8	25.5
31-40	4	35.5
41-50	1	45.5

- مركز الفئة الأدنى: 5
- مركز الفئة الأعلى: 45.5

$$45.5 - 5 = 40.5$$

- المدى: 40.5
- هذا يعكس الاختلاف بين مركز الفئات، مما يعطي فكرة عن مدى انتشار البيانات في جميع الفئات.

الملاحظات:

- استخدام مركز الفئة يساعد في الحصول على تقدير أفضل للمدى عند التعامل مع البيانات المبوبة.
- ومع ذلك، يجب أن نكون على علم بأن هذا الأسلوب قد لا يعكس التشتت الفعلي للبيانات داخل الفئات، ولكنه يوفر تقديرًا مفيدًا.

ثانيا / الانحراف المتوسط Average deviation (AD)

يُعرف الانحراف المتوسط على أنه مقياس لتشتت البيانات يُستخدم لقياس مدى انتشار القيم حول المتوسط الحسابي. يُعتبر الانحراف المتوسط طريقة بسيطة وفعالة لفهم مدى تشتت البيانات دون الحاجة إلى التعامل مع القيم المربعة كما في الانحراف المعياري.

كيفية حساب الانحراف المتوسط

1. حساب المتوسط (المعدل) لمجموعة البيانات.
2. حساب الفرق بين كل قيمة والمتوسط، ثم خذ القيمة المطلقة لهذا الفرق.
3. حساب متوسط الانحرافات المطلقة.

الصيغة:

$$AD = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

إذ أن :

- n هو عدد القيم في مجموعة البيانات.
- x_i هي كل قيمة في مجموعة البيانات.
- $\bar{x}$ هو المتوسط الحسابي.

مثال/

لنفترض أن لدينا مجموعة البيانات التالية:

3 ,5 ,6 ,8 ,4

احسب الانحراف المتوسط للبيانات اعلاه .

الحل /

حساب المتوسط للبيانات $\bar{x}$

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{x} = \frac{4+8+6+5+3}{5} = \frac{26}{5} = 5.2$$

$$|x_i - \bar{x}|$$

حساب الانحرافات المطلقة: $|x_i - \bar{x}|$

$$|4 - 5.2| = 1.2$$
$$|8 - 5.2| = 2.8$$
$$|6 - 5.2| = 0.8$$
$$|5 - 5.2| = 0.2$$
$$|3 - 5.2| = 2.2$$

حساب الانحراف المتوسط:

$$AD = \frac{1.2 + 2.8 + 0.8 + 0.2 + 2.2}{5} = \frac{7.2}{5} = 1.44$$

- الانحراف المتوسط : 1.44

الملاحظات:

- يُعتبر الانحراف المتوسط مقياسًا بسيطًا وسهل الفهم، ولكنه قد لا يُستخدم بشكل شائع مقارنةً بالانحراف المعياري، خصوصًا في الحالات التي تتطلب تحليلاً متقدمًا.
- يمكن أن يكون الانحراف المتوسط مفيدًا في فهم مدى تشتت البيانات في حالات معينة، خاصة عندما تكون البيانات تحتوي على قيم متطرفة.

ثالثاً / الانحراف المعياري (Standard Deviation)

يعد الانحراف المعياري همقياس إحصائي يُستخدم لقياس تشتت مجموعة من القيم حول المتوسط. يُعتبر الانحراف المعياري من أكثر المقاييس شيوعًا في الإحصاء، حيث يُظهر مدى تباعد القيم عن المتوسط، مما يساعد في فهم توزيع البيانات بشكل أفضل.

o الانحراف المعياري للبيانات غير المبوبة :

1- حساب المتوسط (المعدل) لمجموعة البيانات.
2- حساب الفرق بين كل قيمة والمتوسط، ثم قم بتربيع هذا الفرق.
3- حساب متوسط الفروق المربعة.
4- حساب الجذر التربيعي لمتوسط الفروق المربعة للحصول على الانحراف المعياري.

حساب الانحراف المعياري للسكان بالصيغة التالية :

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

- حساب الانحراف المعياري للعينة بالصيغة التالية :

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

إذ أن :

- σ هو الانحراف المعياري للسكان.
- s هو الانحراف المعياري للعينة.
- N هو عدد القيم في السكان.
- n هو عدد القيم في العينة.
- x_i هي كل قيمة في مجموعة البيانات.
- μ هو المتوسط للسكان.
- $\bar{x}$ هو المتوسط للعينة.

مثال/

لنفترض أن لدينا مجموعة البيانات التالية:
[3 ,5 ,6 ,8 ,4]

الحل /

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

حساب المتوسط

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{x} = \frac{4+8+6+5+3}{5} = \frac{26}{5} = 5.2$$

حساب الفروق المربعة

$$(4-5.2)^2 = 1.44$$
$$(8-5.2)^2 = 7.84$$
$$(6-5.2)^2 = 0.64$$
$$(5-5.2)^2 = 0.04$$
$$(3-5.2)^2 = 4.84$$

حساب متوسط الفروق المربعة :

$$= \frac{1.44 + 7.84 + 0.64 + 0.04 + 4.84}{5-1} = \frac{14.8}{4} = 3.7$$

حساب الجذر التربيعي

$$s = \sqrt{3.7} \approx 1.9$$

الملاحظات:

- الانحراف المعياري يُعتبر مقياسًا قويًا للتشتت لأنه يأخذ بعين الاعتبار كل القيم في مجموعة البيانات.
- يُستخدم الانحراف المعياري في العديد من التطبيقات، مثل تحليل البيانات، والبحوث العلمية، والاقتصادية، حيث يساعد في فهم مدى استقرار البيانات ووجود القيم المتطرفة.

عندما تمثل البيانات مجتمعًا كاملاً وليس عينة، يتم حساب الانحراف المعياري بشكل مختلف قليلاً. في هذه الحالة، نقسم مجموع الفروق المربعة على (N) (حيث (N) هو عدد القيم في المجتمع) بدلاً من $n - 1$

مثال /

لنفترض أن لدينا نفس مجموعة البيانات التي تمثل مجتمعًا كاملاً:
[3 ,5 ,6 ,8 ,4]

الحل /

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

حساب المتوسط

$$\mu = \frac{4+8+6+5+3}{5} = \frac{26}{5} = 5.2$$

حساب الفروق المربعة

$$(4-5.2)^2 = 1.44$$
$$(8-5.2)^2 = 7.84$$
$$(6-5.2)^2 = 0.64$$
$$(5-5.2)^2 = 0.04$$
$$(3-5.2)^2 = 4.84$$

حساب مجموع الفروق المربعة :

$$= 1.44 + 7.84 + 0.64 + 0.04 + 4.84 = 14.8$$

$$= \frac{14.8}{5} = 2.96$$

حساب الجذر التربيعي

$$\sigma = \sqrt{2.96} \approx 1.72$$

o الانحراف المعياري للبيانات المبوبة

عند التعامل مع البيانات المبوبة، يُحسب الانحراف المعياري بطريقة مختلفة قليلاً عن البيانات المفردة. يعتمد ذلك على استخدام حدود الفئات وتكراراتها.

<u>خطوات حساب الانحراف المعياري للبيانات المبوبة:</u>

1- قم بتحديد الفئات (أو الفواصل) وتكرارات كل فئة.

2-احسب نقطة الوسط (أو القيمة المركزية) لكل فئة

3- حساب المتوسط المرجح

$$\bar{x} = \frac{\sum (f_i \cdot x_i)}{N}$$

إذ أن :

f_i هو التكرار لكل فئة

N هو إجمالي التكرارات.

4- حساب التباين

$$s^2 = \frac{\sum (f_i \cdot (x_i - \bar{x})^2)}{N}$$

5- حساب الانحراف المعياري

$$s = \sqrt{s^2}$$

<u>مثال / افترض أن لدينا البيانات المبوبة التالية:</u>

الفئة C	التكرار fi	مركز الفئة x_i
0-10	5	5
10-20	10	15
20-30	15	25
30-40	8	35

الحل / حساب المتوسط المرجح

$$\bar{x} = \frac{(5 \times 5) + (10 \times 15) + (15 \times 25) + (8 \times 35)}{38}$$

$$\bar{x} = \frac{25 + 150 + 375 + 280}{38} = \frac{830}{38} \approx 21.84$$

حساب التباين

$$s^2 = \frac{(5 \times (5 - 21.84)^2) + (10 \times (15 - 21.84)^2) + (15 \times (25 - 21.84)^2) + (8 \times (35 - 21.84)^2)}{38}$$

- حساب كل جزء

$(5 - 21.84)^2 \approx 283.65$
$(15 - 21.84)^2 \approx 47.77$
$(25 - 21.84)^2 \approx 9.84$
$(35 - 21.84)^2 \approx 172.74$

- استبدال القيم

$$s^2 = \frac{(5 \times 283.65) + (10 \times 47.77) + (15 \times 9.84) + (8 \times 172.74)}{38}$$

$$s^2 = \frac{1418.25 + 477.7 + 147.6 + 1381.92}{38} = \frac{3425.57}{38} \approx 90.65$$

حساب الانحراف المعياري

$$s = \sqrt{90.65} \approx 9.53$$

<u>الملاحظات:</u>

- الانحراف المعياري للبيانات المبوبة يعطي فكرة عن مدى تشتت البيانات ضمن الفئات.
- من المهم استخدام القيم المركزية لتقدير التشتت بدقة.

<u>رابعاً / التباين (Variance)</u>

يعد التباين مقياس إحصائي يُستخدم لقياس مدى تشتت مجموعة من القيم حول المتوسط. **<u>يُعتبر التباين هو مربع الانحراف المعياري</u>**، مما يعني أنه يوفر معلومات مشابهة حول تشتت البيانات، لكنه يعبر عنها بشكل مختلف.

للسكان

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

- للعينة :

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

إذ أن :

- σ^2 هو التباين للسكان.
- s^2 هو التباين للعينة.
- N هو عدد القيم في المجتمع.
- n هو عدد القيم في العينة.
- x_i هي كل قيمة في مجموعة البيانات.
- μ هو المتوسط للسكان.
- $\bar{x}$ هو المتوسط للعينة.

ويمكن تطبيق الامثله السابقة المتعلقة في مقياس الانحراف المعياري على مقياس التباين

الارتباط

(دراسة العلاقة بين متغيرين)

الارتبـــاط (دراسة العلاقة بين متغيرين)

Correlation (study of the relationship between two variables)

تناولنا في الفصول السابقة دراسة متغير واحد أو ظاهرة معينة من حيث قياس متوسطها وحساب مقياس تشتتها. ومع ذلك، في الحياة العملية، يحتاج الباحث غالبًا إلى دراسة العلاقة بين ظاهرتين أو متغيرين لفهم مدى ارتباطهما ونوع هذا الارتباط. على سبيل المثال، قد يسعى الباحث إلى معرفة ما إذا كان هناك علاقة بين التدخين والإصابة بأمراض الرئة، أو بين مستوى التعليم والدخل، أو بين الحالة التعليمية والوضع الاجتماعي للناخب.

كما هو واضح، هناك العديد من الأمثلة في مجالات متنوعة، وقد يرغب الباحث في دراسة العلاقة بين أكثر من متغيرين في وقت واحد. على سبيل المثال، قد يهتم الباحث بتأثير مستوى التعليم والدخل وحجم الأسرة على درجة الوعي السياسي للفرد. في هذا السياق، يسعى الباحث لفهم تأثير ثلاثة متغيرات على متغير رابع.

هذا العام، سنركز على دراسة العلاقة بين متغيرين فقط، وهو ما يعرف بالارتباط "البسيط" (Simple Correlation). أما الحالات التي تشمل دراسة أكثر من متغيرين، فتُعرف بالارتباط المتعدد (Multiple Correlation)، وهي خارج نطاق دراستنا لهذا العام.

اولاً / انواع العلاقة بين المتغيرات

تتعدد أنواع العلاقات بين المتغيرات، ويمكن تصنيفها إلى عدة فئات رئيسية:

1- العلاقة الإيجابية او الطردية

- تحدث عندما يزيد أحد المتغيرات بزيادة المتغير الآخر. على سبيل المثال، زيادة مستوى التعليم ترتبط بزيادة الدخل.

2- العلاقة السلبية او العكسية

- تحدث عندما يزيد أحد المتغيرات بينما ينقص الآخر. على سبيل المثال، زيادة التدخين قد تكون مرتبطة بانخفاض مستوى الصحة العامة.

3- العلاقة غير الخطية

- هذه العلاقة لا تتبع نمطًا خطيًا، بل يمكن أن تكون معقدة. على سبيل المثال، قد يرتبط مستوى القلق بأداء أكاديمي بشكل غير خطي، حيث يتسبب القلق المعتدل في تحسين الأداء، بينما يؤدي القلق الزائد إلى تدهور الأداء.

4- العلاقة العشوائية

- تحدث عندما لا توجد علاقة واضحة بين المتغيرين، بحيث لا يؤثر أحدهما على الآخر. على سبيل المثال، قد لا توجد علاقة بين عدد الساعات التي يشاهد فيها الشخص التلفاز ودرجة حرارته.

5- العلاقة السببية

- تحدث عندما يؤثر أحد المتغيرات بشكل مباشر على الآخر. على سبيل المثال، قد يؤدي ارتفاع درجات الحرارة إلى ذوبان الجليد.

6- العلاقة التبادلية

- تحدث عندما يؤثر كل من المتغيرين على الآخر. على سبيل المثال، قد يؤثر مستوى الدخل على مستوى التعليم، وفي نفس الوقت، يؤثر مستوى التعليم على مستوى الدخل.

7- العلاقة المتوسطة

- في بعض الحالات، قد يؤثر متغير ثالث على العلاقة بين متغيرين آخرين. على سبيل المثال، قد يؤثر مستوى الدعم الاجتماعي على العلاقة بين التوتر والصحة النفسية.

فهم هذه الأنواع يساعد الباحثين في تحليل البيانات واستخلاص النتائج الدقيقة حول الظواهر المختلفة. وفي هذا الفصل سوف ندرس درجة الارتباط ومدى تأثير كل متغير على المتغير الاخر .

ثانياً / معامل الارتباط البسيط simple correlation coefficient

يعد معامل الارتباط مقياس إحصائي يُستخدم لتحديد قوة واتجاه العلاقة بين متغيرين. يُعبر عنه عادةً برقم يتراوح بين -1 و +1. إليك تفاصيل أكثر حوله:

أنواع معاملات الارتباط

1. معامل الارتباط بيرسون (Pearson Correlation Coefficient)

- يُستخدم لقياس العلاقة الخطية بين متغيرين مستمرين.
- يتراوح من -1 (ارتباط سلبي كامل) إلى +1 (ارتباط إيجابي كامل).
- 0 يعني عدم وجود علاقة.

2. معامل الارتباط سبيرمان (Spearman's Rank Correlation Coefficient)

- يُستخدم لقياس العلاقة بين متغيرين عندما لا تكون البيانات في توزيع طبيعي.
- يعتمد على ترتيب القيم بدلاً من قيمها الفعلية.
- يتراوح أيضًا من -1 إلى +1.

تفسير معاملات الارتباط

- +1: علاقة إيجابية قوية. كلما زاد أحد المتغيرات، زاد الآخر.
- 0.5 إلى 0.9: علاقة إيجابية متوسطة إلى قوية.
- 0 إلى 0.5: علاقة إيجابية ضعيفة.
- 0: عدم وجود علاقة.
- -0.5 إلى -0.9 : علاقة سلبية ضعيفة إلى متوسطة.
- -1: علاقة سلبية قوية. كلما زاد أحد المتغيرات، انخفض الآخر.

استخدامات معامل الارتباط

- تحليل البيانات: لفهم العلاقة بين المتغيرات المختلفة.
- توقع النتائج: في المجالات مثل الاقتصاد، الطب، وعلم النفس.
- تحسين النماذج: في تحليل الانحدار والتنبؤ.

ملاحظات

- معامل الارتباط لا يدل على السببية؛ فقد يكون هناك عوامل أخرى تؤثر على العلاقة بين المتغيرين.
- يجب استخدامه بحذر مع البيانات التي تحتوي على قيم متطرفة أو توزيع غير طبيعي.

- **معامل الارتباط بيرسون (Pearson Correlation Coefficient)**

معامل الارتباط بيرسون (Pearson Correlation Coefficient) هو مقياس إحصائي يُستخدم لتحديد قوة واتجاه العلاقة الخطية بين متغيرين مستمرين. يُعتبر من أكثر معاملات الارتباط شيوعًا في الإحصاء ويُستخدم في العديد من التطبيقات البحثية.

- القيمة: يتراوح بين -1 و +1:
- +1: ارتباط إيجابي كامل (كلما زاد أحد المتغيرات، زاد الآخر).
- 0 : عدم وجود ارتباط.
- -1 : ارتباط سلبي كامل (كلما زاد أحد المتغيرات، انخفض الآخر).

يمكن حساب معامل الارتباط باستخدام الصيغة التالية:

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

إذ أن :
-(r): معامل الارتباط بيرسون.
-(n): عدد الأزواج من البيانات.
-(x): القيم للمتغير الأول.
-(y): القيم للمتغير الثاني.

- **خطوات الحساب**
1. جمع البيانات : قم بجمع القيم للمتغيرين.
2. حساب المجموعات : احسب مجموع القيم $(\sum x)$،$(\sum y)$،$(\sum xy)$،$(\sum x^2)$،$(\sum y^2)$.
3. تطبيق الصيغة: استخدم القيم المحسوبة في الصيغة أعلاه للحصول على معامل الارتباط.

- **استخدامات معامل الارتباط بيرسون**
  - تحليل العلاقات بين المتغيرات في علم النفس، الاقتصاد، العلوم الاجتماعية، والعديد من المجالات الأخرى.
  - تحسين النماذج التنبؤية وفهم البيانات.

- **ملاحظات**
  - الافتراضات: يجب أن تكون البيانات في توزيع طبيعي وأن تكون العلاقة خطية.
  - الاستجابة للقيم المتطرفة: يمكن أن تؤثر القيم المتطرفة بشكل كبير على معامل الارتباط، لذا يجب الانتباه لها.
  - لا يدل على السببية: معامل الارتباط لا يعني أن أحد المتغيرات يسبب الآخر؛ فقد تكون هناك عوامل أخرى تؤثر على العلاقة.

مثال /

لنفترض أن لدينا بيانات عن درجات الطلاب في اختبارين: اختبار الرياضيات واختبارالاحصاء . احسب معامل الارتباط بيرسون

الطالب	درجات الرياضيات	درجات الاحصاء
1	85	78
2	90	88
3	75	70
4	80	85
5	95	90

الحل /

الطالب	درجات الرياضيات (x)	درجات الاحصاء (y)	xy	x^2	y^2
1	85	78	6630	7225	6084
2	90	88	7920	8100	7744
3	75	70	5250	5625	4900
4	80	85	6800	6400	7225
5	95	90	8550	9025	8100
المجموع	**425**	**411**	**35150**	**36375**	**34053**

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

$$r = \frac{5(35150) - (425)(411)}{\sqrt{[5(38375) - (425)^2][5(34053) - (411)^2]}}$$

- البسط :

$$5 \times 35150 = 175750$$

$$425 \times 411 = 174975$$

$$175750 - 174975 = 775$$

- المقام ، الجزء الاول :

$$5 \times 36375 = 181{,}875$$

$$425^2 = 180625$$

$$181875 - 180625 = 11250$$

-
المقام الجزء الثاني:

$$5 \times 34053 = 170265$$

$$411^2 = 168921$$

$$170265 - 168921 = 1344$$

- الآن، نعود إلى صيغة معامل الارتباط:

$$r = \frac{775}{\sqrt{11250 \times 1344}}$$

- حساب الجذر:

$$\sqrt{11250 \times 1344} = \sqrt{15120000} \approx 3885.62$$

$$r \approx \frac{775}{3885.62} \approx 0.199$$

معامل الارتباط بيرسون بين درجات الرياضيات ودرجات الاحصاء هو تقريبًا 0.199 ، مما يدل على وجود علاقة إيجابية ضعيفة بين الدرجتين.

مثال /أحساب معامل ارتباط بيرسون من البيانات التالية :

y	x	n
65	70	1
75	80	2
85	90	3
80	85	4
95	95	5

الحل /

y^2	x^2	xy	y	x	n
4225	4900	4550	65	70	1
5625	6400	6000	75	80	2
7225	8100	7650	85	90	3
6400	7225	6800	80	85	4
9025	9025	9025	95	95	5
32500	**35650**	**34025**	**400**	**420**	**المجموع**

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

$$r = \frac{5(34025) - (420)(400)}{\sqrt{[5(36650) - (420)^2][5(32500) - (400)^2]}}$$

$$r = \frac{170125 - 168000}{\sqrt{[183,250 - 176,400][162,500 - 160,000]}}$$

$$r = \frac{2,125}{\sqrt{[183,250 - 176,400][162,500 - 160,000]}}$$

$$r = \frac{2{,}125}{\sqrt{6{,}850 \times 2{,}500}}$$

$$r = \frac{2{,}125}{\sqrt{17125000}}$$

$$r = \frac{2{,}125}{4138.23}$$

$$r = \frac{2{,}125}{4138.23}$$

$$r \approx 0.95$$

معامل الارتباط بيرسون هو تقريبًا 0.95 ، مما يدل على وجود علاقة إيجابية قوية بين قيمة (x) و (y) .

- **معامل الارتباط سبيرمان (Spearman's Rank Correlation Coefficient)**

يعد معامل الارتباط سبيرمان مقياس إحصائي يُستخدم لتحديد قوة واتجاه العلاقة بين متغيرين، مع التركيز على ترتيب القيم بدلاً من القيم الفعلية. يُعتبر مفيدًا بشكل خاص عندما تكون البيانات غير موزعة بشكل طبيعي أو تحتوي على قيم متطرفة.

يمكن حسابه باستخدام الصيغة التالية:

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

إذ أن :

- r_s : معامل الارتباط سبيرمان.
- d_i : الفرق بين رتب المتغيرات (للطالب i).
- n : عدد البيانات.

خطوات الحساب

1. ترتيب القيم: قم بترتيب القيم لكل متغير.
2. حساب الفروق: احسب الفرق بين رتب كل زوج من القيم.
3. تطبيق الصيغة: استخدم الفروق المحسوبة في الصيغة للحصول على معامل الارتباط.

مثال / من البيانات جد معامل الارتباط سبيرمان

n	x	y
1	70	65
2	90	75
3	65	85
4	85	80
5	45	95

الحل /

n	x	y	رتب x	رتب y	الفرق بين الرتب d	مربع الفرق d^2
1	70	65	3	1	2	4
2	90	75	5	2	3	9
3	65	85	2	4	-2	4
4	85	80	4	3	1	1
5	45	95	1	5	-4	16
المجموع					**0**	**34**

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(34)}{5(5^2-1)} = 0.7$$

معامل الارتباط سبيرمان $r_s = 0.7$ ، فهذا يشير إلى وجود علاقة إيجابية قوية بين المتغيرين المدروسين.

مثال / من البيانات جد معامل الارتباط سبيرمان

n	x	y
1	70	65
2	80	75
3	80	85

4	90	80
5	95	95

عندما توجد قيم متكررة في بيانات حساب معامل الارتباط سبيرمان، قد يتطلب الأمر استخدام ترتيب الرتب المقيدة (Rank Tied Ranks). إليك كيف يتم التعامل مع القيم المتكررة:

- **كيفية التعامل مع القيم المتكررة**

1. تحديد الرتب:
   - عند وجود قيم متكررة، يتم تعيين نفس الرتبة لجميع القيم المتكررة.
   - بدلاً من ترتيب كل قيمة بشكل فردي، يتم حساب الرتبة المتوسطة للقيم المتكررة.
2. حساب الرتبة المتوسطة:

- إذا كانت لديك قيم متكررة، على سبيل المثال، إذا كانت القيم 80 متكررة مرتين، فإن الرتبة التي تُعطى لكل من هذه القيم ستكون:
  - الرتبة الأولى: 2
  - الرتبة الثانية: 3
- الرتبة المتوسطة: $2.5 = \frac{2+3}{2}$ لذا، تُعطى القيمة 80 الرتبة 2.5. ثم استخراج الفرق بين الرتب ومربع الرتب

الحل /

n	x	y	رتب x	رتب y	الفرق بين الرتب d	مربع الفرق d^2
1	70	65	1	1	0	0
2	80	75	2.5	2	0.5	0.25
3	80	85	2.5	4	-1.5	2.25
4	90	80	4	3	1	1
5	95	95	5	5	0	0
المجموع					**0**	**3.5**

الحل /

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(3.5)}{5(5^2-1)} = 0.8$$

معامل الارتباط سبيرمان $r_s = 0.8$ ، فهذا يشير إلى وجود علاقة إيجابية قوية بين المتغيرين المدروسين.

<u>مثال / من البيانات جد معامل الارتباط سبيرمان</u>

n	x	y	رتب x	رتب y	الفرق بين الرتب d	مربع الفرق d^2
1	70	65	1	1	0	0

2	80	75	3	2	1	1
3	80	85	3	4	-1	1
4	80	80	3	3	0	0
5	90	95	5	5	0	0
المجموع					**0**	**2**

الحل /

الرتبة المتوسطة

$$\frac{2+3+4}{3} = 3$$

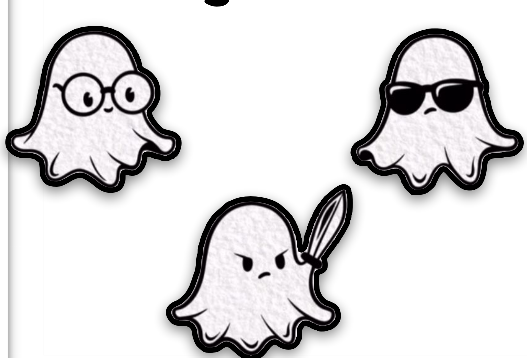
$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(2)}{5(5^2-1)} = 0.9$$

الملخص /

مستويات قوة العلاقة حسب معامل الارتباط سبيرمان (**Spearman's Rank Correlation Coefficient**) تُقسم عادةً إلى الفئات التالية:

1. علاقة ضعيفة : r_s من 0 إلى 0.2 ، تشير إلى وجود علاقة ضعيفة أو غير مهمة بين المتغيرين.
2. علاقة متوسطة : r_s من 0.2 إلى 0.4 ، تشير إلى وجود علاقة متوسطة، حيث يمكن أن تكون هناك بعض التوجهات بين المتغيرين.
3. علاقة قوية : r_s من 0.4 إلى 0.6 ، تشير إلى وجود علاقة قوية، حيث تكون هناك توافق كبير في التغيرات بين المتغيرين.
4. علاقة قوية جدًا : r_s من 0.6 إلى 0.8 ، تشير إلى وجود علاقة قوية جدًا، مما يعني أن التغيرات في أحد المتغيرات ترتبط ارتباطًا وثيقًا بالتغيرات في المتغير الآخر.
5. علاقة ارتباط كاملة : r_s من 0.8 إلى 1 ، تشير إلى وجود علاقة إيجابية كاملة، حيث كل زيادة في أحد المتغيرات يقابلها زيادة في الآخر بشكل متناسق.
6. علاقة سلبية (في الاتجاه المعاكس) : القيم السلبية (مثل -0.8، -0.5، -0.2) تعكس نفس مستويات القوة ولكن في اتجاه سلبي، مما يعني أن زيادة أحد المتغيرات ترتبط بانخفاض في الآخر.

الانحدار الخطي
البسيط



الانحدار الخطي البسيط simple linear regression

يعد الانحدار الخطي البسيط تقنية إحصائية تستخدم لنمذجة العلاقة بين متغيرين: متغير مستقل (X) ومتغير تابع (Y). يهدف الانحدار الخطي البسيط إلى إيجاد خط مستقيم يصف العلاقة بين المتغيرين بحيث يمكن استخدامه للتنبؤ بقيم المتغير التابع بناءً على قيم المتغير المستقل.

تُعبر المعادلة العامة للانحدار الخطي البسيط عن العلاقة بين المتغيرين كما يلي:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

حيث:

- Y : المتغير التابع.
- X : المتغير المستقل.
- β_0: الجزء الثابت (المرتبط بالخط).
- β_1: معامل الانحدار (يمثل التغير في (Y) مع تغيير وحدة واحدة في (X)).
- ϵ : الخطأ العشوائي أو التباين غير المفسر.

- **خطوات تنفيذ الانحدار الخطي البسيط**

1. جمع البيانات: احصل على مجموعة من البيانات التي تحتوي على قيم للمتغيرين المستقل والتابع.
2. تحليل البيانات: قم بإجراء تحليل استكشافي للبيانات، مثل رسم scatter plot لفهم العلاقة بين المتغيرين.
3. تقدير المعاملات: استخدم أساليب إحصائية، مثل طريقة المربعات الصغرى (Ordinary Least Squares)، لتقدير القيم المناسبة لـ β_0 و β_1.
4. تقييم النموذج:
   - استخدم معايير مثل R^2 لتحديد مدى جودة النموذج في تفسير التباين في البيانات.
   - تحقق من افتراضات الانحدار الخطي (مثل خطية العلاقة، استقلال الأخطاء، وتوزيع الأخطاء).
5. التنبؤ: استخدم النموذج للتنبؤ بقيم جديدة لـ(Y) بناءً على قيم جديدة لـ(X).

- تفسير النتائج

- معامل الانحدار β_1 :

- إذا كان إيجابيًا، فهذا يعني أن هناك علاقة إيجابية بين(X) و(Y) (مع زيادة(X) تزيد (Y)).
- إذا كان سالبًا، فهذا يعني أن هناك علاقة سلبية (مع زيادة(X) تنخفض (Y)).

- الجزء الثابت β_0 :

- يمثل قيمة(Y) عندما تكون(X = 0).

لحساب معامل الانحدار (Slope) ومعامل التقاطع (Intercept) لخط الانحدار الخطي البسيط، يمكنك استخدام صيغة المربعات الصغرى (Ordinary Least Squares).

من الصيغ التالية يمكن الوصول الى معادلة الانحدار الخطي البسيط

1- حساب المتوسطات لكل من X,Y

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{Y} = \frac{\sum Y}{n}$$

2- حساب معامل الانحدار (Slope) يمكن حساب معامل الانحدار باستخدام الصيغة التالية:

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

3- ثم يمكنك حساب معامل التقاطع باستخدام الصيغة التالية :

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

ويمكن التعويض معامل الانحدار ومعامل التقاطع في معادلة الانحدار الخطي البسيط

$$Y = \beta_0 + \beta_1 X + \epsilon$$

<u>مثال/ من البيانات التالية ، جد معادلة الانحدار الخطي البسيط</u>

X	Y
30	20
40	25
50	30
60	40
70	50

الحل /

X	Y	$X - X'$	$Y - Y'$	$(X - X')(Y - Y')$	$(X - X')^2$
30	20	-20	-13	260	400
40	25	-10	-8	80	100
50	30	0	-3	0	0
60	40	10	7	70	100

400	340	17	20	50	70
$\sum X-X')=$ **1000**	$\sum (X-X')(Y-Y')=$ **750**			$\sum Y = 165$	$\sum X = 250$

1- حساب المتوسطات

$$\bar{X} = \frac{250}{5} = 50$$
$$\bar{Y} = \frac{165}{5} = 33$$

2- حساب معامل الانحدار (Slope) يمكن حساب معامل الانحدار

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\beta_1 = \frac{750}{1000} = 0.75$$

3- ثم يمكنك حساب معامل التقاطع :

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$
$$= 33 - 0.75 \times 50$$
$$= 33 - 37.5 = -4.5$$

- $\beta_1 = 0.75$
- $\beta_0 = -4.5$

بذلك، المعادلة النهائية للانحدار الخطي هي:

$$Y = -4.5 + 0.75X$$

- **معامل الانحدار ، القيمة:** $\beta_1 = 0.75$

التفسير:

- يمثل معامل الانحدار مقدار التغير في المتغير التابع (الإنفاق) عندما يتغير المتغير المستقل (الدخل) بمقدار وحدة واحدة.
- في هذا المثال، يعني أن كل زيادة بمقدار 1000 دولار في الدخل تُتوقع أن تؤدي إلى زيادة في الإنفاق بمقدار 750 دولار.
- يدل أيضًا على وجود علاقة إيجابية بين الدخل والإنفاق؛ فكلما زاد الدخل، زاد الإنفاق.

- **معامل التقاطع، القيمة** $\beta_0 = -4.5$

التفسير:

- يمثل معامل التقاطع القيمة المتوقعة للمتغير التابع (الإنفاق) عندما يكون المتغير المستقل (الدخل) مساويًا للصفر.
- في هذا السياق، يعني أن إذا كان الدخل صفرًا، فإن الإنفاق المتوقع سيكون -4.5 (وهو عدد غير منطقي في سياق الإنفاق، حيث لا يمكن أن يكون الإنفاق سالبًا).

- هذا يشير إلى أن هناك نفقات ثابتة أو نفقات أخرى لا تعتمد على الدخل، وقد يكون هذا نتيجة للتكاليف الثابتة أو النفقات الأساسية التي يتحملها الأفراد.

- **خلاصة**

- معامل الانحدار يعطيك فكرة عن كيفية تغير الإنفاق مع تغير الدخل.
- معامل التقاطع يوفر نقطة البداية للمعادلة، على الرغم من أن قيمته السلبية قد لا تكون لها دلالة عملية في هذا السياق.

مثال / ماهوتأثير عدد ساعات الدراسة على درجات الاختبار ، بين ذلك من خلال معادلة الانحدار الخطي البسيط من خلال البيانات التالية :-

الطالب	عدد ساعات الدراسة	درجة الاختبار
1	2	65
2	3	70
3	4	75
4	5	80
5	6	85

الحل /

X	Y	$X-X'$	$Y-Y'$	$(X-X')(Y-Y')$	$(X-X')^2$
2	65	-2	-10	20	4
3	70	-1	-5	5	1
4	75	0	0	0	0
5	80	1	5	5	1
6	85	2	10	20	4
20	**375**			**50**	**10**

$$\bar{X} = \frac{20}{5} = 4$$

$$\bar{Y} = \frac{375}{5} = 75$$

$$\beta_1 = \frac{50}{10} = 5$$

$$\beta_0 = 75 - 5 \times 4 = 75 - 20 = 55$$

$$Y = 55 + 5X$$

<u>تفسير النتائج</u>

- معامل الانحدار :$\beta_1 = 5$ يعني أنه مع زيادة عدد ساعات الدراسة بمقدار ساعة واحدة، يُتوقع أن تزيد درجة الاختبار بمقدار 5 درجات.
- معامل التقاطع : $\beta_0 = 55$ يعني أن الطالب الذي لا يدرس (عدد ساعات الدراسة = 0) يُتوقع أن يحصل على درجة 55 في الاختبار.
- العلاقة طردية بين المتغيرين .