وزارة التعليم العالي و البحث العلمي
جامعة الجزائر3
كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université d' Alger 3
Faculté des Sciences Economiques et sciences commerciales et Sciences de Gestion

مطبوعة موجهة لطلبة السنة أولى جذع مشترك

الإحصـــــاء1

دروس وتمارين تطبيقية

إعـداد: د.هــمـال فـــريــدة

السنة الجامعية: 2019-2020

مقدمة

يعد علم الإحصاء من العلوم الضرورية في أي بحث علمي أو دراسة تطبيقية، تهدف إلى الوصول إلى نتائج تعتمد الموضوعية وتتسم بالمصداقية. وتكمن أهميته في استخدامه في شتى المجالات فهو وسيلة يستخدمها معظم الناس في حياتهم وأعمالهم اليومية خاصة متخذي القرار في الإدارات والمؤسسات، ذلك أنه أي قرار لا بد أن يبنى على بيانات ومعلومات دقيقة من أجل الوصول إلى نتائج موضوعية عن الظاهرة محل الدراسة، وهو ما جعله يحظى بالاهتمام الأكاديمي على مستوى المعاهد والجامعات وفي جميع التخصصات تقريبا.

يأتي الإحصاء1 (الإحصاء الوصفي) كمدخل لعلم الإحصاء ويتم تدريسه لطلبة السنة أولى نظرا لاعتماده على مداخل أولية تتعلق بطريقة وصف الظواهر بأساليب إحصائية وصفية في شكل أرقام ورسومات بيانية وجداول توضيحية.

وسعيا منا إلى تقديم ما هو مقرر دراسته في مقياس الإحصاء1 لطلبة العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، قصد لإلمام الطالب لهذا المقياس ارتأينا وضع بين أيديكم مطبوعة الإحصاء1 والتي أتت كثمرة جهد سنوات عديدة في تدريس هذا المقياس، ولقد حرصت في تقديمها على البساطة في السرد والمنهجية في العرض.

وهي عبارة عن سلسلة محاضرات تقدم من خلالها دروس مبسطة ومختصرة مدعمة بالعديد من الأمثلة والتمارين معروضة بحلول نموذجية بالإضافة إلى نماذج امتحانات السنوات السابقة، من خلال ستة فصول أساسية معروفة مرتبة ومتسلسلة وفق مقرر وزارة التعليم العالي والبحث العلمي.

فهرس المحتويات

الفصل الأول

مفاهيم ومصطلحات أساسية

I – تعريف علم الإحصاء:

الإحصاء فرع من فروع الرياضيات يهتم بدراسة الظواهر عن طريق مشاهدة تطوراتها وتفاعلاتها مع بعضها البعض بغية تفسيرها والتنبؤ بها مستقبلا. إن حياتنا اليومية مرتبطة بالإحصاء في العديد من المجالات سواء على مستوى الأفراد والجماعات أو على مستوى الدولة في العديد من الظواهر أو المتغيرات التي نعيشها، منها ما هو ذات طابع اقتصادي أو اجتماعي أو تعليمي أو تربوي أو صحي أو نفسي. فالإحصاء هو علم جمع وترتيب المعلومات الخاصة بظاهرة معينة وعرضها وتحليل معطياتها، ثم التنبؤ بها في المستقبل وفق شروط معينة قصد اتخاذ القرار المناسب.

نتطرق في هذا المقياس إلى الإحصاء الوصفي الذي يعالج الطريقة الإحصائية الوصفية، وهذه الأخيرة مفيدة وضرورية في فحص ودراسة أنواع كثيرة من المشاكل المختلفة ومنها على سبيل المثال:

– اختبار تأثير فعالية دواء معين.

– تأثير السمنة على زيادة احتمال الإصابة بأمراض القلب.

– مراقبة الإنتاج.

–تأثير فائض الإنتاج على مستوى الأرباح.

II– خطوات الطريقة الإحصائية:

تستعمل الطريقة الإحصائية في كل المجالات والتخصصات، في العلوم الاجتماعية والعلوم الاقتصادية .... إلخ وعلى هذا تهدف الطريقة الإحصائية إلى:

- جمع البيانات والمعلومات الخاصة بمختلف الظواهر وتسجيلها في صورة رقمية وتصنيفها في جداول منظمة وتمثيلها بيانيا (وهذا هو الإحصاء الوصفي).
- تحليل البيانات واستخلاص النتائج منها واتخاذ القرارات (الإحصاء الاستدلالي).
- ثم مقارنة الظواهر ببعضها البعض ودراسة العلاقة بينها واستخدامها في فهم حقيقة الظواهر ومعرفة الظواهر التي تسير تبعا لها.

وتتمثل الطريقة الإحصائية في المراحل التالية:

–تحديد الظاهرة المدروسة.

–تحديد المجتمع المدروس.

-جمع المعلومات الإحصائية.

-فرز البيانات.

-العرض الجدولي والبياني للبيانات.

-عرض النتائج وتحليلها وتفسيرها.

وعليه فإن الإحصاء عبارة عن مجموعة الطرق العلمية التي تسمح بجمع البيانات المتعلقة بظاهرة معينة وتبويبها في جداول إحصائية وعرضها في صورة أشكال بيانية وتحليلها باستخدام مقاييس النزعة المركزية ومقاييس التشتت وغيرها من المقاييس الأخرى. حيث يكمن الهدف من التطرق إلى هذه المحاور في تمكين الطالب من التعامل مع الدراسة الإحصائية بسهولة وبساطة، ذلك أنه يحتاج إليها مهما كان تخصصه.

III-المصطلحات الإحصائية:

III-1- الوحدة الإحصائية: هي الوحدة الأساسية لتكوين المجتمع الإحصائي.

III-2- المجتمع الإحصائي: المجتمع هو ليس تلك الكلمة المتداولة، بل المجتمع في قاموس الإحصاء هو عبارة عن مجموعة من المشاهدات و القياسات الخاصة بمجموعة من المفردات تسمى هذه المفردات وحدات إحصائية. إذا هو عبارة عن مجموعة العناصر أو الأفراد التي ينصب عليها الاهتمام في دراسة معينة، بتعريف آخر هو جميع العناصر التي تتعلق بموضوع البحث، وقد يكون مجتمع الدراسة طلاب جامعة معينة، سكان إقليم معين، ....إلخ.

III-3- العينة الإحصائية: هي مجموعة جزئية من المجتمع محل الدراسة، عادة نجري الدراسة على العينة لصعوبة إجرائها على المجتمع ككل.

III-4- الخاصية المدروسة: (الميزة) هي الظاهرة التي يرغب الباحث دراستها أو هي القاسم المشترك بين عناصر المجتمع أو العينة. مثال: أطوال الطلبة. وهي تنقسم إلى نوعين:

III-4-1- الخاصية الكمية: عندما تأخذ الخاصية قيما عددية بمعنى أنها قابلة للعد، عندها نكون بصدد دراسة متغير كمي وهذا الأخير يمكن أن يكون منقطع أو مستمر.

III-4-2- الخاصية الكيفية (نوعية): عندما لا تأخذ الخاصية قيما عددية أي أنها غير قابلة للعد، عندها نكون بصدد دراسة متغير كيفي.

III-5-المتغير الإحصائي: يوضح مختلف القيم التي يمكن أن تأخذها الخاصية المدروسة وهو نوعان: المتغير الإحصائي المنقطع والمتغير الإحصائي المستمر في حالة الخاصية الكمية، وتأخذ توصيفات ورموز في حالة المتغير الإحصائي النوعي.

IV– أنواع البيانات الإحصائية:

كلما كان جمع البيانات دقيقا كلما زادت ثقة الباحث في الاعتماد عليها، كما أنه لا يعتبر تحليل البيانات صحيحا إذا كان هناك أخطاء في جمع البيانات. وهناك نوعان من البيانات، وذلك حسب الظاهرة المدروسة ، كيفية غير قابلة للقياس وكمية قابلة للقياس:

IV – 1– بيانات نوعية: (كيفية أو وصفية)

نحصل على مثل هذه البيانات عندما تكون الخاصية المدروسة هي خاصية نوعية والتي يمكن تصنيفها حسب الأصناف أو الأنواع وليس بقيم عددية. مثال على ذلك تصنيف حسب الجنس، حسب الحالة الاجتماعية، الحالة الاجتماعية المهنية ....الخ. وتستخدم عدة مقاييس لقياس البيانات النوعية منها:

IV –1 –1– التدرج الاسمي: (التصنيف الاسمي)

يصنف هذا المقياس عناصر الظاهرة المدروسة التي تختلف في النوعية لا الكمية، بحيث يستخدم الأعداد لتحديد هوية المفردات وفي هذه الحالة لا يكون للعدد ذلك المدلول الكمي الذي يفهم منه عادة. مثلا نستعمل العددين 0 و 1 للدلالة على التصنيف حسب الجنس، حيث نرمز بـ0 للذكر و 1 للأنثى. لاحظ أن 0 و 1 لا يدلان على قيم عددية

IV – 1–2– التدرج الترتيبي: (التصنيف الترتيبي)

يقع هذا التصنيف على مستوى أعلى من التصنيف الاسمي، فبالإضافة إلى خواص التدرج الاسمي يسمح التصنيف الترتيبي بالمفاضلة، أي بترتيب العناصر حسب سلم معين. مثلا تقديرات الطلبة ترتب من ممتاز إلى ضعيف، أو درجات التأييد إجابة على سؤال ترتب من موافق جدا، موافق، متردد، لا أوافق.

IV – 2– بيانات كمية: (عددية)

نتحصل عليها عندما تكون الخاصية (الصفة) قيد الدراسة قابلة للقياس أو العد. مثلا علامات الطلاب في امتحان ما، كميات السلع والبضائع المنتجة أو المستوردة، عدد العمال في مصنع ما..... إلى غيرها من الأمثلة. وهذه البيانات يمكن أن تكون بيانات كمية منقطعة وبيانات كمية مستمرة، وذلك حسب نوع الخاصية المدروسة.

IV – 2–1– الكمي المنفصل :يأخذ قيمة واحدة التي لا يمكن تجزئتها، أي أنها أعداد صحيحة.

IV – 2–2– الكمي المتصل : يأخذ عدة قيم يمكن تجزئتها و وضعها في مجال.

V – طرق جمع البيانات: هناك عدة طرق لجمع البيانات ، أشهرها:

V – 1– المقابلة الشخصية: هي مقابلة أفراد العينة والتحدث إليهم عن موضوع الدراسة.

V – 2– الملاحظة المباشرة: تستعمل هذه الطريقة عندما يصعب علينا تحديد أفراد العينة، مثلا دراسة كثافة حركة السير في مفترق الطريق، فإنك تقوم بتعداد السيارات التي تمر منه في مدة زمنية معينة (وقت الذروة مثلا) أو أن تقوم بمراقبة تصرف مجموعة من الأطفال أثناء اللعب وتدوين الملاحظات بهدف التعرف على سلوكيات الأطفال في بعض المواقف.

V – 3– الاستبيان: هو وسيلة لجمع البيانات اللازمة للتحقق من فرضيات المشكلة قيد الدراسة، وذلك بالإجابة على الأسئلة المطروحة في الاستبيان.

الفصل الثاني

العرض الجدولي للبيانات الإحصائية

بعد عملية جمع البيانات الإحصائية وفرزها تصبح لدينا مشاهدات على شكل معطيات أو بيانات منفردة أو غير مبوبة، يكون عددها كبير جدا يصعب علينا استيعابها أو استخلاص النتائج منها ما يجعلنا في أمس الحاجة إلى تنظيمها وتبويبها حتى نتمكن من التعامل معها ودراستها والاستفادة منها، فنستعين بعملية تنظيم البيانات في جداول إحصائية وتصنيفها إلى مجموعات متجانسة وتبويبها في جداول تلخيصية.

يقصد بالعرض الجدولي للبيانات الإحصائية وضع البيانات المتحصل على الظاهرة محل الدراسة في جداول خاصة تسمى بجداول التوزيعات التكرارية. يتكون الجدول التكراري من عمودين أساسيين (سطرين) يوضع في العمود الأول (السطر الأول) قيم الخاصية المدروسة والتي نرمز لها بالرمز X_i ، وهي تعبر عن قيم المتغير الإحصائي المدروس، وتكون هذه القيم على شكل صفات، قيم نقطية أو على شكل مجالات. أما العمود الثاني (السطر الثاني) فيخص عدد الوحدات الإحصائية المقابلة لكل قيمة من قيم المتغير الإحصائي ويطلق على عدد الوحدات الإحصائية بالتكرار ونرمز له بالرمز n_i.

I–جداول التوزيعات التكرارية (الجداول الإحصائية):

تختلف الجداول الإحصائية باختلاف نوع البيانات أي حسب نوع المتغير الإحصائي، بين المتغير الإحصائي الكيفي والكمي وإذا كان كمي بين المتغير المنفصل والمتغير المتصل. إذا لا بد من تحديد أولا نوع المتغير قبل تبويب البيانات.

I –1–الجدول التكراري لمتغير كيفي:

يتكون الجدول التكراري لمتغير كيفي من عمودين (أو سطرين)، يحتوي العمود الأول على رموز كتابية للخاصية المدروسة ويحتوي الثاني على تكرارات كل رمز كتابي.

مثال (1–1): أخذت عينة إحصائية تتكون من 200 شخص مغترب في بلد أوروبي لمعرفة جنسية كل مغترب، فتحصلنا على النتائج المدونة في الجدول الإحصائي التالي:

الجدول رقم (1–1): توزيع المغتربين حسب الجنسية

الجنسية (X_i)	تونسية	مغربية	جزائرية	خليجية	مصرية	برتغالية	نيجيرية	المجموع
التكرار (n_i)	40	30	35	25	20	15	35	200

I–2–الجدول التكراري لمتغير كمي منفصل: (منقطع)

المتغير الكمي المنفصل أو المنقطع هو ذلك المتغير الذي لا يمكن تجزئة قيمه، حيث يأخذ قيم صحيحة.

مثال (1–2): لدراسة متوسط عدد أطفال في العائلة لبلد ما، أخذت عينة من 100 أسرة فكانت النتائج كما يلي:

الجدول رقم (1–2): عدد الأطفال في الأسرة

عدد الأطفال (X_i)	0	1	2	3	4	5	6	مجموع
التكرار (n_i)	10	12	16	18	22	14	8	100

I –3–الجدول التكراري لمتغير كمي متصل (مستمر):

المتغير الكمي المتصل هو ذلك المتغير الذي يمكن تجزئة قيمه ووضعها في مجالات (فئات). حيث أن قيمه توضع على شكل مجالات متساوية أو غير متساوية وتسمى فئات، ولكل فئة حد أدنى وحد أعلى.

مثال (1–3): لدراسة ظاهرة تأخر وصول الطلبة إلى مقاعد الدراسة أخذنا عينة من 150 طالب، فكانت النتائج كما يلي:

الجدول رقم (1–3): ظاهرة تأخر وصول الطلبة إلى مقاعد الدراسة

الفئات	$]5-0]$	$]10-5]$	$]15-10]$	$]20-15]$	$]25-20]$	$]30-25]$	مجموع
التكرار (n_i)	10	20	35	55	20	10	150

ويحدد عدد هذه الفئات حسب حجم العينة وحسب توزيع الوحدات الإحصائية في مجال الدراسة.

1–3–1– تحديد الفئات: لتحديد هذه الفئات أو هذه المجالات وضع الإحصائي ستورجس (staurges) قاعدة تجريبية لتحديد طول الفئات وتعتمد هذه القاعدة على مجال الدراسة (المدى العام) وحجم المجتمع أو العينة. وتعطى بالعبارتين التاليتين وذلك حسب اللوغاريتم المستخدم النيبيري أو العشري.

$$K = \frac{E}{1 + 1.32Ln(n)}$$

أو

$$K = \frac{E}{1 + 3.32Log(n)}$$

حيث :

n: حجم العينة أو المجتمع؛ E: المدى العام وهو الفرق بين أكبر قيمة و أصغر قيمة $E = X_n - X_1$

K: طول الفئة ، والعبارة $1 + 1.32Ln(n)$ تمثل عدد الفئات.

ملاحظة: لا يوجد طول متفق عليه لهذه الفئة كما انه لا يوجد عدد متفق عليه من الفئات لتوزيع إحصائي معين ,ولكن يتحدد عدد الفئات وطولها حسب أهمية الدراسة وطول مجال التعريف وحجم العينة أو المجتمع. وللحفاظ على توازن جدول التكرارات من الأفضل أن يتراوح عدد الفئات من 9 إلى 25 فئة.كما أنه يستحسن أن تكون أطوال الفئات متساوية وذلك حتى يمكن مقارنة الظاهرة من مجال إلى مجال آخر. كما يتم تحديد طول الفئة بالعلاقة التالية:

$$K = \frac{E}{L}$$

حيث L يمثل عدد الفئات، ويجب مراعاة المتباينة التالية في تحديد طول الفئة: $E \leq L \times K$

ويمكن تطبيق معادلة يول (YULE) لحساب عدد الفئات بالعلاقة التالية: $L = 2.5\sqrt[4]{n}$

1-3-2- تحديد مراكز الفئات: في حالة تكوين جدول توزيع تكراري لمتغير كمي متصل تضيع القيم الإحصائية الأصلية للمشاهدات ونصبح لا نعرف عنها شيء سوى أنها تنتمي إلى فئة معينة محدودة بحدين معلومين، ولتخطي هذه المشكلة نقوم باستخراج مركز الفئة والذي نحصل عليه بحساب الصيغة التالية:

$$x_i = \frac{x_{min} + x_{max}}{2} = \frac{\text{الحد الأعلى للفئة} + \text{الحدى الأدنى للفئة}}{2}$$

1-3-3- تحديد حدود الفئات: تبدأ حدود الفئة الأولى بوضع أقل قيمة في البيانات في الحد الأدنى ثم يضاف إليها طول الفئة لتحديد الحد الأعلى، ونستمر بهذه الطريقة إلى غاية الوصول إلى الفئة الأخيرة يكون فيها الحد الأعلى لهذه الفئة عبارة عن القيمة العظمى في البيانات.

II-جداول التكرارات التجميعية: نميز هنا بين ثلاثة أنواع من التكرارات

II -1-التكرار المطلق: وهو التكرار العادي، أي عدد المشاهدات الخاصة بكل قيمة التي يمكن أن يأخذها المتغير الإحصائي.

II -2-التكرار النسبي: وهو عبارة عن نسب مئوية نتحصل عليها بقسمة التكرار العادي على مجموع التكرارات $f_{\mathrm{i}} = {n_{\mathrm{i}}}/{\sum \mathrm{n_i}}$. مع العلم أن $\sum f_i = 1$. ويمكن تحويل التكرار النسبي إلى تكرار نسبي مئوي وذلك بضرب التكرار النسبي في 100، ويرمز له بـ $f_{\mathrm{i}}\%$.

يفيد التكرار النسبي في تقليص الشكل البياني في حال كان عدد القيم كبيرا، بينما التكرار النسبي المئوي يفيد في إظهار الشكل عندما يكون عدد القيم صغيرا.

II -3-التكرارات التجميعية: يستخدم هذا النوع من الجداول عندما نحتاج إلى معلومات إضافية عن البيانات، فمثلا قد نحتاج إلى معرفة عدد المفردات أو المشاهدات التي تقل قيمتها أو تزيد عن حد معين. فنقوم لذلك بإيجاد التكرارات التجميعية الصاعدة والنازلة. ومنه تنقسم هذه التكرارات إلى قسمين:

II -3-1- التكرارات التجميعية الصاعدة: وهي عبارة عن تكرار فئة مضافا إليه مجموع تكرارات الفئات السابقة. وهو يمثل مجموع الأفراد الذين تقل قيمتهم الإحصائية عن الحد الأعلى للفئة المقابلة.

II -3-2- التكرارات التجميعية النازلة: عبارة عن مجموع التكرارات ($\sum n_{\mathrm{i}}$) مطروح منه تكرارات الفئات السابقة، ويمثل مجموع الأفراد الذين تزيد قيمتهم الإحصائية عن الحد الأدنى للفئة المقابلة.

تمارين محلولة:

التمرين الأول:

إذا كان لدينا بيانات الحالة المدنية لـ20 شخص في بلدية ما كما يلي:

M	M	C	D	D	C	M	C	D	M
C	D	C	C	C	M	C	M	V	V

حيث:

(C) أعزب، (M) متزوج، (D) مطلق، (V) أرمل.

المطلوب:

1-تحديد الخاصية المدروسة، المجتمع المدروس، الوحدة الإحصائية، ونوع المتغير المدروس.

2-أعد ترتيب هذه البيانات في جدول توزيع تكراري.

الحل:

الخاصية المدروسة: الحالة المدنية للأشخاص. المجتمع المدروس: 20 شخص قاطن في بلدية معينة. الوحدة الإحصائية: شخص. المتغير المدروس: متغير كيفي (نوعي) غير قابل للترتيب.

X_i	n_i	f_i	$f_i\%$
C	8	0,40	40
M	6	0,30	30
D	4	0,20	20
V	2	0,10	10
$\sum$	20	1	100

التمرين الثاني:

ليكن لدينا بيانات حول الحالة الاجتماعية المهنية في مؤسسة معينة و المدونة في الجدول التالي:

Cs	C	C	Ab	Cs	Ab	C	Ab	C	Cs
Ab	Ab	Ab	Ap	Ab	Ap	Ab	Ab	Ap	Ab
C	Ap	Cs	Ab	C	S	Cs	C	Ab	C
Ab	Ap	S	C	Ab	Ap	Ab	S	Ab	Ap
Ab	C	Ap	Ab	Cs	S	C	Ab	C	Ab

حيث:

(Cs) إطار سامي، (C) إطار، (Ab) عون مكتب (Ap) عون متعدد الخدمات، (S) أمين مكتب.

المطلوب:

1-تحديد الخاصية المدروسة، المجتمع المدروس، الوحدة الإحصائية، ونوع المتغير المدروس.

2-أعد ترتيب هذه البيانات في جدول توزيع تكراري.

الحل:

الخاصية المدروسة: الحالة الاجتماعية المهنية لعمال مؤسسة. المجتمع المدروس: 50 عامل.

الوحدة الإحصائية: عامل.

المتغير المدروس: متغير كيفي (نوعي) قابل للترتيب.

X_i	n_i	f_i	$f_i\%$	$F^{\uparrow}$
Cs	6	0,12	12	6
C	12	0,24	24	18
Ab	20	0,40	40	38
Ap	8	0,16	16	46
S	4	0,08	8	50
$\sum$	50	1	100	/

التمرين الثالث:

ليكن لدينا بيانات إحصائية متكون من 50 أسرة في حي سكني، بحيث المتغير المدروس يمثل عدد أفراد هذه الأسر. البيانات معطاة في الجدول التالي:

1	3	1	3	3	2	4	5	3	2
4	3	4	5	1	3	8	1	3	4
8	2	2	3	3	6	2	3	4	6
3	5	3	2	4	4	1	5	2	3
4	2	5	4	3	3	6	2	5	4

المطلوب:

1- حدد نوع المتغير المدروس.

2- بوب هذه البيانات في جدول توزيع تكراري مناسب ثم أحسب كل التكرارات المطلقة والنسبية.

الحل:

نوع المتغير المدروس هو متغير كمي منفصل لأنه يمثل عدد الأفراد.

يحتوي الجدول على:

- التكرارات العادية والتكرارات التجميعية الصاعدة والنازلة على التوالي؛

-التكرارات النسبية والتكرارات التجميعية الصاعدة والنازلة النسبية على التوالي؛

-التكرار النسبي المئوية.

X_i	n_i	$F^{\uparrow}$	$F^{\downarrow}$	f_i	$f_i^{\uparrow}$	$f_i^{\downarrow}$	$f_i\%$
1	5	5	50	0,1	0,1	1	10
2	9	14	45	0,18	0,28	0,9	18
3	15	29	36	0,3	0,58	0,72	30
4	10	39	21	0,2	0,78	0,42	20
5	6	45	11	0,12	0,9	0,22	12
6	3	48	5	0,06	0,96	0,1	6
8	2	50	2	0,04	1	0,04	4
$\sum$	50	/	/	1	/	/	100

التمرين الرابع:

تمثل البيانات أدناه علامات 64 طالبا من طلبة السنة الأولى علوم اقتصادية وتجارية وعلوم التسيير بجامعة الجزائر3، في مقياس الإحصاء1.

19	08	06	20	07	17	08	09
19	08	08	18	15	11	08	08
11	10	07	11	11	06	10	07
15	10	12	11	10	07	10	12
16	15	15	06	07	18	15	15
19	07	17	20	07	19	07	17
20	15	11	08	14	14	15	11
19	12	06	10	12	13	12	06

المطلوب:

-تشكيل جدول التوزيع التكراري المناسب.

الحل:

1-تحديد طول الفئة:

$$K = \frac{E}{1 + 1.32Ln(n)}$$

$$= \frac{20-6}{1+1.32\ln(64)} = 2$$

2-كتابة حدود الفئات:

نبدأ كتابة حدود الفئة الأولى في جدول التوزيع التكراري من أصغر قيمة في البيانات (06) وهي تمثل الحد الأدنى للفئة الأولى ونضيف لها طول الفئة (k=2) لنتحصل على الحد الأعلى للفئة الأولى (8)، وبالتالي الفئة الأولى هي$[6-8[$:، وهكذا مع الفئة الموالية حيث يصبح حدها الأدنى هو الحد الأعلى للفئة السابقة (8) وحدها الأعلى (10)، وبإضافة كل مرة طول الفئة نحصل على فئة جديدة الى أن نكمل عدد الفئات ونغطي جميع البيانات الموجودة.

3-حساب مراكز الفئات:

نقوم بحساب مركز الفئة (x_i) لكل فئة موجودة في جدول التوزيع التكراري، فمثلا الفئة رقم واحد مركزها (x_1) ويحسب بالعلاقة التالية:

$$x_1 = \frac{6+8}{2} = 7$$

وبنفس الطريقة تحسب مراكز الفئات الأخرى وتوضع في جدول التوزيع التكراري.

classes	x_{i}	n_i	f_i	$F^{\uparrow}$	$F^{\downarrow}$	$f_i\%$
6-8	7	14	0.22	14	64	22
8-10	9	07	0.11	21	50	11
10-12	11	13	0.20	34	43	20
12-14	13	06	0.09	40	30	09
14-16	15	10	0.16	50	24	16
16-18	17	04	0.06	54	14	06
18-20	19	10	0.16	64	10	16
$\sum$	/	64	1	/	/	100

الفصل الثالث

العرض البياني للبيانات الإحصائية

بعد التطرق إلى العرض الجدولي للبيانات، ننتقل إلى نوع آخر لتمثيل البيانات الإحصائية وهو العرض البياني. يعطينا العرض البياني صورة عن تطور الظاهرة المدروسة في الزمان والمكان في وقت وجيز، وذلك عن طريق وصف وتلخيص البيانات باستخدام الرسومات البيانية والأشكال الهندسية تمكننا من تحليل سريع للظاهرة المدروسة. وهنا نميز بين نوعين من العروض البيانية، عرض بياني خاص بالمتغير الكيفي وعرض بياني خاص بالمتغير الكمي.

I- العرض البياني الخاص بالمتغير الكيفي: نستعرض أهم العروض البيانية الخاصة بالمتغير الكيفي وهي:

I-1- العرض الدائري: يتمثل في دائرة مقسمة إلى عدة أجزاء نسبيا مع مكونات الخاصية المدروسة، حيث يقابل كل جزء من هذه الدائرة زاوية مركزية تتناسب مع التكرار المقابل لكل خاصية. ويمكن أن نظيف عمود ثالث لجدول المعطيات نضع فيه الزوايا المركزية المقابلة لكل تكرار. يتم حساب الزوايا المركزية باستخدام العلاقة التالية:

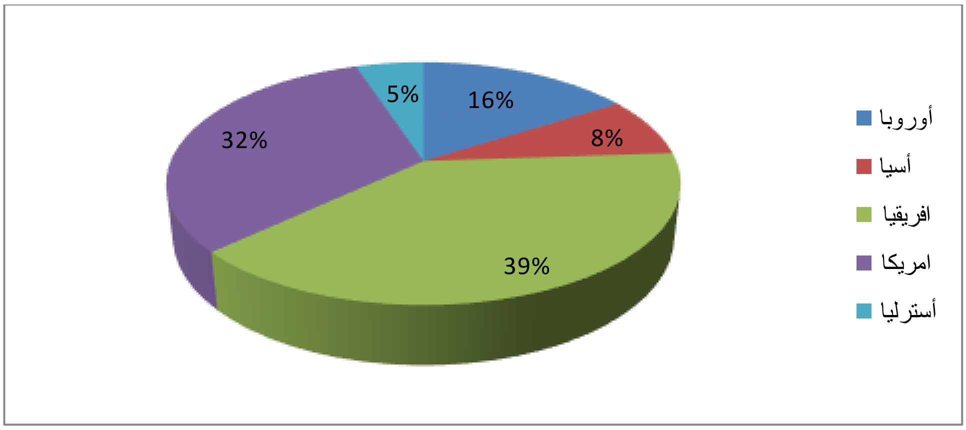
$$\text{الزاوية المركزية} = 360 * {}^{n_i}/_{\sum n_i} = f_i * 360$$

مثال (1-3): يبين الجدول التالي الإنتاج العالمي للذهب حسب القارات.

الجدول رقم (1-3): توزيع إنتاج الذهب حسب القارات.

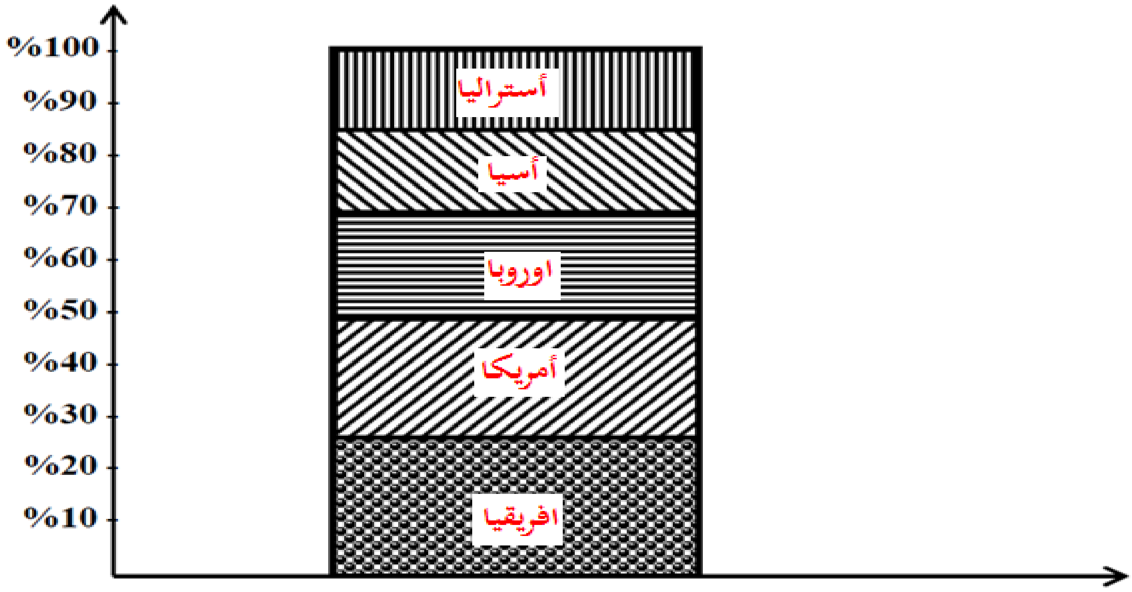
القارات	الذهب (بالطن)	f_i	الزوايا المركزية
اوروبا	176	0.16	°57.5
أسيا	87	0,08	°28,8
افريقيا	431	0,39	°140,4
أمريكا	350	0,32	°145,2
أستراليا	56	0,05	°18
المجموع	1100	100	°360

التمثيل البياني رقم (01): إنتاج كمية الذهب حسب القارات (الدائرة النسبية)



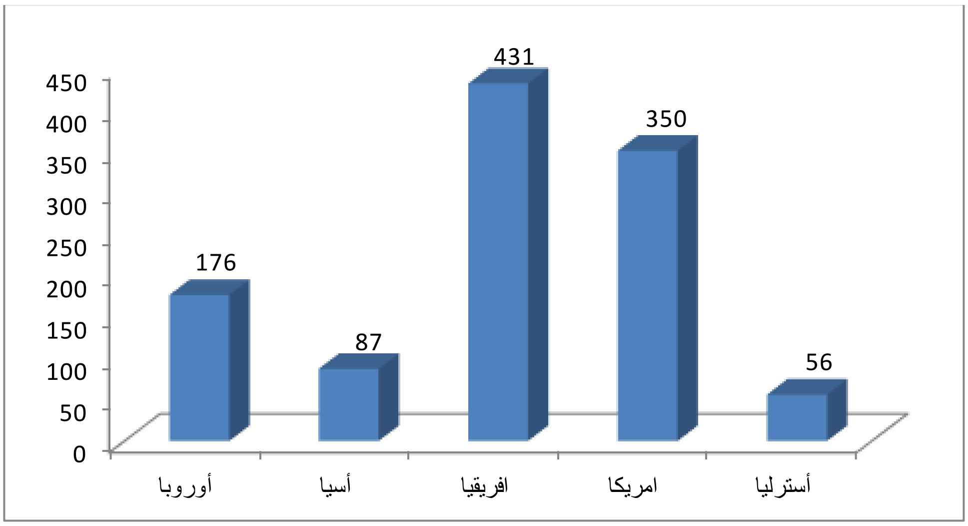
I-2- العمود المجزأ: هو عبارة عن مستطيل مجزأ قاعدة أجزاء، كل جزء يقابل تكرار معين للخاصية المدروسة. لوضع العمود المجزأ نستعمل النسب المئوية المقابلة لكل تكرار حيث أن طول المستطيل يساوي 100%. نأخذ المثال أعلاه.

التمثيل البياني رقم (02): إنتاج كمية الذهب حسب القارات (العمود المجزأ)



I-3- الأعمدة المستطيلة: عبارة عن مستطيلات متباعدة نوعا ما ولها قواعد متساوية وأطوالها تتناسب مع التكرار المقابل لمكونات الخاصية المدروسة. ولرسم التمثيل البياني بالأعمدة المستطيلة نأخذ معطيات المثال السابق.

التمثيل البياني رقم (03): إنتاج كمية الذهب حسب القارات (الأعمدة المستطيلة)



II– العرض البياني للمتغير الكمي المنقطع: نفرق عنا بين نوعين من العروض، العرض البياني للتكرارات البسيطة والعرض البياني للتكرارات المتجمعة.

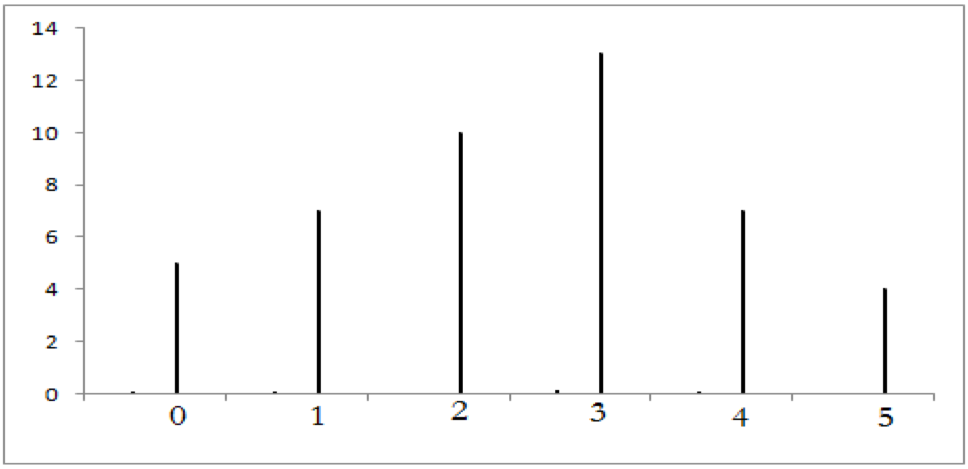
II–1– المستقيمات العمودية: وهو عبارة عن أعمدة بسيطة (مستقيمات عمودية) تتناسب أطوالها مع التكرار المقابل وتسمى الأعمدة البسيطة.

مثال (3–2): يبين الجدول التالي عدد الأطفال في العائلة لعينة متكونة من 46 أسرة. طلب منا عرض هذه البيانات بالأعمدة البسيطة.

الجدول رقم (3–2): توزيع الأسر حسب عدد الأطفال

عدد الأطفال x_i	0	1	2	3	4	5	المجموع
التكرار البسيط n_i	5	7	10	13	7	4	46

التمثيل البياني رقم (04): توزيع الأسر حسب عدد الأطفال (الأعمدة البسيطة)



نلاحظ من خلال التمثيل البياني أن أطول عمود يقابل 3 وتكراره يساوي 13 بمعنى أن أغلبية الأسر لها ثلاثة أطفال.

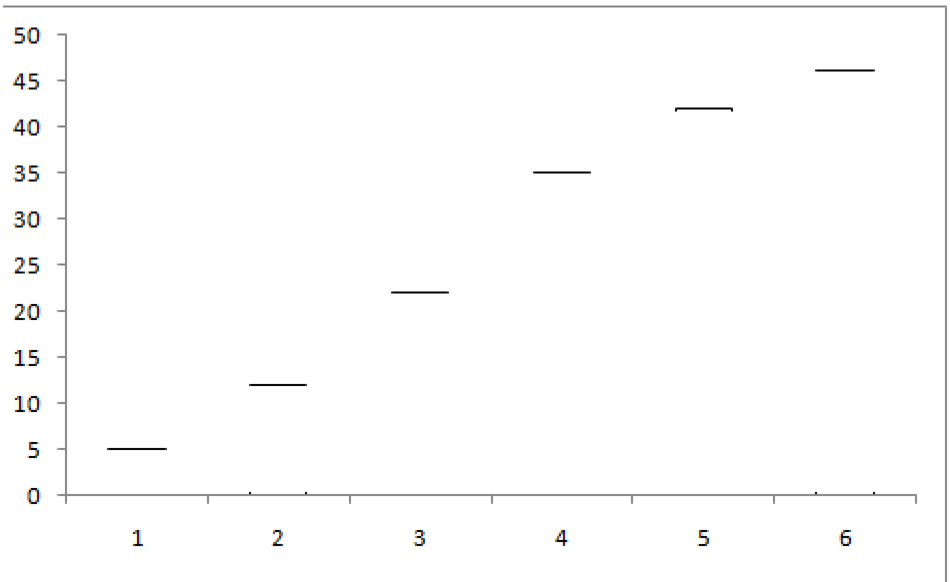
II– 2– المدرج التكراري الصاعد والنازل: لتمثيل التكرارات التجميعية الصاعدة والنازلة نأخذ المثال رقم (3-2) لأكثر توضيح مع إضافة عموديين آخرين في الجدول يحتويان على التكرارات التجميعية الصاعدة. والنازلة على التوالي:

الجدول رقم (3-3): التكرارات التجميعية الصاعدة والنازلة

xi	ni	F↑	F↓
0	5	5	46
1	7	12	41
2	10	22	34
3	13	35	24
4	7	42	11
5	4	46	4
∑	46	/	/

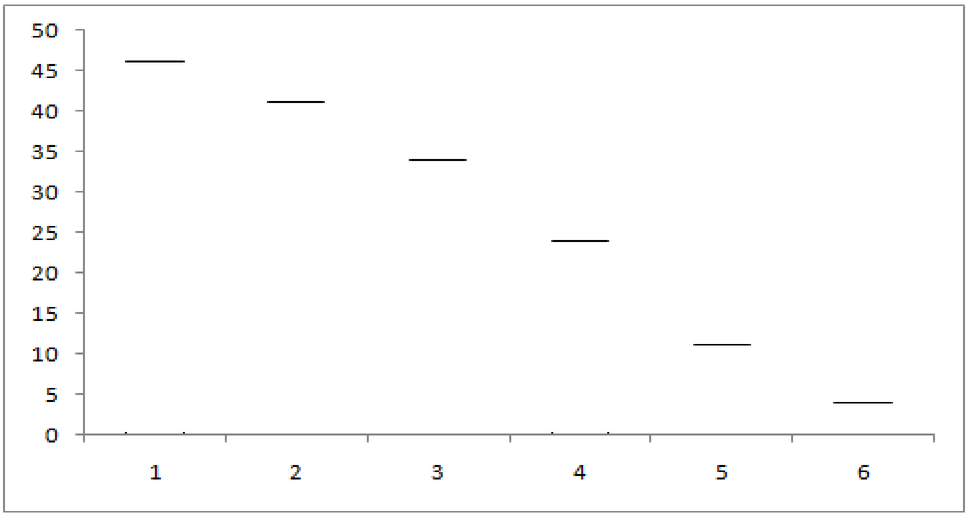
II-2-1- المدرج التكراري الصاعد: هو عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات التجميعية الصاعدة المقابلة لكل قيمة من قيم المتغير المدروسة وهو يبين لنا قوة تسارع أو تباطئ أو ثبات الظاهرة المدروسة.

التمثيل البياني رقم (05): منحنى التكرار التجميعي الصاعد لتوزيع الأسر



II– 2–2– المدرج التكراري النازل: عبارة عن قطع مستقيمة متنازلة حسب تنازل التكرارات التجميعية النازلة، حيث أن القطعة المستقيمة الأولى تقابل مجموع التكرارات وأصغر قيمة للمتغير المدروس، والقطعة الثانية تقابل مجموع التكرارات ناقص التكرار البسيط الأول مع القيمة الثانية للمتغير الإحصائي وهكذا. وبالرجوع إلى الجدول رقم (3–3) فنحصل على التمثيل البياني التالي:

التمثيل البياني رقم (06): منحنى التكرار التجميعي النازل لتوزيع الأسر



III–العرض البياني للمتغير الكمي المستمر: تعتبر العروض البيانية للمتغير الكمي المستمر (المتصل) هي أكثر العروض استعمالا من أهمها ما يلي:

III–1– المدرج التكراري: عبارة عن مستطيلات (أعمدة) متلاصقة أطوالها متناسبة مع التكرارات المقابلة لكل فئة وقواعدها تساوي طول الفئة المقابلة، بحيث توضع الفئات على محور السينات بينما توضع التكرارات على محور العينات. ونشير إلى أنه لا بد من ملاحظة أطوال الفئات قبل رسم المدرج التكراري، وهنا نميز بين حالتين عند رسم هذا الأخير:

III–1–1– المدرج التكراري في حالة تساوي أطوال الفئات: تكون فيه قاعدة المقارنة ثابتة ومتساوية ومن ثمة لا نجري أي تعديل على جدول المعطيات وعليه نقوم برسم المدرج التكراري مباشرة.

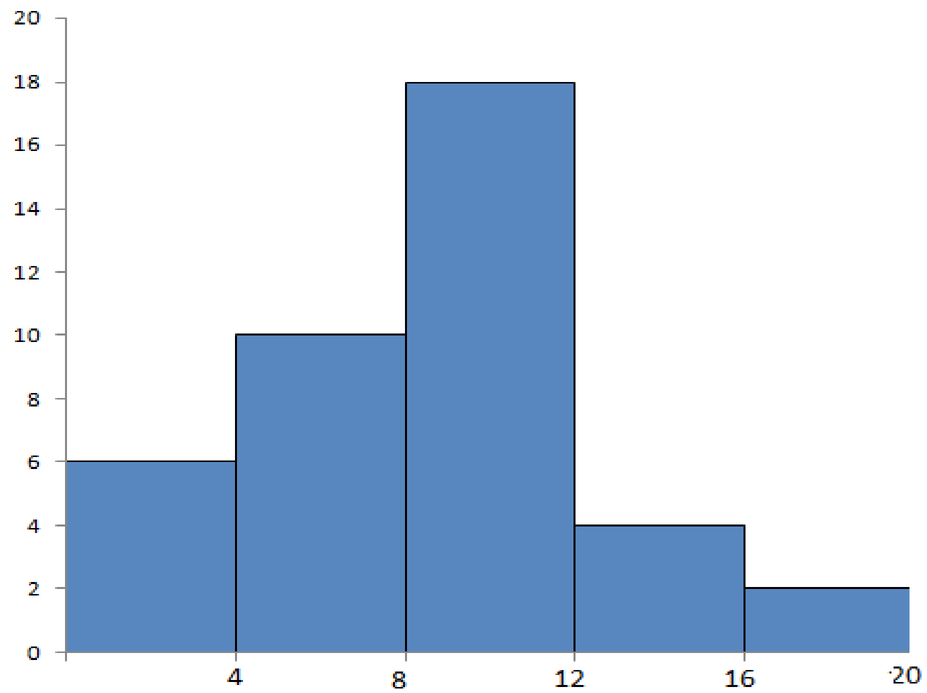
المثال (3–3): نريد دراسة توزيع علامات الطلاب في عينة متكونة من 40 طالب، والجدول التالي يوضح ذلك

المطلوب: تمثيل هذه البيانات بالمدرج التكراري.

الجدول رقم (3–4): توزيع علامات الطلبة.

الفئات	$]4-0]$	$]8-4]$	$]12-8]$	$]16-12]$	$]20-16]$	مجموع
التكرار (n_i)	6	10	18	4	2	40

التمثيل البياني رقم (07): التوزيع التكراري لنقاط الطلاب



نلاحظ أن الظاهرة تزداد حتى تأخذ قيمة عظمى ثم تبدأ في التناقص، ونقول عن هذه الظاهرة أنها ظاهرة طبيعية والفئة $]12-8]$ والتي تقابل أكبر تكرار وأكبر عمود تسمى الفئة المنوالية.

III–1–2– المدرج التكراري في حالة عدم تساوي أطوال الفئات: في حالة وجود فئات غير المتساوية في الجدول التكراري تقوم بتعديل التكرارات، لأن قاعدة المقارنة غير ثابتة حتى يكون هناك تناسب بين طول الفئة والتكرار المقابل لها. ولهذا الغرض نقوم بتعديل التكرارات ونستخدم لذلك المعادلة التالية:

$$n_i^* = \frac{n_i}{K_i} * K_i^*$$

التكرار المعدل: هو عبارة عن النسبة بين التكرار البسيط وطول الفئة المقابلة مضروب في طول الفئة المختار (K_i^*) وهو القاسم المشترك الأكبر لأطوال الفئات.

ملاحظة: تقوم بتعديل التكرارات عندما تكون الفئات غير منتظمة عند رسم المدرج التكراري وفي حالة تحديد الفئة المنوالية وحساب المنوال.

مثال (3–4): الجدول التالي يوزع عينة متكونة من 100 موظف حسب الأجر الشهري.

الجدول رقم (3–5): توزيع العمال حسب الأجر الشهري (الوحدة 10^3دج).

الفئات	[20 – 25[	[25 – 35[	[35 – 40[	[40 – 55[	[55 – 75[	[75 – 80[	المجموع
التكرار	5	15	20	25	30	5	100

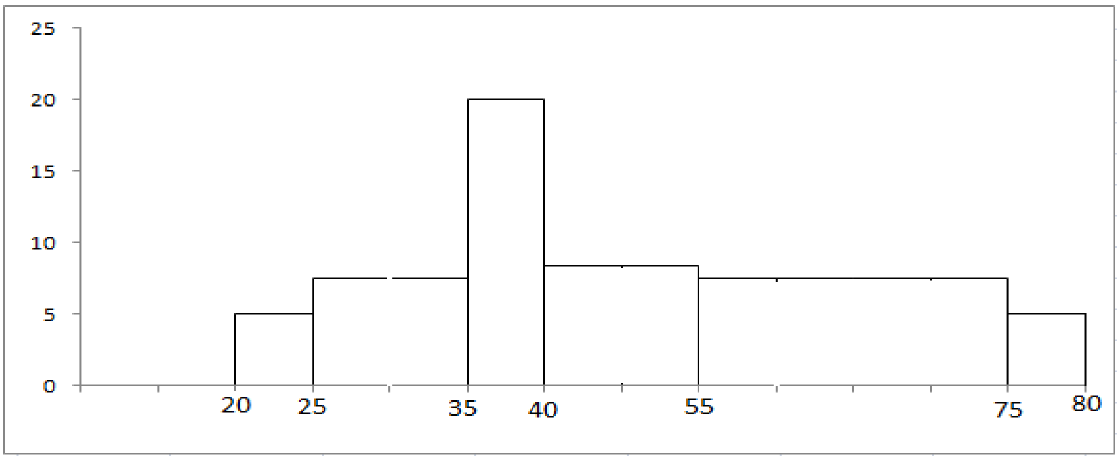
المطلوب: تمثيل هذه البيانات باستخدام المدرج التكراري.

الحل: نلاحظ أن أطوال الفئات غير متساوية ومنه نقوم بتعديل التكرارات أولا قبل رسم المدرج التكراري. نأخذ طول الفئة المختار 5 كأساس لتعديل التكرار (القاسم المشترك الأكبر).

الجدول رقم (3–6): التكرار المعدل

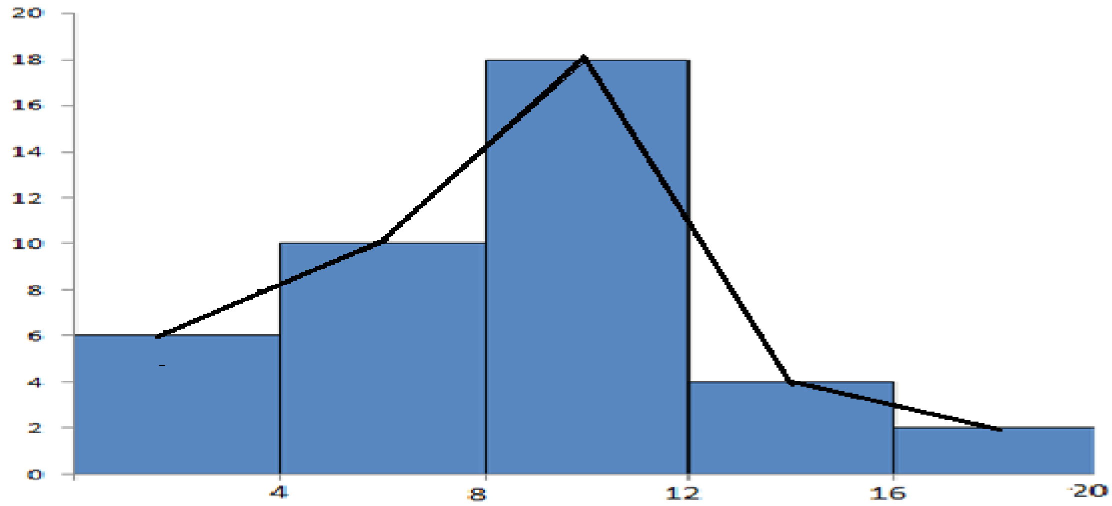
الفئات	[20 – 25[	[25 – 35[	[35 – 40[	[40 – 55[	[55 – 75[	[75 – 80[	المجموع
التكرار	5	15	20	25	30	5	100
K_i	5	10	5	15	20	5	/
n_i^*	5	7.5	20	8.33	7.5	5	/

التمثيل البياني رقم (08): توزيع العمال حسب الأجور



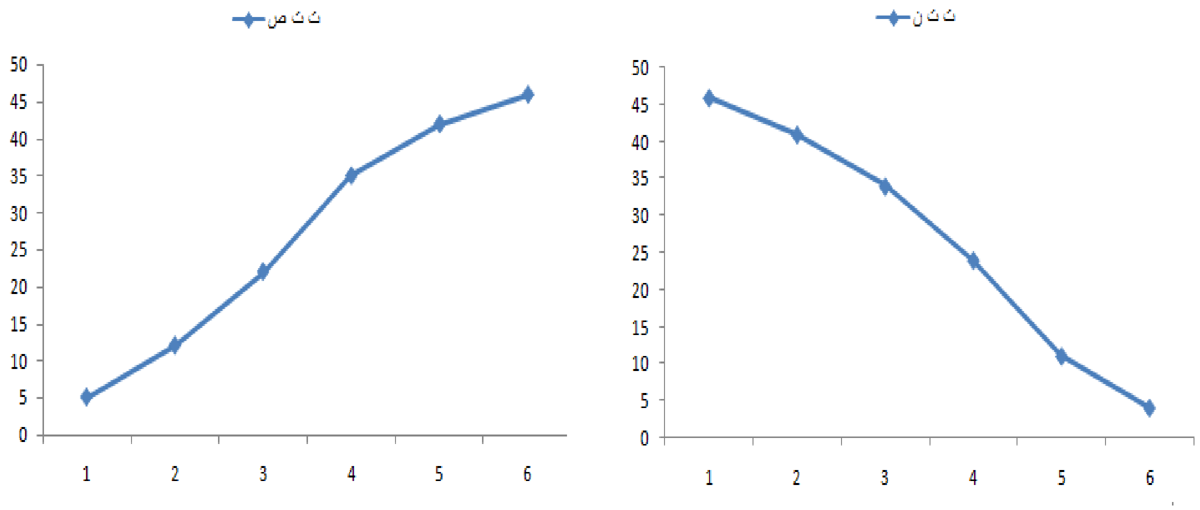
III–2–المضلع التكراري: عبارة عن قطع مستقيمة متصلة ومنكسرة، حيث يمكن رسم كل قطعة مستقيمة ابتدءا من مركز الفئة والتكرار المقابل لها. ولتوضيح ذلك نأخذ المثال رقم (3–3).

التمثيل البياني رقم (09): توزيع علامات الطلبة



III-3-منحنى التكرارات التجميعية الصاعدة والنازلة: نقوم برسم المنحنى التجميعي الصاعد في حالة بيانات المتغير الكمي المتصل باستعمال النقاط ذات الإحداثيات الحد الأعلى للفئة والتكرار التجميعي الصاعد المقابل لها، أما المنحنى التجميعي النازل فيرسم بوصل مجموعة النقاط ذات إحداثيات الحد الأدنى للفئة والتكرار النازل المقابل لها. ولغرض رسم هذا المنحنى نأخذ المثال (3-2).

التمثيل البياني رقم (10): توزيع الأسر حسب عدد الأطفال



ملاحظة: يبين كل من منحنى تكرار التجميعي الصاعد ومنحنى التكرار التجميعي النازل شدة أو ضعف تطور الظاهرة المدروسة، وتسمى فاصلة نقطة تقاطع المنحنى التجميعي الصاعد والنازل الوسيط.

III-4-المضلع التكراري الصاعد: نقوم برسم هذا المضلع باستعمال نقاط ذات إحداثيات مركز الفئة والتكرار التجميعي الصاعد المقابل لها.

III-5-المضلع التكراري النازل: نقوم برسمه وذلك بإيصال مجموعة نقاط ذات الإحداثيات مركز الفئة والتكرار التجميعي النازل المقابل لها.

تمارين محلولة

التمرين الأول: تمثل البيانات التالية توجيه 1080 طالب إلى السنة الثانية حسب الشعبة في ميدان العلوم الاقتصادية والتجارية وعلوم التسيير وهي ملخصة في الجدول أدناه. مثل هذه البيانات بالعرض البياني المناسب لها.

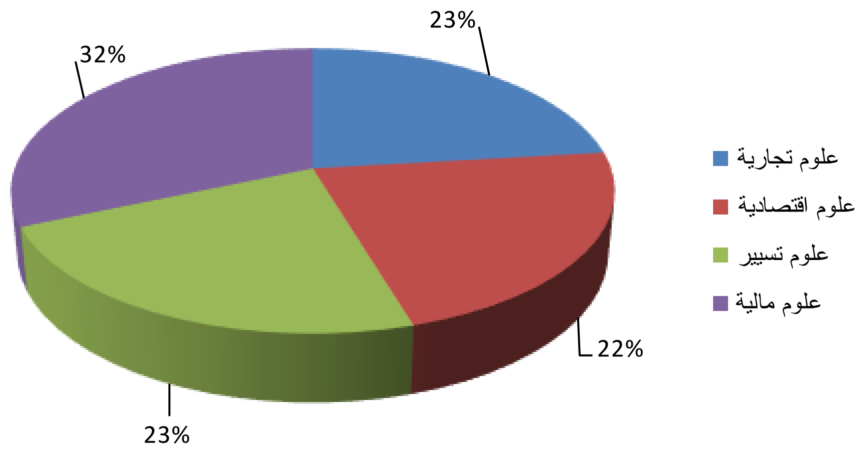
الشعبة	التكرار
علوم تجارية	250
علوم اقتصادية	240
علوم تسيير	250
علوم مالية	340
المجموع	1080

الحل: يمكن تمثيل هذه البيانات بالدائرة النسبية أو الأعمدة البيانية

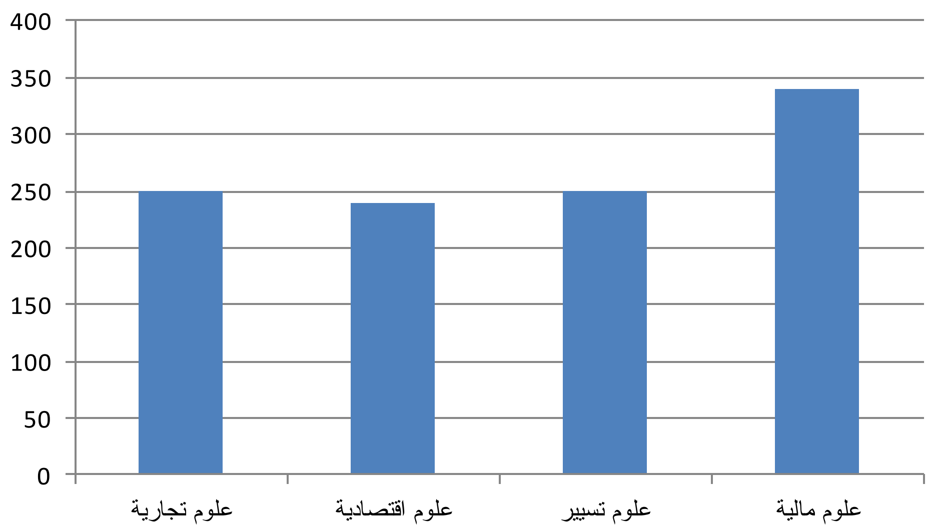
1-الدائرة النسبية: لرسم الدائرة النسبية يجب أولا حساب الزوايا المركزية لكل شعبة وذلك باستخدام العلاقة:

$$\frac{n_i}{\sum \mathrm{n_i}} * 360$$

الشعبة	التكرار	الزوايا المركزية
علوم تجارية	250	°83.33
علوم اقتصادية	240	°80
علوم تسيير	250	°83.33
علوم مالية	340	°113.33
المجموع	1080	°360



2–الأعمدة البيانية:

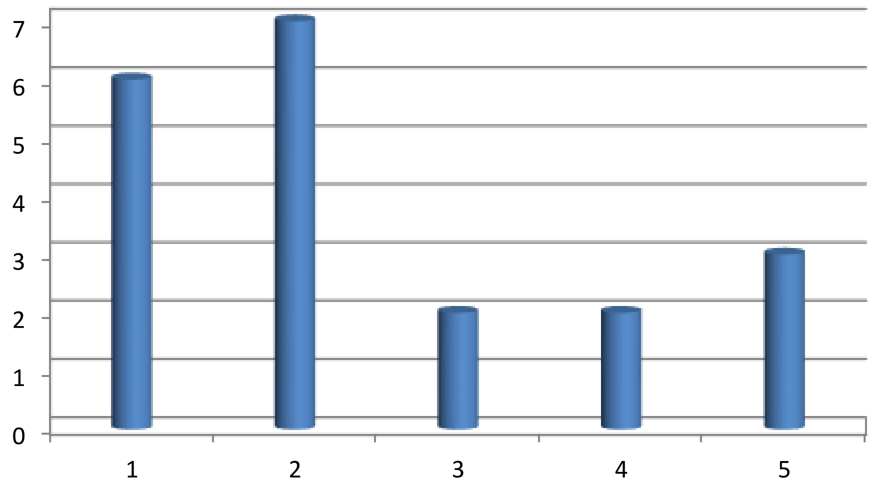


التمرين الثاني: يبين الجدول التالي عدد غيابات 20 طالبا في مقياس الاحصاء1 موزعة كما يلي:

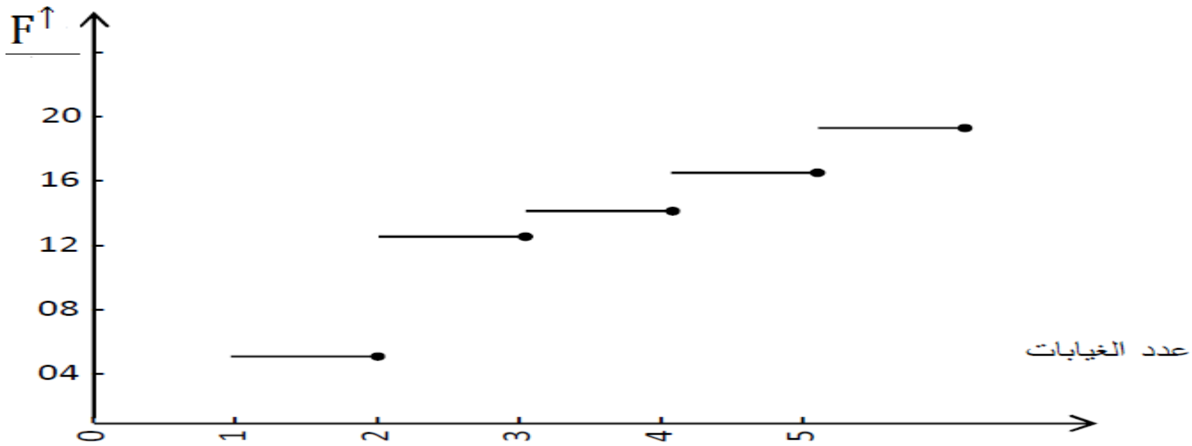
الغيابات	عدد الطلبة	$F\uparrow$	$F\downarrow$
1	06	06	20
2	07	13	14
3	02	15	07
4	02	17	05
5	03	20	03
المجموع	20	/	/

المطلوب: تمثيل هذه البيانات بالأعمدة البسيطة و المدرج التكراري الصاعد والنازل.

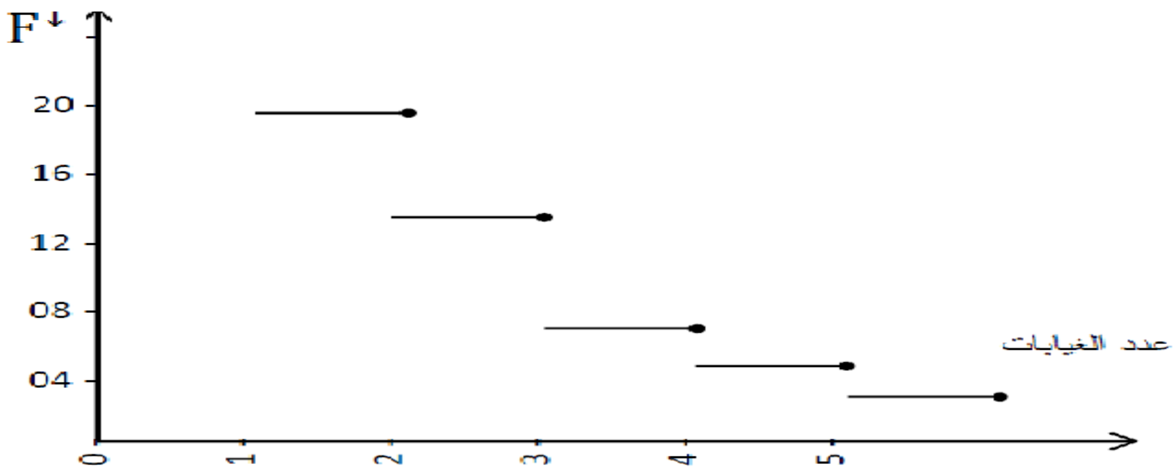
الحل: 1– الأعمدة البسيطة



2–المدرج التكراري الصاعد:



3–المدرج التكراري النازل:

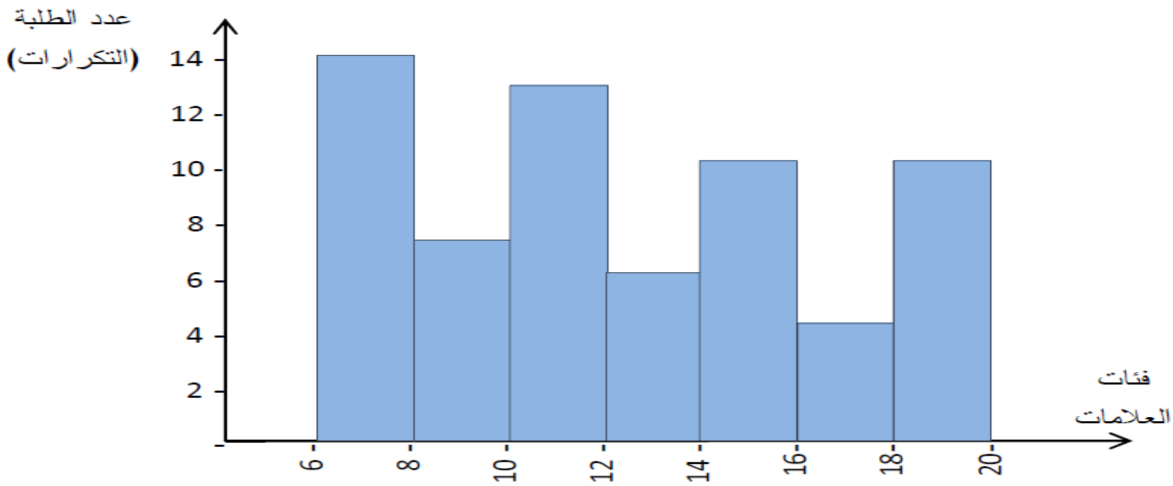


التمرين الثالث: بالرجوع إلى التمرين الرابع من المحور الثاني والذي يمثل علامات 64 طالبا في مقياس الإحصاء1.

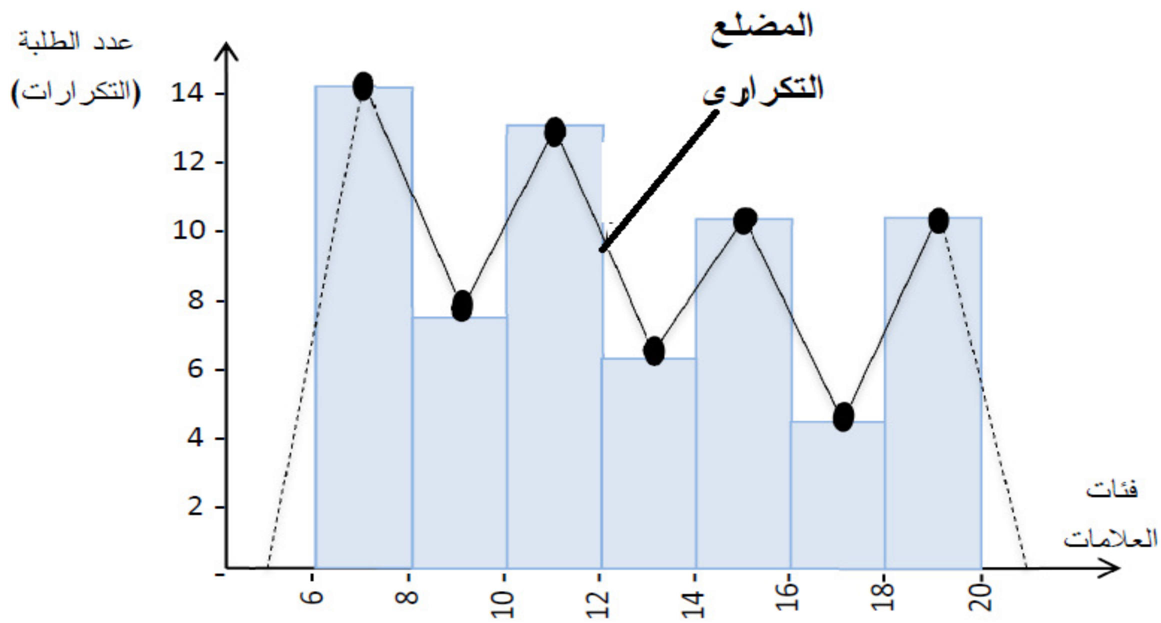
أرسم كل التمثيلات البيانية الخاصة بهذا النوع من المتغيرات الإحصائية.

الحل:

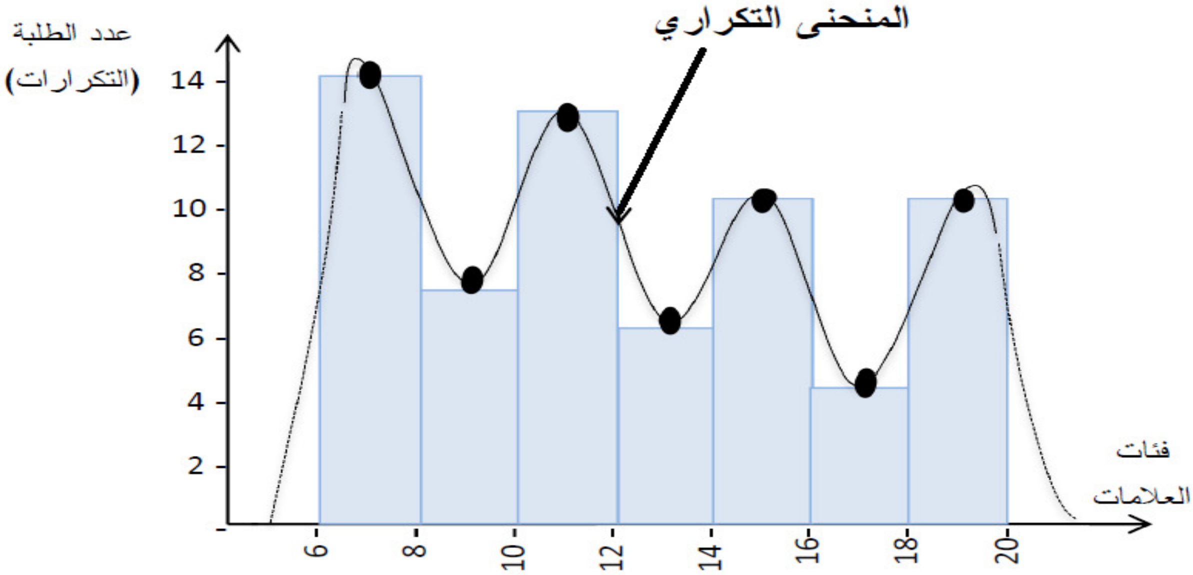
1 -**المدرج التكراري:**



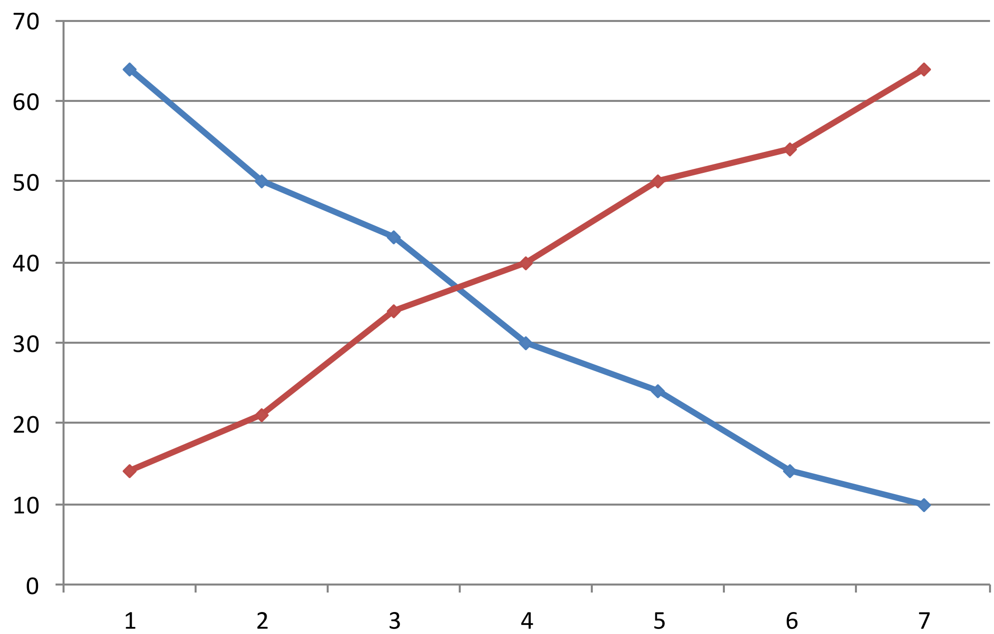
2**-المضلع التكراري:**



3-المنحنى التكراري:



4- المنحنى التكراري الصاعد والنازل:



الفصل الرابع

مقــاييــس النزعة المركزية

حتى نتمكن من تحليل الجداول الإحصائية والتنبؤ بالظاهرة المدروسة ومن ثم اتخاذ القرارات على ضوء هذه الأخيرة لابد من استعمال مقاييس وصفية عددية تكون أكثر فعالية. من خلال ملاحظة بيانات الظاهرة المدروسة سواء كانت في صورتها الأولية أو مبوبة في جداول التوزيعات التكرارية، نجد أن معظم البيانات تميل إلى التمركز حول قيمة معينة وكلما ابتعدنا عن هذا المركز (التجمع)، فإن عدد البيانات يبدأ في التناقص سواء على اليمين أو على اليسار، وتسمى هذه القيمة مركز التوزيع والحصول عليها ضروري ومهم في دراسة خصائص التوزيع ومقارنته مع توزيعات أخرى

مفهوم النزعة المركزية

يلجأ الباحث بعد عملية جمع البيانات عن الظاهرة المدروسة وتبويبها في جداول التوزيعات التكرارية إلى استخدام المقاييس الإحصائية وهي الطرق التي تقوم بحساب القيمة التي تتمركز حولها معظم المشاهدات، نسمي هذه المقاييس بمقاييس بالنزعة المركزية، حيث أنها تبحث في تقدير قيمة تتمركز حولها أغلبية القيم.

يعبر قياس النزعة المركزية عن القيمة المركزية للتوزيع الإحصائي، ولقياسها نستعمل عدة مقاييس من أهمها المتوسط الحسابي، الوسيط والمنوال.

I– المتوسط الحسابي (الوسط الحسابي): يعد المتوسط الحسابي من أشهر وأكثر المقاييس استخداما وشيوعا، **و**هو عبارة عن حاصل قسمة مجموع قياسات الظاهرة المدروسة على مجموع الوحدات الإحصائية. ويرمز له بالرمز $\overline{X}$. ويمكن حسابه في حالة وجود بيانات مبوبة وبيانات غير مبوبة.

I–1– المتوسط الحسابي البسيط: يستخدم هذا المتوسط في حالة بيانات غير مبوبة، أي عندما يكون لقياسات المتغير المدروس نفس الأهمية. إذا كان لدينا سلسلة متكونة من n مشاهدة $x_1, x_2, \dots\dots\dots, x_n$، فإن المتوسط الحسابي لهذه السلسلة هو:

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

مثال (4–1): لدينا علامات طالب لأربع مقاييس لها نفس المعامل والمتكونة من المشاهدات التالية:

12	14	10	15
2	2	2	**2**

المطلوب: حساب متوسط علامات هذا الطالب (المعدل).

الحل:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{4} x_i}{4} = \frac{12 + 14 + 10 + 15}{4} = 12.75$$

يمكن أن نحسب المتوسط الحسابي باستعمال الوسط الفرضي (x_0)، حيث على أساس قيمة فرضية معينة تحدد قيمة المتوسط الحسابي الحقيقي. يفضل أن تكون قيمة الوسط الفرضي مساوية لإحدى مفردات الظاهرة المدروسة. ويعطى المتوسط الحسابي بطريقة الوسط الفرضي بالعلاقة التالية:

$$\overline{X} = x_0 + \frac{\sum_{i=1}^{n}(x_i - x_0)}{n}$$

حيث يمثل ($x_i - x_0$) انحرافات المفردات عن وسطها الفرضي. نأخذ المثال (3-1) ونفرض $x_0 = 12$. قيمة المتوسط الحسابي باستعمال الوسط الفرضي هي:

$$\overline{X} = 12 + \frac{(12-12) + (14-12) + (10-12) + (15-12)}{4} = 12.75$$

I-2- المتوسط الحسابي المرجح: تختلف أهمية قياسات المتغير الإحصائي من قيمة إلى أخرى في كثير من الأحيان، فهي تختلف باختلاف معامل الترجيح الخاص بكل قيمة من قيم المتغير الإحصائي ومن هنا أدخل الترجيح في حساب المتوسط الحسابي. الترجيح يعبر عن أهمية (وزن) قياس معين من قياسات المتغير الإحصائي. وهو عبارة عن حاصل قسمة مجموع محصلة ضرب المشاهدات في تكرارها على مجموع التكرارات ومنه تعطى كل من علاقة المتوسط الحسابي المرجح و المتوسط الحسابي المرجح الفرضي على الشكل التالي على التوالي:

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

$$\overline{X} = x_0 + \frac{\sum_{i=1}^{n} n_i (x_i - x_0)}{\sum_{i=1}^{n} n_{i\,i}}$$

مثال (4-2): لدينا علامات طالب في خمس مقاييس بمعاملات مختلفة هي كالتالي:

15	14	16	10	8
3	2	4	1	2

المطلوب: أحسب المتوسط الحسابي المرجح، والمتوسط الحسابي المرجح الفرضي؟

الحل: قيمة المتوسط الحسابي المرجح هي:

$$\overline{X} = \frac{\sum_{i=1}^{12} n_i \chi_i}{\sum_{i=1}^{12} n_i} = \frac{(8*2)+(10*1)+(16*4)+(14*2)+(11*3)}{12} = 12.58$$

نفرض أن $\chi_0 = 11$ ، ومنه قيمة المتوسط الحسابي المرجح الفرضي هي:

$$\overline{X} = 11 + \frac{2(8-11)+1(10-11)+4(16-11)+2(14-11)+3(11-11)}{12} = 12.58$$

ملاحظة: في حالة بيانات مبوبة ذات المتغير الكمي المتصل نستعمل نفس علاقة الوسط الحسابي المرجح ولكن نستبدل القيم النقطية المتغير (X_i) بمراكز الفئات.

I-3- المتوسط الحسابي المرجح لأوساط حسابية: يستخدم هذا النوع من المتوسطات عند دمج مجموعتين أو أكثر يعرف مسبقا المتوسط الحسابي الخاص بكل مجموعة. إذا كانت لدينا مجموعة من المشاهدات (n_1) وسطها الحسابي $(\overline{X}_1)$ ومجموعة أخرى (n_2) بوسط حسابي $(\overline{X}_2)$ فإن المتوسط الحسابي للمجموعتين معا هو:

$$\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$$

مثال (3-4): الجدول الموالي يبين عدد العمال ومتوسط الغيابات في أربع مديريات خلال سنة. المطلوب حساب متوسط الغيابات في المؤسسة.

مديريات المؤسسة	المديرية الأولى	المديرية الثانية	المديرية الثالثة	المديرية الرابعة
عدد العمال	150	80	125	235
الغيابات (بالأيام)	11	5	16	12

$$\overline{X} = \frac{150*11+80*5+125*16+235*12}{11+5+16+12} = 4114.091$$

I-4- خصائص الوسط الحسابي:

- يتأثر المتوسط الحسابي بالقيم المتطرفة (الشاذة)؛
- أكثر المقاييس استخداما؛
- لا يمكن أن يكون لأي توزيع تكراري أكثر من وسط حسابي؛
- مجموع انحرافات قيم المتغير بالنسبة للمتوسط الحسابي يساوي صفر؛
- لا يمكن استخدامه في الجداول الإحصائية المفتوحة لأنه يعتمد في حسابه على مركز الفئة؛

- لا يمكن حساب المتوسط الحسابي بيانيا.

ملاحظة:

نستعمل المتوسط الحسابي عندما تكون العلاقة خطية بين الخاصية المدروسة والتكرار، أما في حال عدم وجود هذه العلاقة فنستعين بوسائط أخرى وهي المتوسط الهندسي، المتوسط التوافقي والمتوسط التربيعي.

II–المتوسط الهندسي (الوسط الهندسي): كثيرا ما نصادف حالات تكون فيها قيم الظاهرة المدروسة عبارة عن نسب أو معدلات، أو نريد حساب معدلات النمو (معدل النمو الاقتصادي، معدل زيادة الأجور....) فنلجأ إلى استخدام المتوسط الهندسي لدراسة مثل هذه الظواهر. يعرف المتوسط الهندسي (M_G) لـ n قيمة من قيم المتغير الإحصائي بأنه الجذر النوني لجداء هذه القيم. ونفرق هنا بين طريقتين في حساب المتوسط الهندسي، حالة البيانات المبوبة (المتوسط الهندسي المرجح) وحالة البيانات غير المبوبة (المتوسط الهندسي البسيط).

II–1–المتوسط الهندسي البسيط: إذا كان لدينا x_n ,.......... ,x_1,x_2 تمثل قيم مشاهدات ظاهرة ما، فإن المتوسط الهندسي لهذه الأخيرة يعطى بالعلاقة التالية:

$$M_G = \sqrt[n]{x_1 * x_2 * \dots * x_i * \dots .* x_n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

ولتسهيل العمليات الحسابية نقوم بإدخال اللوغاريتم على الصيغة السابقة لتصبح أس المتوسط الحسابي للوغاريتم القيم، بالشكل التالي:

$$\exp LOG M_G = \exp \frac{1}{n} \mathrm{LOG}[x_1 * x_2 * \dots .* x_n]$$

$$= \exp \frac{1}{n}[\mathrm{LOG}x_1 + \mathrm{LOG}x_2 + \cdots . + \mathrm{LOG}x_n]$$

$$M_G = \exp \frac{1}{n} \sum_{i=1}^{n} \mathrm{LOG}x_i$$

مثال (4–4): لتكن لدينا السلسلة الإحصائية التالية:

10 14 16 7 9

المطلوب: أوجد المتوسط الهندسي لهذه السلسلة؟

الحل:

$$M_G = \sqrt[n]{x_1 * x_2 * \dots .* x_n} = \sqrt[5]{9 * 7 * 16 * 14 * 10} = 10.71$$

II-2- المتوسط الهندسي المرجح: إذا كانت لدينا السلسلة $x_1, x_2, \dots\dots\dots, x_k$ ، والتي تمثل قيم متغير الظاهرة المدروسة وكانت $n_1, n_2, \dots\dots\dots, n_k$ تمثل التكرارات المقابلة لها فإن قيمة المتوسط الهندسي في هذه الحالة يعطى كما يلي:

$$M_G = \sqrt[\sum_{i=1}^{k} n_i]{x_1^{n_1} * x_2^{n_2} * \dots * x_k^{n_k}}$$

$$M_G = e^{\frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} n_i LOGx_i}$$

مثال (4-5): أحسب المتوسط الهندسي من خلال البيانات الإحصائية التالية:

X_i	n_i	$LOGX_i$	$n_i LOGX_i$
2	2	0.30	0.60
4	3	0.60	1.80
5	2	0.69	1.38
6	4	0.77	3.08
$\sum$	11		6.86

ومنه المتوسط الحسابي الهندسي هو:

$$LOGM_G = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} n_i LOGx_i = \frac{1}{11} 2.624 = 0.623$$

$$M_G = e^{0.623} = 4.203$$

ملاحظة: في حالة المتغير الكمي المتصل نستعمل مراكز الفئات بدل قيم المتغير الإحصائي.

II-3- خواص المتوسط الهندسي:

- يستخدم كثيرا في الظواهر الاقتصادية (حساب الأرقام القياسية للأسعار، معدل النمو الاقتصادي، معدل الفائدة، ومعدل النمو السكاني)؛

-لا يتأثر بالقيم المتطرفة؛

-لا يمكن حسابه في جداول التوزيع التكرارية المفتوحة؛

-لا يصبح للمتوسط الهندسي أي معنى في حالة وجود قيم سالبة أو معدومة في جدول التوزيع التكراري.

III- المتوسط التوافقي: الوسط التوافقي هو عبارة على مقلوب المتوسط الحسابي لمقلوب قيم المتغير الإحصائي، ونرمز له بالرمز (M_H). يستعمل الوسط التوافقي في حالة وجود علاقة عكسية بين ظاهرتين أو بين

الخاصية المدروسة والتكرار، كما أنه يستخدم في قياس معدلات السرعة. وهنا كذلك نفرق بين الوسط التوافقي البسيط والمرجح.

III-1- المتوسط التوافقي البسيط: إذا كانت $x_1, x_2, \dots\dots\dots, x_n$ تمثل قيم متغير إحصائي لظاهرة معينة، فإن المتوسط التوافقي البسيط لهذه الظاهرة هو:

$$M_H = \frac{n}{{}^1/_{x_1} + {}^1/_{x_2} + \cdots . + {}^1/_{x_i} + \cdots . + {}^1/_{x_n}} = \frac{n}{\sum_{i=1}^{n} {}^1/_{x_i}}$$

مثال (4-6): يقطع دراج مسافة 100 كلم في أربع مراحل، سرعة كل مرحلة على التوالي هي: 10كلم/سا، 30كلم/سا، 40كلم/سا و20كلم/سا. ما هي السرعة المتوسطة لهذا الدراج؟

الحل: بتطبيق الصيغة أعلاه للمتوسط التوافقي نجد:

$$M_H = \frac{400}{{}^1/_{10} + {}^1/_{30} + {}^1/_{40} + {}^1/_{20}} = 19.2\text{km/h}$$

III-2- المتوسط التوافقي المرجح: إذا كانت لدينا السلسلة $x_1, x_2, \dots\dots\dots, x_k$ ، والتي تمثل قيم متغير الظاهرة المدروسة وكانت $n_1, n_2, \dots\dots\dots, n_k$ تمثل التكرارات المقابلة لها، فإن قيمة المتوسط التوافقي يعطى بالصيغة التالية:

$$M_H = \frac{n_{1+} n_{2+} \dots . + n_{i+} \dots . . n_k}{{}^1/_{x_1} + {}^1/_{x_2} + \cdots . + {}^1/_{x_i} + \cdots . + {}^1/_{x_k}} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} {}^{n_i}/_{x_i}}$$

مثال (4-7): نأخذ معطيات المثال (4-5) ونحسب المتوسط التوافقي.

X_i	n_i	${}^{n_i}/_{x_i}$
2	2	1
4	3	0.75
5	2	0.4
6	4	0.66
$\sum$	11	2.81

ومنه قيمة المتوسط التوافقي هي:

$$M_H = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} {}^{n_i}/_{x_i}} = \frac{11}{2.81} = 3.91$$

III-3- خواص المتوسط التوافقي:

- يتأثر بالقيم المتطرفة أقل من تأثر المتوسط الحسابي؛

-لا يمكن حسابه في جداول التوزيع التكرارية المفتوحة؛

-يستخدم في حساب متوسطات الأسعار والسرعة؛

-لا يمكن حسابه في حالة وجود قيم معدومة في جدول التوزيع التكراري.

IV-المتوسط التربيعي: يستعمل المتوسط التربيعي في حالة وجود علاقة مضاعفة بين الخاصية المدروسة والتكرار، وهو عبارة عن الجذر التربيعي للنسبة بين مجموع مربعات قيم المتغير الإحصائي وعدد المشاهدات. وهنا كذلك نفرق بين المتوسط التربيعي البسيط والمتوسط التربيعي المرجح.

IV-1- المتوسط التربيعي البسيط: لتكن $x_1, x_2, \ldots\ldots\ldots, x_k$ تمثل قيم المتغير المدروس فإن المتوسط التربيعي البسيط لهذه البيانات يكتب على الشكل التالي:

$$M_Q = \sqrt{\sum_{i=1}^{n} x_i^2 / n}$$

مثال (4-8): أوجد المتوسط التربيعي لسلسلة البيانات غير المبوبة التالية:

5 4 2 -3 -1

$$M_Q = \sqrt{\sum_{i=1}^{n} x_i^2 / n} = \sqrt{\frac{(-1)^2 + (-2)^2 + (2)^2 + (4)^2 + (5)^2}{5}} = 3.30$$

IV-2- المتوسط التربيعي المرجح: إذا كانت لدينا السلسلة $x_1, x_2, \ldots\ldots\ldots, x_k$ ، والتي تمثل قيم متغير الظاهرة المدروسة وكانت $n_1, n_2, \ldots\ldots\ldots, n_k$ تمثل التكرارات المقابلة لها، فإن قيمة المتوسط التربيعي لهذه الظاهرة يعطى بالصيغة التالية:

$$M_Q = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i}}$$

مثال (4-9): بالرجوع إلى المثال (4-5)، أحسب المتوسط التربيعي.

x_i	n_i	x_i^2	$n_i x_i^2$
2	2	4	8
4	3	16	46
5	2	25	125
6	4	36	216
$\sum$	11	81	409

إذا قيمة المتوسط التربيعي هي:

$$M_Q = \sqrt{\frac{409}{11}} = 6.10$$

ملاحظة: تستعمل الوسائط المذكورة سابقا $(M_H , M_G, \bar{X}, M_Q)$ حسب نوع العلاقة الموجودة بين الخاصية المدروسة والتكرار وهذا انطلاقا من العلاقة العامة التالية:

$$f(x) = x^{\alpha}$$

حيث α تأخذ القيم: 1-؛ 0؛ 1؛ 2

لما $\alpha = -1$، فإن $f(x) = \frac{1}{x}$ وفي هذه الحالة نستعمل المتوسط التوافقي.

لما $\alpha = 0$، فإن $f(x) = 1$ وفي هذه الحالة نستعمل المتوسط الهندسي.

لما $\alpha = 1$، فإن $f(x) = x$ وفي هذه الحالة نستعمل المتوسط الحسابي.

لما $\alpha = 2$، فإن $f(x) = x^2$ وفي هذه الحالة نستعمل المتوسط التربيعي.

كما أن قيم α تبين ترتيب الوسائط الأربعة، بحيث تكون العلاقة التالية دائما محققة.

$$M_Q > \bar{X} > M_G > M_H$$

وحتى نتأكد من صحة العلاقة، نقارن بين المتوسطات المحسوبة في المثال (4-5)، حيث وجدنا القيم كما يلي:

$M_H = 3.91$، $M_G = 4.20$، $\bar{X} = 4.54$، و $M_Q = 6.10$ وعليه فالعلاقة صحيحة.

V- الوسيط

هو ثاني مقياس من مقاييس النزعة المركزية، ويمثل المشاهدات التي تكون التكرارات التي تسبقها تساوي التكرارات التي تليها، بمعنى آخر هي القيمة التي تقسم المجتمع الإحصائي إلى قسمين متساويين.

مثال: إذا كان وسيط أجور عمال شركة ما هو 12.000دج، معنى ذلك أن 50% من العمال لديهم أجر أقل من 12.000دج و50% من العمال لديهم أجر أكبر من 12.000دج.

V-1- الوسيط في حالة بيانات غير المبوبة: الوسيط عبارة عن قيمة مركزية لسلسلة إحصائية، نرمز له بـ M_e. ولحسابه نقوم بترتيب قيم سلسلة المتغير الإحصائي ترتيبا تصاعديا أو تنازليا. وهنا نميز بين حالتين:

V-1-1- الوسيط في حالة بيانات غير المبوبة المفردة: إذا كان عدد المشاهدات (n) عددا فرديا، فإن قيمة الوسيط هي قيمة x التي ترتيبها $\frac{n+1}{2}$ وبالتالي: $M_e = x_{\frac{n+1}{2}}$

مثال (4-10): لدينا السلسلة الإحصائية التالية:

2 1 0 0 1 2 3

المطلوب: أوجد قيمة الوسيط لهذه البيانات؟

الحل: عند ترتيب السلسلة ترتيبا تصاعديا (n=7)، نحصل على:

3 2 2 1 1 0 0

فإن الوسيط هو القيمة التي تتوسط السلسلة المرتبة: ⇧

$$M_e = x_{\frac{n+1}{2}} = 1$$

V-1-2- الوسيط في حالة بيانات غير المبوبة الزوجية: إذا كان عدد المشاهدات (n) عدد زوجي، فإنه يوجد قيمتين في وسط السلسلة والوسيط في هذه الحالة هو متوسط هذان القيمتين، هو قيمة x التي ترتيبها $(\frac{n}{2}+1 ، \frac{n}{2})$، وبالتالي:

$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

مثال (4-11): لدينا السلسلة الإحصائية التالية:

2 4 1 0 0 1 2 3

المطلوب: أوجد قيمة الوسيط لهذه البيانات؟

الحل: عند ترتيب السلسلة ترتيبا تصاعديا (n=8):

4 3 2 2 1 1 0 0

الوسيط هو متوسط القيمتين: ⇧ ⇧

$$M_e = \frac{1+2}{2} = 1.5$$

V-2- الوسيط في حالة بيانات المبوبة: وهنا نميز بين حالتين كذلك:

V-2-1- الوسيط في حالة المتغير الكمي المنفصل: لحساب الوسيط في حالة بيانات مبوبة مع دراسة متغير كمي منفصل نتبع الخطوات التالية:

- نقوم بحساب التكرار المجمع الصاعد؛
- نقوم بتحديد رتبة الوسيط $(\frac{\sum_{i=1}^{k} n_i}{2})$ ؛
- نبحث عن رتبة الوسيط في العمود (السطر) الخاص بالتكرار التجميعي الصاعد بحيث تكون القيمة التي تكرارها يساوي رتبة الوسيط $(\frac{\sum_{i=1}^{k} n_i}{2})$ أو الأعلى منها مباشرة؛ لتكون القيمة التي تقابلها في عمود (سطر) المتغير المدروس هي قيمة الوسيط.

مثال (4–12): تمثل البيانات التالية توزيع العائلات حسب عدد الأطفال.

عدد الأطفال (X_I)	عدد العائلات (n_i)
0	5
1	10
2	15
3	20
4	10
5	5
المجموع	65

المطلوب: إيجاد الوسيط.

الحل: لحساب وسيط هذه البيانات نتبع الخطوات التالية:

- نضيف عمود نحسب فيه التكرار المتجمع الصاعد.

عدد الأطفال (X_i)	عدد العائلات (n_i)	ت.ت.ص
0	5	5
1	10	15
2	15	30
3	20	50
4	10	60
5	5	65
المجموع	65	/

- نحدد رتبة الوسيط:

$$\frac{\sum_{i=1}^{k} n_i}{2} = \frac{65}{2} = 32.5$$

- نبحث في العمود الثالث (ت.ت.ص) عن التكرار الذي يساوي 32.5 أو الذي يأتي بعده مباشرة.
- نلاحظ أن (50 تأتي مباشرة بعد 32.5) وعليه قيمة الوسيط هي قيمة X_i التي تقابل القيمة 50، ومنه فإن قيمة الوسيط هي 3 ($M_e = 3$).

V-2-2-الوسيط في حالة المتغير الكمي المتصل: لحساب الوسيط في هذه الحالة نتبع الخطوات التالية:

- نقوم حساب التكرار المجمع الصاعد؛
- تحديد رتبة الوسيط وهي عبارة عن ($\frac{\sum_{i=1}^{k} n_i}{2}$)؛
- تحديد الفئة الوسيطية التي ينتمي إليها الوسيط وهي الفئة التي تقابل التكرار المجمع الصاعد الذي يساوي رتبة الوسيط أو الأكبر منه مباشرة؛
- حساب علاقة الوسيط و التي تكتب على الشكل التالي:

$$M_e = A + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - F_{n-1}^{\uparrow}}{n_{M_e}} * k$$

حيث:

A : يمثل الحد الأدنى للفئة الوسيطية؛

$F_{n-1}^{\uparrow}$: يمثل التكرار التجميعي الصاعد للفئة ما قبل الوسيطية؛

k : يمثل طول الفئة؛

n_{M_e}: يمثل تكرار الفئة الوسيطة.

مثال (4-13): يبين الجدول التالي ظاهرة تأخر العمال عن الالتحاق بمناصب عملهم.

التأخر بالدقائق (X_i)	]10 – 5]	]15 – 10]	]20 – 15]	]25 – 20]	]30 – 25]	المجموع
عدد العمال (n_i)	12	20	25	15	5	77

المطلوب: أحسب قيمة الوسيط؟

الحل: لتحديد قيمة الوسيط نتبع المراحل التالية:

الفئات	التكرار (n_i)	ت.ت.ص ($F^{\uparrow}$)	ت.ت.ن ($F^{\downarrow}$)
]10 – 5]	12	12	77
]15 – 10]	20	32	65
]20 – 15]	25	57	45
]25 – 20]	15	72	20
]30 – 25]	5	77	5
المجموع	77	/	/

- لتحديد رتبة الوسيط نحسب العبارة:

$$\frac{\sum_{i=1}^{k} n_i}{2} = \frac{77}{2} = 38.5$$

- تحديد الفئة الوسيطية: وهي الفئة $[15-20[$
- نحسب علاقة الوسيط:

$$M_e = A + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - F_{n-1}^{\uparrow}}{n_{M_e}} * k = 15 + \frac{38.5-32}{25} * 5 = 16.3$$

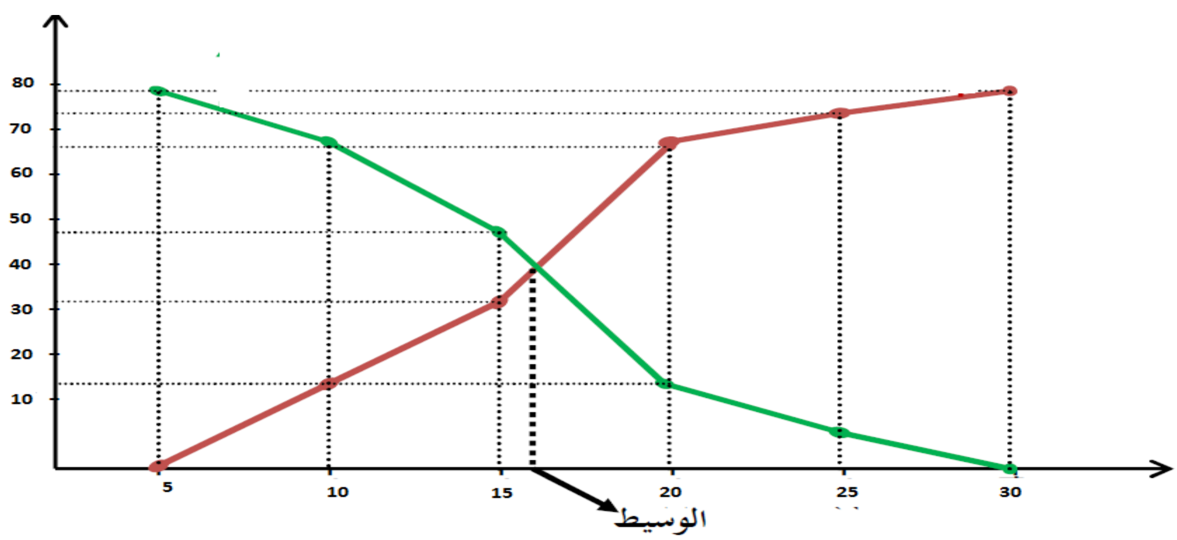
V-3-حساب الوسيط بيانيا:

يمكن حساب الوسيط بيانيا وذلك عن طريق رسم المنحنى التكراري الصاعد أو النازل أو رسم المنحنيين في نفس المعلم. ويكون ذلك بإتباع المراحل التالية:

- حساب التكرار المجمع الصاعد والنازل ؛
- وضع المنحنى التكراري الصاعد والنازل في نفس المعلم؛
- قيمة الوسيط هي نقطة تقاطع المنحنيين بعد إسقاطها على محور الفواصل.

مثال (4-14): نريد حساب قيمة الوسيط بيانيا، نأخذ معطيات المثال (4-13) ونضيف إلى جدول التوزيع التكراري عمود رابع نضع فيه التكرار التجميعي النازل، حتى نتمكن من رسم منحنيين التكرار التجميعي الصاعد والنازل في نفس المعلم.

التمثيل البياني رقم (11): تحديد قيمة الوسيط بيانيا



V–4– خصائص الوسيط:

- يتميز الوسيط بعدم الثبات، ذلك أنه يتغير كلما تغيرت أطوال الفئات؛
- يتحدد الوسيط بعدد البيانات وليس بقيمتها؛
- يمكن حسابه في جداول التكرارات المفتوحة؛
- الوسيط لا يتأثر بالقيم المتطرفة (الشاذة).

VI–الربيعيات: إذا كان الوسيط يقسم البيانات إلى قسمين متساويين، فإن الربيعيات تقسم البيانات إلى أربع أجزاء متساوية. وإذا نستعمل نفس فكرة الوسيط عند حساب الربيعيات (الربيعي الأول، الربيعي الثالث) أما الرباعي الثاني فهو الوسيط، حيث أن مفهوم الربيعي هو تعميم لفكرة الوسيط. ونحسب رتبة الربيعي i بالعلاقة التالية:

$$R_{Q_i} = \frac{i\sum_{i=1}^{k} n_i}{4}$$

حيث: $i = 1,2,3$

VI–1– الربيع الأول: تعتبر قيمة الربيعي الأول هي قيمة المتغير الإحصائي التي تقع في المرتبة 25% $(\frac{\sum_{i=1}^{k} n_i}{4})$ وبرمز لها بالرمز Q_1 وهو يقسم المجتمع إلى قسمين، الأول يحتوي على 25% من الوحدات الإحصائية والثاني يحتوي على 75% من الوحدات الإحصائية.

VI–2– الربيع الثالث: هي قيمة المتغير الإحصائي التي تقع في المرتبة 75% $(\frac{3\sum_{i=1}^{k} n_i}{4})$ وهو يقسم المجتمع إلى قسمين يحتوي الأول على 75% من الوحدات الإحصائية ويحتوي الثاني 25% من الوحدات.

تختلف طرق حساب الربيعيات حسب نوع المتغير المدروس. حيث نفرق دائما بين البيانات غير المبوبة والبيانات المبوبة (المتغير الكمي المنفصل، والمتغير الكمي المتصل).

VI–3– حساب الربيعيات في حالة البيانات غير المبوبة: يوجد العديد من الطرق عند حساب الربيعيات في هذه الحالة، وسنقدم طريقة واحدة من بين هذه الطرق ذلك أن المجال لا يسع لذكرها كلها، وهي تعتبر طرق تقريبية حيث أنه عند تطبيق بعض هذه الطرق على نفس السلسلة الإحصائية يمكن أن نجد اختلاف طفيف في النتيجة. وهنا نميز بين حالتين:

- إذا كانت رتبة الربيعي عدد صحيح فإن قيمة الربيعي عبارة غن متوسط قيمتين وهي تساوي:

$$Q_i = \frac{x_i + x_{i+1}}{2}$$

- أما إذا كانت رتبة الربيعي عدد غير صحيح فإن قيمة الربيعي هي القيمة التي تلي رتبة الربيعي وتساوي:

$$Q_i = x_{i+1}$$

مثال (4–15): لتكن لدينا السلسلة الإحصائية التي التالية (n=12):

34 30 27 25 23 22 19 18 16 15 14 12

المطلوب: أحسب قيمة الربيعي الأول والثاني والثالث.

الحل:

– **الربيع الأول**: نلاحظ أن رتبة الربيع الأول عبارة عن عدد صحيح ($R_{Q_1} = \frac{12}{4} = 3$) ومنه قيمة الربيعي هي:

$$Q_1 = \frac{15 + 16}{2} = 15.5$$

– **الربيع الثاني**: رتبة الربيع الثاني هي: $R_{Q_2} = 2 * \frac{12}{4} = 6$

$$Q_2 = \frac{19 + 22}{2} = 20.5$$

– **الربيع الثالث**: رتبة الربيع الثالث هي: $R_{Q_3} = 3 * \frac{12}{4} = 9$

$$Q_3 = \frac{25 + 27}{2} = 26$$

مثال (4–16): لتكن لدينا السلسلة الإحصائية التي التالية (n=10):

27 25 23 22 19 18 16 15 13 12

المطلوب: أحسب قيمة الربيعي الأول والثاني والثالث.

الحل:

– **الربيع الأول**: نلاحظ أن رتبة الربيع الأول عبارة عن عدد غير صحيح ($R_{Q_1} = \frac{10}{4} = 2.5$) ومنه قيمة الربيعي هي:

$$Q_1 = X_3 = 15$$

– **الربيع الثاني**: رتبة الربيع الثاني هي: $R_{Q_2} = 2 * \frac{10}{4} = 5$

$$Q_2 = \frac{18+19}{2} = 18.5$$

- **الربيع الثالث**: رتبة الربيع الثالث هي: $R_{Q_3} = 3 * \frac{10}{4} = 7.5$

$$Q_3 = X_8 = 23$$

VI-4- حساب الربيعيات في حالة البيانات المبوبة: لحساب الربيعيات نتبع نفس خطوات حساب الوسيط في الحالتين (المتغير الكمي المنفصل والمتغير الكمي المتصل) الشيء الذي يتغير هو رتبة الربيعي (R_{Q_i})، ذلك أنه على أساس الرتبة تتحدد كل المعطيات الأخرى. كما نستعمل نفس علاقة الوسيط في حالة المتغير الكمي المتصل والتي تكتب بالصيغة التالية:

$$Q_i = A + \frac{\frac{i\sum_{i=1}^{k} n_i}{4} - F_{n-1}^{\uparrow}}{n_{Q_i}} * k$$

مثال (4-17): إذا أخذنا معطيات المثال (4-12) أحسب الربيع الأول والثالث.

الحل:

عدد الأطفال (X_i)	عدد العائلات (n_i)	ت.ت.ص
0	5	5
1	10	15
2	15	30
3	20	50
4	10	60
5	5	65
المجموع	65	/

- رتبة الربيع الأول: $R_{Q_1} = \frac{65}{4} = 16.25$ منه قيمة الربيع الأول هي قيمة x_i التي تقابل ت.ت.ص الذي هو مباشرة أكبر من 16.25 (30). وإذا $Q_1 = 2$
- رتبة الربيع الثالث: $R_{Q_3} = 3 * \frac{65}{4} = 48.75$ منه قيمة الربيع الثالث هي قيمة x_i التي تقابل ت.ت.ص الذي هو مباشرة أكبر من 48.75 (50). وإذا $Q_3 = 3$

مثال (4-18): نأخذ نفس معطيات المثال (4-13) ونريد حساب الربيع الأول والثالث.

الحل:

الفئات	التكرار (n_i)	ت.ت.ص $(F^{\uparrow})$	ت.ت.ن $(F^{\downarrow})$
$]10-5]$	12	12	77
$]15-10]$	20	32	65
$]20-15]$	25	57	45
$]25-20]$	15	72	20
$]30-25]$	5	77	5
المجموع	77	/	/

- رتبة الربيع الأول: $R_{Q_1}=\frac{77}{4}=19.25$ منه الفئة الربيعية الأولى هي $]15-10]$

وقيمة الربيع الأول هي:

$$Q_1=10+\frac{19.25-12}{20}*5=11.81$$

- رتبة الربيع الثالث: $R_{Q_3}=3*\frac{77}{4}=57.75$ منه الفئة الربيعية الثالثة هي: $]25-20]$

وقيمة الربيع الأول هي:

$$Q_3=20+\frac{57.75-57}{15}*5=20.25$$

ملاحظة: إضافة إلى الربيعات يمكن أن نستعمل العشيرات والمئويات وهي تحسب بنفس مبدأ الربيعيات.

VII– المنوال: يمثل المنوال قيمة المتغير الإحصائي الأكثر انتشارا أو الأكثر تكرارا، ويرمز له بالرمز M_o. تختلف طريقة حساب قيمة المنوال حسب نوع المتغير الإحصائي، حيث يؤخذ مباشرة من البيانات غير المبوبة والبيانات المبوبة حالة المتغير النوعي والمتغير الكمي المنقطع بينما يحسب بعلاقة رياضية في حالة المتغير الكمي المستمر.

VII–1– المنوال في حالة بيانات غير مبوبة: هي القيمة أو الصفة الأكبر تكرارا مقارنة بالقيم أو الصفات الأخرى.

مثال (4–19): لتكن لدينا السلسلة التالية:

7 10 9 10 11 9 10 15 10 6

المطلوب: حدد قيمة المنوال

الحل: نلاحظ من البيانات أن القيمة الأكثر انتشارا هي القيمة 10 ومنه $M_o = 10$.

VII-2- المنوال في حالة بيانات مبوبة (متغير كمي منقطع): هو عبارة عن قيمة المتغير الإحصائي التي يقابلها أكبر تكرار.

مثال (4-20): ليكن لدينا متغير إحصائي (X_i) يمثل عدد الغرف في البيت لعدد من العائلات القاطنة في حي سكني معين.

X_i	1	2	3	4	5
n_i	7	10	25	15	10

المطلوب: أوجد قيمة المنوال.

الحل: نلاحظ من خلال جدول التوزيع التكراري أن قيمة المتغير التي يقابلها أكبر تكرار هي القيمة 3 ومنه $M_o = 3$.

ملاحظة: يمكن أن يكون في السلسلة أكثر من المنوال وفي هذه الحالة تسمى السلسلة متعددة المنوال، كما يمكن أن لا تحتوي السلسلة على أي منوال وهنا تسمى عديمة المنوال.

VII-3- المنوال في حالة بيانات مبوبة (متغير كمي مستمر): نعتمد في حساب المنوال في حالة متغير كمي مستمر الخطوات التالية:

- تحديد الفئة المنوالية وهي الفئة التي تقابل أكبر تكرار؛
- حساب علاقة المنوال والتي تكتب على الشكل التالي:

$$M_o = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} * k$$

حيث Δ_1 : هو الفرق بين تكرار الفئة المنوالية وتكرار الفئة ما قبل المنوالية.

و Δ_2 : هو الفرق بين تكرار الفئة المنواية وتكرار الفئة ما بعد المنوالية.

A هي الحد الأدنى للفئة المنوالية و k تمثل طول الفئة المنوالية.

لا بد أن نشير هنا أنه في حال وجود فئات غير متساوية الأطوال نقوم بتعديل التكرارات قبل حساب المنوال وأخذ التكرار المعدل بدل التكرار العادي.

مثال (4-21): بأخذ معطيات المثال (4-13)، أوجد قيمة المنوال.

الفئات	التكرار (n_i)
$]10-5]$	12
$]15-10]$	20
$]20-15]$	25
$]25-20]$	15
$]30-25]$	5
المجموع	77

الحل: من خلال جدول التوزيع التكراري نلاحظ أن الفئات متساوية الطول ومنه نحسب المنوال مباشرة ونتبع الخطوات المذكورة أعلاه.

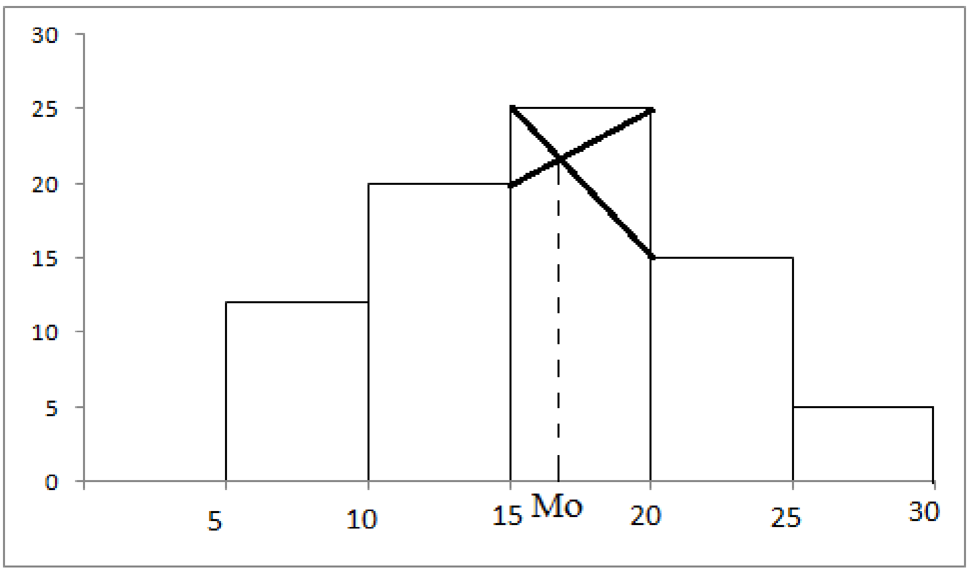
– تحديد الفئة المنوالية: الفئة المقابلة لأكبر تكرار وهي الفئة $]20-15]$ إذا المنوال ينتمي لهذه الفئة.

– حساب علاقة المنوال:

$$\mathrm{M_o} = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} * k = 15 + \frac{(25-20)}{(25-20)+(25-15)} * 5 = 16.666$$

VII–4– حساب المنوال بيانيا: نستطيع استخراج قيمة المنوال بيانيا من خلال رسم المدرج التكراري. في هذه الحالة قيمة المنوال هي عبارة عن فاصلة نقطة تقاطع القطعتين المستقيمتين، حيث القطعة الأولى التي تربط بين الحد الأدنى للفئة المنوالية والحد الأعلى للفئة السابقة والقطعة المستقيمة الثانية هي التي تربط بين الحد الأدنى للفئة المنوالية والحد الأعلى للفئة اللاحقة. نأخذ المثال (4–21).

التمثيل البياني رقم (12): حساب المنوال بيانيا



VII-5- خصائص المنوال:

- تتأثر قيمة المنوال بتكرار الفئة السابقة واللاحقة؛
- لا يتأثر بالقيم المتطرفة؛
- يمكن حسابه في جداول التوزيعات التكرارية المفتوحة؛
- يعتبر من أفضل المقاييس لوصف المتغير الكيفي.

VIII – العلاقة بين مقاييس النزعة المركزية: توصل العلماء إلى وضع علاقة بين المقاييس الثلاثة (المتوسط الحسابي والوسيط والمنوال).

$$\bar{X} - M_o = 3(\bar{X} - M_e)$$

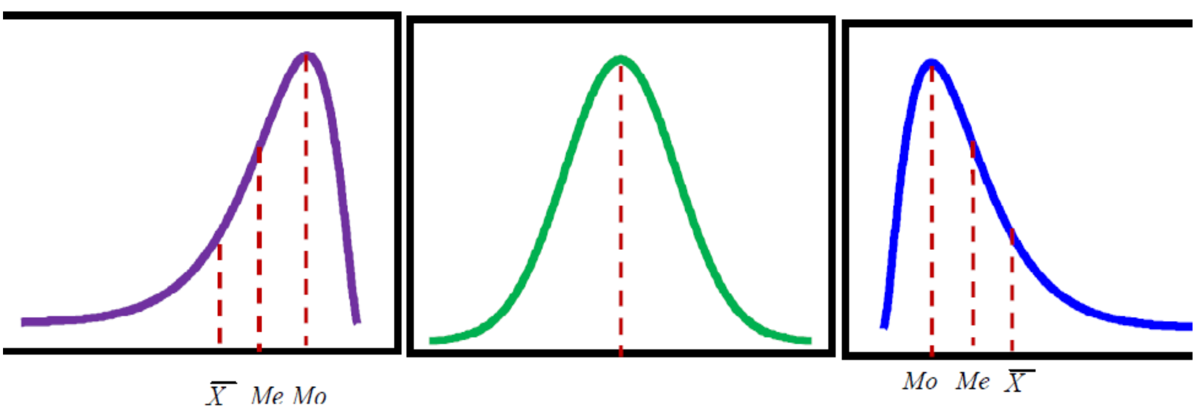
تكون هذه العلاقة صحيحة حسب KARL PEARSON ، إذا كان التوزيع المدروس قريب من المتناظر (المتماثل)، وتوجد ثلاثة أشكال للعلاقة التي تربط بين $\bar{X}$، M_e، M_o .

VIII-1- التوزيع المتماثل ($\bar{X} = M_e = M_o$): إذا كانت مقاييس النزعة المركزية متساوية نقول أن التوزيع التكراري متماثل، وفي هذه الحالة يكون المنحنى البياني لهذا الأخير يشبه شكل الجرس، ويسمى هذا التوزيع بالتوزيع الطبيعي.

VIII-2- التوزيع المائل نحو اليمين: إذا كانت مقاييس النزعة المركزية مرتبة على النحو ($\bar{X} > M_e > M_o$) فإن منحنى التوزيع التكراري يأتي ملتويا قليلا إلى اليمين.

VIII-3- التوزيع المائل نحو اليسار: إذا جاءت مقاييس النزعة المركزية مرتبة على النحو ($\bar{X} < M_e < M_o$) فإن منحنى التوزيع التكراري يأتي ملتويا قليلا إلى اليسار.

التوزيع يميل نحو اليمين **التوزيع المتماثل** **التوزيع يميل نحو اليسار**



تمارين محلولة:

التمرين الأول: تمثل البيانات الآتية توزيع 40 عاملا في ورشة البناء حسب عدد أيام العمل في شهر أ وت.

عدد أيام العمل (X_i)	10	15	18	20	22	المجموع
عدد العمال (n_i)	10	12	8	6	4	40

المطلوب:

-ماهو متوسط أيام العمل في هذه الورشة لشهر أوت.

-أحسب كل كم الوسيط والمنوال.

الحل:

1-حساب متوسط أيام العمل:

(X_i)	(n_i)	$n_i * X_i$	$F^{\uparrow}$
10	10	100	10
←15	12	180	22
18	08	144	30
20	06	120	36
22	04	88	40
المجموع	40	632	/

متوسط أيام العمل في الورشة هو:

$$\bar{X} = \frac{\sum_{i=1}^{5} n_i * x_i}{\sum_{i=1}^{5} n_i} = \frac{632}{40} = 15.8 \cong 16$$

اشتغل عمال الورشة في شهر أوت في المتوسط 16 يوما.

2-حساب الوسيط: بما أن المتغير المدروس عبارة عن متغير كمي منفصل سنتبع الخطوات التالية في تحديد قيمة الوسيط:

-نحدد رتبة الوسيط: $\frac{\sum_{i=1}^{5} n_i}{2} = \frac{40}{2} = 20$

-نلاحظ أن رتبة الوسيط غير موجودة في عمود التكرار المتجمع الصاعد وعليه نأخذ القيمة التي تأتي بعدها مباشرة وهي 22، ومنه تكون قيمة الوسيط المقابلة لهذه القيمة هي: $M_e = 15$.

3-حساب المنوال: بما أن أكبر تكرار هو 12، فإن قيمة المنوال هي قيمة المقابلة لهذا التكرار وعليه $M_o = 15$

التمرين الثاني: يبين الجدول التالي توزيع 100 أسرة حسب نفقاتها الشهرية (الوحدة 10^3دج).

النفقات	20-25	25-35	35-40	40-55	55-75	75-80	المجموع
التكرار	5	15	20	25	30	5	100

المطلوب:

-ما نوع المتغير المدروس؟

-ما هو متوسط الإنفاق الشهري لهذه الأسر؟

-حدد القيمة الوسيطية لهذه النفقات وكذا الربيعي الأول والثالث.

-حدد قيمة المنوال.

-حدد شكل التوزيع الإحصائي من خلال مقاييس النزعة المركزية.

الحل:

-المتغير المدروس: متغير كمي متصل.

لتسهيل العمليات الحسابية نستعين بالجدول.

X_i	n_i	x_i	$n_i x_i$	$F^{\uparrow}$	k	n^*
20-25	5	22,5	112,5	5	5	5
25-35	15	30	450	20	10	7,5
35-40	20	37,5	750	40	5	20
40-55	25	47,5	1187,5	65	15	8,33
55-75	30	65	1950	95	20	7,5
75-80	5	77,5	387,5	100	5	5
$\sum$	100	/	4837,5	/		/

ومنه متوسط الإنفاق الشهري للأسر هو:

$$\bar{X} = \frac{\sum_{i=1}^{6} n_i * x_i}{\sum_{i=1}^{6} n_i} = \frac{4837.5}{100} = 48.375 * 10^3 DA$$

-القيمة الوسيطية هي الوسيط ويحسب بالعلاقة التالية:

$$M_e = A + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - F_{n-1}^{\uparrow}}{n_{M_e}} * k = 40 + \frac{50-40}{25} * 15 = 46 * 10^3 DA$$

نقول أن: 50% من الأسر نفقاتهم أكبر من $10^3 * 46$ دج، و50% نفقاتهم أقل من $10^3 * 46$ دج.

-قيمة الربيعي الأول والثالث: العبارة العامة لحساب الربيعيات هي:

$$Q_i = A + \frac{\frac{i\sum_{i=1}^{k} n_i}{4} - F_{n-1}^{\uparrow}}{n_{Q_i}} * k$$

ومنه Q_1 يحسب كما يلي:

$$Q_1 = A + \frac{\frac{\sum_{i=1}^{k} n_i}{4} - F_{n-1}^{\uparrow}}{n_{Q_1}} * k = 35 + \frac{25-20}{20} * 5 = 36.250 * 10^3 DA$$

ونقول أن 25% من الأسر نفقاتهم $36.250 * 10^3$ دج.

ويحسب Q_3 كما يلي:

$$Q_3 = A + \frac{\frac{3\sum_{i=1}^{k} n_i}{4} - F_{n-1}^{\uparrow}}{n_{Q_3}} * k = 55 + \frac{75-65}{30} * 20 = 61.666 * 10^3 DA$$

ونقول أن 75% من الأسر نفقاتهم $61.666 * 10^3$ دج.

-تحديد قيمة المنوال: بما أن أطوال الفئات غير متساوية، فلا بد من تعديل التكرارات قبل حساب المنوال وذلك بالعبارة:

$$n_i^* = \frac{n_i}{K_i} * K_i^*$$

حيث طول الفئة المختار هو 5 (القاسم المشترك) وتحدد قيمة المنوال بالعلاقة التالية:

$$M_o = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} * k = 35 + \frac{(20-7.5)}{(20-7.5)+(20-8.33)} * 5 = 37.586 * 10^3 DA$$

ونقول أن الإنفاق الشهري الأكثر تكرارا بين الأسر هو $37.586 * 10^3$ دج.

-لتحديد شكل التوزيع من خلال مقاييس النزعة المركزية، نقوم بمقارنة المقاييس الثلاث.

نلاحظ أن مقاييس النزعة المركزية مرتبة على النحو ($\bar{X} > \mathrm{M_e} > \mathrm{M_o}$) ومنه فإن التوزيع مائل إلى اليمين.

الفصل الخامس

مقــاييــس التـشتت

بين مقاييس النزعة المركزية القيمة المركزية للتوزيع المدروس دون أن تظهر كيفية توزيع وانتشار الوحدات الإحصائية حولها أو تعطينا فكرة عن مدى تجانس أو عدم تجانس البيانات، إذا فهي لا تكفي لوصف وإجراء المقارنات بين التوزيعات التكرارية، بينما مقاييس التشتت تبين لنا كيفية انتشار قيم المتغير الإحصائي حول القيمة المركزية. عند إجراء مقارنة بين توزيعين تكراريين يمكن أن يتساوى متوسطهما الحسابي رغم أن انتشار البيانات في التوزيعان مختلف تماما وغير متجانس لهذا الغرض وجدت مقاييس التشتت حتى تعطينا فكرة عن مدى تباعد البيانات عن بعضها البعض.

مثال (5-1): إذا كان لدينا التوزيعان التاليان:

X:	80	100	140	180	200
Y:	120	125	140	150	165

وأردنا حساب المتوسط الحسابي للتوزيعين نلاحظ أن قيمتهما متساوية ($\bar{X} = \mathrm{M_e} = 140$)، غير أن السلسلة الأولى أكثر انتشارا من السلسلة الثانية، تساوي القيمة المركزية لا يعني بالضرورة أن التوزيعان الإحصائيان متماثلان، وبالتالي لدراسة ظاهرة معينة ولمعرفة درجة أهميتها لا نكتفي بدراسة خصائص النزعة المركزية فقط بل علينا دراسة تشتت التوزيع كذلك.

يقصد بالتشتت مدى تباعد قيم المتغير الإحصائي عن بعضها البعض أو عن القيمة المركزية وهي بذلك تعطينا فكرة عن مدى تجانس وتباين القيم، وتعتبر بيانات الظاهرة متجانسة إذا كانت قيمها قريبة من بعضها البعض ونقول في هذه الحالة أن توزيعها غير متشتت أما إذا كانت البيانات مختلفة وغير متجانسة نقول أن التوزيع متشتت. ولقياس تشتت توزيع معين نفرق بين مقاييس التشتت المطلقة والتي تقيس مقدار التشتت حول القيمة المركزية للظواهر التي لها نفس وحدة القياس ومقاييس التشتت النسبية التي تقيس مقدار التشتت حول القيمة المركزية للظواهر التي لها وحدات قياس مختلفة. كما أن هناك بعض المقاييس تقيس تقارب أو تباعد القيم من القيمة المتوسطة والبعض الأخر يقيس تقارب أو تباعد القيم عن بعضها البعض.

I- مقاييس التشتت المطلقة: تقيس مقدار التشتت حول القيمة المركزية للظواهر التي لها نفس وحدة القياس.

I -1- المدى العام: عبارة عن مقياس عام وسريع لتشتت التوزيع، فهو يستعمل لإعطاء فكرة سريعة عن مدى تقارب أو تباعد المفردات. المدى العام هو الفرق بين أكبر قيمة وأصغر قيمة في سلسلة البيانات ويرمز له بالرمز E. ويعطى بالعلاقة التالية:

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

خواصه:

- بسيط وسهل الحساب إلا أنه يعتمد على القيمتين المتطرفتين فقط؛
- يتأثر بالقيم المتطرفة؛
- يستعمل للمقارنة بين توزيعين إحصائيين أو أكثر.

مثال (5–2): أحسب المدى العام باستعمال معطيات المثال (5-1).

الحل:

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

$$E_X = 200 - 80 = 120$$

$$E_Y = 165 - 120 = 45$$

بالرغم من أن القيمة المركزية هي نفسها بالنسبة للتوزيعين، إلا أن قيمة المدى العام تختلف ونلاحظ أن ($E_X > E_Y$) وبالتالي نقول أن التوزيع الثاني أقل تشتت من التوزيع الأول.

I –2– المدى الربيعي: هو عبارة عن الفرق بين الربيعي الثالث والربيعي الأول، فهو بذلك يهمل الربع الأول والربع الأخير من البيانات المرتبة ترتيبا تصاعديا؛ وفي هذه الحالة تكون أكبر قيمة هي الربيعي الثالث وأصغر قيمة في البيانات هي الربيعي الأول، فيعطينا فكرة عن المجال الذي تنتشر فيه نصف عدد البيانات، يستخدم المدى الربيعي في حالة وجود قيم متطرفة جدا أو جداول توزيعات تكرارية مفتوحة، ويرمز له بالرمز (IQ) ويعطى بالصيغة التالية:

$$IQ = Q_3 - Q_1$$

خواصه:

- يستعمل للمقارنة بين توزيعين إحصائيين أو أكثر؛
- يضم 50%من المجتمع؛
- استعمالاته محدودة غير أنه أحسن من المدى العام؛
- لا يتأثر بالقيم المتطرفة؛
- يمكن حسابه في جداول التوزيعات التكرارية المفتوحة.

مثال (3-5): أحسب المدى الربيعي لمعطيات المثال (4-13).

الحل: قمنا سابقا (المثال (4-18)) بحساب الربيع الأول والربيع الثالث ومنه فإن المدى الربيعي هو:

$$IQ = Q_3 - Q_1 = 20.25 - 11.81 = 8.44$$

I – 3- الانحراف الربيعي: هو نصف الفرق بين الربيعي الثالث والربيعي الأول، و هو عبارة عن نصف المدى الربيعي و يرمز له بالرمز (Q). المهم في هذا المقياس هو مدلوله الإحصائي، حيث يعتبر مقياسا لمراكز القيم وهو أفضل من المدى لأنه لا يتأثر بالقيم المتطرفة. تعطى علاقة الانحراف الربيعي كما يلي:

$$Q = \frac{\mathrm{IQ}}{2} = \frac{\mathrm{Q}_3 - \mathrm{Q}_1}{2}$$

مثال (4-5): تبين لنا السلسلة التالية مداخيل عشرة عمال في شركة ما.

11000 12000 13000 15000 15000 16000 16000 17000 18000 20000

المطلوب: أحسب الانحراف الربيعي.

الحل:

$$Q = \frac{\mathrm{Q}_3 - \mathrm{Q}_1}{2} = \frac{17000 - 13000}{2} = 2000$$

مدلول هذه القيمة هو أن 50% من المداخيل تبعد (تنحرف) في المتوسط عن الوسيط بـ 2000 و.ن.

I – 4-الانحراف المتوسط: يفهم من كلمة انحراف، الفرق بين قيمة المتغير الإحصائي وقيمة مقياس من مقاييس النزعة المركزية. في كثير من الأحيان نأخذ بعين الاعتبار المتوسط الحسابي.

ومنه الانحراف المتوسط عبارة عن المتوسط الحسابي لهذه الانحرافات بالنسبة لـ X بالقيمة المطلقة، لأن مجموع الانحرافات بالنسبة للمتوسط الحسابي يساوي الصفر ($\sum_{i=1}^{n}(X_\mathrm{i} - \overline{\mathrm{X}}) = 0$). يرمز للانحراف المتوسط بالرمز ($\mathrm{E}_{\overline{\mathrm{x}}}$) ويعطى بالعلاقة التالية:

$$\mathrm{E}_{\overline{\mathrm{X}}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{X}_\mathrm{i} - \overline{\mathrm{X}}|}{\mathrm{n}}$$

في حالة بيانات غير مبوبة.

أما في حالة بيانات مبوبة تكتب العلاقة على الشكل التالي:

$$\mathrm{E}_{\overline{\mathrm{X}}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{n}_\mathrm{i} |\mathrm{X}_\mathrm{i} - \overline{\mathrm{X}}|}{\mathrm{n}}$$

مثال (5-5) : لتكن لدينا السلسلة التالية ونريد حساب الانحراف المتوسط.

1	2	4	5	7	8	9	10	11	13

الحل:

نحسب أولا المتوسط الحسابي: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{70}{10} = 7$

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

$$E_{\bar{X}} = \frac{|1-7| + |2-7| + |4-7| + |5-7| + |7-7| + |8-7| + |9-7| + |10-7| + |11-7| + |13-7|}{10}$$

$$E_{\bar{X}} = \frac{32}{10} = 3.2$$

خواصه:

- يأخذ بعين الاعتبار كل القيم؛
- يتأثر بالقيم المتطرفة، فهو يعتمد في حسابه على جميع القيم؛
- يتعامل مع القيم المطلقة وتعتبر هذه الخاصية من عقبات هذا المقياس لأنه لا يفرق بين القيم التي تكون أكبر من المتوسط والقيم التي تكون أقل من المتوسط الحسابي؛
- يعتبر أحسن مقياس مقارنة بالمدى العام؛
- يمكن حسابه عن طريق الانحراف المتوسط عن الوسيط، إلا أن قيمة الانحراف المتوسط عن الوسيط أقل من قيمة الانحراف المتوسط عن المتوسط الحسابي.

I-5- التباين والانحراف المعياري: التباين عبارة عن مجموع انحرافات قيم المتغير الإحصائي بالنسبة للمتوسط مقسومة على عدد المشاهدات والانحراف المعياري هو الجذر لتربيعي للتباين. ويرمز للتباين بالرمز (σ^2) أو ($V(X)$) في المجتمع الإحصائي أما في العينة يرمز له بالرمز (S^2).

I-5-1 التباين:

I-5-1-1- في حالة بيانات غير مبوبة :

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

كما يمكن كتابة صيغة التباين على الشكل التالي:

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

I-5-1-2- في حالة بيانات مبوبة:

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

وتكتب كذلك:

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

ملاحظة: في حال أردنا حساب التباين في العينة نستعمل القيمة المصححة وهي مجموع انحرافات قيم المتغير الإحصائي بالنسبة للمتوسط مقسومة على عدد المشاهدات ناقص واحد.

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

I-5-2- الانحراف المعياري: عبارة عن الجذر التربيعي للتباين وهو من أهم وأكثر مقاييس التشتت استعمالا في مجال الإحصاء، يرمز له بالرمز (σ) في المجتمع الإحصائي و (S) في العينة.

$$\sigma = \sqrt{V(x)}$$

خواص الانحراف المعياري:

- هو أكثر مقاييس التشتت استعمالا، لأنه يستعمل في حساب عدة مؤشرات منها:معامل الارتباط، تحديد أشكال التوزيعات الإحصائية، ...الخ؛
- يستعمل للمقارنة بين توزيعين تكراريين أو أكثر؛
- يتأثر بالقيم المتطرفة،
- لا يمكن حسابه في جداول التوزيعات التكرارية المفتوحة؛
- يمكن الاعتماد عليه عند مقارنة تشتت توزيعين أو أكثر بشرط أن يكون لهم نفس المتوسط الحسابي.

ملاحظة: كلما كان الانحراف المعياري صغير، كلما كان التوزيع أو النوع جيدا ونقول أن القيم ليست متباعدة عن وسطها الحسابي وبالتالي فهو أقل تشتتا.

مثال (5-6): لتكن لدينا التوزيع التالي:

9 7 6 5 4 4 3 2

المطلوب: أوجد تباين هذا التوزيع ثم انحرافه المعياري.

الحل:

- المتوسط الحسابي:

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

$$\bar{X} = \frac{2+3+4+4+5+6+7+9}{8} = 5$$

- التباين:

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

$$V(X) = \frac{1}{8}[(2-5)^2 + (3-5)^2 + (4-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2 + (9-5)^2]$$

$$V(X) = \frac{1}{8}[9+4+1+1+0+1+4+16] = \frac{36}{8} = 4.5$$

- العلاقة الثانية للتباين:

$$V(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$V(X) = \frac{1}{8}(2^2 + 3^2 + 4^2 + 4^2 + 5^2 + 6^2 + 7^2 + 9^2) - 5^2$$

$$V(X) = \frac{1}{8}(4+9+16+16+25+36+49+81) - 25$$

$$= \frac{236}{8} - 25 = 4.5$$

- الانحراف المعياري:

$$\sigma = \sqrt{V(x)}$$

$$\sigma = \sqrt{4.5} = 2.12$$

ملاحظة: إذا كان المتغير المدروس متغير كمي متصل فإنه يتم تعويض قيم (X_i) بمراكز الفئات (x_i).

II – مقاييس التشتت النسبية: وهي تقيس مقدار التشتت حول القيمة المركزية للظواهر التي لها وحدات قياس مختلفة لأنها مقاييس خالية من القياس، أو ليس لها نفس المتوسط الحسابي كما أنها تستخدم لمقارنة تشتت أو تجانس بين مجموعتين من البيانات.

II – 1– معامل الاختلاف: هو النسبة بين الانحراف المعياري والوسط الحسابي، يستعمل معامل الاختلاف في المقارنة بين التوزيعات التكرارية ويرمز له بالرمز (CV). كلما كان معامل الاختلاف كبيرا، كلما دل ذلك على قوة التشتت بين وحدات الظاهرة وكلما كان صغيرا، كلما دل ذلك على تجانس مفردات الظاهرة المدروسة. ويحسب بالعبارة التالية:

$$CV = \frac{\sigma}{\bar{X}} * 100$$

مثال (5–7): يتم في أحد المصانع إنتاج نموذجين من العجلات (A) و (B) المدة المتوسطة لاستعمال النموذج الأول هي 10.000 كلم/سا بانحراف معياري قدره 2.000 كلم/سا. أما النوع الثاني فتقدر المدة المتوسطة لاستعماله بـ 11.000 كلم/سا بانحراف معياري قدره 3.000كلم/سا. ما هو أحسن نموذج ؟

الحل: لمعرفة أحسن نموذج تقوم بحساب معامل اختلاف النموذجين.

$$CV_A = \frac{\sigma}{\bar{X}} * 100 = \frac{2.000}{10.000} * 100 = 20\%$$

$$CV_B = \frac{\sigma}{\bar{X}} * 100 = \frac{3.000}{11.000} * 100 = 27\%$$

بمقارنة معامل اختلاف المجموعتين نلاحظ أن النموذج (A) أقل تشتتا من النموذج (B)، فهو بذلك يعتبر أكثر تجانسا من التوزيع الثاني وقيمه قريبة من المتوسط.

II – 2– معامل الاختلاف الربيعي: يستعمل معامل الاختلاف في حالة البيانات المغلقة أو التوزيعات التكرارية المغلقة من الجهتين أما في حالة توزيع تكراري مفتوح نستعمل معامل الاختلاف الربيعي، يرمز له بالرمز ($\acute{CV}$) ويعطى بالعبارة التالية:

$$\acute{CV} = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

مثال (5–8): إذا أخذنا معطيات المثال (3–12) أوجد معامل الاختلاف هذا التوزيع.

الحل: بعد حساب الربيعي الأول والربيعي الثالث تحصلنا على النتائج التالية:

$Q_3 = 20.25$ و $Q_1 = 11.81$

$$\acute{CV} = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

$$= \frac{20.25 - 11.81}{20.25 + 11.81} * 100 = 26.32\%$$

نقول في هذه الحالة أن التشتت ضعيف (أقل من 50%) ومنه فإن التوزيع متجانس وقيم المتغير قريبة من المتوسط.

III – العلاقة بين الانحراف المعياري والمدى الربيعي والانحراف المتوسط:

لتحديد هذه العلاقة نفرض أن التوزيع الإحصائي المدروس يكون قريب من التماثل بالنسبة للقيمة المركزية.

III –1– العلاقة بين الانحراف المعياري والمدى الربيعي: انطلاقا من العلاقة الموجودة بين المجال $[Q_1, Q_3]$ والمجال $[\bar{X} - 0.67\sigma, \bar{X} + 0.67\sigma]$ حيث يضم كل منهما على50% من الوحدات الإحصائية المركزية، وبعد المطابقة بينهما فإن:

- الحد الأدنى للمجال الأول يساوي الحد الأدنى للمجال الثاني.
- الحد الأعلى للمجال الأول يساوي الحد الأعلى للمجال الثاني.

ومنه فإن:

$$\begin{cases} Q_1 = \bar{X} - 0.67\sigma \\ Q_3 = \bar{X} + 0.67\sigma \end{cases} \Rightarrow Q_3 - Q_1 = 1.34\sigma$$

III –2– العلاقة بين الانحراف المعياري والانحراف المتوسط: تعبر هذه العلاقة تجريبية، حيث كلما كان التوزيع الإحصائي أقرب إلى التناظر أو التماثل كلما كانت العلاقة صحيحة. وهي عبارة عن أربعة أخماس الانحراف المعياري وتكتب بالصيغة التالية:

$$E_{\bar{X}} = \frac{4}{5}\sigma$$

III –3– العلاقات أخرى تخص الانحراف المعياري: يستعمل الانحراف المعياري لتحديد مجالات الثقة وهذا إذا فرضنا أن التوزيع التكراري المدروس توزيع متماثل أو متناظر أو قريب من التناظر، فهو يستعمل في تحديد نسب عدد الوحدات الإحصائية حسب مجالات ثقة معطاة، بحيث يعطي كل مجال نسبة توزيع البيانات في المجتمع المدروس.

- **المجال الأول:** $[\bar{X} \pm 0.67\sigma]$ يحتوي على 50% من المجتمع.
- **المجال الثاني:** $[\bar{X} \pm \sigma]$ يحتوي على68,27% من المجتمع.

- **المجال الثالث**: $[\bar{X} \pm 2\sigma]$ يحتوي على 95,45% من المجتمع.
- **المجال الرابع**: $[\bar{X} \pm 3\sigma]$ يحتوي على 99,73% من المجتمع.

تمارين محلولة

التمرين الأول: لتكن لدينا السلسلة الإحصائية التالية:

33	24	19	39	48	45	26	35	38	25

المطلوب: أوجد كل من:

- المدى العام - المدى الربيعي - نصف المدى الربيعي - التباين والانحراف المعياري.

-قس تشتت هذه السلسلة وماذا تلاحظ؟

الحل:

-المدى العام:

$$E = X_{max} - X_{min} = 48 - 19 = 29$$

- المدى الربيعي: يحسب المدى الربيعي بدلالة الربيعيان الأول والثالث، لأجل ذلك نقوم بترتيب السلسلة ترتيبا تصاعديا أولا.

48	45	39	38	35	33	26	25	24	19

رتبة الربيعي الأول عدد غير صحيح ($\frac{10}{4} = 2.5$) فإن قيمة الربيعي هي القيمة التي تلي رتبة الربيعي وهي: $Q_1 = 25$.

رتبة الربيعي الثالث عدد غير صحيح ($\frac{3*10}{4} = 7.5$) فإن قيمة الربيعي هي القيمة التي تلي رتبة الربيعي وهي: $Q_3 = 39$.

ومنه:

$$IQ = Q_3 - Q_1 = 39 - 25 = 14$$

-نصف المدى الربيعي:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{14}{2} = 7$$

-التباين والانحراف: يحسب التباين بدلالة المتوسط الحسابي ولذلك علينا حساب المتوسط الحسابي لهذه السلسلة أولا.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{19+24+25+26+33+35+38+39+45+48}{10} = 33.2$$

$$V(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$= \frac{1}{10}\sum_{i=1}^{10}(19^2 + 24^2 + 25^2 + 26^2 + 33^2 + 35^2 + 38^2 + 39^2 + 45^2 + 48^2) - 33.2^2 = 82.36$$

$$\sigma = \sqrt{V(x)} = \sqrt{82.36} = 9.07$$

-لقياس تشتت هذا التوزيع نحسب معامل الاختلاف.

$$CV = \frac{\sigma}{\bar{X}} * 100 = \frac{9.07}{33.2} * 100 = 27.32\%$$

نلاحظ أن أقل من 50% ومنه نقول أن تشتت هذا التوزيع ضعيف.

التمرين الثاني: إذا كان لديك توزيع علامات الطلبة في مقياس الإحصاء بمتوسط 14 بانحراف معياري قدره 1.8، متوسط علامات الرياضيات 10 بانحراف 1.4. أي التوزيعين أكثر تشتتا؟

الحل:

إذا استخدمنا الانحراف المعياري كمقياس للمقارنة نقول أن توزيع علامات الرياضيات أكثر تجانسا (أقل تشتتا) من توزيع علامات الإحصاء ولكن بما أن المتوسط الحسابي الذي حسب من حوله الانحراف المعياري مختلف فإن المقارنة في هذه الحالة لا تصح ويجب البحث عن مقياس آخر يوحد المقارنة وهو معامل الاختلاف كما يلي:

$$CV_S = \frac{\sigma}{\bar{X}} * 100 = \frac{1.8}{14} * 100 = 12.86\%$$

$$CV_M = \frac{\sigma}{\bar{X}} * 100 = \frac{1.4}{10} * 100 = 14\%$$

ومن هنا تبين أن توزيع علامات الرياضيات أكثر تشتتا من توزيع علامات الإحصاء.

التمرين الثالث: أخذت عينتان من مجتمعين، فأعطتا النتائج التالية:

I	II
$\sum_{i=1}^{50} X_i = 300$	$\sum_{i=1}^{40} Y_i = 280$
$\sum_{i=1}^{50} X_i^2 = 1950$	$\sum_{i=1}^{40} Y_i^2 = 2080$

المطلوب: - أوجد الوسط الحسابي والانحراف المعياري لكل عينة.

- أي العينتين أكثر تجانسا (أقل تشتتا).
- دمجت العينتان، فما هو الوسط الحسابي والانحراف المعياري للمجموعة الناتجة.

الحل:

-حساب الوسط الحسابي والانحراف المعياري لكل عينة:

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{300}{50} = 6$; $V(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = 1950 - 6^2 = 3$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{280}{40} = 7$; $V(Y) = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 = 2080 - 7^2 = 3$

-حساب معامل الاختلاف لكل عينة:

$$CV_1 = \frac{\sigma}{\bar{X}} * 100 = \frac{1.73}{6} * 100 = 28.83\%$$

$$CV_2 = \frac{\sigma}{\bar{Y}} * 100 = \frac{1.73}{7} * 100 = 24.71\%$$

نلاحظ أن معامل اختلاف العينة الثانية أصغر من معامل اختلاف العينة الأولى ومنه العينة الثانية أكثر تجانسا (أقل تشتتا) من العينة الأولى.

-حساب الوسط الحسابي والانحراف المعياري بعد دمج العينتين:

نضع: $Z = X + Y$ ، إذا $\bar{Z} = \bar{X} + \bar{Y}$ ومنه :

$$\bar{Z} = \bar{X} + \bar{Y} = 6 + 7 = 11$$

$$\sigma_Z = \sigma_X + \sigma_Y = 1.73 + 1.73 = 3.46$$

التمرين الرابع: إليك بيانات السلسلة التالية:

18 17 18 14 13 15 16 2 16 17 15 17 10 13 12

المطلوب:

-ضع بيانات السلسلة في جدول توزيع تكراري وأحسب المتوسط الحسابي.

-ما هي قيمة المدى العام؟ وما هي قيمة التشتت النسبي المقابلة لها؟ هل يمكن الاعتماد على هذا المقياس ولماذا؟

-قس تشتت هذا التوزيع باستعمال المدى الربيعي.

الحل:

X_i	2	10	12	13	14	15	16	17	18	المجموع
n_i	1	1	1	2	1	2	2	3	2	15
n_iX_i	2	10	12	26	14	30	32	51	36	213
$F^{\uparrow}$	1	2	3	5	6	8	10	13	15	/

-المتوسط الحسابي:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i = \frac{213}{15} = 14.2$$

-المدى العام والتشتت النسبي للمدى العام:

$$E = X_{max} - X_{min} = 18 - 2 = 16$$

$$e_{\bar{X}} = \frac{E}{\bar{X}} * 100 = \frac{16}{14.2} * 100 = 112$$

نلاحظ أن قيمة التشتت النسبي كبيرة جدا، كما أنه لا يمكن الاعتماد على المدى العام في قياس التشتت لأنه لا يأخذ بعين الاعتبار كل البيانات.

-حساب المدى الربيعي:

$$I_Q = Q_3 - Q_1$$

نعين قيمة الربيع الأول:

$$\frac{\sum_{i=1}^{n} n_i}{4} = \frac{15}{4} = 3.75$$

نبحث عن هذه القيمة في السطر الرابع أو القيمة التي هي أكبر منها ومن تم فإن قيمة الربيع الأول هي قيمة X_i المقابلة لها في السطر الأول، ومنه $Q_1 = 13$

نعين قيمة الربيع الثالث:

$$\frac{3\sum_{i=1}^{n} n_i}{4} = \frac{3 * 15}{4} = 11.25$$

نبحث عن هذه القيمة في السطر الرابع أو القيمة التي هي أكبر منها ومن تم فإن قيمة الربيع الأول هي قيمة X_i المقابلة لها في السطر الأول، ومنه $Q_3 = 17$

$$I_Q = Q_3 - Q_1 = 17 - 13 = 4$$

الفصل السادس

مقـــاييـــس الشكــــل

إذا كانت مقاييس النزعة المركزية تسمح بتلخيص بيانات الظاهرة المدروسة في صورة رقمية، ومقاييس التشتت تسمح بإعطاء فكرة عن درجة تجانس أو تباعد هذه البيانات عن بعضها البعض أو عن القيمة المركزية فإن مقاييس الشكل تعطي فكرة عن شكل توزيع بيانات الظاهرة وانتشارها على شكل منحنى بياني من حيث تطاوله أو التوائه عن الشكل الطبيعي ومقارنة هذا الأخير بالتوزيع المرجعي (التوزيع الطبيعي) وذلك عن طريق حساب مقاييس التفرطح والالتواء.

وقبل التطرق لكيفية حساب مقاييس الشكل فإنه من الضروري التعرف على مقاييس إحصائية تسمح لنا بتحديد كل من معامل الالتواء والتفرطح والتي تعرف باسم العزوم ذلك أن مقاييس الشكل تعتمد في حسابها على العزوم.

I – العزوم:

قد تكون العزوم حول نقطة الأصل أو حول المتوسط أو حول أي نقطة معينة، أما رتبة العزم فتحدد بدرجة القوة (الأس) التي ترفع إليها القيم أو انحرافاتها عن المتوسط الحسابي، ونميز هنا بين نوعان من العزوم البسيطة والمركزية.

I –1– العزوم البسيطة: إذا كانت الظاهرة المدروسة (X) تأخذ القيم الإحصائية $x_1, x_2, \dots x_i, \dots, x_n$ فإن العزم البسيط من الدرجة k لهذا المتغير الإحصائية والذي يرمز له بالرمز m_k .

I –1–1– العزوم البسيطة في حالة بيانات غير مبوبة: يعرف العزم البسيط من الدرجة k في هذه الحالة بالعلاقة التالية:

$$m_k = \frac{\sum_{i=1}^{n} x_i^k}{n} \qquad k = 1,2,3, \dots \dots$$

ومنه فإن:

$$m_0 = 1$$

$$m_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$$

$$m_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

I –1–2– العزوم البسيطة في حالة بيانات مبوبة: يعطى العزم البسيط من الدرجة k في هذه الحالة بالصيغة التالية:

$$m_k = \frac{\sum_{i=1}^{n} n_i x_i^k}{\sum_{i=1}^{n} n_i} \qquad k = 1,2,3, \dots \dots ..$$

I –2– العزوم المركزية: إذا كانت الظاهرة المدروسة (X) تأخذ القيم الإحصائية $x_1, x_2, \dots x_i, \dots, x_n$ فإن العزم المركزي من الدرجة k لهذا المتغير الإحصائية والذي يرمز له بالرمز μ_k .

I –2–1– العزوم المركزية في حالة بيانات غير مبوبة: يعطى العزم المركزي من الدرجة k في وجود بيانات غير مبوبة كما يلي:

$$\mu_k = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^k}{n} \qquad k = 1,2,3, \dots \dots$$

وبالتالي فإن قيم العزوم المركزية الأربعة الأولى تكون كما يلي:

$$\mu_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})}{n} = 0$$

$$\mu_2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} = \sigma^2$$

$$\mu_3 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{n}$$

$$\mu_4 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^4}{n}$$

I –2–2– العزوم المركزية في حالة بيانات مبوبة: يعطى العزم المركزي من الدرجة k في وجود بيانات مبوبة كما يلي:

$$\mu_k = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^k}{\sum_{i=1}^{n} n_i} \qquad k = 1,2,3, \dots \dots$$

I –3– العلاقة بين العزوم المركزية والعزوم البسيطة: يمكن استغلال العلاقة بين العزوم المركزية والعزوم البسيطة وذلك قصد تسهيل العمليات الحسابية والتي تعطى على النحو التالي:

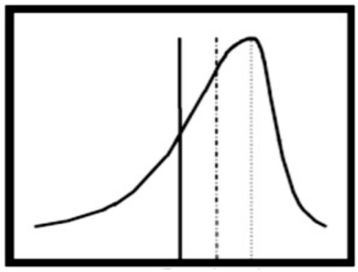
$$\mu_2 = \sigma^2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3$$

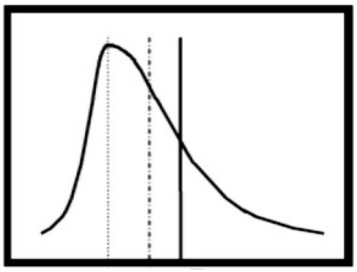
$$\mu_4 = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4$$

II – الالتواء: من أهم الأشكال التي يمكن أن يخذها أي توزيع تكراري وفقا لتوزع البيانات حول قيمة مركزية معينة غالبا ما تكون المتوسط الحسابي هي:

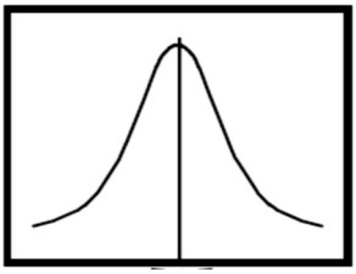
- **توزيع متماثل (متناظر):** إذا كان $\bar{X} = \mathrm{M_e} = \mathrm{M_o}$ بمعنى أن تكون 50% من القيم على يمين المتوسط و 50% على يساره، وهو يعتبر شكلا معياريا طبيعيا تقاس بالنسبة له كل الأشكال المتبقية (كما يسمى توزيع جرسي).
- **توزيع موجب الالتواء:** إذا كان $\bar{X} > \mathrm{M_e} > \mathrm{M_o}$ أي يكون التوزيع ممتد أكثر نحو اليمين.
- **توزيع سالب الالتواء:** إذا كان $\bar{X} < \mathrm{M_e} < \mathrm{M_o}$ فيأتي التوزيع ممتدا أكثر نحو اليسار.



$\bar{X} < \mathrm{M_e} < \mathrm{M_o}$



$\bar{X} > \mathrm{M_e} > \mathrm{M_o}$



$\bar{X} = \mathrm{M_e} = \mathrm{M_o}$

من خلال الأشكال الثلاثة، يمكن تعريف الالتواء على أنه بعد المنحنى عن التماثل وهو إما موجب أو سالب كما يمكن معرفة درجة التوائه (حاد أو بسيط) إلا أن هذا لا يعطينا قياسا رقميا للالتواء ولهذا نقوم بحساب معامل الالتواء الذي يقاس بعدة علاقات.

II –1– معامل فيشر للالتواء: ويطلق عليه كذلك معامل الالتواء العزمي، يعبر عنه رياضيا بالعلاقة التالية:

$$\gamma_F = \frac{\mu_3}{\sigma^3}$$

II –2– معامل بيرسون للالتواء: يمكن حساب معامل بيرسون للالتواء بطريقتين تقريبيتين وهما:

II –2–1– معامل بيرسون الأول للالتواء: يعطى معامل بيرسون الأول للالتواء بالعلاقة التالية:

$$P_1 = \frac{\bar{\mathrm{X}} - \mathrm{M_o}}{\sigma}$$

II –2–2– معامل بيرسون الثاني للالتواء: ويعطى بالعلاقة التالية:

$$P_2 = \frac{3(\bar{\mathrm{X}} - \mathrm{M_e})}{\sigma}$$

II –3– معامل يول للالتواء: ويسمى كذلك معامل الالتواء الربيعي، يستعمل هذا الأخير في قياس الالتواء في التوزيعات المفتوحة. ويحسب بالعلاقة التالية:

$$\gamma = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

وبما أن قيمة معامل التواء التوزيع المتماثل تساوي الصفر، فإنه عندما يكون المنحنى سالب الالتواء فإن إشارة المعامل تكون سالبة وعندما يكون المنحنى موجب الالتواء فإن إشارته تكون موجبة. ومنه يتم تحديد شكل التوزيع كما يلي:

إذا كان: $P = 0$ أو $\gamma = 0$ فإن منحنى التوزيع يكون متماثلا.

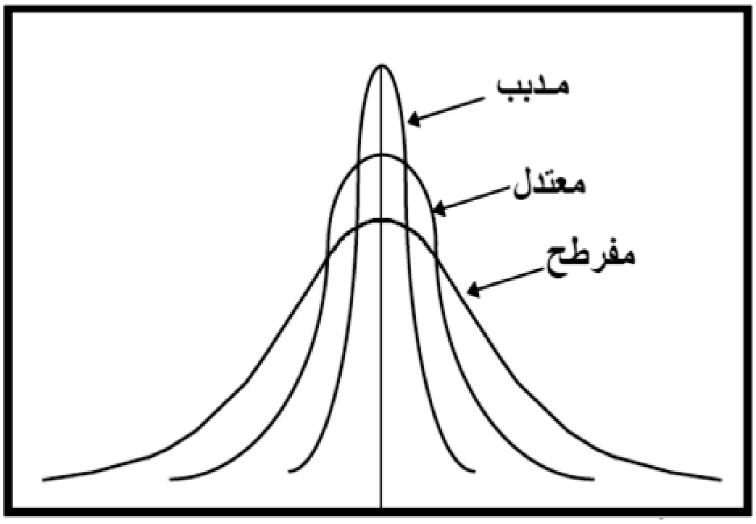
إذا كان: $P > 0$ أو $\gamma > 0$ فإن منحنى التوزيع مائل لليمين.

إذا كان: $P < 0$ أو $\gamma < 0$ فإن منحنى التوزيع مائل لليسار.

III – التفرطح: يقصد به قياس درجة ارتفاع قمة التوزيع، فإذا حدثنا عن شكل التفرطح نقول أن التوزيع التكراري قد تكون:

- متماثل أو معتدل القمة.
- مدبب القمة إذا كانت قمته أعلى من القمة المعتدلة.
- مفرطح القمة إذا كانت قمته أقل من القمة المعتدلة.

والتمثيل البياني التالي يبين ذلك:



وهو يقيس درجة علو قمة التوزيع مقارنة بالتوزيع المعياري (المتماثل) فكلما كان الشكل أكثر ارتفاعا نقول أن الشكل مدبب وكلما كان أقل ارتفاعا من الشكل الطبيعي نقول أنه مفرطح. ويقاس التفرطح معامل لتفرطح.

II –1– معامل بيرسون للتفرطح: حيث يعطى بالعلاقة التالية:

$$\beta_P = \frac{\mu_4}{\mu_2^2}$$

حيث يتم تحديد نوع الشكل كما يلي:

إذا كان: $\beta_P = 3$ فإن منحنى التوزيع معتدل (طبيعي).

إذا كان: $\beta_P > 3$ فإن منحنى التوزيع مدبب.

إذا كان: $\beta_P < 3$ فإن منحنى التوزيع مفرطح.

II –1– معامل فيشر للتفرطح: هو عبارة عن معامل بيرسون للتفرطح مطروح منه القيمة النظرية (3)ويعطى بالعلاقة التالية:

$$\beta_F = \frac{\mu_4}{\mu_2^2} - 3$$

ويتم تحديد شكل التوزيع كما يلي :

إذا كان: $\beta_F = 0$ فإن منحنى التوزيع معتدل (طبيعي).

إذا كان: $\beta_F > 0$ فإن منحنى التوزيع مدبب.

إذا كان: $\beta_F < 0$ فإن منحنى التوزيع مفرطح.

تمارين محلولة

التمرين الأول: تمثل البيانات التالية معدلات مجموعتين من طلبة السنة الأولى جذع مشترك في كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير بعد المداولات النهائية.

المطلوب: حدد شكل توزيع معدلات الطلبة في كل مجموعة.

المعدلات	المجموعة الأولى n_i	المجموعة الثانية n_i
6.90-6.30	03	01
7.50-6.90	07	04
8.10-7.50	10	08
8.70-8.10	13	11
9.30-8.70	15	13
9.90-9.30	16	15
10.50-9.90	15	16
11.10-10.50	13	16
11.70-11.10	10	15
12.30-11.70	07	12
12.90-12.30	03	01
المجموع	112	112

الحل: نقوم أولا بحساب مقاييس النزعة المركزية الثلاث لكل مجموعة.

المجموعة الأولى:

نستعين بجدول التوزيع التكراري لتسهيل الحسابات.

المعدلات	n_i	x_i	$n_i x_i$	$F^{\uparrow}$
6.90-6.30	03	6,6	19,8	3
7.50-6.90	07	7,2	50,4	10
8.10-7.50	10	7,8	78	20
8.70-8.10	13	8,4	109,2	33
9.30-8.70	15	9	135	48
9.90-9.30	16	9,6	153,6	64
10.50-9.90	15	10,2	153	79
11.10-10.50	13	10,8	140,4	92
11.70-11.10	10	11,4	114	102
12.30-11.70	07	12	84	109
12.90-12.30	03	12,6	37,8	112
المجموع	112	/	1075,2	/

-المتوسط الحسابي:

$$\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} n_i x_i}{\sum_{i=1}^{n} n_i} = \frac{1075.2}{112} = 9.6$$

متوسط معدلات طلبة المجموعة الأولى هو 9.60.

-الوسيط:

رتبة الوسيط 56 ومنه الفئة الوسيطية هي: $\left[9.30 - 9.90\right[$

$$M_e = \mathrm{A} + \frac{\frac{\sum_{i=1}^{n} n_i}{2} - F_{n-1}^{\uparrow}}{n_{Me}} * \mathrm{K} = 9.30 + \frac{56 - 48}{16} * 0.6 = 9.60$$

نقول أن 50% من مجموع الطلبة تحصلوا على معدلات أكبر من 9.60، و50% الأخرى معدلاتهم أقل من 9.60.

-المنوال:

الفئة المنوالية هي الفئة التي لها أكبر تكرار وهي: $\left[9.30 - 9.90\right[$

$$M_o = \mathrm{A} + \frac{\Delta_1}{\Delta_1 + \Delta_2} * \mathrm{K} = 9.30 + \frac{(16-15)}{(16-15)+(16-15)} * 0.6 = 9.60$$

نقول أن المعدل الأكثر تكرارا في المداولات هو 9.60.

ملاحظة: نلاحظ أن مقاييس النزعة المركزية متساوية ($M_o = M_e = \overline{\mathrm{X}}$) إذا شكل التوزيع متماثل أو متناظر. وأيضا لو قمنا بحساب كل معاملات الالتواء والتفرطح المعروفة سنجدها معدومة.

المجموعة الثانية:

نستعين بجدول التوزيع التكراري لتسهيل الحسابات.

المعدلات	n_i	x_i	n_ix_i	$F^{\uparrow}$	$n_ix_i^2$	$(X_i-\overline{X})^4$	$n_i(X_i-\overline{X})^4$
6.90-6.30	02	6,6	13,2	2	87,12	122,49	244,98
7.50-6.90	04	7,2	28,8	6	207,36	55,28	221,14
8.10-7.50	08	7,8	62,4	14	486,72	20,46	163,68
8.70-8.10	11	8,4	92,4	25	776,16	5,43	59,77
9.30-8.70	13	9	117	38	1053	0,74	9,59
9.90-9.30	15	9,6	144	53	1382,4	0,01	0,17
10.50-9.90	16	10,2	163,2	69	1664,64	0,01	0,09
11.10-10.50	15	10,8	162	84	1749,6	0,58	8,72
11.70-11.10	14	11,4	159,6	98	1819,44	4,71	65,95
12.30-11.70	12	12	144	110	1728	18,47	221,70
12.90-12.30	2	12,6	25,2	112	317,52	51,07	102,13
المجموع	112	/	1111.8	/	11271.96	/	1097,91

-المتوسط الحسابي:

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{\sum_{i=1}^{n} n_i} = \frac{1111.8}{112} = 9.93$$

متوسط معدلات طلبة المجموعة الأولى هو 9.93.

-الوسيط:

رتبة الوسيط 56 ومنه الفئة الوسيطية هي: $]10.50-9.90]$

$$M_e = A + \frac{\frac{\sum_{i=1}^{n} n_i}{2} - F_{n-1}^{\uparrow}}{n_{Me}} * K = 9.90 + \frac{56-53}{16} * 0.6 = 10.01$$

نقول أن 50% من مجموع الطلبة تحصلوا على معدلات أكبر من 10.01، و50% الأخرى معدلاتهم أقل من 10.01.

-المنوال:

الفئة المنوالية هي الفئة التي لها أكبر تكرار وهي: $]10.50-9.90]$

$$M_o = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} * K = 9.90 + \frac{(16-15)}{(16-15)+(16-15)} * 0.6 = 10.20$$

نقول أن المعدل الأكثر تكرارا في المداولات هو 10.20.

من خلال قيم مقاييس النزعة المركزية نلاحظ أن ($M_o > M_e > \overline{X}$) ما يدل على أن التوزيع سالب الالتواء أي يميل نحو اليسار. سنتأكد من ذلك بحساب معاملات الالتواء لهذا التوزيع.

معامل بيرسون الأول للالتواء: يعطى بالعلاقة التالية: ولكن قبل ذلك سنحسب الانحراف المعياري للتوزيع:

$$\sigma = \sqrt{\frac{(\sum_{i=1}^{n} n_i x_i)^2}{\sum_{i=1}^{n} n_i} - \bar{X}^2} = \sqrt{\frac{11271.96}{112} - (9.93)^2} = 1.43$$

$$P_1 = \frac{\overline{X} - M_o}{\sigma} = \frac{9.93 - 10.20}{1.43} = -0.19$$

معامل بيرسون الثاني للالتواء: ويعطى بالعلاقة التالية:

$$P_2 = \frac{3(\overline{X} - M_e)}{\sigma} = \frac{3(9.93 - 10.01)}{1.43} = -0.17$$

وهو ما يؤكد أن التوزيع سالب الالتواء، يميل نحو اليسار. في المرحلة الموالية سنختبر ما إذا كان التوزيع طبيعي أم لا.

معامل بيرسون للتفرطح: حيث يعطى بالعلاقة التالية:

$$\beta_P = \frac{\mu_4}{\mu_2^2} = \frac{9.80}{4.18} = 2.34 < 3$$

ومنه التوزيع متفرطح.

نماذج امتحانات مقترحة

امتحان (2018/2017)

السؤال1 (**نقطة**) : ماذا نعني بالتكرار المعدل؟

....

السؤال2 (**نقطة**) : برهن أن: .... $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$

التمرين الأول (**06 نقاط**): يبين الجدول الأتي توزيع 280 عائلة حسب عدد الأطفال في بلدية ما.

X_i	n_i
0	30
1	45
2	15
3	60
4	90
5	40
$\sum$	

المطلوب: – حدد المجتمع الإحصائي: ....

– الوحدة الإحصائية:............................ .

– الخاصية المدروسة ونوعها...

– أوجد التكرارات التجميعية الصاعدة والنازلة.

– حدد النسب المئوية للعائلات التي لها: – أكثر من طفلين : ....

– التي لها أقل من خمسة أطفال: ....

التمرين الثاني (**06 نقاط**): أخذت عينتان من مجتمعين، فأعطتنا النتائج التالية:

I	II
$\sum_{i=1}^{50} X_i = 300$	$\sum_{i=1}^{40} Y_i = 280$
$\sum_{i=1}^{50} X_i^2 = 1950$	$\sum_{i=1}^{40} Y_i^2 = 2080$

المطلوب: – أوجد الوسط الحسابي والانحراف المعياري لكل عينة.

– أي العينتين أكثر تجانسا (أقل تشتتا).

– دمجت العينتان، فما هو الوسط الحسابي للمجموعة الناتجة.

التمرين الثالث (**06 نقاط**): إذا كانت المعدلات السنوية في مقياس الإحصاء 1 لـ 40 طالب موزعة كما يلي:

Classes	n_i
4-6	9
6-10	10
10-12	14
12-14	5
14-18	2
$\sum$	

المطلوب: - ما نوع المتغير المدروس: ....

- أحسب مقاييس النزعة المركزية.
- أحسب معامل بيرسون للإلتواء و ماذا تستنتج؟
- إذا علمت أن μ_4=203.455 ، أحسب معامل فيشر للتفرطح، ثم حدد شكل هذا التوزيع.

امتحان رقم (2019/2018)

السؤال1 **(نقطة)** : هل يعرف المتوسط الهندسي على أنه الجذر التربيعي لمجموع قيم المتغير مربعة. علل ؟

....

السؤال2 **(نقطة)** : بين أن: .... $\mu_2 = m_2 - m_1^2$

التمرين الأول **(06 نقاط)**: لتكن لدينا السلسلة الإحصائية المتماثلة التالية:

classes	n_i
1-3	n_1
3-5	6
5-7	10
7-9	n_4
9-11	3
$\sum$	

1- أكمل الجدول.

2- حدد قيمة المتوسط الحسابي واستنتج قيمة الوسيط والمنوال.

3- أحسب الانحراف المعياري.

4- ماذا يمكنك القول عن شكل هذا التوزيع من خلال مقاييس الالتواء والتطاول.

التمرين الثاني **(06 نقاط)**: لتكن لدينا D السلسلة الإحصائية المتماثلة التالية:

classes	n_i
1-3	n_1
3-5	6
5-7	10
7-9	n_4
9-11	3
$\sum$	

المطلوب:

1-أكمل الجدول ثم حدد قيمة المتوسط الحسابي والانحراف المعياري.

2- إذا اعتبرنا أن الوسط الحسابي لسلسلة أخرى E هو 7 والانحراف المعياري لها هو 2.64، قارن بين تشتت التوزيعين.

__التمرين الثالث__ (**06 نقاط**): لتكن لدينا السلسلة الإحصائية التالية:

25-38-35-26-45-48-39-19-24-33

المطلوب: - أوجد متوسط السلسلة، الوسيط، والمنوال.

- أحسب الربيع الأول والربيع الثالث.
- قس تشتت هذه السلسلة وماذا تلاحظ؟

امتحان رقم (2017/2016)

التمرين الأول:(**06**) لتكن لدينا السلسلة الإحصائية التالية: 33-24-19-39-48-45-26-35-25-38

المطلوب:- أوجد متوسط السلسلة، الوسيط، و المنوال.

- أحسب الربيع الأول و الربيع الثالث.

التمرين الثاني:(**06**) حسب الديوان الوطني للإحصائيات فإن حظيرة المركبات الوطنية المسجلة في الجزائر لسنة 2010 هي 2977857 مركبة بأقياس زوايا مرافقة لكل نوع من أنواع المركبات كما يلي:

السيارات السياحية: 210 درجة، الشاحنات: 114 درجة، الدراجات النارية: 36 درجة.

المطلوب:

1-ماهي طبيعة المتغير المدروس؟

2-ما هو عدد كل نوع ؟ مثل هذه البيانات في الجدول المناسب.

3-أي نوع من هذه المركبات تمثل المنوال؟

التمرين الثالث: (**08**) إذا كان لدينا التوزيع التكراري التالي:

classes	n_i
5-10	8
10-15	24
15-20	55
20-25	154
25-30	105
30-35	49
35-40	5
$\sum$	400

المطلوب:

1-ما نوع المتغير المدروس.

2-أحسب مقاييس النزعة المركزية.

3-أحسب معامل بيرسن للإلتواء، وماذا تستنتج؟

4-إذا علمت أن $\mu_4 = 3751.08$، أحسب معامل فيشر للتفلطح ثم حدد شكل هذا التوزيع.

المراجع

المراجع:

-جلاطو جيلالي، **الإحصاء مع تمارين محلولة**، ديوان المطبوعات الجامعية، الجزائر، الطبعة الخامسة، 2005

-مو. أري. ر. شبيخل، **الإحصاء، سلسلة ملخصات شوم**، ترجمة د.شعبان عبد الحميد شعبان' د. أحمد حسن الموازيني، الدار الدولية للاستثمارات الثقافية، الطبعة الثامنة، القاهرة، 2006.

- حيدوشي عاشور، **محاضرات في الإحصاء الوصفي**، مطبوعة موجهة لطلبة السنة أولى جذع مشترك، جامعة أكلي محند أولحاج، البويرة، 2016.

-هشام بورمة، **الإحصاء1 دروس وأمثلة تطبيقية**، مطبوعة موجهة لطلبة السنة أولى جذع مشترك، جامعة محمد الصديق بن يحي، جيجل، 2018.

-Alain BACCINI, **Statistique descriptive élémentaire**, institut de mathématiques de Toulouse, UMR, CNRS, France, 2010.

-Yves TILLE, **Résumé du cours de statistique descriptive**, 2010.