الجمهورية الجزائرية الديمقراطية الشعبية

جامعة غرداية

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قسم العلوم المالية والمحاسبة

محاضرات في الإحصاء 01

موجهة لطلبة السنة أولى جذع مشرك

الدكتور: عبدالرؤوف عباده

أستاذ محاضر صنف أ، جامعة غرداية

abada.abderraouf@univ-ghardaia.dz

السنة الجامعية 2024/2023

المفاهيم والمبادئ الإحصائية

أولا :مفهوم الإحصاء :

الإحصاء مجموعة من الإجراءات والأساليب المستخدمة في جمع وعرض تحليل البيانات التي تبنى عليها القرارات في مواجهة عدم التأكد أو في مواجهة معلومات ناقصة .وفي الوقت الحاضر نجد أن التحليل الإحصائي يستخدم تقريبا في كل مهنة .فالاقتصادي يستخدم لاختبار كفاءة الإنتاج المختلفة، ورجل الأعمال قد يستخدمه لاختبار تصميم أو تغليف المنتج بما يعظم المبيعات، والباحث الاجتماعي يستخدمه لتحليل نتائج عقار معين على برنامج تأهيلي، وعالم النفس الصناعي لدراسة استجابات العمال لظروف العمل بالمصنع، والعالم السياسي للتنبؤ بأنماط التصويت، والطبيب لاختبار فعالية عقار جديد، والكيمائي لإنتاج أسمدة أرخص، وهكذا.

ثانيا :الإحصاء الوصفي Descriptive Statistics :يختص بوصف خصائص البيانات المستخدمة في البحث الإحصائي، فإذا كانت لدينا بعض البيانات الخاصة بظاهرة معينة، فعلم الإحصاء الوصفي بين لنا كيف يتم توزيع هذه البيانات وما إذا كانت تتمركز حول قيمة معينة أم إنها متباينة وإذا كانت هناك علاقة ين الظاهرة وظاهرة أخرى وما قوة هذه العلاقة[1].

ثالثا :المجتمع الإحصائي: Population وهو عبارة عن مجموع الوحدات أو الأفراد التي تشترك في صفة، والمقصود بالمجموع، المجموعة من الناحية الرياضية، مثل :مجموعة الطلبة، مجموعة الأدوات، مجموعة الحيوانات، مجموعة الكراسي.... الخ .حيث نقسم بدوره إلى قسمين:

أ- **المجتمع المحدود Finite population** :وهو المجتمع الذي يمكن حصر عدد مفرداته، كما هو الحال في أطوال موظفي المؤسسة أو عدد الوحدات الإنتاجية في مصنع ما في يوم معين.

ب- **المجتمع غير المحدود Infinite population** :وهو المجتمع الذي من الصعب أو المستحيل حصر مفرداته مثل :الأسماك في النهر أو البحر أو النجوم في الفضاء الخارجي.

رابعا :الوحدة الإحصائية Statistical unit :هي الوحدة الأساسية لتكوين المجتمع الإحصائي أو هي العنصر الأولي محل الملاحظة مهما كانت طبيعته.

خامسا :العينة Sample :وهي الجزء الذي يمثل المجتمع أحسن تمثيل ويختلف حجمها حسب أهمية الدراسة وحسب الإمكانيات المادية والبشرية المتاحة للقيام بهذه الدراسة وأسلوب العينة متبع في اغلب

---

[1]- إسماعيل محمد بن قانة، **الإحصاء الوصفي والحيوي " دروس وتطبيقات "**، دار أسامة للنشر والتوزيع، الطبعة الأولى، الأردن، 2011، ص 9.

الدراسات الميدانية وذلك لاستحالة جمع المعلومات الإحصائية من كل الوحدات التي تشكل المجتمع المدروس أو ما يسمى بالحصر الشامل.

سادسا :المتغيرات الإحصائية

المتغيرات: هي الخاصية المدروسة أو الظاهرة الإحصائية المدروسة في المجتمع الإحصائي.

تنقسم المتغيرات الإحصائية إلى قسمين:

1-متغيرات كيفية: هي تلك المتغيرات التي لاينكن قياسها أو غير قابلة للقياس مثل الجنسية، الحالة العائلية، الحالة المدنية .... الخ.

2-متغيرات كمية: هي تلك الخصائص التي يمكن قياسها وهي أكثر المتغيرات انتشارا واستعمالا لأن لغة الإحصاء هي لغة الأرقام، مثال لذلك: الإنتاج، الاستهلاك، الاستثمار، الوزن .... الخ. وتنقسم المتغيرات الكمية بدورها إلى قسمين: متغيرات منفصلة (متقطعة)، متغيرات مستمرة.

أ-المتغيرات منفصلة (متقطعة): وهي تلك المتغيرات التي تأخذ قيما صحيحة، لا يمكن تجزئتها مثلا عدد الأطفال في العائلة، عدد قطع الغيار المنتجة .... الخ.

ب-متغيرات مستمرة: وهي تلك المتغيرات التي تأخذ كل القيم الممكنة لمجال الدراسة، ونظرا للعدد غير المتناهي لهذه القيم تقسم مجال الدراسة إلى مجالات جزئية تسمى الفئات.

مثال: حدد المجتمع الإحصائي، الوحدة الإحصائية، الصفة وطبيعتها والمتغير الإحصائي ونوعه للمعطيات التالية:

- الأجور السنوية للمعلمين في مدرسة معينة.
- لون حقائب الطلبة في قسم معين.
- الرياضة المفضلة لكل طالب في مدرسة ما.
- عدد الأطفال في كل أسرة داخل حي سكني معين.

الحل: الجدول التالي يحدد كل من المجتمع، الوحدة، الصفة، والمتغيرة الإحصائية

المجتمع الإحصائي	الوحدة الإحصائية	طبيعة الصفة	المتغيرة الإحصائية
المعلمين	المعلم	كمية	مستمرة: الأجور
الحقائب	الحقيبة	كيفية: اللون	-
الطلبة	الطالب	كيفية: الرياضة	-
الأسر	أسرة	كمية	متقطعة: عدد الأطفال

تنظيم وعرض البيانات

يمكن تنظيم البيانات في شكلين أساسيين هما الجداول الإحصائية والرسوم البيانية، والجداول الإحصائية الأكثر استخداما هي جداول البيانات المبوبة في فئات، أما الجدول التكراري المزدوج فهو عبارة عن جدول تكراري يجمع ببين متغيرين، وهو نادرا ما يستخدم.

أولا :العرض الجدولي:

يقوم الباحث بتفريغ النتائج المتحصل عليها في جداول أولية بعد الانتهاء من عملية جمع البيانات، ثم يلخصها في جداول نهائية تسمى جداول المعطيات:

1-جداول التوزيع التكراري البسيط

هو عبارة عن وضع المعلومات الإحصائية في جداول نهائية، يحتوي كل منها على عمودين (أو سطرين) بين العمود (السطر) الأول قيم الظاهرة أو المتغير المدروس، وتكون هذه القيم على شكل قيم نقطية أو على شكل مجالات، أما العمود (السطر) الثاني فيحتوي على تكرارات هذه القيم أو المجالات.

1-1-جدول التكرارات في حالة بيانات المتغيرات الكمية المنفصلة (متقطعة)

المتغير الكمي المنفصل (المتقطع) وهو ذلك المتغير الذي لا يمكن تجزئة قيمه، حيث يأخذ قيما صحيحة.

مثال: لدراسة متوسط عدد الأطفال في العائلة في مجتمع ما، أخذت عينة عشوائية حجمها n=30 عائلة، فكانت النتائج التالية:

5	1	4	3	2	2	0	5	4	4
1	3	2	4	0	5	4	3	2	1
4	2	1	5	2	4	3	1	4	0

المطلوب: تبويب هذه البيانات في جدول توزيع تكراري؟

المجموع	5	4	3	2	1	0	عدد الأطفال (x_i)
	IIII	~~IIII~~ III	IIII	~~IIII~~ I	~~IIII~~	IIII	التفريغ
20	4	8	4	6	5	3	عدد العائلات (n_i)

حيث x_i :قيم المتغير الإحصائي، n_i تكرارات هذه القيم

1-2-جدول التوزيع التكراري في حالة المتغيرات الكمية المتصلة (مستمرة)

تبويب البيانات الكمية (المستمرة) المتصلة في شكل جدول تكراري بسيط يتطلب تكوين جدول تفريغ مكون من ثلاثة أعمدة، العمود الأول فيمثل ما يطلق عليه الفئات، والعمود الثاني يمثل العلامات والعمود الثالث يمثل التكرارات

الخطوات المتبعة لبناء التوزيع التكراري:

أولا: المدى للبيانات

يتم تقسم المدى بين أصغر قيمة تأخذها الظاهرة وأكبر قيمة تأخذها الظاهرة إلى عدة تقسيمات فئات المدى :أكبر قيمة للظاهرة أصغر قيمة للظاهرة.

ثانيا: عدد الفئات

يجب أن يكون عدد الفئات وأطوال كل فئة من الفئات مناسب، بحيث لا يكون عدد الفئات كبيرا أو طول الفئة صغير وبالتالي يفقد التلخيص أهميته، وألا يكون عدد الفئات صغيرا وأطوال الفئات كبير فيفقد التوزيع كثير من تفاصيله ويمكن حساب عدد الفئات كما يلي:

$$k = 1 + 3.32 \log n$$

حيث :K :عدد الفئات

1 و3.32: ثوابت

$\log n$: لوغاريتم عشري لحجم العينة[2].

ثالثا: طول الفئة

يرمز لو بالرمز Δ وهو حاصل قسمة المدى العام على عدد الفئات، ثم تقريب الجواب على أعلى بحيث يكون عدد الأرقام المعنوية يساوي أو يقل عن عدد الأرقام المعنوية المستعملة في البيانات.

$$\Delta = \frac{\text{المدى}}{\text{عدد الفئات}}$$

طول الفئة :الحد الأعلى للفئة – الحد الأدنى للفئة

رابعا: نعين الحد الأدنى لأول فئة ويجب أن يكون هذا الحد أصغر من أو مساويا لأصغر قيمة في البيانات.

[2]- يستخرج اللوغاريتم العشري للعينة بالآلة الحاسبة وذلك بكتابة حجم العينة ثم الضغط على الزر log n

خامسا: نعين الحد الأعلى للفئة الأولى وذلك بإضافة طول الفئة إلى الحد الأدنى لتلك الفئة

الحد الأعلى للفئة الأولى = الحد الأدنى للفئة الأولى + طول الفئة

سادسا: نعين الحدود الدنيا والحدود العليا لجميع الفئات الباقية وذلك بإضافة طول الفئة لكل حد .

سابعا: نعين مراكز الفئات :نجد المركز لكل فئة بقسمة مجموع حديها على 2

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{2}$$

ويمكن إيجاد مراكز الفئات الأخرى بإضافة طول الفئة إلى مركز الفئة التي قبلها

ثامنا: نفرغ البيانات المعطاة لدينا على الفئات التي أنشأنها وذلك باستعمال خط عمودي لكل قراءة وخط أفقي للقراءة الخامسة في كل فئة وذلك لتسهيل جمع التكرارات.

مثال: إذا كانت لدينا البيانات التالية لمتغير كمي متصل:

59	60	48	45	46	43	42	33	34	43
54	49	30	33	40	45	42	44	43	42
59	60	40	44	35	37	39	50	41	52

المطوب: تبويب هذه البيانات في جدول توزيع تكراري

الحل:

1-إيجاد المدى :المدى للبيانات = أكبر قيمة -أصغر قيمة

30 − 60 = 30

2-عدد الفئات: K=1+3.32 log n

الملاحظة أن حجم العينة30 =n ، وعليه فإن:

$$\log(30) = 1.477$$
$$k = 1 + (3.32 \times 1.477) = 5.90 \cong 6$$

وبما أن البيانات معطاة لأقرب عدد صحيح (لا يوجد كسور عشرية) نأخذ6=k

3-إيجاد طول الفئة:

$$\Delta = \frac{\text{المدى}}{\text{عدد الفئات}}$$

$$\Delta = \frac{30}{6} = 5$$

4–تحديد جدول الفئات

الفئة الأولى :الحد الأدنى للفئة الأولى هو أصغر قيمة في البيانات وهو30

الحد الأعلى للفئة الأولى = الحد الأدنى للفئة الأولى + طول لفئة

الحد الأعلى للفئة الأولى هو 30 + 5 = 35، وعليه فإن الفئة الأولى هي: [30 – 35[

الفئة الثانية :الحد الأدنى للفئة الثانية هو35

الحد الأعلى للفئة الثانية هو 35 + 5 = 40، وعليه فإن الفئة الأولى هي: [35 – 40[

ونستمر بهذه الطريقة حتى نكون الست فئات.

5–تعيين مركز الفئات X_j

$$\text{مركز الفئة الأولى} = \frac{\text{الحد الأدنى الفئة الأولى} + \text{الحد الأعلى الفئة الأولى}}{2}$$

$$x_1 = \frac{30+35}{2} = 32.5$$

مركز الفئة الثانية = مركز الفئة الأولى + طول الفئة

$$x_2 = 32.5 + 5 = 37.5$$

مركز الفئة الثالثة:

$$x_3 = 37.5 + 5 = 42.5$$

وهكذا......

6–نفرغ البيانات المعطاة في جدول التكراري

الفئات	مركز الفئة X_j	إفراغ البيانات	التكرار n_i
[35 – 30]	32.5	IIII	4
[40 – 35]	37.5	III	3
[45 – 40]	42.5	I ~~IIII~~ ~~IIII~~	11
[50 – 45]	47.5	~~IIII~~	5
[55 – 50]	52.5	III	3
[60 – 55]	57.5	IIII	4
المجموع (∑)			30

تساعد الجداول الإحصائية البسيطة أو جداول التوزيع التكراري الباحث على تكوين فكرة سريعة ومبسطة عن توزيع الظاهرة المدروسة، وللتدقيق أكثر يلجأ الباحث إلى حساب التكرار النسبي والتكرار النسبي المئوي.

2- التكرارات النسبية:

إن التكرار النسبي لكل فئة هو نسبة تكرار تلك الفئة إلى مجموع التكرارات، ويرمز له بالرمز P_i

$$P_i = \frac{n_i}{N}$$

حيث أن: $N = \sum n_i$ مجموع التكرارات، و n_i تكرار لكل فئة i .

إن الجدول الذي يعطينا الفئات (أو مراكزها) مع تكراراتها النسبية يسمى التوزيع التكراري النسبي.

الفئات	التكرار النسبي P_i
$]35 - 30]$	0.13
$]40 - 35]$	0.1
$]45 - 40]$	0.37
$]50 - 45]$	0.17
$]55 - 50]$	0.1
$]60 - 55]$	0.13
المجموع $(\sum)$	1

نلاحظ أن مجموع التكرارات النسبية يجب أن يساوي 1 أي أن$\sum \boldsymbol{P_i} = 1$ لأن $\sum n_i = N$

3- التكرار النسبي المئوي

يرمز لو بالرمز $P_i\%$ ، وهو يساوي تكرار الفئةf_x ، مقسم على حجم العينة N، ضرب الناتج في100.

$$P_i\% = \frac{n_i}{N} \times 100$$

التكرار النسبي المئوي $P_i\%$	الفئات
13	$]35-30]$
10	$]40-35]$
37	$]45-40]$
17	$]50-45]$
10	$]55-50]$
13	$]60-55]$
100	المجموع $(\sum)$

ويمكن وضع ما سبق في جدول واحد كما هو موضح في الجدول التالي

التكرار النسبي المئوي $P_i\%$	التكرار النسبي P_i	التكرار n_i	مركز الفئة X_j	الفئات
13	0.13	4	32.5	$]35-30]$
10	0.1	3	37.5	$]40-35]$
37	0.37	11	42.5	$]45-40]$
17	0.17	5	47.5	$]50-45]$
10	0.1	3	52.5	$]55-50]$
13	0.13	4	57.5	$]60-55]$
100	1	30		المجموع $(\sum)$

التوزيع التكراري المتجمع:

في كل من الأحيان نحتاج لمعرفة عدد المفردات (التكرارات) التي تقل قيمتها عن حد معين أو تلك التي تساوي قيمتها أو تزيد عن حد معين ويتم ذلك عن طريق التكراري المتجمع حيث ينقسم إلى نوعين:

- التكرار المتجمع الصاعد.
- التكرار المتجمع النازل.

أولا: التكرار المتجمع الصاعد: التكرار المتجمع الصاعد لأي فئة أو قيمة من قيم المتغير هو عبارة عن تكرار هذه الفئة (القيمة) ضافا إليه مجموع تكرارات الفئات السابقة (القيم السابقة)، ويرمز له بالرمز $n_i \uparrow$

ثانيا: التكرار المتجمع النازل :التكرار المتجمع النازل لأي فئة أو قيمة من قيم المتغير هو عبارة عن مجموع التكرارات $\sum n_i$ مطروحا منه تكرارات الفئات السابقة (القيم السابقة)، ويرمز له بالرمز $n_i \downarrow$

مثال : نفس المثال السابق، المطلوب وضع التكرار المتجمع الصاعد والنازل.

الفئات	مركز الفئة X_j	التكرار n_i	$n_i \uparrow$	$n_i \downarrow$
]35 − 30]	32.5	4	4	30
]40 − 35]	37.5	3	7	26
]45 − 40]	42.5	11	18	23
]50 − 45]	47.5	5	23	12
]55 − 50]	52.5	3	26	7
]60 − 55]	57.5	4	30	4
المجموع ($\sum$)		30		

أما إذا استعملنا التكرارات النسبية بدلا من التكرارات فإننا نحصل على التوزيع التكراري المجتمع النسبي ونحصل على هذه النتيجة نفسها لو قسمنا التكرار المجتمع لكل فئة على مجموع التكرارات.

وكذلك إذا استعملنا التكرارات المئوية بدلا من التكرارات فإننا نحصل على التوزيع التكراري المجتمع المئوي.

1-3-جدول التكرارات في حالة بيانات المتغيرات الكيفية

إنشاء جداول تكرارية بسيطة في حالة البيانات الكيفية يتطلب إنشاء جدول التفريغ من ثلاثة أعمدة يطلق على العمود الأول عمود الصفات (توضع به الصفات حسب تنسيق أو ترتيب معين)، والعمود الثاني يوضح العلامات وذلك بوضع علامة (I) أمام كل صفة ونستمر في عملية التبويب (التفريغ) حتى ننتهي من تفريغ جميع الصفات، ويلاحظ أن كل خمسة علامات تكون ما يسمى بالحزمة IIII ، لتسهيل عد العلامات، أما العمود الثالث والذي نطلق عليه التكرار ويمثل عدد العلامات أمام كل صفة من الصفات.

مثال :فيما يلي تقديرات في مادة الإحصاء لعدد 30 طالبا في إحدى الفرق الدراسية:

ممتاز	جيد	ضعيف	مقبول	مقبول
جيد	ضعيف	ضعيف	جيد جدا	**ممتاز**

مقبول	جيد جدا	ضعيف	جيد	مقبول
جيد جدا	جيد جدا	جيد	مقبول	ضعيف
ممتاز	جيد	مقبول	مقبول	جيد
جيد	جيد	جيد	جيد	جيد

التقديرات هنا عبارة عن صفات يمكن تبوبها على أساس هذه الصفات كما يلي

الصفة (التقدير)	عملية التفريغ	التكرار
ممتاز	III	3
جيد جدا	IIII	4
جيد	I ~~IIII~~ ~~IIII~~	11
مقبول	II ~~IIII~~	7
ضعيف	~~IIII~~	5
إجمالي التكرارات		30

2-جدول التوزيع التكراري المزدوج

يستعمل الجدول المزدوج عند دراسة خاصيتين في نفس الوقت في مجتمع ما، وتوضع المعلومات الإحصائية بالشكل التالي: نضع أفقيا (الأسطر) الخاصية الأولى، ونضع عموديا (الأعمدة) الخاصية الثانية[3].

نرمز لقيم الخاصية الأولى بالرمز y_j حيث j تتراوح من 1 إلى m، وبالرمز x_i لقيم الخاصية الثانية، حيث i يأخذ القيم من 1 إلى n.

يمكن وضع الجدول المزدوج بأربعة أبعاد حسب الشكل التالي:

	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$	$\boldsymbol{y_4}$	$\boldsymbol{n_i}$
$\boldsymbol{x_1}$	n_{11}	n_{12}	n_{13}	n_{14}	$\boldsymbol{n_1}$
$\boldsymbol{x_2}$	n_{21}	n_{22}	n_{23}	n_{24}	$\boldsymbol{n_2}$
$\boldsymbol{x_3}$	n_{31}	n_{32}	n_{33}	n_{34}	$\boldsymbol{n_3}$
$\boldsymbol{x_4}$	n_{41}	n_{42}	n_{43}	n_{44}	$\boldsymbol{n_4}$
$\boldsymbol{n_{.j}}$	$\boldsymbol{n_{.1}}$	$\boldsymbol{n_{.2}}$	$\boldsymbol{n_{.3}}$	$\boldsymbol{n_{.4}}$	$\sum \boldsymbol{n_i} = \sum \boldsymbol{n_{.j}}$

[3]- جلاطو جيلالي، **الإحصاء مع تمارين محلولة**، الطبعة التاسعة، ديوان المطبوعات الجامعية، الجزائر، 2012، ص 18.

مثال لذلك: لدراسة مستوى معيشة العائلات يمكن أن نتطرق إلى خاصيتين، مهنة رب العائلة والإنفاق الاستهلاكي، نلاحظ الخاصية الأولى كيفية أما الخاصية الثانية فإنها قابلة للقياس (كمية).

ثانيا العرض البياني

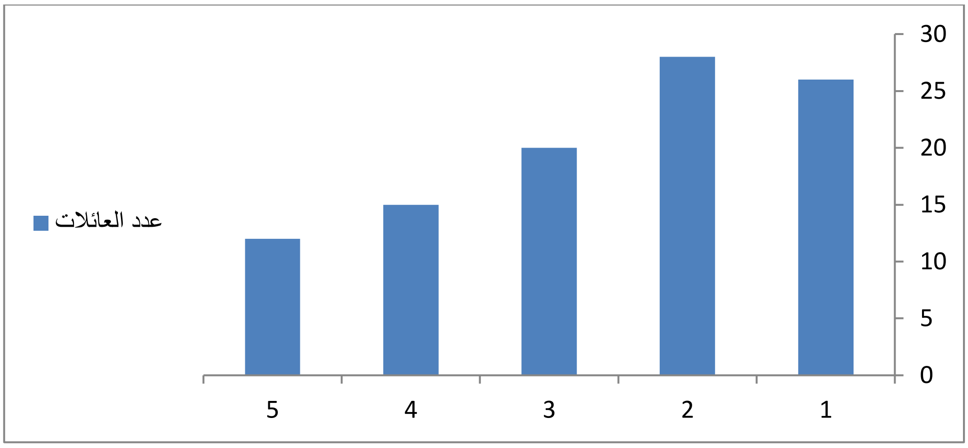
1- العرض البياني في حالة متغير كمي منفصل

1-1- العرض البياني للتكرارات البسطة: هو عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة معينة للمتغير المدروس وتسمى بالأعمدة.

مثال1: مثل بيانيا بيانات الجدول التوزيع التكراري لعدد الأطفال في العائلة لعينة تتكون من 100 عائلة

عدد الأطفال	1	2	3	4	5	المجموع
عدد العائلات	26	28	20	15	12	100

التمثيل البياني:



2- العرض البياني في حالة متغير كمي متصل (مستمر)

إن العروض البيانية للمتغير الكمي المتصل هي أكثر العروض استعمالا ومن أهمها:

2-1- المدرج التكراري (Frequency Histogram):

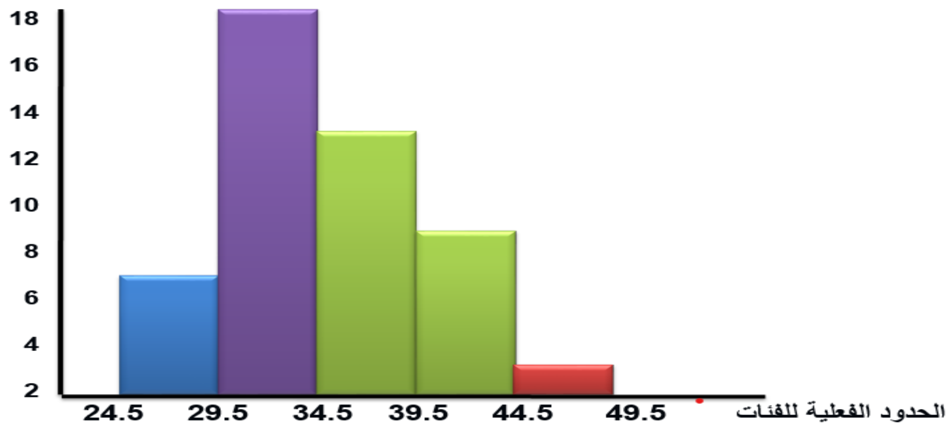
المدرج التكراري هو التمثيل البياني للجدول التكراري البسيط الخاص بالبيانات الكمية، وهو عبارة عن أعمدة بيانية متلاصقة، حيث تمثل التكرارات على المحور العمودي، بينما تمثل قيم المتغير (الفئات) على المحور الأفقي، ويتم تمثيل كل فئة بعمود ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول ال الفئة[4].

ملاحظة: إذا كانت أطوال الفئات غير متساوية فإنه يتوجب حساب التكرار المعدل وفق القانون التالي:

$$\text{التكرار المعدل} = \frac{\text{تكرار الفئة}}{\text{طول الفئة}}$$

مثال: أرسم المدرج التكراري للتوزيع التكراري المعطي في الجدول التالي:

الفئات	]29.5–24.5]	]34.5–29.5]	]39.5–34.5]	]44.5–39.5]	[49.5–44.5]	المجموع
التكرار	7	18	13	9	3	50

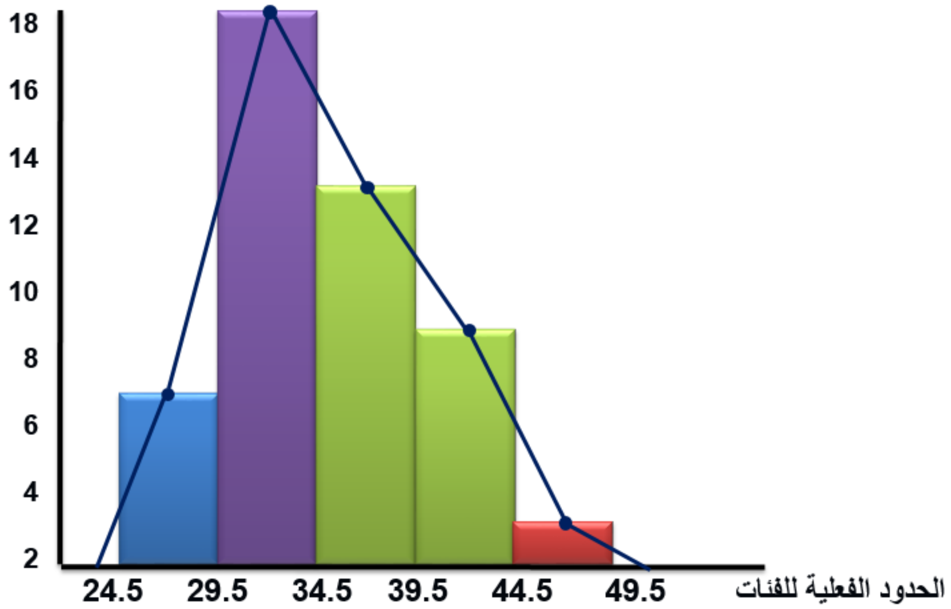


2_2- المضلع التكراري (Frequency Polygon)

المضلع التكراري هو مضلع مغلق نحصل عليه بتنصيف الأضلاع العلوية للمستطيلات في المدرج التكراري ثم يوصل هذه النقاط بعضها مع البعض، ولكي نغلق الخط المنكسر الذي حصلنا عليه نعتبر أن هناك فئتين متطرفتين واحدة إلى أقصى اليسار والثانية إلى أقصى اليمين وتكرار كل منها صفر أي

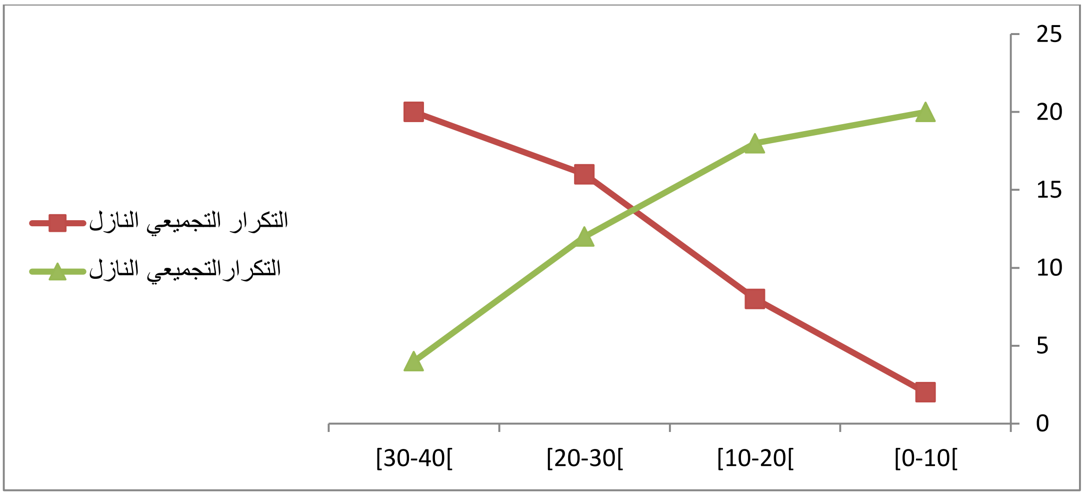
4- عبان عبدالقادر، **الإحصاء الوصفي للعلوم الاجتماعية والعلوم الاقتصادية**، الطبعة الأولى، دار ومضة للنشر والتوزيع والترجمة، الجزائر، 2022، ص 38.

ارتفاع كل من المستطيلين المقامين على هاته الفئتين صفر، نأخذ مركز هاتين الفئتين ونغلق المضلع[5]، كما في الشكل التالي الذي يمثل المضلع التكراري للتوزيع التكراري السابق.



2-3- منحنى التكرارات المتجمعة الصاعدة والنازلة

يتم رسم منحنى التكرار المتجمع الصاعد عن طريق إيصال مجموعة من النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرارات المتجمعة الصاعدة المقابل لها، أما منحنى التكرار المتجمع النازل فيتم رسمه بإيصال مجموعة النقاط ذات الإحداثيات: الحدود الدنيا للفئات والتكرارات المتجمعة النازلة المقابلة لها، وتسمى نقطة التقاطع بينهما بالوسيط.



[5]- محمد صبحي أبو صالح، عدنان محمد عوض، **مقدمة في الإحصاء**، ديوان المطبوعات الجامعية، الجزائر، 1984، ص 18.

3- العرض البياني في حالة متغير كيفي:

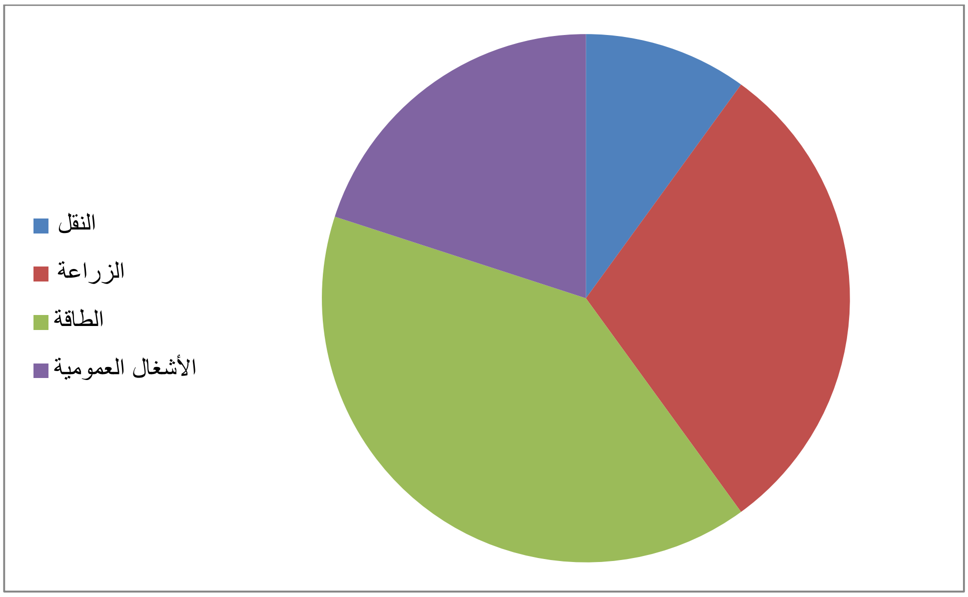
3-1-العرض الدائري:

العرض الدائري هو عبارة عن دائرة مقسمة إلى عدة اجزاء كل جزء يقابل زاوية مركزية تتناسب مع التكرارات المقابلة لكل خصية من الخصائص المدروسة، حيث نقوم بإضافة عمود يحتوي على الزاوية المركزية المقابلة لكل تكرار.

يتم حساب الزوايا المركزية باستخدام العلاقة التالية:

$$\text{الزاوية المركزية} = \frac{\text{تكرار الخاصية}}{\text{مجموع التكرارات}} \times 360^{\circ} = 360^{\circ} \times n_i$$

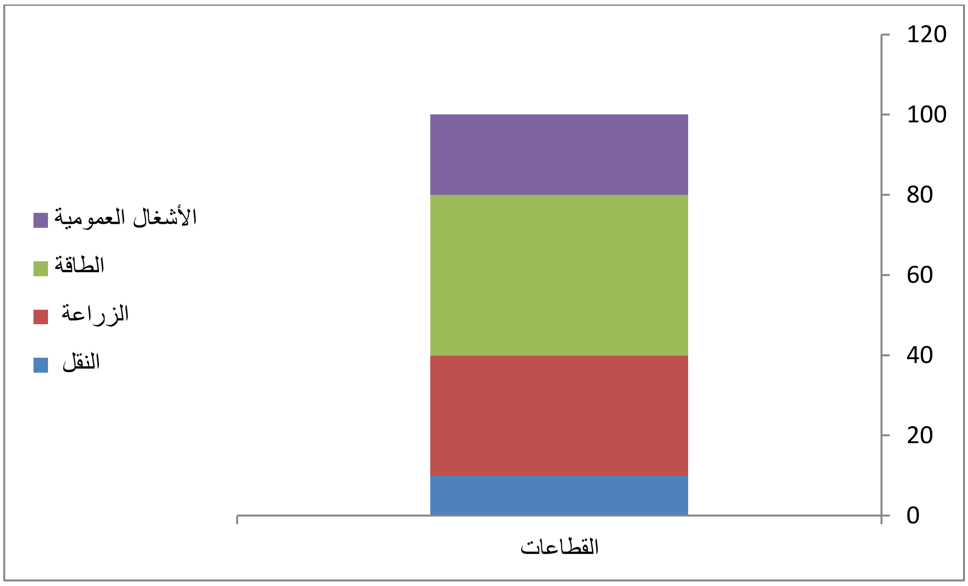
الفئات	التكرار n_i	التكرار النسبي P_i	الزاوية المركزية
النقل	2	0.1	36
الزراعة	6	0.3	108
الطاقة	8	0.4	144
الأشغال العمومية	4	0.2	72
المجموع	20		



3-2- العمود المجزأ:

العمود المجزأ هو عبارة عن مستطيل مقسم إلى عدة أجزاء، كل جزء منه يقابل تكرار معين للخاصية المدروسة، ومن الأفضل عند رسم العمود المجزأ استعمال النسب المئوية المقابلة لكل تكرار حيث أن طول المستقيم هو 100%.

القطاعات	التكرار n_i	التكرار النسبي P_i	التكرار النسبي المئوي $P_i\%$
النقل	2	0.1	10
الزراعة	6	0.3	30
الطاقة	8	0.4	40
الأشغال العمومية	4	0.2	20
المجموع	20		100

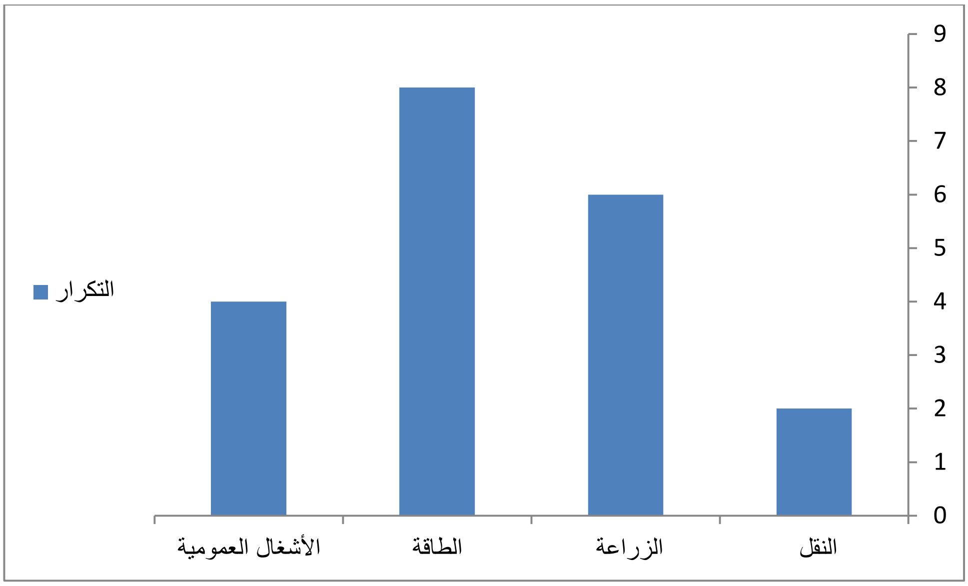


3-3- الأعمدة المستطيلة:

تستخدم الأعمدة البيانية للتعبير البياني عن العلاقة بين متغيرين وتستخدم في البيانات الوصفية، وهي عبارة عن أعمدة او مستطيلات قواعدها متساوية وتمثل الصفة الوصفية ويمثل ارتفاعاتها التكرار المقابل لكل منها[6].

[6]- عبد الحليم عشماوي، صلاح جلال، محمد حسين صادق، **الإحصاء الحيوي وتصميم التجارب**، المكتبة الأكاديمية، مصر، 2008، ص26.

الفئات	النقل	الزراعة	الطاقة	الأشغال العمومية	المجموع
التكرار	2	6	8	4	20



مقاييس النزعة المركزية

مقياس النزعة المركزية هي تلك المقاييس التي تبحث عن تقدير قيمة تتمركز حولها أغلبية القيم وهذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر عن أو يمثل جميع بيانات تلك المجموعة

قياس النزعة المركزية : يعبر قياس النزعة المركزية عن مركز التوزيع الإحصائي، لقياس هذه النزعة نستعمل عدة مقاييس من بينها:

1- المتوسط الحسابي Arithmetic Mean

يعتبر المتوسط الحسابي من أشهر و أكثر متوسطات النزعة المركزية استخداما و شيوعـــا و الإحصاء و هو مركز التوازن لأي ظاهرة و يرمز له بالرمز $\bar{X}$

1-1- خواص المتوسط الحسابي:

- مجموع إنحرفات جملة من اقياسات عن متوسطها يساوي الصفر.

ولتوضيح ذلك نرمز ب d_i للإنحراف $x_i - \bar{x}$ أي إنحراف القياس x_i عن المتوسط $\bar{x}$ ، ولنحسب مجموع الإنحرافات $\sum d_i$ فنجد [7]:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$$

وهذه الخاصة توضح الدور المركزي الذي يلعبه المتوسط.

- يكون مجموع مربعات انحرافات القياسات عن قمية a أصغر ما يمكن عندما يكون $a = \bar{x}$

لنأخذ مجموع مربعات انحرافات القياسات عن قيمة ما a ، أي $\sum_{i=1}^{n}(x_i - \bar{x})^2$ ، فيمكن كتابة:

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

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

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

7- أنيس إسماعيل كنجو، **الإحصاء والاحتمال**، الطبعة الأولى، مكتبة العبيكان، السعودية، 2000، ص47.

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - a)\sum_{i=1}^{n}(x_i - \bar{x}) + \mathrm{n}(\bar{x} - a)^2$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - a)\sum_{i=1}^{n}(x_i - \bar{x}) + \mathrm{n}(\bar{x} - a)^2$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \mathrm{n}(\bar{x} - a)^2 \geq \sum_{i=1}^{n}(x_i - \bar{x})^2$$

ذلك لأن $\mathrm{n}(\bar{x} - a)^2$ كمية غير سالبة. أي أن مجموع مربعات الإنحرافات عن قيمة ما (a) هو دائما اكبر من مجموع مربعات الإنحرافات عن المتوسط $\bar{x}$ أويساويه.

1-2- حساب المتوسط الحسابي من البيانات الأولية (البيانات غير المبوبة)

يستعمل هذا المتوسط في حالة البيانات غير المبوبة أي عندما يكون لقياسات المتغير المدروس نفس المستوي عن الاسمية. فإذا كانت لدينا: χ_n ، .... ،χ_2،χ_1 تمثل قيم ظاهرة ما، فإن المتوسط الحسابي لهذه القيم هو مجموع هذه القيم مقسوما على عدد ما، و تعطي علاقة المتوسط الحسابي بالصيغة التالية:

$$\bar{\mathrm{X}} = \frac{\chi_1 + \chi_2 + \cdots . + \chi_n}{\mathrm{n}} = \frac{\sum_{0=1}^{\mathrm{n}} \chi_{\mathrm{i}}}{\mathrm{n}}$$

حيث n تمثل عدد القيم

مثال: تمثل السلسلة الإحصائية التالية ما أنفقه خمسة زبائن عند تناولهم لفطور الصباح بأحد المقاس.

5 ، 10 ، 20 ، 15 ، 40

المطلوب حساب متوسط الإنفاق

$$\bar{\chi} = \frac{\sum_{0=1}^{n} \mathrm{x_i}}{n} = \frac{\sum_{0=1}^{n} \mathrm{x_i}}{5} = \frac{5 + 10 + 20 + 15 + 40}{5} = \frac{90}{5} = 18$$

1-3- حساب المتوسط الحسابي من البيانات المبوبة:

بعض الحالات، تكون القيم المراد حساب المتوسط الحسابي لها ليس لها نفس الأهمية، بل أهميات نسبية تختلف باختلاف عامل الترجيح الخاص بها: مثل هذه الحالات يتطلب الأمر استخدام صيغة أخرى تسمى بصيغة المتوسط الحسابي المرجح.

1-3-1- حالة متغير كمي منفصل:

إذا كانت: χ_k ، .... ، χ_2 ، χ_1 تمثل مفردات (قيم) ظاهرة معينة χ، و كانت n_k ، ... ، n_2 ، n_1 تمثل التكرارات المقابلة لها، β هذه تعطي علاقة المتوسط الحسابي بالصيغة

$$\bar{\chi} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n_1 + n_2 + \ldots\ldots + n_k} = \frac{\sum_{i=0}^{k} x_i n_i}{\sum_{0=1}^{k} ni}$$

حيث: χ_i تمثل القيمة المفردة i $(i = 1 , 2 , \ldots. , k)$

n_i تمثل تكرار المفردة i

k عدد المفردات المختلفة عن بعضها البعض

مثال:

تابع للمثال السابق: إيجاد متوسط عدد الأطفال لدى العائلات:

عدد الأطفال x_i	عدد تكرار العائلات n_i	$x_i n_i$
1	26	25
2	28	56
3	20	60
4	15	60
5	12	60
المجموع $\sum n_i$	100	261

$$\bar{x} = \frac{\sum_{0=1}^{5} x_i n_i}{\sum n_i}$$

$$= \frac{261}{100}$$

$$\bar{x} = 2.61 \simeq 3$$

1-3-2- حالة متغير كمي متصل:

في هذه الحالة يكون جدول التوزيع التكراري على شكل فئات عدد k، فإذا كانت: x_1، x_2، ... ، x_k تمثل مراكز الفئات، و كانت n_1 ، n_2 ، ... ، n_k تمثل التكرارات المقابلة لتلك الفئات فإننا نعتبر التكرارات تتجمع حول مراكز الفئات و عليه نستخدم الصيغة التالية

$$\bar{\chi} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n_1 + n_2 + \ldots\ldots + n_k} = \frac{\sum_{j=0}^{k} x_j n_j}{\sum_{0=1}^{k} nj}$$

حيث: x_j تمثل القيمة j $(j = 1 , 2 , \ldots . , k)$

n_j تمثل تكرار j

k : عدد الفئات

مثال: تابع للمثال السابق: إيجاد المتوسط الحسابي:

$x_j n_j$	مركز الفئة x_j	التكرار n_j	الفئات
162.5	32.5	5	]35 − 30]
112.5	37.5	3	]40 − 35]
680	42.5	16	]45 − 40]
522.5	47.5	11	]50 − 45]
262.5	52.5	5	]55 − 50]
230	57.5	4	[60 − 55]
1970		44	المجموع $\sum n_i$

$$\bar{x} = \frac{\sum_{j=0}^{k} x_j n_j}{\sum_{0=1}^{k} nj}$$

$$= \frac{\sum_{j=0}^{k} x_j n_j}{100} = \frac{1970}{44}$$

$$\bar{x} = 44.77$$

2- الوسيط Median

الوسيط مجموعة من البيانات هو القيمة التي تقسم تلك البيانات المرئية ترتيبا تصاعديا أو تنازليا إلى قسمين متساويين، حيث يكون عدد القيم الأكبر منه مساويا لعدد القيم الأصغر منه، ويرمز للوسيط بالرمز Me

2- 1- الوسيط للبيانات الأولية (غير المبوبة)

لحساب الوسيط من البيانات الأولية يجب أولا ترتيب البيانات (القيم) تصاعديا أو تنازليا وهنا نميز حالتين:

2-1-1- الوسيط للبيانات غير المبوبة المفردة

إذا كان عدد قيم (n) فردي فإن الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$، و بالتالي قيمة الوسيط هي المفردة (القيمة) التي ترتيبها يقابل رتبة الوسيط $Me = x_{\frac{n+1}{2}}$

مثال: أوجد الوسيط للبيانات التالية:

9 ، 15 ، 12 ، 10 ، 3 ، 6 ، 8

الحل:

1- ترتيب البيانات ترتيبا تصاعديا:

x_7	x_6	x_5	x_4	x_3	x_2	x_1
15	12	10	9	8	6	3

$$Me = x_4 = 9$$

عدد القيم ($n = 7$ عدد فردي)، و عليه فإن رتبة الوسيط $\frac{7+1}{2} = \frac{n+1}{2} = 4$

2-1-2- الوسيط للبيانات غير المبوبة الزوجية

إذا كان عدد القيم (n) زوجي فرتبة الوسيط هي ($\frac{n}{2}+1$ ، $\frac{n}{2}$) و بالتالي قيمة الوسيط تساوي المتوسط الحسابي للمفردتين المتقابلتين لرتبة الوسيط $Me = \frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$

مثال: أوجد الوسيط للبيانات التالية:

7 ، 13 ، 1 ، 3 ، 10 ، 8 ، 6 ، 16

الحل: ترتيب البيانات (القيم) تصاعديا أو تنازليا (يفضل ترتيب البيانات تصاعديا)

x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1
16	13	10	8	7	6	3	1

عدد القيم $(n = 8)$ عدد زوجي) ، و عليه فإن رتيبة الوسيط $\frac{8}{2}$، $\frac{8}{2}$+1، و بالتالي قيمة الوسيط تساوي المتوسط الحسابي للمفردتين المتقابلتين لرتبة الوسيط أي:

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_{\frac{8}{2}} + X_{\frac{8}{2}+1}}{2} = \frac{x_4 + x_5}{2} = \frac{7+8}{2} = 7.5$$

2- 2- الوسيط للبيانات المبوبة

لحساب الوسيط من البيانات المبوبة فإننا نميز بين حالتين

2-2-1 حالة متغير كمي منفصل:

لحساب الوسيط لبيانات متغير كمي منفصل مبوبة في جدول تكراري نتبع الخطوات التالية:

- أولا: نقوم بحساب التكرار المتجمع الصاعد
- ثانيا: تحديد رتبة الوسيط $\left(\frac{N}{2}\right)$ حيث: $N = \sum_{i=1}^{k} n_i$
- رابعا: إيجاد قيمة الوسيط: بحيث العمود (السطر) الخاص بالتكرار المتجمع الصاعد عن القيمة التي تكرارها المتجمع الصاعد يساوي رتبة الوسيط $\left(\frac{N}{2}\right)$ أو أعلى منها مباشرة لتكون القيمة التي تقابلها هي قيمة الوسيط

مثال: تابع للمثال السابق المطلوب: إيجاد الوسيط

عدد الأطفال x_i	التكرار n_i	$N^{\uparrow}$
1	26	25
2	28	53
3	20	73
4	15	88
5	12	100
المجموع $\sum n_i$	100	

نحدد رتبة الوسيط $\frac{N}{2} = \frac{100}{2} = 50$

53 أكبر مباشرة من 50

و عليه تكون قيمة الوسيط هي قيمة x_i التي تقابل التكرار المتجمع الصاعد 53 فإن قيمة الوسيط $Me = 2$

2-2-2- حالة متغير كمي متصل:

لتحديد الوسيط لبيانات مبوبة في جدول توزيع تكراري بفئات (حالة متغير كمي متصل) نتيح الخطوات التالية:

- أولا: نقوم بحساب التكرار المتجمع الصاعد
- ثانيا: نحدد رتبة الوسيط $\frac{N}{2}$ حيث: $N = \sum_{i=1}^{k} n_i$
- ثالثا: نحدد الفئة الوسيطية (الفئة التي ينتمي إليها الوسيط)، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتيبة الوسيط أو أعلى منها مباشرة
- رابعا: نحدد و نحسب الوسيط بتطبيق العلاقة الإحصائية للوسيط:

$$Me = L_{infMe} + \frac{\frac{N}{2} - N^{\nearrow}_{Me-1}}{n_{Me}} \times a$$

حيث: L_{infMe}: الحد الأدنى للفئة الوسيطية

N: مجموع التكرارات: $N = \sum_{i=1}^{k} n_i$

$N_{Me-1}^{\nearrow}$: التكرار المتجمع الصاعد للفئة قبل الوسيطية

Me : تكرار الفئة الوسيطية

n_{Me} : طول الفئة الوسيطية

مثال: أحسب الوسيط للبيانات المبوبة في الجدول التكراري:

الفئات	التكرار n_i	$N^{\uparrow}$
]35 − 30]	5	5
]40 − 35]	3	8
]45 − 40]	16	24
]50 − 45]	11	35
]55 − 50]	5	40
[60 − 55]	4	44
المجموع	44	

الحل:

1− نقوم بحساب التكرار المجمع الصاعد

2− نحدد رتبة الوسيط $(\frac{N}{2} = \frac{44}{2} = 22)$ حيث $N = \sum n_i = 44$

3− نحدد الفئة الوسيطية (الفئة التي ينتمي إليها الوسيط، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة الوسيط أي أعلى منها مباشــرة، (24 أكبر مباشرة من الوسيط 22) و عليه فإن الفئة الوسيطية هي [40 − 45]

4− نحدد و نحسب الوسيط بتطبيق العلاقة الإحصائية للوسيط

$$Me = L_{infMe} + \frac{\frac{N}{2} - N_{Me-1}^{\nearrow}}{n_{Me}} \times a$$

$$= 40 + \frac{\frac{44}{2} - 8}{16} \times 5$$

$$Me = 44.375$$

2-3- الطريقة البيانية لإيجاد الوسيط

الوسيط بيانيا هو نقطة تقاطع كل من منحنى التكرار المتجمع الصاعد والنازل كما يمكن تحديده باستخدام منحنى التكرار المتجمع الصاعد أو النازل فقط، ويكون ذلك من خلال إتباع الخطوات التالية:

- أولا: رسم منحنى التكرار المتجمع الصاعد أو النازل
- ثانيا: تحديد الوسيط $(\frac{N}{2})$ على محور العمودي (التكرار المتجمع الصاعد أو النازل)، ثم رسم مستقيم أفقي ينطلق من رتبة الوسيط على المحور العمودي حتى يلامس منحنى التكرار المتجمع الصاعد أو النازل.
- ثالثا: رسم مستقيم عمودي ينطلق من نقطة التماس السابقة وينتهي مع ملامسة المحور الأفقي، حيث تعطي نقطة التماس مع محور الأفقي (الفئات) فهي قيمة الوسيط.

3- المنوال Mode

تعبر قيمة المنوال على المشاهدة الأكثر تكرار فهو بمثابة القمة الشائعة و قد يكون للبيانات منوال واحد و يمكن أن يكون أكثر من منوال و يعبر المنوال أفضل مقياس لوصف البيانات النوعيـــة و يرمز له بالرمز Mo

3 – 1 – المنوال للبيانات الأولية

المنوال هو القيمة أو الصفة الأكثر تكرار لمقارنة القيم أو الصفات، و على ضوء هذا التعريف فإن المنوال لمجموعة من البيانات قد لا يكون قيمة أو صفة واحدة.

مثال: البيانات التالية تمثل تقديرات التي تحصل عليها 10 طلاب

ممتاز ، جيد ، جيد جدا ، متوسط ، فوق المتوسط ، جيد ، ضعيف ، جيد جدا ، جيد

المطلوب: إيجاد المنوال لهذه البيانات

الحل: نلاحظ أن الصفة جيد هي الصفة الأكثر تكرار من بين الصفات و عليه فإن المنوال هو الصفة جيد (Mo = جيد)

مثال: إذا كان x متغير إحصائي يمثل عدد الفرق في البيت لعدد من العائلات القاطنة بحي معين، أوجد المنوال؟

3 ، 2 ، 5 ، 3، 3 ، 4 ، 5 ، 2 ، 4 ، 3

الحل:

المنوال هو القيمة الأكثر تكرار مقارنة ببقية القيم، و عليه نلاحظ أن القيمة 3 هي الأكثر تكرار (العدد 3 يتكرر 4 مرات)، و بالتالي فإن المنوال هنا هو(Mo = 3)

3 – 2 – المنوال للبيانات المبوبة

المنوال هو الأكثر شيوعا أو تكرار من بين القيم

3–2–1– المنوال لمتغير كمي منفصل

المنوال:_ هو قيمة المتغير X_i المقابلة لأكبر تكرار في جدول التوزيع التكراري

مثال: أحسب المنوال للبيانات التالية:

عدد الأطفال x_i	التكرار n_i
1	25
2	28
3	20
4	15
5	12

من خلال الجدول السابق نلاحظ أن

أكبر تكرار هو 28 و عليه فإن المنوال هو (Mo = 2)

3–2–2– المنوال لمتغير كمي متصل:

لإيجاد منوال لبيانات مبوبة في جدول توزيع تكراري (حالة متغير كمي متصل)، نقوم بإتباع الخطوات التالية:

- الفئة المنوالية هي الفئة الأكبر تكرار n_i
- يتم حساب المنوال باستخدام العلاقة التالية:

$$Mo = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a$$

حيث:

L_1 : الحد الأدني للفئة المنوالية.

Δ_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها.

Δ_2 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها.

a : طول الفئة المنوالية.

مثال: أحسب المنوال للبيانات المبوبة في جدول التوزيع التكراري التالي:

الفئات	التكرار n_i
]35 − 30]	5
]40 − 35]	3
]45 − 40]	16
]50 − 45]	11
]55 − 50]	5
[60 − 55]	4
المجموع	44

الحل:

1− تحديد الفئة المنوالية

نلاحظ أن أكبر تكرار هو 16 إذن الفئة المنوالية هي [40 − 45[

2− حساب المنوال: نقوم بحساب المنوال باستخدام العلاقة التالية:

$$Mo = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a$$

$$Mo = 40 + \frac{(16-3)}{(16-3)+(16-11)} \times 5$$

$$Mo = 40 + \frac{13}{13+5} \times 5$$

$$Mo = 43.61$$

3-3- تحديد المنوال بيانيا

يمكن تحديد المنوال بيانيا وذلك من خلال رسم المدرج التكراري للفئة المنوالية والفئتين السابقة اللاحقة لها، حيث نقوم بإيصال بداية المستطيل للفئة المنوالية ببداية المستطيل للفئة اللاحقة لهـا، ونهاية المستطيل للفئة المنوالية بنهاية المستطيل للفئة السابقة لها، ومن نقطة التقاطع نسقط عمود على المحور الأفقي (محور الفئات) ونقطة تقاطعه مع هذا المحور تعطي قيمة المنوال.

4- الوسط الهندسي X_G (Geometric Mean)

4-1-الوسط الهندسي X_G في حالة البيانات غير المبوبة:

الوسط الهندسي X_G لمجموعة من القيم $x_1, x_2, x_3, \cdots, x_n$ هو الجذر النوني لحاصل ضرب هذه القيم أي بطريقة مباشرة $X_G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \cdots \times x_n}$

كما يمكن حساب الوسط الهندسي بإستخدام اللوغاريتمات كالتالي:

$$ln(X_G) = \frac{1}{n}\sum_{i=1}^{n} ln(x_i) = \frac{1}{n}\big(ln(x_1) + ln(x_2) + \cdots + ln(x_n)\big)$$

$$X_G = e^{ln(X_G)}$$

مثال : أحسب الوسط الهندسي للبيانات للبيانات التالية:

3 ،6 ،8 ،7 ،9 ، 12

وعليه فإن الوسط الهندسي يساوي:

$$X_G = \sqrt[6]{3 \times 6 \times 8 \times 9 \times 7 \times 12}$$
$$X_G = \sqrt[6]{108864}$$
$$X_G = 6{,}91$$

4-2- الوسط الهندسي X_G في حالة البيانات المبوبة:

إذا كانت $x_1, x_2, x_3, \cdots, x_k$ هي بيانات مبوبة في جدول تكراري لمتغير كمي منفصل أو مراكز فئات لمتغير كمي متصل وكانت $n_1, n_2, n_3, \cdots, n_k$ تكرارا موافقة لها، فإن الوسط الهندسي X_G يحسب بالصيغة التالية :

$$X_G = \left(\prod_{i=1}^{K}(x_i)^n\right)^{\frac{1}{N}} = \sqrt[N]{x_1^{n_1} \times x_2^{n_2} \times \cdots \times x_K^{n_k}}$$

حيث $N = \sum_{i=1}^{k} n_i$

كما يمكن حساب الوسط الهندسي بإستخدام الصيغة اللوغاريتمية كالتالي:

$$ln(X_G) = \frac{1}{n}\sum_{i=1}^{n} n_i \times ln(x_i) = \frac{1}{n}\left(n_1 ln(x_1) + n_2 ln(x_2) + \cdots + n_k ln(x_n)\right)$$

$$X_G = e^{ln(X_G)}$$

حيث أن: $e = 2,7182818$

مثال: حساب المتوسط الهندسي لبيانات متغير كمي متصل التالية:

$n_i \times ln(x_i)$	لوغاريتمات $ln(x_i)$	مركز الفئة x_i	ni	الفئات
3.2188	1.6094	5	2	]10-0]
16.2483	2.7080	15	6	]20-10]
25.7510	3.21885	25	8	]30-20]
14.2213	3.555348061	35	4	]40-30]
59.4395			20	المجموع

$$ln(X_G) = \frac{1}{n}\sum_{i=1}^{n} n_i \times ln(x_i)$$

$$ln(X_G) = \frac{1}{20} \times 59.4395$$

$$ln(X_G) = 2.9719$$

$$X_G = e^{ln(X_G)} = e^{2.9719} = 19.5289$$

مثال: حساب المتوسط الهندسي لبيانات متغير كمي منفصل التالية:

عدد الأطفال x_i	**التكرار n_i**	$ln(x_i)$	$n_i \times ln(x_i)$
1	26	0	0
2	28	0.6931	19.4081
3	20	1.0986	21.9722
4	15	1.3862	20.7944
5	12	1.6094	19.3132
المجموع $\sum n_i$	100		81.4880

$$ln(X_G) = \frac{1}{n}\sum_{i=1}^{n} n_i \times ln(x_i)$$

$$ln(X_G) = \frac{1}{100} \times 81.4880$$

$$ln(X_G) = 0.81488$$

$$X_G = e^{ln(X_G)} = e^{0.81488}$$

$$X_G = 2.2589$$

5− الوسط التوافقي X_H (*Harmonic Mean*) :

5−1− الوسط التوافقي X_H في حالة البيانات غير المبوبة:

الوسط التوافقي X_H لمجموعة من القيم $x_1, x_2, x_3, \cdots, x_n$ هو مقلوب الوسط الحسابي لمقلوبات هذه القيم ،أي بطريقة مباشرة

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

$$X_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

مثال : حساب الوسط التوافقي للبيانات التالية:

3 ،6 ،8 ،7 ،9 ، 12

وعليه فإن الوسط التوافقي يساوي

$$X_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{6}{\frac{1}{3}+\frac{1}{6}+\frac{1}{8}+\frac{1}{7}+\frac{1}{9}+\frac{1}{12}}$$

$$X_H = \frac{6}{0.9623} = 6.235$$

$$X_H = 6.235$$

5-2- الوسط التوافقي X_H في حالة البيانات المبوبة:

إذا كانت $x_1, x_2, x_3, \cdots, x_k$ هي بيانات مبوبة في جدول تكراري لمتغير كمي منفصل أو مراكز فئات لمتغير كمي متصل وكانت $n_1, n_2, n_3, \cdots, n_k$ تكرارا موافقة لها، فإن الوسط التوافقي X_H يحسب بالصيغة التالية :

$$X_H = \frac{N}{\sum_{i=1}^{k}\frac{n_i}{x_i}} = \frac{N}{\frac{n_1}{x_1}+\frac{n_2}{x_2}+\cdots+\frac{n_k}{x_k}}$$

حيث $N = \sum_{i=1}^{k} n_i$

مثال : حساب الوسط التوافقي لبيانات متغير كمي متصل:

الفئات	n_i	مركز الفئة x_j	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$
]10−0]	2	5	0.2	0.4
]20−10]	6	15	0.06666667	0.4
]30−20]	8	25	0.04	0.32
]40−30]	4	35	0.02857143	0.11428571
المجموع	20			1.23428571

$$X_H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

$$X_H = \frac{20}{\sum_{i=1}^{4} \frac{n_i}{x_i}} = \frac{20}{1.2342}$$

$$X_H = 16{,}2037$$

مثال : حساب الوسط التوافقي لبيانات متغير كمي متصل:

عدد الأطفال x_i	التكرار n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$
1	26	1	26
2	28	0.5	14
3	20	0.33333333	6.66666667
4	15	0.25	3.75
5	12	0.2	2.4
المجموع $\sum n_i$	100		52.8166667

$$X_H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

$$X_H = \frac{100}{\sum_{i=1}^{4} \frac{n_i}{x_i}} = \frac{100}{52.8166667}$$

$$X_H = 1{,}8933$$

6- الواسط التربيعي X_Q (*Quadratic Mean*) :

6-1- الوسط التربيعي X_Q في حالة البيانات غير المبوبة:

الوسط التربيعي X_Q لمجموعة من القيم $x_1, x_2, x_3, \cdots, x_n$ هو الجذر التربيعي للوسط الحسابي لمربعات هذه القيم ،أي بطريقة مباشرة :

$$X_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i^2)}$$

$$X_Q = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

مثال : أحسب الوسط التربيعي للبيانات التالية:

3 ،6 ،8 ،7 ،9 ، 12

وعليه فإن الوسط التربيعي يساوي:

$$X_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i^2)}$$

$$X_Q = \sqrt{\frac{3^2 + 6^2 + 8^2 + 7^2 + 9^2 + 12^2}{6}} = \sqrt{\frac{383}{6}}$$

$$X_Q = 7{,}9895$$

6-2- الوسط التربيعي X_Q في حالة البيانات المبوبة:

إذا كانت $x_1, x_2, x_3, \cdots, x_k$ هي بيانات مبوبة في جدول تكراري لمتغير كمي منفصل أو مراكز فئات لمتغير كمي متصل وكانت $n_1, n_2, n_3, \cdots, n_k$ تكرارا موافقة لها، فإن الوسط التربيعي X_Q يحسب بالصيغة التالية :

$$X_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i(X_i^2)}$$

$$X_Q = \sqrt{\frac{n_1x_1^2 + n_2x_2^2 + \cdots + n_kx_K^2}{N}}$$

حيث $N = \sum_{i=1}^{k} n_i$

مثال : حساب الوسط التربيعي لبيانات متغير كمي متصل:

$n_i(X_i^2)$	X_i^2	مركز الفئة x_i	ni	الفئات
50	25	5	2	]10−0]
1350	225	15	6	]20−10]
5000	625	25	8	]30−20]
4900	1225	35	4	]40−30]
11300			20	المجموع

$$X_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i(X_i^2)}$$

$$X_Q = \sqrt{\frac{11300}{20}} = \sqrt{565}$$

$$X_Q = 23{,}7697$$

مثال : حساب الوسط التربيعي لبيانات متغير كمي منفصل:

$n_i(X_i^2)$	X_i^2	التكرار n_i	عدد الأطفال x_i
26	1	26	1
112	4	28	2
180	9	20	3
240	16	15	4
300	25	12	5
858		100	المجموع $\sum n_i$

$$X_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i(X_i^2)}$$

$$X_Q = \sqrt{\frac{858}{100}} = \sqrt{8.58}$$

$$X_Q = 2,9291$$

7- الربيعيات (Quartiles)

الربيعي : يكون في ربع العينة هو الذي يقسم البيانات بعد ترتيبها إلى أربعة أجزاء متساوية يتحدد بين كل منها 25% أي ربع المشاهدات ويرمز لها ب Q_1، Q_2، Q_3 حيث يسمى Q_1 الربيع الأول،Q_2 الربيع الثاني (الوسيط)، Q_3 الربيع الثالث.

لتحديد الربيعي لبيانات مبوبة في جدول توزيع تكراري بفئات (حالة متغير كمي متصل) نتيح الخطوات التالية:

- أولا: نقوم بحساب التكرار المتجمع الصاعد
- ثانيا: نحدد رتبة الربيع Q_k : $\frac{k\,\mathrm{N}}{4}$ حيث: $N = \sum_{i=1}^{n} n_i$ وأن : $k = 1,2,3$
- ثالثا: نحدد الفئة الربيع Q_k (الفئة التي ينتمي إليها الربيع Q_k)، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتيبة الربيع Q_k أو أعلى منها مباشرة
- رابعا: نحدد و نحسب الربيع Q_k بتطبيق العلاقة الإحصائية للربيع Q_k:

$$Q_k = L_{infQ_k} + \frac{\frac{kN}{4} - N^{\nearrow}_{Q_k-1}}{n_{Q_k}} \times a$$

حيث: L_{infQ_k}: الحد الأدنى لفئة الربيع k

N: مجموع التكرارات: $N = \sum_{i=1}^{k} n_i$

$N^{\nearrow}_{Q_k-1}$: التكرار المتجمع الصاعد للفئة قبل فئة الربيع k

n_{Q_k} : تكرار فئة الربيع k

a : طول فئة الربيع k

8- العشيرات (Deciles)

العشير: هو الذي يقسم البيانات بعد ترتيبها إلى 10 أجزاء متساوية يتحدد بين كل منها 10% أي عشر المشاهدات ويرمز لها ب D_1، D_2،....، D_9 حيث يسمى D_1 العشير الأول، D_2 العشير الثاني،....، D_5 العشير الخامس (الوسيط)،....، D_9 العشير التاسع.

لتحديد العشسر لبيانات مبوبة في جدول توزيع تكراري بفئات (حالة متغير كمي متصل) نتبع الخطوات التالية

- أولا: نقوم بحساب التكرار المتجمع الصاعد
- ثانيا: نحدد رتبة العشير D_k : $\frac{k\,N}{10}$ حيث: $N=\sum_{i=1}^{n} n_i$ وأن : $k=1,2,3,\cdots,9$
- ثالثا: نحدد العشير D_k (الفئة التي ينتمي إليها العشير D_k)، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة العشير D_k أو أعلى منها مباشرة
- رابعا: نحدد و نحسب العشير D_k بتطبيق العلاقة الإحصائية العشير D_k:

$$D_k = L_{infD_k} + \frac{\frac{kN}{10} - N^{\nearrow}_{D_k-1}}{n_{D_k}} \times a$$

حيث: L_{infD_k}: الحد الأدنى لفئة العشير k

N: مجموع التكرارات: $N=\sum_{i=1}^{k} n_i$

$N^{\nearrow}_{D_k-1}$: التكرار المتجمع الصاعد للفئة قبل فئة العشير k

n_{D_k} : تكرار فئة العشير k

a : طول فئة العشير k

9- المئينات (Centiles)

المئيني: هو الذي يقسم البيانات بعد ترتيبها إلى 100 جزء متساوي يتحدد بين كل منها 1% ، ويرمز لها ب C_1، C_2،....، C_{99} حيث يسمى C_1 المئيني الأول، C_2 المئيني الثاني،....، C_{50} المئيني الخمسون (الوسيط)،....، C_{99} المئيني التاسع.والتسعون.

لتحديد المئيني C_k لبيانات مبوبة في جدول توزيع تكراري بفئات (حالة متغير كمي متصل) نتبع الخطوات التالية:

- أولا: نقوم بحساب التكرار المتجمع الصاعد
- ثانيا: نحدد رتبة المئيني C_k : $\frac{k\,N}{100}$ حيث: $N = \sum_{i=1}^{n} n_i$ وأن : $k = 1,2,3,\cdots,99$
- ثالثا: نحدد المئيني C_k (الفئة التي ينتمي إليها المئيني C_k)، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة المئيني C_k أو أعلى منها مباشرة
- رابعا: نحدد و نحسب المئيني C_k بتطبيق العلاقة الإحصائية المئيني C_k:

$$C_k = L_{infC_k} + \frac{\frac{kN}{100} - N^{\nearrow}_{C_k-1}}{n_{C_k}} \times a$$

حيث: L_{infC_k}: الحد الأدنى لفئة المئيني k

N: مجموع التكرارات: $N = \sum_{i=1}^{k} n_i$

$N^{\nearrow}_{C_k-1}$: التكرار المتجمع الصاعد للفئة قبل فئة المئيني k

n_{C_k} : تكرار فئة المئيني k

a : طول فئة المئيني k

مثال:

الفئات	*ni*	مركز الفئة x_i	$N_i^{\nearrow}$
]10−0]	2	5	2
]20−10]	6	15	8
]30−20]	8	25	16
]40−30]	4	35	20
المجموع	20		

أولا: حساب الربيع الثالث

- نقوم بحساب التكرار المجمع الصاعد
- نحدد رتبة الربيع الثالث إي K=3 $(\frac{3N}{4}=\frac{60}{4}=15)$ حيث $N=\sum n_i=20$
- نحدد فئة الربيع الثالث Q_3 (الفئة التي ينتمي إليها Q_3، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة Q_kأي أعلى منها مباشــرة، (16 أكبر مباشرة من الوسيط 15) و عليه فإن الفئة الوسيطية هي [20 −30[
- نحدد و نحسب الربيع الثالث بتطبيق العلاقة الإحصائية للربيع:

$$Q_3=L_{infQ_3}+\frac{\frac{3N}{10}-N^{\nearrow}_{Q_3-1}}{n_{Q_3}}\times a$$

$$=20+\frac{\frac{3*20}{4}-8}{8}\times 10=28.75$$

ثانيا: حساب العشير السابع

- نقوم بحساب التكرار المجمع الصاعد
- نحدد رتبة العشير السابع إي K=7 $(\frac{7N}{10}=\frac{140}{10}=14)$ حيث $N=\sum n_i=20$
- نحدد فئة العشير السابع D_7 (الفئة التي ينتمي إليها D_7، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة D_7أي أعلى منها مباشــرة، (16 أكبر مباشرة من رتبة العشير السابع 14) و عليه فإن الفئة الوسيطية هي [20 −30[
- نحدد و نحسب العشير السابع بتطبيق العلاقة الإحصائية للعشير:

$$D_7=L_{infD_7}+\frac{\frac{7N}{10}-N^{\nearrow}_{D_7-1}}{n_{D_7}}\times a$$

$$=20+\frac{\frac{7*20}{10}-8}{8}\times 10=27.5$$

- **ثالثا: حساب الربيع الثالث**
- نقوم بحساب التكرار المجمع الصاعد

- نحدد رتبة الربيع الثالث إي K=3 $(\frac{3N}{4}=\frac{60}{4}=15)$ حيث $N=\sum n_i=20$
- نحدد فئة الربيع الثالث Q_3 (الفئة التي ينتمي إليها Q_3، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة Q_3أي أعلى منها مباشــرة، (16 أكبر مباشرة من رتبة الربيع الثالث 15) و عليه فإن فئة Q_3هي [20 −30]
- نحدد و نحسب الربيع الثالث بتطبيق العلاقة الإحصائية للربيع:

$$Q_3=L_{infQ_3}+\frac{\frac{3N}{10}-N^{\nearrow}_{Q_3-1}}{n_{Q_3}}\times a$$

$$=20+\frac{\frac{3*20}{4}-8}{8}\times 10=28.75$$

رابعا: حساب المئيني الخامس والأربعون

- نقوم بحساب التكرار المجمع الصاعد
- نحدد رتبة المئيني الخامس والأربعون إي K=45 $(\frac{45N}{100}=\frac{900}{100}=9)$ حيث $N=\sum n_i=20$
- نحدد فئة المئيني الخامس والأربعون C_{45} (الفئة التي ينتمي إليها C_{45}، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة C_{45}أي أعلى منها مباشــرة، (16 أكبر مباشرة من رتبة المئيني الخامس والأربعون 9) و عليه فإن فئة المئيني الخامس والأربعون هي [20 −30]
- نحدد و نحسب المئيني الخامس والأربعون بتطبيق العلاقة الإحصائية للمئيني:

$$C_{45}=L_{infC_{45}}+\frac{\frac{45N}{100}-N^{\nearrow}_{C_{45}-1}}{n_{C_{45}}}\times a$$

$$=20+\frac{\frac{45*20}{10}-8}{8}\times 10=20.125$$

10− خصائص مقاييس النزعة المركزية ومقارنة بين صفات المتوسط الحسابي والويط والمنوال[8]:

- الوسط الحسابي هو متوسط لقيم المجموعة وليس لمنازل المجموعة كما هو الحال في الوسيط والمنوال.

[8]- محمد حسين محمد رشيد، **الإحصاء الوصفي والتطبيقي والحيوي**، الطبعة الأولى، دائرة المكتبة الوطنية، الأردن، 2008، ص 91.

- يتأثر الوسط الحسابي بالقيم المتطرفة لذلك يصبح مضللا في بعض الحالات.
- مجموع الإنحرافات عن الوسط الحسابي يساوي الصفر.
- مجموع مربعات إنحرفات القيم عن وسطها الحسابي، أقل من مجموع مربعات إنحرفات القيم عن أي قيمة أخرى.
- الوسط الحسابي هو نقطة إتزان للمدرج التكراري، وبما ان الوسط الحسابي هو نقطة إتزان للتوزيع، فإنه إذا اضفنا عدد من القيم التي قيمتها مساوية للوسط الحسابي إلى البيانات فإن هذه الإضافة لا تؤثر ولكن إذا أضفنا مفردات تختلف قيمتها عن قيمة الوسط الحسابي فإن قيمته تتغير.
- لا يمكن حساب الوسط الحسابي في حالة الجداول المفتوحة، لذا نلجأ في حالة الجداول المفتوحة إلى حساب الوسيط والمنوال.
- الوسيط سهل التعريف وسهل الحساب ولا يتأثر بالقيم الشاذة
- يعتبر الوسيط مقياس موضع فإنه لا يعتمد على جميع القيم دائما فتغير بعض القيم قد تؤثر عليه وقد لا تؤثر عليه.
- يستعمل الوسيط خاصة في البيانات التي يعرف ترتيبها ولا تعرف قيمها وكذلك في البيانات الناقصة، لذلك يمكن إستخراجه في الجداول المفتوحة.
- إذا أخذت عينة من مجمتمع ما وأخذت عينة أخرى من نفس المجتمع فإننا نجد تقاربا بين الوسطين الحسابين لهاتين العينتين أكثر من التقارب بين وسطيهما لذلك فإن الوسط الحسابي أكثر ثبوتا من الوسيط.
- المنوال لا يتأثر بالقيم الشاذة (المتطرفة) لذلك يمكن حسابه من الجداول المفتوحة.
- يعتبر المنوال المقياس الوحيد الذي يمكن إستخدامه لإيجاد متوسط للظواهر التي لا يمكن قياسها رقميا (كميا) مثل الصفات فهو الصفة الأكثر شيوعا.
- يفضل إستخدام الوسط الحسابي إذا كان التوزيع متماثلا وإهتمامنا منصب على القيمة العددية لجميع القيم (البيانات) بدلا من قيمة نموذجية.
- يفضل استخدام الوسيط إذا أردنا إيجاد قيمة نموذجية (ممثلة) وإذا كان التوزيع ملتويا.

مقاييس التشتت

Measures of dispersion

يعتبر التشتت صفة طبيعية في جميع الظواهر لتوضيح مدى قرب أو بعد البيانات عن بعضها البعض، أو بصورة أخرى مدى قرب أو بعد البيانات عن قيمة متوسطة كقيمة ممثلة لبيانات تلك الظاهرة، ومن الطبيعي أنه يوجد اختلاف في أطوال الناس أو في أعمارهم أو أوزانهم أو دخولهم. وهكذا وكلما زاد هذا الاختلاف كلما زادت درجة التشتت والعكس صحيح.

ولا تبين مقاييس النزعة المركزية بمفردها مدى انتشار قيم الظواهر وقياس المتغيرات وتوصيفها وصفا تاما، ولذلك كان من اللازم دراسة أنواع أخرى من المقاييس التي تحدد لنا مدى قرب أو بعد البيانات عن بعضها أو عن أحد المتوسطات، وتنقسم مقاييس التشتت إلى نوعين إلى جانب بعض العلاقات الأخرى التي تساعد على قياس درجة التشتت:

اولا: مقاييس التشتت المطلقة:

هناك عدة مقاييس إحصائية لقياس التشتت المطلق تختلف فيما بينها من حيث الدقة والسهولة، والأساس النظري الذي يبنى عليه كل منهما ومن أهمها:

1- المدى Range

يعتبر المدى أبسط مقاييس التشتت، وهو عبارة عن الفرق بين أكبر قيمة في المجموعة وأصغر قيمة في المجموعة، وهو أول مؤشر يقيس القيم.

$$E = X_{max} - X_{min}$$

وفي البيانات المبوبة في جدول التكرار يعرف المدى بأنه الفرق بين الحد الأعلى للفئة العليا وبين الحد الأدنى للفئة الدنيا.

$$E = U_k - L_1$$

مثال: الجدول الموالي يمثل درجات الحرارة اليومية في مدينة ما وعلى ضوء هذه المعطيات أحسب المدى؟

درجات الحرارة °C	37	38	39	40	41

الحل :

$$E = X_{max} - X_{min} = 41 - 37 = 4$$

مثال: إذا كان لدينا جدول التوزيع التكراري التالي المطلوب أحسب المدى؟

الفئات	]35 – 30]	]40 – 35]	]45 – 40]	]50 – 45]	]55 – 50]	[60 – 55]	المجموع $\sum n_i$
التكرار n_i	5	3	16	11	5	4	44

الحل:

$$R = U_k - L_1 = 60 - 30 = 30$$

المدى الربيعي: هو الفرق بين الربيعي الثالث Q_3 والربيعي الأول Q_1 ويرمز له $I.Q$:

$$I.Q = Q_3 - Q_1$$

ويتميز المدى بالخصائص التالية:

- يضم 50% من المجتمع مهما كان التوزيع الإحصائي.
- يتغير طوله مقارنة بالمدى العام حسب طبيعة التوزيع.
- استعمالاته محدودة نظرا لبسطاته، غير أنه أحسن من المدى العام.
- يستعمل في المقارنة بين توزيعيين إحصائيين أو أكثر.

2– نصف المدى الربيعي:

نصف المدي الربيعي هو الفرق بين الربيعي الثالث Q_3 والربيعي الأول Q_1 تقسيم إثنان، ويعطى بالصيغة التالية :

$$\frac{I.Q}{2} = \frac{Q_3 - Q_1}{2}$$

3- النسبة بين المدى الربيعي والمدى العام:

يبين هذ المقياس تشتت 50% من الواحدات الإحصائية المركزية (أي التي تقع في مركز التوزيع) حول الوسيط مقارنة بالمدى العام. وتكتب علاقة هذا المقياس بالشكل التالي :

$$\Re = \frac{Q_3 - Q_1}{E} \times 100$$

من مميزات هذا المقياس ما يلي:

- يستعمل لقياس تشتت 50% من الواحدات التي تقع حول القيمة المركزية بالنسبة لنفس التوزيع.
- إذا كان $\Re = 50\%$ يكون التوزيع الإحصائي متماثل أو متناظر، أما إذا كان هذا المقياس أكبر من 50% فإن التوزيع يكون قوي التشتت بالنسبة للقيمة المركزية، وإذا كان عكس ذلك فإن التوزيع يكون قليل التشتت يالنسبة للقيمة المركزية (الوسيط)

4- الإنحراف المتوسط (The Mean Deviation)

إن أحد مقاييس التشتت التي تخطر على البال هو مجموع انحرافات البيانات عن وسطها الحسابي أي $\sum_{i=1}^{n}(X_i - \bar{X})$ ولكن هذا المجموع يساوي صفرا دائما لأن مجموع الانحرافات الموجبة عن الوسط الحسابي يساوي مجموع الانحرافات السالبة، لذلك لا بد من حذف الإشارة السالبة لنحصل على مقياس ذي معنى.

إن إحدى الطرق للتخلص من الإشارة السالبة هي بأخذ القيمة المطلقة (Absolute Value) التي تعرف بالمعادلة[9]:

$$|X| = \begin{cases} X & X \geq 0 \\ -X & X < 0 \end{cases}$$

وباستعمال القيمة المطلقة نحصل على تعريف الانحراف المتوسط

الإنحراف المتوسط لقيم ظاهرة ما هو عبارة عن المتوسط الحسابي للقيم المطلقة لإنحراف هذه القيم عن وسطها الحسابي، ويرمز له بالرمز $E_{\bar{X}}$.

[9] - محمد صبحي أبو صالح، عدنان محمد عوض، **مرجع سبق ذكره**، ص 39.

4-1- حساب الإنحراف المتوسط من البيانات الأولية (البيانات غير المبوبة)

إذا كانت $x_1, x_2, x_3, \cdots, x_n$ قيم الظاهرة X ، فإن الإنحراف المتوسط يكون بالصيغة التالية:

$$E_{\bar{X}} = \frac{\sum_{i=1}^{n} |x_i - \bar{X}|}{n}$$

حيث:- n : هي مجموع المشاهدات (حجم العينة).

- $\bar{X}$: المتوسط الحسابي للبيانات

مثال: أوجد الانحراف المتوسط للبيانات التالية:

7، 9، 4، 5، 8، 10، 7، 6.

الحل:

نحسب أولا المتوسط الحسابي فنجد:

$$\bar{X} = \frac{\sum_{i=1}^{8} X_i}{n} = \frac{6+7+10+8+5+4+9+7}{8}$$

$$\bar{X} = \frac{56}{8} = 7$$

إذن الانحراف المتوسط:

$$E_{\bar{X}} = \frac{\sum_{i=1}^{8} |x_i - \bar{X}|}{n}$$

$$E_{\bar{X}} = \frac{|6-7| + |7-7| + |10-7| + |8-7| + |5-7| + |4-7| + |9-7| + |7-7|}{8}$$

$$E_{\bar{X}} = \frac{1+0+3+1+2+3+2+0}{8} = \frac{12}{8}$$

$$E_{\bar{X}} = 1.5$$

4-2- حساب الانحراف المتوسط من البيانات المبوبة:

إذا كانت $x_1, x_2, x_3, \cdots, x_k$ قيم الظاهرة X ، أو مراكز فئات التوزيع التكراري للظاهرة X، وكانت التكرارات المقابلة لها n_k ، ... ، n_2 ، n_1 ، فإن الإنحراف المتوسط يكون بالصيغة التالية:

$$E_{\bar{X}} = \frac{\sum_{i=1}^{k} |x_i - \bar{X}| n_i}{n}$$

حيث: $\bar{X}$: المتوسط الحسابي، n : مجموع التكرارات، x_i : قيم الظاهرة x أو مراكز الفئات

مثال: أحسب الانحراف المتوسط للتوزيع التكراري التالي:

x	**1**	**2**	**3**	**4**	**5**
التكرار n_i	26	28	20	15	12

الحل:

x	**التكرار n_i**	$x_i n_i$	$x_i - \bar{X}$	$\lvert x_i - \bar{X} \rvert$	$\lvert x_i - \bar{X} \rvert n_i$
1	26	25	−1.61	1.61	41.86
2	28	56	−0.61	0.61	17.08
3	20	60	0.39	0.39	7.8
4	15	60	1.39	1.39	20.85
5	12	60	2.39	2.39	28.68
المجموع $\sum n$	100	261			116.27

$$\bar{x} = \frac{\sum_{0=1}^{5} x_i n_i}{\sum n_i} = \frac{261}{100} = 2.61$$

$$E_{\bar{X}} = \frac{\sum_{i=1}^{5} |x_i - \bar{X}| n_i}{n} = \frac{116.27}{100} = 1.1627$$

مثال : أحسب الانحراف المتوسط للتوزيع التكراري التالي:

الفئات	]35 − 30]	]40 − 35]	]45 − 40]	]50 − 45]	]55 − 50]	[60 − 55]	المجموع $\sum n_i$
التكرار n_i	5	3	16	11	5	4	44

الحل:

الفئات	$\mathbf{n_i}$	مركز الفئة x_j	$x_i n_i$	$x_i - \bar{X}$	$\lvert x_i - \bar{X}\rvert$	$\lvert x_i - \bar{X}\rvert n_i$
]35 − 30]	5	32.5	162.5	−12.27	12.27	61.35
]40 − 35]	3	37.5	112.5	−7.27	7.27	21.81
]45 − 40]	16	42.5	680	−2.27	2.27	36.32
]50 − 45]	11	47.5	522.5	2.73	2.73	30.03
]55 − 50]	5	52.5	262.5	7.73	7.73	38.65
[60 − 55]	4	57.5	230	12.73	12.73	50.92
المجموع $\sum n_i$	44		1970			239.08

$$\bar{x} = \frac{\sum_{0=1}^{6} x_i n_i}{\sum n_i} = \frac{1970}{44} = 44.77$$

$$E_{\bar{X}} = \frac{\sum_{i=1}^{5} |x_i - \bar{X}| n_i}{n} = \frac{239.08}{44} = 5.43$$

5− التباين والانحراف المعياري:

التباين: يعرف التباين على أنه الوسط الحسابي لمربعات إنحراف القيم عن وسطها الحسابي ويعتبر أهم مقاييس التشتت وأكثرها تطبيقا، ويرمز للتباين بالرمز σ^2 أو $V(x)$ في حالة بيانات المجتمع و S^2 في حالة بيانات العينة.

الإنحراف المعياري: الإنحراف المعياري هو الجذر التربيعي للتباين، ويعتبر من اهم المقاييس الإحصائية للتشتت وأكثرها إستخدما في النظريات والقوانين الإحصائية، ويرمز له بالرمز σ في حالة بيانات المجتمع و S في حالة بيانات العينة.

5-1- حساب التباين والإنحراف المعياري من البيانات الأولية

تباين المجتمع: إذا كانت القيم $x_1, x_2, x_3, \cdots, x_k$ تمثل بيانات مجتمع ما، فإن التباين لهذا المجتمع يعطى بالعلاقة الأتية:

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

حيث μ هو المتوسط الحسابي للمجتمع، أي أن : $\mu = \frac{\sum_{i=1}^{k} x_i}{N}$

الإنحراف المعياري للمجتمع:

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

تباين العينة: إذا كانت القيم $x_1, x_2, x_3, \cdots, x_k$ تمثل بيانات عينة ما، وكان $\bar{X}$ متوسطها الحسابي، فإن التباين يعطى بالعلاقة الأتية:

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

الإنحراف المعياري للعينة:

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

مثال: أحسب التباين الإنحراف المعياري للبيانات 2، 7، 9، 4، 8، 5، 7.

الحل:

الوسط الحسابي:

$$\bar{X} = \frac{\sum_{i=1}^{7} x_i}{n} = \frac{7+5+8+4+9+7+2}{7} = \frac{42}{7} = 6$$

التباين:

$$S^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{X})^2}{n-1}$$
$$= \frac{(7-6)^2 + (5-6)^2 + (8-6)^2 + (4-6)^2 + (9-6)^2 + (7-6)^2 + (2-6)^2}{7-1}$$
$$S^2 = \frac{(1)^2 + (-1)^2 + (2)^2 + (-2)^2 + (3)^2 + (1)^2 + (4)^2}{6}$$
$$S^2 = \frac{1+1+4+4+9+1+16}{6} = \frac{36}{6} = 6$$

الانحراف المعياري:

$$S = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{X})^2}{n-1}} = \sqrt{S^2} = \sqrt{6} = 2.45$$

5-2- حساب التباين والإنحراف المعياري من البيانات المبوبة

تباين المجتمع: : إذا كانت $x_1, x_2, x_3, \cdots, x_k$ تمثل قيم أو مراكز الفئات لبيانات مجتمع ما (حجمه N عنصر)، و $n_1, n_2, n_3, \cdots, n_k$ تمثل التكرارات المقابلة لها فإن تباين المجتمع يعطى بالعلاقة التالية :

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

الإنحراف المعياري للمجتمع :

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

تباين العينة : : إذا كانت $x_1, x_2, x_3, \cdots, x_k$ تمثل قيم أو مراكز الفئات لبيانات عينة ما (حجمه n عنصر)، و $n_1, n_2, n_3, \cdots, n_k$ تمثل التكرارات المقابلة لها، وكان $\bar{X}$ متوسطها الحسابي، فإن تباين العينة يعطى بالعلاقة التالية :

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

الإنحراف المعياري للعينة:

$$S = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{X})^2 n_i}{n-1}}$$

مثال: حسب التباين والإنحراف المعياري للتوزيع التكراري التالي:

$(x_i - \bar{X})^2 n_i$	$(x_i - \bar{X})^2$	$x_i - \bar{X}$	$x_i n_i$	التكرار n_i	x
67.34	2.59	−1.61	25	26	1
10.36	0.37	−0.61	56	28	2
3	0.15	0.39	60	20	3
28.95	1.93	1.39	60	15	4
68.52	5.71	2.39	60	12	5
178.17			261	100	المجموع $\sum n_i$

المتوسط الحسابي

$$\bar{x} = \frac{\sum_{0=1}^{5} x_i n_i}{\sum n_i} = \frac{261}{100} = 2.61$$

التباين:

$$S^2 = \frac{\sum_{i=1}^{5}(x_i - \bar{X})^2 n_i}{n-1} = \frac{178.17}{100-1} = 1.80$$

الإنحراف المعياري:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{X})^2 n_i}{n-1}} = \sqrt{1.80} = \mathbf{1.34}$$

مثال: حساب التباين والانحراف المعياري للتوزيع التكراري التالي:

$(x_i-\bar{X})^2n_i$	$(x_i-\bar{X})^2$	$x_i-\bar{X}$	$x_i n_i$	x_i مركز الفئة	n_i	الفئات
752.75	150.55	−12.27	162.5	32.5	5	]35 − 30]
158.55	52.85	−7.27	112.5	37.5	3	]40 − 35]
82.4	5.15	−2.27	680	42.5	16	]45 − 40]
81.95	7.45	2.73	522.5	47.5	11	]50 − 45]
298.75	59.75	7.73	262.5	52.5	5	]55 − 50]
248.2	162.05	12.73	230	57.5	4	[60 − 55]
1622.6			1970		44	المجموع $\sum n_i$

المتوسط الحسابي:

$$\bar{x}=\frac{\sum_{0=1}^{6} x_i n_i}{\sum n_i}=\frac{1970}{44}=44.77$$

التباين

$$S^2=\frac{\sum_{i=1}^{5}(x_i-\bar{X})^2n_i}{n-1}=\frac{1622.6}{44-1}=37.73$$

الانحراف المعياري:

$$S=\sqrt{S^2}=\sqrt{\frac{\sum_{i=1}^{k}(x_i-\bar{X})^2n_i}{n-1}}=\sqrt{37.73}=6.14$$

الصيغة المختزلة لحساب التباين:

باستخدام خواص المجموع، ومن تعريف المتوسط يمكن أن نكتب ما يلي[10]:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}\left(x_i^2-2\bar{x}x_i+\bar{x}^2\right)$$

$$=\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2\bar{x}x_i+\sum_{i=1}^{n}\bar{x}^2$$

[10]- أنيس إسماعيل كنجو، **مرجع سبق ذكره**، ص86.

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2$$

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

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

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

ومنه:

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

ونكتب العبارة المختزلة السابقة، أحيانا، على الشكل:

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

ثانيا مقاييس التشتت النسبية

1− معامل الاختلاف (Coefficient of Variation)

إن مقاييس التشتت التي بحثناها تعتمد كلها على الوحدة المستعملة في البيانات ولكي نحصل على مقياس لا يعتمد على الوحدة المستعملة نعرف معامل الاختلاف، وإن أهم استعمالات معامل الاختلاف هو للمقارنة بين التغير في عدة مجموعات أو توزيعات تكرارية ولا فرق في ذلك إذا كانت الوحدات المستعملة مختلفة أو نفسها[11].

ومعامل الاختلاف هو:

[11] محمد صبحي أبو صالح، عدنان محمد عوض، **مرجع سبق ذكره**، 1984، ص 45.

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

حيث : $-$ S : الإنحراف المعياري للعينة

- $\bar{X}$: المتوسط الحسابي للعينة

مثال: أيهما أكثر تغيرا (اختلافا) البيانات المجموعة الأولى او المجموعة الثانية؟

	المتوسط الحسابي	الانحراف المعياري
المجموعة الأولى	2.61	1.34
المجموعة الثانية	44.77	6.14

الحل:

نحسب معامل الاختلاف لكل مجموعة من البيانات.

للمجموعة الأولى:

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

$$C.V = \frac{1.34}{2.61} \times 100\% = 51.34\ \%$$

وللمجموعة الثانية:

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

$$C.V = \frac{6.14}{44.77} \times 100\% = 13.71\ \%$$

ومنه نستنتج ان الاختلاف في المجموعة الأولى أكبر منه في المجموعة الثانية.

2- معامل التغير الربيعي: (Coefficient of Quartile Variation)

يستخدم معامل التغير الربيعي لقياس درجة التشتت النسبي لمجموعة من القيم أو لتوزيع معين، ويمكن إسجاده وفقا للصيغة التالية:

$$Q.C.V = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

في حالة الجداول التكرارية المفتوحة والتي لا يمكننا عندها حساب المتوسط الحسابي والإنحراف المعياري، لذلك نستخدم صيغة اخرى تعتمد على الربيعين الأعلى والأذني.

مثال: إذا كان لدينا حدول التوزيع التكراري التالي:

الفئات	n_i	$N_i^{\nearrow}$
]6,5−4.5]	2	2
]8,5−6.5]	5	7
]10,5−8.5]	8	15
]12,5−10.5]	4	19
]14,5−12.5]	1	20
المجموع	20	

المطلوب : حساب معامل التغير الربيعي

أولا: حساب الربيع الأول

- نقوم بحساب التكرار المجمع الصاعد
- نحدد رتبة الربيع الأول إي K=1 $\left(\frac{1N}{4} = \frac{20}{4} = 5\right)$ حيث $N = \sum n_i = 20$
- نحدد فئة الربيع الثالث Q_1 (الفئة التي ينتمي إليها Q_1، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة Q_1أي أعلى منها مباشـرة، (7 أكبر مباشرة من الربيع الأول 5) و عليه فإن الفئة الوسيطية هي]8,5−6.5]
- نحدد ونحسب الربيع الثالث بتطبيق العلاقة الإحصائية للربيع:

$$Q_1 = L_{infQ_1} + \frac{\frac{1N}{4} - N^{\nearrow}_{Q_1-1}}{n_{Q_1}} \times a$$

$$= 6.5 + \frac{\frac{1*20}{4} - 2}{5} \times 2 = 7.7$$

ثانيا: حساب الربيع الثالث

- نقوم بحساب التكرار المجمع الصاعد
- نحدد رتبة الربيع الثالث إي K=3 $(\frac{3N}{4} = \frac{60}{4} = 15)$ حيث $N = \sum n_i = 20$
- نحدد فئة الربيع الثالث Q_3 (الفئة التي ينتمي إليها Q_3، و هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي رتبة Q_3 أوأعلى منها مباشـرة، (15 تساوي مباشرة رتبة الربيع الثالث 15) و عليه فإن الفئة الوسيطية هي [8.5−10,5[
- نحدد ونحسب الربيع الثالث بتطبيق العلاقة الإحصائية للربيع:

$$Q_3 = L_{infQ_3} + \frac{\frac{3N}{10} - N^{\nearrow}_{Q_3-1}}{n_{Q_3}} \times a$$

$$= 8.5 + \frac{\frac{3 * 20}{4} - 7}{8} \times 2 = 10.5$$

ثالثا: حساب معامل التغير الربيعي:

$$Q.C.V = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Q.C.V = \frac{10.5 - 7.7}{10.5 + 7.7}$$

$$Q.C.V = 0.1538$$

مقاييس الشكل

أولا: مقايس الإلتواء (Measures of SKewness)

يكون التوزيع التكراري ملتويا نحو اليمين أو موجب الإلتواء إذا كان ممتدا أكثر نحو اليمين، أما إذا كان له طرف ممتد أكثر نحو اليسار فيقال إنه سالب الإلتواء أو ملتوى نحو اليسار.

عند ما يكون التوزيع ملتويا نحو اليمين فإن القيم المتطرفة نحو اليمين تؤثر على الوسط الحسابي وتسحبه نحو اليمين وبذلك يكون الوسط الحسابي أكبر من الوسيط أما إذا كان التوزيع ملتويا نحو اليسار فإن الوسط الحسابي يكون أصغر من الوسيط [12].

ولذلك يمكننا إستعمال التعريف التالي:

تعريف 01: نعرف مقياس الإلتواء لتوزيع تكراري أو مجموعة من البيانات بالمعادلة التالية:

$$\gamma_1 = \frac{3(\bar{X} - Me)}{S}$$

حيث : $\bar{X}$: المتوسط الحسابي

Me : الوسيط

S : الإنحراف المعياري

ويستفاد من مقياس الإلتواء في أمرين أولهما معرفة نوعية التواء التوزيع التكراري. فإذا كان مقياس الإلتواء موجبا فهذا يعني أن الوسط الحسابي أكبر من الوسيط وأن الطرف الأيمن ممتد أكثر وبالتالي يكون الإلتواء نحو اليمين، أما إذا كان مقياس الإلتواء سالبا فهذا يعني ان الإلتواء نحو اليسار والطرف الأيسر هو الممتد أكثر.

والفائدة الثانية تكون للمقارنة بين التواء توزيعين تكراريين أو مجموعتين من البيانات، فالمجموعة التي مقياس الإلتواء لها أكبر يكون توزيعها ملتويا أكثر من توزيع المجموعة الأخرى.

أما القسمة على الإنحراف المعياري في تعريف مقياس الإلتواء فذلك لجعله غير معتمد على وحدة القياس المستعملة في البيانات.

ويستعمل في العادة مقياس اخر للإلتواء يعتمد في نعريفه على العزم الثالث حول الوسط الحسابي ولذلك نعرف العزوم أولا.

تعريف 2: العزم k للبيانات $x_1, x_2, x_3, \cdots, x_k$ حول وسطها الحسابي هو

[12]- محمد صبحي أبو صالح، عدنان محمد عوض، **مرجع سبق ذكره،** ص47.

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

والعزم K للتوزيع التكراري ذات الفئات حول وسطه الحسابي هو:

$$m_k = \frac{\sum_{i=1}^{k}(X_I - \bar{X})^k n_i}{n}$$

اما العزم k حول نقطة الأصل فيعرف بالمعادلة

$$m_k = \frac{\sum_{i=1}^{k} X_i^k n_i}{n}$$

تعريف 3: مقياس الإلتواء لتوزيع تكراري أو مجموعة من البيانات هو النسبة بين العزم الثالث حول الوسط الحسابي ومكعب الإنحراف المعياري، أي

$$\gamma_2 = \frac{m_3}{S^3}$$

وهذا المقياس لا يعتمد على وحدة قياس البيانات .

هناك مقاييس اخرى للإلتواء منها:

2- معامل الإلتواء الربيعي (Quartile Coefficient of Skewness)

يستعمل الربيعات في تعريفه ويعطي بالصيغة التالية:

$$\gamma_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

3- معامل الإلتواء المئيني (Percentile Coefficient of Skewness)

يستعمل المئينات في تعريفه ويعطي بالصيغة التالية:

$$\gamma_4 = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

ثانيا: التفرطج (Kurtosis)

المقصود بالتفرطح هو درجة علو قمة التوزيع بالنسبة للتوزيع الطبيعي فكلما كانت البيانات قريبة من بعضها (قل التباين في قيمها) كلما ارتفعت قمة التوزيع، والعكس صحيح فكلما كانت قيم البيانات ذات تشتت أكبر (كبر التباين بين القيم) انخفضت قمة التوزيع وإزداد تفلطحه[13].

التفرطح هو قياس درجة علو قمة التوزيع بالنسبة للتوزيع الطبيعي عادة[14].

ومن مقاييس التفرطح المقياس الذي يستعمل الربيعات والمئينات في تعريفه ويسمى : معامل التفرطح المئيني (**The Percentile Coefficient of Skewness**)

1- معامل التفرطح المئيني:

$$k_1 = \frac{1}{2}\frac{Q_3 - Q_1}{P_{90} - P_{10}}$$

يعتمد في تعريفه على العزم الرابع حول الوسط الحسابي ويسمى

2- معامل التفرطح العزومي (Moment Coefficient of Skewness)

يعتمد في تعريفه على العزم الرابع حول الوسط الحسابي ويعطي بالصيغة التالية:

$$k_2 = \frac{m_4}{S^4}$$

حيث : m_4 : العزم الرابع حول الوسط الحسابي

S^4 : مربع التباين

فإذا كانت قيمة معامل التفرطح كبيرة سمي التوزيع كبير التفرطح (Platykurtic) وإذا كانت قيمته صغيرة أي كان للتوزيع قمة عالية سمي التوزيع مدببا أو قليل التفرطح (Leptokurtic).

أما إذا كان معامل التفرطح متوسط القيمة سمي التوزيع متوسط التفرطح (Mesokurtic).

---

[13]- ص 163.

[14]- محمد صبحي أبو صالح، عدنان محمد عوض، **مرجع سبق ذكره،**،ص 49.

مثال: أحسب مقياس الإلتواء γ_2 ومقياس التفرطح k_2 للتوزيع التكراري التالي:

x_i	التكرار n_i
5	5
8	6
11	6
14	4
17	3
20	2
المجموع	26

الحل:

x_i	n_i	$x_i n_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 n_i$	$(X_i - \bar{X})^3$	$(X_i - \bar{X})^3 n_i$	$(X_i - \bar{X})^4$	$(X_i - \bar{X})^4 n_i$
5	5	25	−6	36	180	−216	−1080	1296	6480
8	6	48	−3	9	54	−27	−162	81	486
11	6	66	0	0	0	0	0	0	0
14	4	56	3	9	36	27	108	81	324
17	3	51	6	36	108	216	648	1296	3888
20	2	40	9	81	162	729	1458	6561	13122
المجموع	26	286			540		972		

الوسط الحسابي

$$\bar{X} = \frac{\sum x_i n_i}{n} = \frac{286}{26} = 11$$

التباين :

$$S^2 = \frac{\sum (X_i - \bar{X})^2 n_i}{n-1} = \frac{540}{25} = 21.6$$

الإنحراف المعياري:

$$S = \sqrt{S^2} = \sqrt{21.6} = 4.6475$$

العزم الرابع حول الوسط الحسابي يساوي:

$$m_3 = \frac{(X_i - \bar{X})^3 n_i}{n} = \frac{972}{26} = 37.38$$

العزم الرابع حول الوسط الحسابي يساوي:

$$m_4 = \frac{(X_i - \bar{X})^4 n_i}{n} = \frac{24300}{26} = 934.62$$

مقياس الإلتواء:

$$\gamma_2 = \frac{m_3}{S^3} = \frac{37.38}{(4.6475)^3} = \frac{37.38}{100.388} = 0.372$$

ويتضح منه أن التوزيع قليل الإلتواء

مقياس التفرطح

$$k_2 = \frac{m_4}{S^4} = \frac{934.62}{(4.6475)^4} = 2.00$$

الأرقام القياسية

الأرقام القياسية هي أدوات رياضية تُستخدم لقياس التغيرات في متغيرات اقتصادية أو اجتماعية أو مالية عبر فترات زمنية مختلفة. تُستخدم الأرقام القياسية بشكل واسع في العديد من المجالات مثل الاقتصاد، والإحصاء، والتسويق، وتحليل الأسواق المالية. تهدف الأرقام القياسية إلى مقارنة التغيرات في الأسعار أو الكميات أو القيم الأخرى بين فترات زمنية معينة، مما يساعد في تحليل أداء القطاعات الاقتصادية أو البضائع والخدمات.

1-تعريف الأرقام القياسية: الأرقام القياسية هي مقياس نسبي يُستخدم لتمثيل التغيرات في متغيرات معينة مقارنة بفترة معينة تم اختيارها كنقطة مرجعية. يحدد الرقم القياسي التغير بين القيم في فترتين زمنيتين، وهناك عدة أمثلة على الأرقام القياسية منها:

- **الرقم القياسي للأسعار** :يقيس التغير في الأسعار على مر الزمن (مثل مؤشر أسعار المستهلك).
- **الرقم القياسي للإنتاج** :يقيس التغير في مستوى الإنتاج الصناعي أو الزراعي.
- **الرقم القياسي للعمالة** :يستخدم لقياس التغير في معدلات التوظيف أو البطالة.

2- متطلبات حساب الرقم القياسي :

لحساب أي رقم قياسي لابد من توفير المتطلبات التالية قبل القيام بعملية الحساب، وهذه المتطلبات هي[15] :

- تحدي طبيعة الظاهرة التي يراد قياس تغيرها
- تحدي المفردات التي تتألف منها تلك الظاهرة فيما إذا كانت متشابهة أو مختلفة.
- تحديد فترة الأساس – في حالة الأساس الثابت.
- تحديد الأوزان المناسبة إذا كانت مفردات الظاهرة ذات أوزان مختلفة.
- تحديد المصادر التي تستقى منها المعلومات، وتحديد طبيعة البيانات الإحصائية التي يجب أو يمكن جمعها.

[15]- عبدالحسين زيني، **الأرقام القياسية**، الطبعة الأولى، دار الحامد للنشر والتوزيع، الأردن، 2012، ص28.

- تحديد صيغة الرقم القياسي المناسب. وهذا يتوقف على طبيعة الظاهرة التي يراد قياسها من ناحية، وطبيعة البيانات الإحصائية من ناحية اخرى والهدف من القياس من ناحية ثالثة.

3-خصائص الأرقام القياسية:

- **النسبية** :الأرقام القياسية تمثل نسبة أو مقياس بين قيمتين في فترتين زمنيتين مختلفتين.
- **تعدد الاستخدامات** :يمكن استخدامها لقياس التغيرات في مجموعة واسعة من المجالات مثل الأسعار، الإنتاج، الكميات، وغيرها.
- **المرجعية** :تُقاس الأرقام القياسية دائمًا بالنسبة لنقطة مرجعية تُعتبر كقيمة أساسية (عادةً ما تكون 100).

4-أنواع الأرقام القياسية: هناك نوعان رئيسيان من الأرقام القياسية:

أ. الرقم القياسي التجميعي البسيط:

- تُستخدم لقياس التغير في كميات أو أسعار أحد السلع أو الخدمات.
- يتم حسابها باستخدام متوسط الأسعار أو الكميات.

ب. الرقم القياسي التجميعي المرجح:

- تُستخدم لقياس التغير في مجموعة من السلع أو الخدمات أو المتغيرات.
- يتم حسابها باستخدام متوسط مرجح للعديد من السلع والخدمات.

حيث هناك نوعان رئيسيان من الرقم القياسي المرجح التجميعي المرجح:

ب.1. الرقم القياسي التجميعي المرجح بأوزان ثابتة: وأهم أنوعه[16]:

- الرقم القياسي التجميعي المرجح بأوزان موضوعة.
- الرقم القياسي التجميعي المرجح بأوزان فترة الأساس (صيغة لاسبير).
- الرقم القياسي التجميعي المرجح بأوزان إحدى سنوات المقارنة أو غيرها.

ب.2. الرقم القياسي التجميعي المرجح بأوزان متغيرة: وأهم أنواعه:

- الرقم القياسي التجميعي المرجح بأوزان السنوات المقارنة (صيغة باش).
- الرقم القياسي التجميعي المرجح بأوزان مشتركة (ووسطها الحسابي أو الهندسي) من المقارنة والأساس (صيغة مارشال – ايجورث).

---

[16]- عبدالحسين زيني، **مرجع سبق ذكره**، 2012، ص92.

- الوسط الهندسي لصيغتي لاسبير وباش – الرقم القياسي الأمثل (صيغة فيشر).

5–طرق حساب الأرقام القياسية :

تختلف طرق حساب الأرقام القياسية حسب نوع الرقم القياسي وطبيعة البيانات المتاحة. أهم الطرق تشمل:

5–1– طريقة المتوسط البسيط:

تستخدم هذه الطريقة لقياس التغير في الأسعار أو الكميات عبر فترة معينة باستخدام المتوسط الحسابي.

الرقم القياسي باستخدام المتوسط البسيط للأسعار

$$R = \frac{\sum P_1}{\sum P_0} \times 100$$

حيث:

- P_1 هو السعر في الفترة الحالية.
- P_0 هو السعر في الفترة الأساسية (المرجعية).

5–2– طريقة المتوسط المرجح:

تستخدم هذه الطريقة عندما يكون لدينا سلع أو خدمات ذات أوزان أو أهمية مختلفة. يتم حساب الرقم القياسي باستخدام أوزان محددة لكل عنصر.

الرقم القياسي باستخدام المتوسط المرجح للأسعار

$$R = \frac{\sum P_1 \times q}{\sum P_0 \times q} \times 100$$

حيث:

- P_1 هو السعر في الفترة الحالية.
- P_0 هو السعر في الفترة الأساسية (المرجعية).
- q هو الوزن أو الأهمية النسبية للسلعة.

لكي نراعي أهمية كل سلعة عند حساب الرقم القياسي التجميعي فإننا نعطي كل سلعة وزنا يمثل الكميات المنتجة أو المستهلكة منها. وهناك أكثر من أسلوب للمفاضلة بين إختيار كميات سنة الأساس أو كميات سنة المقارنة أو المتوسط لكميات سنتي الأساس والمقارنة. وفيما يلي عرض موجز لهذه الطرق[17]:

[17]- غضبان عبدالله البدران، **الإحصاء الاستدلالي مع تطبيقات**، الطبعة الأولى، دار المناهج للنشر والتوزيع، 2021، ص 305.

5-2-1- إستخدام كميات سنة الأساس كأوزان (رقم لاسبير)

إذا فرضنا أن الكميات المستهلكة من السلع في سنة الأساس هي q_i حيث $i = 1, \cdots, r$ فإن الرقم التجميعي المرجح بكميات الأساس والذي يطلق عليه رقم لاسبير التجميعي للأسعار (Aggregate Lasperyres Price Index (A.LP.I) يعرف بالمعادلة التالية:

$$\mathrm{A.LP.I} = \frac{p_{n1}q_{o1} + \cdots + p_{nr}q_{0r}}{p_{01}q_{01} + \cdots + p_{0r}q_{0r}} \times 100$$

$$\mathrm{A.LP.I} = \frac{\sum p_n q_o}{\sum p_0 q_0} \times 100$$

وبالتأمل في هذا الرقم نجد أنه متحيز إلى اعلى، ذلك لأن الكميات المستهلكة من السلع المختلفة لا تتغير بنفس النسبة شأنها شأن الأسعار التي لاتتغير بنفس النسبة، وعلى ذلك فإن السلع التي إرتفعت في السعر تعطي وزنا أكبر من أهميتها الحقيقية بينما تعطي السلع التي انخفضت في السعر وزنا أقل من أهميتها الحقيقية في وقت تركيب الرقم القياسي. وحيث أننا نعطي وزنا أكبر مما يجب للسلع التي إرتفعت في الثمن ووزنا أقل مما يجب للسلع التي إنخفضت في الثمن فإن الرقم القياسي يكون متحيزا إلى أعلى.

5-2-2- إستخدام كميات سنة المقارنة كأوزان (رقم باش)

إذا فرضنا أن كميات سنة المقارنة هي على التوالي: $q_{n1}, q_{n2}, \cdots, q_{nr}$ فإن الرقم التجميعي المرجح بهذه الكميات (والذي يطلق عليه رقم باش التجميعي للأسعار A.P.P.I) تعرف بالمعادلة الأتية:

$$\mathrm{A.P.P.I} = \frac{p_{n1}q_{n1} + p_{n2}q_{n2} + \cdots + p_{nr}q_{nr}}{p_{01}q_{n1} + p_{02}q_{n2} + \cdots + p_{0r}q_{nr}} \times 100$$

$$\mathrm{A.P.P.I} = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100$$

وإذا كانت الطريقة السابقة تجعل الرقم متحيزا إلى اعلى، فإن هذه الطريقة تجعله متحيزا إلى أدنى، وذلك لن السلع التي انخفضت أثمانها تزداد الكمية المستهلكة منها عادة ولذلك فهي تعطي أهمية أكثر مما يجب لمجرد أن ثمنها قد إنخفض حتى كأن تغير الثمن يقرر الأوزان التي تستخدم لقياس تغير الثمن نفسه.

5-2-3- استخدام متوسط كميات سنة المقارنة وسنة الأساس كأوزان (رقم مرشال)

يسمى برقم مارشال (Marshal idex)، فهو الذي إقترح إستخدام متوسط كميات السنة المقارنة وسنة الأساس كأوزان وذلك للتخلص من صعوبة إستخدام كميات سنة الأساس أو كميات سنة المقارنة والتي

تجعل من الرقم القياسي متحيزا على أعلى أو متحيزا إلى أدني على التوالي. ويعرف رقم مرسال للأسعار كما يلي:

$$M.P.I = \frac{\sum p_n(q_n + q_0)}{\sum p_0(q_n + q_0)} \times 100$$

5-2-4- الرقم القياسي التجميعي الأمثل (رقم فيشر)

وتتلخص هذه الطريقة في تركيب رقمين قياسين مستقلين يستخدم في أحدهما اوزان سنة الأساس كما فعلنا في أولا، ويستخدم في الثاني أوزان سنة المقارنة كما فعلنا في ثانيا والوسط الهندسي لهذين الرقمين هو الرقم القياسي الأمثل (والذي يطلق عليه رقم فيشر الأمثل F.I.P.I) ويعرف كتالي:

$$F.I.P.I = \sqrt{\left(\frac{\sum p_n q_o}{\sum p_0 q_0}\right)\left(\frac{\sum p_n q_n}{\sum p_0 q_n}\right)} \times 100$$

6- استخدامات الأرقام القياسية:

الأرقام القياسية لها تطبيقات متنوعة في عدة مجالات، ومنها:

6-1- في الاقتصاد :

- قياس مؤشرات الأسعار مثل مؤشر أسعار المستهلك ومؤشر أسعار المنتجين.
- قياس مستوى الإنتاج (مثل الرقم القياسي للإنتاج الصناعي).
- قياس نمو الناتج المحلي الإجمالي.

6-2- في الأعمال والتسويق :

- تحليل التغيرات في المبيعات أو الأرباح عبر فترات زمنية معينة.

6-3- في التجارة الدولية :

- تحليل التغيرات في أسعار الصادرات والواردات.

7- مثال تطبيقي:

لنأخذ مثالًا بسيطًا عن **الرقم القياسي للأسعار** باستخدام طريقة المتوسط البسيط

نفترض أن لدينا بيانات أسعار سلع في فترتين:

- الفترة المرجعية (العام 2020) :
  - سلعة A: 10 دولار
  - سلعة B :20 دولار

- الفترة الحالية (العام 2021) :
  - سلعة A :12 دولار
  - سلعة B :22 دولار

حساب الرقم القياسي للأسعار:

$$R = \frac{(22+12)}{(20+10)} \times 100 = \frac{34}{30} \times 100 = 113.33$$

الرقم القياسي للأسعار هو **113.33**، مما يعني أن الأسعار قد ارتفعت بنسبة 13.33% في الفترة الحالية مقارنة بالفترة المرجعية.

8- الأمثلة التوضيحية حول كيفية استخدام الأرقام القياسية في مختلف المجالات:

1- الرقم القياسي للأسعار (مثل مؤشر أسعار المستهلك):

مثال: نفترض أن لدينا مؤشر أسعار المستهلك لعدد من السلع في فترة معينة (سنة 2020) مقارنة بسنة 2021.

السلعة	2020 (السعر)	2021 (السعر)
الخبز	2	2.5
الحليب	3	3.5
الأرز	5	6
السكر	4	4.5

لحساب الرقم القياسي للأسعار باستخدام طريقة المتوسط البسيط:

مجموع الأسعار في :2020

$$2 + 3 + 5 + 4 = 14$$

مجموع الأسعار في :2020

$$2.5 + 3.5 + 6 + 4.5 = 16.5$$

الرقم القياسي للأسعار:

$$R = \frac{16.5}{14} \times 100 = 117.86$$

النتيجة :الرقم القياسي للأسعار لعام 2021 مقارنة بعام 2020 هو 117.86، مما يعني أن الأسعار قد زادت بنسبة 17.86%.

2-الرقم القياسي للإنتاج الصناعي:

مثال: نفترض أن هناك مصنعًا ينتج ثلاثة أنواع من السلع :المنتج A والمنتج B والمنتج C في فترتين (2020 و **2021**).

المنتج	2020 (الإنتاج بالوحدات)	2021 (الإنتاج بالوحدات)
المنتجA	500	600
المنتجB	300	400
المنتجC	200	250

لحساب الرقم القياسي للإنتاج الصناعي باستخدام طريقة المتوسط المرجح:

الإجمالي في :2020

$$500 + 300 + 200 = 1000$$

الإجمالي في 2021:

$$600 + 400 + 250 = 1250$$

الرقم القياسي للإنتاج:

$$R = \frac{1250}{1000} \times 100 = 125$$

النتيجة :الرقم القياسي للإنتاج الصناعي لعام 2021 مقارنة بعام 2020 هو **125**، مما يعني أن الإنتاج الصناعي زاد بنسبة **.25%**

3- الرقم القياسي للأجور:

مثال: لنأخذ مثالًا على قياس التغير في الأجور في قطاع معين بين عامين: 2019 و2020:.

القطاع	2019 (الأجر بالدينار)	2020 (الأجر بالدينار)
قطاع التعليم	5000	5500
قطاع الصحة	6000	6500
قطاع الصناعة	7000	7500

لحساب الرقم القياسي للأجور باستخدام طريقة المتوسط البسيط:

مجموع الأجور في 2019:

$$5000 + 6000 + 7000 = 18000$$

مجموع الأجور في 2020:

$$5500 + 6500 + 7500 = 19500$$

الرقم القياسي للأجور:

$$R = \frac{19500}{18000} \times 100 = 108.33$$

النتيجة :الرقم القياسي للأجور لعام 2020 مقارنة بعام 2019 هو **108.33**، مما يعني أن الأجور قد زادت بنسبة **8.33%.**

4-الرقم القياسي لأسعار الصادرات والواردات:

مثال: نفترض أن لدينا بيانات حول أسعار السلع المصدرة والمستوردة في دولتين خلال عامين.

السلعة	السعر في 2020 (الواردات)	السعر في 2021 (الواردات)	السعر في 2020 (الصادرات)	السعر في 2021 (الصادرات)
النفط	40	45	50	55
القمح	25	30	20	22
السيارات	100	120	90	110

لحساب الرقم القياسي لأسعار الواردات:

مجموع أسعار الواردات في :2020

$$40 + 25 + 100 = 165$$

مجموع أسعار الواردات في :2021

$$45 + 30 + 120 = 195$$

الرقم القياسي لأسعار الواردات:

$$R = \frac{195}{165} \times 100 = 118.18$$

النتيجة :الرقم القياسي لأسعار الواردات لعام 2021 مقارنة بعام 2020 هو 118.18، مما يعني أن أسعار الواردات ارتفعت بنسبة 18.18%.

5-الرقم القياسي للبطالة:

مثال :لنفرض أن لدينا بيانات حول معدلات البطالة في دولتين بين عامين.

الدولة	2020 (معدل البطالة)	2021 (معدل البطالة)
دولةA	6%	7%
دولةB	4%	5%
دولةC	10%	12%

لحساب الرقم القياسي للبطالة باستخدام طريقة المتوسط البسيط:

مجموع معدلات البطالة في :2020

$$6 + 4 + 10 = 20$$

مجموع معدلات البطالة في 2021:

$$7 + 5 + 12 = 24$$

الرقم القياسي للبطالة:

$$R = \frac{24}{20} \times 100 = 120$$

النتيجة :الرقم القياسي للبطالة لعام 2021 مقارنة بعام 2020 هو **120**، مما يعني أن معدلات البطالة قد ارتفعت بنسبة **20%.**

الارتباط الخطي

Linear Correlation

يمثل الارتباط (أو على وجه التحديد معامل الارتباط) الأسلوب (أو المقياس) الذي يمكن من دراسة نوع وقوة العلاقة بين المتغيرات من حيث الاتجاه والقوة، هذه العلاقة بين المتغيرات التي قد تكون خطية أو غير خطية.

1- مفهوم الارتباط

الارتباط يعني وجود علاقة بين عاملين أو أكثر، كل منهما يؤثر على الأخر أي أن التأثير متبادل بينهما، وقد يكون التأثير موجب إذا كانت الزيادة في أحدهما يتبعه زيادة في الأخر وتسمى هذه الحالة بالارتباط الموجب. كما قد يكون التأثير سالبا إذا كانت الزيادة في أحدهما يتبعها نقص في الأخرى وتسمى هذه الحالة بالارتباط السالب[18].

إن الارتباط قد يكون بسيطا إذا كانت العلاقة بين متغيرين أو عاملين فقط وقد يكون متعدد إذا كانت العلاقة بين أكثر من متغيرين أو عاملين، وعادة ما يرمز لمعامل الارتباط بالرمز (r) إذا كانت محسوبا من العينة ويرمز له بالرمز (R) إذا كان محسوبا من المجتمع.

تتراوح قيمة الارتباط ما بين −1 و +1 فإذا كانت القيمة هي (+1) دل ذلك على وجود ارتباط موجب، وكلما اقتربت قيمة معامل الارتباط من الصفر دل ذلك على وجود ارتباط تتناقص قيمته تدريجيا إلى ان يتلاشى وجوده، أما إذا وصلت قيمة الارتباط عند (−1) أعتبر ذلك على أنه ارتباط سالب، لذلك فإنه من الأهمية بمكان ملاحظة إشارة معامل الارتباط مع قيمته.

2- أهمية دراسة الارتباط، أنواعه

2-1- أهمية دراسة الارتباط: تهدف دراسة الارتباط إلى وصف درجة التغير الاقتراني بين المتغيرات وتفيد في:

- تحديد قوة الارتباط بين المتغيرين، أي بيان ما إذا كان الارتباط قوي أو ضعيف أو منعدم.
- تحديد اتجاه العلاقة بين المتغيرين، أي بيان ما إذا كانت العلاقة طردية أو عكسية.

[18]- إسماعيل محمد بن قانة، **مرجع سبق ذكره**، ص140.

- إن دراسة الارتباط تعد الأساس لدراسة وتحليل علاقات السببية.
- تعطي مؤشرات لإمكان تقدير المتغيرات بدلالة أخرى.

2-2- أنواع الارتباط:

يمكن تقسيم الارتباط حسب العديد من الخصائص التي تميزه، حيث يمكن إجمالها فيما يأتي:

2-2-1- من حيث قوته: ويمكن أن نميز:

- ارتباط كامل (قيمته 1 أو -1) وهذا يعني أحد المتغيرات يتوقف كليا على تغير الأخر.
- ارتباط جزئي وهذا يعني أنه يوجد ارتباط ولكن ليس بقوة الارتباط السابق.

2-2-2- من حيث عدد المتغيرات التي تؤخذ بعين الاعتبار: ويمكن أن نميز:

- ارتباط بسيط أين ندرس العلاقة بين متغيرين فقط.
- ارتباط متعدد أين ندرس العلاقة بين أكثر من متغيرين.

2-2-3- من حيث العلاقة الرياضية التي تربط بين المتغيرات: ونميز منها:

- ارتباط خطي.
- ارتباط غير خطي.

3- قياس الارتباط: يمكن قياس الارتباط بين المتغيرات على مرحلتين:

- **في المرحلة الأولى:** يتم التعرف عن طريق الوصف على مدى الترابط بين المتغيرين وذلك من خلال سحابة النقاط، إذ يمكن القول من خلال ملاحظة سحابة النقاط أن الارتباط قوي إيجابي أو سلبي ضعيف.
- **في المرحلة الثانية:** يتم تقدير قوة أو ضعف الارتباط من حيث قرب أو بعد النقاط عن الخط المستقيم فكلما اقتربت من الخط المستقيم كانت قوية وكلما ابتعدت عنه كان الارتباط ضعيفا وبشكل عام فإن قيمة الارتباط تتراوح بين 1، و -1 ولحساب هذه القيمة نلجأ إلى استخدام أحد المقاييس التالية:

3-1- معامل ارتباط بيرسون أو معامل ارتباط حاصل ضرب العزم:

يوضح معامل ارتباط بيرسون درجة العلاقة بين المتغيرات، التي تهدف إلى تحديد مدى جودة وصف معادلة خطية، فإذا كانت جميع قيم المتغيرات تحقق معادلة ما بالضبط فإنها ستحقق ارتباطا كاملا بين المتغيرات. وإذا كان شكل الانتشار أو سحابة النقاط تبدو كخط فإن الارتباط يسمى خطيا. ويعطى معامل ارتباط بيرسون بين متغيرين X و Y وفق الصيغة التالية:

$$r_{x,y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

حيث وبعد إجراء بعض المعالجات الجبرية يمكن أن نستنتج:

$$r = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{\sqrt{n(\sum X_i^2) - (X_i)^2}\sqrt{n(\sum Y_i^2) - (Y_i)^2}} \ or$$

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum X_i^2 - n\bar{X}^2}\sqrt{\sum Y_i^2 - n\bar{Y}^2}}$$

علما ان :

$\bar{X}$: تمثل الوسط الحسابي للبيانات X.

$\bar{Y}$: تمثل الوسط الحسابي للبيانات Y.

σ_x : تمثل الانحراف المعياري للبيانات X.

σ_y : تمثل الانحراف المعياري للبيانات Y.

ملاحظة:عند حساب معامل الارتباط لبيرسون يشترط أن يكون التوزيع لكلا المتغيرين طبيعي (أي يتبع التوزيع الاحتمالي الطبيعي أو لابلاس-قوس) وأن تكون العينة عشوائية وقيم الفرد لا تعتمد على قيم فرد أخر (استقلالية أفراد العينة). وفي حالة عدم خضوع المتغيرين للتوزيع الطبيعي نستخدم معامل ارتباط اخر هو معامل ارتباط سبيرمان أو كندال "تاو".

مثال: يمثل الجدول التالي ساعات الدراسة لأحد الطلاب لخمسة امتحانات و العلامة التي حصل عليها.

عدد ساعات الدراسة(x)	6	5	7	8	3
العلامة (Y)	30	28	35	42	25

المطلوب : أحسب معامل الارتباط بين عدد ساعات الدراسة و العلامة المحصل عليها ؟ ماذا تستنتج ؟

الحل:حساب معامل الارتباط بين عدد ساعات الدراسة (x) و العلامة المحصل عليها (y)

i	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	6	30	0.2	−2	−0.4	0.04	4
2	5	28	−0.8	−4	3.2	0.64	16
3	7	35	1.2	3	3.6	1.44	9
4	8	42	2.2	10	22	4.84	100
5	3	25	−2.8	−7	19.6	7.84	49
المجموع	29	160			48	14.8	178

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{N}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{29}{5}=5.8$$

$$\bar{x}=\frac{\sum_{i=1}^{n}y_i}{N}=\frac{\sum_{i=1}^{5}y_i}{5}=\frac{160}{5}=32$$

$$r_{x,y}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}=\frac{48}{\sqrt{14.8}\sqrt{178}}=0.9351$$

$$r_{x,y}=0.9351$$

من $r_{x,y}=0.9351$ نستنتج أن هناك علاقة قوية طردية بين عدد ساعات الدراسة والعلامة المحصل عليها.

3-2- معامل الارتباط الرتبي Rank corrélation coefficient:

يستخدم معامل الارتباط الرتبي لدراسة الارتباط في حالة متغيرات كيفية، حيث نستعمل ترتيبا تصاعدية أو تنازلية عوضا عن القيم العديدية للمتغيرات المدروسة. فإذا كان الترتيب التصاعدي للمتغير المستقل يرافقه ترتيب تصاعدي للمتغير التابع نقول أن الارتباط موجب (علاقة طردية)، أما إذا كان الترتيب التصاعدي للمتغير المستقل يرافقه ترتيب تنازلي للمتغير التابع نقول أن الارتباط بين المتغيرين سالب (علاقة عكسية)

قياس الارتباط الرتبي:

لقياس الارتباط الرتبي بين مفردات المتغير التابع والمستقل نرتب كلاهما حسب أهمية خصائصهما ثم نحسب مجموع مربعات الفروق بين كل الرتب المتقابلة، لتصبح قياسات المتغيرين عبارة عن متوالية حسابية. ولاستخراج العلاقة الإحصائية لمعامل الارتباط الرتبي نعوض في علاقة الارتباط الخطي البسيط بقيم هاتين المتواليتين حسب التغيرات التالية[19]:

- مجموع n عدد مرتبا ترتيبا تصاعديا هو: $\sum X_i = \sum Y_i = \frac{n(n+1)}{2}$
- مجموع مربعات هذه الأعداد هو: $\sum X_i^2 = \sum Y_i^2 = \frac{n(n+1)(2n+1)}{2}$
- متوسط n عددا مرتبا ترتيبا تصاعديا هو: $\bar{X} = \bar{Y} = \frac{n+1}{2}$
- أما تباين هذه القيم فهو:
  - $V(X) = V(Y) = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$
- مجموع مربعات الفروق بين رتب المتغيرين هو: $\sum d_i^2 = \sum (X_i - Y_i)^2$

حسب هذه المعطيات تصبح علاقة معامل الارتباط الرتبي بالشكل التالي:

$$r = \frac{\frac{(n+1)(2n+1)}{6} - \frac{\sum d_i^2}{2n} - \left(\frac{n+1}{2}\right)^2}{\frac{n^2-1}{12}}$$

$$r = \frac{\frac{\sum d_i^2}{2n}}{\frac{n^2-1}{12}} = 1 - \frac{\sum d_i^2 \,.\, 12}{2n(n^2-1)}$$

[19]- جلاطو جيلالي، مرجع سبق ذكره، ص 152.

إذن $r = 1 - \frac{6.\sum d_i^2}{n(n^2-1)}$ ، تسمى هذه العلاقة بعلاقة: Spearman

مثال: احسب معامل الارتباط الرتبي للعلاقة بين أسعار الفائدة ومعدل البطالة للبيانات التالية

السنة	سعر الفائدة(%)	ترتيب سعر الفائدة	معدل البطالة(%)	ترتيب معدل البطالة
2019	4.5	2	7.0	3
2020	3.0	4	8.5	4
2021	5.0	1	6.0	2
2022	2.5	5	5.5	1
2023	3.5	3	6.2	5

الخطوات:

أولا: ترتيب البيانات :تم ترتيب البيانات بناء على أسعار الفائدة ومعدل البطالة.

ثانيا: حساب الفروق بين الرتب:

$$d_1 = 2 - 3 = -1$$

$$d_2 = 4 - 4 = 0$$

$$d_3 = 1 - 2 = -1$$

$$d_4 = 5 - 1 = 4$$

$$d_1 = 3 - 5 = -2$$

ثالثا: تربيع الفروق:

$$d_1^2 = (-1)^2 = 1$$

$$d_2^2 = (0)^2 = 0$$

$$d_3^2 = (-1)^2 = 1$$

$$d_4^2 = (4)^2 = 16$$

$$d_5^2 = (-2)^2 = 4$$

رابعا: حساب معامل الإرتباط الرتبي:

$$\boldsymbol{r} = 1 - \frac{6.\sum d_i^2}{n(n^2-1)} = 1 - \frac{6.(1+0+1+16+4)}{5(25-1)} = -0.1$$

$$\boldsymbol{r} = -0.1$$

التفسير :معامل الارتباط الرتبي $\boldsymbol{r} = -0.1$ يشير إلى علاقة ضعيفة وسلبية بين أسعار الفائدة ومعدل البطالة، مما يعني أن العلاقة بين المتغيرين ليست قوية في هذه البيانات.

3-3- معامل الارتباط المتعدد Multiple corrélation coefficient

يستخدم لقياس العلاقة بين أكثر من متغيرين، إلا أن إشارة معامل الارتباط لا تدل على اتجاه العلاقة هنا لأن هذا الاتجاه لا يكون موحدا لجميع المتغيرات. وصيغة حسابه في حالة 3 متغيرات لإيجاد العلاقة بين x_1 وكل من x_3 و x_2 هي[20]:

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}}$$

حيث أن: r_{23}، r_{13}، r_{12} هي معاملات ارتباط يتم إيجادها بموجب صيغة الارتباط البسيط.

4-3- معامل الارتباط الجزئي Partial corrélation coefficient

يستخدم لقياس العلاقة بين زوج من المتغيرات عندما تكون باقي المتغيرات ثابتة. وبذلك فإن الفرق بين الارتباط البسيط والارتباط الجزئي هو أن الأول يقيس العلاقة بين متغيرين ضمن تأثير المتغيرات الأخرى، في حين يقيس الثاني العلاقة بين متغيرين مع استبعاد تأثير المتغيرات الأخرى.

وصيغة حساب معامل الارتباط الجزئي بين y و x_2 مع ثبات x_1 مثلا هي:

$$r_{y\,2.1} = \frac{r_{y2} - (r_{y1})(r_{12})}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}}$$

[20]– عبدالحميد عبدالمجيد البلداوي، **الأساليب التطبيقية لتحليل وإعداد البحوث العلمية مع حالات دراسية باستخدام برنامج SPSS** ، دار الشروق للنشر والتوزيع، عمان الأردن، 2008، ص177.

حيث أن: r_{y2} و r_{y1} و r_{12} هي معاملات يتم إيجادها بموجب صيغة الارتباط البسيط.

3-5- الإرتباط بين الصفات Correlation between Attributes :

وهو الإرتباط الذي يقترن بوجود توزيع تكراري مزدوج لمتغيرين من النوع الوصفي (كلاهما أو احدهما من النوع الوصفي) مما يتعذر إستخدام معامل إرتباط سبيرمان (لإنها تكون غير قابلة للترتيب تصاعديا أو تناوليا) ويستوجب مقياس أخر ملائم لهذه الحالة مثل معامل التوافق أومعامل الإقتران وهي كما يلي[21]:

3-5-1- معامل التوافق Coefficient of Contingency :

يمكن إيجاد معامل التوافق للبيانات المبوبة الوصفية غير القابلة للترتيب (او قابلة للترتيب) من خلال تكوين جدول التوافق الأتي:

		مستويات y				
		y_1	y_2	$\cdots$	y_m	$T_{j.}$
مستويات x	x_1	n_{11}	n_{12}	$\cdots$	n_{14}	T_1
	x_2	n_{21}	n_{22}	$\cdots$	n_{24}	T_2
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	x_k	n_{41}	n_{42}	$\cdots$	n_{44}	T_k
المجموع	$T_{.i}$	$T_{.1}$	$T_{.2}$	$\cdots$	$T_{.m}$	$\sum n_{j.} = \sum n_{.i} = T_{km} = n$

فإن معامل التوافق C يمكن أن نحصل عليه بالاعتماد على صفوف (او عمدة) جدول التوافق كما في الصيغة الأتية:

$$C = \sqrt{\frac{r-1}{r}}$$

حيث r نحصل عليه من خلال جمع قيم الصفوف r_j ، أي أن $r = \sum_{j=1}^{k} r_j$

وأن قيم r_j نحصل عليها من خلال الصيغة التالية:

[21]- طه حسين الزبيدي، **مبادئ الإحصاء**، الطبعة الأولى، دار غيداء للنشر والتوزيع، 2012،ص 175.

$$r_j = \frac{1}{T_{j.}}\sum_{i=1}^{m}\frac{n_{ji}^2}{n_{.i}}$$

مثال: الجدول التالي يبين عدد حوادث الطرق التي وقعت في إحدى المدن خلال فترة زمنية موزعة حسب نوع الحادث وحالة الطقس، المطلوب حساب معامل التوافق.

حالة الطقس / نوع الحادث	دهس	اصطدام	انقلاب	المجموع
صحو	15	8	5	28
ممطر	5	25	15	45
ضباب	10	23	20	53
المجموع	30	56	40	126

الحل: نحسب أولا قيم r_j من الصيغة التالية:

$$r_j = \frac{1}{T_{j.}}\sum_{i=1}^{m}\frac{n_{ji}^2}{n_{.i}}$$

$$r_1 = \frac{1}{T_{1.}}\sum_{i=1}^{3}\frac{n_{1i}^2}{n_{.i}} = \frac{1}{28}\left(\frac{15^2}{30}+\frac{8^2}{56}+\frac{5^2}{40}\right) = \frac{9.2679}{28} = 0.331$$

$$r_2 = \frac{1}{T_{2.}}\sum_{i=1}^{3}\frac{n_{2i}^2}{n_{.i}} = \frac{1}{45}\left(\frac{5^2}{30}+\frac{25^2}{56}+\frac{15^2}{40}\right) = \frac{17.619}{45} = 0.3915$$

$$r_3 = \frac{1}{T_{3.}}\sum_{i=1}^{3}\frac{n_{3i}^2}{n_{.i}} = \frac{1}{53}\left(\frac{10^2}{30}+\frac{23^2}{56}+\frac{20^2}{40}\right) = \frac{22.7798}{53} = 0.4298$$

لذلك فإن:

$$r = \sum_{j=1}^{k} r_j = r_1 + r_2 + r_3 = 0.331 + 0.3915 + 0.4298 = 1.1523$$

فإن معامل التوافق نحصل عليه من خلال:

$$C = \sqrt{\frac{r-1}{r}} = \sqrt{\frac{1.1523-1}{1.1523}} = 0.3636$$

وهذا يعني ان هنالك نسبة توافق مقدارها 36.36% بين حالة الطقس ونوع الحادث.

3-5-2- معامل الاقتران Coefficient of Association

هناك حالتان في عملية حسابه وهي كما يلي:

أولا: في حالة وجود مستويين لكل متغير:

معامل الإقتران يقيس العلاقة بين متغيرين وصفيين غير قابلين للترتيب (او ربما أحدهما او كلاهما قابل للترتيب) مفرغة بياناتهما في جدول توافق بحجم 2x2 مثل الجنس (ذكر، أنثى) وكفاءة الأداء في العمل (جيد، سئ) وجدول التوافق يكون كالأتي:

		مستويات y		
مستويات x		y_1	y_2	$T_{j.}$
	x_1	n_{11}	n_{12}	T_1
	x_2	n_{21}	n_{22}	T_2
المجموع	$T_{.i}$	$T_{.1}$	$T_{.2}$	n

وبالاعتماد على جدول التوافق نحصل على معامل الإقتران من خلال تطبيق الصيغة الأتية:

$$C.A_2 = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}$$

مثال: الجدول التالي يبين عدد الذكور والإناث المشتركين في امتحان كفاءة التدريس والمطلوب حساب معامل الاقتران بين الجنس وكفاءة التدريس.

الجنس	كفاءة التدريس	جيد	سيئ	المجموع
الذكور		40	25	65
الإناث		30	15	45
المجموع		70	40	110

الحل: يمكن حساب معامل الإقتران من خلال الصيغة التالية:

$$C.A_2 = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}$$

$$C.A_2 = \frac{(40)(15) - (25)(30)}{(40)(15) + (25)(30)} = \frac{-150}{1350}$$

$$C.A_2 = -0.1111$$

وهذا يعني هناك ارتباط عكسي ضعيف بين الجنس وكفاءة التدريس.

ملاحظة: قيمة معامل الإقتران تكون محصورة مابين 1 و -1

ثانيا: في حالة وجود مستويين او اكثر لكل متغير:

يمكن حساب معامل الاقتران بالاعتماد على جدول التوافق k x m الموضح سابقا من خلال تطبيق الصيغة التالية:

$$C.A_G = \frac{\sum_{i=1}^{m} n'_{ij} + \sum_{j=1}^{k} n''_{ij} - T'_{.i} - T'_{j.}}{2n - \left(T'_{.i} + T'_{j.}\right)}$$

حيث أن:

n'_{ij} يمثل أكبر تكرار في العمود i

n''_{ij} يمثل أكبر تكرار في العمود j

$T'_{.i}$ يمثل أكبر مجموع من بين مجاميع الأعمدة.

$T'_{j.}$ يمثل أكبر مجموع من بين مجاميع الصفوف.

مثال: بإستعمال بيانات جدول التوافق في المثال السابق أحسب معامل الاقتران

الحل: يتم تكوين الجدول الاتي:

حالة الطقس / نوع الحادث	دهس	إصطدام	إنقلاب	المجموع	اكبر تكرار في العمود	أكبر مجموع للصفوف
صحو	15	8	5	28	15	…
ممطر	5	25	15	45	25	…
ضباب	10	23	20	53	23	53
المجموع	30	56	40	126	63	…
أكبر تكرار في العمود	15	25	20	60		
اكبر مجموع للأعمدة	…	56	…	…		

من خلال جدول التوافق وتطبيق الصيغة التالية نحصل على معامل الاقتران وكما يلي:

$$C.A_G = \frac{\sum_{i=1}^{m} n'_{ij} + \sum_{j=1}^{k} n''_{ij} - T'_{.i} - T'_{j.}}{2n - \left(T'_{.i} + T'_{j.}\right)}$$

$$C.A_G = \frac{60 + 63 - 56 - 53}{2(126) - (56 + 53)} = \frac{14}{143} = 0.0979$$

الانحدار الخطي البسيط

أولا: مفهوم الانحدار الخطي البسيط: يستخدم النموذج الخطي ذي المتغيرين، أو تحليل الانحدار الخطي البسيط، لاختبار الفروض حول العلاقة بين متغير تابع y، ومتغير مستقل أو مفسر x، والتنبؤ. ويبدأ الانحدار البسيط برسم مجموعة قيم XY في شكل انتشار تم التحديد بالنظر ما إذا كانت هناك علاقة خطية تقريبية [22].

$$Y_i = b_0 + b_1 X_i \dots\dots\dots\dots\dots\dots\dots.(1)$$

وحيث أنه من غير المتوقع أن تقع النقاط تماما على الخط، فإن العلاقة الخطية التامة في المعادلة (1) يجب أن تعدل لكي تضم **حد تشويش عشوائي** أو **خطأ** أي "**عنصر عشوائي**"، ε_i.

$$Y_i = b_0 + b_1 X_i + \varepsilon_i \dots\dots\dots\dots\dots\dots\dots.(2)$$

ثانيا: الفرضيات الخاصة بالمتغير العشوائي[23]:

1- أن المتغير العشوائي (ε_i) هو متغير تعتمد قيمته في أية فترة زمنية على عامل الصدفة، فقد تكون أكبر أو أصغر أو مساوية إلى الصفر، إلا أنها في المتوسط تساوي الصفر، أي $E(\varepsilon_i)=0$.

2- أن المتغير العشوائي يتوزع توزيعا طبيعيا، حول القيمة المتوقعة أو حول الوسط الحسابي المساوي للصفر عند كل قيمة من قيم المتغير المستقل x أي بشكل جرس.

3- أن تباين (Vriance)المتغير العشوائي (حد الخطأ)، حول الوسط الحسابي مقدار ثابت عند كل قيمة من قيم X أي: $var(\varepsilon_i) = E(\varepsilon_i)^2 = \sigma^2$

الفرضيات الثلاثة السابقة يمكن جمعها بشكل مختصر وتمثيلها كالآتي:

$$\varepsilon_i \sim N(0, \sigma^2)$$

بمعنى أن الخطأ العشوائي،ε_i، يتوزع، ~، توزيعا طبيعيا،N ، بوسط حسابي مساوي للصفر، 0، وتباين ثابت قيمته σ^2 .

[22]- دومينيك سالفاتور، ترجمة سعدية حافظ منتصر، **الإحصاء والاقتصاد القياسي**، سلسلة ملخصات شوم، الطبعة السادسة ، الدار الدولية للإستثمارات الثقافية، 2012، ص138.

[23]- حسين علي بخيت، سحر فتح الله، **الاقتصاد القياسي**، الطبعة الأولى، دار اليازوري، الأردن، 2009، ص38- ص41.

4- أن قيم ε_i غير مرتبطة بأي من المتغيرات المستقلة ، أي انعدام التباين المشترك Covariance بين ε_i و X_i أي : $Cov(\varepsilon_i X_i) = 0$.

5- القيم المختلفة للمتغير العشوائي (ε_i) تكون مستقلة عن بعضها البعض بعبارة أخرى قيم التباين المشترك ل ε_i مع ε_j مساوية للصفر، وعليه فإن قيمة المتغير العشوائي في أي فترة لا تعتمد على قيمته في فترة أخرى أي: $(i, j = 1,2,3, \dots, n,\ i \neq j)$ $\quad Cov(\varepsilon_i \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0$

6- انعدام العلاقة بين المتغيرات المستقلة.

ثالثا: تقدير نموذج الانحدار الخطي البسط بطريقة المربعات الصغرى العادية

طريقة المربعات الصغرى العادية (OLS) هي أسلوب لتوفيق " أفضل" خط مستقيم لعينة مشاهدات XY، وهو يتضمن تصغير مجموع المربعات لانحرافات النقاط (الرأسية) عن الخط إلى أدنى حد ممكن:

$$Min \sum \left(Y_i - \hat{Y}_i\right)^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots. (3)$$

حيث تشير Y_i إلى المشاهدات الفعلية، وتشير $\hat{Y}_i$ إلى القيم "الموافقة" المناظرة، بحيث تكون $Y_i - \hat{Y}_i =$ ε_i هي البواقي.

ويعطي هذا الأسلوب المعادلتين الطبيعتين التاليتين[24]:

$$\sum Y_i = n\hat{b}_0 + \hat{b}_1 \sum X_i$$

$$\sum X_i Y_i = \hat{b}_0 \sum X_i + \hat{b}_i \sum X_i^2$$

حيث n عدد المشاهدات، $\hat{b}_0$، $\hat{b}_1$ هي مقدرتان للمعلمتين الحقيقيتين b_0،b_1 .

وبحل المعادلتين أنيا، نحصل على:

$$\hat{b}_1 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

ونحصل على قيمة $\hat{b}_0$ كما يلي :

[24] – دومينيك سالفاتور، ترجمة سعدية حافظ منتصر، **مرجع سبق ذكره**، ص139.

$$\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}$$

ومن المفيد عادة استخدام صيغة مكافئة لتقدير $\hat{b}_1$: $\hat{b}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

حيث $x_i = X_i - \bar{X}$ ، $y_i = Y_i - \bar{Y}$ وتكون معادلة الإنحدار المقدرة بطريقة المربعات الصغرى (OLS):

$$\hat{Y}_i = \hat{b}_0 - \hat{b}_1\bar{X}$$

رابعا: خواص مقدرات طريقة المربعات الصغرى العادية

مقدرات المربعات الصغرى العادية (OLS) هي أفضل مقدرات خطية غير متحيزة (BLUE).

وعدم التحيز يعني $E(\hat{b}) = b$

بحيث أن: $Bias = E(\hat{b}) - b$

أما وصف مقدر بأنه " **أفضل مقدر غير متحيز** " أو انه **مقدر كفؤ** فيعني أنه ذو أصغر تباين. وبالتالي فإن مقدرات OLS هي الأفضل من بين كل المقدرات الخطية غير المتحيزة. وتعرف هذه الخاصية بنظرية " جاوس ماركوف "، وهي تمثل أهم مبرر لاستخدام OLS.

أحيانا، قد يرغب الباحث ان يقبل بعض التحيز في مقابل تباين أصغر، بتصغير متوسط مربع الخطأ، MSE

$$MSE(\hat{b}) = E(\hat{b} - b)^2 = var(\hat{b}) + (bias\hat{b})^2$$

ويكون المقدر **متسقا** إذا اقتربت قيمته من المعلمة الحقيقية مع اقتراب حجم العينة من ما لانهاية (بمعنى أنه غير متحيز في اللانهاية) وينتهي توزيعه إلى المعلمة الحقيقية.

مثال: الجدول يوضح معدل الدخل السنوي x ومقدار التوفير السنوي (بالدينار) y لعينة تتكون من 10 أسر:

Xi	12	15	9	22	16	36	10	11	25	18
Yi	0.6	1.1	0.2	2.4	1.2	3.6	0.4	0.6	1.7	1.1

$$\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y}) = 75.54 \text{ ، } \sum_{i=1}^{10} y_i = 12.9 \text{ ، } \sum_{i=1}^{10} x_i = 174 \text{ ، } \sum_{i=1}^{10} x_i y_i = 300$$

$$\sum_{i=1}^{10}(y_i-\bar{y})^2 = 9.749 \text{ ، } \sum_{i=1}^{10} x_i^2 = 3656 \text{ ، } \sum_{i=1}^{10} y_i^2 = 26.39 \text{ ، } \sum_{i=1}^{10}(x_i-\bar{x})^2 = 628.4$$

المطلوب:

1- كتابة النموذج الخطي ؟

2- تقدير معادلة الانحدار الخطية باستخدام طريقة المربعات الصغرى مستخدما معدل الدخل السنوي كمتغير مستقل ؟

الحل:

1- كتابة النموذج الخطي

$$Y_i = b_0 + b_1 X_i + \varepsilon_i \qquad i = 1, \cdots, 10$$

2- تقدير معادلة الانحدار الخطية باستخدام طريقة المربعات الصغرى مستخدما معدل الدخل السنوي كمتغير مستقل

$$\hat{Y}_i = \hat{b}_0 - \hat{b}_1 X_i$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{10}(x_i-\bar{x})^2} = \frac{75.54}{628.4} = 0.12$$

$$\bar{X} = \frac{\sum_1^{10} x_i}{N} = \frac{174}{10} = 17.4 \qquad \bar{Y} = \frac{\sum_1^{10} y_i}{N} = \frac{12.9}{10} = 1.29$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} = 1.29 - (0.12) * 17.4 = -0.798$$

إذن المعادلة المقدرة هي:

$$\hat{Y}_i = -0.798 + 0.12\, X_i \qquad i = 1, \cdots, 10$$

قائمة المراجع:

1-إسماعيل محمد بن قانة، **الإحصاء الوصفي والحيوي " دروس وتطبيقات "**، دار أسامة للنشر والتوزيع، الطبعة الأولى، الأردن، 2011.

2-جلاطو جيلالي، **الإحصاء مع تمارين محلولة**، الطبعة التاسعة، ديوان المطبوعات الجامعية، الجزائر، 2012.

3-عبان عبد القادر، **الإحصاء الوصفي للعلوم الاجتماعية والعلوم الاقتصادية**، الطبعة الأولى، دار ومضة للنشر والتوزيع والترجمة، الجزائر، 2022.

4-محمد صبحي أبو صالح، عدنان محمد عوض، **مقدمة في الإحصاء**، ديوان المطبوعات الجامعية، الجزائر، 1984.

5-عبد الحليم عشماوي، صلاح جلال، محمد حسين صادق، **الإحصاء الحيوي وتصميم التجارب**، المكتبة الأكاديمية، مصر، 2008.

6-أنيس إسماعيل كنجو، **الإحصاء والاحتمال**، الطبعة الأولى، مكتبة العبيكان، السعودية، 2000.

7-محمد حسين محمد رشيد، **الإحصاء الوصفي والتطبيقي والحيوي**، الطبعة الأولى، دائرة المكتبة الوطنية، الأردن، 2008.

8-عبد الحسين زيني، **الأرقام القياسية**، الطبعة الأولى، دار الحامد للنشر والتوزيع، الأردن، 2012.

9-غضبان عبدالله البدران، **الإحصاء الاستدلالي مع تطبيقات**، الطبعة الأولى، دار المناهج للنشر والتوزيع، 2021.

10-عبدالحميد عبدالمجيد البلداوي، **الأساليب التطبيقية لتحليل وإعداد البحوث العلمية مع حالات دراسية باستخدام برنامج SPSS** ، دار الشروق للنشر والتوزيع، عمان الأردن، 2008.

11-طه حسين الزبيدي، **مبادئ الإحصاء**، الطبعة الأولى، دار غيداء للنشر والتوزيع، 2012.

12-دومينيك سالفاتور، ترجمة سعدية حافظ منتصر، **الإحصاء والاقتصاد القياسي**، سلسلة ملخصات شوم، الطبعة السادسة، الدار الدولية للإستثمارات الثقافية، 2012.

13-حسين علي بخيت، سحر فتح الله، **الاقتصاد القياسي**، الطبعة الأولى، دار اليازوري، الأردن، 2009.