جامعة صالحي أحمد بالنعامة

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قسم التعليم المشترك

مطبوعة بيداغوجية في مقياس:

الإحصاء 1 (الإحصاء الوصفي)

محاضرات وتمارين

موجهة لطلبة السنة الأولى (جذع مشترك)

- ميدان: العلوم الاقتصادية و العلوم التجارية وعلوم التسيير
- الطور :ليسانس نظام(LMD)
- السداسي :الأول

من اعداد الدكتور: زقاي وليد

أستاذ محاضرب

السنة الجامعية 2025-2026

مقدمة:

يُعد علم الإحصاء أداة حيوية وأساسية في عالم الاقتصاد وإدارة الأعمال، حيث يوفر منهجًا منظمًا لجمع البيانات، تحليلها، وتفسيرها. في الاقتصاد، يُستخدم الإحصاء لفهم الاتجاهات الكلية مثل التضخم والبطالة والنمو الاقتصادي، مما يساعد الحكومات والمؤسسات على اتخاذ قرارات استراتيجية مبنية على حقائق. أما في إدارة الأعمال، فيُمكّن الإحصاء المديرين من تحليل أداء الشركات، تقييم المخاطر، التنبؤ بطلب المستهلكين، وتحسين العمليات الإنتاجية، وهو ما يساهم في تحقيق أهداف المؤسسة بكفاءة أعلى.

يهدف هذا المقياس إلى تمكين الطالب من اكتساب القدرات التحليلية اللازمة لمعالجة البيانات وتفسيرها ضمن سياق أكاديمي ومنهجي. يتمثل الهدف الأول في تزويد الطالب بالآليات الكفيلة ب**وصف وتصنيف وتلخيص البيانات** بشكل دقيق، مما يمثل حجر الزاوية في أي دراسة إحصائية. يليه هدف أساسي آخر، وهو تطوير مهارة **العرض البصري للبيانات** عبر استخدام الجداول والرسوم البيانية المناسبة، والتي تعد وسيلة أساسية لتوصيل المعلومات المعقدة بوضوح وفاعلية. وأخيرًا، يسعى المقياس إلى تدريب الطالب على **حساب وتفسير المقاييس الإحصائية** المختلفة، مما يمكنه من استخلاص خصائص المتغيرات قيد الدراسة وتحليل سلوكها في المجتمع الإحصائي.

لتحقيق أقصى استفادة من هذا المقياس، يُتوقع من الطالب أن يمتلك **معرفة أساسية بالعمليات والقواعد الرياضية** التي تم تناولها في المرحلتين المتوسطة والثانوية. لا يهدف المقياس إلى التعمق في الجانب الرياضي النظري، بل إلى توظيف هذه الأدوات كمبادئ لتمكين الفهم الإحصائي. وبالتالي، فإن الخلفية الرياضية المطلوبة لا تتجاوز الحدود المألوفة، مما يسمح بالتركيز على الجوانب التطبيقية والتحليلية للمفاهيم الإحصائية التي سيتم استعراضها.

يهدف هذا المقياس إلى تزويد الطالب بأساس متين في علم الإحصاء الوصفي، وذلك عبر استعراض سلسلة من المحاور المترابطة التي تبدأ من المفاهيم الأساسية وتنتهي بالتطبيقات المتقدمة:

1. **نظرة عامة حول علم الإحصاء:**
   - ما المقصود بعلم الإحصاء؟ لماذا ندرس الإحصاء؟
   - تطبيقات الإحصاء في الاقتصاد وإدارة الأعمال.
   - أنواع البيانات وتصنيف المتغيرات.
   - مصادر البيانات ومفاهيم حول العينات وطرق الحصول عليها.

مقدمة:

2. **العرض الجدولي للبيانات** :العرض الجدولي للمتغيرات الكمية المستمرة- العرض الجدولي للمتغيرات الكمية المتقطعة. العرض الجدولي للمتغيرات النوعية. التكرار المطلق. التكرار النسبي. التكرارات المجمعة الصاعدة والنازلة.
3. **العرض البياني للبيانات** :العرض البياني للمتغيرات الكمية المستمرة - العرض البياني للمتغيرات الكمية المتقطعة. العرض البياني للمتغيرات النوعية.
4. **مقاييس النزعة المركزية** :المتوسط الحسابي، المتوسط الهندسي، المتوسط التوافقي والمتوسط التربيعي. الوسيط وأشباه الوسيط (المئينات، العشيريات والربيعيات)، المنوال. مقارنة بين مقاييس النزعة المركزية.
5. **مقاييس التشتت** :مقاييس التشتت المطلقة (المدى العام، المدى الربيعي، الانحراف المتوسط، التباين والانحراف المعياري). مقاييس التشتت النسبي (معامل الاختلاف، البيانات المعيارية.(
6. **مقاييس الشكل** :حساب العزوم، مقاييس الالتواء (بيرسون، فيشر، .....)، مقاييس التفرطح (بيرسون، فيشر، .....).
7. **مقاييس التمركز** :منحنى لورنز- مؤشر جيني.
8. **الأرقام القياسية** :الأرقام القياسية البسيطة، الأرقام القياسية المجمعة، الأرقام القياسية المرجحة.
9. **الارتباط والانحدار** :توزيعات المتغيرات ثنائية التغير (جداول التوافق والتكرارات المشتركة، الهامشية والشرطية). الارتباط بين متغيرين كيفيين (إحصاء كاي مربع، ومعامل الارتباط الخطي). الارتباط بين متغيرين مستمرين (سحابة النقاط ومعامل الارتباط الخطي، الانحدار الخطي البسيط).

يُقيّم أداء الطالب في هذا المقياس بناءً على نظام شامل يجمع بين **التقييم المستمر** و**الامتحان النهائي** .هذا النظام يهدف إلى قياس مدى استيعابك للمفاهيم بشكل تراكمي. سيُحتسب المعدل النهائي للمادة بناءً على وزن ترجيحي، حيث تُشكل الدروس النظرية نسبة **60%** من المعدل، بينما تُشكل الأعمال الموجهة والتطبيقات العملية نسبة **40%**. هذا التوزيع يضمن أن يُعكس تقييمك ليس فقط معرفتك النظرية، بل أيضًا قدرتك على تطبيق هذه المعارف عمليًا.

باختصار، يمثل هذا المقياس بوابة أساسية إلى عالم التحليل الإحصائي، حيث يوفر لك الأدوات والمفاهيم اللازمة لتحويل البيانات الخام إلى معلومات قيمة. إن الإلمام بهذه المبادئ لا يقتصر على النجاح الأكاديمي، بل هو استثمار حقيقي في مهاراتك التحليلية التي لا غنى عنها في العديد من المجالات المهنية. نأمل أن تكون هذه الرحلة التعليمية ممتعة ومفيدة، وأن تكتشفوا من خلالها القوة الحقيقية للإحصاء في فهم العالم من حولنا.

الفهرس

الفهرس

الفصل الأول: مدخل عام إلى علم الإحصاء

يُعد علم الإحصاء من العلوم الأساسية لفهم الظواهر الاقتصادية والاجتماعية والإدارية، إذ يهتم بجمع البيانات وتنظيمها وتلخيصها وعرضها بطرق كمية وبيانية تسهّل تفسيرها واستخلاص خصائص المجتمع محل الدراسة. وقد تطور من كونه أداة لتعداد السكان وحصر الموارد إلى وسيلة علمية محورية في البحث وصنع القرار. لا تقتصر أهميته على توفير المؤشرات، بل تكمن في تبسيط البيانات المعقدة وتحويلها إلى استنتاجات دقيقة تساعد على اتخاذ قرارات رشيدة، مما يجعله عنصرًا أساسيًا في تكوين الطالب الجامعي خاصة في الاقتصاد والتسيير.

1 ماهية علم الإحصاء:

الإحصاء علم رياضي تطبيقي يختص بجمع البيانات ومعالجتها وتحويلها إلى مؤشرات قابلة للتحليل. وهو اليوم أداة أساسية لفهم الظواهر وصنع القرار في المجالات الاقتصادية والاجتماعية والإدارية.

1.1 تعريف وأصل علم الإحصاء: يعود أصل كلمة "إحصاء (Statistics) "إلى الكلمة اللاتينية "status" والتي تعني "الأرض" أو "الدولة". تطور المفهوم لاحقاً لتظهر كلمة "statisticum" التي تشير إلى كل ما هو متعلق بالدولة. ظل هذا المفهوم محدوداً في البداية، مقتصراً على وصف حالة أو وضع الدولة بالأرقام، كعدد السكان، القدرة العسكرية، أو الثروات (تيلوت، 2009، ص. 9). شهد علم الإحصاء الحديث، كعلم قائم بذاته، بدايته الحقيقية مع مطلع القرن الثامن عشر، بفضل إسهامات باحثين رياضيين بارزين أمثال **قوس (Gauss)** و**بيرنولي (Bernoulli)** و**لابلاس (Laplace)**، الذين وضعوا الأسس الأولى للتحليل الإحصائي وأرسوا القوانين الاحتمالية. استمر هذا العلم في التطور خلال القرن العشرين، حيث اكتملت أسسه ليصبح علماً شاملاً يهتم بدراسة جمع المعطيات، تحليلها، معالجتها، تفسير النتائج وتمثيلها، بهدف جعل هذه المعطيات المشاهدة قابلة للفهم من قبل الجميع (راتول، 2006، ص. 12).

يمكن تعريف علم الإحصاء بأنه: "فرع من فروع الرياضيات التطبيقية ومجموعة من التقنيات والطرق التي تدرس الظواهر من خلال جمع البيانات ومعالجتها وتحليلها وعرضها وتفسير النتائج وجعلها مفهومة للجميع" (عزوز، 2010، ص. 15). يساهم علم الإحصاء في تحقيق مجموعة من الأهداف الأساسية، تشمل (جلاطو، 2001، ص. 20):

- عد وحصر الأشياء المراد توفير بيانات إحصائية عنها.
- تقليل الأعداد الكبيرة إلى بُعد يسهل فهمها واستيعابها.
- جمع البيانات اللازمة للوصول إلى الأهداف المحددة.

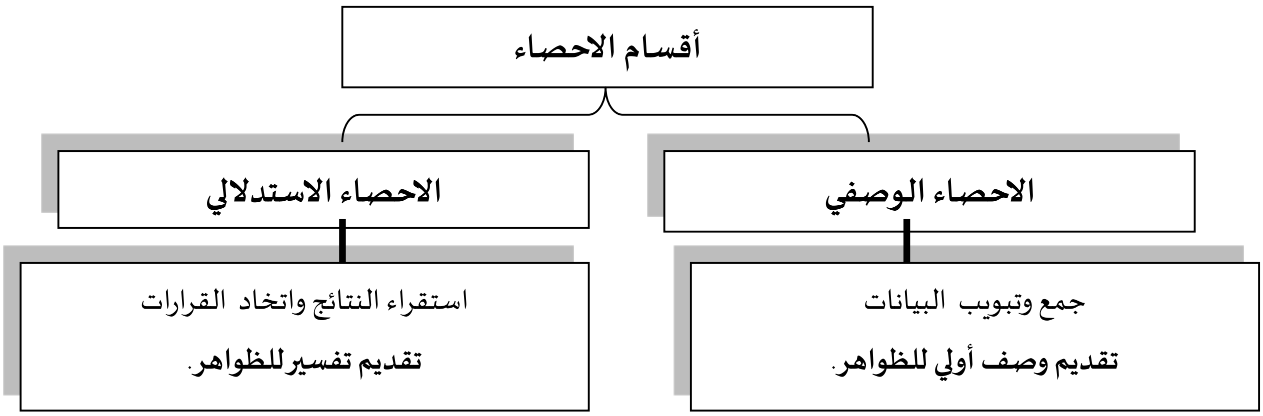
- دراسة البيانات وتحليلها للوصول إلى النتائج المنطقية.
- المساهمة الفعالة في اتخاذ القرار المناسب لكل مشكلة أو ظاهرة.

2.1 أنواع علم الإحصاء: ينقسم علم الإحصاء إلى قسمين رئيسيين، يتميز كل منهما بأهدافه ومنهجياته الخاصة (انظر الشكل رقم 01):

أ. الإحصاء الوصفي: (Descriptive Statistics) هو القسم من علم الإحصاء الذي يهدف إلى وصف المعطيات المجمعة وتلخيصها بطريقة موجزة وواضحة، وذلك من خلال دراسة بعض الظواهر (ذات الطابع الاقتصادي، الاجتماعي، أو التجريبي، إلخ). يتم ذلك عبر تنظيم البيانات في جداول وعرضها بيانياً، أو حساب بعض المؤشرات الإحصائية البسيطة التي تدل على طبيعة البيانات وتوزيعها. يسهل هذا القسم على الباحث الحصول على معلومات مركزة ثم تحليلها للوصول إلى نتائج أولية عن الظواهر المتعددة. إن الهدف الجوهري للإحصاء الوصفي هو وصف البيانات المرصودة بطريقة تركيبية وذات معنى لتحليلها بشكل أفضل(Hamdani, 2005, p. 18) .

ب. الإحصاء الاستدلالي: (Inferential Statistics) يُعرف أيضاً بالإحصاء الرياضي أو الإحصاء الاستقرائي. يهدف الإحصاء الاستدلالي إلى فهم الظواهر المتعلقة بمجتمع معين بالاستناد إلى ملاحظة ودراسة جزء محدود من هذا المجتمع، يُطلق عليه **العينة**. فالإحصاء الاستدلالي هو عملية تحليلية للنتائج تهدف إلى تفسير الظواهر، والتوصل إلى استنتاجات معممة من البيانات المتوفرة، والتي تُستخدم لاحقاً في التنبؤ والاستشراف واتخاذ القرارات .(Anderson et al., 2015, p. 25)

الشكل رقم 01: أقسام علم الاحصاء.



المصدر :من إعداد الأستاذ بالاعتماد على المراجع المذكورة أنفا.

3.1 أهمية الإحصاء في الاقتصاد وإدارة الأعمال:

للإحصاء أهمية جوهرية في جميع المجالات العلمية والتطبيقية، وحتى في حياتنا اليومية. وقد كان له دور كبير ومحوري في تطور العلوم الاقتصادية وعلوم التسيير (الإدارة):

أ. **علاقة علم الإحصاء بعلم الاقتصاد** :تُعد الإحصاءات ضرورية للتنمية الاقتصادية وجهود تحقيق الأهداف الإنمائية. يُعتبر علم الإحصاء أداة أساسية للتحليل الاقتصادي، حيث يمكّن من التحقق من "القوانين" والنظريات المختلفة لسلوك الأعوان الاقتصاديين. يقوم علم الإحصاء بتحليل المشاكل الاقتصادية وتوفير المعلومات اللازمة عنها، ثم يقوم بتحليلها ودراستها بشكل منطقي، مما يوفر الحلول التي تساعد على اتخاذ القرارات السليمة لحل تلك المشكلات. لقد أثمرت العلاقة الوثيقة بين علم الإحصاء وعلم الاقتصاد عن ظهور نظرية مستقلة تُعرف بـ **علم الاقتصاد القياسي(Econometrics)** ، والذي يختص بقياس الظواهر والمشكلات الاقتصادية باستخدام الأدوات والأساليب الإحصائية (عزوز، 2011، ص. 30).

ب. **الإحصاء وإدارة الأعمال** :تكمن أهمية الإحصاء في إدارة الأعمال بشكل عام في مساهمته الفعالة في عملية **اتخاذ القرارات** المتعلقة بمستقبل ومصير المؤسسة، وذلك بناءً على البيانات التي تم جمعها ومعالجتها. فالإحصاء يساعد على توفير البيانات اللازمة التي تُسهم في تشخيص وتحليل واقع المؤسسة، مما يسهم في تحديد مكامن الضعف والقوة، وبالتالي تحقيق الاستعمال الأمثل لمواردها(Goldfarb & Pardoux, 2011, p. 45). كما يمكّن الإحصاء المؤسسة من التموقع بفاعلية في بيئتها التنافسية، بالاعتماد على الإحصائيات المتعلقة بالمبيعات والأرباح التي تحققها المؤسسات المنافسة. وتُعتبر المعطيات والبيانات الإحصائية أداة مهمة لعمليات **التخطيط** وتحديد الأهداف بناءً على واقع المؤسسة، من خلال تحليل البيئة الداخلية والخارجية. وأخيراً، يسمح علم الإحصاء بتحديد البدائل المتاحة، من خلال دراسة مختلف الاحتمالات المتعلقة بظاهرة معينة. أي، باستخدام **نظرية الاحتمالات،** يمكن توقع نتائج متعددة، وتحديد كافة الإجراءات التي تصاحب هذه النتائج، مما يجعل الاحتمال طريقة قوية للسيطرة بشكل كبير على المخاطر التي قد تتعرض لها المؤسسة في المستقبل.(Anderson et al., 2015, p. 60)

2 المفاهيم الأساسية و أنواع البيانات في الإحصاء:

قبل الغوص في تفاصيل التحليل الإحصائي، من الضروري فهم مجموعة من المفاهيم الأساسية التي تشكل حجر الزاوية في أي دراسة إحصائية:

1.2 المفاهيم الأساسية: تشكل هذه المفاهيم القاعدة النظرية التي تُبنى عليها جميع المراحل اللاحقة من جمع البيانات وتحليلها وتفسيرها (انظر الشكل رقم 02):

1.1.2 الوحدة الإحصائية أو الفرد(Statistical Unit / Individual) : هي وحدة الملاحظة أو القياس التي يتم جمع البيانات منها أو استخلاصها. وبالتالي، فإن الوحدة الإحصائية هي العنصر الأساسي في عملية تجميع وتبويب البيانات الإحصائية. على سبيل المثال، في دراسة عن دخل الأسر، تكون "الأسرة" هي الوحدة الإحصائية (راتول، 2006، ص. 25).

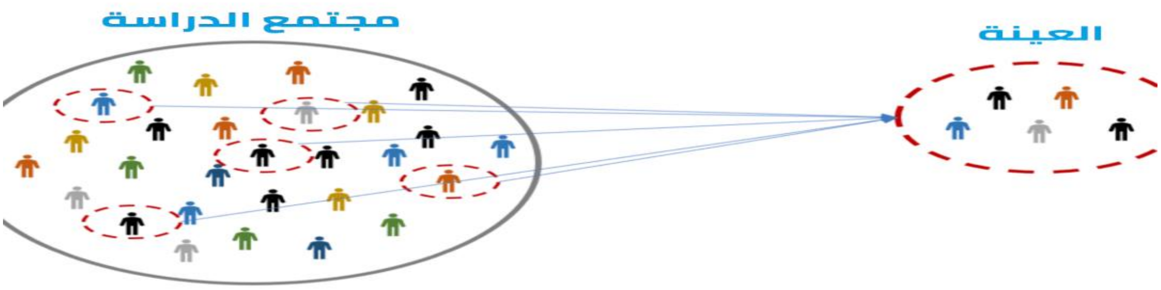
2.1.2 المجتمع الإحصائي(Population): المجتمع الإحصائي هو كل العناصر أو المفردات التي تشترك في خاصية محددة تهم الباحث في دراسته. يمثل "الكل" الذي يتم البحث فيه، مثل جميع سكان مدينة أو كل طلاب كلية معينة في عام دراسي محدد. يُرمز للعدد الإجمالي لعناصر المجتمع بالرمز(**N**) (تيلوت، 2009، ص. 15).

3.1.2 العينة (Sample) :نظرًا لصعوبة دراسة المجتمع الإحصائي ككل (سواء لكبر حجمه أو لارتفاع التكلفة أو استغراق الوقت)، فإننا غالبًا ما نقتصر على دراسة جزء من هذا المجتمع يُسمى بـ "العينة .(Échantillon) " يجب أن تكون العينة ممثلة للمجتمع الذي سُحبت منه، وأن تستجيب لشروط ومعايير محددة لضمان صلاحية النتائج المستخلصة وتعميمها على المجتمع بأكمله (عزوز، 2010، ص. 40).

4.1.2 الصفة الإحصائية (Characteristic) :تُعرف الصفة الإحصائية بأنها الخاصية التي يمكن ملاحظتها وقياسها في الأفراد أو المجتمعات، وتُستخدم للتمييز بين وحدات المجتمع الإحصائي .لكل صفة عدة "كيفيات" تمثل الحالات التي يمكن أن تظهر عليها .على سبيل المثال، صفة "الجنس" لها كيفيتان هما "ذكر" و "أنثى"، بينما صفة "اللون" لها كيفيات مثل "أبيض" و "أسود." (جلاطو، 2001، ص. 35).

5.1.2 المتغير الإحصائي (Variable) :يعتبر المتغير الإحصائي هو التعبير الكمي أو النوعي للصفة التي يأخذها العنصر الإحصائي (الوحدة الإحصائية) عند دراسة مجموعة أو مجتمع معين. يأخذ المتغير قيمًا معينة وفقًا للوحدات الإحصائية التي تمت دراستها (مثل الطول المقاس بالسنتيمتر، لون العين كفئة). اعتمادًا على القيم التي يمكن أن يأخذها المتغير، فإنه يُقسم إلى فئتين رئيسيتين: الكمية والنوعية(Hamdani, 2005, p. 30) .

الشكل رقم 02: المفاهيم الأساسية في الاحصاء



المصدر :من إعداد الأستاذ بالاعتماد على المراجع المذكورة أنفا.

2.2. أنواع البيانات وتصنيف المتغيرات(Types of data and variables) :

من التعريف السابق لعلم الإحصاء، يتضح أنه العلم الذي يهتم بجمع البيانات. يُعد نوع البيانات وطريقة قياسها من أهم العوامل التي تحدد نوع التحليل الإحصائي المناسب (راتول، 2006، ص. 35).تُصنف البيانات بشكل عام إلى مجموعتين رئيسيتين هما :البيانات الكمية **(Quantitative Data)** والبيانات النوعية **(Qualitative Data)** (كما هو موضح في الشكل رقم 03).

1.2.2. المتغيرات الكمية(Quantitative Data) : هي المتغيرات التي يمكن قياسها عددياً، وتتميز بقيم رقمية تعبر عن كميات قابلة للعد أو القياس. تُقسم المتغيرات الكمية إلى نوعين:

أ. **المتغيرات المستمرة (Continuous Variables)** :هي المتغيرات التي يمكن أن تأخذ أي قيمة ضمن مجال أو فترة معينة من المشاهدات. تنتمي هذه القيم إلى مجموعة الأعداد الحقيقية وتحوي عددًا لا نهائيًا من القيم الممكنة بين أي نقطتين محددتين.أمثلة: درجات الحرارة (يمكن أن تكون 25.5 درجة مئوية)، الطول (170.2 سم)، الوزن (68.7 كيلوغرام)، الدخل الشهري، الوقت...الخ

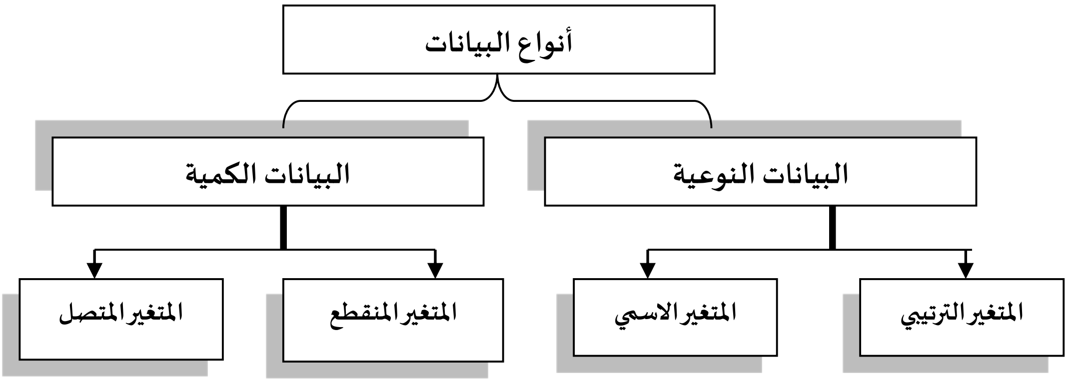
ب. **المتغيرات المنفصلة (Discrete Variables)** :هي متغيرات عددية تأخذ فقط قيمًا صحيحة ومحددة، وغير قابلة للتجزئة أو الكسور ضمن مدى معين، وتنتمي عادةً إلى مجموعة الأعداد الصحيحة. أمثلة: عدد أفراد الأسرة (لا يمكن أن يكون 3.5 أفراد)، عدد الحوادث الأسبوعية، عدد الكتب في مكتبة ....الخ

2.2.2 البيانات الوصفية أو النوعية (Qualitative Data): هي بيانات غير رقمية في جوهرها، أو بيانات رقمية تُستخدم لتمثيل فئات أو مستويات محددة. تُقاس البيانات الوصفية بمعيارين رئيسيين (Goldfarb & Pardoux, 2011, p. 50):

أ.   **بيانات وصفية مقاسة بمعيار اسمي (Nominal Scale)** :وهي بيانات غير رقمية تتكون من مجموعات متناهية، لكل مجموعة خصائص تميزها عن المجموعات الأخرى. الأهم هو أن هذه المجموعات لا يمكن المفاضلة أو الترتيب بينها بشكل طبيعي (أي لا يوجد ترتيب منطقي). أمثلة: الجنسية (جزائري، تونسي، مصري)، الأعمال الحرة (طبيب، مهندس، محامي)، فصيلة الدم(A, B, AB, O) ، الحالة الاجتماعية (أعزب، متزوج، مطلق، أرمل).

ب. **بيانات وصفية مقاسة بمعيار ترتيبي (Ordinal Scales)** :وتتكون هذه البيانات من مستويات أو فئات يمكن ترتيبها تصاعديًا أو تنازليًا حسب خاصية معينة، لكن الفروقات بين المستويات قد لا تكون متساوية أو ذات معنى كمي. أمثلة: **المستوى التعليمي** :متغير وصفي تقاس بياناته بمعيار ترتيبي: (ابتدائية، متوسطة، ثانوي، جامعية،.....)، **تقدير الطالب** :(ضعيف، مقبول، جيد، ممتاز...)، **مستوى الرضا** : (غير راضٍ، راضٍ جزئياً، ..)

الشكل رقم 03: أنواع البيانات الاحصائية



المصدر :من إعداد الأستاذ بالاعتماد على المراجع المذكورة أنفا.

<u>تمرين تطبيقي: تحديد المفاهيم الأساسية و أنواع البيانات</u>

في الأمثلة التالية، قم بتحديد :**المجتمع الإحصائي، الوحدة الإحصائية، المتغير الإحصائي (الصفة(، طبيعة ونوع المتغير الإحصائي، والكيفيات الممكنة:**

1-الحالة العائلية لسكان مدينة النعامة. 2-نوع الجنس لأفراد القرية.3- عدد الأطفال لكل أسرة في حي 256 سكن بولاية النعامة.4- ظروف العمل في مصنع البسكويت. 5- الأجر الشهري لعمال مصنع البسكويت.6- أعمار فريق كرة القدم لولاية النعامة أكابر.7- أوزان طلبة السنة الأولى علوم اقتصادية للمركز الجامعي أحمد صالحي.

حل التمرين التطبيقي:

الرقم	المجتمع الاحصائي	العنصر الاحصائي	المتغير الإحصائي(الصفة)	طبيعة ونوع المتغيرات	الكيفيات
1	سكان مدينة النعامة	كل فرد من سكان مدينة النعامة	الحالة العائلية	**نوعي اسمي**	متزوج، مطلق، أرمل، أعزب
2	سكان القرية	كل فرد من القرية	الجنس	**نوعي اسمي**	ذكر، أنثى
3	الأسر في حي 256 سكن بالنعامة	كل أسرة من أسر حي 256 سكن	عدد الأطفال في كل أسرة	**كمي منفصل**	0، 1، 2، 3،......الخ
4	عمال مصنع البسكويت	العامل في مصنع البسكويت	ظروف العمل	**نوعي ترتيبي**	جيدة، متوسطة، سيئة
5	عمال مصنع البسكويت	العامل في مصنع البسكويت	الأجر الشهري	**كمي مستمر (متصل)**	18000دج، 35000دج، .....
6	أفراد فريق كرة القدم أكابر	كل فرد من فريق كرة القدم	السن	**كمي مستمر (متصل)**	18 سنة، 20 سنة، ....
7	طلبة السنة الأولى علوم اقتصادية	كل طالب من طلبة السنة الأولى علوم اقتصادية	الوزن	**كمي مستمر (متصل)**	50كلغ، 60كلغ ، 76.5كلغ، 90كلغ،....

3 جمع البيانات ومفاهيم حول العينات وطرق الحصول عليها.

تعتمد الطريقة الإحصائية لحل المشكلات أو دراسة الظواهر على مجموعة من المراحل المنهجية التي تضمن الوصول إلى نتائج دقيقة وموثوقة (موسى، 2016، ص. 35):

1) **جمع البيانات** :هي الخطوة الأولى التي يتم فيها الحصول على المعلومات الأولية، أو قيم المشاهدات، أو القياسات المتعلقة بالدراسة.

2) **تنظيم البيانات وعرضها** :تتضمن هذه المرحلة ترتيب المعلومات في جداول منظمة وعرضها بطرق مناسبة باستخدام الأشكال البيانية والرسوم التوضيحية لتسهيل فهمها.

3) **تحليل البيانات** :يتم في هذه المرحلة تطبيق الأساليب الإحصائية المختلفة لإيجاد قيم لمقاييس ومؤشرات تُحدد من البيانات المدروسة.

4) **استقراء النتائج و اتخاذ القرارات** :هي المرحلة النهائية التي يتم فيها التوصل إلى الاستنتاجات، والتي قد تكون على شكل تقديرات، أو تنبؤات، أو تعميمات، أو قرارات تتعلق بقبول أو رفض الفرضيات الإحصائية.

1.3. جمع البيانات: يُعتبر جمع البيانات الممثلة للمجتمع خطوة جوهرية في أي بحث إحصائي، إذ يشترط لتأسيس نتائج صحيحة تحديد الظاهرة المدروسة والمجتمع محل الدراسة بدقة قبل الشروع في عملية الجمع (السعود، 2018، ص. 40).

1.1.3.مصادر جمع البيانات: يمكن الحصول على البيانات من مصدرين رئيسيين: (انظر الشكل رقم 04).

أ. **المصادر الأولية :(Primary Sources)** هي المصادر التي يتم الحصول منها على البيانات بشكل مباشر، حيث يقوم الباحث بنفسه بجمع البيانات من المفردة محل البحث. تتميز بالدقة والثقة العالية، لأن الباحث هو من يجمعها مباشرة. ومع ذلك، تتطلب وقتًا وجهدًا كبيرين، وتعد مكلفة (العاني وآخرون، 2017، ص. 50).

ب. **المصادر الثانوية(Secondary Sources)** : هي تلك التي تُستقى منها البيانات بشكل غير مباشر، حيث جُمعت مسبقًا من طرف أفراد أو هيئات متخصصة مثل نشرات الديوان الوطني للإحصاءات. وتمتاز بتقليل الوقت والجهد والتكلفة، غير أن مستوى دقتها قد يكون أقل من المصادر الأولية **(Hamdani, 2005, p. 70)**.

2.1.3. أسلوب جمع البيانات : من المهم التمييز بين مجتمع الدراسة والعينة قبل تحديد أساليب جمع البيانات.

- **المجتمع(Population)** : المجتمع الإحصائي هو الإطار الكلي الذي يضم جميع المفردات ذات الخصائص المشتركة محل الدراسة. ويُعد الكيان الشامل الذي يسعى الباحث إلى توصيفه أو استنتاج خصائصه، وقد يكون منتهيًا يمكن حصر عناصره أو غير منتهٍ يتعذر إحصاؤه بدقة(Hamdani, 2005, p. 25) .

- **العينة (Sample)** :هي مجموعة جزئية من المجتمع يتم اختيارها بطريقة معينة للدراسة. تُستخدم العينة نظرًا لصعوبة دراسة المجتمع ككل (لأسباب تتعلق بالوقت، الجهد، التكلفة، أو استحالة الحصر). يجب أن تكون العينة ممثلة للمجتمع بشكل صادق لضمان إمكانية تعميم النتائج (موسى، 2016، ص. 40).

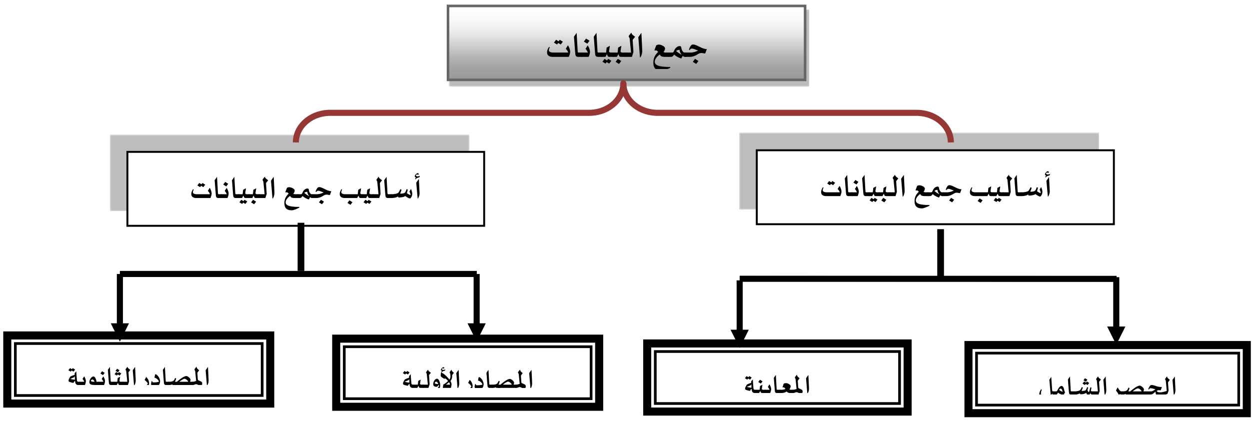
يتم جمع البيانات وفق **أسلوبين إحصائيين رئيسيين**: أسلوب الحصر الشامل وأسلوب المعاينة (انظر الشكل 04).

أ. **أسلوب الحصر (المسح) الشامل (Census Method)** :يتضمن جمع البيانات من جميع عناصر المجتمع الإحصائي. يمتاز بالدقة العالية والمصداقية لشموله جميع المفردات. لكن عيوبه تشمل التكاليف المرتفعة، الحاجة إلى وقت وجهد كبيرين، وتطلب عددًا كبيرًا من الباحثين (العاني وآخرون، 2017، ص. 65).

ب. **أسلوب المعاينة :(Sampling Method)** يعتمد على دراسة جزء من المجتمع الكلي. يتطلب هذا الأسلوب حرصًا شديدًا عند اختيار العينة لتكون ممثلة، بهدف تعميم نتائجها على المجتمع. يتميز بتقليل الوقت والجهد

والتكلفة، ويمكن من الحصول على بيانات أكثر تفصيلاً. يُفضل استخدامه عندما يكون الحصر الشامل صعبًا أو مستحيلاً (السعود، 2018، ص. 70).

الشكل رقم 04: مصادر وأساليب جمع البيانات

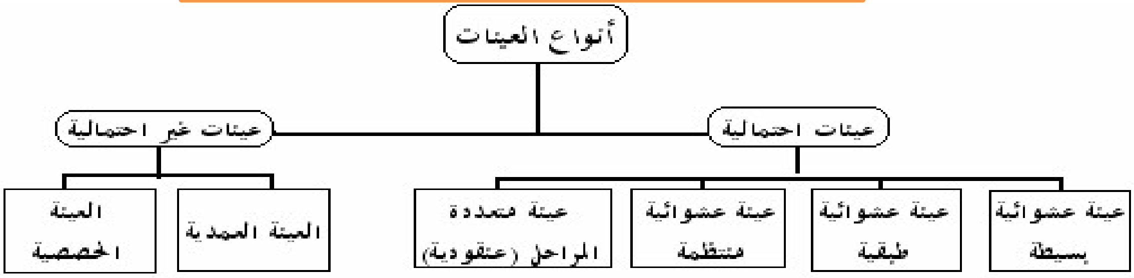


المصدر :من إعداد الأستاذ بالاعتماد على المراجع المذكورة أنفا.

2.3 أنواع العينات

يتوقف نجاح استخدام أسلوب المعاينة على عدة عوامل، منها: كيفية تحديد حجم العينة، طريقة اختيار مفردات العينة، ونوع العينة المختارة. تُقسم العينات، وفقًا لأسلوب اختيارها، إلى نوعين رئيسيين (Hamdani, 2005, p. 80): العينات الاحتمالية والعينات غير الاحتمالية (انظر الشكل رقم 05).

الشكل رقم 05: أنواع العينات الاحصائية



المصدر :موسى، محمد. (2016).

1.2.3العينات الاحتمالية(Probability Samples) : هي العينات التي يتم اختيار مفرداتها وفقًا لقواعد الاحتمالات، أي بطريقة عشوائية بحتة. تهدف هذه الطريقة إلى تجنب التحيز وتُمكن من تعميم النتائج على المجتمع المدروس بثقة إحصائية (موسى، 2016، ص. 85). ومن أهم أنواعها:

أ. **العينة العشوائية البسيطة** :تُختار بحيث تكون فرص الاختيار متكافئة لجميع أفراد المجتمع، مما يضمن إعطاء كل فرد نفس الفرصة للظهور في العينة (العاني وآخرون، 2017، ص. 90).

ب. **العينة العشوائية الطبقية** :يُقسم المجتمع أولاً إلى مجموعات فرعية متجانسة تُسمى **طبقات** .يتم اختيار مفردات من كل طبقة بما يتناسب مع حجمها. تُستخدم عندما يكون المجتمع مكونًا من طبقات غير متداخلة (مثلاً: ذكور/إناث) (السعود، 2018، ص. 95).

$$\textbf{حجم العينة} = \frac{\text{حجم الطبقة * حجم العينة}}{\text{حجم المجتمع}}$$

ت. **العينة العشوائية المنتظمة**: تُختار بترقيم عناصر المجتمع ثم تحديد فاصل ثابت للاختيار بعد اختيار أول عنصر عشوائيًا(Hamdani, 2005, p. 90) .مثال: لاختيار 5 طلاب من 50 (فاصل 10)، نختار رقمًا عشوائيًا بين 1-10 (مثلاً 4)، ثم نختار الطلاب 4، 14، 24، 34، 44.

ث. **العينة العنقودية أو المتعددة المراحل (Cluster Sample / Multi-stage Sample)** :يُقسم المجتمع إلى مجموعات طبيعية (عناقيد)، ثم يتم اختيار عينات عشوائية من هذه العناقيد، وتُدرس جميع المفردات ضمن العناقيد المختارة. يمكن أن يتم الاختيار على عدة مراحل (موسى، 2016، ص. 100).

2.2.3 العينات غير الاحتمالية(Non-Probability Samples)

يتم اختيار مفرداتها بطريقة غير عشوائية، بناءً على حكم الباحث أو سهولة الوصول. لا يمكن تعميم نتائجها على المجتمع بأكمله بثقة إحصائية (العاني وآخرون، 2017، ص. 110). أهم أنواعها:

أ. **العينة العمدية (Judgmental / Purposive Sample)** :تُختار بحيث تتوافر في كل عنصر شروط محددة يراها الباحث ضرورية للدراسة، ويعتمد الاختيار على خبرة الباحث.

ب. **العينة الحصصية (Quota Sample)** :تُختار لتشمل خصائص المجتمع بنسب معطاة مسبقًا، لكن اختيار الأفراد داخل كل حصة يكون غير عشوائي. مثال: اختيار 20 ذكرًا و30 أنثى من عينة 50 شخصًا لتمثيل نسبة 40% ذكور و60% إناث في المجتمع.

الفصل الثاني: العرض الجدولي للبيانات الاحصائية

إن عملية جمع البيانات الأولية، على أهميتها، لا تُجدي نفعًا ما لم يتم تبويبها وتحليلها بشكل منهجي .وفي هذا السياق، يُعد **العرض الجدولي للبيانات** أداة محورية في الإحصاء الوصفي، حيث يحوّل البيانات الخام إلى معلومات منظمة يسهل فهمها واستخلاص الاستنتاجات منها .يُعرف هذا الإجراء بإنشاء **جداول التوزيع التكراري**، التي تُعتبر حجر الزاوية في التحليل الإحصائي (عبد الفتاح، 2012، ص27). تختلف بنية هذه الجداول وطريقة تصميمها تبعًا لطبيعة المتغيرات الإحصائية، سواء كانت كمية منفصلة، كمية مستمرة، أو نوعية.

1 العرض الجدولي للمتغيرات الكمية المنفصلة

يُعد العرض الجدولي للمتغيرات الكمية المنفصلة خطوة أساسية لتلخيص البيانات الخام، حيث يُسجّل كل قيمة فريدة للمتغير مع عدد مرات تكرارها. تُشكل هذه الجداول حجر الزاوية في التحليل الإحصائي، وتُستخدم لتجسيد التوزيعات التكرارية المطلقة والنسبية.

1.1. التوزيع التكراري المطلق (Frequency Distribution)

يُمثّل التوزيع التكراري المطلق طريقة لتلخيص بيانات المتغير العشوائي وتصنيفها ضمن مجموعات. يُعرض هذا التوزيع في جدول بسيط يتكون من عمودين رئيسيين (طيار، 2019، ص. 91).

أ. **المتغير الاحصائي Xi**: في هذا العمود، تُقسم البيانات إلى فئات مصنفة حسب النوع أو القيمة العددية:

- **للمتغير الكيفي (النوعي)** :تُوضع في هذه الخانات صفة المتغير (مثل الجنس، الحالة الاجتماعية).
- **للمتغير الكمي المنفصل** :تُكتب قيم المتغير، ويجب أن تكون مرتبة تصاعديًا أو تنازليًا (مثل عدد الأطفال، عدد المنتجات المعيبة).

ب. **التكرار المطلق** (ni): **و يمثل عمود التكرارات (Frequency) ويتضمن** عدد المرات التي تكررت فيها الفئة أو القيمة. يُرمز له بالرمز (ni). في هذه الخانات.

الشكل رقم 01: الشكل العام لجدول التوزيع التكراري

في حالة متغير كيفي نضع في هذه الخانات صفة المتغير.
في حالة متغير كمي منفصل نكتب قيم المتغير ويجب أن ترتب تصاعديا أو تنازليا.
في حالة متغير كمي متصل على شكل فئات.

المتغير X_i	التكرار المطلق n_i
X_1	N_1
X_2	N_2
.	.
X_k	N_k
المجموع Σ	$N = \Sigma ni$

نضع في هذه الخانات عدد المفردات المقابلة لكل صفة أو قيمة أو فئة للمتغير الإحصائي

المصدر: عبد الفتاح، عز حسن. (2012).

2.1 .جداول التوزيعات التكرارية النسبية

يُحسب التكرار النسبي لأي فئة بقسمة تكرارها المطلق على العدد الكلي للمشاهدات، ويُعبّر عنه بالصيغة التالية:

التكراري النسبي $fi = \dfrac{ni \text{ تكرار الفئة}}{\Sigma ni \text{(التكرارات) العدد الكلي للمشاهدات}}$

ويمكن تحويله إلى **تكرار مئوي** بضرب النتيجة في 100، على النحو التالي:

$$fi\ \% = \frac{ni * 100}{\Sigma ni}$$

<u>مثال تطبيقي رقم 2-1:</u> لدينا بيانات عدد الغرف في عينة تتكون من 80 أسرة:

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

<u>المطلوب:</u>

1. انشأ جدول تكراري ثم أحسب التكرار النسبي.
2. اشرح معنى n_4 و f_3

<u>الحل:</u> قبل انشاء الجدول التكراري، يجب ترتيب البيانات تصاعديا ثم نحسب تكراراتها كالتالي:

/111111111/2222222222222222222222/33333333333333333333333333/ /444444444444/

5555555555

عدد الغرف Xi	عدد الأسر ni	التكرارات النسبية FI	التكرارات النسبية المئوية FI %
1	9	(9/80)=0.11	11%
2	23	0,29	29%
3	26	0,33	33%
4	12	0,15	15%
5	10	0,13	12%
المجموع	80	1,00	100%

n_4= 12 : تعني أن 12أسرة (عدد التكرارات للفئة الرابعة) لديها 4غرف.

F_3 =0.33 : تعني أن 32.5% من الأسر (التكرار النسبي المئوي للفئة الثالثة) لديها 3غرف.Σ ni=100

3.1. التوزيعات التكرارية المتجمعة

تُستخدم التوزيعات المتجمعة (الصاعدة والنازلة) للإجابة على تساؤلات حول عدد المفردات التي تقل عن قيمة معينة أو تزيد عنها (طبيّه، 2008، ص. 82).

أ. **التوزيع التكراري المتجمع الصاعد ↑Ni (على الأكثر أي أقل أو يساوي):** يوضح عدد المفردات التي قيمتها أقل من أو تساوي الحد الأعلى للفئة، ويُحسب بتجميع التكرارات المطلقة تصاعديًا.

التكرار المتجمع الصاعد	القيمة المقابلة	حساب التكرار المتجمع الصاعد	
التكرار الصاعد الأول N_1↑	هو نفسه تكرارها المطلق الأول	أي: $N_1 \uparrow = n_1$	
التكرار الصاعد الثاني $N2$↑	أقل أو يساوي X_2	أي: $N_2 \uparrow = n_1 + n_2$	$N_2 \uparrow = N_1 \uparrow + n_2$
التكرار الصاعد الثالث N_3↑	أقل أو يساوي X_3	أي: $N_3 \uparrow = n_1 + n_2 + n_3$	$N_3 \uparrow = N_2 \uparrow + n_3$
....	....	....	....
التكرار الصاعد الأخير	أقل أو يساوي X_k	أي: $N_k \uparrow = n_1 + n_2 + n_3 + .. n_k$	$N_k \uparrow = n_1 + n_2 + ... + n_k$

يمكننا الحصول على التكرار الصاعد النسبي ↗ fi ، إذا استخدمنا التكرارات النسبية fi بدلاً من المطلقة ni. ويمكن حساب التكرار النسبي الصاعد المئوي بضرب التكرار النسبي لكل مشاهده في مائة. لحساب التكرار المتجمع الصاعد و التكرار النسبي الصاعد و التكرار النسبي الصاعد المئوي، نأخذ بيانات المثال السابق كالتالي:

	عدد	التكرارات	التكرارات	التكرارات	-1 n_3 ↑ صاعد: هناك

عدد الأطفال Xi	الأسر ni	الصاعدة Ni↑	النسبية الصاعدة Fi↑	النسبية الصاعدة المئوية Fi↑%	58 أسرة تملك 3 غرف على الأكثر(أقل أو يساوي 3 غرف)
أقل أو يساوي 1	9	**9**	0.11	11%	2- يوضح f4↑ صاعد أن 88% من الأسر تملك 4 غرف على الأكثر(أقل أو يساوي 4غرف).
أقل أو يساوي 2	23	**32**=23+9	0.40	40%	
أقل أو يساوي 3	26	**58**=26+32	0.73	73%	
أقل أو يساوي 4	12	**70**=12+58	0.88	88%	
أقل أو يساوي 5	10	**80**=10+70	1	100%	
المجموع	**80**	-		-	

ب. **التوزيع التكراري المتجمع الهابط Ni↓ (على الأقل أي أكثر أو يساوي):** يُظهر عدد المفردات التي قيمتها أكثر من أو تساوي الحد الأدنى للفئة، ويُحسب عن طريق الطرح من المجموع الكلي أو من التكرار المتجمع الهابط السابق.

التكرار المتجمع النازل	**القيمة المقابلة**	**حساب التكرار المتجمع النازل**	
التكرار النازل الأول $N_1\downarrow$	هو مجموع التكرارات	أي: $N_1\downarrow = \Sigma ni$	
التكرار النازل الثاني $N2\downarrow$	أكثر أو يساوي X_2	أي: $N_2\downarrow = \Sigma n - n_1$	⬅ $N_2\downarrow = N_1\downarrow - n_1$
التكرار النازل الثالث $N_3\downarrow$	أكثر أو يساوي X_3	أي : $N_3\downarrow = n - n_1 - n_2$	⬅ $N_3\downarrow = N_2\downarrow - n_2$
....	....	....	....
التكرار النازل الأخير	أكثر أو يساوي X_k	أي: $N_k\downarrow = n_1 - n_2 - n_3 + \ldots n_{i-1}$	⬅ $N_k\downarrow = n_1 - n_2 - \ldots - n_{i-1}$

ملاحظة: يمكننا أيضًا حساب **التكرار النسبي المتجمع الهابط** (Fi↓) والتكرار **المئوي المتجمع الهابط (Fi↓%)** بنفس الطريقة التي حسبنا بها التكرارات المتجمعة الصاعدة النسبية والمئوية كالتالي :

عدد الأطفال Xi	عدد الأسر ni	التكرارات النازلة Ni↓	التكرارات النسبية النازلة Fi↓	التكرارات النسبية النازلة المئوية Fi↓%	N4↓= 22 :يمثل أن 22أسرة تملك 4غرف على الأقل (أي أكثر من أو يساوي 4 غرف) .
أكثر أو يساوي 1	9	**80**	1	%100	يوضح أن 60% :60% = F3↓% من الأسر تملك 3غرف على الأقل (أي أكثر من أو يساوي 3 غرف).
أكثر أو يساوي 2	23	**71**=9-80	.087	%89	
أكثر أو يساوي 3	26	**48**=23-71	0.60	%60	
أكثر أو يساوي 4	12	**22**=26-48	0.30	%28	
أكثر أو يساوي 5	10	**10**=12-22	0.14	%13	
المجموع	**80**	-	-	-	

2 العرض الجدولي للمتغيرات الكمية المستمرة

تُعد المتغيرات الكمية المستمرة، مثل الأوزان أو الأطوال، الأكثر شيوعًا في الدراسات الإحصائية .نظرًا لأن هذه المتغيرات يمكن أن تأخذ عددًا لا نهائيًا من القيم ضمن مجال معين، فمن غير العملي إدراج كل قيمة فردية في جدول، على عكس المتغيرات الكمية المنفصلة (التي تأخذ قيمًا محددة .(لذلك، نلجأ إلى تجميع هذه البيانات في فئات (Classes)أو مجالات، وعادة ما تكون على الشكل [أ، ب [. و من أجل انشاء جدول التوزيع التكراري للمتغيرات الكمية المتصلة ، فإننا نتبع الخطوات التالية (طيار، 2019، ص. 106):

أ. **<u>الفئات class (C)</u>**: الفئات هي المجموعات التي يتألف منها التوزيع التكراري. تُحدد كل فئة بحدين:

- **الحد الأدنى للفئة :(Lower Class Limit)** ويرمز له بالرمز .Xmin
- **الحد الأعلى للفئة :(Upper Class Limit)** ويرمز له بالرمز .Xmax

مثال :في الفئة [1، 5 [الحد الأدنى هو 1 والحد الأعلى هو 5.

من المهم فهم طبيعة حدود الفئات:

- **حد فعلي :(Inclusive)** عندما تكون الفئة من[أ إلى ب]، فإن القيمة ب تنتمي إلى الفئة. هذا يعني أن الفئة تشمل حدها الأعلى.
- **حد غير فعلي :(Exclusive)** عندما تكون الفئة من [أ إلى أقل من ب[، فإن القيمة ب لا تنتمي إلى الفئة. هذا يعني أن الفئة لا تشمل حدها الأعلى، والقيم التي تساوي ب تندرج في الفئة التالية. هذا هو الأسلوب الأكثر شيوعًا للمتغيرات المستمرة لتجنب تداخل القيم بين الفئات.

ب. **<u>المدى الكلي (Total Range):</u>** يُعرف المدى الكلي بأنه الفرق بين أكبر قيمة وأصغر قيمة في المجموعة الإحصائية. يرمز له بالرمز .E

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

ت. **<u>عدد الفئات(K):</u>** يُفضل أن يتراوح عدد الفئات بين 5 و15 فئة، اعتمادًا على حجم مجموعة البيانات، حيث يجب تجنب الفئات التي تحتوي على تكرارات قليلة جدًا، إن أمكن، لضمان تمثيل جيد للبيانات . يمكن حساب عدد الفئات K باستخدام إحدى القاعدتين الشائعتين (طبيّه، 2008، ص. 139):

- **قاعدة ستورج(Sturge's Rule) :**

$$K=1+3.3\log_{10}(n)$$

أو

- قاعدة يول(Yule's Rule) :

$$k=2.5\ (n)^{1/4}$$

حيث : K : عدد الفئات. / N : العدد الكلي للقيم (حجم العينة).

ث. **تحديد طول الفئة(L):** يُحدد طول الفئة بقسمة المدى الكلي (E) على عدد الفئات (K) يجب تقريب طول الفئة إلى أقرب عدد صحيح يناسب طبيعة البيانات (عادةً للأعلى) لضمان تغطية جميع القيم.

$$\text{طول الفئة } L = \frac{\text{المدى الكلي } \mathbf{E}}{\text{عددالفئات}(\mathbf{K})}$$

ج. **مركز الفئة Ci:** يعتبر مركز الفئة القيمة الواقعة في منتصف الفئة، و يرمز لها بالرمز Ci. و تحسب كالتالي :

$$\text{مركزالفئة } \mathbf{Ci} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{\mathbf{2}}$$

ح. **تكرار الفئة (Classes frequency) :** هو عدد القيم التي تقع في مدى تلك الفئة، ويرمز له بالرمز(ni)

- n1 : يمثل تكرار الفئة الأولى.
- n2 : يمثل تكرار الفئة الثانية، وهكذا.

يجب أن يكون مجموع التكرارات (Σni) دائمًا مساويًا للعدد الكلي لقيم المفردات (N) في مجموعة البيانات.

بالإضافة إلى هذه الأساسيات، يمكن حساب التكرارات الأخرى المشتقة بنفس طريقة المتغيرات الكمية المنفصلة:

- **التكرار المتجمع الصاعد (Ni↑)** : يمثل مجموع الأفراد **الذين تقل قيمتهم الإحصائية عن أو تساوي الحد الأعلى للفئة المقابلة.**
- **التكرار المتجمع النازل (Ni↓)** :يمثل مجموع الأفراد **الذين تساوي أو تزيد قيمتهم الإحصائية عن الحد الأدنى للفئة المقابلة.**
- **التكرار النسبي (fi)** : وهو تكرار الفئة مقسومًا على المجموع الكلي للتكرارات.
- **التكرار المئوي (fi%)** : وهو التكرار النسبي مضروبًا في 100.

- **التكرار النسبي المتجمع الصاعد** :($Fi\uparrow$) التكرار المتجمع الصاعد مقسومًا على المجموع الكلي.
- **التكرار النسبي المتجمع النازل ($Fi\downarrow$)** :التكرار المتجمع النازل مقسومًا على المجموع الكلي.

__مثال تطبيقي رقم 2-2:__ تمثل البيانات التالية أوزان الملاكمين المشاركين في البطولة الوطنية للملاكمة مرتبة تصاعديا من الأقل وزنا الى الأكثر وزنا:

62.8	62.7	61.5	60.5	59.7	59	58	58	56.8	56.2	55.8	55	54	53	52.3	52
72.9	72.5	71.5	70.4	70.2	70	69	68.2	66.8	66.7	66.3	65.3	65	64	64.2	64
83.1	82.5	82.1	81.2	80.5	80	80	79.5	78.6	78.1	77.4	77	76	76	74.3	73.8
93.5	93.1	92	91.5	91.2	90	89	88.1	87.8	87.2	86.4	85.9	85	84	83.9	83.5

__المطلوب:__

1. حدد المجتمع الاحصائي، الصفة، الكيفيات ثم حولها الى متغير نوعي.
2. احسب عدد الفئات، طول الفئة، مركز الفئة،
3. احسب التكرار المتجمع الصاعد التكرار المتجمع النازل، التكرار النسبي، التكرار النسبي الصاعد و النازل.
4. اشرح : **n_3، f_5، $N_3\downarrow$، $N_6\uparrow$، $F_4\uparrow$ ، $F_3\downarrow$**

__الحل:__

1. **تحديد المجتمع الإحصائي والصفة والكيفيات:**

أ. **المجتمع الإحصائي** :جميع الملاكمين المشاركين في البطولة الوطنية للملاكمة.

ب. **الصفة المدروسة** :أوزان الملاكمين.

ت. **الكيفيات (القيم المحتملة)** :الأوزان الموزعة في مجالات (فئات) تتراوح بين أصغر وأكبر وزن.

ث. **التحويل إلى متغير نوعي** :يمكن تصنيف هذه الأوزان ضمن فئات وزن الملاكمة التقليدية (مثل وزن الديك، الريشة، الخفيف، إلخ.)، مما يحولها إلى متغير نوعي ترتيبي.

2. **حساب عدد الفئات، طول الفئة، ومركز الفئة:**

العدد الكلي للملاكمين (n) : 64 / **أصغر قيمة (Xmin)**: 52 / **أكبر قيمة(Xmax)**: 93.5

أ. **حساب عدد الفئات (K) باستخدام قاعدة ستورج** :

$K=1+3.3\log_{10}(64) = K=1+3.3\times1.806 = K=1+5.9598\approx7.01$: نُقرّب K إلى **7 فئات**.

ب. **حساب المدى الكلي:(E)**

E=X max−X min=93.5−52=41.5

ت. حساب طول الفئة(L) :

: L= E / K =41.5/7≈5.92   نُقرّب L إلى 6 لضمان تغطية جميع الأوزان بشكل كامل.

3. بناء الجدول التكراري:

أوزان الملاكمين Xi	المتغير النوعي	مركز الفئة	التكرارات ni	التكرارات النسبية Fi	التكرارات النسبية المئوية Fi %	التكرارات النازلة Ni↓	التكرارات الصاعدة Ni↑	التكرارات النسبية الصاعدة Fi↑	التكرارات النسبية النازلة Fi↓
]58-52]	الديك (52 كغ)	55	8	0,125	%12,50	64	8	0,1	1,0
]64-58]	الريشة (58 كغ)	61	8	0,125	%12,50	56	16	0,3	0,9
]70-64]	الخفيف (64 كغ)	67	10	0,16	%15,63	**48**	26	0,4	0.75
]76-70]	الوسط (70 كغ)	73	8	0,13	%12,50	38	34	**0,5**	0,6
]82-76]	المتوسط (76 كغ)	79	11	0,17	%17,19	30	45	0,7	0,5
]88-82]	خفيف الثقيل (82 كغ)	85	11	0,17	%17,19	19	**56**	0,9	0,3
]94-88]	الثقيل (91 كغ)	91	8	0,13	%12,50	8	64	1,0	0,1
المجموع	-		64	1	%100	-	-	-	-

4. الشرح:

- **10**= N_3: هناك 10 ملاكمين أوزانهم تتراوح بين 58و 64 كلغ.
- **F_5=0.17** : %17 من الملاكمين أوزانهم تتراوح بين 76و 82 كلغ.
- **N_3↓= 48**: أوزان 48 ملاكما من بين 64، أكبر أو يساوي 64 كلغ.
- N_6↑= 56: أوزان 56 ملاكما من بين 64، أقل من 88 كلغ.
- **F_4↑= 0.5**: أوزان نصف الملاكمين أقل من 76 كلغ.
- F_3↓= 0.75 : أوزان %80 من الملاكمين أكبر أو يساوي 64 كلغ.

3   العرض الجدولي للمتغيرات النوعية.

تُعبّر المتغيرات النوعية عن صفات أو أنواع لا تأخذ قيمًا عددية (مثل الجنس، الحالة الاجتماعية). يتم إنشاء جدول تكراري لها برصد كل فئة نوعية في عمود، مع تسجيل عدد المفردات التابعة لها في **عمود التكرار المطلق** (ni)(عبد الفتاح، 2012، ص. 35).

$$N = \sum_{i=1} \text{ni} = \text{n1} + n2 + \cdots nk$$

المحور الثاني: العرض الجدولي للبيانات الاحصائية

تنقسم المتغيرات النوعية إلى:

- **متغيرات نوعية ترتيبية** :يمكن ترتيب فئاتها بشكل منطقي (مثال: مستويات التعليم).
- **متغيرات نوعية اسمية** :لا يمكن ترتيب فئاتها بأي شكل (مثال: الجنسيات، الألوان).

المتغير xi	متغير نوعي ترتيبي (نتائج الطلبة)	متغير نوعي اسمي (أعمال حرة)	التكرار المطلق
الصفة 1: X_1	ضعيف	طبيب	n_1
الصفة 2: X_2	متوسط	محامي	n_2
....	حسن	ميكانيكي	....
الصفة K: X_K	جيد		n_k
المجموع Σ	-	-	N

يمكن حساب التكرارات المتجمعة (الصاعدة والنازلة)، والتكرار النسبي، والتكرار المئوي، والتكرار النسبي الصاعد والنازل للمتغيرات النوعية بنفس الطريقة التي حسبنا بها للمتغيرات الكمية، خاصة إذا كانت ترتيبية.

الفصل الثالث: العرض البياني للبيانات الاحصائية

تعتبر الرسوم البيانية أداة قوية وفعالة لوصف البيانات وتحليلها، حيث توفر نظرة سريعة وواضحة على شكل التوزيع ومدى تمركز البيانات. في العديد من التطبيقات، يكون العرض البياني أسهل وأسرع في فهم الظواهر قيد الدراسة مقارنة بالجداول الرقمية. تتنوع طرق عرض البيانات بيانيًا حسب نوع البيانات (سواء كانت نوعية أو كمية، متصلة أو منفصلة) (عزوز، 2010، ص. 45).

1 العرض البياني للمتغيرات النوعية.

يمكن عرض البيانات الخاصة بمتغير وصفي سواء القابل للترتيب أو الاسمي، في شكل دائرة بيانية أو أعمدة مستطيلة بيانية أو أعمدة مجزأة ، يمكن من خلاله وصف ومقارنة مجموعات أو مستويات هذا المتغير.

1.1 الدائرة البيانية (Diagramme Circulaire):

تُستخدم الدائرة البيانية لعرض بيانات المتغير الوصفي عن طريق تقسيم الدائرة (360 درجة) إلى قطاعات، يمثل كل قطاع تكرارًا نسبيًا لمجموعة معينة من المتغير. يُحدد مقدار الزاوية لكل قطاع باستخدام المعادلة التالية (السيد، 2020، ص. 120):

مقدار الزاوية الخاصة بتلك المجموعة = تكرار الفئة (ni) / مجموع التكرارات (N) * 360°

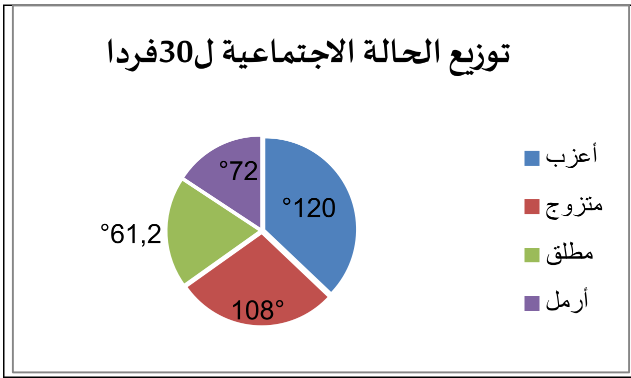
أو ببساطة:

مقدار الزاوية الخاصة بتلك المجموعة = التكرار النسبي للمجموعة (Fi) × 360°

<u>مثال 3-1:</u> لنفترض أن لدينا عينة مكونة من 30 فردًا من سكان إحدى البلدات موزعين حسب الحالة الاجتماعية. لرسم الدائرة البيانية، نقوم بإنشاء جدول نحسب فيه التكرار النسبي لكل مجموعة، ثم نضرب هذا التكرار في 360 درجة للحصول على مقدار الزاوية داخل الدائرة، كما هو موضح في الجدول التالي:

الشكل رقم 3-1: الرسم البياني لتوزيع الحالة الاجتماعية ل 30فردا باستعمال الدائرة النسبية.

الحالة الاجتماعية Xi	عدد الأفراد ni	التكرارات النسبية Fi	مقدار الزاوية
أعزب	10	0,33	°120=°360*0.33
متزوج	09	0,25	°108 =°360*0.25
مطلق	05	0,17	°61.2=°360*017
أرمل	06	0.20	°72=°360*0.20



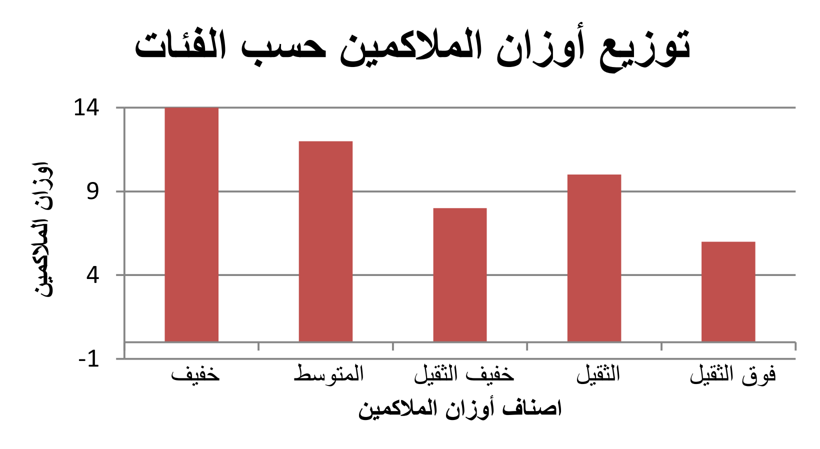
المجموع	30	1	360°	

2.1. الأعمدة المستطيلة Diagramme en Colonnes: عبارة عن مستطيلات متباعدة بمسافات ثابتة ولها قواعد متساوية، تتناسب أطوالها مع التكرارات المقابلة لمكونات الخاصية المدروسة (عبد الفتاح، 2012، ص. 75).

ولرسم التمثيل البياني الأعمدة المستطيلة نأخذ معطيات المثال التالي(مثال 3-2): لدينا أوزان 50 ملاكما موزعين حسب الفئات التالية:

الشكل رقم3-2: الرسم البياني لتوزيع أوزان الملاكمين حسب الفئات باستعمال الأعمدة المستطيلة.

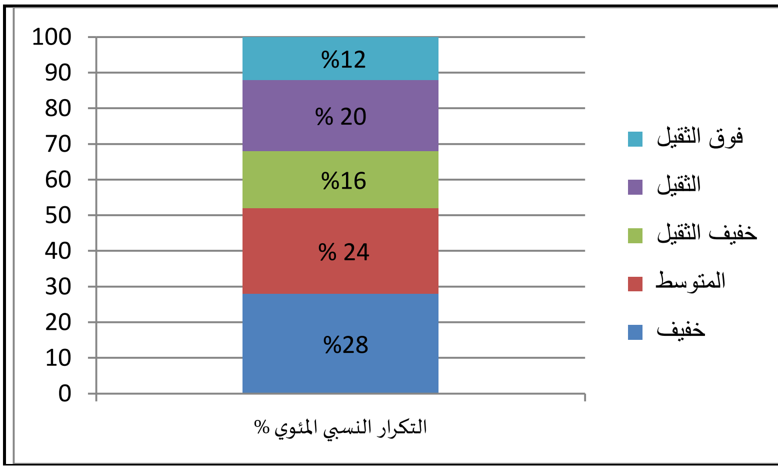
أوزان الملاكمين Xi	التكرارات ni
خفيف	14
المتوسط	12
خفيف الثقيل	8
الثقيل	10
فوق الثقيل	6
المجموع	50



1.3. الأعمدة المجزأة Diagramme en Barres: عبارة عن مستطيل واحد مقسم إلى عدة أجزاء، يمثل كل جزء تكرارًا معينًا للخاصية المدروسة. من الأفضل عند رسم العمود استخدام النسب المئوية المقابلة لكل تكرار، حيث أن طول المستطيل الكلي يمثل 100% (تيلوت،2009، ص 38). لرسم الأعمدة المجزأة، نأخذ معطيات المثال السابق (مثال 3-2)، ولكننا نحول التكرارات إلى نسب مئوية أولًا:

الشكل رقم3-3: الرسم البياني لتوزيع أوزان الملاكمين حسب الفئات باستعمال الأعمدة المجزأة.

أوزان الملاكمين Xi	التكرارات النسبية المئوية Fi %
خفيف	**%28**
المتوسط	**%24**
خفيف الثقيل	**%16**
الثقيل	**%20**
فوق الثقيل	**%12**
المجموع	**%100**

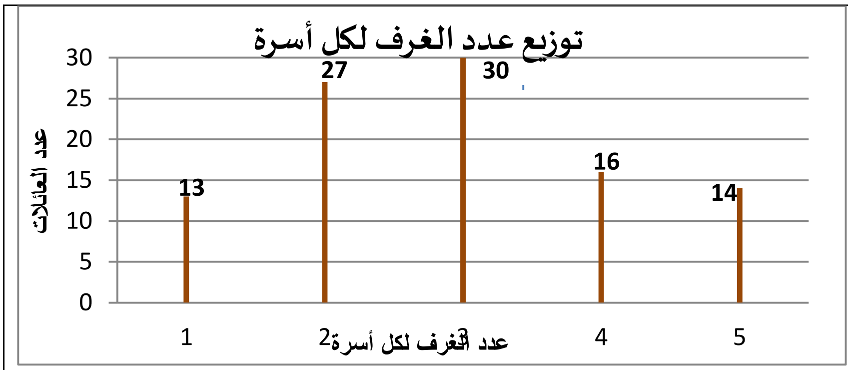


2 العرض البياني للمتغيرات الكمية المتقطعة

يتم رسم المتغيرات الكمية المتقطعة بيانيًا من خلال نوعين من العروض البيانية: الأعمدة البسيطة، والعرض البياني للتكرار المتجمع الصاعد والنازل.

1.2. الأعمدة البسيطة: تُستخدم الأعمدة البسيطة لتمثيل المتغيرات الكمية المتقطعة، حيث يتناسب طول كل عمود مع تكرار القيمة المقابلة. وتُوظف للمقارنة بين القيم حسب الزمن أو التصنيفات.

الشكل رقم3-4: الرسم البياني لتوزيع الأسر حسب عدد الغرف باستعمال الأعمدة البسيطة



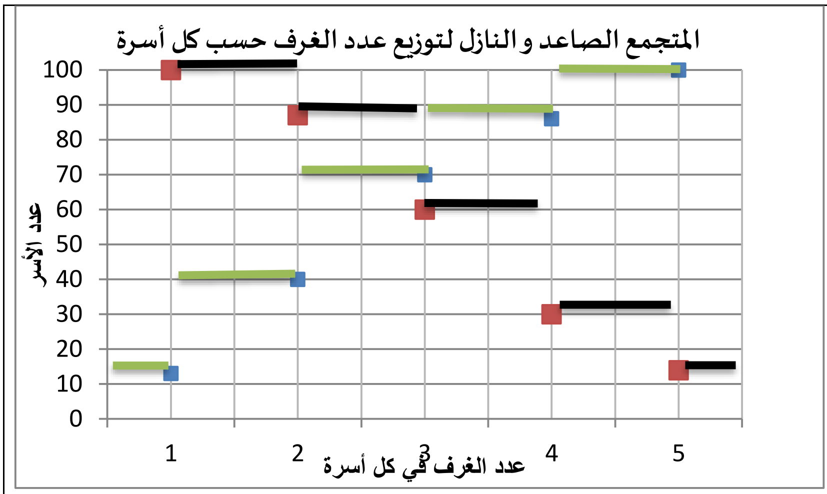
عدد الغرف لكل عائلة Xi	التكرارات ni
1	13
2	27
3	30
4	16
5	14
المجموع	**100**

2.2. العرض البياني للتكرار المتجمع الصاعد والنازل: هذا النوع من الرسوم البيانية يوضح التكرارات المتجمعة، إما تصاعديًا أو تنازليًا(تيلوت،2009، ص 40).

أ. **التكرارات المتجمعة الصاعدة Effectifs cumulés croissants**(أقل أو يساوي): هو عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات التجميعية الصاعدة المقابلة لكل قيمة من قيم المتغير الإحصائي المدروس. إذا أخذنا المثال السابق حول عدد الغرف لكل أسرة، فإنه، على سبيل المثال، لرسم القطعة المستقيمة المقابلة للقيمة 2 والتي يساوي تكرارها الصاعد 40، نضع قطعة مستقيمة عند إحداثيات النقطة (2، 40) تبدأ من مستوى القيمة 2 لتنتهي عند مستوى القيمة 1.

ب. **التكرارات المتجمعة النازلة Effectifs cumulés décroissants(أكثر أو يساوي): هو عبارة عن قطع** مستقيمة متنازلة حسب تنازل التكرارات التجميعية النازلة. إذا أخذنا المثال السابق حول عدد الغرف لكل أسرة، فإنه، على سبيل المثال، لرسم القطعة المستقيمة المقابلة للقيمة 3 والتي يساوي تكرارها النازل 60، نضع قطعة مستقيمة عند إحداثيات النقطة (3، 60) تبدأ من مستوى القيمة 3 لتنتهي عند مستوى القيمة 4. لإنشاء الرسم البياني للتكرار المتجمع الصاعد والنازل، نحتاج إلى حساب التكرارات المتجمعة الصاعدة والنازلة أولًا:

الشكل رقم3-5: منحنى التكرار المتجمع الصاعد و النازل لتوزيع الأسر حسب عدد الغرف.



عدد الغرف لكل عائلة Xi	التكرارات الصاعدة Ni↑	التكرارات النازلة Ni↓
1	13	100
2	40	87
3	70	60
4	86	30
5	100	14
المجموع	-	-

3 العرض البياني للمتغيرات الكمية المستمرة

تُعد العروض البيانية للمتغيرات الإحصائية المستمرة من أكثر الطرق البيانية استخدامًا وفعالية في تمثيل البيانات. من أهم هذه العروض: **المدرج التكراري، المضلع التكراري، والمنحنى التكراري، بالإضافة إلى منحنى التكرارات المتجمعة الصاعدة والنازلة** (السيد، 2020، ص. 117):

1.3. المدرج و المضلع التكراري L'Histogramme et le Polygone :

المدرج التكراري هو عبارة عن مستطيلات متلاصقة، حيث تمثل التكرارات على المحور الرأسي، بينما تمثل قيم المتغير (حدود الفئات) على المحور الأفقي. كل مستطيل منها يمثل فئة من فئات التوزيع التكراري، ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة. **أما المضلع التكراري** فهو خط منكسر يصل بين نقاط تمثل كل نقطة منها فئة من فئات التوزيع، حيث تكون إحداثي أي نقطة (x ، y) كما يلي

X : مركز الفئة التي تمثلها النقطة.

Y : تكرار هذه الفئة.

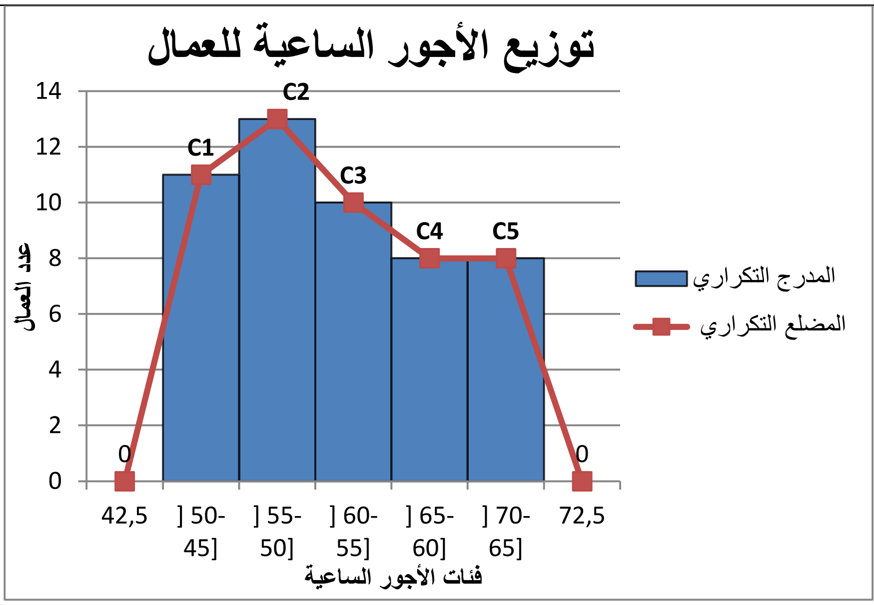
يجب أن يُغلق المضلع التكراري من طرفيه على المحور الأفقي. يتم ذلك بإضافة فئة افتراضية قبل الفئة الأولى وفئة أخرى بعد الفئة الأخيرة. يكون تكرار كلتا الفئتين الافتراضيتين صفرًا، وبالتالي تقع النقطتان الخاصتان بهاتين الفئتين على المحور الأفقي، مما يؤدي إلى غلق المضلع التكراري. من المهم قبل رسم المدرج والمضلع التكراري ملاحظة ما إذا كانت أطوال الفئات متساوية أم لا، حيث نميز حالتين عند رسم المدرج التكراري:

أ. **في حالة المدرج التكراري في حالة فئات متساوية الطول:** عندما تكون الفئات متساوية الطول، تكون قاعدة المقارنة ثابتة ومتساوية. في هذه الحالة، يتم رسم المدرج التكراري مباشرة. لتوضيح ذلك، **نأخذ المثال التالي**

(المثال 3-4): الجدول التالي يمثل الأجور الساعية لعينة من 50 عاملًا في مصنع البسكويت (الوحدة: 10 دج).

الشكل رقم 3-6: الرسم البياني لتوزيع الأجور الساعية لعينة من العمال باستعمال المدرج و المضلع التكراري.

أجور العمال الساعية Xi	مركز الفئة Ci	عدد العمال في كل فئة التكرارات ni
]45 -40]	42.5	0
]50 -45]	47,5	11
]55 -50]	52,5	13
]60 -55]	57,5	10
]65 -60]	62,5	08
]70 -65]	67,5	08
]75 -70]	72.5	0
المجموع	-	50



ب. **المدرج التكراري في حالة فئات غير متساوية الطول** :إذا كانت فئات التوزيع غير متساوية الطول نقوم بتعديل التكرارات لأن قاعدة المقارنة غير ثابتة، وذلك حتى يكون هناك تناسب بين طول الفئة والتكرار المقابل لها.

ولغرض تعديل التكرارات نستخدم المعادلة الآتية:

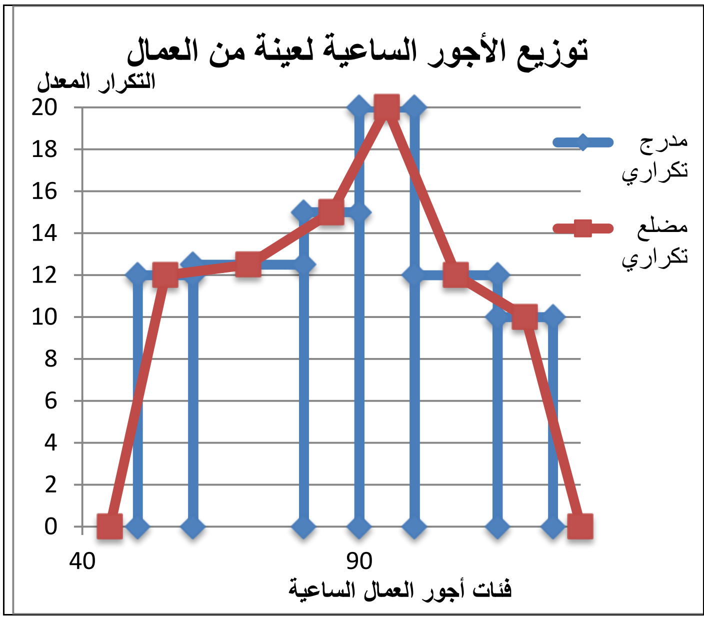
$$ni *= \frac{\text{تكرارالفئة } ni \;\; X \;\; \text{طول المختارالفئة } L *}{\text{طول الفئة } L}$$

ملاحظة : طول الفئة المختار (L*) هو القاسم المشترك الأكبر لأطوال الفئات أو التكرار الشائع.

ولتوضيح ذلك نأخذ المثال التالي**(المثال 3 -5)**: الجدول التالي يمثل الأجور الساعية لعينة من 100 عاملا في مصنع الأجور (الوحدة 10دج) كالتالي:

الشكل رقم 3-7: الرسم البياني لتوزيع الأجور الساعية لعينة من العمال باستعمال المدرج و المضلع التكراري (حالة الفئات الغير منتظمة).

الأجور X الساعية	طول الفئة	مركز الفئة	التكرارا ت ni	التكرار المعدل
] 50- 40]	10	45	0	0
] 60- 50]	10	55	12	12
] 80- 60]	20	70	25	12,5
] 90- 80]	10	85	15	15
] 100- 90]	10	95	20	20
] 115- 100]	15	107,5	18	12
] 125- 115]	10	120	10	10
] 135- 125]	10	130	0	0
المجموع			100	



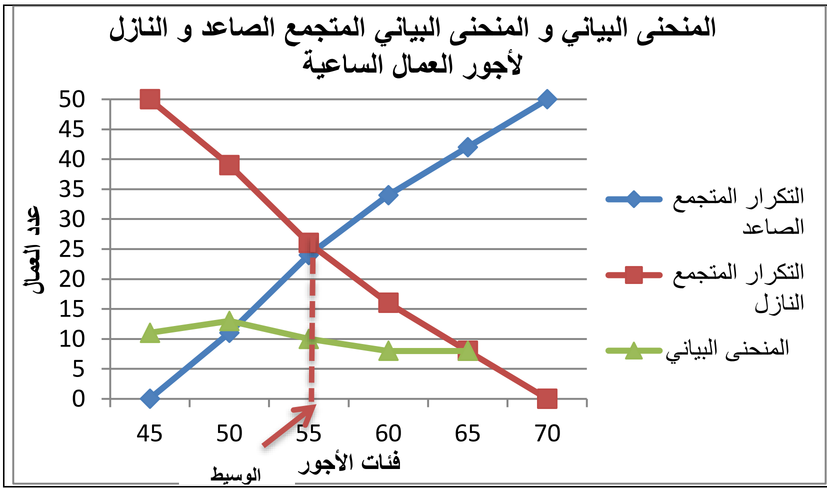
2.3. المنحنى التكراري و منحنى التكرارات المتجمعة الصاعدة والنازلة:

أ. **المنحنى التكراري Courbe de fréquence:** يرسم المنحنى التكراري باتّباع الخطوات نفسها التي أجريناها في حالة المضلع التكراري والفرق بين الحالتين هو في طريقة توصيل النقط التي تمثل الفئات ، ففي حالة المنحنى التكراري نصل بين النقط بمنحنى ممهد مستمر بحيث يمر بأكبر عدد ممكن من النقط ويمر من خلال الباقي بتوازن. كذلك في حالة المنحنى التكراري لا نستعين بالفئتين :السابقة للأولى واللاحقة للأخيرة.

ب. **منحنى التكرارات المتجمعة الصاعدة والنازلة:** يرسم منحنى التكرار المتجمع الصاعد بربط النقاط التي تمثل الحدود العليا للفئات مع تكراراتها المتجمعة، بينما يُرسم المنحنى النازل بربط الحدود الدنيا للفئات مع التكرارات النازلة. يمثل تقاطع المنحنيين قيمة الوسيط الإحصائي. ويُستخدم هذا الأسلوب لتوضيح توزيع البيانات كما في مثال الأجور الساعية للعمال.

الشكل رقم 3-7: المنحنى البياني و المنحنى البياني المتجمع الصاعد و النازل لتوزيع الأجور الساعية لعينة من العمال.

الأجور X الساعية	مركز الفئة Ci	التكرارات ni	التكرار المتجمع الصاعد	التكرار المتجمع النازل
] 50- 45]	47,5	11	11	50
] 55- 50]	52,5	13	24	39
] 60- 55]	57,5	10	34	26
] 65- 60]	62,5	8	42	16
] 70- 65]	67,5	8	50	8
المجموع		50	-	-



الفصل الرابع: مقاييس النزعة المركزية

تُعدّ مقاييس النزعة المركزية من الركائز الأساسية في التحليل الإحصائي، إذ تختزل البيانات في قيمة واحدة تمثل مركز التوزيع وتعكس ميله العام. وتمكّن هذه الخاصية الباحث من تبسيط كميات كبيرة من المعطيات، وتحديد موقع تركزها، مما يوفّر أداة دقيقة للمقارنة بين المجموعات المختلفة، ويساعد في صياغة قرارات مبنية على معطيات كمية موثوقة. ومن ثمّ فهي لا تقتصر على الوصف، بل تؤدي دورًا تحليليًا في الكشف عن طبيعة البيانات واتجاهها العام. وتتجسد هذه المقاييس أساسًا في المتوسط الحسابي الذي يناسب البيانات المتجانسة لكنه يتأثر بالقيم المتطرفة، والوسيط الذي يُستخدم عند وجود انحرافات أو قيم شاذة، والمنوال الذي يُبرز أكثر القيم تكرارًا. ويتيح توظيف هذه المقاييس معًا قراءة أكثر شمولية لشكل التوزيع ودرجة انحرافه، مع ضرورة مراعاة القيود الخاصة بكل مقياس واختيار الأنسب بحسب طبيعة البيانات وأهداف التحليل، الأمر الذي يمنح النتائج صلابة علمية ودقة استنتاجية.

1 المتوسط الحسابي(Mean):

1.1. مفهوم المتوسط الحسابي (Mean) : المتوسط الحسابي، والمعروف أيضًا بالوسط الحسابي أو المعدل، هو **أحد أهم مقاييس النزعة المركزية وأكثرها شيوعًا** .يمثل هذا المقياس **القيمة النموذجية أو المركزية لمجموعة من البيانات**، حيث يعمل كنقطة توازن تتجمع حولها معظم القيم. يمكن تخيله كـ "نقطة ارتكاز" للميزان الذي تُمثل فيه كل قيمة من البيانات ثقلاً. يستخدم المتوسط الحسابي على نطاق واسع في مجالات متنوعة، من الإحصاء الأكاديمي والبحث العلمي إلى التطبيقات اليومية في الأعمال والاقتصاد، لأنه يوفر **خلاصة سريعة ومفهومة لمجموعة كبيرة من الأرقام** .يُرمز له عادةً بالرمز ($\overline{X}$). هناك عدة طرق لحساب المتوسط الحسابي:

1.1. المتوسط الحسابي للبيانات الأولية (البيانات غير المبوبة): البيانات الأولية هي **البيانات الخام التي لم يتم تنظيمها أو تجميعها في جداول تكرارية** .لحساب المتوسط الحسابي لهذه البيانات، توجد طريقتان رئيسيتان:

أ. **الطريقة المباشرة**: هذه هي **الطريقة الأكثر بساطة ووضوحًا** لحساب المتوسط الحسابي. تعتمد على مبدأ أساسي :**جمع كل القيم الموجودة في المجموعة، ثم قسمة هذا المجموع على العدد الكلي لهذه القيم** .تخيل أن لديك درجات امتحان لعدد من الطلاب؛ لكي تعرف متوسط أداء الصف، تقوم بجمع درجات جميع الطلاب وتقسمها

$$\frac{\text{مجموع هذه القيم}}{\text{عددها}}$$

ب. على عدد الطلاب. هذه الطريقة تعطينا **قيمة دقيقة تعكس المركز العام للبيانات**، وهي مفيدة بشكل خاص عندما تكون مجموعة البيانات صغيرة أو متوسطة الحجم.

الوسط الحسابي ($\overline{X}$) لمجموعة قيم =

$$\overline{x} = \frac{x1 + x2 + x3 + \cdots xk}{n} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$: أي

<u>مثال (4-1) :</u> لدينا علامات أحد الطلبة في الامتحانات الخاصة بثمانية مقاييس كالتالي: 09- 12- 11- 15 - 12- 13- 10- 14

المطلوب : احسب الوسط الحسابي لعلامات هذا الطالب.

- الحل: لإيجاد الوسط الحسابي للدرجات تطبق المعادلة السابقة كما يلي:

$$\overline{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} = \frac{09 + 12 + 11 + 15 + 12 + 13 + 10 + 14}{8} = \frac{96}{8} = 12$$

وهذا يعني أن الوسط الحسابي لعلامات هذا الطالب في الامتحانات الثمانية (n=8) هو: **12**

ت. طريقة الانحرافات (طريقة الوسط الفرضي): تُعرف هذه الطريقة أيضًا بـ "طريقة الوسط التقديري" أو "الوسط المفترض". هي **طريقة بديلة وأحيانًا أكثر كفاءة للحسابات اليدوية، خاصة مع مجموعات البيانات الكبيرة أو القيم ذات الأرقام الكبيرة** .تعتمد الفكرة على **اختيار "وسط فرضي" أو تقديري (X0)** من بين القيم (أو قيمة قريبة منها). ثم نحسب **انحراف كل قيمة عن هذا الوسط الفرضي**، ونقوم بجمع هذه الانحرافات. إذا كان مجموع الانحرافات موجبًا، فهذا يعني أن المتوسط الحقيقي أكبر من الوسط الفرضي، والعكس صحيح إذا كان سالبًا. تُضاف بعد ذلك متوسط هذه الانحرافات إلى الوسط الفرضي للحصول على المتوسط الحسابي الحقيقي. هذه الطريقة تقلل من حجم الأرقام التي نتعامل معها في العمليات الحسابية، مما يقلل من احتمالية الأخطاء.

$$\overline{x} = Xo + \frac{\Sigma di}{n}$$

علما أن x_0 :الوسط الفرضي، Σdi مجموع الانحرافات، حيث أن $di=(x_i-x_0)$ ، nعدد البيانات.

باستعمال هذه الطريقة للمثال السابق نحسب الوسط الحسابي كالتالي:

- نفرض أن الوسط الفرضي $x_0 = 10$
- مجموع الانحرافات Σdi = ((10-09) +(10-12)+ (10-11)+(10-15)+(10-12)+(10-13)+(10-10)+(10-14)) = 16

$$\bar{x} = Xo + \frac{\Sigma di}{n} = 10 + \frac{16}{8} = 12$$

ومنه قيمة الوسط الحسابي هي: $\bar{x}$**=12**

كما تلاحظ، **النتيجة هي نفسها** (12)، مما يؤكد صحة كلا الطريقتين. هذه الطريقة قد تكون مفيدة بشكل خاص في الحسابات اليدوية المعقدة.

2-1 المتوسط الحسابي للبيانات المبوبة: تُستخدم البيانات المبوبة عندما تكون لدينا **كمية كبيرة من البيانات التي تم تنظيمها وتجميعها في جداول تكرارية** (مثل الفئات والتكرارات). في هذه الحالة، يُطلق على المتوسط الحسابي غالبًا اسم **"المتوسط المرجح"**، لأن كل قيمة (أو مركز فئة) يتم "ترجيحها" بتكرارها. هناك حالتان رئيسيتان:

1-2-1 في حالة متغير متقطع(منفصل): المتغير المتقطع هو المتغير الذي يأخذ قيمًا محددة ومنفصلة)مثل عدد الأطفال في الأسرة، أو عدد الغرف). لا يمكن أن يأخذ قيمًا كسرية بين هذه الأرقام الصحيحة. عند تبويب هذه البيانات، تُظهر الجداول كل قيمة فريدة مع عدد مرات تكرارها.

أ. **الطريقة المباشرة:** في هذه الحالة، لا نتعامل مع قيم فردية فقط، بل مع **قيم وتكراراتها** .لذلك، يجب أن نأخذ في الاعتبار عدد مرات ظهور كل قيمة. يتم ذلك عن طريق **ضرب كل قيمة (Xi) بتكرارها المقابل(ni)** ، **ثم جمع كل هذه النواتج** .بعد ذلك، نقسم هذا المجموع على **العدد الكلي للتكرارات** (وهو في الأساس إجمالي عدد الملاحظات). هذه الطريقة تعطينا متوسطًا يعكس التوزيع الفعلي للبيانات المبوبة. **(Goldfarb, B., & Pardoux, C. . p. 35-37).**

$$\bar{x} = \frac{x1n1 + x2n2 + x3n3 + \cdots . x\,k\,n\,k}{n1 + n2 + n3 + \cdots \ldots . nk} = \frac{\sum_{i=1}^{i=k} \mathrm{ni}\, \mathrm{x_i}}{\sum \mathrm{ni}}$$

حيث:

- Xi: **القيمة الفريدة للمتغير المتقطع** (مثل عدد الغرف).

- ni: **التكرار المقابل** لتلك القيمة، أي عدد مرات ظهورها في البيانات.
- $\sum Xini$: **مجموع حواصل ضرب كل قيمة بتكرارها.**
- $\sum ni$: **العدد الكلي للتكرارات**، وهو يمثل حجم العينة أو المجتمع.

مثال (1-2) : أحسب المتوسط الحسابي لتوزيع عدد الغرف لعينة من 80 أسرة حسب الجدول التالي:

عدد الغرفXi	1	2	3	4	5	**المجموع**
عدد الأسر ni	9	19	30	12	10	80
$\sum X_i n_i$	**9**=9*1	**38**=19*2	**90**=30*3	**48**=12*4	**50**=10*5	235
$\sum d_i$	1-3=**-2**	2-3=**-1**	3-3=**0**	4-3=**1**	5-3=**2**	0
$\sum d_i n_i$	-18	-19	0	12	20	-5

$$\bar{x} = \frac{\sum_{i=1}^{i=k} ni\ x_i}{\sum\ ni} = \frac{235}{80} = 2.9375 \approx 3$$

وهذا يعني أنه في هذه العينة المكونة من **80 عائلة، فان كل عائلة تملك 3 غرف في المتوسط.**

ب. طريقة الانحرافات(طريقة الوسط الفرضي):هذه الطريقة هي **امتداد لطريقة الانحرافات للبيانات غير المبوبة، ولكنها تطبق على البيانات المبوبة ذات المتغيرات المتقطعة** .الفكرة الأساسية هي نفسها: اختيار وسط فرضي (X0) ثم حساب الانحرافات، ولكن هنا، يتم **ترجيح كل انحراف بتكراره المقابل** .(ni)وهذا يجعل الحسابات أكثر دقة لأنها تأخذ في الاعتبار أهمية كل قيمة (بناءً على عدد مرات تكرارها) في مجموع الانحرافات .

$$\bar{x} = Xo + \frac{\sum_{i=1}^{i=k} d_i\ ni}{\sum\ ni}$$

حيث:

- X_0: **الوسط الفرضي** .يفضل اختيار القيمة ذات التكرار الأكبر، لأنها تمثل نقطة مركزية قوية في التوزيع، مما يقلل من قيم الانحرافات ويجعل الحسابات أبسط.
- $di=(X_i-X_0)$: **الانحراف** لكل قيمة عن الوسط الفرضي.
- ni : **التكرار المقابل** للقيمة.(Xi)
- $\sum d_i n_i$: **مجموع حواصل ضرب الانحرافات في تكراراتها.**
- $\sum n_i$: **العدد الكلي للتكرارات.**

باستعمال هذه الطريقة للمثال السابق نحسب الوسط الحسابي كالتالي:

- نختار قيمة الوسط الفرضي x_0 ويفضل أن تكون القيمة ذات اكبر تكرار : $\boldsymbol{x_0} = 3$
- $(x_i - x_0) = di$
- ضرب الانحرافات di بالتكرار المقابل $\boldsymbol{\Sigma\, d_i n_i = -5}$

$$\overline{\boldsymbol{x}} = \boldsymbol{Xo} + \frac{\sum_{i=1}^{i=k} d_i \,\mathrm{ni}}{\sum \ \mathrm{ni}} \ = \boldsymbol{3} + \frac{-5}{80} = \boldsymbol{2.9375 \approx 3}$$

ومنه قيمة الوسط الحسابي هي نفسها : $\boldsymbol{\bar{x} = 3}$

كما ترى، **النتيجة هي نفسها** تمامًا، مما يؤكد مرونة ودقة هذه الطريقة في التعامل مع البيانات المبوبة.

2-2-1 في حالة متغير متصل (مستمر): المتغير المتصل هو المتغير الذي **يمكن أن يأخذ أي قيمة ضمن مدى معين**)مثل الطول، الوزن، المعدلات، درجات الحرارة). في الجداول التكرارية، يتم تمثيل هذه البيانات على شكل **فئات أو فترات** (مثال: [10-12)، [12-14)). لحساب المتوسط الحسابي لهذه البيانات، لا يمكننا استخدام القيم الفردية مباشرة، بل يجب أن نستخدم **نقطة تمثل كل فئة**، وهي **مركز الفئة**.

أ. الطريقة المباشرة: في حالة المتغير المتصل، نتبع نفس منطق الطريقة المباشرة للمتغير المتقطع، ولكن مع استبدال القيم الفردية Xiبـ **مراكز الفئات .(Ci)** مركز الفئة هو **نقطة المنتصف لكل فئة**، ويُحسب بجمع الحد الأدنى والحد الأعلى للفئة وقسمة الناتج على 2. بعد حساب مراكز الفئات، نضرب كل مركز فئة بتكراره المقابل، نجمع هذه النواتج، ثم نقسم على إجمالي عدد التكرارات. هذه الطريقة تعطينا **تقديرًا جيدًا للمتوسط الحسابي** في حالة البيانات الفئوية. (راتول، محمد. 2006،. ص. 25-27).

$$\overline{\boldsymbol{x}} = \frac{\boldsymbol{c1n1 +\ c2n2 + c3n3 + \cdots . c\ k\ n\ k}}{\boldsymbol{n1 + n2 + n3 + \cdots \ldots . nk}} = \ \frac{\sum_{i=1}^{i=k} \mathrm{ni}\ c_i}{\sum \ \ \mathrm{ni}}$$

حيث:

- Ci :**مركز الفئة**، وهو نقطة المنتصف لكل فترة =Ci .(الحد الأدنى للفئة+الحد الأعلى للفئة/2)
- ni :**التكرار المقابل** لتلك الفئة.
- $\sum$Cini :**مجموع حواصل ضرب كل مركز فئة بتكراره.**

- العدد الكلي للتكرارات: $\sum ni$.

مثال (1-3) : أحسب المتوسط الحسابي لتوزيع المعدلات المتحصل عليه في شهادة الباكالوريا لعينة من طلبة المركز الجامعي –النعامة، حسب الجدول التالي:

معدلات شهادة الباكالوريا Xi	]12-10]	]14-12]	]16-14]	]18-16]	]20-18]	المجموع
مركز الفئة C_i	**11**	**13**	**15**	**17**	**19**	-
التكرارات ni	12	9	6	5	4	36
$\Sigma\, C_i n_i$	**12*11=132**	**9*13=117**	**6*15=90**	**5*17=85**	**4*19=76**	**500**
di	**11-11=0**	**13-11=2**	**15-11=4**	**17-11=6**	**19-11=8**	-
Σ dini	**0*12=0**	**2*9=18**	**4*6=24**	**6*8=30**	**8*4=32**	**104**

$$\bar{x} = \frac{\sum_{i=1}^{i=k} ni\, C_i}{\sum \; ni} = \frac{500}{36} = 13.88$$

وهذا يعني أنه في هذه العينة المكونة من 36 طالبًا، فإن **متوسط معدلات الطلبة في شهادة البكالوريا هو 13.88**. هذا يعطينا مؤشرًا عن المستوى الأكاديمي العام للطلبة في هذه العينة.

ب. طريقة الانحرافات(طريقة الوسط الفرضي): تُطبق هذه الطريقة بنفس المبدأ السابق لطريقة الانحرافات، ولكن هنا، **نستخدم مراكز الفئات (Ci) بدلاً من القيم الفردية (Xi).** يتم اختيار وسط فرضي(X0) ، غالبًا ما يكون مركز الفئة ذات أعلى تكرار، ثم تُحسب الانحرافات لكل مركز فئة عن هذا الوسط الفرضي، وتُضرب في التكرارات المقابلة قبل جمعها. (جلاطو، جيلالي. 2001. ص. 30-32).

$$\bar{x} = Xo + \frac{\sum_{i=1}^{i=k} d_i\, \mathrm{ni}}{\sum \; \mathrm{ni}}$$

حيث:

- x_0: **الوسط الفرضي**. يفضل اختيار مركز الفئة ذات أعلى تكرار لأنه غالبًا ما يكون قريبًا من المتوسط الحقيقي ويقلل من حجم الانحرافات.
- $di = (Ci - X_0)$: **انحراف مركز الفئة** عن الوسط الفرضي.
- ni: **التكرار المقابل** لتلك الفئة.

باستعمال هذه الطريقة للمثال السابق نحسب الوسط الحسابي كالتالي:

- نختار قيمة الوسط الفرضي x_0 ويفضل أن تكون القيمة ذات اكبر تكرار : $x_0 = 11$.

- ضرب الانحرافات di بالتكرار المقابل $\sum d_i n_i = 104$

$$\overline{x} = Xo + \frac{\sum_{i=1}^{i=k} d_i \, ni}{\sum \quad ni} = 11 + \frac{104}{36} = 13.88$$

ومنه قيمة الوسط الحسابي هي نفسها: **13.88** = $\overline{x}$

مرة أخرى، النتيجة متطابقة، مما يؤكد فعالية هذه الطريقة في حساب المتوسط الحسابي للبيانات المتصلة.

3-1 خصائص الوسط الحسابي :يمتلك الوسط الحسابي عدة خصائص أساسية تؤثر على كيفية استخدامه وتفسير نتائجه:

أ. يتأثر الوسط الحسابي **بالقيم المتطرفة أو الشاذة** (القيم التي تقع في طرفي مجال الدراسة.):هذه الخاصية هي **نقطة ضعف رئيسية للمتوسط الحسابي** .القيم المتطرفة هي **القيم التي تبتعد بشكل كبيرعن بقية البيانات**، سواء كانت عالية جدًا أو منخفضة جدًا. عندما توجد هذه القيم، فإنها **تسحب المتوسط الحسابي في اتجاهها**، مما يجعله لا يمثل "مركز" البيانات بشكل دقيق أو عادل. تخيل أنك تحسب متوسط دخل 10 أشخاص، و9 منهم يتقاضون 10000 دينار شهريًا، ولكن الشخص العاشر يتقاضى مليون دينار شهريًا. سيقفز المتوسط الحسابي بشكل كبير ليصبح رقمًا لا يمثل دخل أي شخص تقريبًا في المجموعة. في هذه الحالات، قد تكون مقاييس نزعة مركزية أخرى مثل **الوسيط** أكثر ملاءمة.

ادا كانت لنا البيانات التالية: 10-11-13-16 -70، فان المتوسط الحسابي لها هو كالتالي:

$$\overline{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} = \frac{10+11+13+16+70}{5} = \frac{120}{5} = 24$$

نلاحظ أن القيمة 70 تُعد قيمة متطرفة رفعت المتوسط من 12.5 إلى 24، مما جعله غير ممثل فعليًا **لغالبية** البيانات، ويُظهر كيف يمكن للقيم الشاذة أن تؤثر سلبًا على دقة المتوسط الحسابي. هذا يوضح كيف أن **المتوسط يصبح مضللاً** في وجود القيم المتطرفة.

ب. يستعمل الوسط الحسابي في حالة المتغيرات الكمية أي القابلة للقياس: يعني ذلك أن المتوسط الحسابي **لا يمكن استخدامه للمتغيرات الوصفية أو الفئوية**)مثل الجنس، اللون المفضل، الحالة الاجتماعية). لكي يكون المتوسط الحسابي ذا معنى، يجب أن تكون البيانات عبارة عن **أرقام يمكن جمعها وقسمتها**، أي بيانات كمية. على سبيل المثال، لا يمكن حساب متوسط الألوان المفضلة لأنها ليست قيمًا عددية.

ت. لا يمكن حساب المتوسط الحسابي في حالة البيانات المفتوحة (أكثر من و أقل من): البيانات المفتوحة هي تلك الفئات التي **ليس لها حدود واضحة في أحد طرفيها**، مثل "أقل من 10" أو "أكثر من 50". في هذه الحالات، لا يمكننا تحديد مركز الفئة (أو قيمة X_i (بدقة، وبالتالي **لا يمكننا تطبيق صيغة المتوسط الحسابي** لأننا لا نعرف القيمة الدقيقة لـ X_i أو C_i في تلك الفئات.

ث. إن مجموع الانحرافات قيم المتغير الإحصائي بالنسبة للوسط الحسابي تساوي الصفر و نرمز لهذه الانحرافات بالرمز ei حيث : $\Sigma ei = \Sigma(x_i - \bar{x}) = 0$. من خلال المثال السابق (3-1) نتأكد من هذه الخاصية مع العلم أن $\bar{x} = 3$ كالتالي:

عدد الغرف X_i	1	2	3	4	5	المجموع
$\Sigma(x_i - \bar{x})$	2-	1-	0	1	2	0

تدل هذه الخاصية على أن المتوسط الحسابي يقع في مركز البيانات.

ج. لا يمكن أن يكون لأي توزيع تكراري أكثر من وسط حسابي واحد: هذه الخاصية تعني أن **المتوسط الحسابي قيمة فريدة ومحددة** لأي مجموعة بيانات معينة. بمجرد أن تكون لديك مجموعة من الأرقام، لن يكون هناك سوى متوسط حسابي واحد فقط لها. هذه الخاصية تجعل المتوسط الحسابي مقياسًا **موضوعيًا وواضحًا**، على عكس مقاييس أخرى قد يكون لها أكثر من قيمة (مثل المنوال).

ح. مجموع مربعات الانحرافات بالنسبة للوسط الحسابي أصغر ما يمكن مقارنة ببقية القيم الأخرى: هذه خاصية رياضية مهمة توضح أن **المتوسط الحسابي هو نقطة التوازن الفريدة لمجموعة البيانات** .إذا طرحت المتوسط الحسابي من كل قيمة في المجموعة، ثم جمعت جميع هذه الفروقات (الانحرافات)، فإن المجموع سيكون دائمًا صفرًا. هذا يعني أن **مجموع الانحرافات الموجبة (القيم التي تقع فوق المتوسط) يساوي تمامًا مجموع الانحرافات السالبة (القيم التي تقع تحت المتوسط)**

$\Sigma(x_i - \bar{x})^2 < \Sigma(x_i - x_\alpha)^2$ حيث أن x_i لا يساوي X_α

تدل هذه الخاصية على أن **قيمة المتوسط الحسابي هي الأقرب إحصائيًا إلى جميع بيانات X_i مجتمعة** من أي قيمة أخرى يمكن اختيارها كمركز. وهذا ما يجعله مقياسًا مثاليًا لتمثيل "المركز" من منظور الانحرافات المربعة، وهي أساس العديد من الاختبارات والتحليلات الإحصائية الأخرى.

2 الوسيط(Médiane):

الوسيط هو أحد أهم مقاييس النزعة المركزية، ويُستخدم لتحديد القيمة التي تفصل مجموعة البيانات إلى نصفين متساويين من حيث عدد القيم: نصف القيم أقل منها، والنصف الآخر أكبر منها، ونرمز له ب (**Me**). يختلف الوسيط عن المتوسط الحسابي بكونه يعتمد على **رتب القيم** لا على مجموعها، لذا فهو مفيد جدًا عندما تكون البيانات غير متماثلة (ملتوية) أو تحتوي على **قيم متطرفة (شاذة)**.(A. V. Kuzin, 2016)

يُفضل استخدام الوسيط عندما تكون البيانات مشتتة أو تحتوي على انحرافات أو قيم شاذة، لأنه أقل تأثراً بها مقارنة بالوسط الحسابي. هذا يعني أن وجود قيمة عالية جدًا أو منخفضة جدًا في مجموعة البيانات لن يؤثر بشكل كبير على قيمة الوسيط، بينما قد يؤثر بشكل كبير على المتوسط الحسابي(2011) .Ghasem, B .

مثال توضيحي إذا كان الوسيط لأوزان مجموعة من الطلبة يساوي 70 كغ، معنى ذلك أن 50 % من الطلبة أوزانهم أقل من 70 كغ، و أن 50 % من الطلبة أوزانهم أكثر من 70 كغ.

1.2. الوسيط للبيانات الأولية (البيانات غير المبوبة): يتم تحديد الوسيط في حالة بيانات غير مبوبة حسب الخطوات التالية:

1-نرتب قيم السلسلة الإحصائية ترتيبا تصاعديا أو تنازليا.

2-نحدد ترتيب الوسيط، وهنا لابد أن نميز بين حالتين:

أ. عندما يكون عدد القيم N **فردي**، في هذه الحالة رتبة الوسيط هو عبارة عن:

$\frac{n+1}{2}$ أي $Me = X_{(x+1)/2}$

ب. اذا كان **عدد القيم N زوجيا**، فإن الوسيط هو متوسط القيمة التي رتبتها $\frac{n}{2}$ و القيمة التي رتبتها $\frac{n}{2}+1$ أي :

$$Me = \frac{X\left(\frac{x}{2}+1\right) + X(x/2)}{2} = \frac{\text{القيمة التي رتبتها } 2N + 2N + 1 \text{ القيمة التي رتبتها}}{2}$$

<u>مثال: 2-2 : في</u> **حالة N فردي**، لدينا السلسلة الإحصائية التالية حول علامات 07 طلبة في مقياس الإحصاء كالتالي:

13 ; 14 ; 7 ; 15 ; 9 ; 12 ; 10 . المطلوب تحديد الوسيط :

أولا :نرتب العلامات ترتيبا تصاعديا : 7;9 ;10 ; **12** ; 13; 14 ; 15

ثانيا : نحدد ترتيب الوسيط: $\frac{7+1}{2} = 4$: مرتبة الوسيط هي المرتبة الرابعة **إذن** Me = 12

الشرح 50% :من الطلبة علاماتهم أقل من أو تساوي 12، و 50% من الطلبة علاماتهم أكبر من أو تساوي 12. هذا يعني أن نصف الطلبة حصلوا على 12 أو أقل، والنصف الآخر حصلوا على 12 أو أكثر.

مثال: 2-3 : **في حالة N زوجي**، لدينا السلسلة الإحصائية التالية حول علامات 08 طلبة في مقياس الإحصاء كالتالي:

18، 16، 10، 13، 13،11،14،15

أولا: نرتب العلامات ترتيبا تصاعديا : 10، 11، 13، 13، 14، 15،16،18

ثانيا: مرتبة القيم التي تدخل في حساب الوسيط: $\frac{8}{2} = 4$ ، $\frac{8}{2} + 1 = 4 + 1 = 5$

ثالثا : تحديد قيمة الوسيط $\mathbf{Me} = \frac{X(5) + X(4)}{2} = \frac{14+13}{2} = \mathbf{13.5}$

الشرح: 50 % من الطلبة علاماتهم أقل من 13.5 ، و 50 % من الطلبة علاماتهم أكثر من 13.5.

2.2. الوسيط في حالة بيانات مبوبة أو في حالة توزيع تكراري: تختلف طريقة حساب الوسيط باختلاف نوعية المتغير الكمي سواء كان متغيرا كميا منقطعا أو متغيرا كميا منفصلا: (M. A. Al-Hajji & R. M. N. Al-Mousawi, 2008)

1.2.2 في حالة متغير متقطع(منفصل): لحساب الو سيط الحسابي في هذه الحالة، نتبع الخطوات التالية:

أ. تحديد التكرار التجميعي الصاعد.

ب. تحديد **رتبة الوسيط** وهو عبارة عن نصف مجموع التكرارات Σ ni/2.

ت. تحديد **فئة الوسيط**، وهي الفئة التي تقابل التكرار التجميعي الصاعد الذي يساوي ترتيب الوسيط أو الأكبر منه مباشرة. هناك حالتين:

- **الحالة الأولى**: اذا كانت قيمة (Σ ni/2) غير موجودة في التكرار المتجمع الصاعد، فهناك قيمة واحدة للوسيط Me، و هي قيمة المتغير المقابلة لأول قيمة اكبر من(Σ ni/2) في التكرار المتجمع الصاعد.

- **الحالة الثانية:** اذا كانت قيمة (Σ ni/2) موجودة في التكرار المتجمع الصاعد، فان هناك رتبتين للوسيط **الرتبة الاولى** هي المقابلة ل(Σ ni/2) في التكرار المتجمع الصاعد و **الرتبة الثانية** هي التي تلي قيمة (Σ ni/2) في التكرار المتجمع الصاعد. و قيمة الوسيط تحسب بالمتوسط الحسابي للقيمتين المقابلتين لرتبة2/ (Σ ni) و القيمة التي تلها.

لحساب الوسيط في **الحالة الأولى**، نأخذ نفس المثال **(1-3)** حول توزيع عدد الغرف لعينة من 80 أسرة حسب الجدول التالي:

المحور الرابع: مقاييس النزعة المركزية

عدد الغرف Xi	1	2	3	4	5	المجموع
عدد الأسر ni	9	19	30	12	10	80
التكرارات الصاعدة Ni↑	9	28	58	70	80	-

الحل:

1-تحديد التكرار التجميعي الصاعد في العمود الثالث من الجدول.

2-ترتيب الوسيط : $(\Sigma\ ni)/2=80/2=40$.

3-الفئة الوسيطية هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي 40 أو أكبر من 40 مباشرة. نلاحظ أن العدد 40 غير موجود من بين ت.م.ص ولكن العدد الأكبر منه مباشرة هو 58 إذن الفئة التي تقابل ت.م.ص 58 هي الفئة الوسيطية: 3

ادن قيمة الوسيط هي Me=3

الشرح: 50 % من الأسر تملك أقل من 03 غرف، و 50 % من الأسر تملك أكثر من 03 غرف.

لحساب الوسيط **في الحالة الثانية، نأخذ المثال(2-4)** حول توزيع عدد الغرف لعينة من 100 أسرة حسب الجدول التالي:

عدد الغرف Xi	1	2	3	4	5	المجموع
عدد الأسر ni	9	19	22	28	22	**100**
التكرارات الصاعدةNi↑	**9**	**28**	**50**	**78**	**100**	-

الحل:

- ترتيب الوسيط : $(\Sigma\ ni)/2=100/2=50$.

- الفئة الوسيطية هي الفئة التي تقابل التكرار المتجمع الصاعد الذي يساوي 50 مباشرة، اذن توجد رتبتين للوسيط:

الرتبة الاولى هي المقابلة ل($\Sigma\ ni/2$) في التكرار المتجمع الصاعد الذي يساوي 50 و التي يقابلها المتغير $X_i=3$ و الرتبة الثانية هي التي تلي قيمة($\Sigma\ ni/2$) في التكرار المتجمع الصاعد،التي تساوي 78 و التي تقابلها المتغير 4 . الوسيط هو المتوسط الحسابي للقيمتين 4 و 3 :

و بالتالي $Me=(3+4)/2=3.5$

الشرح: 50 % من الأسر تملك أقل من 3.5 غرفة، و 50 % من الأسر تملك أكثر من 3.5 غرفة.

2.2.2. في حالة متغير متصل (مستمر): اذا كان لدينا جدول توزيع تكراري على شكل فئات، نقوم بحساب **الوسيط بطريقة المد الداخلي:**

$$Me = Lim_{me} + \left(\frac{\frac{\Sigma ni}{2} - Nme-1\uparrow}{nme}\right) * L_{me}$$

حيث أن: Lim_{me} الحد الأدنى للفئة الوسيطية ، L_{me} طول الفئة الوسيطية ، $N_{me-1}\uparrow$ التكرار التجميعي الصاعد السابق للفئة الوسيطية ، n_{me} التكرار الأصلي للفئة الوسيطية.

لحساب الوسيط نأخذ نفس المثال (1-4) حول توزيع المعدلات المتحصل عليه في شهادة الباكالوريا لعينة من طلبة المركز الجامعي –النعامة، حسب الجدول التالي:

معدلات شهادة الباكالوريا Xi	[12-10[	[14-12[	[16-14[	[18-16[	[20-18[	المجموع
التكرارات ni	12	9	6	5	4	36
التكرارات الصاعدة Ni↑	**12**	**21**	**27**	**32**	**36**	-

1- تحديد التكرار التجميعي الصاعد في العمود الثالث من الجدول.

2- حساب رتبة الفئة الوسيطية : $(\Sigma ni/2) = 36/2 = 18$.

3- الفئة الوسيطية هي **[14-12[** و منه الحد الادنى للفئة الوسيطية **(Lim_{me}) هو 12**

4- التكرار المتجمع الصاعد الذي يسبق الفئة الوسيطية**($N_{me-1}\uparrow$) هو 12**

5- طول الفئة الوسيطية **(L_{me}) هو 12-14 =2**

6- تكرار الفئة الوسيطية **(n_{me}) هو 9**

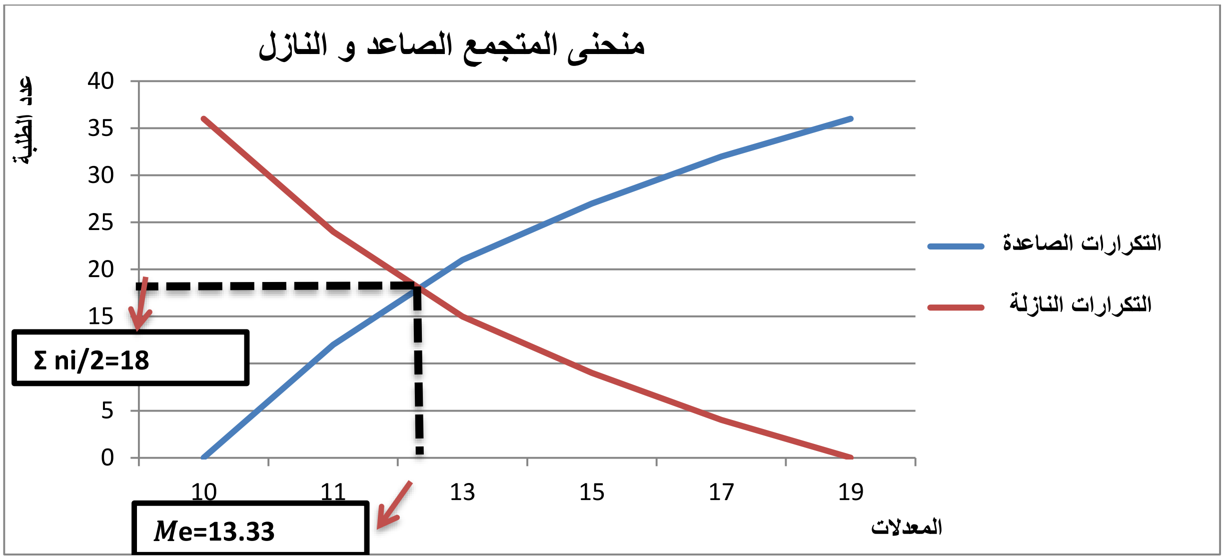
$$Me = Lim_{me} + \left(\frac{\frac{\Sigma ni}{2} - Nme-1\uparrow}{nme}\right) * L_{me} = 12 + \left(\frac{18-12}{9}\right) * 2 = 13.33$$

<u>الشرح:</u> 50 % من الطلبة حصلوا على معدلات **أقل من 13.33** ، و 50 % من الطلبة حصلوا على معدلات **أكثر من 13.33.**

2.2.3. حساب الوسيط بالطريقة البيانية: يتم الاعتماد في إيجاد قيمة الوسيط في حالة البيانات الكمية المستمرة باستخدام الطريقة البيانية على منحنيي كل من التكرار المتجمع الصاعد و التكرار المتجمع النازل. و بالاعتماد على جدول المثال السابق تم رسم كل من التكرار التجميعي الصاعد و النازل كما هو مبين في الشكل الموالي:

لحساب المنوال بيانا، نأخذ نفس **المثال السابق (1-4)**، ثم نرسم المدرج التكراري كما هو موضح أدناه:

الشكل رقم 2-1: المنحنى التكرار التجميعي الصاعد و النازل لتوزيع معدلات شهادة الباكالوريا لعينة من 40 طالبا من طلبة المركز الجامعي.



من خلال الشكل الوارد أعلاه نلاحظ أن فاصلة نقطة تقاطع منحنى التكرار التجميعي الصاعد مع منحنى التكرار التجميعي النازل هي قيمة الوسيط و تساوي إلى $M_e=13.33$ ، أما ترتيبها فينصف لنا حجم العينة و الذي يساوي إلى : $\Sigma ni/2=18$.

4.2.2. خواص الوسيط: يتميز الوسيط، كمقياس للنزعة المركزية، بعدة خصائص تجعله مفيدًا في تطبيقات معينة:

- **سهولة التعريف والحساب** :مفهوم الوسيط مباشر وسهل الفهم، فكل ما يتطلبه الأمر هو ترتيب البيانات والعثور على القيمة الوسطى أو متوسط القيمتين الوسطيتين.
- **يعتمد على القيم الوسطى فقط** :هذه الخاصية تعني أن الوسيط **لا يتأثر بالقيم القصوى أو المتطرفة** في مجموعة البيانات. سواء كانت القيمة الأعلى 100 أو 100000، فإن الوسيط سيبقى كما هو طالما أن ترتيب القيم الوسطى لم يتغير. هذا يجعله مقياسًا قويًا (Robust) ضد القيم الشاذة، على عكس المتوسط الحسابي الذي يتأثر بها بشدة.
- **لا يتأثر بوجود قيم متطرفة** :هذه النقطة هي جوهر قوة الوسيط. إذا كانت هناك أخطاء في إدخال البيانات أو وجود قيم غير طبيعية (مثل دخل شخص ثري جدًا في عينة من ذوي الدخل المتوسط)، فإن الوسيط سيبقى مقياسًا موثوقًا للنزعة المركزية.

- **يمكن إيجاد الوسيط بيانيًا** :كما رأينا في القسم السابق، توفر الطريقة البيانية وسيلة بصرية قوية لتحديد الوسيط وفهم توزع البيانات.
- **يستعمل الوسيط في الحالات التي تكون فيها بعض البيانات ناقصة بشرط أن نعرف ترتيبها** :على سبيل المثال، إذا كنا نعرف أن قيمة معينة هي "أقل من 10" ولكن لا نعرف القيمة الدقيقة، يمكننا استخدام الوسيط طالما أننا نعرف ترتيب هذه القيمة بالنسبة لبقية البيانات. هذه مرونة لا تتوفر دائمًا في مقاييس أخرى مثل المتوسط الحسابي الذي يتطلب معرفة جميع القيم بدقة.

3 المنوال (Le mode).

يُقصد بالمنوال **القيمة الأكثر تكرارًا** في مجموعة البيانات، ونرمز له بـ (Mo) .يُستخدم بكثرة في حالة البيانات الوصفية لمعرفة النمط (المستوى) الشائع، أو عندما تكون البيانات متقاربة بشكل كبير. قد يكون لمجموعة البيانات منوال واحد، وتُسمى حينئذٍ بيانات **وحيدة المنوال** (Unimodal) . قد يكون لها أكثر من منوال، وتُسمى **متعددة المنوال** (Multimodal) .وقد لا يكون لمجموعة البيانات أي منوال، وبذلك تُسمى **عديمة المنوال** (Amodal أو No Mode).

1.3. حساب المنوال في حالة البيانات غير المبوبة: في هذه الحالة لا توجد أي عمليات حسابية لإيجاد المنوال، كل ما يتطلبه إيجاد المنوال هو معرفة القيمة التي تتكرر أكثر من غيرها، وذلك كما هو واضح في الأمثلة التالية:

- قيمة المنوال للبيانات التالية : 11 ، 12 ، 9 ، 10 ، 8 ، 11 ، 13 هي: M_O=11 لأنها الأكثر تكرار من غيرها، يوجد منوال واحد و بالتالي هي **بيانات أحادية المنوال**.
- قيمة المنوال للبيانات التالية : 11 ، 12 ، 9 ، 10 ، 9 ، 11 ، 15 : يوجد منوالان: M_O=9 وM_O=11 و بالتالي هي **بيانات ثنائية المنوال**
- قيمة المنوال للبيانات التالية : 11 ، 12 ، 9 ، 10 ، 9 ، 11 ، 12 : يوجد 3 منوالات: ، M_O=9 وM_O=12 و بالتالي هي **بيانات متعددة المنوال**.
- قيمة المنوال للبيانات التالية : 15 ، 12 ، 9 ، 10 ، 8 ، 11 ، 13 : لا يوجد منوال و بالتالي هي **بيانات عديمة المنوال**.

كما يمكن ايجاد المنوال للمتغيرات النوعية بنفس الطريقة وذلك كما هو واضح في السلسلة التالية:

- قيمة المنوال للبيانات التالية : متوسط ، ضعيف ، متوسط ، حسن ، جيد ، ممتاز، ضعيف جدا: هي القيمة الأكثر تكرارا من غيرها هي : **متوسط** =M_O . يوجد منوال واحد و بالتالي هي **بيانات أحادية المنوال**.

3.2. حساب المنوال في حالة توزيع تكراري لمتغير إحصائي منفصل:

لحساب المنوال في حالة البيانات المبوبة ذات المتغيرات الكمية المنقطعة، نستخرج القيمة التي تقابل أكبر تكرار من تكرارات القيم المجاورة لها. لحساب المنوال، نأخذ المثال **(رقم 3-1)** و المتعلق بتوزيع الغرف لعينة تتكون من 30 أسرة في مجمع سكني في ولاية النعامة:

الجدول رقم 3-1 : توزيع الغرف لعينة تتكون من 30 أسرة في مجمع سكني في ولاية النعامة:

عدد الغرفXi	1	2	3	4	المجموع
عدد الأسر ni	5	8	10	7	30
كما اشرنا سابقا المنوال هو القيمة التي تقابل أكبر تكرار الذي هو 10، ومن خلال الجدول أعلاه، فقيمة المنوال هي $M_O=3$ أي أن غالبية الأسر تملك 3 غرف.					

3.3. حساب المنوال في حالة توزيع تكراري لمتغير إحصائي متصل: لحساب المنوال لبيانات مبوبة في جداول تكرارية منتظمة أي فئاتها متساوية الطول نتبع الخطوات التالية: (Goldfarb, B., & Pardoux, C. (2011). d. p. 45)

- نحدد أولاً الفئة التي تحتوي المنوال ويطلق عليها الفئة المنوالية وهي الفئة المقابلة لأكبر تكرار.
- نحسب قيمة المنوال باستخدام القانون التالي:

$$Mo = A + (\frac{D1}{D1 + D2} * L)$$

حيث أن:

- A : الحد الأدنى للفئة المنوالية(أو الحد الأدنى الحقيقي في حالة البيانات التي تمثل متغيراً منفصلاً.)
- **D1**: الفرق الأول= تكرار الفئة المنوالية – تكرار الفئة السابقة لها.
- **D2**: الفرق الثاني = تكرار الفئة المنوالية – تكرار الفئة اللاحقة لها.
- **L** : طول الفئة المنوالية .

ملاحظة :إذا كان جدول التوزيع التكراري غير منتظم، يجب تعديل تكراراته قبل تطبيق خطوات إيجاد المنوال التي أشرنا إليها.

لحساب المنوال، نأخذ المثال (3-2) حول توزيع معدلات شهادة الباكالوريا لعينة من 40 طالبا من طلبة المركز الجامعي –النعامة، حسب الجدول التالي:

الجدول رقم 3-2: توزيع معدلات شهادة الباكالوريا لعينة من 40 طالبا من طلبة المركز الجامعي –النعامة:

معدلات شهادة الباكالوريا Xi	]12-10]	]14-12]	] 16-14]	]18-16]	]20-18]	المجموع
التكرارات عدد الطلبةni	6	9	14	7	4	40

الحل:

من خلال الجدول أعلاه نلاحظ أن الفئة المنوالية هي [16-14] كونها تحتوي على أكبر تكرار(تكرار الفئة المنوالية هو 14 طالبا) وبالتالي فإن الحد الأدنى للفئة A= 14 و طول الفئة هو L=2 و بالتالي يمكن حساب كل من:

- **D1**: الفرق الأول= تكرار الفئة المنوالية – تكرار الفئة السابقة لها = 14- 9= 5
- **D2**: الفرق الثاني = تكرار الفئة المنوالية – تكرار الفئة اللاحقة لها = 14-7=7

وعليه نحسب قيمة المنوال باستخدام القانون التالي:

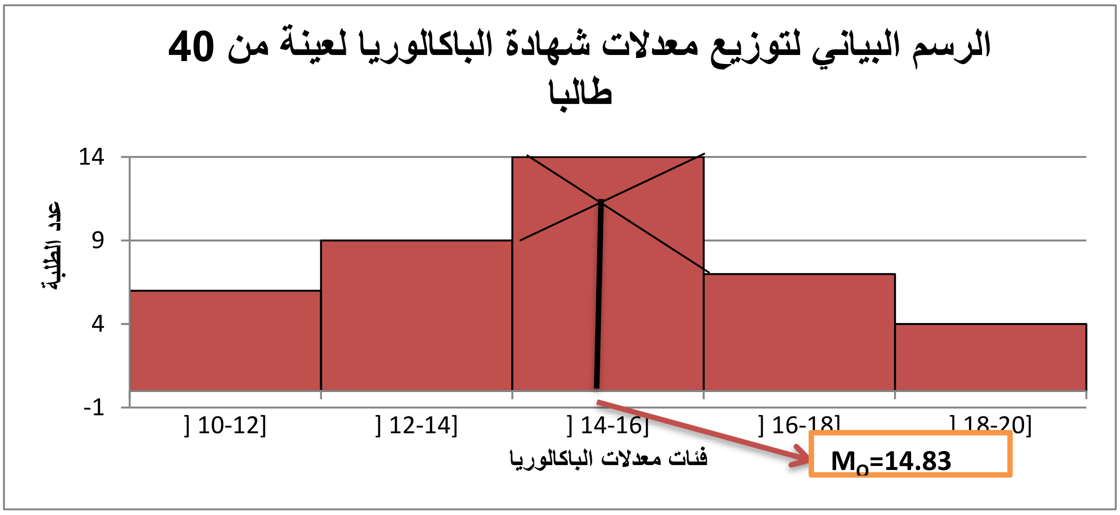
$$Mo = A + (\frac{D1}{D1+D2} * L) = 14 + (\frac{5}{5+7} * 2) = 14.83$$

M_O=14.83

الشرح: معدلات الباكالوريا لأغلبية الطلبة تقدر ب 14.83.

4.3. حساب المنوال بالطريقة البيانية: في هذه الطريقة، يتم رسم المدرج التكراري. ثم يتم أولاً الوصل بين النقطة التي تمثل الحد الأعلى للفئة السابقة للفئة المنوالية والنقطة التي تمثل الحد الأعلى للفئة المنوالية. ثم ثانيًا، يتم الوصل بين النقطة التي تمثل الحد الأدنى للفئة المنوالية والنقطة التي تمثل الحد الأدنى للفئة اللاحقة للفئة المنوالية. يتقاطع الخطان في نقطة، شاقولها على المحور الأفقي يمثل القيمة المنوالية. لحساب المنوال بيانيًا، نأخذ نفس المثال السابق (3-2)، ثم نرسم المدرج التكراري كما هو موضح أدناه: Anderson, D. R., Sweeney, D. J., Camm, J. D., Williams, T. A., & Cochran, J. J. (2015).. p. 55-56

الشكل رقم 3-1: الرسم البياني لتوزيع معدلات شهادة الباكالوريا لعينة من 40 طالبا من طلبة المركز



5-3 خصائص المنوال : يتميز المنوال بعدة خصائص:

1. **أسهل مقاييس النزعة المركزية في حسابه**، ولا يتأثر بوجود قيم متطرفة.
2. يمكن حسابه في حالة **التوزيعات التكرارية المفتوحة** بشرط ألا تكون الفئة المفتوحة هي الفئة المنوالية.
3. يمكن إيجاد المنوال **للبيانات النوعية**، كما يمكن إيجاد المنوال **بيانيًا**. ((Hamdani, H. (2005، ص -40 42).
4. **ليس له معنى إذا كانت البيانات قليلة العدد** وقد لا يوجد أصلاً. أما في حالة البيانات كثيرة العدد، فله معنى معقول وله أهمية كبيرة وخاصة في عملية التسويق، فمثلاً شركات تسويق الألبسة لا تهتم بالوسط الحسابي أو بالوسيط، بل تهتم بالمقياس الأكثر شيوعًا وهو المنوال.
5. يتأثر كثيرًا **بطريقة اختيار الفئات التكرارية للتوزيع**، فإذا غيرنا تقسيم الفئات لنفس التوزيع، يحدث تغيير في التكرارات، وفي الغالب يحدث تغيير في موقع الفئة المنوالية، ولذلك نحصل على قيم مختلفة للمنوال.

4 مشتقات المتوسط الحسابي: المتوسط الهندسي، المتوسط التو افقي والمتوسط التربيعي.

4 .1. المتوسط الهندسي(Moyenne Géométrique)

يعرف المتوسط الهندسي لمجموعة من البيانات عددها n بأنه الجذر النوني لحاصل ضرب هذه البيانات ويرمز له بالرمز G_X. يعتبر المتوسط الهندسي وسطا مناسبا للأرقام القياسية خاصة المعدلات أو تقدير عدد السكان أو حساب النسب(Guisnel, H. (2018 .

أ. المتوسط الهندسي في حالة البيانات غير المبوبة:

إذا كان لدينا عدد n من البيانات $x_1 . x_2 . \; .... x_n$ من الأعداد الموجبة، فإن المتوسط الهندسي لهذه الأعداد يعرف كما يلي :

$$G_x = \sqrt[n]{x_1 . x_2 . \; .... x_n}$$

لتسهيل العمليات الحسابية، نقوم بإدخال اللوغاريتم على الصيغة السابقة، لتصبح بالشكل التالي:

= 1/n(log x1, log x2, log x3,.......log xn) = 1/n log Xi Log G_x = 1/n log (x1, x2, x3,.......xn)

$$log \quad (G_x) = \frac{\sum_{i=1}^{n} Log(\text{X})}{n}$$

ومن خلال استخدام معكوس دالة اللوغارتم تصبح صيغة المتوسط الهندسي كما يلي

$$G_x = \text{antilog}[\; \frac{\sum_{\text{i}=1}^{\text{n}} \log_{10}(\text{X})}{\text{n}}]$$

<u>مثال رقم 4-1:</u> لتكن السلسلة الإحصائية التالية 8; 11; 12; 6; 10 ;14المطلوب حساب المتوسط الهندسي:

$$G_x = \sqrt[6]{8 * 11 * 12 * 6 * 10 * 14} = 9.80$$

أو نقوم بتطبيق الصيغة اللوغاريتمية على المثال السابق:

log 8	log 11	log 12	log 6	log 10	log 14	log X_i
0.903	1.04	1.08	0.77	1	1.14	5.948

5.948/6 = 0.99 **Log G_x** = 1/6(log 8+log 11+ log 12+log 6+ log 10+ log 14)=

ومن خلال استخدام معكوس دالة اللوغارتم تصبح صيغة المتوسط الهندسي كما يلي

إذن قيمة المتوسط الهندسي هي: **G_x = · $10^{0.99}$ = 9.80**

ب. المتوسط الهندسي للبيانات المبوبة: يحسب **المتوسط الهندسي للبيانات المبوبة** وفق العلاقة بالشكل التالي:

$$G_x = \sqrt[n]{C_1{}^{n1} . C_2{}^{n2} \; .... C_n{}^{ni}}$$

حيث ان C_i تمثل متوسط الفئة i والى k ، ويتم **تعويض C_i ب X_i في حالة المتغير المنفصل.**

عندما تكون السلاسل أو الجداول تحمل أرقاما كبيرة، فان حساب المتوسط الحسابي يتم بالصيغة **اللوغاريتمية** بالشكل التالي:

$$Log(Gx) = \frac{\sum_{i=1}^{k} \text{ni Log Ci}}{\sum_{i=1}^{k} n_i}$$

ومن خلال استخدام معكوس دالة اللوغارتم تصبح صيغة المتوسط الهندسي كما يلي

$$Gx = antilog[\frac{\sum_{i=1}^{k} ni \; Log(C_i)}{\sum_{i=1}^{k} n_i}]$$

مثال 4-2: اوجد المتوسط الهندسي للتوزيع التكراري التالي:

n_i Log C_i	Log C_i	$C_i^{n_i}$	n_i	C_i	الفئات
1,431	0.477	27	3	3	2-4
2,796	0.699	625	4	5	4-6
5,07	0,845	117649	6	7	6-8
7,632	0,954	43046721	8	9	8-10
5,205	1,041	161051	5	11	10-12
22,134	-	43326073	25	-	المجموع

ومن خلال تطبيق الصيغة العادية، نحسب المتوسط الهندسي بالشكل التالي:

$$\boldsymbol{G_x = \sqrt[n]{C_1{}^{n1} * C_2{}^{n2} \ldots C_n{}^{ni}} \quad G_x = \sqrt[25]{3^3 * 5^4 * 7^6 * 8^9 *} 11^5 = 7.68}$$

ومن خلال تطبيق **الصيغة اللوغاريتمية**، نحسب المتوسط الهندسي بالشكل التالي:

$$\boldsymbol{G_x} = antilog\left[\frac{\Sigma niLog\ Ci}{\sum_{i=1}^{k} n_i}\right] = antilog\left(\frac{22.134}{25}\right) = \boldsymbol{7.68}$$

2. 4 . المتوسط التوافقي (Moyenne Harmonique) : المتوسط التوافقي للقيم الإحصائية هو مقلوب الوسط الحسابي لمقلوب قيم هذه المتغيرات. يستعمل المتوسط التوافقي في حالة وجود علاقة عكسية بين ظاهرتين، مثلا كلما ارتفعت الأسعار انخفضت القدرة الشرائية.(Droesbeke, J. J., & Lejeune, M. 2014))

أ. **المتوسط التوافقي البسيط** :لتكن X_1 ;X_2 ;... ....X_n : قيم المتغير الإحصائي ، حيث نرمز للمتوسط التوافقي بالرمز H_m ، و يحسب المتوسط التوافقي للبيانات غير المبوبة بالعلاقة التالية:

$$\boldsymbol{H_m = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}}$$

مثال 4-3: احسب المتوسط التوافقي للبيانات التالية (11,21,27,31,33) .

$$\boldsymbol{H_m} = \frac{5}{\frac{1}{11} + \frac{1}{21} + \frac{1}{27} + \frac{1}{31} + \frac{1}{33}} = \frac{5}{0.2381} = \boldsymbol{20.9995}$$

ب. **المتوسط التوافقي للبيانات المبوبة** : عند وجود k من المشاهدات $X_1, X_2, X_3, \ldots X_k$ وبالتكرارات $n_1, n_2, n_3, \ldots n_k$ فان المتوسط التوافقي للبيانات يعرف بالعلاقة التالية

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

في حالة **المتغير المستمر** نستبدل X_i ب C_i التي تمثل متوسط الفئة i والى k. **لحساب** المتوسط التوافقي في هذه الحالة، نأخذ المثال السابق **(4-2)** كالتالي:

المجموع	10-12	8-10	6-8	4-6	2-4	**الفئات**
-	11	9	7	5	3	C_i
26	5	8	6	4	3	n_i
4	0,455	0,889	0,857	0.8	1	n_i/C_i

$$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{X_i}} = \frac{26}{4.001} = 6.498$$

3. 4 المتوسط التربيعي: المتوسط التربيعي عبارة عن الجذر التربيعي لمتوسط مربعات قيم المتغير الإحصائي حيث نرمز له بالرمز **QM**:

أ. **المتوسط التربيعي البسيط:** لتكن : X1 ;X2 ;... ....Xn قيم المتغير الإحصائي، نحسب المتوسط التربيعي للبيانات غير المبوبة بالعلاقة التالية:

$$\boldsymbol{QM} = \sqrt{\frac{X_1^2 + X_2^2 + \cdots \dots + X_n^2}{n}} = \sqrt{\frac{\sum X_i^2}{n}}$$

مثال 4-4: احسب المتوسط التربيعي للبيانات التالية (10,24,34,46,50) .

$$\boldsymbol{QM} = \sqrt{\frac{10^2 + 24^2 + 34^2 + 46^2 + 50^2}{5}} = \sqrt{\frac{6448}{5}} = 35.91$$

ب. **المتوسط التربيعي للبيانات المبوبة :** عند وجود k من المشاهدات $X_1, X_2, X_3, \dots\dots X_k$ وبالتكرارات $n_1, n_2, n_3, \dots n_k$ فان المتوسط التربيعي للبيانات يعرف بالعلاقة التالية:

$$\boldsymbol{QM} = \sqrt{\frac{n_1X_1^2 + n_2X_2^2 + \cdots \dots + n_kX_n^2}{\sum_{i=1}^{k} n_i}} = \sqrt{\frac{\sum niX_i^2}{\sum_{i=1}^{k} n_i}}$$

في حالة **المتغير المستمر** نستبدل X_i ب C_i التي تمثل مركز الفئة i والى k.

لحساب المتوسط التربيعي في هذه الحالة، نأخذ المثال السابق **(4-2)** كالتالي:

المجموع	10-12	8-10	6-8	4-6	2-4	**الفئات**
-	11	9	7	5	3	C_i
25	5	8	6	4	3	n_i
	121	81	49	25	9	C_i^2
1674	605	648	294	100	27	niX_i^2

$$QM=\sqrt{\frac{n_1C_1^2+n_2C_2^2+\cdots\ldots+n_kC_n^2}{\sum_{i=1}^{k}n_i}}=\sqrt{\frac{\sum niC_i^2}{\sum_{i=1}^{k}n_i}}=\sqrt{\frac{1674}{25}}=\mathbf{8.18}$$

4-4العلاقة بين المتوسط الحسابي و مشتقاته: من خلال حساب المتوسط الحسابي و مشتقاته في الأمثلة السابقة، نلاحظ أن **8.18> 7.92 <7.68 > 6.24** ، أي أنه مهما تكن البيانات، فانه في كل الأحوال:

$$H_m < GX < \bar{X} < QM$$

تجدر الإشارة الى أن الوسطين الحسابي و التربيعي يتأثران بالقيم الكبيرة، بينما الهندسي و التوافقي أكثر تحيزا للقيم الصغيرة(Bouhadiba, A. 2017 , P 52) .

الفصل الخامس : مقاييس التشتت (Mesures de la dispersion)

عند مقارنة مجموعتين من البيانات، لا تكفي التمثيلات البيانية أو مقاييس النزعة المركزية وحدها لإبراز الفروق الحقيقية بينهما، إذ قد تتساوى القيم المركزية رغم وجود تباين كبير في تشتت القيم. لذلك تُستخدم **مقاييس التشتت** لقياس مدى تقارب أو تباعد القيم عن بعضها البعض أو عن القيمة المركزية. وتُظهر هذه المقاييس درجة تجانس البيانات: فإذا كانت القيم متقاربة دلّ ذلك على ضعف التشتت، أما تباعدها فيشير إلى تشتت كبير. ومن أبرز مقاييس التشتت المستعملة في الإحصاء :**المدى، الانحراف الربيعي، الانحراف المتوسط، التباين، الانحراف المعياري، ومعامل الاختلاف**) طبيّه، 2008، ص. 64).

1 المدى العام (Range):

يمثل المدى العام زوج البيانات الذي يضم اقل القيم واكبرها ويحسب المدى بعد ترتيب البيانات بالفرق ما بين اكبر قيمة واصغرها وبالصيغة التالية:

المدى = اعلى قيمة – أدنى قيمة $R = X_{max} - X_{min}$

مثال: المدى للبيانات التالية : (15,12,10,9,8,7,5,3,2): اكبر قيمة هي 15 واصغر قيمة هي 2 وعليه فان

R = 15 -2 = 13

اما في **حالة البيانات المبوبة**، فان المدى يمثل الفرق بين الحد الاعلى للفئة الاخيرة والحد الادنى للفئة الاولى . من خلال المثال السابق (3-2) حول توزيع معدلات شهادة الباكالوريا، فان المدى العام يساوي:

R = الحد الادنى للفئة الأولى - الحد الاعلى للفئة الاخيرة **=20 -10 = 10**

ملاحظة: نلاحظ أن المدى يهتم بالقيمتين المتطرفتين اللتين قد تكونان منفصلتين عن باقي افراد المجموعة ولذلك فان المدى المطلق لا يعطي دلالة واضحة لمدى انتشار وتوزيع القيم (جلاطو، 2001).

في المقابل، المدى يتأثر بوحدة القياس ولا يعكس دائمًا درجة الانتشار بشكل دقيق. لذلك نستخدم **المدى النسبي** لقياس التشتت بشكل مقارن:

$$R\% = \frac{R}{\bar{x}} * 100$$

وهو يبيّن مدى التشتت نسبةً إلى الوسط الحسابي، ما يسمح بالمقارنة بين توزيعات مختلفة حتى لو كانت وحدات القياس غير متجانسة.

إذا كان المتوسط الحسابي للسلسلة السابقة هو:

$$= 7.89\bar{x} = \frac{2+3+5+7+8+9+10+12+15}{9} = \frac{71}{9}$$

فإن:

$$*100 = 1.65 R\% = \frac{13}{7.89}$$

أي أن المدى النسبي يعادل 1.65 ضعف المتوسط الحسابي.

2 المدى الربيعي

يعتبر **نصف المدى الربيعي** مقياساً مفيداً للتشتت، عندما يكون الهدف من استخدامه وصف البيانات ومعرفة هذا التشتت بصورة عامة ومبسطة. يستخدم هذا المقياس أيضا في حالة وجود قيم شاذة او متطرفة لأنه في مثل هذه الحالة لا يمكن ان يعطينا المدى صورة صادقة عن الفروق بين الدرجات.

قبل التعرض لكل من المدى الربيعي و نصف المدى الربيعي، يجب التعرض **لمفهوم الربيعات**. إن الربيعات هي كل مجموعة من البيانات الممكن تقسيمها إلى أربعة أقسام متساوية بعد ترتيبها تصاعديا، أين يفصل بين كل قسم ما يسمى بالربيع، و يمكن التمييز بربيعين أساسين هما الربيع الأدنى(**الربيع الأول**)والربيع الأعلى(**الربيع الثالث**)، أما **الربيع الأوسط** فهو عبارة عن الوسيط، ونرمز للربيعات ب **Q_i حيث i=1,2,3**

- **الربيع الأول** :هو قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى قسمين، يحتوي القسم الأول على 25% من الوحدات الإحصائية، أما القسم الثاني فيحتوي على 75 % من هذه الوحدات . تقع قيمة الربيع الأول في نهاية الربع الأول من التوزيع الإحصائي، وفي هذا الموضع تكون مرتبته تساوي 25 % أو $(\frac{\Sigma ni}{4})$ حسب الترتيب التصاعدي لقيم المتغير الإحصائي، ويرمز له بالرمز Q_1 . و يعطى بالعلاقة التالية:

$$Q1 = \text{الحد الأدنى لفئة الربيع الأول} + \left(\frac{\frac{\Sigma ni}{4} - \text{التكرار المتجمع الصاعد للفئة قبل فئة الربيع الأول}}{\text{تكرار فئة الربيع الأول}}\right) * \text{طول الفئة} =$$

$$\text{Lim}_{Q1} + \left(\frac{\frac{\Sigma ni}{4} - NQ-1\uparrow}{nQ1}\right) * L_{Q1}$$

- **الربيع الثالث :**هو قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى قسمين، يحتوي القسم الأول على 75 % من المجتمع المدروس، أما القسم الثاني فيحتوي على 25 % المتبقية منه حسب الترتيب التصاعدي للمتغير المدروس، ويرمز له بالرمز $Q3$ ، ومرتبته تساوي 75 % أو$(\frac{3\Sigma ni}{4})$. وتعطى علاقته الإحصائية بالعلاقة التالية:

$$Q3 = \text{الحد الأدنى لفئة الربيع الثالث} + \left(\frac{\frac{3\Sigma\, ni}{4} - \text{التكرار المتجمع الصاعد للفئة قبل فئة الربيع الثالث}}{\text{تكرار فئة الربيع الثالث}}\right) * \text{طول الفئة}$$

$$= \text{Lim}_{Q3} + \left(\frac{\frac{3\Sigma\, ni}{4} - NQ1-1\uparrow}{n\, Q3}\right) * L_{Q3}$$

وبعد التعرف على الربيعات و كيفية حسابها، فانه يمكن التعرف عن كل من المدى الربيعي و نصف المدى الربيعي:

- **المدى الربيعي :** هو عبارة عن الفرق بين الربيع الثالث و الربيع الأول و يرمز له بالرمز(**RQ**):

$$\mathbf{RQ} = \boldsymbol{Q3} - \boldsymbol{Q1}$$

- **نصف المدى الربيعي:** يعرف كذلك بالانحراف الربيعي نصف الفرق بين قيمتي الربيع الاول والربيع الثالث.

$$\mathbf{HRQ} = \left(\frac{\boldsymbol{Q3} - \boldsymbol{Q1}}{\mathbf{2}}\right)$$

3 الانحراف المتوسط

يعرف الانحراف المتوسط لمجموعة من البيانات بانه **متوسط الانحر افات المطلقة لتلك الدرجات عن وسطها** الحسابي، ويرمز له بالرمز **Dm** . **الانحراف المتوسط أكثر دقة من المدى والانحراف الربيعي** لشموله كل القيم ولكنه محدود الاستخدام لتأثره بالقيم الشاذة وتجاهله الإشارة السالبة. يحسب المتوسط الحسابي بالطريقة التالية:

- **في حالة سلسلة احصائية:**

$$\mathbf{Dm} = \frac{\sum_{i=1}^{i=n} |x_i - \bar{x}|}{\Sigma ni}$$

- **في حالة التوزيع التكراري:**

$$\mathbf{Dm} = \frac{\sum_{i=1}^{i=k} ni|x_i - \bar{x}|}{\Sigma ni}$$

ولحساب الانحراف المتوسط **في حالة سلسلة احصائية**، نأخذ المثال (رقم ......) لدينا سلسلة البيانات التالية:

20 , 16 , 12 , 9 , 6 , 4 , 3

الحل:

- نستخرج قيمة الوسط الحسابي: $\overline{x} = \frac{20+16+12+9+6+4+3}{7} = \frac{70}{7} = 10$
- ثم بعد ذلك نجد الفرق المطلق بين قيمة الوسط الحسابي , وكل درجة من الدرجات. وتستخرج قيمة الانحراف المتوسط بقسمة حاصل جمع الانحرافات المطلقة على عدد الدرجات.

$$Dm = \frac{(3-10)+(4-10)+(6-10)(9-10)+(12-10)+(16-10)+(20-10)}{7}$$

$$\frac{7+6+4+1+2+6+10}{7} = \frac{36}{7} = 5.14$$

غير أن هذا المقياس يتأثر بوحدة القياس. لذلك نستخدم الانحراف المتوسط النسبي:

$$Dm\% = \frac{Dm}{\overline{x}}$$

الذي يوضح مقدار التشتت بالنسبة إلى المتوسط، ويتيح إجراء المقارنة بين سلاسل إحصائية مختلفة.

وباستخدام المثال السابق فان : $Dm\% = \frac{5.14}{10} = 0.514\ (51.4\%)$

يعني أن **الانحر افات المطلقة تعادل 51.4% من قيمة المتوسط الحسابي**، أي أن التشتت متوسط الحجم مقارنة بوسط البيانات.

4 التباين والانحراف المعياري

أ. **التباين (variance) :** هو أحد مقاييس التشتت الأوسع انتشارا، و يعرف بأنه متوسط مربع انحرافات القيم عن وسطها الحسابي، وعادة ما يرمز له بالرمز V(x) أو σ^2 ، ويحسب التباين بالعلاقة التالية:

- **في حالة سلسلة احصائية:**

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

- **في حالة التوزيع التكراري:**

$$V(x) = \frac{\sum_{i=1}^{i=k} ni\,(x_i - \overline{x})^2}{\Sigma ni}$$

ويمكن تبسيط المعادلة بحساب التكرار النسبي لتصبح بالصيغة التالية (Doane & Seward, 2016) :

$$(\mathbf{x}) = \sum_{i=1}^{i=k} fi\,(x_i - \bar{x})^2$$

ب. **__الانحراف المعياري:__** ويسمى في بعض الاحيان بالانحراف القياسي، حيث يعرف بأنه الجذر التربيعي لمعدل مجموع مربعات انحرافات القيم عن متوسطها ، اي ان الانحراف المعياري (S_X) يمثل جذر التباين كالاتي **(Hurlin, 2015)**:

- **في حالة سلسلة احصائية:**

$$s_x = \sqrt{\frac{\sum_{i=1}^{i=n}(x_i-\bar{x})^2}{ni}} \quad \text{أو} \quad s_x = \sqrt{s_x^2}$$

- **في حالة التوزيع التكراري:**

$$s_x = \sqrt{\sum_{i=1}^{i=k} fi\,(x_i-\bar{x})^2} \quad \text{أو} \quad s_x = \sqrt{\frac{\sum_{i=1}^{i=k} ni\,(x_i-\bar{x})^2}{\Sigma ni}}$$

5 معامل الاختلاف (coefficient of variation)

والذي يعتبر وسيلة ملائمة لوصف الانحراف المعياري ويحسب من خلال قسمة الانحراف المعياري على المتوسط وهو من مقاييس التشتت النسبية الخالية من وحدة القياس، **حيث يسمح بإجراء المقارنة ما بين المجموعات بوحدات قياس متشابهة او مختلفة وفق الصيغة التالية .**

$$cv = \frac{s}{\bar{x}} \times 100\%$$

كلما كان هذا المعامل صغيرا كلما دل ذلك على انتشار البيانات في مدى ضيق و أقرب للوسط الحسابي ويستدل منه على أن البيانات أكثر تجانسا.

6 تمرين تطبيقي لمقاييس التشتت:

لحساب مختلف مقاييس التشتت التي رأيناها سابقا، نأخذ المثال السابق **(المثال 3 -4):** الجدول التالي يمثل الأجور الساعية لعينة من 50 عاملا في مصنع البسكويت (**الوحدة 10دج**).

الأجور X الساعية	مركز الفئة Ci	التكرارات ni عدد العمال	التكرارات الصاعدة Ni↑	Ni *Ci	$ni\|xi-\bar{x}\|$	ni (ci- $\bar{x}$)^2
] 50- 45]	47,5	11	11	522,5	97,9	871,31
] 55- 50]	52,5	13	24	682,5	50,7	365,04
] 60- 55]	57,5	10	34	575	11	41,14
] 65- 60]	62,5	8	42	500	48,8	1562,82
] 70- 65]	67,5	8	50	540	88,8	6160,5
المجموع	-	50	-	2820	297,2	9000,81

الحل:

1- حساب المتوسط الحسابي:

$$\bar{x} = \frac{\sum_{i=1}^{i=k} ni\, C_i}{\Sigma ni} = \frac{2820}{50} = 56.4$$

الشرح: متوسط الأجور الساعية لهذه العينة هو **560** دج $\bar{x}=$

2- حساب المدى العام :

R = الحد الاعلى للفئة الاخيرة - الحد الادنى للفئة الأولى = **70 -45 = 25**

الشرح: الفرق بين الأجر الساعي الأكبر و الأصغر هو **250**دج =R

3- المدى الربيعي:

لحساب المدى الربيعي، نحسب الربيع الأول و الثالث على الشكل التالي:

أ. <u>**حساب الربيع الأول:**</u>

رتبة الفئة الربيعية الأولى: (Σ ni)/4=50/4=12.5، الفئة الربيعية الأولى هي **[55- 50]** و منه الحد الادنى للفئة الربيعية الأولى (Lim $_{Q1}$) هو 50، التكرار المتجمع الصاعد الذي يسبق الفئة الربيعية الأولى (N_{Q-1}↑) هو 11، طول الفئة الربيعية الأولى (L_{Q1}) هو 55-50= 5، تكرار الفئة الربيعية الأولى (n_{Q1})هو 13.

$$Q1 = \text{الحد الأدنى لفئة الربيع الأول} + \left(\frac{\frac{\Sigma ni}{4} - \text{التكرار المتجمع الصاعد للفئة قبل فئة الربيع الأول}}{\text{فئة تكرار الربيع الأول}}\right) * \text{طول الفئة}$$

$$= 50 + \left(\frac{\frac{50}{4} - 11}{13}\right) * 5 = 50.57$$

<u>**الشرح:**</u> 25 % من الأجور الساعية أقل من **505.7**دج ، 75 % من الأجور الساعية **أكثر من 505.7**دج.

ب. **حساب الربيع الثالث :**

رتبة الفئة الربيعية الثالثة: $(3\Sigma ni)/4=150/4=37.5$، الفئة الربيعية الثالثة هي] **65- 60]** و منه الحد الادنى للفئة الربيعية الثالثة (Lim_{Q3}) هو 60، التكرار المتجمع الصاعد الذي يسبق الفئة الربيعية الثالثة ($N_{Q-3}\uparrow$) هو 11، طول الفئة الربيعية الثالثة (L_{Q3}) هو 65-55 =5، تكرار الفئة الربيعية الثالثة (n_{Q3})هو 8.

$$Q3 = \text{الحد الأدنى لفئة الربيع الثالث} + \left(\frac{\frac{3\Sigma ni}{4} - \text{التكرار المتجمع الصاعد للفئة قبل فئة الربيع الثالث}}{\text{فئة تكرار الربيع الثالث}}\right) * \text{طول الفئة}$$

$$= 60 + \left(\frac{\frac{150}{4} - 34}{8}\right) * 5 = 62.18$$

الشرح: 75 % من الأجور الساعية أقل من **621.8دج** ، 25 % من الأجور الساعية **أكثر من 621.8دج.**

و منه يمكننا حساب المدى الربيعي ونصف المدى الربيعي ، على الشكل التالي:

$$\mathbf{RQ = Q3 - Q1 = 62.18 - 50.57 = 11.61}$$

$$\boldsymbol{HRQ = ((Q3 - Q1)/2) = ((11.62)/2) = 5.81}$$

الشرح: 50 % من أجور العمال الساعية محصورة ما بين 505 و 621 دج، بمدى يقدر ب 116دج.

4- الانحراف المتوسط:

$$\mathbf{Dm} = \frac{\sum_{i=1}^{i=k} ni|C_i - \bar{x}|}{\Sigma ni} = \frac{297,2}{50} = 5.94$$

الشرح: يقدر الانحراف المتوسط عن المتوسط الحسابي للأجور الساعية ب 59.4دج= **Dm**

6- التباين (variance) :

$$V(x) = \frac{\sum_{i=1}^{i=k} ni\,(x_i - \bar{x})^2}{\Sigma ni} = \frac{9000}{50} = 180$$

7- الانحراف المعياري:

$$s_x = \sqrt{\frac{\sum_{i=1}^{i=k} ni\,(x_i - \bar{x})^2}{\Sigma ni}} = \sqrt{V_x^2} = \sqrt{180} = 13.41$$

الشرح: الفرق المعياري بين الأجور الساعية للعمال و الأجر المتوسط يقدر ب 134 دج = S_x

8-معامل الاختلاف(الانحراف المعياري النسبي) CV

لمقارنة معاملات الاختلاف، نفترض أنه في دراسة مماثلة على عينة أخرى، كان متوسط الأجور الساعية لهذه العينة هو **550 دج** =$\bar{x}$ و الانحراف المعياري يساوي 138 دج = S_x

الدراسة الأولى:

$$cv = \frac{s}{\bar{x}} \times 100\% = \frac{13.41}{56.4} \times 100\% = \mathbf{23.77}\%$$

<u>**الدراسة الثانية:**</u>

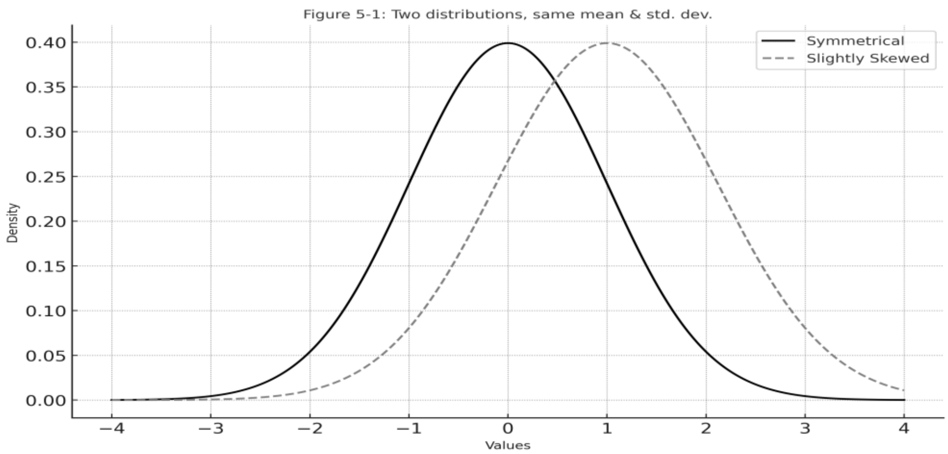
$$cv = \frac{s}{\bar{x}} \times 100\% = \frac{13.8}{55} \times 100\% = \mathbf{25}\%$$

نلاحظ أن معامل الاختلاف كبير، مما يعني وجود تشتت **وتباين كبير للأجور الساعية للعمال**. وبمأن معامل الاختلاف في الدراسة الثانية أكبر من الدراسة الأولى، فإننا نستنتج أن **تشتت الأجور الساعية في الدراسة الثانية أكبر منه في الدراسة الأولى**.

الفصل السادس: مقاييس الشكل

تلعب الإحصاءات الوصفية دورًا أساسيًا في تحليل البيانات وفهم خصائصها، إذ تتيح مقاييس النزعة المركزية تحديد موقع القيم الأكثر تمثيلاً، فيما تقدم مقاييس التشتت فكرة عن مدى انتشارها. غير أن هذه الأدوات وحدها قد لا تكفي، فقد تتشابه مجموعتان في المتوسط والانحراف المعياري لكنهما تختلفان في شكل التوزيع، وهو ما يعكس خصائص مميزة للسلوك الإحصائي للظاهرة. هنا تبرز أهمية **مقاييس الشكل** التي تمنح رؤية أعمق لطبيعة التوزيع الاحتمالي. فالالتواء (Skewness) يكشف عن درجة تناظر أو انحراف التوزيع، بينما يوضح التفرطح (Kurtosis)مدى حدّة أو تسطّح القمة مقارنة بالتوزيع الطبيعي. وتعتمد هذه المقاييس على مفهوم العزوم (Moments)التي تصف السمات الرياضية الجوهرية للتوزيع. ومن خلال هذه المؤشرات، يمكن اختيار النماذج الإحصائية الأنسب للتحليل، بما يضمن تفسيرًا أكثر دقة وواقعية للبيانات(Le Hay & Chanvril-Ligneel, 2014 ، ص. 87).

الشكل (6-1) :مثال لمجموعتين لهما نفس المتوسط والانحراف المعياري، لكن بأشكال توزيع مختلفة (إحدى المجموعتين ملتوية نحو اليمين، والأخرى متناظرة).



1 العزوم(Moments) :

1.2. تعريف العزوم: يُعد **العزم** قيمة إحصائية تُستخدم لتلخيص خصائص التوزيع (Lecoutre, 2023 ، ص. 45). ويُعرّف بأنه القيمة المتوقعة لمقدار مرفوع لأس معيّن. ويمكن صياغته من **الدرجة r** حول قيمة اختيارية A كما يلي:

- بالنسبة للبيانات غير المبوبة :

$$M_{r(A)} = \left(\frac{1}{n}\right) \Sigma\ (X_i - A)^r$$

- للبيانات المبوبة:

$$M_{r(A)} = \left(\frac{1}{\Sigma} n_i\right) \Sigma\ [\ n_i (X_i - A)^r]$$

حيث:

- ni : تكرار الفئة i.
- Xi : مركز الفئة i.
- A : القيمة الاختيارية التي يُحسب العزم حولها.

2.2. أنواع العزوم:

أ. **العزوم حول الصفر (العزوم الأولية) :**يُحسب هذا النوع من العزوم عندما تكون القيمة الاختيارية A=0. وتُعرف بالصيغة :

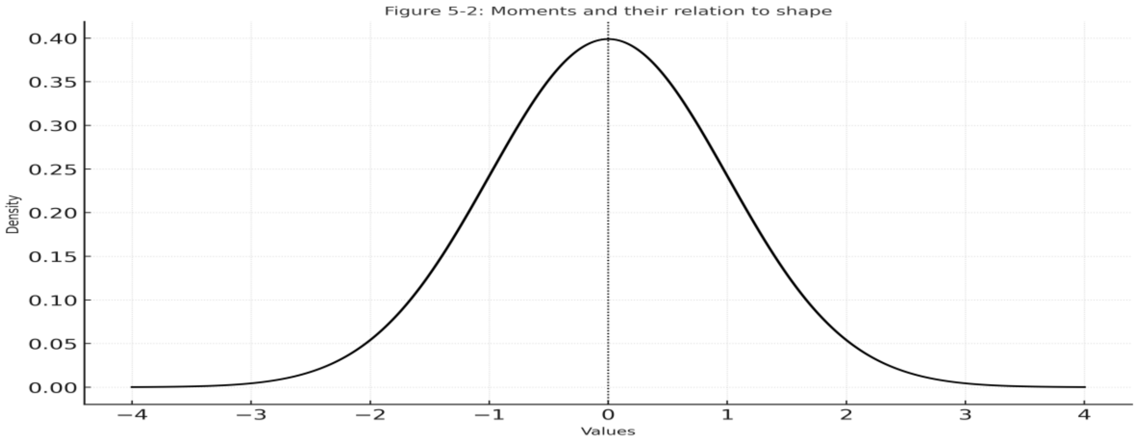
$$\mathrm{m}'_{\mathrm{r}} = \Sigma\ [\ \mathrm{n_i} * \mathrm{X_i^r}]$$

العزم الأولي من الدرجة الأولى (r=1) يمثل المتوسط الحسابي: m_1' = X⁻

ب. **العزوم المركزية :**هذا هو النوع الأكثر أهمية في دراسة مقاييس الشكل. تُحسب العزوم المركزية حول المتوسط الحسابي (X⁻) أي (A=X). ويُرمز لها بالرمز μr. يصف كل عزم مركزي خاصية محددة للتوزيع:

- **العزم المركزي الأول :(μ1)** يكون دائمًا مساويًا للصفر(μ1=0) ، لأنه يمثل مجموع الانحرافات عن المتوسط.
- **العزم المركزي الثاني :(μ2)** يُعرف بالتباين، وهو مقياس رئيسي للتشتت.(μ2=σ2) .
- **العزم المركزي الثالث :(μ3)** يُستخدم لقياس الالتواء، حيث يعتمد على الاتجاه الذي تتراكم فيه الانحرافات المكعبة من المتوسط.
- **العزم المركزي الرابع :(μ4)** يُستخدم لقياس التفرطح، حيث يعتمد على الانحرافات المرفوعة للأس الرابع من المتوسط.

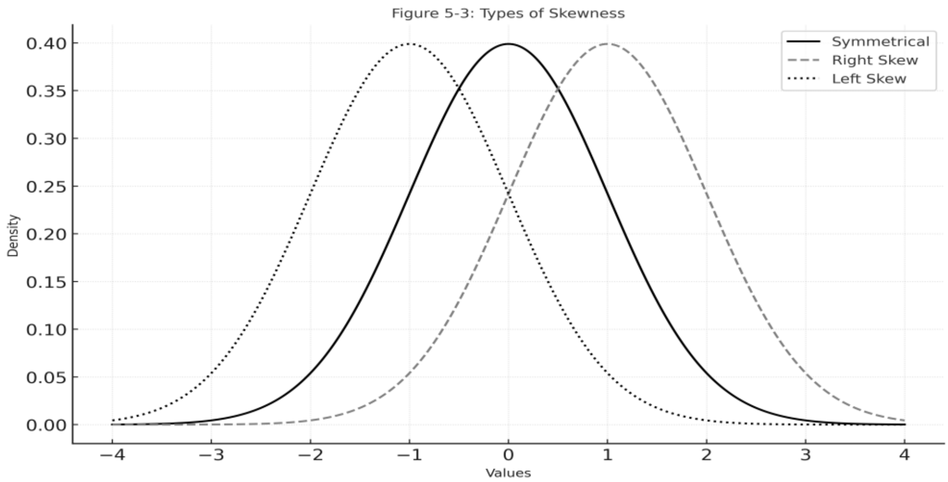
الشكل (6-2): رسم توضيحي للعزوم الأربعة وعلاقتها بالشكل العام للتوزيع.



2 مقاييس الالتواء(Skewness)

1.2. تعريف الالتواء: الالتواء هو مقياس لعدم تماثل التوزيع الإحصائي .(Baggio et al., 2022) يُعطي فكرة عن مدى انحراف البيانات عن الشكل المتماثل الشبيه بالجرس.

الشكل (6-3) :أشكال الالتواء الثلاثة (موجب – متناظر – سالب).



- **التوزيع المتناظر :(Symmetrical Distribution) يكون المنحنى متماثلًا على جانبي المتوسط. في هذه الحالة، تتساوى مقاييس النزعة المركزية :المنوال = الوسيط = المتوسط الحسابي.**
- **التوزيع الملتوي نحو اليمين :(Positive Skewness) يمتد ذيل المنحنى الطويل نحو اليمين. يحدث هذا غالبًا عندما توجد قيم متطرفة عالية تؤثر على المتوسط. في هذه الحالة، تكون العلاقة بين مقاييس النزعة المركزية هي :المنوال >الوسيط >المتوسط الحسابي.**

- **التوزيع الملتوي نحو اليسار :(Negative Skewness) يمتد ذيل المنحنى الطويل نحو اليسار. يحدث هذا عندما توجد قيم متطرفة منخفضة. في هذه الحالة، تكون العلاقة بين مقاييس النزعة المركزية هي : المنوال <الوسيط <المتوسط الحسابي.**

2.2. مقاييس الالتواء:

أ. **معامل بيرسون للالتواء(Pearson's Skewness Coefficient) :** يعتمد هذا المعامل على العلاقة بين المتوسط والمنوال أو المتوسط والوسيط. يُعدّ مقياسًا بسيطًا وسريعًا ولكنه أقل دقة من معامل فيشر لأنه لا يأخذ في الاعتبار جميع قيم البيانات (أبو صالح، 1998). يحسب هذا المعامل وفق **الصيغ التالية:**

$$P2 = \frac{(Me - \bar{X})^{\ 3}}{\sigma} \qquad P1 = \frac{\mathrm{Mo} - \bar{X}}{\sigma}$$

حيث: $\bar{X}$: المتوسط الحسابي./ Mo : المنوال. / Me :الوسيط. / σ : الانحراف المعياري.

التفسير:

- $P>0$:التوزيع ملتوي نحو اليمين.
- $P<0$:التوزيع ملتوي نحو اليسار.
- $P=0$:التوزيع متناظر.

ب. **معامل فيشر للالتواء α_F (Fisher's Skewness Coefficient) :** يُعتبر هذا المعامل أكثر دقة لأنه يعتمد على العزم المركزي من الدرجة الثالثة(μ_3) ، والذي يأخذ في الاعتبار جميع البيانات. حسب **الصيغة: التالية:** $\alpha_F = \frac{\mu_3}{\sigma^3}$

التفسير:

- $\alpha_F>0$:التوزيع ملتوي نحو اليمين.
- $\alpha_F<0$:التوزيع ملتوي نحو اليسار.
- $\alpha_F=0$:التوزيع متناظر.

3.2. تمرين تطبيقي على الالتواء:

يمثل الجدول ادناه توزيع رواتب 40 موظفًا (بالألف دينار):

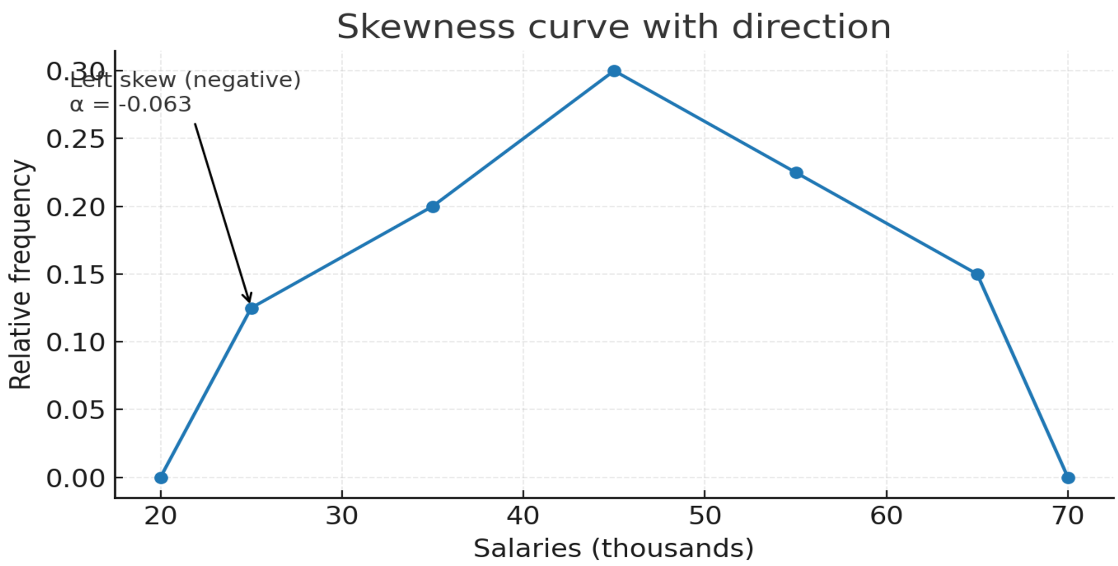
الفئة Xi	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
التكرارNi	5	8	12	9	6

المطلوب:

1. حساب المتوسط الحسابي والانحراف المعياري.
2. حساب معامل فيشر للالتواء وتفسيره.

الفئة Xi	التكرار Ni	مركز الفئة Ci	Ni *Ci	$(ci-\bar{x})^2$	$ni\,(ci-\bar{x})^2$	$ni\,(ci-\bar{x})^3$
20–30	5	25	125	430.5625	2152.8125	-44717.4
30–40	8	35	280	115.5625	924.5	-9938.32
40–50	12	45	540	0.5625	6.75	-5.04
50–60	9	55	495	85.5625	770.0625	7123.05
60–70	6	65	390	370.5625	2223.375	42895.56
المجموع	40	—	1830	—	6077.5	-4642.15

الشكل (6-4): منحنى التوزيع التكراري للتمرين مع إبراز اتجاه الالتواء.



الحل:

1- حساب المتوسط الحسابي:

$$\overline{x} = \frac{\sum_{i=1}^{i=k} ni\, C_i}{\Sigma ni} = \frac{1830}{40} = 45.75$$

2- الانحراف المعياري:

$$= \sqrt{V_x^2} = \sqrt{151.94} = 12.33 \quad s_x = \sqrt{\frac{\sum_{i=1}^{i=k} ni\,(x_i-\overline{x})^2}{\Sigma ni}} = \sqrt{\frac{6077,5}{40}}$$

3- العزم المركزي الثالث

$$\mu_3 = \frac{ni\,(x_i - \overline{x})^3}{\Sigma ni} = \frac{-4642.15}{40} = -116.05$$

4- معامل فيشر للالتواء

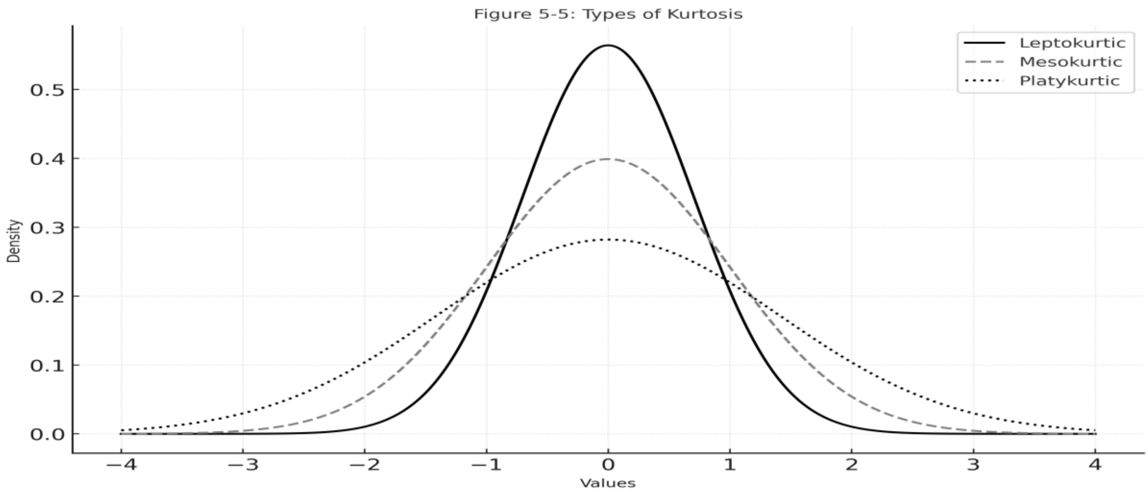
$$\alpha_F = \frac{\mu_3}{\sigma^3} = \frac{-116.05}{12.32^3} = -0.063$$

بما أن $\alpha F<0$ وقيمته صغيرة حوالي(-0.63)، فإن التوزيع ملتوٍ قليلًا نحو اليسار وليس نحو اليمين.

3 مقاييس التفرطح(Kurtosis) :

1.3. تعريف التفرطح : هو مقياس لدرجة "تسطح" أو "حدة" قمة التوزيع .(Bailly & Carrère, 2015) يُعطي فكرة عن مدى تمركز البيانات حول المتوسط، وعن سُمك ذيول التوزيع (أي احتمالية وجود قيم متطرفة). يُستخدم هذا المقياس بشكل أساسي مع التوزيعات المتناظرة.

الشكل (6-5): منحنيات مقارنة للتوزيع المتطاول، الطبيعي، والمتفرطح



- **التوزيع المتطاول :(Leptokurtic)** قمة حادة ومدببة، مما يدل على أن معظم البيانات متمركزة بشكل كبير حول المتوسط وأن هناك ذيول سميكة (وجود قيم متطرفة أكثر).

- **التوزيع الطبيعي :(Mesokurtic)** قمة معتدلة تشبه شكل الجرس، وهو معيار يُستخدم للمقارنة.
- **التوزيع المتفرطح :(Platykurtic)** قمة مسطحة، مما يدل على أن البيانات أكثر تشتتًا وتبعثرًا حول المتوسط وأن هناك ذيول رفيعة (قيم متطرفة أقل).

2.3. معامل فيشر للتفرطح(Fisher's Kurtosis Coefficient) : يُعتبر هذا المعامل الأدق لقياس التفرطح، ويُقارن بالتوزيع الطبيعي الذي قيمته صفر. يعتمد على العزم المركزي من الدرجة الرابعة(μ4) .

$$\mu_4 = \frac{\boldsymbol{ni\ (x_i - \overline{x})^4}}{\boldsymbol{\Sigma ni}}$$

معامل فيشر للتفرطح $\boldsymbol{\beta_F}$

$$\boldsymbol{\beta_F = \frac{\mu_4}{\sigma^4} - 3}$$

التفسير:

β_F **:** 0=التوزيع طبيعي(Mesokurtic) .

β_F<0: التوزيع متفرطح(Platykurtic) .

3.3. تمرين تطبيقي على التفرطح: سنعمل على نفس بيانات التمرين السابق للالتواء (رواتب 40 موظفًا) :

1- العزم المركزي الرابع

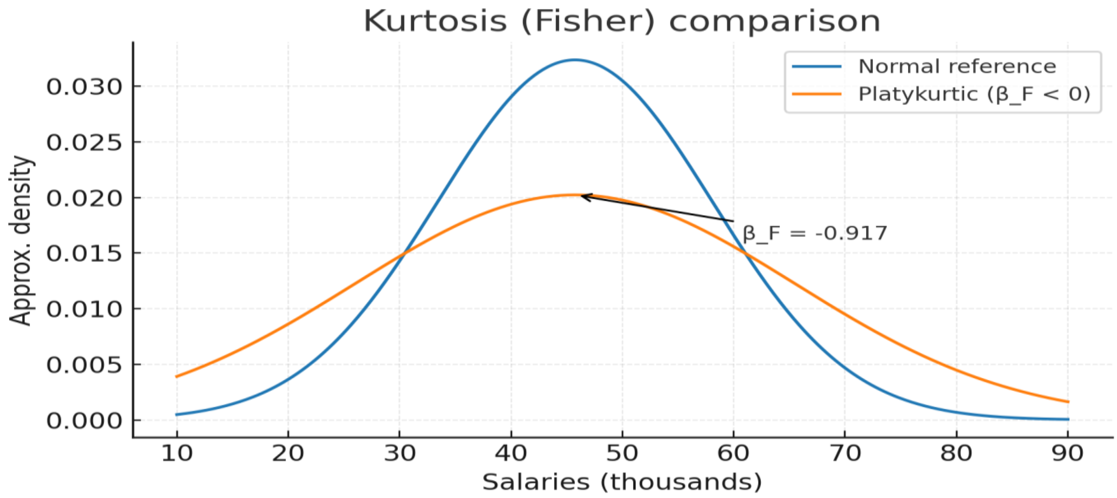
$$\boldsymbol{\mu_4 = \frac{ni\ (x_i - \overline{x})^4}{\Sigma ni} = \frac{1,926,013.87}{40} = 48,150.35}$$

2- معامل فيشر للتفرطح $\boldsymbol{\beta_F}$

$$\beta_F = \frac{\mu_4}{\sigma^4} - 3 = \frac{\boldsymbol{48,150.35}}{\boldsymbol{23085}} - 3 = 2.0867 - 3 = \boldsymbol{-0.913}$$

بما أن قيمة ***βF*** سالبة *(0.9133−)* ، فإن التوزيع **متفرطح** *(Platykurtic).* هذا يعني أن قمة التوزيع **مسطحة نسبياً** وذيوله **أرق** مقارنةً بالتوزيع الطبيعي.

الشكل (6-6): منحنى التوزيع للتمرين مع إبراز درجة التفرطح



تُعد مقاييس الشكل (الالتواء والتفرطح) أدوات إحصائية أساسية لفهم طبيعة التوزيع التكراري. فبينما تخبرنا المقاييس التقليدية أين يقع مركز البيانات ومدى تشتتها، تُعطينا مقاييس الشكل فكرة عن تناظرها وشكل قمتها. هذا الفهم الشامل لخصائص البيانات يساعد الباحثين في اتخاذ قرارات أفضل وتحديد النماذج الإحصائية الأنسب.

الفصل السابع: مقاييس التمركز: منحنى لورنز ومؤشر جيني

تُعد مقاييس التمركز أدوات إحصائية بالغة الأهمية في مجال الاقتصاد والعلوم الاجتماعية، حيث تُستخدم لتقييم مدى عدالة توزيع الدخل أو الثروة في مجتمع ما (السيد، 2020). لا تقتصر أهميتها على التحليل الاقتصادي فحسب، بل تمتد إلى دراسة التفاوت في توزيع الظواهر المختلفة، مثل الأراضي الزراعية أو الممتلكات. يُعد كل من **منحنى لورنز** و**مؤشر جيني** من أبرز هذه المقاييس، حيث يوفران رؤية شاملة حول الفرق بين التوزيع الفعلي للبيانات والتوزيع العادل النظري.

1 منحنى لورنز (Lorenz Curve) :

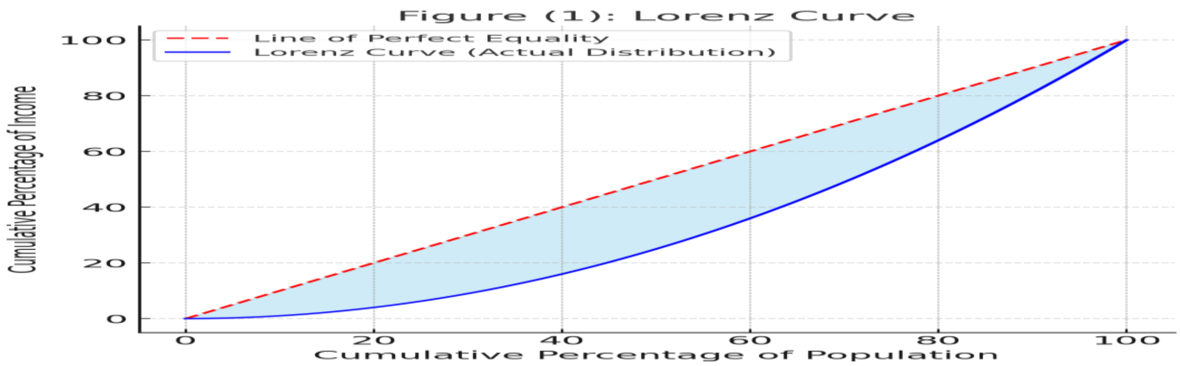
1.1. التعريف: منحنى لورنز هو تمثيل بياني لتوزيع الدخل أو الثروة. طوره الاقتصادي الأمريكي ماكس أوتو لورنز في عام 1905 لتمثيل عدم المساواة في توزيع الثروة. هو أحد الأساليب الكارتوغرافية لقياس التركز وبيان شكل وعدالة في توزيع ظاهرة معينة .(Milanovic, 2016) يعتمد المنحنى على متغيرين: متغير مستقل يمثل على محور الفواصل (المحور الأفقي)، ومتغير تابع يمثل على محور التراتيب (المحور الرأسي).

- **المحور الأفقي (س):** يمثل التكرار التجميعي الصاعد للمتغير المستقل، مثل النسبة المئوية للسكان (من الأقل دخلاً إلى الأعلى دخلاً).
- **المحور الرأسي (ص):** يمثل التكرار التجميعي الصاعد للمتغير التابع، مثل النسبة المئوية للدخل أو الثروة.

2.1. خصائص المنحنى وتفسيره:

- **خط المساواة التامة :**يُسمى أيضًا خط العدالة أو خط التماثل. هو خط مستقيم بزاوية 45 درجة يمتد من نقطة الأصل (0,0) إلى الزاوية العليا اليمنى (100,100). يمثل هذا الخط حالة التوزيع المثالي، حيث يحصل كل فرد على نفس نسبة الدخل أو الثروة.
- **منحنى التوزيع الفعلي :**يقع دائمًا أسفل خط المساواة التامة، ويعكس التوزيع الفعلي غير المتساوي. كلما ابتعد منحنى لورنز عن هذا الخط المستقيم، كان التوزيع أكثر عدم مساواة. على سبيل المثال، إذا كان منحنى لورنز يظهر أن 20% من السكان يمتلكون 5% فقط من الدخل، فهذا يشير إلى تفاوت كبير.

شكل (1) منحنى لورنز



2 مؤشر جيني (Gini Index)

1.2 التعريف: يُعد مؤشر جيني أحد المقاييس الأكثر شيوعًا في قياس عدالة توزيع الدخل القومي .(Giraud, 2015) تعتمد فكرته على منحنى لورنز، ولكنه يُميز بأنه يعطي قياسًا رقميًا لدرجة عدم المساواة في التوزيع، مما يسهل المقارنة بين الفترات الزمنية أو الدول المختلفة. طوره الإحصائي الإيطالي كورادو جيني في عام 1912.

2.2 طريقة الحساب: توجد عدة طرق لحساب **مؤشر جيني**:

أ. **الطريقة الهندسية**: تعتمد على حساب **المساحة المحصورة بين منحنى لورنز وخط المساواة التامة** (وتسمى **منطقة عدم التماثل** A) ثم مقارنتها بالمساحة الكلية للمثلث الواقع تحت خط المساواة(**A+B**) .

- تمثل **A+B** مساحة المثلث أسفل خط المساواة التامة، والتي تساوي:

$$\text{المساحة} = \frac{\text{الارتفاع} * \text{القاعدة}}{2} = \frac{1 * 1}{2} = 0.5$$

- يحسب مؤشر جيني بالعلاقة التالية:

$$\text{مؤشر جيني} = \frac{\boldsymbol{A}}{\boldsymbol{A+B}}$$

حيث:

- **A**: منطقة عدم التماثل (بين خط المساواة ومنحنى لورنز)
- **A+B** : المساحة الكلية للمثلث تحت خط المساواة.

وبما أن A+B=0.5 دائمًا، يمكن تبسيط الصيغة إلى:

G=1−2B حيث **B** هي المساحة تحت منحنى لورنز.

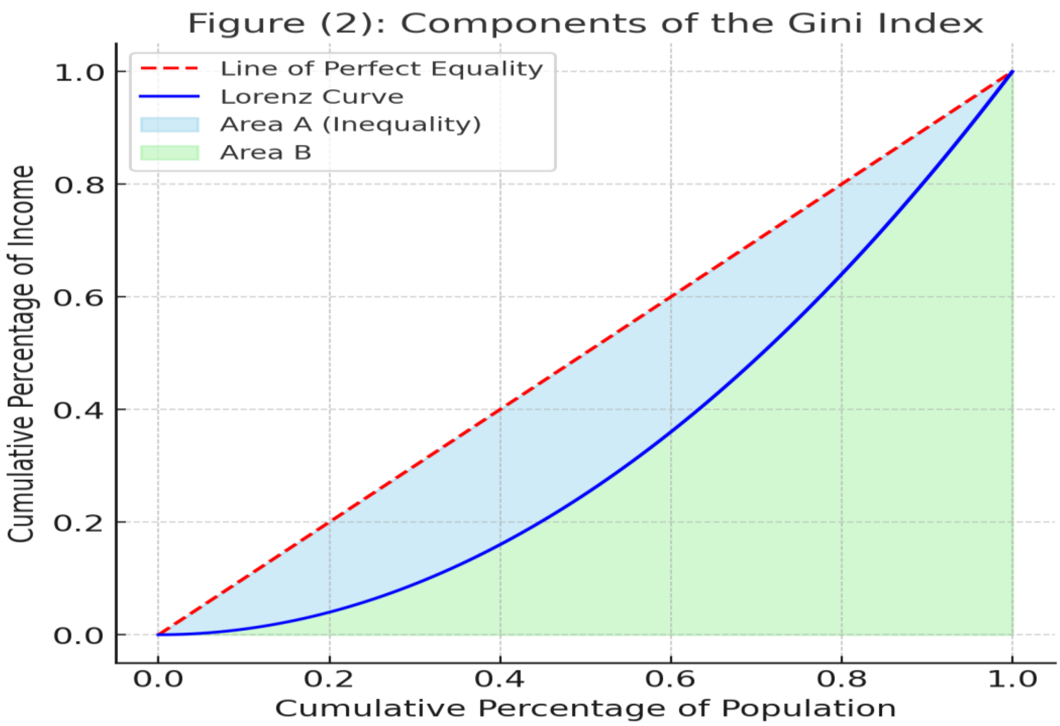
ب. **الطريقة الحسابية:** يمكن حساب المساحة B تحت منحنى لورنز باستخدام قاعدة شبه المنحرف:

$$(x - X_{n-1}) * (y + Y_{n-1})^n / 2 \ = \ B$$

حيث : Xi: النسبة التراكمية للأسر. Yi : النسبة التراكمية للدخل.

ثم نحسب مؤشر جيني كما يلي: **G=2B− 1**

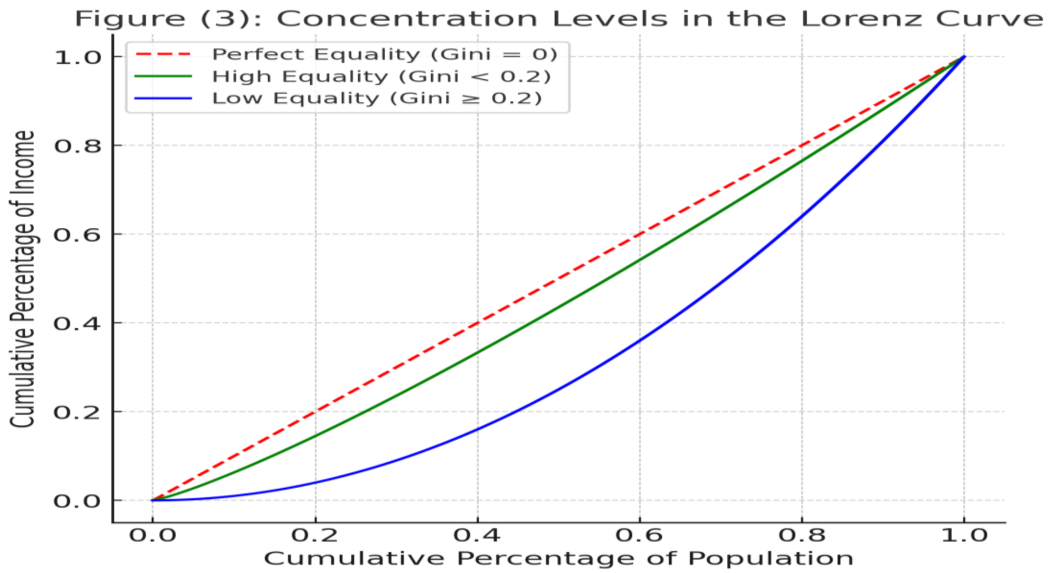
شكل (2) مكونات مؤشر جيني



3.2. خصائص المؤشر وتفسيره

- تتراوح قيمة مؤشر جيني بين الصفر والواحد.(0≤IGini≤1)
- **IGini=0:** يعني وجود مساواة مطلقة، حيث ينطبق منحنى لورنز على خط العدالة.
- **IGini=1:** يعني وجود عدم مساواة مطلقة، حيث ينطبق منحنى لورنز على المحورين.
- إذا كان المؤشر IGini<0.2، فإن التوزيع يعتبر أكثر عدالة، حيث يقترب منحنى لورنز من خط العدالة (Boudet, 2019).
- إذا كان المؤشر IGini≥0.2، فإن التوزيع يعتبر أقل عدالة، حيث يبتعد منحنى لورنز عن خط العدالة.

شكل (3) مستويات التركز في منحنى لورنز



3 **تمرين تطبيقي:**

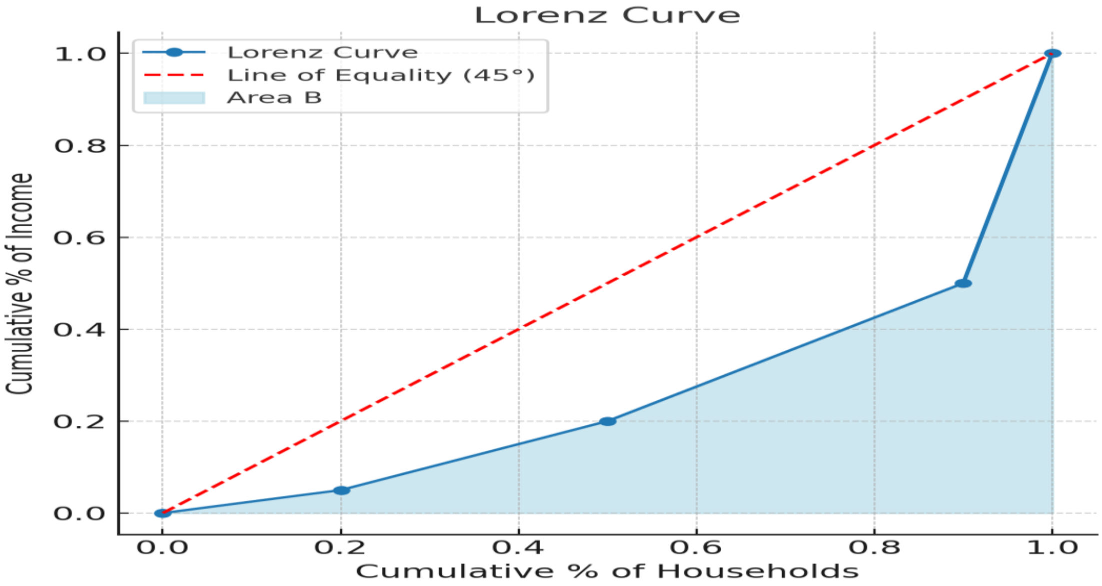
يمثل الجدول ادناه البيانات حول توزيع الدخل في مدينة ما، ونرغب في رسم **منحنى لورنز** و حساب **مؤشر جيني**.

فئة السكان	عدد الأسر	الدخل (مليون)	التراكمي للأسر %	التراكمي للدخل %
1	20	5	20%	5%
2	30	15	50%	20%
3	40	30	90%	50%
4	10	50	100%	100%

الحل:

1. **بناء جدول التوزيع التكراري التراكمي** :كما هو موضح أعلاه، نقوم بحساب النسب المئوية التراكمية لكل من الأسر والدخل.
2. **رسم منحنى لورنز** :نرسم المحور الأفقي للنسبة المئوية التراكمية للأسر، والمحور الرأسي للنسبة المئوية التراكمية للدخل. ثم نرسم خط المساواة التامة بزاوية 45 درجة ونرسم منحنى التوزيع الفعلي باستخدام النقاط من الجدول.

شكل (4) : منحنى لورنز



3. **حساب مؤشر جيني** :هناك عدة طرق لحساب مؤشر جيني، منها الطريقة الهندسية والحسابية .الطريقة الهندسية تعتمد على حساب المساحة المحصورة بين منحنى لورنز وخط العدالة.

- **بعد التحويل إلى كسور عشرية:**

x=[0, 0.20, 0.50, 0.90, 1.00], y=[0, 0.05, 0.20, 0.50, 1.00]

- **حسابB :**

B=0.005+0.0375+0.14+0.075=0.2575

- **حساب مؤشر جيني:**

G=1−2 (0.2575)=0.485 (أي48.5 48.5%)

قيمة مؤشر جيني المحسوبة هي 0.485 .هذا الرقم يقع بين 0 و1 .مؤشر 0 يرمز للمساواة التامة، و1 يرمز لعدم مساواة تامة .بما أن 0.485 أكبر من 0.2، فإن هذا يشير إلى أن التوزيع أقل عدالة ، ويعكس تباينًا ملحوظًا لكن ليس أقصى حد، مما يعني وجود درجة متوسطة إلى عالية من عدم المساواة في توزيع الدخل بين الأسر في هذه المدينة .

يُعد كل من منحنى لورنز ومؤشر جيني أدوات أساسية لقياس وتقييم التفاوت الاقتصادي. يوفر منحنى لورنز رؤية بيانية واضحة، بينما يقدم مؤشر جيني قياسًا رقميًا مُبسطًا يسهل المقارنة. يُكمل المقياسان بعضهما البعض، ويُستخدمان على نطاق واسع في الدراسات الاقتصادية والاجتماعية لفهم وتحليل توزيع الثروات والدخول.

الفصل الثامن: الأرقام القياسية

تُعد **الأرقام القياسية** أداة إحصائية حيوية تستخدم لقياس التغير النسبي في ظاهرة معينة (مثل الأسعار، الكميات، أو القيم) خلال فترة زمنية معينة أو بين مناطق جغرافية مختلفة. تُبنى فكرتها على مقارنة قيمة المتغير في "فترة المقارنة" بقيمته في "فترة الأساس"، والتي تُعتبر نقطة مرجعية ثابتة (أبو صالح وعوض، 2007). يتم التعبير عن النتائج عادة كنسبة مئوية، مما يسهل تفسيرها ومقارنتها. على سبيل المثال، إذا كان الرقم القياسي للأسعار 120، فهذا يعني أن الأسعار قد ارتفعت بنسبة 20% مقارنة بفترة الأساس.

1 الأرقام القياسية البسيطة(Simple Index Numbers) :

تُستخدم الأرقام القياسية البسيطة لقياس التغير في قيمة متغير واحد فقط، مثل سعر سلعة واحدة أو كمية إنتاجها (طيار، 2019). تُعتبر هذه الأرقام أساساً لفهم كيفية تغير قيمة عنصر فردي بمرور الوقت. يقيس **الرقم القياسي للسعر البسيط** التغير في سعر سلعة معينة بين فترتين، باستعمال الصيغة التالية:

$$\mathrm{Pn} = \left(\frac{\mathrm{P1}}{\mathrm{P0}}\right) * 100$$

مثال توضيحي :إذا كان سعر كيلوغرام من الأرز في عام 2020 هو 100 د.ج (فترة الأساس) وفي عام 2023 أصبح 120 د.ج (فترة المقارنة)، فإن الرقم القياسي للسعر يكون Pn=(120/100)*100=120: وهذا يعني أن سعر الأرز قد ارتفع بنسبة 20% في عام 2023 مقارنة بعام 2020.

2 الأرقام القياسية المجمعة(Aggregative Index Numbers)

تُستخدم هذه الأرقام لقياس التغير في مجموعة من السلع أو الخدمات .(Anderson et al., 2015) تُعتبر أداة مفيدة لإعطاء صورة عامة عن التغيرات، ولكنها قد تكون مضللة لأنها لا تأخذ في الحسبان الأهمية النسبية لكل سلعة (لوكايون ولابروس). على سبيل المثال، قد يؤدي ارتفاع سعر سلعة غير مهمة اقتصادياً إلى التأثير على الرقم القياسي المجمع بنفس قدر ارتفاع سعر سلعة أساسية، وهذا لا يعكس الواقع.

يُحسب **الرقم القياسي المجمع للأسعار** عن طريق جمع أسعار السلع في فترة المقارنة وقسمتها على مجموع أسعارها في فترة الأساس، و يحسب بالصيغة التالية:

$$\text{Pn} = \left(\frac{P1}{\text{P0}}\right) * 100$$

السلعة	سعر 2020 (P_0)	سعر 2023 (P_1)
أرز	100 د.ج	120 د.ج
سكر	80 د.ج	100 د.ج
زيت	150 د.ج	170 د.ج

- **مجموع الأسعار في عام 2020:** 100+80+150=330 $(\sum P_0)$
- **مجموع الأسعار في عام 2023: 120+100+170=390** $(\sum P_1)$
- **الرقم القياسي المجمع للأسعار :**

$$\text{Pn} = \left(\frac{390}{330}\right) * 100 = \mathbf{118.18}$$

وهذا يعني أن متوسط أسعار هذه السلع قد ارتفع **بنحو 18.18% بشكل إجمالي.**

3 الأرقام القياسية المرجحة(Weighted Index Numbers)

تُعتبر هذه الأرقام أكثر دقة وواقعية لأنها تأخذ في الاعتبار الأهمية النسبية (الوزن) لكل سلعة في المجموعة (أبو صالح وعوض، 2007). يتم تحديد الوزن بناءً على كمية الاستهلاك أو الإنتاج، مما يعكس الأهمية الاقتصادية للسلعة .

1.3. الرقم القياسي المرجح للأسعار (طريقة لاسبير-Laspeyres): يستخدم كميات فترة الأساس (Q_0) كأوزان. تُعبر هذه الطريقة عن تكلفة سلة السلع الأساسية في فترة المقارنة مقارنة بتكلفتها في فترة الأساس، و يحسب بالصيغة التالية:

$$PL = \left(\frac{\sum P1Q0}{\sum P0Q0}\right) * 100$$

2.3. الرقم القياسي المرجح للأسعار (طريقة باش– Paasche) : يستخدم كميات فترة المقارنة (Q1) كأوزان.

تُعبر هذه الطريقة عن تكلفة "سلة" السلع الجديدة في فترة المقارنة مقارنة بتكلفتها لو تم شراؤها بأسعار فترة الأساس، ويحسب بالصيغة التالية:

$$\boldsymbol{PP = \left(\frac{\sum P1Q1}{\sum P0Q1}\right) * 100}$$

4 تمرين تطبيقي:

لدينا بيانات حول أسعار وكميات ثلاث سلع أساسية في مدينة معينة خلال عامين:

السلعة	السعر 2020(P0)	الكمية 2020(Q0)	السعر 2023(P1)	الكمية 2023(Q1)
خبز	100د.ج	50وحدة	120د.ج	60وحدة
حليب	150د.ج	30وحدة	180د.ج	35وحدة
سكر	80د.ج	20وحدة	100د.ج	25وحدة

المطلوب:

1. احسب الرقم القياسي لسعر الخبز في عام 2023 و احسب الرقم القياسي لكمية الحليب في عام 2023.
2. احسب الرقم القياسي المجمع للأسعار لمجموعة السلع الثلاث.
3. احسب الرقم القياسي للأسعار باستخدام **طريقة لاسبير** ثم باستخدام **طريقة باش**.

الحل:

1. **الأرقام القياسية البسيطة:**

الرقم القياسي لسعر الخبز: $$\mathbf{Pn = \left(\frac{P1}{P0}\right) * 100 = \left(\frac{120}{100}\right) * 100 = 120}$$

التحليل :سعر الخبز ارتفع بنسبة **20%** من عام 2020 إلى عام 2023.

- **الرقم القياسي لكمية الحليب:**

$$\mathbf{Pn = \left(\frac{P1}{P0}\right) * 100 = \left(\frac{35}{30}\right) * 100 = 116.67}$$

التحليل :كمية الحليب المستهلكة زادت بنسبة **16.67%** في عام 2023 مقارنة بعام 2020.

2. **الرقم القياسي المجمع للأسعار:**

- **حساب المجاميع:**
  - مجموع الأسعار في فترة الأساس: $(\sum P0) = 100+150+80=33$
  - مجموع الأسعار في فترة المقارنة: $(\sum P1) = 120+180+100=400$

$$Pn = \left(\frac{\sum P1}{\sum P0}\right) * 100 = \left(\frac{400}{330}\right) * 100 = 121.21$$

التحليل :مجموع أسعار السلع قد ارتفع بنسبة **21.21%** بشكل عام.

3. **الأرقام القياسية المرجحة:** لتحليل أعمق، نحتاج إلى حساب قيمتي: $\sum P0Q0, \sum P1Q0, \sum P0Q1, \sum P1Q1$

السلعة	P_0Q_0	P_1Q_0	P_0Q_1	P_1Q_1
خبز	100*50=5000	120*50=6000	100*60=6000	120*60=7200
حليب	150*30=4500	180*30=5400	150*35=5250	180*35=6300
سكر	80*20=1600	100*20=2000	80*25=2000	100*25=2500
المجموع	$\sum P_0Q_0=11100$	$\sum P_1Q_0=13400$	$\sum P_0Q_1=13250$	$\sum P_1Q_1=16000$

- **الرقم القياسي بطريقة لاسبير:**

$$PL = \left(\frac{\sum P1Q0}{\sum P0Q0}\right) * 100 = \left(\frac{13400}{11100}\right) * 100 = 120.72$$

التحليل :باستخدام أوزان عام 2020، ارتفعت تكلفة السلة الاستهلاكية بنسبة **20.72%**.

- **الرقم القياسي بطريقة باش:**

$$PP = \left(\frac{\sum P1Q1}{\sum P0Q1}\right) * 100 = \left(\frac{16000}{13250}\right) * 100 = 120.75$$

التحليل :باستخدام أوزان عام 2023، ارتفعت تكلفة السلة الاستهلاكية بنسبة **20.75%**.

الفصل التاسع: الارتباط والانحدار(Correlation and Regression)

يُعد تحليل **الارتباط (Correlation)** و**الانحدار (Regression)** من أهم أدوات الإحصاء التطبيقي، حيث يتيحان دراسة العلاقة بين متغيرين أو أكثر. يُستخدم الارتباط لقياس قوة واتجاه العلاقة، بينما يُستخدم الانحدار لوصف هذه العلاقة في نموذج رياضي يمكن استخدامه للتنبؤ. هذا الفصل يوضح كيفية التعامل مع أنواع مختلفة من البيانات، سواء كانت نوعية أو كمية، لفهم العلاقات المتبادلة بينها(Hurlin, 2015) .

1 توزيعات المتغيرات ثنائية التغير(Bivariate Distributions)

تُعد دراسة **توزيعات المتغيرات ثنائية التغير** خطوة أولى وحاسمة في أي تحليل إحصائي يهدف إلى فهم العلاقة بين متغيرين. فبدلاً من تحليل كل متغير بمعزل عن الآخر، تسمح لنا هذه التوزيعات بفحص سلوكهما المشترك، مما يكشف عن أنماط وعلاقات قد لا تكون واضحة عند النظر إلى كل متغير على حدة. تُعتبر هذه التوزيعات بمثابة خريطة طريق للبيانات، حيث تُقدم نظرة عامة شاملة قبل الغوص في التفاصيل المعقدة للارتباط والانحدار.

1.1. جداول التوافق(Contingency Tables) : أداة قوية لتلخيص وتصنيف البيانات النوعية (التصنيفية). تُستخدم هذه الجداول عندما نريد دراسة كيفية توزيع المشاهدات بين فئات متغيرين مختلفين .كل خلية في هذا الجدول تمثل تقاطع فئة من المتغير الأول مع فئة من المتغير الثاني، وتحتوي على **التكرار المشترك**، وهو عدد المشاهدات التي تقع ضمن هذا التقاطع المحدد. على سبيل المثال، إذا كنا ندرس العلاقة بين الجنس ومستوى التعليم، فإن الخلية التي تتقاطع فيها فئة "ذكر" مع فئة "جامعي" ستُظهر عدد الذكور الذين يحملون شهادة جامعية. هذه التكرارات المشتركة هي اللبنة الأساسية لتحليل العلاقة بين المتغيرات النوعية (جلاطو، 2001).

مثال رقم 9-1 : من خلال دراسة حول عينة من 450 موظفاً في شركة ما، قمنا بجمع بيانات عن الجنس (ذكر/أنثى) ومستوى التعليم (جامعي/ثانوي). يمكن تنظيم هذه البيانات في جدول التوافق التالي:

الجنس / مستوى التعليم	جامعي	ثانوي	المجموع (التكرار الهامشي)
ذكر	150	100	250
أنثى	120	80	200
المجموع (التكرار الهامشي)	270	180	450

يشير الجدول إلى أن **150موظفًا جامعيًا هم من الذكور**، في حين أن توزيع حاملي الشهادات الثانوية متقارب بين الجنسين (100 ذكر مقابل 80 أنثى).

2.1. التكرارات الهامشية (Marginal Frequencies) والتكرارات الشرطية(Conditional Frequencies) : تُعد دراسة التكرارات الهامشية والشرطية خطوة أساسية في فهم العلاقة بين متغيرين كيفيين، حيث تُقدم رؤى أعمق من مجرد التكرارات المشتركة.

أ. **التكرارات الهامشية :**تُشير إلى التوزيع الإجمالي لكل متغير على حدة. تُحسب بجمع كل صف أو عمود في جدول التوافق، مما يعطينا فكرة عن التوزيع العام لكل فئة من المتغيرين. **مثال :**في دراسة حول 200 شخص، إذا كان 120 شخصًا من الذكور و80 من الإناث، فإن **التكرارات الهامشية للجنس** هي 120 (ذكور) و80 (إناث). وبالمثل، إذا كان 150 شخصًا يُفضلون القهوة و50 يُفضلون الشاي، فإن **التكرارات الهامشية للمشروب المفضل** هي 150 (قهوة) و50 (شاي).

ب. **التكرارات الشرطية :**تُقدم نظرة أكثر تفصيلاً، حيث تُوضح توزيع متغير واحد بشرط أن يأخذ المتغير الآخر قيمة معينة. تُستخدم هذه التكرارات، التي عادة ما تكون على شكل نسب مئوية، للمقارنة بين المجموعات المختلفة واكتشاف وجود أي ارتباط محتمل. **مثال :**لنفترض أن 80 من أصل 120 ذكرًا يُفضلون القهوة. فإن **التكرار الشرطي** هو نسبة الذكور الذين يُفضلون القهوة من إجمالي الذكور، أي .%66.7≈100*(80/120)وبالمثل، إذا كان 70 من أصل 80 أنثى يُفضلن القهوة، فإن النسبة هي .%87.5=100*(70/80)تُظهر هذه المقارنة أن الإناث في هذه العينة يملن إلى تفضيل القهوة أكثر من الذكور.

تُكمل التكرارات الهامشية والشرطية بعضها البعض لتقديم صورة شاملة عن التوزيعات ثنائية التغير، مما يُمهد الطريق لتحليل الارتباط بشكل أكثر دقة.

2 **الارتباط بين متغيرين كيفيين**(Association Between Qualitative Variables)

عندما نتعامل مع البيانات النوعية، مثل لون العين (أزرق، بني، أخضر) أو حالة الحالة الاجتماعية (أعزب، متزوج)، لا يمكننا استخدام معامل ارتباط بيرسون لأنه مصمم للمتغيرات الكمية. بدلاً من ذلك، نستخدم أدوات إحصائية مصممة خصيصًا لقياس قوة العلاقة بين المتغيرات الفئوية. الهدف هنا هو تحديد ما إذا كان توزيع فئات أحد المتغيرين يعتمد على توزيع فئات المتغير الآخر أم لا.

1.2. إحصاء كاي مربع($\chi 2$) : اختبار كاي مربع يُستخدم لتحليل العلاقة بين متغيرين نوعيين من خلال مقارنة التكرارات المشاهدة بالتكرارات المتوقعة. يهدف إلى تحديد ما إذا كان هناك فرق دال إحصائيًا يشير إلى وجود علاقة بين المتغيرين، وذلك وفق **الفرضيات الإحصائية التالية:**

- **الفرضية الصفرية :(H0)** لا يوجد ارتباط بين المتغيرين (أي أنهما مستقلان). على سبيل المثال، أن الجنس ومستوى التعليم مستقلان عن بعضهما البعض.
- **الفرضية البديلة :(H1)** يوجد ارتباط بين المتغيرين (أي أنهما غير مستقلين). على سبيل المثال، أن مستوى التعليم يعتمد على الجنس.

ب. **حساب إحصاء كاي مربع** :

$$\chi 2 = \frac{\sum(Oij - Eij)2}{Eij}$$

- **Oij** : هو التكرار المشاهد في الخلية (i,j)في جدول التوافق و اهي القيم الفعلية التي تم جمعها من العينة.
- **Eij**: هو التكرار المتوقع لكل خلية، والذي يُحسب بافتراض أن المتغيرين مستقلان. يتم حسابه لكل خلية بالصيغة :

$$Eij = \frac{(\text{مجموع العمود الهامشي}) \times (\text{مجموع الصف الهامشي})}{\text{المجموع الكلي}}$$

- بعد حساب Oijو Eijلكل خلية، يتم حساب الفرق بينهما وتربيعه، ثم قسمته على التكرار المتوقع. تُجمع هذه القيم للحصول على إحصاء كاي مربع .($\chi 2$) إذا كانت قيمة $\chi 2$المحسوبة كبيرة، فهذا يعني أن الفروقات بين ما لاحظناه وما توقعناه كبيرة جدًا بحيث لا يمكن أن تكون نتيجة للصدفة، مما يدفعنا إلى رفض الفرضية الصفرية وقبول وجود علاقة ارتباط.(Hurlin, 2015)

2.2. معاملات الارتباط(Correlation Coefficients) : يُظهر اختبار كاي مربع (Chi-Square Test) ما إذا كانت هناك علاقة بين متغيرين نوعيين، لكنه لا يوضح **قوة العلاقة** .لذلك، طُوّرت معاملات ارتباط مشتقة من إحصاء كاي مربع لتقدير قوة الترابط بشكل عددي.

أ. **معامل فاي(ϕ)** : يُستخدم هذا المعامل في حالة الجداول الثنائية (2×2). صيغته:

$$\varphi = \sqrt{\frac{\chi 2}{n}}$$

حيث: x^2 :قيمة إحصاء كاي مربع. n : الحجم الكلي للعينة.

تتراوح قيمة ϕ بين 0 و1؛ كلما اقتربت من 1 دلّ ذلك على علاقة قوية، وكلما اقتربت من 0 دلّ على علاقة ضعيفة.

ب. معامل التو افق (Contingency Coefficient, C) يُستخدم هذا المعامل مع الجداول الأكبر من (2×2).

صيغته:

$$C = \sqrt{\frac{\chi 2}{\chi 2 + n}}$$

حيث :x^2 : قيمة كاي مربع. / n : الحجم الكلي للعينة. تتراوح قيمة C بين 0 و1، إلا أنها لا تصل إلى 1 أبدًا.

مثال رقم 9-2: إذا حصلنا على φ =0.65 في دراسة العلاقة بين "الجنس" و"مستوى التعليم"، فهذا يشير إلى علاقة قوية نسبيًا. أما إذا كان C = 0.30 عند دراسة "الحالة الاجتماعية" و"مكان الإقامة"، فهذا يدل على علاقة أضعف.

3 **الارتباط بين متغيرين مستمرين** (Correlation Between Continuous Variables) :

عندما تكون البيانات كمية (أو مستمرة)، مثل الطول والوزن، الدخل وسنوات الخبرة، أو درجات الحرارة ومبيعات الآيس كريم، يمكننا تحليل العلاقة بينها بشكل أكثر دقة باستخدام أدوات إحصائية متقدمة. هذه الأدوات لا تحدد فقط وجود العلاقة، بل تقيس قوتها واتجاهها، وتُمكننا من بناء نماذج للتنبؤ.

1.3. سحابة النقاط(Scatter Plot) : تُعد سحابة النقاط الأداة البيانية الأساسية والأكثر أهمية في دراسة العلاقة بين متغيرين مستمرين (جلاطو، 2001)، وهي ببساطة رسم بياني يُمثل كل زوج من القيم (xi,yi) كنقطة في المستوى الإحداثي. يُمكّن هذا التمثيل من الحصول على انطباع بصري فوري حول شكل العلاقة ونوعها:

- **ارتباط خطي إيجابي** :تتجمع النقاط في شكل خط مائل يتجه من أسفل اليسار إلى أعلى اليمين. هذا يعني أن زيادة قيمة المتغير المستقل (x) يرافقها زيادة في قيمة المتغير التابع(y) .
- **ارتباط خطي سلبي** :تتجمع النقاط في شكل خط مائل يتجه من أعلى اليسار إلى أسفل اليمين. هذا يعني أن زيادة قيمة x يرافقها نقصان في قيمة y.
- **عدم وجود ارتباط** :تكون النقاط متناثرة بشكل عشوائي دون أي نمط واضح.
- **ارتباط غير خطي** :تتجمع النقاط في شكل منحني (مثل U أو J)، مما يشير إلى وجود علاقة غير خطية.

2.3. معامل الارتباط الخطي(Linear Correlation Coefficient) : بينما تُعطينا سحابة النقاط فكرة بصرية، فإن **معامل الارتباط الخطي** يُقدم قياسًا كميًا وموضوعيًا لقوة واتجاه العلاقة الخطية. أشهر هذه المعاملات هو **معامل ارتباط بيرسون(Pearson's r)** ، الذي يُعتبر المقياس القياسي للارتباط الخطي (طبيه، 2008)، وفق الصيغة التالية :

$$r = \frac{\sum(xi - x^{-})(yi - y^{-})}{\sqrt{(xi - x^{-})2 \ \ (yi - y^{-})2}}$$

- **تفسير القيمة:**
  - **r=+1** : ارتباط خطي إيجابي تام. جميع النقاط تقع على خط مستقيم واحد.
  - **r=−1** :ارتباط خطي سلبي تام. جميع النقاط تقع على خط مستقيم واحد.
  - **r=0** :لا يوجد ارتباط خطي.
  - **قيمة r بين 0 و1 (أو بين 0 و-1)** :تُشير إلى وجود ارتباط خطي، وقوة هذا الارتباط تتناسب طرديًا مع اقتراب القيمة من 1 أو -1.
  - يجب الانتباه إلى أن الارتباط لا يعني السببية؛ فوجود علاقة قوية بين متغيرين لا يعني بالضرورة أن أحدهما يسبب الآخر.

4 الانحدار الخطي البسيط(Simple Linear Regression) :

يُعد **الانحدار الخطي البسيط** خطوة أعمق من الارتباط. بينما يصف الارتباط قوة العلاقة، فإن الانحدار يصف العلاقة في نموذج رياضي يمكن استخدامه للتنبؤ. يُستخدم لوصف العلاقة الخطية بين متغير مستقل واحد **(X)**ومتغير تابع واحد **(Y)** (Doane & Seward, 2016) .

1.4. معادلة خط الانحدار :الهدف هو إيجاد أفضل خط مستقيم يمر عبر سحابة النقاط، ويُعبر عنه بالمعادلة :

$$\mathrm{y} = a + \mathrm{bx}$$

- **y^** : هي القيمة المتوقعة للمتغير التابع، وليست القيمة الفعلية.
- **a** : يُعرف بـ **الجزء المقطوع** (intercept)، ويمثل القيمة المتوقعة لـ y^عندما تكون قيمة xتساوي صفرًا.
- **b** : يُعرف بـ **ميل الخط** (slope)، ويمثل مقدار التغير المتوقع في y لكل وحدة زيادة فيX.
- **x** : المتغير المستقل الذي نستخدمه للتنبؤ.

2.4. حساب معاملات الانحدار (bوa) يتم حساب هذه المعاملات باستخدام طريقة المربعات الصغرى لتقليل المسافة العمودية بين النقاط وخط الانحدار.

- **ميل الخط(b)** : $b = \frac{\sum(x^{-}-xi)(y^{-}-yi)}{(x^{-}-xi)2}$
- **الجزء المقطوع(a)**: $\mathbf{a = y^{-} - bx^{-}}$

بمجرد حساب هذه المعاملات، يمكن استخدام المعادلة للتنبؤ بقيمة yلأي قيمة معطاة لـ x. على سبيل المثال، إذا كانت المعادلة y^=2+5x، فهذا يعني أن كل زيادة بوحدة واحدة في xستؤدي إلى زيادة بـ 5 وحدات في y، وأن قيمة y المتوقعة ستكون 2 عندما يكون x=0.

5 **تمرين تطبيقي:** لدينا بيانات عن **الدخل المتاح** (x) و **الإنفاق الاستهلاكي** (y) لعائلة خلال 10 أشهر (الوحدة= آلاف دينار جزائري).

الشهر	1	2	3	4	5	6	7	8	9	10
X (دخل، أ.دج)	5	8	10	3	6	9	4	7	11	2
Y (إنفاق، أ.دج)	75	85	90	60	80	88	70	82	95	55

المطلوب

1. حساب **معامل الارتباط الخطي (بيرسون)** بين xو y.
2. إيجاد **معادلة الانحدار الخطي البسيط** (y^=a+bx).
3. توقع ^y عندما يكون x =7000.دج

<u>الحل:</u>

1- حساب المتوسطات: $y = \frac{780}{10} = 78$ $X = \frac{65}{10} = 6.5$

2- **المجاميع اللازمة** (باستخدام جدول الانحرافات)

- ∑ (xi−x⁻)(yi−y⁻)=346.0 / ∑(xi−x⁻)2=82.5 ∑(yi−y⁻)2=1528

3- معامل الارتباط (بيرسون):

$$r = \frac{\sum(xi - x^{-})(yi - y^{-})}{\sqrt{(xi - x^{-})2 \ \ (yi - y^{-})2}} = \frac{346}{\sqrt{1528 * \ 82.5}} = 0.9745$$

4- معادلة الانحدار البسيط:

- الميل $b = \frac{(x^{-} - xi)(y^{-} - yi)}{(x^{-} - xi)2} = \frac{346}{82.5} \approx 4.194$
- الجزء المقطوع(a): $a = y^{-} - bx^{-} = 78.0 - 4.1939 \times 6.5 \approx 50.7394$
- المعادلة: $y = 50.7394 + 4.1939x$

5- التنبؤ عندما x=7:

$$y = 50.7394 + 4.1939(7) \approx 80.10$$

النتيجة :عند دخل مقداره 7 آلاف دينار جزائري، من المتوقع أن يكون الإنفاق الاستهلاكي حوالي 80.101 **دينار جزائري**.

الخاتمة

تعتبر هذه المطبوعة رحلة معرفية عميقة في عالم الإحصاء الوصفي، حيث تتجاوز مجرد سرد الأرقام لتقدم فهمًا شموليًا لكيفية تحليل البيانات واستخلاص الرؤى منها. لقد بدأنا المسار بالتعرف على الإحصاء كعلم حيوي لا غنى عنه في مختلف المجالات، خاصة الاقتصاد وإدارة الأعمال، حيث يساعد في اتخاذ القرارات الاستراتيجية. لم يقتصر الأمر على الجانب النظري، بل امتد إلى التطبيق العملي من خلال تعلم أساليب جمع البيانات وتصنيفها، وكيفية عرضها بأسلوب منهجي ومبسط سواء في جداول تكرارية دقيقة أو رسوم بيانية جذابة تسهل فهم الظواهر المعقدة.

انتقلت المطبوعة بعد ذلك إلى صلب الإحصاء الوصفي، وهو فن تلخيص البيانات عبر مقاييس متنوعة من خلال تعلم كيفية تحديد **مقاييس النزعة المركزية** مثل المتوسط الحسابي، والوسيط، والمنوال، القيمة التي تتركز حولها البيانات، مما يعطينا فكرة سريعة عن متوسط الأداء أو القيمة الأكثر شيوعًا. وفي المقابل، سلطت **مقاييس التشتت** الضوء على مدى انتشار البيانات وتباعدها عن بعضها البعض، مما يضيف عمقًا للتحليل، فليس كافيًا معرفة المتوسط فقط، بل يجب أيضًا فهم مدى التجانس أو التباين بين القيم.

لم تكتف المطبوعة بذلك، بل قدمت أدوات تحليلية متقدمة لتوفير رؤية أعمق، مثل **مقاييس الشكل** (الالتواء والتفرطح) التي تصف لنا شكل التوزيع الاحتمالي للبيانات، مما يساعد في تحديد ما إذا كانت البيانات متماثلة أو منحرفة، وهل قمتها حادة أم مسطحة. وفي فصول لاحقة، تم استكشاف **مقاييس التمركز** مثل منحنى لورنز ومؤشر جيني، التي تعد ضرورية في العلوم الاجتماعية والاقتصادية لقياس مدى عدالة توزيع الدخل أو الثروة. كما تم التطرق إلى **الأرقام القياسية** كأداة حاسمة لقياس التغير النسبي للظواهر عبر الزمن، مثل التضخم أو نمو الإنتاج، مما يسهل عملية المقارنة الزمنية. وأخيرًا، اختتمت المطبوعة بالتعرف على **الارتباط والانحدار**، وهما من أهم الأدوات التي تسمح لنا بدراسة طبيعة وقوة العلاقة بين متغيرين أو أكثر، مما يمهد الطريق للتنبؤ بالاتجاهات المستقبلية.

في الختام، فإن هذه المطبوعة لم تكن مجرد دليل إحصائي، بل كانت خريطة طريق متكاملة، مكنت الطالب من اكتساب المهارات اللازمة لتحليل البيانات من جميع جو انبها. لقد تحولت البيانات الخام بفضل هذه الأدوات إلى معلومات قيمة، وأصبح الطالب قادرًا على استخلاص استنتاجات منطقية وموضوعية، وهو ما يمثل جوهر الفكر الإحصائي وضرورة حقيقية في عالم يعتمد بشكل متزايد على البيانات.

قائمة المراجع

المراجع باللغة العربية

1. **أبو صالح، كمال وعوض، كمال (2007) .** *مدخل إلى علم الإحصاء* .بيروت: دار الحكمة للنشر والتوزيع.
2. **تيلوت، علي. (2009).** *مبادئ الإحصاء الوصفي والاستدلالي* .الجزائر: دار العلوم للنشر والتوزيع.
3. **حمداني، رياض. (2005).** *الإحصاء التطبيقي* .دمشق: دار الفكر.
4. **راتول، محمد. (2006).** *أساسيات الإحصاء* .القاهرة: مكتبة الأنجلو المصرية.
5. **السيد، يوسف. (2020).** *التحليل الإحصائي للبيانات الاجتماعية* .عمان: دار المسيرة للنشر.
6. **عبد الفتاح، محمود. (2012).** *الإحصاء التطبيقي للمبتدئين* .الرياض: دار عالم الكتب.
7. **العاني، عبد الرحمن وآخرون. (2017).** *مناهج البحث العلمي* .عمان: دار الكتب العلمية.
8. **عزوز، كريم. (2010).** *الإحصاء في العلوم الاقتصادية* .الجزائر: ديوان المطبوعات الجامعية.
9. **موسى، خالد. (2016).** *أساليب جمع البيانات وتحليلها* .بيروت: دار الحافظ.
10. **طبيّه، خالد. (2008).** *المدخل إلى الإحصاء* .دمشق: دار الحكمة.
11. **طيار، جمال. (2019).***تحليل البيانات الكمية* .الجزائر: دار الخلدون للنشر

المراجع باللغة الأجنبية

- **Hurlin, C. (2015).** *Introduction to Econometrics*. Paris: Hachette.
- **Le Hay, F. & Chanvril-Ligneel, A. (2014).** *Statistiques Descriptives*. Bruxelles: De Boeck.
- **Milanovic, A. (2016).** *Income Inequality and the Gini Index*. New York: Penguin Press.