جامعة يحي فارس بالمدية
كلية العلوم الاقتصادية والعلوم
التجارية وعلوم التسيير

مطبوعة مقدمة من طرف:

الـدكـتـور:

معـوشـي عيـماد

محاضرات في:

الإحصاء الوصفي

STATISTIQUE DESCRIPTIVE

(مقرر: إحصاء 1)

مقدمة لطلبة السنة الأولى ليسانس

ميدان: علوم اقتصادية وعلوم تجارية وعلوم تسيير

السنة الجامعية : 2019-2018

الإحصاء الوصفي

STATISTIQUE DESCRIPTIVE

(مقرر: إحصاء 1)

إعداد: د. معوشي عيماد

2019-2018

الفهرس

الفهرس

الفهرس

قوائم الأشكال، الجداول والملاحق

مقدمة ..... أ

- **الفصل الأول: الإطار المفاهيمي للإحصاء الوصفي**.....**1-24**

تمهيد.....2

- المبحث الأول: مفهوم علم الإحصاء.....3
- المبحث الثاني: البيانات.....6
- المبحث الثالث: أنواع العينات.....12
- تمارين 21

- **الفصل الثاني: عرض البيانات الإحصائية**.....**25-64**

تمهيد.....26

- المبحث الأول: العرض الجدولي للبيانات.....27
- المبحث الثاني: العرض البياني للبيانات.....44
- تمارين 61

- **الفصل الثالث: مقاييس النزعة المركزية**.....**65-106**

تمهيد.....66

- المبحث الأول: مفهوم النزعة المركزيةومقاييسها.....67
- المبحث الثاني: المتوسطات وأنواعها.....69
- المبحث الثالث: الوسيط وشبيهات الوسيط.....83
- المبحث الرابع: المنوال.....95
- المبحث الخامس: العلاقة بين بعض المتوسطات.....102
- تمارين 105

- **الفصل الرابع: مقاييس التشتت**.....**107-126**

تمهيد.....108

- المبحث الأول: مفهوم مقاييس التشتت وأنواعها.....109
- المبحث الثاني: مقاييس التشتت المطلق.....111
- المبحث الثالث: مقاييس التشتت النسبي والقيمة المعيارية.....121

- تمارين 124
- **الفصل الخامس: العزوم ومقاييس الشكل.....151–127**

تمهيد.....128
- المبحث الأول: حساب العزوم.....129
- المبحث الثاني: مقاييس الشكل.....135
- المبحث الثالث: مقاييس التفلطح (التطاول، الإنبساط) 147
- تمارين 150

المراجع.....152

الملاحق 156

قوائم

الأشكال، الجداول

والملاحق

- **قائمة الأشكال**

الرقم	عنوان الشكل	الصفحة
1-1	العلاقة بين مستويات القياس وأنواع البيانات	11
2-1	العينة والمجتمع	12
3-1	طرق إختيار العينات	20
1-2	سلوك عدد الفئات حسب طريقتي "ستورج" (العشري والنيبيري) و"يول"	31
1-3	التمثيل البياني للمنوال بطريقة المدرج التكراري	96
2-3	التمثيل البياني للمنوال بطريقة الرافعة.	98
1-5	التوزيع المتماثل	135
2-5	توزيع ملتوي يمينا	136
3-5	توزيع ملتوي يسارا	136
4-5	حالات الإلتواء لمقياس "فيشر"	137
5-5	الشكل الصندوقي أفقيا	142
6-5	الشكل الصندوقي عموديا	142
7-5	حالات الالتواء في الشكل الصندوقي (أفقيا)	143
8-5	حالات الالتواء في الشكل الصندوقي (عموديا)	143
9-5	وصف تركز البيانات في الشكل الصندوقي	144
10-5	تحديد القيم الشاذة والمتطرفة باستخدام التمثيل الصندوقي	144
11-5	حالات التطاول في التوزيع التكراري	147

- **قائمة الجداول**

الرقم	عنوان الجدول	الصفحة
1-2	جدول توزيع تكراري لعينة حجمها n	27
2-2	الشكل العام لجدول التوزيع التكراري المزدوج	42
1-3	مقارنة بين الوسيط وشبيهات الوسيط	94
1-5	بناء مثلث باسكال	133

- **قائمة الملاحق**

الرقم	عنوان الملحق	الصفحة
1	الأحرف الإغريقية	157
2	جدول الأرقام العشوائية	158
3	مثلث باسكال	159

مقدمة

مقدمة

تعد البيانات أحد أهم الركائز التي تستند عليها الدراسات التطبيقية المعاصرة سواء في العلوم الدقيقة أو الإجتماعية عموما والإقتصادية بصفة خاصة، حيث ومنذ القدم إهتمت الدول والمجتمعات بأساليب مختلفة لجمعها قصد الإستفادة منها في تسيير الحياة اليومية للأفراد نظرا للمتطلبات التي تقتضيها الأمور التنظيمية للدولة سواء القديمة أو الحديثة، وعملية جمع البيانات هي وسيلة للوصول إلى بعض المعلومات التي تستطيع توضيحها لنا، قصد استخدامها في إتخاذ القرار، إن العلم الذي يعنى بعملية جمع وتنظيم وتلخيص البيانات بغرض تحليلها يطلق عليه الإحصاء الوصفي، حيث يقف الباحث على عملية وصف البيانات من خلال تحكمه في أدوات هذا العلم، في هذه المطبوعة نتطرق إلى أهم المفاهيم المتعلقة بالإحصاء الوصفي من خلال خمسة فصول أتت بأسلوب بسيط غير مخل بالمحتوى العلمي لهذه المادة خاصة وأنها في الغالب موجهة إلى طلبة السنوات الأولى جذع مشترك في مختلف التخصصات وخاصة التخصصات ذات الصلة بالعلوم الإقتصادية، التجارية وعلوم التسيير، حيث تم تخصيص الفصل الأول للمفاهيم الإصطلاحية لمختلف المصطلحات التي يصادفها الدارس للإحصاء الوصفي، أما الفصل الثاني فخصص لشرح طريقتي عرض البيانات سواء الجدولية أو البيانية، أما الفصلين الثالث والرابع فخصص لأهم مقاييس الإحصاء الوصفي وهما مقاييس النزعة المركزية ومقاييس التشتت على الترتيب، أما مقاييس الشكل والتي تعتبر احد أهم مقاييس الإحصاء الوصفي أيضا نظرا لأنها تعطينا فكرة تصورية لسلوك البيانات فتم تركها للفصل الخامس والأخير بعدما تم التعريج في بداية هذا الفصل على حساب العزوم التي تعتبر ركيزة أساسية لحساب مقاييس الشكل، وتجدر الإشارة إلى أنه تم الإستعانة بمجموعة من التمارين التطبيقية في كل بند من البنود الخاصة بهذه المطبوعة بالإضافة إلى مجموعة من التمارين غير المحلولة في نهاية كل فصل لتمكين الطالب من إختبار قدراته على فهم محتوى هذه المادة.

الفصل الأول

مفاهيم عامة

في الإحصاء الوصفي

تمهيد

في هذا الفصل نتطرق إلى المفاهيم العامة التي يحتاجها أي دارس لعلم الإحصاء بصفة عامة والإحصاء الوصفي بصفة خاصة، حيث في بادئ الأمر نعرف ماهية الإحصاء ثم نحدد أنواعه ووظائفه وكذلك عرض لخطوات البحث العلمي الذي يعتمد أساسا على المنهجية الإحصائية، لأخذ فكرة عامة عن البنود الرئيسية التي يجب معالجتها في هذه المطبوعة بصفة عامة، حيث في هذا الفصل نعرض من بين هذه الخطوات مفهوم البيانات وأنواعها وكذالك مستويات قياسها بالإضافة إلى مصادرها وطرق جمعها ، وهذه الأخيرة –طرق الجمع– تظهر لنا مصطلحا هاما لدارسي الإحصاء وهو العينة والتي خصصنا لها مبحثا كاملا لتعريفها ومعرفة أنواعها وطريقة سحبها غير أن هذا الأمر يكون بإختصار نظرا لعدم اتساع المقام لذكرها بالتفصيل.

المبحث الأول

مفهوم علم الإحصاء

من المفاهيم الشائعة بين الناس عن الإحصاء، ما هي إلا أرقام وبيانات رقمية فقط، كأعداد السكان، أعداد المواليد، أعداد الوفيات، أعداد المزارعين، وأعداد المزارع، ...، ومن ثمة ارتبط مفهوم الناس عن الإحصاء بأنه عد أو حصر الأشياء والتعبير عنها بأرقام، غير أن هذا يعتبر المفهوم المحدود لعلم الإحصاء، ولكن **الإحصاء كعلم** هو الذي يهتم بطرق جمع البيانات، تبويبها وتلخيصها بشكل يمكن الإستفادة منها في وصف البيانات وتحليلها للوصول إلى قرارات سليمة في ظل ظروف عدم التأكد[1].

1- وظائف علم الإحصاء:

من التعريف السابق يمكن تحديد أهم وظائف علم الإحصاء في ثلاثة وظائف وهي: وصف البيانات، الإستدلال الإحصائي، التنبؤ.

1-1- وصف البيانات: تعتبر طريقة جمع البيانات وتبويبها وتلخيصها من أهم وظائف علم الإحصاء بصفة عامة، وهو ما نطلق عليه خلال هذه المرحلة **الإحصاء الوصفي** وهو موضوع هذه المطبوعة، إذ لا يمكن الإستفادة من البيانات الخام ووصف الظواهر المختلفة محل إهتمام الباحث إلا إذا تم جمع البيانات وعرضها في شكل جدولي أو بياني من ناحية وحساب بعض المقاييس الإحصائية البسيطة التي تدلنا على طبيعة البيانات من ناحية أخرى.

1-2- الاستدلال الإحصائي: يستند الإستدلال الإحصائي على فكرة إختيار جزء من المجتمع يسمى العينة بطريقة علمية مناسبة، بغرض إستخدام بيانات هذه العينة في التوصل إلى نتائج يمكن تعميمها على مجتمع الدراسة، وهذا الفرع من الإحصاء يدعى **الإحصاء الإستدلالي** ويهتم بموضوعين هما:

أ. **نظرية التقدير:** وتهتم بحساب مقاييس من بيانات العينة تسمى **إحصاءات** (جمع إحصاءة) تستخدم كتقدير لمؤشرات المجتمع التي تسمى **معالم** (جمع معلمة)، ويطلق على المقاييس الإحصائية المحسوبة من بيانات العينة في هذه الحالة **بالتقدير بالنقطة**، كما يمكن أيضا استخدام المقاييس الإحصائية المحسوبة من بيانات العينة في تقدير المدى أو المجال الذي يمكن أو نأمل أن تقع داخله معلمة المجتمع بإحتمال معين وهذا ما يطلق علية **التقدير بالفترة**، ونلاحظ أن عملية وصف البيانات (المرحلة الأولى) والتي هي موضوع الإحصاء الوصفي القاعدة الأساسية لفهم ودراسة مواضيع الإستدلال الإحصائي.

ب. **إختبار الفرضيات:** وفيه يتم استخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفرضيات التي يضعها الباحث حول معالم المجتمع.

1-3- التنبـــؤ: وفيه يتم استخدام نتائج الاستدلال الإحصائي، والتي تدلنا على سلوك الطاهرة المدروسة في الماضي في معرفة ما يمكن أن يحدث لها في الحاضر والمستقبل. وهناك العديد من الأساليب الإحصائية المعروفة التي

تستخدم في التنبؤ، ومن أبسطها أسلوب الإتجاه العام، وهي معادلة رياضية يتم تقدير معاملاتها باستخدام بيانات العينة، ثم بعد ذلك تستخدم المعادلة المقدرة في التنبؤ بما يمكن أن يحدث للظاهرة محل الدراسة في المستقبل[2].

2- مراحل البحث العلمي (المنهج الإحصائي) :

البحث العلمي المعتمد على المنهجية الإحصائية والذي تمتاز به كثير من العلوم سواء في الرياضيات التطبيقية أو العلوم الإنسانية والإجتماعية، عادة ما يمر بمجموعة من الخطوات أو المراحل نذكرها بالترتيب فيما يلي:

2-1- تحديد مشكلة موضوع البحث: يتم تحديد نوع المشكلة وفي أي مجال من العلوم المختلفة التي ذكرناها سابقا والتي تستحق البحث والتقصي فدور الباحث هنا كيف يتم إختيار المشكلة المناسبة لدراسة الظواهر المألوفة والغريبة، وذلك بالتعبير عنها بعلاقة ما بين المتغيرات أو المشاهدات بحيث تمكن الباحث من إجراء التحليل الإحصائي أو الإستنتاجي.

2-2- صياغة الفرضيات: خلال هذه المرحلة يتم صياغة فرضيات بغرض اختبار صحتها أو خطئها بعد المرور بالمراحل الموالية.

2-3- قياس الظاهرة المدروسة: يتم هنا ضبط المقياس المناسب لقياس الظاهرة المدروسة بالشكل الذي يسمح بجمع المشاهدات حولها. ويتم التطرق لمستويات القياس للبيانات فيما بعد.

2-4- جمع البيانات: تعتمد هذه المرحلة بدورها على بعض الأساليب الإحصائية ونتطرق لها لاحقا.

2-5- تبويب وعرض البيانات: هذه المرحلة تكون عبارة عن تنظيم البيانات المجمعة وعرضها بطريقة تسمح بأخذ فكرة عامة حول سلوك الظاهرة ويكون هذا التنظيم في شكل عرض بياني أو في شكل عرض جدولي وتتعدد طرق العرض في كلا الصنفين حسب نوع البيانات والتي نتطرق إليها في المباحث الموالية.

2-6- تحليل البيانات: هذه المرحلة يتم استخدام الأساليب الإحصائية المختلفة حيث يتم حساب بعض المقاييس الإحصائية، ونذكر في هذا الصدد على سبيل المثال، مقاييس النزعة المركزية، مقاييس التشتت، مقاييس الشكل، ...الخ وهذا ما يتم إدراجه في الفصول اللاحقة بنوع من التفصيل.

2-7- تفسير البيانات: هذا يعتمد على عملية ربط الحقائق أو الأمور التي حدّدت من خلال تحليل البيانات مع المؤشرات والغرض من تحليل البيانات، مع مراعاة أن المعلومات التي تم الحصول عليها وجمعت تتحول إلى أدلة للإجابة على مشكلة البحث التي تم طرحها.

2-8- تلخيص النتائج: استنادا على التحليل الإحصائي لبيانات العينة يتم وصف النتائج حول المجتمع ومن ثمة اقتراح الحلول ووضع التوصيات المختلفة[3].

3- تعريف بعض المصطلحات الإحصائية:

3-1- المجتمع الإحصائي: هو مجموعة المشاهدات والقياسات الخاصة بمجموعة من الوحدات الإحصائية والتي تخص ظاهرة من الظواهر القابلة للقياس، أو هو تلك المفردات تحت الدراسة أو البحث ويهدف تعريف

المجتمع الإحصائي إلى تعيين الحدود الصريحة لعملية جمع البيانات ولعملية الإستقراء أو الإستنتاجات التي يمكن الحصول عليها من الدراسة.

ويمكن أن تكون عناصر المجتمع أفراد، عائلات، موظفين، طلاب، مؤسسات، ...الخ. يجب أن تكون معرفة ومحددة بحدود الزمن بحيث يستطيع الباحث معرفة إنتماء أي عنصر من عدم إنتمائه لذلك المجتمع وهذه المفردات التي يتكون منها المجتمع تسمى الوحدات الإحصائية، ولذلك فإن المجتمع الإحصائي يتمثل حجمه بعدد الوحدات الداخلة فيه وعادة ما يرمز لحجم المجتمع N.

3-2- الوحدة الإحصائية: هي الوحدة الأساسية لتكوين المجتمع الإحصائي، فإذا قلنا مجتمع المؤسسات الصغيرة في ولاية المدية، فنقول أن المؤسسة الصغيرة الواحدة من نفس الولاية، تعتبر وحدة إحصائية هنا.

3-3- الظاهرة الإحصائية: هي الخاصية المدروسة أو المتغير المدروس في المجتمع الإحصائي، والذي يأخذ مجموعة من القياسات. وقد نعرف أكثر من متغير إحصائي بالنسبة لنفس المجتمع، فمثلا في مجتمع طلاب السنة الأولى لكلية العلوم بجامعة المدية نستطيع تعريف متغيرات عديدة وأخذ القياسات بشأنها والتي تمثل لنا البيانات – نتطرق للبيانات لها بالتفصيل فيما بعد– ومن أمثلة هذه المتغيرات هنا طول الطلبة، أوزانهم، معدلاتهم، تفضيلاتهم من سلعة معينة، ... الخ.

3-4- البيانات: بشكل عام فالبيانات هي مجموعة من الحروف أو الكلمات أو الأرقام أو الرموز أو الصور (الخام) المتعلقة بموضوع معين. مثال على ذلك: بيانات الموظفين (الأسماء – الأرقام الوظيفية – المهن – الصور) بدون ترتيب ، وينتج عن هذه البيانات بعد المعالجة ما يطلق عليه مصطلح معلومات.

3-5- العينة: هي جزء من المجتمع الإحصائي ولكن ليس أي جزء بل ذلك الجزء الذي يمثل المجتمع أحسن تمثيل. يختلف حجم العينة حسب أهمية ونوعية الدراسة وكذا حسب الإمكانيات المادية والبشرية المتاحة للباحث للقيام بدراسته. تدعى طرق استخراج العينة بعملية المعاينة ونتطرق إليها لاحقا وكذا نتعرف على أسباب إعتماد أغلب الدراسات الميدانية على العينة بدل الحصر الشامل للمجتمع[4].

المبحث الثاني

البيـــــــــانات

نتطرق في هذا المبحث إلى أهم عنصر عند جمع البيانات وهم تحديد طبيعة البيانات وكذلك أداة القياس وهذا يصبح سهلا على دارس علم الإحصاء عند إمكانية تحديده لنوع البيانات التي يتعامل معا وكذا لمستوى قياسها، ونحاول أن نزيل اللبس في الخلط الواقع بين المفهومين بعد التطرق لكل منهما على حدى.

1- مصادر جمع البيانات:

يعتمد الباحثون على مصدرين للحصول على البيانات الإحصائية الخاصة بظاهرة معينة وهما المصادر الأولية وتسمى أحيانا بالمصادر المباشرة، والثانية هي المصادر الثانوية والتي تسمى المصادر غير المباشرة ونذكرهما فيما يلي[5]:

1-1- المصادر الأولية (الميدانية): تلك المصادر التي نحصل منها على البيانات بصفة **مباشرة**، حيث يقوم الباحث نفسه بجمع البيانات من المفردة الإحصائية محل البحث مباشرة، فإذا اهتم الباحث مثلا بجمع بيانات عن أسرة ما يقوم فإجراء مقابلة مع رب الأسرة ليتم الحصول منه مباشرة على بيانات تخص أسرته مثل بيانات المنطقة التابع لها، الحي الذي يسكن فيه، الجنسية، المهنة، الدخل الشهري، عدد أفراد أسرته، والمستوى التعليمي،... الخ ما يميز هذا النوع من المصادر هو الدقة والثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة الإحصائية محل البحث مباشرة، غير أنه يعاب عليها أنها تحتاج إلى وقت وجهد كبير بالإضافة إلى أنها مكلفة من الناحية المادية.

أ. المقابلة الشخصية :حيث يقوم الباحث بالاتصال المباشر بالأشخاص محل الدراسة، وبهذا يستطيع الباحث أن يحقق أعلى درجات الدقة في جمع البيانات، إلا أن هذه الطريقة ورغم ما تمتاز به من دقة المعلومات قد تكون مكلفة، ولاسيما في حالة العينات الكبيرة الحجم.

ب. الاستمارة الإحصائية :وفيها يقوم الباحث بتصميم استمارة تشمل العديد من الأسئلة الرئيسية والفرعية التي تحقق جميع الأهداف للبحث محل الدراسة، ويراعى في هذه الاستمارة الشروط التالية:

- يجب أن تكون الأسئلة واضحة وسهلة.
- يجب أن تكون الاستمارة غير طويلة وذلك حتى لا يصاب الشخص محل الدراسة بالملل والضيق.
- يجب التأكيد على سرية البيانات للشخص محل الدراسة حتى لا تكون إجاباته بعيدة عن الواقع.

- يجب أن تشمل الاستمارة على بعض الأسئلة المكررة في أكثر من موضع تحقيقا لمصداقية الشخص محل الدراسة.
- يجب أن تحقق الاستمارة الأهداف محل الدراسة.

يجب على الباحث بعد إعداد الاستمارة أن يقوم باختبارها عن طريق أخذه لعينة صغيرة جدا، وذلك للتحقق من جميع الشروط السابقة[6].

1-2- المصادر الثانوية (التاريخية): تلك المصادر التي نحصل منها على البيانات بشكل **غير مباشر**، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين، أجهزة وهيئات رسمية متخصصة كنشرات وزارة الفلاحة، نشرات الديوان الوطني للإحصائيات، نشرات منظمة الأمم المتحدة للأغذية والزراعة FAO، تقارير البنك الدولي، تقارير صندوق النقد الدولي FMI،...الخ.

يميز هذا النوع من المصادر توفير الجهد والوقت والمال، إلا أن درجة الثقة في البيانات ليست كما في حالة المصادر الأولية[7].

2- أساليب جمع البيانات:

يتحدد الأسلوب المستخدم في جمع البيانات، حسب الهدف من البحث وحجم المجتمع محل الدراسة ويوجد نوعين من أساليب جمع البيانات هما أسلوب الحصر الشامل وأسلوب المعاينة ونذكرهما فيما يلي:

2-1- أسلوب الحصر الشامل: يستخدم هذا الأسلوب إذا كان الغرض من البحث هو حصر جميع مفردات المجتمع وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا إستثناء، كحصر جميع المزارع التي تنتج التمور، حصر البنوك الزراعية في بلد ما، و**يتميز** أسلوب الحصر الشامل بالشمول ودقة النتائج وبالتالي لا يحتاج إلى أساليب وتعقيدات التعميم، ولكن **يعاب** عليه أنه يحتاج إلى الوقت الطويل والجهد (فريق عمل كبير وفي حالة المجتمعات الكبيرة) والتكلفة العالية ويتعرض لخطر التحيز والذي نتناوله لاحقا.

2-2- أسلوب المعاينة: ينبني هذا الأسلوب على معاينة جزء من المجتمع محل الدراسة، يتم إختياره بطريقة علمية سليمة ومن ثمة دراسته ثم تعميم نتائج العينة على المجتمع وما **يتميز** به هذا الأسلوب هو إقتصاده للوقت وتقليله للجهد وتدنيته للتكلفة بالإضافة إلى حصوله على بيانات أكثر تفصيلا وخاصة إذا جمعت البيانات من خلال استمارة إستبيان، كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب معها إجراء حصر شامل لأن الوحدات الإحصائية تتعرض للتلف عند دراستها، مثل معاينة دم المريض كله يودي بحياته لذا يجب أخذ عينة فقط من الدم أو تعداد الأسماك في البحر يستوجب أخذ عينة منه أو دراسة لجودة المصابيح الكهربائية لمصنع ما يتوجب أخذ عينة ودراستها عوض التضحية بجميع الإنتاج نظرا لتشابه خصائص المصابيح سواء من حيث المواد أو طريقة الصنع المنتجه في المصنع الواحد.

ويعاب على أسلوب المعاينة أن النتائج التي تعتمد على هذا الأسلوب تكون أقل دقة من سابقتها وخاصة إذا لم تكن العينة ممثلة للمجتمع، وكذلك تستوجب دراسة العينات الإلمام بتقنيات الإستدلال الإحصائي حتى يكون تعميم النتائج سليما، ويتعرض هذا الأسلوب لخطر أخطاء الصدفة، ونذكر في البند الموالي خطأين شائعين كثيرا ما يرتبطان بأسلوبي جمع البيانات[8].

3- الأخطاء الشائعة في عملية جمع البيانات:

عموما هنا خطأين يتعرض لهما الباحث العلمي عند جمعه للبيانات المعنية بدراسته وهما: خطأ التحيز وخطأ الصدفة ونذكرهما فيما يلي:

3-1- خطأ التحيز: يقصد بالتحيز إنحراف متوسط جميع التقديرات الممكنة لمعلمة المجتمع عن قيمتها الحقيقية (حتى وإن كانت مجهولة) ومثل هذا الخطأ يظهر في طريقة الحصر الشامل، وخاصة عندما لا تتوافر الإمكانيات لدى الباحث، وهذا النوع من الأخطاء قد يرجع إلى عوامل إنسانية ناشئة عن تصرف القائمين على البحث، لا سيما عند عملية جمع البيانات والغريب في الأمر أن طريقة العينات تتعرض بدرجة أقل لمثل هذه الأخطاء وخاصة إذا ما إعتنى العاملون في البحث بإختيار العينة ولكن رغم هذا يمكن تلافي هذا النوع من الأخطاء عن طريق حذف مصدر الخطأ، إذ كثيرا ما نجد أن مصادر هذا الخطأ، تأتي من عدم دقة الإشراف على جمع البيانات أو قلة التدريب أو ضعف الوعي الإحصائي عند الأفراد القائمين بعملية جمع البيانات.

3-2- خطأ الصدفة: الصدفة هو كل حدث يأتي عرضا (بدون سبب)، وهذا النوع من الخطأ غالبا ما نجده في العينات، إن أن من شروط إختيار العينة في البحوث العلمية أن تكون ممثلة للمجتمع الإحصائي تمثيلا صادقا وصحيحا، إلا أن إختيار عدد غير محدود من مفردات المجتمع بأسلوب عشوائي لا يعني بالضرورة الحصول على عينة ممثلة له، تتمثل فيها كل خصائص وصفات المجتمع الذي أخذت منه. وهذا يعني أن هذه العينة ستتعرض نتائجها إلى نوع من الخطأ يسمى بخطأ الصدفة، وذلك لأن حدوث هذا الخطأ جاء نتيجة لعوامل الصدفة البحتة -عرضيا- ويمكن التقليل من خطأ الصدفة كلما كبر حجم العينة وتزايد بصغر حجم العينة، وعموما ينجم هذا الخطأ عن أحد أو أكثر هذه الأسباب:

- أخذ العينة من مصدر خاطئ، وهذا يعني سوء إختيار العينة، إذ يتم إختيارها بشكل شخصي مما يجعل الباحث منحازا.
- التحيز الشخصي عند أخذ العينة وإختيارها، فقد يكون متعمدا خاصة إذا تجاهل بعض المفردات لعدة أسباب قد تعنيه لشخصه.
- جمع بيانات ناقصة عن المفردات طريق ومن ثمة إسقاط هذه المفردات عن العينة.
- عند اختيار العينات العشوائية قد تكون المفردات لا تمثل المجتمع صدفة فقط ولا دخل للجانب الشخصي للباحث، وليس له يد في ذلك بالرغم من كون هذا الأمر نادر جدا خاصة عندما تكون العينة كبيرة بما يكفي[9].

4- أنواع البيانات:

تقسم البيانات حسب طبيعتها أي حسب درجة التعبير عنها رقيما إلى قسمين، بيانات كيفية وبيانات كمية كما يلي:

4-1- البيانات الكيفية (الوصفية): وهي تلك البيانات التي لا يمكن التعبير عنها رقميا أو عدديا لأن قيمها تكون في شكل أوصاف مثل الألوان: أحمر، أصفر، أسود، ...، والجنس: ذكر، أنثى، أو تكون في شكل رتب، مثل ترتيب طلبة الدفعة المتخرجة من تخصص معين: الأول، الثاني، الثالث، ....، ومستوى الذكاء: ذكي، متوسط الذكاء، قليل الذكاء.

4-2- البيانات الكمية: هي تلك البيانات التي يمكن التعبير عنها رقميا أو عدديا، ويسهل التعامل معها أحسن من الأولى نظرا لأنها تتعامل مع أرقام ونعلم أن معالجة الأرقام رياضيا يكون أسهل من معالجة الأوصاف. وما تجدر الإشارة إليه أن البيانات الكمية تنقسم إلى قسمين أحدها منفصل والآخر متصل كما يلي:

أ. **البيانات الكمية المنفصلة (المتقطعة)**: هي تلك البيانات الكمية التي تأخذ قيما صحيحة ولا يمكن تجزأتها -لا تكون بالفاصلة- مثل عدد أفراد الأسرة (1، 2، 3، ...) وعدد الطلبة في كل فوج لا يمكن القول 20 طالب ونصف.

ب. **البيانات الكمية المتصلة (المستمرة)**: هي تلك البيانات الكمية التي لها إمكانية تجزئة قيمها -تكون بالفاصلة أحيانا- مثل الدخل الشهري، الطول قد يأخذ قيمة 1م و20سم و3مم، أو نقول مثلا 1.5 متر، 2.25 متر، ...، وكذلك الأوزان والكتل والحجوم، ... الخ[10].

5- مستويات قياس البيانات:

في هذا البند نتطرق إلى تلك المسألة التي تتعلق بمدى إجراء العمليات الحسابية على قيم المتغيرات للبيانات المجمعة وهذا راجع إلى مستوى قياس البيانات لظاهرة المدروسة وميز الإحصائيون بين أربعة مستويات للقياس، حيث تزداد إمكانية إجراء العملياات الرياضية كلما كان مستوى القياس أعلى ونذكره فيما يلي:

5-1- المقياس الإسمي: ولا يوجد أي ترتيب، أو تحديد لكمية، أو نسبة في هذا المقياس، لذلك يعبر عنه بأنه من أضعف مستويات القياس أو أبسطها. ومثاله تصنيف الذكور والإناث، فنقول مثلا لدينا جنس العامل ذكر أو أنثى فلا مفاضلة بين القيمتين ولا أي قيمة عددية. أو ألوان السيارات، أخضر، أسود، أبيض فيتم تصنيف السيارات بحسب ألوانها، لا يمكن ترتيبها ولا تكميمها. فلا تستخدم العمليات الحسابية في هذا النوع من القياس، فلا نستطيع أن نجري أي عمليات حسابية على هذه الأعداد، ولكن بالإمكان عد الأفراد في كل فئة، مثلا 10 ذكور، 12 إناث. وهو يستخدم للبيانات الكيفية (النوعية).

5-2- المقياس الترتيبي: ويستخدم هذا المقياس لترتيب القيم الكيفية (لا تأخذ قيما عددية) للمتغيرة المدروسة. و من أمثلته ترتيب الأوائل في أحد الفصول الدراسية ، بأن نقول خالد هو الأول ، حسن هو الثاني، و عصام

الثالث، ففي هذه الحالة قمنا بترتيب الأشخاص لكن دون أن نبين الفرق أو المسافة بين الأول والثاني، فيمكن أن يكون معدل خالد 13 أو 14 أو 16. فهو يفيد الترتيب بين الأفراد ولكن ليس من الضروري أن تكون الفروق في مقدار أو درجة الخاصية بين كل رتبتين متجاورتين منتظمة، أي أن وحدات القياس غير متساوية، والفرق بين الأول والثاني في الخاصية، لا يساوي بالضرورة الفرق بين الثاني والثالث. مثاله: ترتيب الطلاب من حيث الدافعية للتعلم. وهو يستخدم للبيانات الكيفية (النوعية).

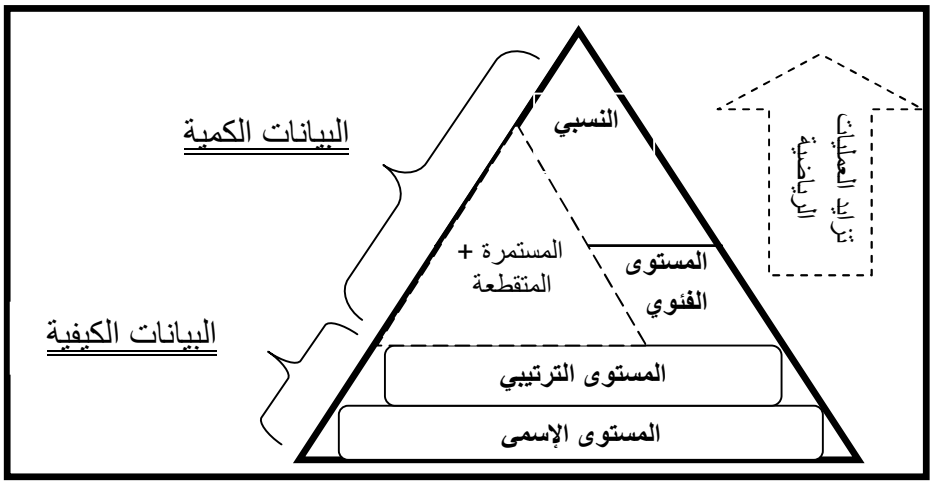
5-3- المقياس الفئوي (الفتري): المقياس الفئوي ويسمى بالمقياس الفتري أو مقياس المسافة، حيث تكون الوحدات هنا متساوية، ولكن لا يوجد صفر حقيقي وإنما هو افتراضي لا يعني انعدام الخاصية عندما تأخذ القيمة "0"، ويكون هناك سالب تحت هذا الصفر الافتراضي، ولهذا لا يصح أن ننسب قيمة إلى أخرى على أنها ضعفها أو نصفها مثل درجة الحرارة (القول أن درجة الحرارة 0، لا يعني إنعدام في ظاهرة الحرارة). كذلك فالطالب الذي يحصل على درجة صفر، لا يعني أنه لا يملك أي معلومة. لهذا ينبغي علينا مراعاة قيمة الصفر في هذا النوع من البيانات للتمييز بينه وبين المقياس النسبي الموالي. وهو يستخدم للبيانات الكمية.

5-4- المقياس النسبي: المقياس النسبي له جميع خصائص المقياس الفئوي بالإضافة إلى وجود صفر حقيقي، ولذلك يمكن أن ننسب قيمة إلى أخرى، مثل الوزن والطول (القول أن وزن شئ ما هو 0، يعني أنه لا يوجد هذا الشئ أصلا أي أن الظاهرة تنعدم فعلا)، ولا يكون هناك سالب تحت الصفر الحقيقي، وهو يستخدم للبيانات الكمية[11].

6- العلاقة بين مستويات القياس وأنواع البيانات:

كثيرا ما يتم الخلط بين مستويات القياس وأنواع البيانات، وفي هذا الصدد نقول أن أنواع البيانات يعنى بتصنيف البيانات حسب إمكانية التعبير عنها رقميا أم لا من جهة ومن جهة أخرى بيان إمكانية تجزئة قيم المتغير أم لا إن كان قيمه رقمية من جهة أخرى، أما مستويات القياس فتعنى بتحديد إمكانية إجراء العمليات الحسابية على البيانات كحساب مقاييس الإحصاء الوصفي والتي نتطرق إليها لاحقا بالتفصيل، وهذا يتأتى من معرفة المقياس المناسب للظاهرة عند عملية جمع البيانات والتي تعد من خطوات البحث العلمي كما ذكرنا سابقا. والشكل الموالي يعطينا فكرة عن مطابقة مستويات القياس مع أنواع البيانات:

شكل 1-1: العلاقة بين مستويات القياس وأنواع البيانات



المصدر: جامعة الملك فيصل، يوم: 2017/09/30 [على الخط] الرابط:
https://www.kfu.edu.sa/sites/Home

نلاحظ من خلال الشكل أن البيانات الكيفية يكون مستوى قياسها إما إسميا أو ترتيبيا فقط، أما البيانات الكمية فسواء كانت متقطعة أو مستمرة لها مستويين للقياس لا علاقة لهما بصنف البيانات الكمية، حيث قد تكون البيانات كمية متقطعة ولها مستوى قياس إما فئوي أو نسبي، وكذلك قد تكن البيانات كمية مستمرة ولها مستوى قياس إما فئوي أو نسبي، وكثيرا ما يحدث الإلتباس عند التفرقة بين هذين المقياسين، وأفضل طريقة للتميز بينهما هو الرجوع إلى القيمة الصفرية للظاهرة المدروسة فإذا عبرت عن إنتفاء الظاهرة -إنعدامها- نقول أن الصفر حقيقيا ويكون مستوى القياس نسبيا (الطول، الوزن، العمر، ...)، اما إذا كانت القيمة الصفرية لا تؤدي إلى انتفاء الظاهرة المدروسة -الظاهرة غير منعدمة- فنقول أن الصفر هنا إفتراضيا أو غير حقيقي، ويكون مستوى القياس فتريا -فئوي- مثل درجة الحرارة تأخذ القيمة صفر عن حالة التجمد وهو مقياس افتراضي يعبر عن درجة حرارة التجمد أي أن الظاهرة لا تزال قائمة.

وكما هو مبين من خلال الرسم أن ترتيب مستويات القياس تم من خلال درجة تكميم وإجراء العمليات الحسابية حيث كلما صعدنا للأعلى تزيد إمكانية تطبيق العمليات الرياضية على البيانات. وكذلك كل مقياس يأخذ خصائص المقياس الذي قبلة ويزيد عليه. فأضعف مقياس هو الإسمي، وأفضلها هو النسبي. وهذا يفيد في عملية الرفع من مستوى قياس متغير ما عند تعذر التعامل معه إحصائيا فمثلا عدم إمكانية التعامل مع الجنس الذي يأخذ قيمتين 'ذكر، أنثى' يمكن ترميز ذكر بـ 1 وانثى بـ 2، وهكذا يتم تسهيل العمليات الحسابية ولكن يجب توخي الحذر مع الترميزات المستخدمة خاصة عند تطبيق المقاييس الوصفية عليها أو تفسيرها.

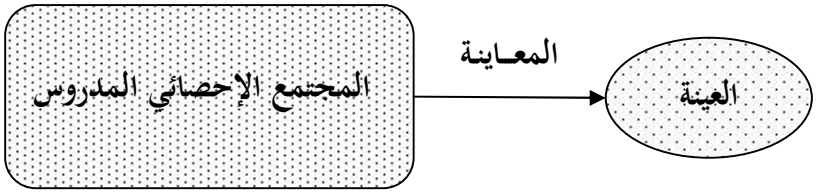
المبحث الثالث

أنـواع العـينـــــات

نتطرق في هذا المبحث إلى ذلك الجزء من المجتمع الذي يمثله أحسن تمثيل حيث يختار بغرض إجراء الدراسة عليه لتعميم نتيجتها على المجتمع الإحصائي، حيث نجد في هذا الصدد نوعين من طرق إختيار العينة أحدهما طريقة عشوائية والأخرى غير عشوائية وكلا منهما يحوى مجموعة من الأنواع تختار حسب طبيعة المجتمع وهدف الباحث بالإضافة إلى أسباب أخرى نذكرها في هذا المبحث.

ويقصد بالمعاينة تلك العملية التي يتم بموجبها إختيار العينة والرسم الموالي يوضح ذلك:

شكل 1-2 : العينة والمجتمع



المصدر: إعداد الباحث

1- العينات العشوائية (الإحتمالية):

هي العينات التي يتم إختيار مفرداتها وفقا لقواعد الإحتمالات، بمعنى آخر هي التي يتم إختيار مفرداتها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيز، ومن أهم العينات الإحتمالية أو العشوائية ما يلي:

1-1- العينة العشوائية البسيطة:

يتم إختيار العينة العشوائية البسيطة بطريقتين وذلك يرجع إلى حجم المجتمع الإحصائي، حيث أنه إذا كان المجتمع حجمه يساوي أو **أقل من 25** مفردة يتم ترقيم هذه المفردات من 1 إلى 25، ثم تكتب على بطاقات ويتم إختيار وبطريقة عشوائية العدد المطلوب لتشكيل العينة المرغوبة وهذا بدون إرجاع حتى لا تتكرر مفردة ما أكثر من مرة. أما في المجتمعات الكبيرة **والتي تزيد عن 25** مفردة يتم إستخدام جدول الأرقام العشوائية (أنظر **الملحق -2-**). والمثال التالي يبين طريقة ذلك:

- **مثال:** مجتمع حجمه 5000 مفردة نريد سحب عينة عشوائية حجمها 50 مفردة من هذا المجتمع، لإجراء عملية السحب بإستخدام جدول للأرقام العشوائية نقوم بما يلي:

1- نرقم مفردات المجتمع من 1 إلى 50.

2- بما أن حجم المجتمع ذو 4 منازل لابد من التأكد من أن جدول الأرقام العشوائية مكون من 4 منازل وفي حالة كونه يحتوي على 5 منازل نقوم بحذف عمود (خانة) الآحاد من هذا الجدول،

3- ثم نبدأ بقراءة الأعداد من هذا الجدول إبتداءا من أقصى اليسار ومن أعلى العمود الأول آخذين الأعداد التي تقل عن 5000 وغير مكررة،.

4- نتابع هذه العملية بشكل متسلسل وكلما انتهينا من عمود نبدأ من العمود المجاور له -الذي يليه او على يمينه- حتى نتحصل على العينة المرغوبة.

- **مميزات العينات العشوائية البسيطة :**

- تعطى جميع مفردات المجتمع الأصلي نفس الفرصة المتكافئة في الاختيار .
- لا تتقيد بترتيب معين أو نظام مقصود
- لا تتطلب معرفة سابقة بخصائص مفردات المجتمع الأصلي
- تتفادى التحيز لاعتمادها على حد كبير على قانون الاحتمالات
- سهولة سحب مفرداتها حيث لا تتطلب سوى قوائم تتضمن بيانات عن مجتمع الدراسة
- انخفاض خطأ المعاينة حيث تشترط تجانس مجتمع الدراسة
- استخدام برامج الحاسب الآلي في الاختيار يوفر كثير من الجهد والوقت والمال للباحث.

- **عيوب العينات العشوائية البسيطة :**

- استخدام جداول الأرقام العشوائية لتحديد كل مفردة عملية متعبة وخاصة إذا كانت العينة كبيرة وقد تأخذ جهداً كبيراً.
- عدم توفر قوائم مسبقة لمجتمع الدراسة ، وفى حالة توافرها فهى عادة ما تكون غير دقيقة.
- في حالة عدم التأكد من تجانس مجتمع الدراسة يمكن أن تقود النتائج التى يتم التوصل إليها إلى أخطاء.

1-2- العينة العشوائية المنتظمة: يتم فيها اختيار الحالة الأولى من العينة بطريقة عشوائية ثم يمضى الباحث في اختيار بقية الحالات على أبعاد رقمية منتظمة أو متساوية بين الحالات ، بحيث تكون المسافة بين أي وحدتين متتاليتين ثابتة في جميع الحالات.

ونتبع الخطوات التالية :

1- تحديد المجتمع الأصلى ذو الحجم (N) وترقيم وحداته من 1 إلى N

2- تحديد حجم العينة المرغوب فيه (n)

3- تحديد المسافة بين أفراد العينة من خلال $k= N/n$ ، حيث k عدد طبيعي (يتم تقريبه إذا كان كسرا)

4- اختر عشوائياً عدداً (r) ينحصر بين 1 وقيمة k، أي: $r \in \mathbb{N}\,; r \in [1\,;\ k]$

5- أضف إلى العدد المختار قيمة k بشكل منتظم حتى نحصل على العينة التي حجمها n

ترتيب العنصر المختار في المجتمع $[r] \quad [r+k] \quad [r+2k] \quad \cdots \quad [r+(n-1)k]$

عدد المفردات المختار $(1) \quad (2) \quad (3) \quad \cdots \quad (n)$

حيث : $[r+(n-1)k] \leq N$

- **مثال** : إذا كان مجتمع الأصلي يتكون من 100 مفردة ، ونريد الحصول على عينة منه عددها 20 مفردة، كيف يمكنك تحقيق ذلك باستخدام العينة العشوائية المنتظمة ؟

لدينا $N=100$ و $n=20$ إذن: $K=100/20=5$

الآن نختار عدداً ينحصر بين 1 و5 وليكن 4 وذلك بشكل عشوائي ، ونجعل العدد 4 نقطة الانطلاق، ونضيف له $k=5$ بشكل ثابت منتظم، حتى يصبح حجم العينة المختارة 20 مفردة.

أي: **4**، 4+5=**9**، 9+5=**14**، 14+5=**19**، **24**، **29**،... **99** : $4 + (20 - 1)5 = \mathbf{99} \leq 100$

- **مميزات العينات العشوائية المنتظمة** :

- تعد من أسهل العينات العشوائية في التطبيق
- لا تحتاج إلى عملية إعداد مسبق لمفردات الدراسة خاصة إذا كانت مجموعات داخل مجتمع الدراسة.
- لا تحتاج إلى الرجوع في كل مرة يتم فيها سحب المفردات إلى مرجع أو دليل فيكتفي بالمفردة الأولى أما باقي المفردات فتحدد تلقائياً عن طريق صيغة رياضية سهلة ومبسطة

- **عيوب العينات العشوائية المنتظمة:**

- تستلزم توفر قائمة حديثة تشمل كافة أسماء مفردات المجتمع الأصلي.
- قد تكون العينة المختارة غير متجانسة ، وذلك حينما تختار مفردات على أبعاد منتظمة يصادف أن يكونوا من طبقة معينة أو من ذوى خصائص وصفات مميزة وغير متشابهة مع بقية المفردات.
- يشترط في المجتمع الأصلي أن يكون الأفراد في تسلسل منسق وتدرج من حيث التنوع.
- لا تحدث احتمالية فرصة التمثيل لمفردات مجتمع الدراسة إلا مرة واحدة وهى عند اختيار المفردة الأولى.
- في حالة كون طول الفئة كبيراً وهناك مجموعات داخل مجتمع الدراسة عددها أقل من طول الفئة فإن احتمال تمثيل هذه المجموعة في العينة يكون محدوداً.

1-3- العينة العشوائية الطبقية: إن المعاينة العشوائية البسيطة كثيراً ما تؤدى إلى الحصول على عينة تبتعد فى خصائصها عن خصائص المجتمع مما يترتب عليه خطأ المعاينة . ولزيادة احتمال تمثيل خصائص المجتمع في العينة، فإننا نلجأ إلى العينة العشوائية الطبقية. وهى نوع آخر من العينات العشوائية غير أنها تتعامل مع مجتمع غير متجانس ونتبع الخطوات التالية :

1- تحديد وتعريف المجتمع الأصلي .

2-تحديد حجم العينة .

3-تحديد الطبقات الفرعية بناء على خصائص المجتمع الأصلى .

4-تصنيف أفراد المجتمع وفقاً للطبقات الفرعية السابق تحديدها وبحيث ينتمى كل فرد لمجموعة واحدة فقط، وذلك حتى لا تتداخل المجموعات .

5-اختيار عينة عشوائية بسيطة من كل طبقة فرعية.

- **طريقة تحديد العدد المسحوب من كل طبقة:**

إذا كان لدينا مجتمع إحصائي حجمه N وبه k من الطبقات حيث بكل طبقة N_i من المفردات أي :

$$\sum_{i=1}^{k} N_i = N$$

ونريد سحب عينة حجمها n، فالعدد n_i المطلوب سحبه من كل طبقة N_i يمكن أستنتاجه بالطريقة الثلاثية:

$$\begin{cases} N \to n \\ N_i \to \boldsymbol{n_i} \end{cases} \quad \Rightarrow \quad \boldsymbol{n_i = \frac{N_i}{N} \times n}$$

حيث نحصل على :

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

- **مثال:** ليكن لدينا مجتمع به 400 عامل يتوزعون على أربعة مهن 100 في الفلاحة، 150 في التجارة، 80 في الصناعة، 70 مهن أخرى، ونريد سحب عينة حجمها 50 عامل بغرض إجراء دراسة معينة، فكيف يتم ذلك؟

- **الحل:**

لدينا مجتمع غير متجانس مقسم إلى أربعة طبقات حيث: $100 + 150 + 80 + 70 = 400$

1- العدد المسحوب من طبقة الفلاحة:

$$n_2 = \frac{N_2}{N} n = \frac{100}{400} 50 = 12.5 \cong \mathbf{12}$$

2- العدد المسحوب من التجارة:

$$n_2 = \frac{N_2}{N} n = \frac{150}{400} 50 = 18.75 \cong \mathbf{19}$$

3- العدد المسحوب من الصناعة:

$$n_2 = \frac{N_2}{N} n = \frac{80}{400} 50 = \mathbf{10}$$

4- العدد المسحوب من المهن الأخرى:

$$n_2 = \frac{N_2}{N} n = \frac{70}{400} 50 = 8.75 \cong \mathbf{9}$$

وبعد معرفة العدد المسحوب من كل طبقة يتم اختيار هذا العدد من كل طبقة إما عن طريق العينة العشوائية البسيطة أو المنتظمة .

- **مميزات العينات العشوائية الطبقية :**

– يتحقق التمثيل، ليس فقط للمجتمع الأصلي، بل لكل طبقاته الفرعية مهما كان بعضها يشكل أقلية صغيرة

– أدق من العينة العشوائية البسيطة، لأنها تجمع العشوائية وبالتالي تحقق التكافؤ بين الأفراد، والحياد في الاختيار ، والغرضية ، فنضمن عدم خلوها من خصائص المجتمع الأصلي

– تتميز بالدقة الإحصائية وانخفاض نسبة حدوث الخطأ المعيارى، خاصة كلما كانت المجموعات أو الطبقات متجانسة داخلياً.

- **عيوب العينات العشوائية الطبقية:**

– تتطلب من الباحث التعرف وبشكل جيد على مجتمع دراسته لتحديد المجموعات التي يتكون منها.

– تتطلب إجراءات كثيرة يجب على الباحث القيام بها قبل الشروع في استخدام أي من العينات العشوائية البسيطة أو المنتظمة

– يقوم الباحث بسحب عدد من العينات تبعاً لعدد مستويات المتغير الذي يتعامل معه مما يؤدى إلى مضاعفة الجهد الذي يقوم به.

1–4– العينة العشوائية المتعددة المراحل (العنقودية): عبارة عن مجموعة من العينات العشوائية البسيطة أو المنتظمة المستخدمة لسحب مفردات مجتمع دراسة واحد ، هذه المجموعة من العينات لا تقل عن مرحلتين وتزيد حسب طبيعة الدراسة وفى كل مرحلة يتم سحب عينة ، وفى حالة وجود عينة عشوائية واحدة لا نطلق عليها عينة عنقودية لأنها في هذه الحالة إما أن تكون عينة عشوائية منتظمة أو بسيطة. كما يتم فيها في اختيار مجموعات وليس مفردات. فمثلا يكون الاختيار العشوائى لمناطق أو مجموعات أو تجمعات مختلفة مثل المدارس أو الفصول الدراسية أو المناطق التعليمية. وتتصف هذه التجمعات في أن لكل أعضائها نفس الخصائص. و يمكن تنفيذها في مراحل، تتضمن اختيار عناقيد ضمن عناقيد أخرى وتسمى العينة متعددة المراحل.

- **مثال :**

يهدف باحث إلى التعرف على آراء معلمي المرحلة الإبتدائية حول دور المشرف التربوي في العملية التعليمية، ويتكون المجتمع من جميع المعلمين في الجزائر، موزعين على المدراس الإبتدائية، ويريد الحصول على عينة مكونة من 500 معلم كيف يتم ذلك باستخدام العشوائية العنقودية ؟

فهنا يقوم الباحث باختيار ولاية أو مجموعة من الولايات بطريقة عشوائية، ثم بلدية أو بعض البلديات من كل ولاية مختارة بطريقة عشوائية، ثم يقوم بإختيار مدرسة أو بعض المدارس من كل البلديات المختارة وبطريقة عشوائية، ثم يقوم بإختيار مجموعة من المعلمين من كل مدرسة وبطريقة عشوائية (بالعينة البسيطة على سبيل المثال)

- **مميزات العينات العشوائية العنقودية:**

– تتعامل مع كل المجتمعات المتجانسة بغض النظر عن حجمها بشرط أن يكون مجتمع الدراسة موزعاً فى أكثر من مكان جغرافي.

- أن جميع المجتمعات الفرعية المكونة لمجتمع الدراسة الأصلي تتشابه في الخصائص العامة بصورة كبيرة.
- تناسب المجتمعات الكبيرة المتناثرة التي تشغل حيزاً جغرافياً شاسعاً.
- يمكن استخدام كل من العينة العشوائية البسيطة والمنتظمة عند الانتقال من مرحلة إلى أخرى.

- عيوب **العينات العشوائية العنقودية** :

- تتطلب خطوات كثيرة تبعاً لعدد المراحل كما تتطلب سحب عينات كثيرة أيضاً "عينة في كل مرحلة.-
- احتمال كبير ألا تكون العينة ممثلة للمجتمع
- تحليل بياناتها غير مناسب باستخدام معظم أساليب الإحصاء الاستدلالى.

2- العينات غير العشوائية (غير الإحتمالية):

في التي يتم اختيار مفرداتها بطريقة غير عشوائية، حيث يقوم الباحث بإختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة ومن أهم أنواع العينات غير الإحتمالية نذكر فيما يلي:

1-1- العينة الصدفية: هي العينة التي يتم فيها اختيار مفردات الدراسة نتيجة لعامل الصدفة وليس لأي عامل آخر.أي أنا سميت العينة الصدفية بهذا الاسم لأن الباحث يعتمد في اختيار مفردات العينة على الصدفة، وتتمثل إجراءات اختيار العينة في أن يقابل الباحث أول من يصادفهم من الطلاب بعد أدائهم الاختبار في مادة معينة، أو آخر من يخرج من المسجد عقب أداة صلاة معينة، أو من يصل إلى السوق، وهكذا، إلى أن يصل المجموع إلى العدد المقرر أن تتكون منه العينة. على الرغم مما يبدو من سهولة اختيار مفردات هذه العينة، إلا أن ما يعيبها هو أنها قد لا تمثل المجتمع تمثيلا صحيحًا، بمعنى أن أول الواصلين إلى السوق أو أول الخارجين من صالة الاختبار قد لا يمثلون كل الخصائص الفعلية التي تتمثل في أفراد المجتمع الأصلي.

- **ومن خصائصها:**

- من أضعف العينات اللإحتمالية بوجه عام من حيث قدرتها على الوصول بنتائج دقيقة نظراً لارتفاع نسبة التحيز لدى الباحث وانخفاض نسبة التمثيل لمجتمع الدراسة
- تتصف بسهولة التطبيق ولا تتطلب أي إجراء مسبق.
- تستخدم في البرامج الإعلامية والتليفزيونية أو قياس اتجاهات الرأي العام حول قضية ما وسؤال من نقابله مصادفة.

- **ملاحظة:** نلاحظ أنه بالرغم من تسميتها صدفية إلا أنها لا تنضوي تحت العينات العشوائية لكون أن إختيار المفردات لا يكون بالطريقة الاحتمالية أو العشوائية نظرا لأن المفردات لا تكون لها نفس الحظ في الظهور.

1-2- العينة العمدية (القصدية أو الغرضية): اختيار الحالات بناء على هدف خاص لدى الباحث، مثل: دراسة 3 مؤسسات معينة من بين كل المؤسسات الموجودة في المنطقة.

أي أنها سميت بهذا الاسم لأن الباحث يتعمد وضع خصائص معينة، يحرص على توافرها في مفردات الدراسة، مثلا يرغب إجراء الدراسة على من تتوفر فيهم الشروط التالية:

- أن يكون حاصلا على شهادة جامعية في الهندسة الصناعية.
- أن لا يقل معدله السنوي عن 7.
- أن يجيد لعبة التيكوندو.
- أن يكون حاصلا على دورات تدريبية في مجال الأمن والسلامة.

- **ويوجد أنواع كثيرة من العينات العمدية أو الغرضية القصدية أهمها :**

أ. **العينة الحصصية:** تتطلب معرفة مسبقة لمجتمع الدراسة من حيث تكوين المجموعات داخله وعملية الاختيار في كل مجموعة لا ترتبط بقواعد معينة ولكن لقناعة الباحث بشرط أن تمثل كل مجموعة في العينة حسب تمثيلها في مجتمع الدراسة. وهي من أفضل العينات غير الإحتمالية لأن الباحث يختار العينة وفقاً لخصائص محددة مسبقاً لأفراد المجتمع. إلا أنه يصعب فيها الحصول على عينة ممثلة للمجتمع

وتشبه إلى حد كبير العينة العشوائية الطبقية، وتتمثل أوجه الاختلاف في أنها غير عشوائية، وأن المجتمع يكون في هذه الحالة غير معروفا، إلا أنه يمكن الحصول على الإحصائيات الخاصة بنسبة تواجد كل حصة من الحصص، فعلى سبيل المثال لو كان عدد أفراد المجتمع 12000، وكانت نسبة المتزوجين إلى العزاب هي 7 إلى 5، وأراد الباحث أن يختار عينة تتكون من 1000 شخص مثلا، عندها يرسم خطة تتضمن اختيار 7 متزوجين مقابل 5 عزاب، وهكذا حتى تكتمل العينة. أو أن يحدد الباحث فئات المجتمع (ذكور وإناث) ثم يختار عدد ثابت من كل فئة إذ يقرر مثلا اختيار عشرة ذكور وعشرة إناث.

ب. **عينة كرة الثلج "الشبكية"**: فيها يتعرف الباحث على فرد من المجتمع الأصلي ، يقوده لفرد آخر وهكذا يتسع نطاق معرفة الباحث بهذا المجتمع، وتسمى بالعينة المتضاعفة. حيث تتطلب قدرة من الباحث على إقناع من يتعرف إليهم من مجتمع الدراسة بالتعاون معه في إرشاده إلى مفردات أخرى. و تستخدم في حالة عدم توفر قائمة بكل أفراد المجتمع الأصلي.

- **مثال:** يريد الباحث دراسة مجتمع المدمنين في مدينة ما لا يجد أمامه إلا من هو في السجن أو مصحة علاجية، أو التعرف على أحدهم وتكوين علاقة معه فسوف يقوده إلى مجموعة من زملاءه المدمنين

ت. **عينة الحالات المتطرفة** : تبرز الظاهرة موضع اهتمام الباحث بشكل كبير، ويطلق عليها عينة الحالات الشاذة. يتم فيها اختيار حالات غير ممثلة لمجتمع الأصلي.

- **مثال** : يريد باحث التعرف على الخصائص النفسية والاجتماعية للمتسربين من التعليم الابتدائي وأهم العوامل التي أدت إلى تسربهم، فالعينة هنا حالات المتسربين

ث. العينة المتتابعة : تشبه العينة العمدية مع وجود فرق هو أنه في العينة **العمدية** يحاول الباحث الحصول على أكبر عدد ممكن من الحالات المناسبة التي تقع في نطاق تعريفه للمتغيرات التي يدرسها حتى يستنفد ما لديه من جهد ومال ووقت. فالمبدأ الأساسي هو الحصول على كل حالة ممكن الحصول عليها. أما في العينة **المتتابعة** فإن الباحث يظل يجمع الحالات حتى تملأ المعلومات أو الحالات التي يحصل عليها، الفراغ الذي لديه. وفى هذه الحالة يجمع عدداً من الأفراد ويدرسهم، ثم يجمع عدداً آخر ويدرسهم وهكذا بالتتابع حتى يحقق الهدف الذي يريد الوصول إليه من دراسة العينة.

- **مثال** : يريد باحث دراسة حالات الرسوب في الثانوية العامة ولتحقيق ذلك يجمع الباحث عدداً من الراسبين ويدرسهم وفقاً لمتغيرات بحثه، ثم يجمع عدداً آخر ويدرسهم، وهكذا حتى يصل إلى نقطة التشبع ولا يعود يحصل على بيانات جديدة، فيتوقف ويعتبر أنه حصل على العينة التي يريدها.

ج.العينة النظرية : يرتكز اختيار الباحث لأفراد العينة على قاعدة نظرية تحدد أبعاد السمة موضوع الدراسة، والتى على أساسها سيتم الاختيار .يجمع الحالات بناء على التوجه النظري الذي توصل إليه من دراسة سابقة لمبادئ النظرية

- **مثال** : يريد الباحث تحليل سلوكيات العاملين في مؤسسة صناعية في ضوء نظرية العمل المهني وقواعده، فيبدأ بملاحظة مجموعة من الأفراد طوال أيام الأسبوع. ويحلل تصرفاتهم في ضوء هذه النظرية.

ح.العينة الاجتهادية: اختيار الأفراد بناء على الخبرة الذاتية والاجتهاد من الباحث . تؤدى إلى الحصول على عينة متحيزة .

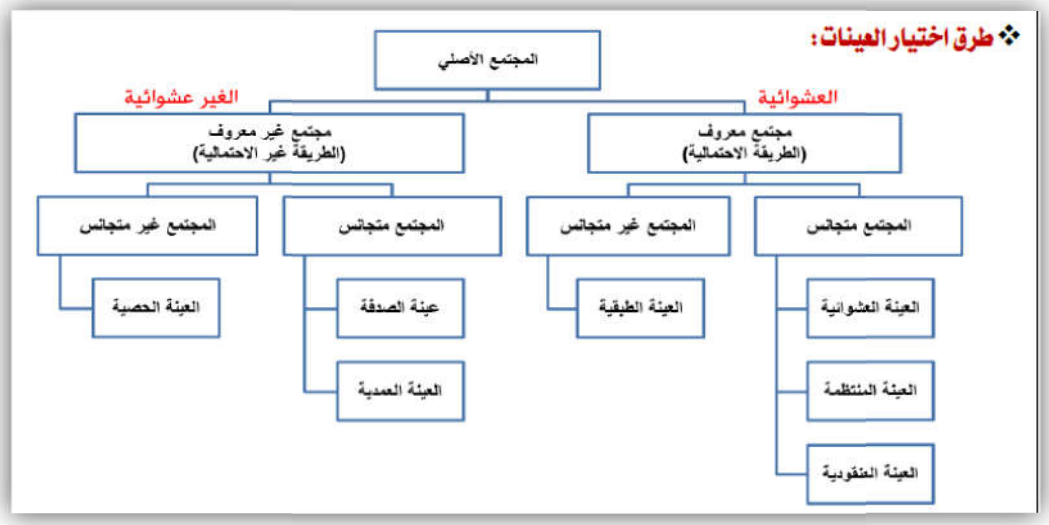
- **مثال:** اختيار الرؤساء في العمل أو الزملاء أو الأصدقاء

خ.العينة الكتلية: يختار الأفراد من القسم الذي يدرس فيه أو العمارة التي يسكن بها. الهدف الوحيد للاختيار هو سهولة الحصول على البيانات. تعتبر عينة متحيزة لا يمكن التعميم منها وإنما نتائجها لا تنطبق إلا على الكتلة التي اختارها الباحث فقط .

د. العينة المعيارية: عينة ينتقى الباحث أفرادها في ضوء معايير، مثل الأطفال من سن (3-5) سنوات ذوى الإعاقة البصرية، وتعد هذه المعايير مصدر طمأنينة للباحث إلى حد كبير أن العينة ستمثل المجتمع الأصلي

ذ. العينة المكثفة أو الشديدة: عينة تمدنا بمعلومات وفيرة عن حالات تعبر عن الظاهرة بوضوح، لكن ليس بشكل حاد كما في العينة المتطرفة، ومثال ذلك اختيار الطالب ذي المستوى الجيد أو المتوسط[12].

شكل 1-3 : طرق إختيار العينات



المصدر: جامعة الملك فيصل، يوم: 2017/09/26 [على الخط] الرابط:

https://www.kfu.edu.sa/sites/Home

- **ملاحظات هامة حول العينات :**

1- من الممكن الحصول على عينة عشوائية ولكنها غير ممثلة لخصائص المجتمع، فالعشوائية لا تضمن التمثيل، ولكنها تتيح فقط فرصة متساوية لأي فرد من أفراد المجتمع أن يختار في العينة.

2- عندما يتجه الباحث إلى العينات العشوائية أو الاحتمالية فإنه إما أن يبدأ بالعينة العشوائية البسيطة أو المنتظمة أو ينتهي بهما عندما يكون مجتمع الدراسة متجانس. أما في حالة المجتمعات غير المتجانسة فعليه بالعينة العشوائية الطبقية.

3- قبل اختيار أفراد الدراسة يجب وبكل وضوح أن نقرر ما هو الحجم المناسب للعينة التي تفي بمتطلبات الدراسة، وعلينا مراعاة العوامل التي تؤثر في حجم العينة

4- قبل البدء في سحب مفردات العينة يجب على الباحث الإحاطة بمجتمع الدراسة، ومعرفة خصائصه.

تمارين

- **التمرين 01**: حدد المجتمع الإحصائي، الوحدة الإحصائية، المتغير الإحصائي ونوعية المتغير في الأمثلة التالية:
  - التخصص المختار من طرف الناجحون الجدد في شهادة البكالوريا .
  - نسبة النجاح في شهادة البكالوريا بثانويات ولاية المدية.
  - عدد التلاميذ المرشحين لشهادة البكالوريا بولايات الجمهورية
  - مكان الاقامة للطلبة الجدد بجامعة المدية .
  - تصنيف مناطق الوطن من حيث معدل سقوط الأمطار السنوي.
  - تصنيف الأحزاب السياسة من حيث عدد الأصوات المحصل عليها.
  - الأجر السنوي لمعلمي مدرسة ما.
  - بعد مقرات الولايات عن مدينة الجزائر.
  - هواية طلبة السنة الأولى LMD .
  - آثار الحضارات القديمة في دول آسيا الوسطى.
  - أعلى القمم الجبلية في دول أوربا.
  - أطول الأنهار في الدول الأفريقية .
  - ماركات السيارات في الحضيرة الوطنية للسيارات.
- **التمرين 02**: في مؤسسة جامعية بها 12000 طالب موزعون على6 تخصصات مختلفة على النحو التالي:3600 3000 2400 1600 1000 400 . وبهدف دراسة أين يرغب الطلبة قضاء أوقات فراغهم قررنا سحب عينة تتكون من 240 طالب.

المطلوب :

ماهي طبيعة المجتمع الإحصائي المدروس.

ما نوعية العينة المسحوبة من هذا المجتمع ؟ .

ما هو عدد الوحدات الإحصائية الواجب سحبها من كل تخصص حتى تكون هذه العينة تمثل المجتمع أحسن تمثيل؟

- **التمرين 03**: في دراسة لجدوى احد المشروعات في بلدية يقدر عدد سكانها بـ 12000 نسمة كان السؤال المطروح هو " موافق أو غير موافق" ، تقدم إلى البلدية 200شخص للتحبير عن موقفهم من المشروع.

المطلوب : حدد المجتمع الإحصائي ، الوحدة الإحصائية ، المتغير الإحصائي ونوعية المتغير.

- **التمرين 04** : سمح إحصاء السكان لسنة 1987 بتحديد جنسية الأجانب المتواجدين بالجزائر والتي كانت كالتالي:

مغاربة، تونسيون، فرنسيون، ألمان، إيطاليون، فلسطينيون وآخرين .

المطلوب: حدد المجتمع الإحصائي ، الوحدة الإحصائية ، الصفة المدروسة ونوعية الصفة.

- **التمرين 05** : صنف المتغيرات التالية حسب نوعيتها (كيفية ، كمية مستمرة أو غير مستمرة).

الطول ، العمر ، الجنسية ، القامة ، الحالة العائلية ، الوزن ، لون البشرة ، الدخل الفردي الشهري ، عدد الأطفال لكل أسرة ، عدد الغرف بكل مسكن .

- **التمرين 06**: حدد نوع المتغيرات التالية:

المتغير	وصفي	كمي منفصل	كمي متصل
نوع العمل			
العمر بالسنوات			
فصيلة الدم			
دخل رب الاسرة			
درجة الذكاء			
درجة الطالب في الامتحان			
نوع الكلية			
رقم الهاتف			
زمن انتظار الخدمة			
عدد اطفال الاسرة			
لون البشرة			
ترتيب الطلبة في الامتحان			
المستوى الاجتماعي			
الانفاق الاسري			
رأي الافراد في احد القوانين			
درجة الحرارة			

- **التمرين 07**: حدد مستوى قياس المتغيرات التالية:

المتغير	اسمي	ترتيبي	فترة	نسبة
نوع العمل				
العمر بالسنوات				
فصيلة الدم				
دخل رب الاسرة				
درجة الذكاء				
درجة الطالب في الامتحان				
نوع الكلية				
رقم الهاتف				
زمن انتظار الخدمة				
عدد اطفال الاسرة				
لون البشرة				
ترتيب الطلبة في الامتحان				
المستوى الاجتماعي				
الانفاق الاسري				
رأي الافراد في احد القوانين				
درجة الحرارة				

- **التمرين 08**: إليك جدول البيانات التالي:

	النوع	**التخصص الاساسي**	**تقدير الكالريوس**	**العمر**	**درجة الامتحان**	**عدد اللغات**
1	انثى	علوم سياسية	جيد جدا	23	8	3
2	ذكر	إحصاء	امتياز	22	9	2
3	ذكر	هندسة ميكانيكية	جيد جدا	24	7	1
	—	—	—	—	—	—
	—	—	—	—	—	—
14	انثى	اقتصاد	جيد جدا	26	8	2
15	ذكر	علوم سياسية		25	9	2

المطلوب:

1- ما هو عدد المفردات الكلية ؟

2- ماهي المتغيرات المتضمنة مجموعة البيانات؟

3- وما هي أنواع المتغيرات

4- ما هو مستوى القياس المناظر لكل متغير من متغيرات الدراسة

5- هل يمكن إجراء عمليات حسابية على متغير "تقدير البكالوريوس"؟ ولماذا؟

- **مراجع وإحالات الفصل الأول:**

[1] عبد اللطيف يوسف الصديقي "تاريخ الإحصاء" المؤسسة الجامعية للدراسات والنشر والتوزيع، ط 1، 2006، لبنان، ص 6.
[2] شرف الدين خليل "الإحصاء الوصفي" شبكة الأبحاث والدراسات الإقتصادية، بدون سنة، بدون بلد، ص 8.
[3] دلال القاضي، وآخرون "الإحصاء -للإدرايين والإقتصاديين-" دار الحامد، 2005، الأردن، ص 16، 17.
[4] جلاطو جيلالي "الإحصاء مع تمارين ومسائل محلولة" د. م. ج، 1998، الجزائر، ص 5.
[5] أحمد عبد السميع طبية "مبادئ الإحصاء" دار البداية، ط 1، 2008، الأردن، ص 13.
[6] أماني موسى محمد "التحليل الإحصائي للبيانات" مركز تطوير الدراسات العليا والبحوث في العلوم الهندسية، ط 1، 2008، مصر، ص 7، 8.
[7] شرف الدين خليل "الإحصاء الوصفي" سبق ذكره، ص 10، 11.
[8] شرف الدين خليل، سبق ذكره، ص 12.
[9] موسوعة البحوث والمواضيع المدرسية، [على الخط] 2017/09/30 :
http://bohotti.blogspot.com/2014/12/blog-post_444.html
[10] جلاطو جيلالي، سبق ذكره، ص 6.
[11] دلال القاضي وآخرون، سبق ذكره، ص 19.
[12] جامعة الملك فيصل، م. ع. السعودية، [على الخط] 2017/09/20:
https://www.kfu.edu.sa/sites/Home

الفصل الثاني

عرض البيانات الإحصائية

تمهيد

رأينا في الفصل السابق أهم المفاهيم التي تصادف الدارس لعلم الإحصاء، ومن بين هذه المفاهيم تعد البيانات أحد أهم المصطلحات الإحصائية، حيث تطرقنا إلى أنواعها، طرق جمعها وقياسها بالإضافة إلى مصادرها وتعتبر الطريقة التي يتمّ من خلالها عرض البيانات في العديد من المجالات من أهم الأمور، فهناك طرق تشجع المتلقي على التفاعل بشكل أكبر من الطرق الأخرى؛ بسبب سهولة عرض البيانات، وسهولة استيعابها، وتخضع البيانات على اختلافها في العادة إلى العديد من العمليات التي تتمّ عليها والتي تعمل على معالجتها وتحويلها إلى قراءات يمكن توظيفها في المجال الذي يهتم بها، وقد تستعمل البيانات في المناهج المدرسية، وفي قطاع الأعمال، والاقتصاد، والأبحاث، وفي الإحصاءات، وفي العديد من المجالات الأخرى، لهذا فهناك العديد من الطرق المتبعة والتي يمكن من خلالها عرض البيانات بالطريقة المثلى وبشكل يؤدّي الغرض كاملاً، ومن أهم طرق عرض البيانات المستعمل على نطاق واسع من قبل العديد من الأفراد نجد العرض الجدولي الذي يكون في شكل جداول متنوعة، وكذا العرض البياني في شكل صور أو أشكال ...الخ في هذا الفصل نتطرق إلى أهم الطرق السائدة لعرض مجموعة البيانات بإختلاف أنواعها.

المبحث الأول
العرض الجدولي للبيانات

عرض البيانات الإحصائية هو أهم خطوة بعد جمع البيانات حيث يتسني عن طريقة إعطاء فكرة أولية عن سلوك البيانات، ويكون العرض الجدولي الذي نتطرق إليه في هذا المبحث هو الركيزة الأساسية لتحليل البيانات المبوبة وعرضها بيانيا أيضا.

1- جدول التوزيع التكراري البسيط:

هي تلك الجداول التي تتوزع فيها البيانات حسب صفة واحدة ويتألف عادة من عمودين (أو سطرين)، الأول يمثل الظاهرة أو الخاصية المدروسة والثاني يمثل عدد المفردات التابعة لكل مشاهدة وهو ما يسمى بالتكرارات[1].

جدول رقم 2-1 : جدول توزيع تكراري لعينة حجمها n

الرقم	المتغير x_i (الخاصية المدروسة)	التكرار المطلق n_i
1	x_1	n_1
2	x_2	n_2
3	x_3	n_3
⋮	⋮	⋮
k	x_k	n_k
المجموع: $\sum_{i=1}^{k} n_i$		n

المصدر: إعداد الباحث

1-1- حالة متغير كيفي: إذا كان لدينا متغيرا كيفيا، نكون هنا أمام جداول توزيع تكرارية بسيطة يتم بناؤها كما هو موضح في الأمثلة التالية:[2]

- **مثال (1)**: حالة المتغيرة الكيفية :

تم دراسة 30 زبون لإحدى العلامات التجارية للسيارات حول أفضل لون يفضلونه لسيرتهم المستقبلية من بين الألوان التالية: الأبيض، الأسود، الفضي، الأحمر، البني وهي الألوان التي توفرها هذه العلامة التجارية. فكانت النتائج كما يلي:

أحمر أسود أسود أحمر فضي فضي بني أبيض أبيض أحمر أسود أسود أحمر بني أبيض أبيض أحمر بني أبيض أبيض أحمر أحمر أسود أسود أحمر فضي فضي بني أبيض أبيض

المطلوب: وضع البيانات المتاحة في جدول توزيع تكراري.

- **الحل:** حالة متغير كيفي إسمي:

نلاحظ أنه لدينا خاصية مدروسة واحدة والتي تمثل اللون المفضل لكل زبون، يتبقى لنا الآن حساب مختلف التكرارات لكل مشاهدة ليتم وضعها في جدول توزيع تكراري كما يلي:

المتغير (اللون) x_i	أبيض	أسود	فضي	أحمر	بني	**المجموع**
التشطيب	//// ///	//// /	////	//// ///	////	
التكرارات n_i	8	6	4	8	4	**30**

- **ملاحظات:** نلاحظ أننا أضفنا سطر سميناه التشطيب والغرض منه هو المساعدة في حساب التكرات فقط حيث توضع كل 5 مشاهدات في حزمة بعد شبها من قائمة البيانات لتسهيل عملية العد، كما يمكن وضع الجدول بطريقة أفقية أو عمودية. ولا يهم ترتيب قيم المتغير في الجدول إذا كان المتغير كيفي ومستوى قياسه إسمي -كما في المثال السابق-، بعكس المتغير الكيفي ذو مستوى القياس الترتيبي الذي تكون فيه المشاهدات مرتبا إما تصاعديا أو تنازليا -كما في المثال الموالي- غير أنه يفضل الترتيب **التنازلي** خاصة عند حساب التكرارات المتجمعة الصاعدة والنازلة بمختلف أنواعها لكي لا تختلط الأمور والقصد من وراء هذه التكرارات.

- **مثال (2) :** حالة متغير كيفي ترتيبي.

تم جمع بيانات لفوج معين يتكون من 30 طالب بقسم سنة أولى علوم إقتصادية حول تقدير البكالوريا المحصل عليه فكانت النتائج كما في جدول التوزيع التكراري التالي:

تقدرير البكالوريا (المتغير)	مقبول	جيد	جيد جدا	ممتاز	**المجموع**
عدد الطلبة (التكرار)	18	7	4	1	30

- **ملاحظة:** مثلما أشرنا سلفا تم ترتيب قيم المتغير تصاعديا نظرا لكونه متغيرا رتبيا.

1-2- حالة متغير كمي متقطع مع تبويب بدون فئات: إذا كان لدينا متغير كمي متقطع ولكن عدد المشاهدات كان قليلا نوعا ما بحيث لا تحتاج إلى التصنيف في فئات، نكون هنا أمام جداول توزيع تكرارية بسيطة بدون فئات، كما هو موضح في المثال التالي:

- **مثال:** حالة المتغيرة الكمية المتقطعة المبوبة بدون فئات:

لدراسة متوسط عدد الأطفال في العائلة في مجتمع ما، أخذت عينة عشوائية حجمها 50 عائلة، وتم تعداد عدد الأطفال في كل منها فكانت النتائج كمايلي:

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

المطلوب: وضع البيانات في جدول توزيع تكراري.

- **الحل:**

نلاحظ أنه لدينا خاصية مدروسة واحدة تتمثل في عدد الأطفال المشاهد في كل أسرة، وهذا المتغير كمي متقطع ونلاحظ أن عدد المشاهدات هو 6 (بغض النظر عن التكرارات) أي 0 1 2 3 4 5 6 وبالتالي لا يتطلب هذا التصنيف في فئات وبالتالي نحن أمام جدول توزيع تكراري بسيط بدون فئات، يتبقر حساب عدد الأسر المنتمية إلى كل مشاهدة من المشاهدات الستة وهذا ما يسمى بالتكرارات والنتيجة في الجدول التالي:

المتغير (عدد الأطفال) x_i	0	1	2	3	4	5	**المجموع**
التشطيب	/////	// /////	/// ///// /////	///// /////	//// /////	/ /////	
التكرارات (عدد الأسر) n_i	5	7	13	10	9	6	**50**

- **ملاحظة:** في جداول التوزيع التكرارية الخاصة بالمتغيرات الكمية المتقطعة يتم الترتيب في قيم المتغير تصاعديا من الأصغر إلى الأكبر.

1-3- حالة متغير كمي متصل: رأينا في البند السابق المتغير الكمي المتقطع وفي حالة كون عدد المشاهدات قليل (بغض النظر عن التكرارات) يتم توزيع التكرارات على المشاهدات الأصلية بدون تصنيفها في فئات، غير أنه إذا كان المتغير متقطع وكان عدد المشاهدات المذكور كبير يتم تصنيفها في فئات لكي لا نخل بخاصية التلخيص التي يتميز بها علم الإحصاء، أما إذا كان المتغير كمي مستمر فيكون التصنيف في فئات هو الأرجح دائما.

إن عدد الفئات يكون من إختيار الباحث ويعتمد على خبرته من جهة ويتناسب مع حجم البيانات والأهداف المتوخاة من التحليل الإحصائي من جهة أخرى، ولكن عموما يكون محصورا بين 5 و 15 فئة، حيث أن المحددات الأساسية التي تضبط هذا الأمر هو خاصية التلخيص وإظهار التوزيع، حيث أن الإعتماد على عدد قليل من الفئات يجعل من التلخيص جيد غير أن إظهار توزيع التكرارات لكل فئة يكون غير واضح نظرا للتجميع الكبير للقيم داخل الفئات ومن جهة أخرى الإفراط في عدد الفئات يجعل من خاصية التلخيص غير محققة نظرا للعدد الكبير للفئات في حين إظهار التوزيع التكراري يكون جيد، ومن هذا المنطلق يجب الموازنة بين هذين الخاصيتين-إظهار التوزيع والتلخيص- لكي يتحقق توزيع تكراري واضح وبحد مقبول من التلخيص.كذلك يتحدد طول الفئة وعدد الفئات بعوامل أخرى لا تقل أهمية عن هذين الأمرين، وذلك حسب أهمية الدراسة، وطول مجال التعريف (المدى) وحجم العينة أو المجتمع، ويستحسن أن تكون أطوال الفئات متساوية حتى يسهل مقارنة مختلف

مراحل الظاهرة (الفئات)، حيث أن الأطوال غير المتساوية تستلزم فقط عند التمثيلات البيانية أو إستخراج المنوال تعديلا للتكرارات لكي لا تكون الرسومات والنتائج مظللة، وهو ما نراه لاحقا.

في هذا الصدد وضع الإحصائيون طرقا يمكن الإستناد إليها عند تحديد عدد الفئات وتبقى طرقا غير إجبارية على الباحث كما سلف الذكر، ومن بينها نذكر كل من طريقة "يول" وطريقة "ستورج" ونذكرهما فيما يلي:

أ. **طريقة "يول" *Yule***: يحدد **عدد الفئات (k)** حسب هذه الطريقة بالصيغة التالية:

$$k = 2.5 \times \sqrt[4]{n}$$

حيث:

n: حجم العينة أو المجتمع

ب. **طريقة "ستورج" *Sturge***: حسب هذه الطريقة يحدد **عدد الفئات (k)** بالصيغة التالية:

$$k = 1 + 1.32\ ln(n) \quad \text{(اللوغاريتم النيبيري)}$$

أو

$$k = 1 + 3.32\ log(n) \quad \text{(اللوغاريتم العشري)}$$

حيث:

n: حجم العينة أو المجتمع

بعد تحديد عدد الفئات سواء بالطريقتين السابقتين أو إعتمادا على خبرة الباحث يتبقى تحديد **طول الفئات (C)** والذي عادة ما يكون متساوي إذا تم إستعمال القاعدة التالية:

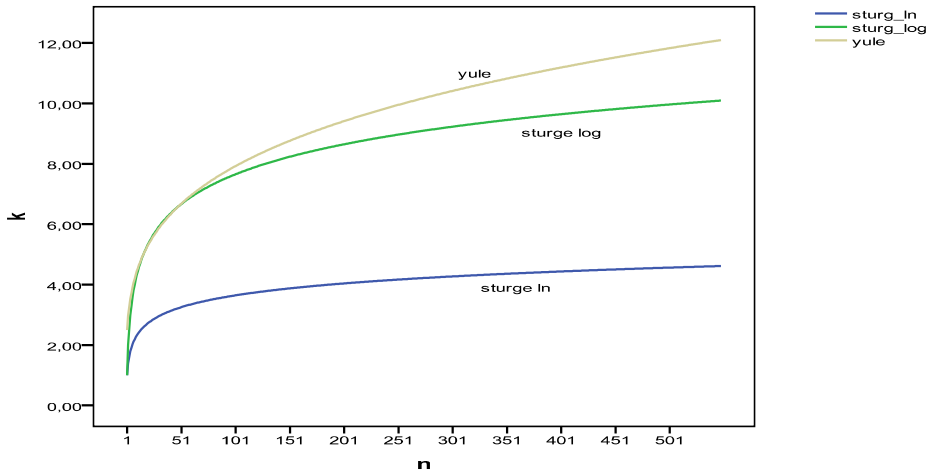
$$C = \frac{R}{k}$$

حيث:

R: المدى، ويحسب بالفرق بين القيمة العليا والقيمة الدنيا للبيانات كما يلي: $R = x_{max} - x_{min}$

- **ملاحظة:** قد تكون الفئات غير متساوية خاصة عند التشتت الكبير للبيانات حيث يتم توسيع الفئات التي تحوي بيانات متطرفة أو متباعدة أو يتم تقليص في حجم الفئات إذا كان هناك تركز كبير للتكرارات في فئات معينة، وهذا على سبيل المثال فقط لا الحصر ويبقى تقدير ذلك للباحث وخبرته وكذا طبيعة الدراسة والمقصود منها والمقاييس المطبقة عليها.

شكل 2-1: سلوك عدد الفئات حسب طريقتي "ستورج" (العشري والنيبيري) و"يول"



المصدر: إعداد الباحث بالإعتماد على برمجية SPSS v.17

من خلال التمثيل البياني بالاعتماد على صيغ حساب عدد الفئات المذكورة سلفا، نلاحظ بأنه من أجل العينات الصغيرة n (اقل من 30 مثلا) تتقارب كل الطرق في إعطاء عدد الفئات، لكن بعد حجم العينة 100 تقريبا تبدأ في التباعد قليلا عن بعضها البعض وطريقة ستورج باللوغاريتم العشري تتوسط الطريقتين الأخريتين لذا نفضل إستخدامها لإيجاد عدد الفئات وإن كانت غير إجبارية، ومما تجدر الإشارة إليه أن عدد الفئات k يميل إلى الاستقرار عند حجوم العينات الكبيرة (يتزايد بنسب ضئيلة جدا) وهذا ليحقق احد أهداف الإحصاء الوصفي وهو التلخيص، ولكي لا يكون عدد الفئات كبير، وهذا الأمر الذي يميز كلا من الطرق المذكورة.

- **مثال (1) حالة متغير كمي متقطع:**

في دراسة لأحد الباحثين لموسم بداية العطلة الصيفية تم تعداد الزبائن الوافدين في أحد الفنادق ذات 3 نجوم، تم ومن خلال السجلات المحتفظ بها في الفندق الرجوع إلى شهري ماي وجوان من عام 2017 والقيام بإستخراج البيانات اليومية وكانت النتائج كما يلي:

بيانات يومية	9	10	11	11	12	12	15	16	20	22	23	27	30	35	35	35
لشهر ماي	36	37	39	40	44	45	46	46	46	47	47	48	48	49	50	
بيانات يومية	50	51	51	52	50	50	49	49	48	47	46	45	44	44	44	
لشهر جوان	45	45	46	47	50	51	50	45	45	45	44	43	40	35	33	

المطلوب:

1- ضع جدولا للتوزيع التكراري للمتغيرة المدروسة باستخدام فئات طولها 10 من شكل [5-15[وهكذا

2- ما هو طول الفئة الذي نستخدمه في جدول توزيع تكراري باستخدام عدد فئات يساوي 8.

3- ما هو عدد الفئات بإستخدام طريقتي Yule وSturge ؟

- **الحل:**

1- جدول توزيع تكراري بطول فئة يساوي 10 من الشكل [5-15]:

لدينا متغيرة إحصائية من النوع الكمي المنفصل لكن نلاحظ أن قيم المتغير الإحصائي تأخذ مجالا واسعا لذا يجب تبويبها في فئات لكي يتم تلخيص البيانات بشكل جيد:

بمأن طول الفئة معطى فإنه يمكن تحديد عدد الفئات التي نكونها كمايلي:

$$C = \frac{R}{k} \Rightarrow k = \frac{R}{C} = \frac{x_{max} - x_{min}}{C} = \frac{50 - 9}{10} = \frac{41}{10} = 4.10 \cong 5$$

نلاحظ أننا قربنا إلى **أكبر** عدد صحيح عند تحديد عدد الفئات مهما كان العدد وراء الفاصلة لأن طول الفئة معطى مسبقا وثابت ولأن كذلك طول الفئات متساوي. لأن التقريب إلى العدد الأصغر يستوجب توسيع الفئة الأخيرة حتى تضم القيم المتبقية.

عند تحديد حدود الفئات يكون الفرق بين الحد الأعلى لكل فئة والفئة التي تليها يساوي الواحد، لأن المتغير منفصل.

رقم	فئات المتغيرة المدروسة (عدد الزبائن) x_i	التكرارات (عدد الأيام) n_i
1	[05-15]	7
2	[16-26]	4
3	[27-37]	9
4	[38-48]	28
5	[49-59]	13
المجموع		**61**

- **مثال (2):** بإعتبار المتغير المدروس في المثال السابق هو أجور 61 عامل أخذوا بطريقة عشوائية فيكون نوعه كمي مستمر (متصل)، فإن الإجابة على السؤال الأول تختلف في حدود الفئات وذلك وفق حالتين:

- **حالة (1):** حيث تكون نهاية الفئة الأولى بمجال مفتوح هي بداية الفئة التي تليها بمجال مغلق أي أن القيمة التي تقع في ذلك الحد تنتمي إلى الفئة الثانية وليس الأولى وهكذا إلى غاية الحد الأخير في الفئة الأخيرة يتم غلق المجال، مما يغير في التكرارات بطبيعة الحال كما يلي:

n_i	الفئات	الرقم
6	]15-05]	1
5	]25-15]	2
3	]35-25]	3
15	]45-35]	4
32	[55-45]	**5**
61	**Σ**	

- **حالة** (**2**): أو تكون حدود الفئات بشكل عادي كما ذكرنا في المتغير المنفصل ولكن مع إضافة عمود نبين فيه الحدود الفعلية (الحقيقية) عن طريق إنقاص 0.5 للحدود الدنيا للفئات وإضافة 0.5 للحدود العليا للفئات وتكون المجالات كما في الحالة (1) السابقة بالنسبة للحدود الفعلية للفئات، كما يلي:

التكرارات n_i	**الحدود الفعلية للفئات**	الفئات	الرقم
7	]15.5-04.5]	[15-05]	1
4	]26.5-15.5]	[26-16]	2
9	]37.5-26.5]	[37-27]	3
28	]48.5-37.5]	[48-38]	4
13	[59.5-48.5]	[59-49]	**5**
61	**المجموع**		

2- في هذا المثال لما يكون عدد الفئات يساوي 8: يتبقى إيجاد طول الفئات وبتطبيق القانون المذكور سابقا نجد:

$$C = \frac{R}{k} = \frac{x_{max} - x_{min}}{k} = \frac{50 - 9}{8} = \frac{41}{8} = 5.12 \cong 5$$

نلاحظ أنه إذا قربنا إلى أصغر عدد يمكن أن يزيد عدد الفئات بفئة واحدة والعكس إذا قربنا إلى العدد الأكبر يمكن أن ينقص عدد الفئات أيضا بفئة واحدة، ويتم التقريب إلى عدد صحيح نظرا لأن المتغير كمي منفصل، وليس متصلا، حيث في هذا الأخير تكون لنا الحرية في التقريب أولا نظرا لطبيعة المتغير ويرجع هذا لطبيعة الدراسة وخبرة الباحث.

3- عدد الفئات بإستخدام طريقتي Yuleو Sturge:

أ. طريقة Yule:

$$k = 2.5 \times \sqrt[4]{n} = 2.5\sqrt[4]{61} = 6.98 \cong 7$$

إذن عدد الفئات هو 7 فئات، حسب هذه الطريقة.

ب. طريقة Sturge: ويتم الإختيار بين استخدام اللوغاريتم العشري أو النيبيري وليس بالضرورة أن تعطي نفس النتائج ولكن تكون تقريبية كما يلي:

- الحالة (1):

$$k = 1 + 3.32\, log(n) = 1 + 3.32\, log(61) = 6.92 \cong 7$$

- الحالة (2):

$$k = 1 + 1.32\, ln(n) = 1 + 1.32\, ln(61) = 6.42 \cong 6$$

نلاحظ أننا قربنا عدد الفئات بطريقة عادية لأن طول الفئة غير معطى ويمكن إستخراجه بطريقة عادية من القانون المذكور سابقا.

يستحسن أن يبدأ بكتابة الحد الأدنى للفئة الأولى بقيمة اصغر مفردة أو اقل من ذلك بقليل وتنتهي بالحد الأعلى للفئة الأخيرة بقيمة اكبر مفردة أو أكثر من ذلك بقليل.

- **مثال (3) حالة متغير كمي مستمر:**

في دراسة لظاهرة تأخر العمال عن العمل أجريت الدراسة على 50 عامل وتم تقدير التأخرات بالدقائق في أحد الأيام وكانت النتائج كما يلي:

53	30	27	25	23	7	13	40	14	6	49	5	4	4	3	2	1
52	11	10	9	8	2	2	41	22	36	5	3	35	3	33	1	16
	9	51	50	4	0	21	12	0	6	0	1	0	1	19	2	31

المطلوب:

1- أنشئ جدولا للتوزيع التكراري باستخدام طريقة *Yule*

2- باستخدام طريقة *Sturge* ما هو عدد الفئات وطول الفئة اللازمين لإنشاء جدول توزيع تكراري.

3- أنشئ جدولا للتوزيع التكراري باستخدام طول فئة يساوي 5 .

4- أنشئ جدولا للتوزيع التكراري باستخدام عدد فئات يساوي 4.

الحل:

1- جدول توزيع تكراري بإستخدام طرية Yule لإيجاد عدد الفئات:

$$k = 2.5 \times \sqrt[4]{n} = 2.5\, \sqrt[4]{50} = 6.64 \cong 7$$

ومنه يمكن تحديد طول الفئات كما يلي:

$$C = \frac{R}{k} = \frac{53 - 0}{7} = 7.57$$

يمكن تقريب العدد أو إبقائه كما هو نظرا لأن المتغير كمي متصل، وهذا حسب طبيعة الدراسة حيث في بعض الدراسات أين تكون الأرقام يتطلب فيها الدقة الشديدة وأخذ الأرقام بعد الفاصلة مهم (مثل علم الكيمياء،

العلوم التجريبية، ...) قد تؤخذ بدون تقريب أو يكون التقريب في الأرقام بعد الفاصلة فقط، مثل تقريب 7.57 إلى 7.50 مثلا، وأحيانا أخرى يتم التقريب إلى أعداد صحيحة لتسهيل الأمر وهذا يرجع إلى خبرة الباحث. كما في مثالنا يتم التقريب إلى العدد 8 كطول فئة.

الرقم	الفئات	التكرار المطلق
1	]08-00]	23
2	]16-08]	8
3	]24-16]	5
4	]32-24]	5
5	]40-32]	3
6	]48-40]	2
7	[56-48]	4
مجموع التكرارات (حجم العينة)		50

2- الجداول المغلقة والمفتوحة:

2-1- الجداول المغلقة: نصادف في جداول التوزيع التكراري جداولا مغلقة وهي الجداول التي يكون فيها الحد أدنى للفئة الأولى والحد الأعلى للفئة الأخيرة معروفين (بغض النظر عن المجالات مفتوحة أو مغلقة)،

- **مثال:**

الرقم	الفئات	التكرار المطلق
1	]08-00]	23
2	]16-08]	8
3	]24-16]	5
4	[32-24]	5

2-2- الجداول المفتوحة: ونصادف في جداول التوزيع التكراري الأخرى جداولا مفتوحة وهي الجداول التي يكون فيها الحد أدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة غير معروفين أو أحدهما فقط غير معروف (بغض النظر عن المجالات مفتوحة أو مغلقة)، ونذكر الحالات الثلاثة التالية:

أ. **جداول مفتوحة من الأسفل:** وهي الجداول التي يكون فيها الحد الأدنى للفئة الأولى غير محدد.

- **مثال :**

الرقم	الفئات	التكرار المطلق
1	08-	23
2	16-08	8
3	24-16	5
4	32-24	5

ب. **جداول مفتوحة من الأعلى:** وهي الجداول التي يكون فيها الحد الأعلى للفئة الأخيرة غير محدد.

- **مثال :**

الرقم	الفئات	التكرار المطلق
1	08-00	23
2	16-08	8
3	24-16	5
4	-24	5

ج. **جداول مفتوحة من الطرفين:** وهي الجداول التي يكون فيها الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة غير محددين[3].

- **مثال:**

الرقم	الفئات	التكرار المطلق
1	08-	23
2	16-08	8
3	24-16	5
4	-24	5

3- حساب مختلف التكرارات:

من خلال التكرار المطلق لأي ظاهرة يمكن حساب مختلف التكرارات والتي لها دلالات معينة يمكن التطرق غليها في البنود الموالية:

2-1- التكرارات المتجمعة الصاعدة ($n_i \nearrow$) والنازلة ($n_i \swarrow$): تحسب التكرارات المتجمعة الصاعدة والنازلة لمختلف المتغيرات ما عدا تلك الكيفية المقاسة بالمقياس الإسمي .

ويجيبنا التكرار المتجمع الصاعد عند قيمة معينة للمتغير عن ما هو مجموع التكرارات السابقة عندما يأخذ المتغير هذه القيمة أو أقل في حالة المتغير الكيفي ذو المقياس الترتيبي والمتغير الكمي المتقطع بدون فئات، أما في حالة

الفئات (متغير منفصل أو متصل) فهو مجموع التكرارات لما يأخذ المتغير قيمة الحد الأعلى للفئة المقابلة أو اقل، ويمكن صياغة هذا رياضيا كما يلي:

$$n_m \nearrow = \sum_{i=1}^{m} n_i$$

ويجيبنا التكرار المتجمع النازل عند قيمة معينة للمتغير عن ما هو مجموع التكرارات اللاحقة عندما يأخذ المتغير هذه القيمة أو أكثر في حالة المتغير الكيفي ذو المقياس الترتيبي والمتغير الكمي المتقطع بدون فئات ، أما في حالة الفئات (متغير منفصل أو متصل) فهو مجموع التكرارات لما يأخذ المتغير قيمة الحد الأدنى للفئة المقابلة أو أكثر ، ويمكن صياغة هذا رياضيا كما يلي:

$$\begin{cases} n_1 \swarrow = n \\ n_m \swarrow = n - \sum_{i=1}^{m-1} n_i \ ; m \neq 1 \end{cases}$$

حيث:

i هو أرقام السطور (أرقام قيم المتغير بالترتيب) في الجدول

m هو رقم السطر الذي نريد حساب فيه التكرار المتجمع

n حجم العينة (مجموع التكررات)

- **مثال (1)** : يبين الجدول الموالي توزيع تكراري مع طريقة حساب التكرار المتجمع الصاعد والنازل وحدود المتغير عند حسابة في حالة متغيرة منفصلة، ويكون الأمر مشابها في حالة مغيرة كيفية ترتيبية:

المتغير	التكرار المطلق	**التكرار المتجمع الصاعد (ت م ص)**		**التكرار المتجمع النازل (ت م ن)**	
x_i	n_i	**حدود المتغير**	$n \nearrow$	**حدود المتغير**	$n \swarrow$
0	2	**0 أو أقل**	2	**0 أو أكثر**	25
1	5	**1 أو أقل**	7	**1 أو أكثر**	23
2	10	**2 أو أقل**	17	**2 أو أكثر**	18
3	5	**3 أو أقل**	22	**3 أو اكثر**	8
4	3	**4 أو أقل**	25	**4 أو أكثر**	3
Σ	25			**5 أو أكثر**	**0**

- **مثال (2):** يبين الجدول الموالي توزيع تكراري وطريقة حساب التكرار المتجمع الصاعد والنازل وحدود المتغير عند حسابهما في حالة بيانات مبوبة بالفئات:

المتغير x_i	التكرار المطلق n_i	التكرار المتجمع الصاعد (ت م ص)		التكرار المتجمع النازل (ت م ن)	
		حدود المتغير العليا	$n\nearrow$	حدود المتغير الدنيا	$n\swarrow$
05-00	20	**5 أو أقل**	2	**0 أو أكثر**	250
10-05	50	**10 أو أقل**	7	**5 أو أكثر**	230
20-10	100	**20 أو أقل**	17	**10 أو أكثر**	180
30-20	50	**30 أو أقل**	22	**20 أو اكثر**	80
40-30	30	**40 أو أقل**	25	**30 أو أكثر**	30
Σ	250			**40 أو أكثر**	**0**

2-2- التكرارات النسبية f_i والنسبية المئوية $f_i^{\%}$: تعرف التكرارات النسبية بأنها حصيلة قسمة التكرار المطلق على مجموع التكرارت، حيث يكون مجموع التكرارات النسبية في الأخير يساوي الواحد الصحيح، وله نفس مفهوم الإحتمال حيث يكون إحتمال الحصول على القيمة المعنية للمتغير أو أي قيمة من الفئة المعنية (في حالة الفئات) عند اختيار عنصر أو مفردة بطريقة عشوائية مساويا للتكرار النسبي المقابل لتلك القيمة أو الفئة[4].

$$f_i = \frac{n_i}{\sum_{i=1}^{k} n_i}$$

حيث k هو عدد الفئات أو قيم المتغير بغض النضر عن التكرارات.

و $\sum_{i=1}^{k} n_i = n$ أي حجم العينة. أو $\sum_{i=1}^{k} n_i = N$ أي حجم المجتمع.

أما التكرار النسبي المئوي فهو حصيلة ضرب التكرار النسبي في 100 % ويعني النسبة المئوية للقيمة المقابلة أو الفئة المعنية من حجم العينة.

$$f_i^{\%} = \frac{n_i}{\sum_{i=1}^{k} n_i} \times 100\ \% = f_i \times 100\ \%$$

- **مثال** : يبين المثال التالي توزيع تكراري مع التكرار النسبي والنسبي المئوي.

المتغير	ت مط n_i	**التكرار النسبي (ت ن)** f_i	**التكرار النسبي المئوي (ت ن م)** $f_i^{\%}$
0	2	**0.08**	**8 %**
1	5	**0.20**	**20 %**
2	10	**0.40**	**40 %**
3	5	**0.20**	**20 %**
4	3	**0.12**	**12 %**
Σ	25	**1**	**100 %**

2-3- التكرارات المتجمعة النسبية والنسبية المئوية: تحسب التكرارات المتجمعة النسبية الصاعدة والنازلة إما من التكرار النسبي وتكون حدود المتغير فيها مشابهة تماما لكل من ت م ص ، و ت م ن، على الترتيب، وتكون طريقة الجمع والطرح مماثلة لهما أيضا:

بالنسبة للتكرار المتجمع الصاعد النسبي يعطى بالصيغة:

$$f_m \nearrow = \sum_{i=1}^{m} f_i$$

بالنسبة للتكرار المتجمع النازل النسبي يعطى بالصيغة:

$$\begin{cases} f_1 \swarrow = 1 \\ f_m \swarrow = 1 - \sum_{i=1}^{m-1} f_i \ ; m \neq 1 \end{cases}$$

وإما تحسب من التكرارات المتجمعة الصاعدة والنازلة على الترتيب عن طريق قسمتهما على مجموع التكرارات:

بالنسبة للتكرار المتجمع الصاعد النسبي يعطى بالصيغة:

$$f_i \nearrow = \frac{n_i \nearrow}{\sum_{i=1}^{k} n_i}$$

بالنسبة للتكرار المتجمع النازل النسبي يعطى بالصيغة:

$$f_i \swarrow = \frac{n_i \swarrow}{\sum_{i=1}^{k} n_i}$$

وتحسب التكرارات المتجمعة النسبية المئوية الصاعدة والنازلة إما من التكرار النسبي المئوي وتكون حدود المتغير فيها مشابهة تماما لكل من ت م ص ، و ت م ن، على الترتيب، وتكون طريقة الجمع والطرح مماثلة لهما أيضا:

بالنسبة للتكرار المتجمع الصاعد النسبي المئوي يعطى بالصيغة:

$$f_m{}^{\%} \nearrow = \sum_{i=1}^{m} f_i{}^{\%}$$

بالنسبة للتكرار المتجمع النازل النسبي المئوي يعطى بالصيغة:

$$\begin{cases} f_1{}^{\%} \swarrow = 100\ \% \\ f_m{}^{\%} \swarrow = (100 - \sum_{i=1}^{m-1} f_i{}^{\%})\ \% \quad ; m \neq 1 \end{cases}$$

وإما تكون التكرارات المتجمعة النسبية المئوية الصاعدة والنازلة هي حصيلة ضرب كل من التكرار النسبي الصاعد والنازل على الترتيب في 100 % كما يلي:

$$f_i^{\%} \nearrow = \frac{n_i \nearrow}{\sum_{i=1}^{k} n_i} \times 100\ \% = f_i \nearrow \times 100\ \%$$

$$f_i^{\%} \swarrow = \frac{n_i \swarrow}{\sum_{i=1}^{k} n_i} \times 100\ \% = f_i \swarrow \times 100\ \%$$

- **مثال (1)**: يمثل الجدول التالي توزيع تكراري تم حساب له التكرارات المتجمعة النسبية والنسبية المئوية الصاعدة :

المتغير x_i	ت مط n_i	ت م ص $n \nearrow$	ت ن f_i	ت ن م $f_i^{\%}$	**ت م ن م ص** $\boldsymbol{f_i} \nearrow$	**ت م ن ص** $\boldsymbol{f_i^{\%}} \nearrow$
0	2	2	0.08	8 %	**0.08**	**8 %**
1	5	7	0.20	20 %	**0.28**	**28 %**
2	10	17	0.40	40 %	**0.68**	**68 %**
3	5	22	0.20	20 %	**0.88**	**88 %**
4	3	25	0.12	12 %	**1**	**100 %**
Σ	25		1	100		

- **مثال** (**2**): يمثل الجدول التالي توزيع تكراري تم حساب له التكرارات المتجمعة النسبية والنسبية المئوية النازلة:

المتغير x_i	ت مط n_i	ت م ن $n \swarrow$	ت ن f_i	ت ن م $f_i^{\%}$	ت م ن ن $f_i \swarrow$	ت م ن م ن $f_i^{\%} \swarrow$
0	2	25	0.08	8 %	**1**	100 %
1	5	23	0.20	20 %	**0.92**	92 %
2	10	18	0.40	40 %	**0.72**	72 %
3	5	8	0.20	20 %	**0.32**	32 %
4	3	3	0.12	12 %	**0.12**	12 %
Σ	25	0	1	100 %	**0**	**0 %**

- **مثال** (**3**): تم حساب جميع أنواع التكرارات في التوزيع التكراري التالي الذي يمثل عدد أطفال أسر 25 عامل في مؤسسة الأجر والقرميد في بلدة معينة:

المتغير x_i	**ت مط** n_i	**ت م ص** $n \nearrow$	**ت م ن** $n \swarrow$	**ت ن** f_i	**ت ن م** $f_i^{\%}$	**ت م ن ص** $f_i \nearrow$	**ت م ن ن** $f_i \swarrow$	**ت م ن م ص** $f_i^{\%} \nearrow$	**ت م ن م ن** $f_i^{\%} \swarrow$
0	**2**	2	25	0.08	8 %	0.08	1	8 %	100 %
1	**5**	7	23	0.20	20 %	0.28	0.92	28 %	92 %
2	**10**	17	18	0.40	40 %	0.68	0.72	68 %	72 %
3	**5**	22	8	0.20	20 %	0.88	0.32	88 %	32 %
4	**3**	25	3	0.12	12 %	1	0.12	100 %	12 %
Σ	**25**		0	1	100 %		0		0 %

4- جدول التوزيع التكراري المزدوج (التقاطعي):

رأينا فيما سبق جدول التوزيع التكراري البسيط حيث كان يعبر عن خاصية مدروسة واحدة أو متغير واحد بالإضافة إلى التكرارات المقابلة (هنا لدينا جدول ذو بعدين) لكن أحيانا كثير يتم إضافة خاصية مدروسة أو متغير آخر حيث يصبح لدينا خاصيتين تتوزع التكرارات تقاطعيا بينهما أي يصبح لدينا خاصيتين بالإضافة إلى التكرارات المقابلة (هنا جدول دو ثلاثة أبعاد)

بإفتراض لدينا متغيرتين X، Y تأخذان مجموعة من القيم، فالشكل العام للجدول المزدوج للتوزيع التكراري لهما يأخذ الشكل التالي:

جدول 2–2 : الشكل العام لجدول التوزيع التكراري المزدوج.

$\sum_{j=1}^{m} n_{ij}$		y_m	⋯	y_j	⋯	y_4	y_3	y_2	y_1	y_j / x_i
التكرارات الهامشية للمتغيرة X	$n_1.$	n_{1m}	⋯	n_{1j}	⋯	n_{14}	n_{13}	n_{12}	n_{11}	x_1
	$n_2.$	n_{2m}	⋯	n_{2j}	⋯	n_{24}	n_{23}	n_{22}	n_{21}	x_2
	$n_3.$	n_{3m}	⋯	n_{3j}	⋯	n_{34}	n_{33}	n_{32}	n_{31}	x_3
	⋮	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋮
	$n_i.$	n_{im}	⋯	n_{ij}	⋯	n_{i4}	n_{i3}	n_{i2}	n_{i1}	x_i
	⋮	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋮
	$n_n.$	n_{nm}	⋯	n_{nj}	⋯	n_{n4}	n_{n3}	n_{n2}	n_{n1}	x_n
$\sum_{i=1}^{n} n_{i.} = \sum_{j=1}^{m} n_{.j}$		$n_{.m}$	⋯	$n_{.j}$	⋯	$n_{.4}$	$n_{.3}$	$n_{.2}$	$n_{.1}$	$\sum_{i=1}^{n} n_{ij}$
		التكرارات الهامشية للمتغيرة Y								

n_{ij} : التكرارات الثنائية.

المصدر: جلاطو جيلالي "الإحصاء مع تمارين ومسائل محلولة" د. م. ج، 1998، الجزائر، ص 16. (بتصرف)

من خلال الجدول نلاحظ أن التكرارات تكون مشتركة بين الخاصيتين المدروستين حيث مجموعها الكلي يمثل حجم العينة n ويمكن التعبير عنه بعدة طرق ، ونكتب الصيغة التالية:

$$\sum_{i=1}^{n} n_{i.} = \sum_{j=1}^{m} n_{.j} = n = \sum_{i=1}^{n}\sum_{j=1}^{m} n_{ij}$$

- **مثال:** أخذت عينة تتكون من 720 طالب من ذكور وإناث من كلية الإقتصاد بجامعة المدية فكانت تخصصاتهم وتوزيع الذكور والإناث على التخصصات موضح في الجدول الموالي:

		التخصص			المجموع حسب الجنس
		ع. إقتصادية	ع. تسيير	ع. تجارية	
الجنس	ذكور	100	150	80	330
	إناث	120	180	90	390
المجموع حسب التخصص		220	330	170	**720**

يمثل الجدول السابق جدولا ذو ثلاثة أبعاد (جنس، تخصص، تكرار) حيث يسمى أحيانا جدولا تقاطعيا، أو جدولا مزدوجا.

5- جدول التوزيع التكراري المتعدد (التقاطعي المتعدد):

في هذا النوع من الجداول يتم إدراج ثلاثة خاصيات أو أكثر، تتقاطع عن طريق التكرارات المقابلة، وتكون المجاميع التكرارية واضحة بالنسبة لكل خاصية على حدى[5].

- **مثال:** في أحد المراكز الجامعية وفي كلية الإقتصاد تحديدا والتي تضمن دراسة ثلاثة تخصصات، كان عدد الطلبة المسجلين خلال ثلاثة سنوات دراسية موضح حسب جنس الطالب في الجدول التالي:

السنوات الدراسية		2015/2014				2016/2015				2017/2016			
التخصصات		إقتصاد	تسيير	تجارة	**مج**	إقتصاد	تسيير	تجارة	**مج**	إقتصاد	تسيير	تجارة	**مج**
الجنس	ذكور	100	150	80	**330**	110	160	90	**360**	120	150	100	**370**
	إناث	120	180	90	**390**	125	185	95	**405**	110	150	80	**340**
المجاميع (ت، س)		220	330	170	**720**	235	345	185	**765**	230	300	180	**710**

من خلال الجدول نرى أن هناك ثلاثة متغيرات تتقاطع فيما بينها (السنة الدراسية، الجنس، التخصص) تتقاطع فيما بينها عن طريق التكرارات المقابلة. حيث بلغ حجم العينة 720، 765، 710 كل سنة دراسية على التوالي.

المبحث الثاني
العرض البياني للبيانات

تختلف طرق عرض البيانات حسب نوعية البيانات التي نحن بصدد عرضها من ناحية الكيفية والكمية، من جهة ومن حيث أنها كمية متصلة أو منفصلة من جهة أخرى، نتطرق فيما يلي لكل نوع على حدى.

1- العرض البياني للبيانات الكيفية:

يتم تمثيل البيانات الكيفية (النوعية) عن طريق الدائرة والأعمدة البيانية بمختلف أنواعها، نتطرق فيما يلي مختلف الأصناف وضمنيا نذكر حالتين مهمتين عند التمثيل وهي عندما نكون أمام متغير إحصائي واحد فقط وحالة أكثر من متغير في آن واحد.

1-1- التمثيل بالدائرة : يتم في هذا النوع من التمثيلات حساب زاوية القطاع المقابلة لكل تكرار كما يلي[6]:

زاوية القطاع=التكرار النسبي × 360°

وهذا يأتي من الآتي:

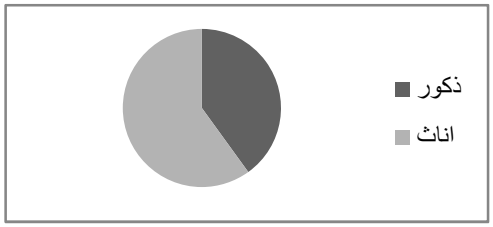
$$\begin{cases} n \to 360° \\ n_i \to x° \end{cases} \Rightarrow x° = \frac{n_i}{n} \times 360° = f_i \times 360°$$

- **مثال:** ليكن التوزيع التكراري التالي الذي يبين متغيرة الجنس لعينة تتكون من 100 فرد، والمطلوب تمثيله بواسطة تمثيل دائري:

المتغيرة (الجنس)	التكرار المطلق	التكرار النسبي	زاوية القطاع
ذكور	40	0.4	°144
إناث	60	0.6	°216
المجموع	100	1	°360

إذن بعد حساب زاوية القطاع للذكور والإناث، نقوم وبإستخدام أداة "منقلة" برسم الزاويتين، مع توضيح مفتاح الرسم والعنوان المناسب (وإن كان هناك مصدر أو برنامج الرسم يجب توضيحه أسفل التمثيل):

دائرة بيانية تمثل توزيع الجنس لعينة من الأفراد



المصدر: بيانات افتراضية بإستخدام برمجية *Excel*

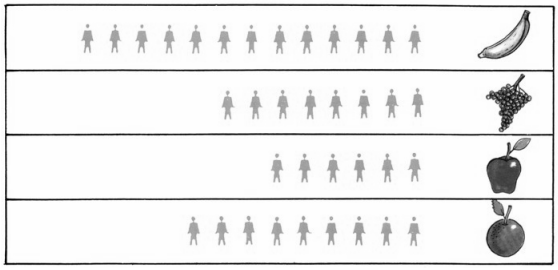
1-2- التمثيل بالصور البيانية والرسومات: يتم في هذا النوع من التمثيلات لإعطاء صورة بسيطة كرمز لنوع الحالة التي تتكرر[7]، مثل صورة لشعار شركة معينة أو صورة تعبر عن فرد، أو شجيرات ... الخ.

- **مثال:** الجدول الموالي يوضح الفاكهة المفضلة عند 358 شخص:

نوع الفاكهة (المتغير)	التكرار (عدد الأفراد)
الموز	130
العنب	80
التفاح	61
البرتقال	88
المجموع	358

التمثيل عن طريق الصور: نلاحظ أنه في بعض التكرارات كانت تقريبية فقط. مثل 61، 88. حيث يظهر من خلال الرسم أنها 60، 90 على الترتيب:

العنوان: صور بيانية تمثل الفاكهة المفضلة لعينة من الأفراد



تمثل 10 أشخاص

المصدر: بيانات افتراضية

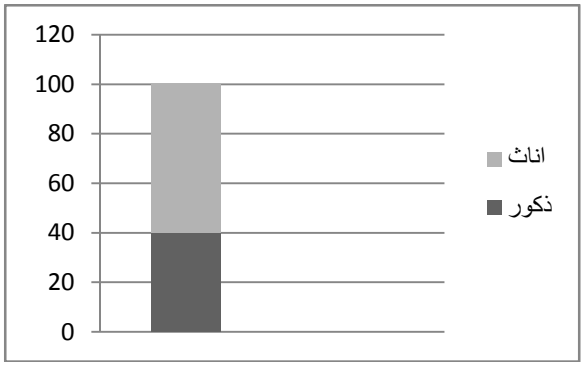
1-3- العمود المجزأ العادي: نقوم عنا بحساب المجموع الكلي للتكرارات (حجم العينة) ثم نرسمه في عمود واحد ويتم تجزأته حسب التكرارات المطلقة المقابلة.

- **مثال:** ليكن التوزيع التكراري التالي، والمطلوب تمثيله عن طريق العمود المجزأ العادي:

الجنس	التكرار
ذكور	40
اناث	60
المجموع	100

يتم استخدام التكرارات المطلقة مباشرة عند التمثيل البياني للعمود المجزأ العادي، كما يلي:

العنوان: عمود مجزأ يمثل توزيع الجنس لعينة من الأفراد



المصدر: بيانات افتراضية إعتمادا على برمجية *Excel*

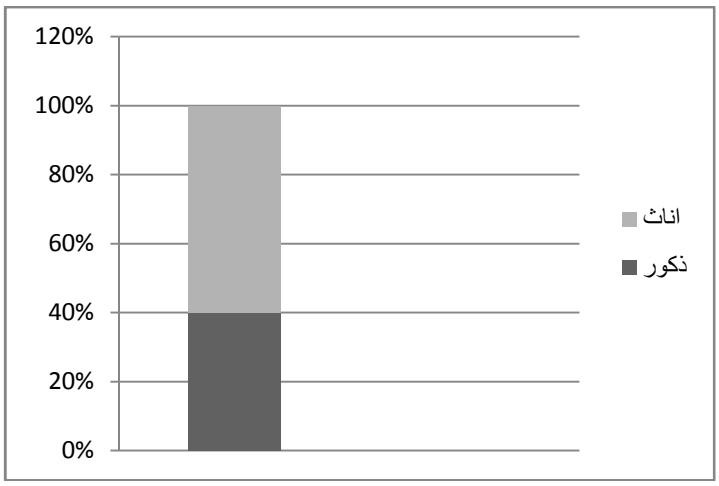
1-4- العمود المجزأ النسبي [متغير واحد] : يتم الاعتماد في التمثيل من هذا النوع على التكرار النسبي المئوي وليس المطلق، وهو مفيد في المقارنة بين الرسومات التي ليس لها نفس وحدة القياس، لأن النسبة المئوية تقوم بتحييد وحدة القياس (توحيدها)

- **مثال:** ليكن التوزيع التكراري التالي والمطلوب تمثيله عن طريق العمود المجزأ النسبي:

المتغيرة	التكرار المطلق	التكرار النسبي المئوي
ذكور	40	40 %
اناث	60	60 %
المجموع	100	100 %

بعد حساب التكرار النسبي المئوي نعتمد عليه أثناء الرسم:

العنوان: عمود مجزأ نسبي يمثل توزيع الجنس لعينة من الأفراد



المصدر: بيانات افتراضية إعتمادا على برمجية *Excel*

1-5- الأعمدة البيانية المستطيلة: ويتم تخصيص عمود مستطيل لكل قيمة يأخذها المتغير ويكون طولها مساويا للتكرار المطلق المقابل[8]:

- **مثال:** ليكن التوزيع التكراري التالي، والمطلوب تمثيله عن طريق الأعمدة البيانية المستطيلة:

الجنس	التكرار المطلق
ذكور	40
اناث	60

ويأخذ التمثيل البياني الشكل التالي:

العنوان: أعمدة بيانية مستطيلة تمثل توزيع الجنس لعينة من الأفراد



المصدر: بيانات افتراضية إعتمادا على برمجية *Excel*

2- العرض البياني للبيانات الكمية المنفصلة (بدون فئات):

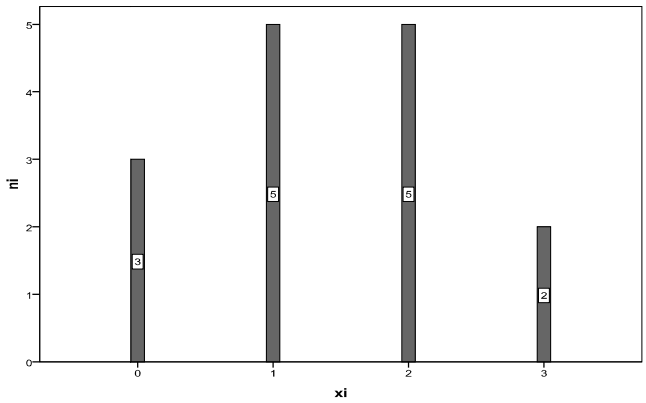
2-1- الأعمدة البيانية البسيطة: حيث يتم رسم عمود بسيط لكل قيمة يأخذها المتغير المدروس، وطول هذه الأعمدة يتناسب مع التكرار المطلق المقابل لكل قيمة.

- **مثال:** ليكن التوزيع التكراري التالي، والمطلوب تمثيله بيانيا:

عدد الأطفال (المتغير) x_i	عدد الأسر (التكرار) n_i
0	3
1	5
2	5
3	2
المجموع	15

نلاحظ أن المتغيرة كمية منفصلة، إذن نمثلها عن طريق الأعمدة البيانية البسيطة، كما يلي:

عنوان الرسم: أعمدة بيانية بسيطة تمثل توزيع عدد الأطفال لعينة من الأسر



مصدر الرسم: بيانات افتراضية إعتمادا على برمجية *SPSS* [*]

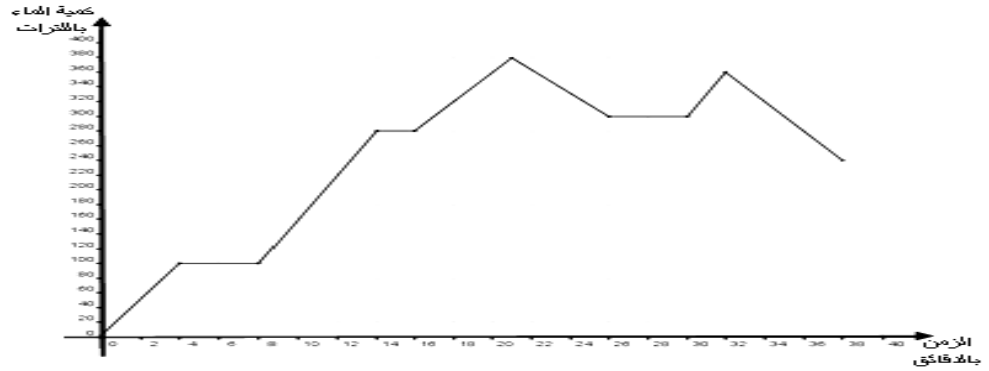
2-2- الخط البياني [سلاسل زمنية]: يعتبر الخط البياني من أشهر التمثيلات البيانية، حيث عندما تكون ظاهرة تتطور بدلالة الزمن، يتم تمثيل الزمن على محور الفواصل والمتغيرة المدروسة على محور التراتيب ثم يتم رسم النقاط المقابلة للإحداثيات (قيمة الظاهرة، الزمن)، ومن ثمة إيصال هذه النقاط بقطع مستقيمة، لتشكل لنا خطا بيانيا يصف لنا تطور الظاهرة عبر الزمن[9].

- **مثال**: لتكن السلسلة الزمنية الإفتراضية التالية، والمطلوب تمثيلها بيانيا:

الزمن بالدقائق	0	5	10	.....	40
عدد الليترات المسقاة	0	40	45	.....	240

بما أن الظاهرة تتطور عبر الزمن نمثلها عن طريق الخط البياني كما يلي:

العنوان: خط بياني يمثل تطور عدد الليترات المسقاة حسب الزمن



المصدر: بيانات افتراضية

[*] برمجية SPSS هي برمجية خاصة بالحساب في مجال الإحصاء والإقتصاد القياسي، وتسمى "الحزمة الإحصائية للعلوم الإجتماعية"، وحروف اسم البرمجية هي إختصار لإسم البرمجية باللغة الإنجليزية التالية: **S**tatistical **P**ackage for the **S**ocial **S**cences.

3- العرض البياني للمتغيرات الكمية المتصلة والمنفصلة المبوبة بالفئات:

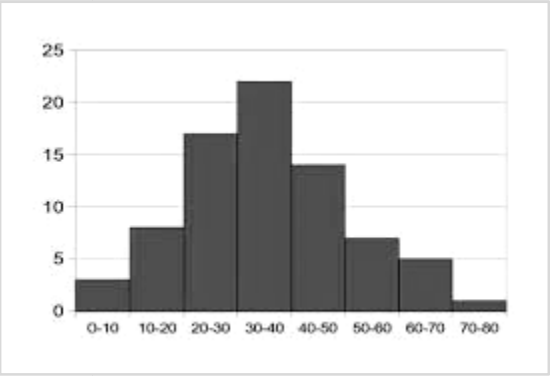
3-1- المدرج التكراري: يشبه الأعمدة البيانية المستطيلة إلا أنها تكون متلاسقة مع بعضها، نظرا لطبيعة المتغير، حيث يتم رسم محورين، يتم في المحور الأفقي (الفواصل) رسم الحدود الفعلية للفئات، وفي المحور العمودي (التراتيب) يتم تدريجه وفق سلم مناسب ليتم إسقاط التكرارات المطلقة عليه ومن ثمة تكون إستطالة كل عمود مقابل لفئة معينة حسب تكرارها المطلق[10]، وفيما يلي مثالا بسيطا على هذا النوع من التمثيلات البيانية:

- **مثال تطبيقي:** ليكن التوزيع التكراري :

فئات المتغير	التكرار المطلق
10-00	3
20-10	8
30-20	12
40-30	23
50-40	14
60-50	11
70-60	5
80-70	1
المجموع	57

بما أن البيانات نوعها كمي، ومبوبة بالفئات، وتجدر الإشارة إلى أن الفئات متساوية، حيث إذا لم تكن متساوية يستوجب إجراء تعديلا للتكرارات عند التمثيل البياني، وهذا نتطرق إليه لاحقا. إذن فإن التمثيل عن طريق المدرج التكراري يأخذ الشكل التالي:

العنوان: مدرج تكراري يمثل متغيرة ما



المصدر: بيانات افتراضية

3-2- المضلع التكراري: حيث يتم رسم محورين، يتم في المحور الأفقي (الفواصل) رسم مراكز الفئات، وفي المحور العمودي (التراتيب) يتم تدريجه وفق سلم مناسب ليتم إسقاط التكرارات المطلقة عليه، ثم رسم النقاط التي

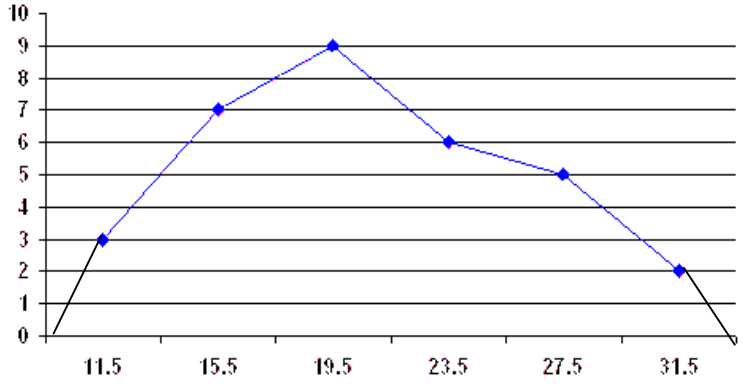
إحداثياتها مراكز الفئات والتكرارات المقابلة لها ثم نصل هذه النقاط بقطع مستقيمة، لتشكل لنا مضلعا مفتوحا، وبغرض غلق هذا المضلع، يتم إضافة مركزي فئة وهميين قبلي (لفئة يكون طولها مساوي لطول الفئة الأولى) وبعدي (لفئة يكون طولها مساويا لطول الفئة الأخيرة) يكون تكراراهما مساويا للصفر ليتم غلق المضلع التكراري[11]، وفيما يلي مثالا بسيطا على هذا النوع من التمثيلات البيانية:

- **مثال تطبيقي**: ليكن التوزيع التكراري :

فئات المتغير	التكرار المطلق	مراكز الفئات x_i
13-9	3	11.5
17-13	7	15.5
21-17	9	19.5
25-21	6	23.5
29-25	5	23.5
33-29	2	31.5
المجموع	32	/

بما أن البيانات نوعها كمي، ومبوبة بالفئات، وتجدر الإشارة إلى أن الفئات متساوية، حيث وكما أشرنا سابقا إذا لم تكن متساوية يستوجب إجراء تعديلا للتكرارات عند التمثيل البياني، وهذا نتطرق إليه لاحقا عند التحدث عن الجداول غير المنتظمة وعند حساب مقياس النزعة المركزية المسمى : المنوال. إذن فإن التمثيل عن طريق المضلع التكراري يأخذ الشكل التالي:

العنوان: مدرج تكراري يمثل متغيرة ما



المصدر: بيانات افتراضية

3-3- المنحنى التكراري: حيث يتم رسم محورين، يتم في المحور الأفقي (الفواصل) رسم مراكز الفئات، وفي المحور العمودي (التراتيب) يتم تدريجه وفق سلم مناسب ليتم إسقاط التكرارات المطلقة عليه ثم رسم النقاط التي إحداثياتها مراكز الفئات والتكرارات المقابلة لها ثم نصل هذه النقاط بخطوط منحنية وفق المسار العام لتوزع هذه

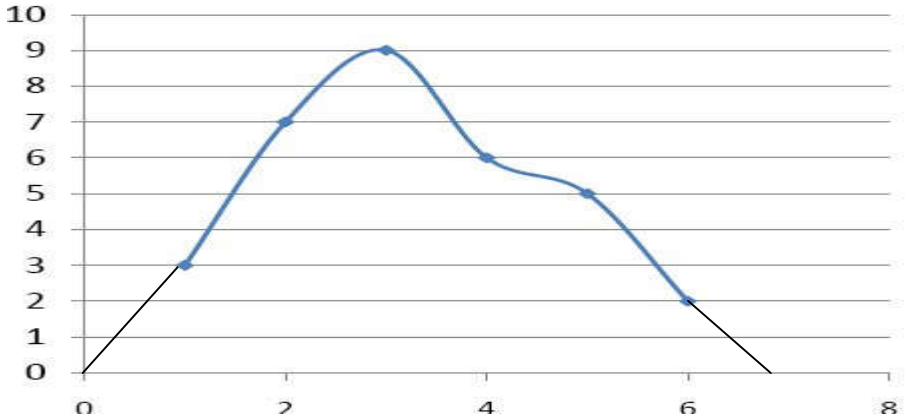
النقاط تراتبيا، لتشكل لنا منحنى مفتوحا، وبغرض غلق هذا المنحنى، نفعل ما فعلناه مع المضلع التكراري وذلك بتحديد مركزي فئة وهميين قبلي (لفئة يكون طولها مساوي لطول الفئة الأولى) وبعدي (لفئة يكون طولها مساويا لطول الفئة الأخيرة) يكون تكراراهما مساويا للصفر ليتم غلق المضلع التكراري [12]، وفيما يلي مثالا بسيطا على هذا النوع من التمثيلات البيانية:

- **مثال تطبيقي**: ليكن التوزيع التكراري :

فئات المتغير	التكرار المطلق	مراكز الفئات x_i
1.5-0.5	3	1
2.5-1.5	7	2
3.5-2.5	9	3
4.5-3.5	6	4
5.5-4.5	5	5
6.5-5.5	2	6
المجموع	32	/

بما أن البيانات نوعها كمي، ومبوبة بالفئات، وتجدر الإشارة إلى أن الفئات متساوية، حيث وكما أشرنا سابقا إذا لم تكن متساوية يستوجب إجراء تعديلا للتكرارات عند التمثيل البياني وذلك حتى لا يكون التمثيل مضللا، وهذا نتطرق إليه لاحقا. إذن فإن التمثيل عن طريق المنحنى التكراري يأخذ الشكل التالي:

العنوان: منحنى تكراري يمثل ظاهرة معينة



المصدر: بيانات افتراضية

3-4- حالة الجداول غير المنتظمة (أطوال الفئات غير متساوية):

إذا كانت هناك فئات التوزيع غير متساوية وأردنا التمثيل البياني نقوم بتعديل التكرارات، حتى يكون هناك تناسب بين طول الفئة والتكرار المقابل لها، وقيم تعديل التكرارات باستخدام المعادلة الآتية:

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

- **مثال تطبيقي:** يبين الجدول التالي توزيع عينة من 100 عامل حسب الأجر اليومي المطلوب تمثيل هذه البيانات باستخدام المدرج التكراري؟.

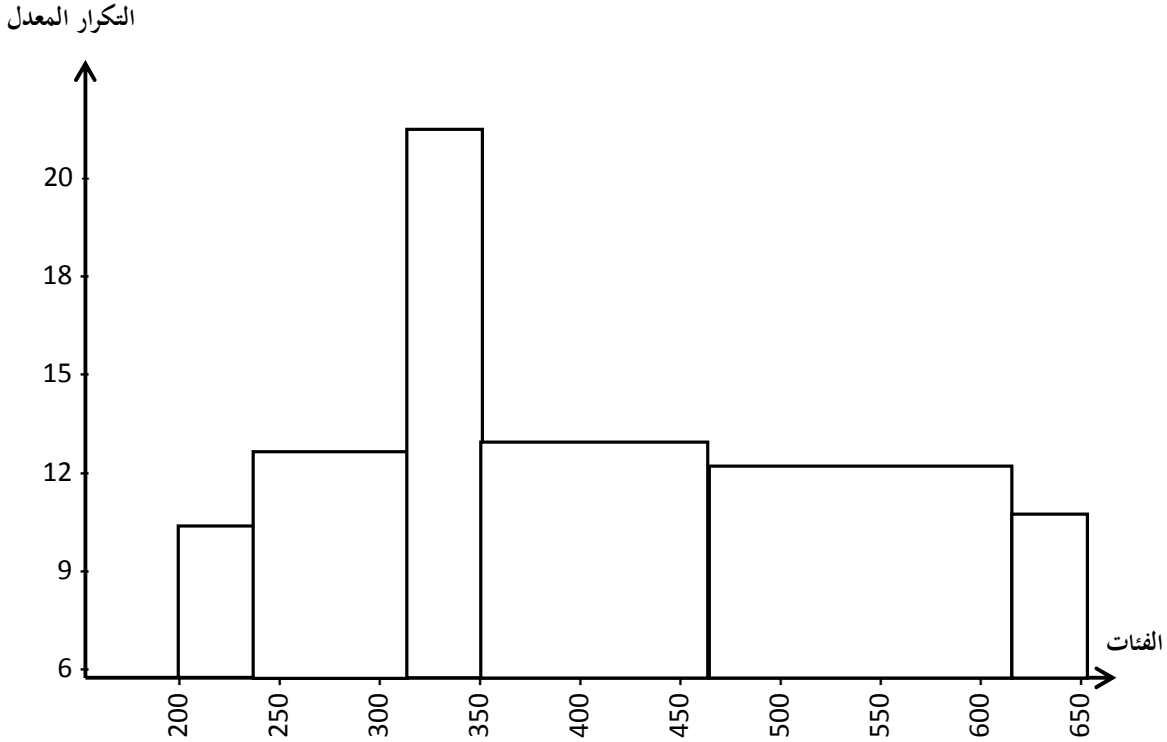
فئة الأجر	عدد العمال
]250 – 200]	5
]350 – 250]	15
]400 – 350]	20
]550 – 400]	25
]750 – 550]	30
[800 – 750]	5
المجموع	100

- **الحل:** بما أن فئات التوزيع غير متساوية فإننا من أجل رسم المدرج التكراري نقوم بتعديل تكرار هذه الفئات وفقا للمعادلة السابقة.

فئة الأجر	عدد العمال	طول الفئة	**التكرار المعدل**
]250 – 200]	5	**50**	5
]350 – 250]	15	10	7.5
]400 – 350]	20	**50**	20
]550 – 400]	25	15	8.33
]750 – 550]	30	20	7.5
[800 – 750]	5	**50**	5
المجموع	100	/	/

تم تعديل التكرارات بالإعتماد على العلاقة السابقة، حيث نلاحظ أن طول الفئة الذي يتكرر كثيرا هو 50، ونلاحظ من خلال التكرارات المعدلة أن الفئات التي طولها 50 لم يتغير تكرارها في حين التي طولها اقل من 50 خفض والعكس بالنسبة للتي طولها اكبر من50 حيث ارتفع تكرارها المعدل بالنسبة لتكرارها المطلق، ويتم التمثيل بيانيا كما يلي: (نختار المدرج التكرار)

العنوان: مدرج تكراري بالتكرارات المعدلة لمتغيرة ما



المصدر: بيانات افتراضية

- **ملاحظة:** نقوم بتعديل التكرارات في حالة الفئات غير المتساوية في الحالتين التاليتين:
  - عند رسم المدرج التكراري (المضلع والمنحنى التكراريين أيضا)
  - عند تحديد الفئة المنوالية وحساب المنوال (نعود له لاحقا)

4- العرض البياني للجداول المزدوجة:

ما يميز الجداول المزدوجة هي أنها تحتوي على متغيرتين، مما يجعل من عملية التمثيل البياني أعقد من الحالة الأولى (الجداول البسيطة) حيث ينبغي معرفة نوع كل خاصة -متغيرة- حيث كما تكونان من نفس النوع قد تكونان من نوعين مختلفين، فمثلا قد تكون واحدة كيفية ترتيبية والأخرى كيفية اسمية، أو واحدة كمية والأخرى كيفية بغض النظر عن أي صنف فيهما، ونتناول أهم التمثيلات البيانية فيما يلي:

4-1- الأعمدة المجزأة:

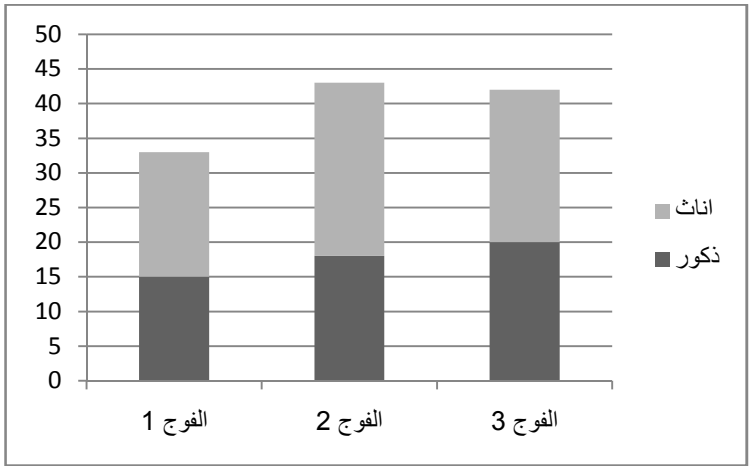
حيث إذا كان هناك متغيرين يتم الاستعانة بالأعمدة المجزأة حيث يتم إعطاء المحور الأفقي لخاصة واحدة ترسم قيمها بأعمدة بيانية مستطيلة يكون طولها حسب التكرارات الهامشية للخاصية الأولى (يتم حساب المجاميع لهذه الخاصية)، وتجزأ حسب الخاصية الثانية ليكون المحور العمودي يعبر عن التكرارات، وذلك باستعمال التكرارات المطلقة (يمكن الاستعانة بالتكرارات النسبية المئوية وتسمى الأعمدة المجزأة النسبية)[13].

- **مثال تطبيقي**: ليكن التوزيع التكراري التالي، والمطلوب تمثيله بواسطة الأعمدة المجزأة:

	الفوج 1	الفوج 2	الفوج 3
ذكور	15	18	20
اناث	18	25	22
المجموع	33	43	42

يتم تخصيص المحور الأفقي لمتغيرة رقم الفوج، ويجزأ كل عمود حسب الجنس، أما محور التراتيب فيعبر عن التكرار المطلق، ويكون للرسم عنوانا واضحا مع مصدر كما جرت العادة، كما يلي:

العنوان: أعمدة بيانية مجزأة تعبر عن توزيع الجنس حسب الفوج



المصدر: بيانات افتراضية من الباحث (تم الاستعانة ببرمجية *Excel*)

ويمكن قلب المتغيرتين في الرسم، حيث يبقى للباحث حرية الاختيار أي من المتغيرين يضعه على المحور الأفقي حتى يكون الرسم أوضح.

4-2- الأعمدة المتلاصقة:

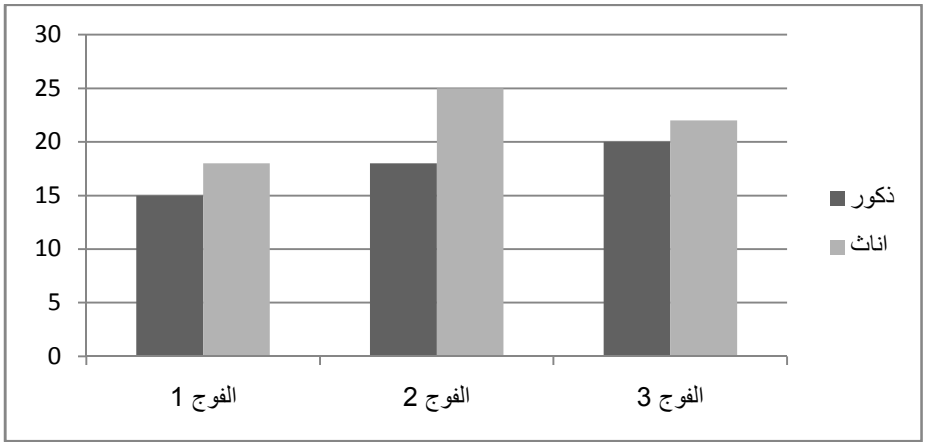
حيث إذا كان هناك متغيرين يتم الاستعانة بالأعمدة المتلاصقة، حيث يتم إعطاء المحور الأفقي لخاصة واحدة ترسم قيمها بأعمدة بيانية مستطيلة متلاصقة حسب عدد القيم للخاصية الأولى، ويكون طول كل واحدة منها حسب التكرارات المقابلة لقيم الخاصية الثانية، ليكون المحور العمودي يعبر عن التكرارات، وذلك باستعمال التكرارات المطلقة (يمكن الاستعانة بالتكرارات النسبية المئوية وتسمى الأعمدة المجزأة النسبية)[14].

- **مثال تطبيقي**: ليكن التوزيع التكراري التالي، والمطلوب تمثيله بواسطة الأعمدة المتلاصقة:

	الفوج 1	الفوج 2	الفوج 3
ذكور	15	18	20
اناث	18	25	22

يتم تخصيص المحور الأفقي لمتغيرة رقم الفوج، ونسم عمودين عند كل فوج حسب الجنس، أما محور التراتيب فيعبر عن التكرار المطلق، ويكون للرسم عنوانا واضحا مع مصدر كما جرت العادة، كما يلي:

العنوان: أعمدة بيانية متلاصقة تعبر عن توزيع الجنس حسب الفوج



المصدر: بيانات افتراضية من الباحث (تم الاستعانة ببرمجية *Excel*)

ويمكن قلب المتغيرتين في الرسم، حيث يبقى للباحث حرية الاختيار أي من المتغيرين يضعه على المحور الأفقى حتى يكون الرسم أوضح. مع مراعاة أن يكون للرسم عنوانا واضحا ومصدرا إما للبرمجية المستخدمة في الرسم او لمصدر البيانات أو لمصدر الرسم إن كان منقولا.

4-3- الخطوط البيانية المتعددة:

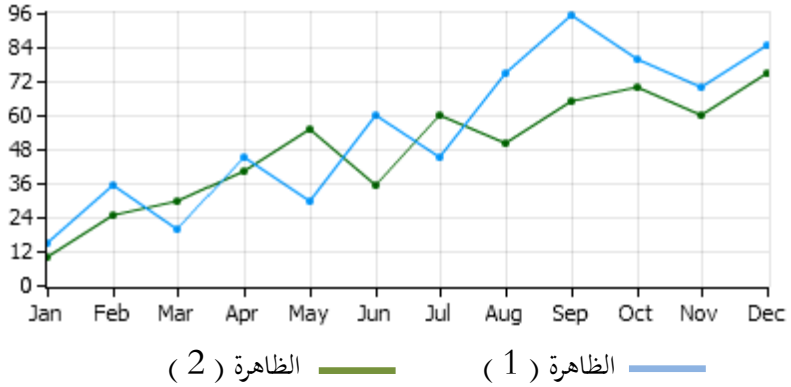
عندما تكون لدينا ظاهرة أو أكثر تتطوران بالنسبة للزمن، يتم الاستعانة بالخط البياني المتعدد، حيث يرسم الزمن على المحور الأفقي، ليتم تخصيص المحمور العمودي للظواهر المدروسة، ثم عن طريق رسم نقاط التقاطع للزمن مع قيم الظواهر، يتم وصل نقاط كل خاصية على حدى.

● **مثال تطبيقي**: ليكن جدول البيانات الافتراضية التالي، الذي يمثل تطور ظاهرتين اقتصاديتين خلال سنة 2017 حسب الأشهر:

الأشهر (2017)	جانفي	فيفري	مارس	افريل	ماي	جوان	جويلية	أوت	سبتمبر	اكتوبر	نوفمبر	ديسمبر
الظاهرة (1)	15	36	24	48	28	60	48	74	96	76	72	84
الظاهرة (2)	11	24	27	40	58	36	60	48	55	72	60	73

التمثيل البياني التالي يمثل تطور الظاهرتين الإفتراضيتين، حسب 12 شهر لسنة معينة (في محور الفواصل) ولتكن 2017، وتكون قيم الظاهرتين مأخوذة من المحور العمودي بعد إسقاطها عليه، حيث تم رسم الظاهرة (1) بلون مخالف للظاهرة (2) للتمييز بينهما ويذكر ذلك في مفتاح الرسم.

العنوان: خطوط بيانية تمثل تكور ظاهرة (1) ، والظاهرة (2) حسب الأشهر لسنة 2017



المصدر: بيانات افتراضية

المبحث الثالث

العرض البياني للتكرارات المتجمعة

التكرارات المتجمعة بكل أنواعها سواء الصاعدة، النازلة والنسبية، لها أهمية كبيرة في عملية تحليل البيانات واستخلاص النتائج وعملية تمثيلها لا تقل أهمية عن التمثيلات البيانية للظاهرة المدوسة عن طريق التكرارات العادية، ونذكر أهم التمثيلات فيما يلي.

1- حالة بيانات مبوبة بدون فئات (كمية منفصلة، ترتيبية):

يرسم منحنى التكرار المتجمع الصاعد عن طريق قطع مستقيمة أفقية تنتهي عند مجموعة النقاط ذات الإحداثيات التالية: قيمة المتغير والتكرار المتجمع الصاعد المقابل لها، وتبدأ من القيمة السابقة لكل قيمة، ويرسم منحنى التكرار المتجمع النازل عن طريق قطع مستقيمة أفقية تبدأ من مجموعة النقاط التي إحداثياتها: قيمة المتغير والتكرار المتجمع النازل مقابل لها، وينتهي طول القطع عندما تبدأ القيمة التالية للمتغير.

يبين كل من منحنى التكرار المتجمع الصاعد ومنحنى التكرار المتجمع النازل شدة أو ضعف تطور الظاهرة المدروسة عن مستوى معين من مجال الدراسة.

- **مثال تطبيقي:** أرسم على نفس المعلم كل من منحنى التكرار المتجمع الصاعد ومنحنى التكرار المتجمع النازل لبيانات التكراري الآتي:

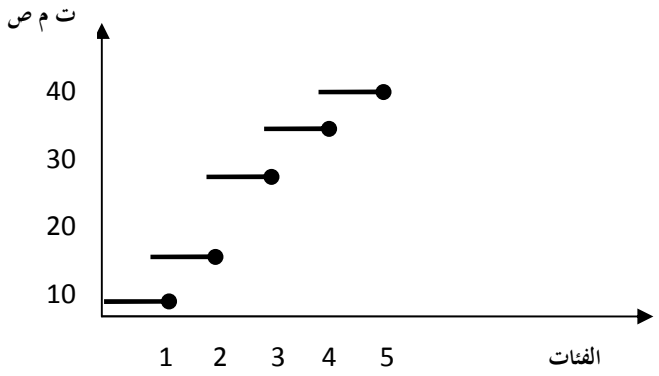
المتغير	التكرار
1	4
2	10
3	12
4	6
5	5
المجموع	37

الحل:

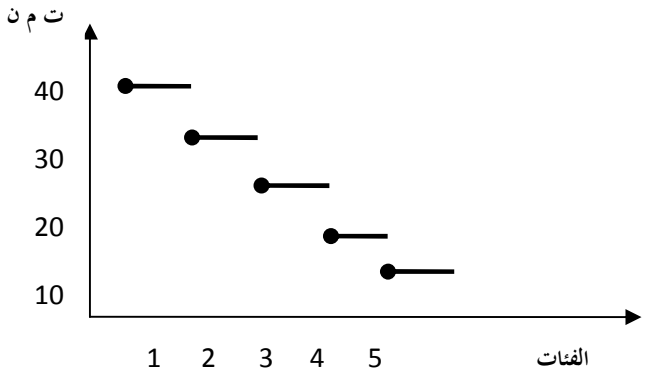
نحسب كل من التكرار المتجمع الصاعد والتكرار المتجمع النازل، للتوزيع المعطى كما يلي:

$n_i \swarrow$	$n_i \nearrow$	n_i	المتغير
37	4	4	1
33	14	10	2
23	26	12	3
11	32	6	4
5	37	5	5
0	/	37	المجموع

عند رسم التكرار المتجمع الصاعد يتم الإستعانة بقطع مستقيمة، تنتهي عند نقاط تقاطع قيمة المتغير مع ت. م. ص. المقابل لها، وتبدأ من حيث تنتهي القطعة السابقة كما يلي:



عند رسم التكرار المتجمع النازل يتم الإستعانة بقطع مستقيمة، تبدأعند نقاط تقاطع قيمة المتغير مع ت. م. ن. المقابل لها، وتنتهي عند بداية القطعة الموالية، كما يلي:



2- حالة بيانات كمية مبوبة بالفئات:

ويتم التمثيل عن طريق المضلع أو عن طريق المنحنى، حيث يرسم منحنى التكرار المتجمع الصاعد عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرار المتجمع الصاعد المقابل لها، ويرسم منحنى التكرار المتجمع النازل بإيصال مجموعة النقاط التي إحداثياتها: الحدود الدنيا للفئات والتكرار المتجمع النازل مقابل لها.

يبين كل من منحنى التكرار المتجمع الصاعد ومنحنى التكرار المتجمع النازل شدة أو ضعف تطور الظاهرة المدروسة عن مستوى معين من مجال الدراسة. إن فاصلة نقطة تقاطع المنحنيين تسمى بالوسيط[15].

- **مثال تطبيقي** : أرسم على نفس المعلم كل من منحنى التكرار المتجمع الصاعد ومنحنى التكرار المتجمع النازل لبيانات التكراري الآتي:

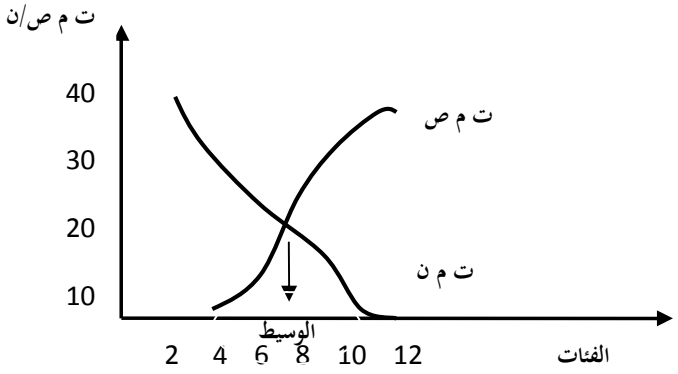
الفئة	التكرار
$[4-2]$	4
$[6-4]$	10
$[8-6]$	12
$[10-8]$	6
$[12-10]$	5
المجموع	37

الحل:

نحسب كل من التكرار المتجمع الصاعد والتكرار المتجمع النازل للتوزيع السابق:

الفئة	n_i	$n_i \nearrow$	$n_i \swarrow$
$[4-2]$	4	4	37
$[6-4]$	10	14	33
$[8-6]$	12	26	23
$[10-8]$	6	32	11
$[12-10]$	5	37	5
المجموع	37	/	0

نرسم النقاط التي إحداثياتها قيمة المتغير و ت. م. ص. (ت. م. ن.) ثم يتم وصل النقاط مع بعضها منحنى متزايد (متناقص) يمثل المنحنى المتجمع الصاعد (المنحنى المتجمع النازل) ونظهر ذلك من خلال رسم واحد نظرا لتشابه سلم الرسم في كليهما، كما يلي:



تظهر نقطة التقاطع في محور الفواصل بقيمة حوالي 7 وهي تمثل لنا الوسيط، وهو ما نتناوله في الفصل الثالث من هذه المطبوعة، تحت عنوان مقاييس النزعة المركزية

تمارين

- **التمرين 01**: تمثل البيانات التالية التقديرات التي تحصل عليها طلبة أحد الأقسام النهائية بإحدى الثانويان خلال الثلاثي الأول لعام 2005 :

ممتاز ، مقبول ، جيد ، ضعيف ، ضعيف ، مقبول ، ضعيف ، مقبول ، ممتاز ، جيد جدا ، ضيف ، ممتاز ، مقبول ، جيد ، جيد ، جيد جدا ، جيد ، ضعيف ، ممتاز ، ممتاز ، ضيف ، مقبول ، مقبول ، ضعيف ، مقبول ، مقبول ، جيد ، مقبول.

1- ما هو المجتمع الإحصائي المدروس ؟.

2- ما هو المتغير الإحصائي المدروس ؟.

3- ضع جدولا للتوزيع التكراري المطلق والنسبي والمئوي لهذه الظاهرة .

- **التمرين 02**: البيانات التالية تمثل الحالة الاجتماعية لعينة متكونة من 400 شخص بإحدى الولايات :

الحالة الاجتماعية / الجنس	لم يسبق له الزواج	متزوج	مطلق	أرمل	المجموع
ذكور	75	70	30	25	200
إناث	60	54	50	36	200
المجموع	135	124	80	61	400

1- ماهو المجتمع الإحصائي المدروس ؟.

2- ما هو المتغير الإحصائي المدروس ؟ وما طبيعته ؟.

3- مثل بيانيا هذا المتغير بالأعمدة المجزأة والأعمدة المتلاصقة.

- **التمرين 03**: في استطلاع حول توزيع اليد العاملة بإحدى البلديات حصلنا على المعطيات التالية :

الفلاحة : 1700 عامل ، الصناعة : 1250 عامل ، التجارة : 1100 عامل ، الإدارة : 750 عامل ، نشاطات أخرى : 200 عامل .

1- ما هو المجتمع الإحصائي المدروس ؟.

2- ما هو المتغير الإحصائي المدروس ؟ وما طبيعته ؟.

3- ضع جدولا مبينا فيه المتغير الإحصائي وتكراره المطلق والنسبي .

4- مثل بيانيا هذا المتغير الإحصائي في تمثيل دائري.

- **التمرين 04**: البيانات التالية تمثل الصادرات العالمية من المعادن في سنة ما (بآلاف الوحدات النقدية)

المناطق	أوربا الغربية	أوربا الشرقية	أمريكا الشمالية	أمريكا الجنوبية	آسيا	أفريقيا	استراليا
الصادرات 10^3	908	730	854	228	228	193	323

- **المطلوب** : عرض البيانات السابقة بيانيا بطريقة مناسبة

- **التمرين 05**: لتكن السلسلة الإحصائية التالية التي تمثل عدد الأطفال لكل أسرة

1، 2 ، 3 ، 4 ، 2 ، 1 ، 2 ، 3 ، 4 ، 5 ، 6 ، 5 ، 6 ، 2 ، 3 ،

4 ، 1 ، 4 ، 2 ، 3 ، 2، 1 ، 3، 4 ، 2 ، 4 ، 3 ، 6 ، 5 .

1- حدد المجتمع الإحصائي ، الوحدة الإحصائية ، المتغير الإحصائي ثم نوعية المتغير .

2- ضع البيانات السابقة في جدول إحصائي يتضمن : القيم المختلفة التي يأخذها المتغير الإحصائي ، والتكرارات : المطلقة ، النسبية ، النسبية المئوية ، المتجمعة الصاعدة والنازلة المقابلة لكل قيم.

3- ما هو التمثيل البياني المناسب لهذه السلسلة ؟.

- **التمرين 06**: البيانات التالية تمثل عدد الطلبة ي إحدى المؤسسات الجامعية:

السنة	الأولى	الثانية	الثالثة	الرابعة
عددالطلبة	9000	7500	7000	6500

استخدم طريقة الدائرة لعرض البيانات الواردة في الجدول أعلاه.

- **التمرين 07** : البيانات التالية تمثل فصائل الدم لعشرين مريضا أجريت لهم عمليات جراحية في خلال أسبوع بإحدى المصحات:

O A AB O O B O B A O A O B A O B A B O AB

المطلوب :

1- عرض البيانات في جدول إحصائي.

2 - عرض البيانات بيانيا بالشكل المناسب

- **التمرين 08** : البيانات التالية تمثل مجموع علامات 20 طالب في 05 مقاييس :

65 ، 90 ، 60 ، 86 ، 57 ، 83 ، 56 ، 80 ، 54 ، 77 ، 50 ، 74 ، 47 ،

89 ، 46 ، 67 ، 42 ، 40 ، 94 ، 93

المطلوب :

1- عرض البيانات السابقة في جدول توزيع تكراري ذي 05 فئات.

2- التمثيل البياني للتوزيع بـ :

- المدرج التكراري والمضلع التكراري
- ارسم المنحنيين المتجمع الصاعد والمتجمع

- **التمرين 09** :لتكن السلسلة الإحصائية التالية التي تمثل مساحة مجموعة من المستثمرات الفلاحية :

المساحة(هكتار)	0 - 1	1 - 2	2 - 3	3 - 5	5 - 10	10 -20	20-40
عدد المستثمرات	27	35	29	54	105	70	40

المطلوب :

1- تحديد التكرارات النسبية والمئوية والمتجمعة الصاعدة والنازلة.

2- عرض البيانات السابقة بيانيا بالطرق المناسبة الممكنة.

3- عدد المستثمرات تقل مساحتها أو تساوي 15 هكتار ، 03 هكتار.

4- عدد المستثمرات الفلاحية التي تساوي مساحتها أو تزيد عن 10 هكتار ، 22 هكتار.

- **التمرين 10**: البيانات التالية تمثل أعمار مجموعة من الأشخاص :

15 ، 19 ، 31 ، 30 ، 21 ، 76 13 ، 35 ، 27 ، 32 ، 77 ، 35 ، 25 ،18 ، 18 ، 15، 45 ، 76 ، 81، 27 ، 86 ، 23 ، 18 ، 18 ، 75 ، 15 ، 69 ، 14 ، 75 ، 63 ، 29، 19 ، 81 ، 15 ، 29، 81 ، 45 ، 17، 15 ، 41 ، 18 ، 31.

- ضع البيانات السابقة في جدول توزيع تكراري مع تحديد التكرارات n_i والتكرارات النسبية f_i والتكرارات المتجمعة الصاعدة والنازلة ، التكرارات المئوية الصاعدة والنازلة.

ملاحظة : تحدد الفئات بالشكل [10- 15 [علما أن طول الفئة هو 5.

- **التمرين 11**: البيانات التالية متعلقة بأسباب الوفاة لعينة متكونة من أطفال وشباب وكهول:

سبب الوفاة	A	B	C	D	E	F	G
أطفال	06	09	02	03	16	04	05
شباب	10	08	07	07	24	20	03
كهول	14	06	13	09	10	05	01

المطلوب:

1- ما نوعية هذا المجتمع ؟ وما هي الخاصية المدروسة في هذا التوزيع ؟ وما طبيعتها ؟

2- ضع مختلف التكرارات لهذا التوزيع.

3- ما هو التمثيل البياني المناسب لهذا التوزيع ؟

- **التمرين 12**: الجدول الموالي يوضح توزيع الأجور الأسبوعية لعمال مؤسسة ما

الفئات (^{2}10)	[0 - 10 [	[10-20 [	[20-30 [	[30- 50[	[50-60]	المجموع
التكرارات n_i	N_1	8	12	14	N_5	50
التكرارات النسبية f_i	0.08	…	…	…	…	1

المطلوب:

1- حدد الخاصية المدروسة ثم اذكر نوعها .

2- أكمل الجدول الإحصائي ثم احسب التكرار المئوي والتكرار المتجمع الصاعد والنازل.

3- احسب نسبة العمال الذين يتقاضون اجر أقل من 25×10^2 د ج ثم عدد العمال الذين تتراوح أجورهم بين 15 و 40×10^2 د ج.

- **التمرين 13** : أكمل الجدول التالي:

الفئات	مراكز الفئات x_i	التكرار النسبي f_i	$f_i\uparrow$	$f_i\downarrow$
[10 -20[	15	.	0.08	1
[20- 30[	25	.	0.29	.
[30- 40[	35	0.26	.	.
[40- 60[	50	0.31	.	.
[60- 80]	70	0.14	.	.

- **التمرين 14** : لتكن مجموعة البيانات المبوبة في جدول توزيع تكراري والمتعلقة بأجور 100 عامل كالآتي:

مراكز الفئات الاجرية	11000	13000	15000	17000	1900	21000	23000	25000	27000	29000	المجموع
عدد العمال	12	14	24	15	13	07	06	04	03	02	**100**

المطلوب:

1- تحديد الفئات الأجرية.

2- تحديد التكرارات المختلفة .

3- تقديم العرض البياني المناسب لهذا التوزيع.

- **مراجع وإحالات الفصل الثاني:**

[1] دلال القاضي وآخرون "الإحصاء -للإداريين والإقتصاديين-" دار الحامد، 2005، الأردن، ص 24.
[2] أحمد شيبات "الإحصاء الوصفي" دار الهدى، 2001، الجزائر، ص 33.
[3] عوض منصور، علي قوقزة، عزام صبري "أساسيات علم الإحصاء الوصفي" دار صفاء، ط 1، 1999، الأردن، ص34.
[4] دومنيك سلفاتور "الإحصاء والإقتصاد القياسي" ترجمة : سعدية حافظ منتصر، الدار الدولية للنشر والتوزيع، مصر، ص 15.
[5] لورانس سي هاملتون "الإحصاء مع برنامج ستاتا" ترجمة: رمضان مفتاح الفيتوري، الفيتوري مفتاح الفيتوري، دار المريخ، 2015، م. ع. السعودية، ص 232.
[6] P. S. MANN «Introductory Statistics » WILEY, 7eme Ed., USA, P 33.
[7] نبيل جمعة صالح النجار " الإحصاء" دار الحامد، ط1، 2010، الأردن، ص 56.
[8] عوض منصور وآخرون، مرجع سابق، ص 45.
[9] عوض منصور وآخرون، مرجع سابق، ص 46.
[10] نبيل جمعة صالح النجار، مرجع سابق، ص 73.
[11] المرجع السابق، ص 77.
[12] المرجع السابق، ص 77.
[13] عبد الرحمن بن محمد سليمان أبو عمة، محمود محمد ابراهيم هندي "الإحصاء التطبيقي" العبيكان، 1995، م. ع. السعودية، ص 42.
[14] عوض منصور وآخرون، مرجع سابق، ص 47.
[15] مبارك أسبر ديب "مبادئ في الاحتمالات والإحصاء" جامعة تشرين، كلية العلوم، 2009/2008، سوريا، ص 53.

الفصل الثالث

مقاييس النزعة المركزية

تمهيد

في هذا الفصل نتطرق إلى تلك الحالة التي حينما نحاول أن نلخص مجموعة من البيانات الرقمية في عدد واحد يرمز لها ويدل عليها ويوضح نزعتها للتجمع فان هذا العدد يكون هو العدد التي تنزع هذه البيانات للتجمع حلوله، فيزداد عدد القيم كلما قربت من هذا العدد ويقل عددها كلما بعدت عنه أي أن للقيم نزعة أو ميلا لقيمة موجودة في مركزها وهذه الخاصية هي التي يطلق عليها النزعة المركزية والقيمة التي تتجمع حولها البيانات يمكن اتخاذها كممثل للمجموعة كلها على أساس أنها غير متطرفة وهي أولى من غيرها بصفة التمثيل لهذه البيانات والمقاييس المستخدمة لقياس هذه القيمة الواقعة في مركز أو وسط البيانات تسمى بمقاييس النزعة المركزية أو نادرا المتوسطات وعليه فإن مقياس النزعة المركزية هو عبارة عن رقم واحد يلخص ويصف خصائص مجموعة من البيانات. ومقاييس النزعة المركزية على أنواع هي (المتوسط الحسابي، الوسيط، المنوال، المتوسط الهندسي، المتوسط التوافقي، المتوسط التربيعي، والمتوسط الهندسي) وان أكثر هذه المقاييس استعمالا هو (المتوسط أو الوسط الحسابي) ويأتي بعد ذلك (الوسيط والمنوال) أما المقاييس الأخرى فقليلا ما تستخدم نظرا لعدم وضوح مدلولاتها لغير المختصين على عكس الثلاثة مقاييس المذكورة ونميز عند حساب مختلف المقاييس بين البيانات التي تكون غير مبوبة وتلك التي تكون مبوبة حيث تظهر لنا التكرارات بشكل جلي في هذه الأخيرة.

المبحث الأول
مفهوم النزعة المركزية ومقاييسها

في هذا المبحث نتطرق إلى مفهوم النزعة المركزية في الإحصاء الوصفي تعتبر مقاييس النزعة المركزية من أهم المقاييس الإحصائية التي يفكر الباحث في حسابها، بل هي أول مقاييس إحصائية يفكر فيها الباحث السياسي عموماً. فمقياس النزعة المركزية لظاهرة إقتصادية ما تعني التعرف على القيمة التي تقع عادة عند مركز التوزيع العددي للقيم المبحوثة.، بالإضافة إلى أنواع المقاييس التي تعنى بحساب مركز البيانات، أما طريقة الحساب فنتركها للمباحث الباقية في هذا الفصل.

1- مفهوم مقاييس النزعة المركزية: Mesures de Tendance Centrale

إن قيم أي ظاهرة (فيزيائية، إقتصادية، إجتماعية ...) تميل أو تنزع للتركز حول قيمة معينة هي متوسط هذه الظاهرة أو مقياس نزعتها المركزية.

فأطوال البالغين تتركز حول رقم معين هو متوسط الطول، وكذلك أوزانهم ومعدلات ذكائهم، وأي ظاهرة أخرى. فالمتوسط – بصفة عامة – هو الذي يعبر عن المستوى العام للظاهرة أي هو الذي يعبر عن جميع قيمها، بمعنى أنه القيمة التي تتركز حولها باقي القيم.

و أهم مقاييس النزعة المركزية التي نتناولها في هذه المطبوعة، نذكرها فيما يلي وبدون شرح ليكون ذلك بمثابة نظرة شاملة عما يتم شرحه بالتفصيل في المباحث اللاحقة من هذه المطبوعة حيث يمكن إختصار تقسيمات مقاييس النزعة المركزية إلى ثلاثة أقسام وهي :المتوسطات بأنواعها، الوسيط وشبيهاته والمنوال:

أ. المتوسطات: المتوسط (العادي والمرجح)، المتوسط الهندسي، المتوسط التوافقي، المتوسط التربيعي.

ب. الوسيط وشبيهات الوسيط (الربيعيات، العشيريات والمئينات)

ت. المنوال (مع مختلف الطرق لحسابه)[1].

وتجدر الإشارة إلى أنه بالرغم من كثرة مقاييس النزعة المركزية وتنوعها إلا أن المتوسط الحسابي يعتبر من أهم المقاييس للنزعة المركزية نظرا لسهولة التعامل الجبري معه وخصائصه الفريدة وسهولة فهمه، بالإضافة إلى أنه يعتبر عنصرا أساسيا لحساب بعض المقاييس الأخرى كمقاييس التشتت، العزوم ومقاييس الشكل التي تكون موضوع الفصلين القادمين.

كما نتطرق فيما بعد إلى طريقة حساب كل مقياس سواء في البيانات غير المبوبة أو تلك المبوبة سواء مبوبة بدون فئات أو مبوبة بالفئات حيث في هذه الأخير يتم حساب ما يعرف بمركز اللفئة وهو القيمة التي تنوب عن كل الفئة في الحساب لمختلف المقاييس، كذلك وما لا مكن إغفاله هو التطرق إلى المزايا لكل مقياس من المقاييس المذكورة بالإضافة إلى عيوبها

2- أهمية تنوع مقاييس النزعة المركزية:

إن مقاييس النزعة المركزية تختلف في فلسفتها حول إيجاد القيمة المركزية للبيانات، فمن المقاييس التي تبحث في القيمة المتساوية التي يمكن توزيعها على كل مفردات العينة مثل المتوسط الحسابي، ومنها ما يبحث في القيمة التي تتوسط البيانات بعد ترتيبها مثل الوسيط ومنها ما يبحث في القيمة التي تتكرر كثيرا مثل المنوال بالإضافة إلى مقاييس أخري غير ظاهرة في معناها بشكل مباشر بل تحتاج إلى دراسة لطبيعة البيانات (مثل المتوسط الهندسي والمتوسط التوافقي ...)، وتكمن أهمية هذا التنوع في مقاييس النزعة المركزية في أنه لا يمكن حساب جميع المقاييس لبعض أنواع البيانات مثلما لا يمكن حساب المتوسط الحسابي في حالة الجداول المفتوحة، وكذلك لا بمكن حساب المتوسط الهندسي للقيم الصفرية (ليس له معنى) والقيم السالبة. حيث يمكن الإستعانة ببعض المتطابقات والعلاقات بين بعض المقاييس لتقدير البعض الآخر وهذا الأمر نتطرق إليه في نهاية هذا الفصل.

المبحث الثاني
المتوسطات وأنواعها

نتطرق في هذا المبحث إلى المتوسطات بكل أنواعها، حيث أن بعض المتوسطات (المتوسط الحسابي) حسابها له طرقا مختلفة تعطي نتائج متساوية ويتم سردها كأنواع للمتوسط، أما بعض المتوسطات هي طرق رياضية مختلفة في طريقة حسابها ومبادئها في نفس الوقت ولا تعطي نتائج متساوية مثل المتوسط التربيعي، الهندسي، التوافقي حيث لها استخدامات محددة وعند بعض الحالات ولتفادي بعض المشاكل كتطرف البيانات.

1- المتوسط الحسابي: Moyenne Arithmétique

هو عبارة عن مجموع القياسات الخاصة بظاهرة معينة على عدد هذه القياسات، وتعطى صيغته حسب طبيعة البيانات(مبوبة أو غير مبوبة) كما يلي:

1-1- المتوسط الحسابي للبيانات غير المبوبة: ويعطى بالعلاقة التالية[2]:

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

- **مثال تطبيقي:** المتوسط الحسابي للبيانات التالية:

5 4 7 4 8 9 6 5

يحسب كما يلي:

$$\bar{X} = \frac{5+4+7+4+8+9+6+5}{8} = \frac{48}{8} = 6$$

1-2- المتوسط الحسابي للبيانات المبوبة: في حالة البيانات المبوبة (جداول توزيع تكراري) سواء بالفئات أو بدون فئات، تعطى صيغة المتوسط الحسابي بالشكل أدناه مع مراعاة أنه في حالة الفئات تصبح x_i تعبر عن مراكز الفئات[3]:

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i\, x_i}{\sum_{i=1}^{k} n_i}$$

في حالة الفئات يحسب x_i مركز الفئة i بالعلاقة التالية:

$$x_i = \frac{A_i + B_i}{2}$$

A_i : الحد الأدنى للفئة رقم i

B_i : الحد الأعلى للفئة i

- **مثال تطبيقي (1):** يحسب المتوسط الحسابي في حالة بيانات مبوبة ولكن بدون فئات كما يلي:

المتغيرة	التكرار المطلق	$n_i\, x_i$
2	3	6
3	7	21
5	5	25
2	5	10
المجموع	20	62

$$\bar{X} = \frac{62}{20} = 3.1$$

- **مثال تطبيقي (2):** يحسب المتوسط الحسابي في حالة بيانات مبوبة بفئات كما يلي:

فئات المتغيرة	التكرار المطلق	مراكز الفئات x_i	$n_i\, x_i$
]20-10]	2	15	30
]30-20]	3	25	75
]40-30]	3	35	105
[50-40]	2	45	90
المجموع	10	/	300

$$\bar{X} = \frac{300}{10} = 30$$

- **مزايا الوسط الحسابي:**

✓ شائع الاستخدام، لأنه مفهوم وسهل الفهم.

✓ يأخذ جميع القيم في الاعتبار.

✓ لا يحتاج إلى لإعادة ترتيب البيانات.

✓ يستخدم للمقارنة بين حالات معينة مثل الأطوال.

✓ يستخدم في حساب كثير من المقاييس والإختبارات الأخرى، (مثل مقاييس التشتت، مقاييس الشكل، ...)

✓ المجموع الجبري لانحرافات القيم (تسمى القيم المعيارية : $\sum_{i=1}^{n} e_i = 0$) عن وسطها الحسابي يساوي صفرا، أي (حسب مبوبة، غير مبوبة) :

$$\sum_{i=1}^{n} (x_i - \bar{X}) = 0 \quad \text{أو} \quad \sum_{i=1}^{k} n_i (x_i - \bar{X}) = 0$$

✓ يكون مجموع مربعات هذه الانحرافات حول المتوسط الحسابي أو القيم المعيارية أصغر ما يمكن من أي انحرافات حول أي قيمة أخرى من البيانات.

$$\sum_{i=1}^{n} (x_i - \bar{X})^2 \sim Min \quad \text{أو} \quad \sum_{i=1}^{k} n_i (x_i - \bar{X})^2 \sim Min$$

أو بطريقة أخرى إذا كان a أحد البيانات[4]:

$$\sum_{i=1}^{n}(x_i-\bar{X})^2 < \sum_{i=1}^{n}(x_i-a)^2 \quad si\ a \neq \bar{X}$$

$$\sum_{i=1}^{n} n_i(x_i-\bar{X})^2 < \sum_{i=1}^{k} n_i(x_i-a)^2 \quad si\ a \neq \bar{X}$$

✓ إذا كان a عددا حقيقيا و $\bar{X}$ المتوسط الحسابي للسلسلة: $x_1, x_2, \dots, x_n$ فإن المتوسط الحسابي للسلسلة: $a \cdot x_1, a \cdot x_2, \dots, a \cdot x_n$ هو: $a \cdot \bar{X}$، نفس الشئ مع القسمة يكفي إعتبار أن $a = \frac{1}{b}$ ولسلسلة: $(x_1 + a), (x_2 + a), \dots, (x_n + a)$ هو: $(\bar{X} + a)$، نفس الشئ بالنسبة للطرح، حيث يمكن إعتبار أن: $a = -b$.

- **مثال تطبيقي**: ولتوضيح الفكرتين الأخيرتين نعرض الجدول التالي الذي يبين متغيرة X تأخذ مجموعة من القيم x_i (بيانات غير مبوبة):

i	x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$(x_i - 2)^2$	$(x_i - 3)^2$	$(x_i - 5)^2$	$(x_i - 9)^2$
1	2	-2,73	7,45	0	1	9	49
2	3	-1,73	2,99	1	0	4	36
3	5	0,27	0,07	9	4	0	16
4	6	1,27	1,61	16	9	1	9
5	2	-2,73	7,45	0	1	9	49
6	3	-1,73	2,99	1	0	4	36
7	3	-1,73	2,99	1	0	4	36
8	5	0,27	0,07	9	4	0	16
9	9	4,27	18,23	49	36	16	0
10	8	3,27	10,69	36	25	9	1
11	6	1,27	1,61	16	9	1	9
	المجاميع						
	52	0	56,18	138	89	57	257
	المتوسط الحسابي: 52/11 = 4.73						

من خلال الجدول نلاحظ أن مجموع الانحرافات حول المتوسط الحسابي تساوي الصفر، بالإضافة على أن مربعات الانحرافات حول المتوسط الحسابي لهذه العينة 4.73 كان 56.18 وهو أقل من مجموع الانحرافات من بعض قيم العينة 2، القيمة الدنيا حيث كان 138، و9 القيمة العليا حيث كان 89 ، وكذلك لبعض القيم مثل 3 كان 57 و5 كان 257.

- **عيوب الوسط الحسابي:**

✓ يتأثر بالقيم المتطرفة (الكبيرة والصغيرة).

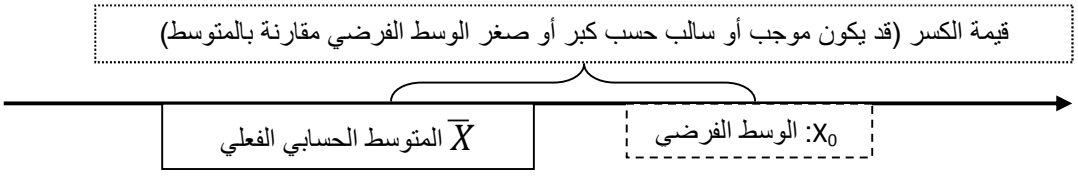
- ✓ يستخدم في المتغيرات الكمية فقط ولا يستخدم في البيانات الوصفية.
- ✓ يستغرق حسابه وقتا كبيرا من البيانات غير المبوبة.
- ✓ لا يمكن حسابه من الجداول المفتوحة في أحد الطرفين.
- ✓ لا يكون قيمة مشاهدة (من البيانات) إلا نادرا[5].

2- الوسط الفرضي:

ما يميز طريقة الوسط الفرضي لحساب المتوسط الحسابي، أنها تعطي نتيجة مطابقة للطريقة الإعتيادية في حساب المتوسط الحسابي.

2-1- للبيانات عير المبوبة: حيث يتم إختيار قيمة معينة من البيانات لا على التعيين ويفضل أن تكون قيمتها تقترب من وسيط البيانات لكي لا تضخم لنا الحسابات عند أخذ الفروقات، وبالتالي تكون النتيجة أكثر دقة، وتعطى الصيغة الرياضية كما يلي[6]:

$$\bar{X} = X_0 + \frac{\sum_{i=1}^{n}(x_i - X_0)}{n}$$



- **مثال تطبيقي:** نأخذ نفس المثال السابق (عند حساب المتوسط الحسابي) لنقارن النتائج:

5 4 7 4 8 9 6 5

نأخذ القيمة الافتراضية $X_0 = 6$ ، ونجري الحسابات:

$$\bar{X} = 6 + \frac{(5-6)+(6-6)+(9-6)+(8-6)+(4-6)+(7-6)+(4-6)+(5-6)}{8}$$

$$= 6 + \frac{-1+0+3+2-2+1-2-1}{8} = 6 + 0 = 6$$

وهي نفس النتيجة بالطريقة العادية.

2-2- للبيانات المبوبة: إذا كنا أمام جداول توزيع تكراري سواء بالفئات أو بدون فئات يتم أخذ كذلك قيمة معينة افتراضية X_0 من وسط الجدول (لتبسيط الحسابات) ، غير أنه إذا كان الجدول بفئات يتم أخذ القيمة الافتراضية من مراكز الفئات ونأخذ مركزا للفئة من وسط الجدول[7] .

$$\bar{X} = X_0 + \frac{\sum_{i=1}^{n} n_i(x_i - X_0)}{\sum_{i=1}^{n} n}$$

- **مثال تطبيقي:** أحسب المتوسط الحسابي بطريقة الوسط الفرضي من جدول التوزيع التكراري التالي (نفس التمرين عندما حسبنا المتوسط الحسابي بالطريقة العادية في حالة البيانات المبوبة بفئات):

فئات المتغيرة	التكرار المطلق
]20-10]	2
]30-20]	3
]40-30]	3
[50-40]	2

- **الحل:** نختار مركز للفئة بشكل افتراضي من وسط الجدول (لتسهيل الحساب): وليكن $X_0=35$

فئات المتغيرة	التكرار المطلق n_i	مراكز الفئات x_i	x_i-X_0	n_i (x_i-X_0)
]20-10]	2	15	10-	20-
]30-20]	3	25	0	0
]40-30]	3	35	10	30
[50-40]	2	45	20	40
المجموع	10	/		50

$$\bar{X} = X_0 + \frac{\sum_{i=1}^{n} n_i(x_i - X_0)}{\sum_{i=1}^{n} n} = 25 + \frac{50}{10} = 30$$

وهي نفس نتيجة الطريقة العادية في حساب المتوسط

3- المتوسط الحسابي بطريقة الترميز (طريقة الوسط الفرضي المختصرة):

إذا كانت البيانات مبوبة وفي حالة وجود توزيع تكراري منتظم-فئات متساوية- نحدد الفئة التي تقع في الوسط ف حالة عدد البيانات الفردي، أو نختار من بين الفئتين الوسطيتين واحدة لا على التعيين في حالة عدد الفئات زوجيا، ثم نضيف عمودا للرمز μ_i حيث يكون مساويا للصفر على مستوى الفئة الوسيطية، وننقص واحدا فواحدا بالنسبة للأسفل، ونضبف واحدا فواحد بالنسبة للأعلى، كما هو موضح في الجدول الموالي:

x_i مراكز الفئات	الرمز μ_i	التكرار المطلق n_i	$n_i \times \mu_i$
⋮	⋮	⋮	⋮
	+2		
	+1		
مركز الفئة الوسطية	0		0
	−1		
	−2		
⋮	⋮	⋮	⋮
المجموع			

وتعطى الصيغة الرياضية للمتوسط الحسابي بطريقة الترميز كما يلي:

$$\bar{X} = X_o + \frac{\sum_{i=1}^{k} n_i \cdot \mu_i}{\sum_{i=1}^{k} n_i} \times C$$

C : طول الفئة الوسيطية.

X_0 : مركز الفئة الوسيطية المقابلة لـ $\mu = 0$

- **مثال تطبيقي**: أحسب المتوسط الحسابي بطريقة الترميز للتوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	2
]30-20]	3
]40-30]	3
[50-40]	2
المجموع	10

- **الحل**: نقوم بالحسابات الضرورية في الجدول

فئات المتغيرة	n_i	مراكز الفئات x_i	μ_i	$n_i\ \mu_i$
]20-10]	2	15	**2−**	4−
]30-20]	3	25	**1−**	3−
]40-30]	3	**35**	**0**	0
[50-40]	2	45	**1+**	2+
المجموع	10	/	/	5−

$$\bar{X} = X_o + \frac{\sum_{i=1}^{k} n_i \cdot \mu_i}{\sum_{i=1}^{k} n_i} \times C = 35 + \frac{-5}{10} 10 = 30$$

نلاحظ أن النتيجة تكون مساوية للطريقة العادية، حيث أننا نذكر بأن طول الفئات يكون متساوي عند إستخدام طريقة الوسط الفرضي المختصر.

4- المتوسط الحسابي المرجح (الموزون):

من حيث الفكرة يماثل الوسط الحسابي الاعتيادي ولكن الوسط الاعتيادي يعتبر مفردات العينة قيد الدراسة لها نفس الأهمية والتأثير في حساب أي مؤشر إحصائي ولكن في بعض الحالات تكون بعض المفردات أكثر أهمية من غيرها مما يستوجب استخدام مؤشر أخر لحساب المعدل مع الأخذ بنظر الاعتبار أهمية كل مفردة من مفردات العينة وهذا المؤشر هو الوسط الحسابي المرجح أو الموزون ومن هنا تكون قيمة هذا الوسط أكثر دقة من الوسط الاعتيادي غير انه اقل استخداما منه.

حيث تعطى للبيانات ترجيحات w_i معينة، مما يجعل هذه الترجيحات أو الأوزان تلعب دور التكرارات تقريبا غير أنه إذا توفرت التكرارات يتم ضرب كل تكرار على حدة والقسمة على حاصل ضرب الأوزان في التكرارات.

4-1- للبيانات غير المبوبة: ويتم حساب المتوسط الحسابي المرجح كما يلي[8]:

$$\bar{X} = \frac{\sum_{i=1}^{s} w_i \cdot x_i}{\sum_{i=1}^{s} w_i}$$

- **مثال تطبيقي:** علامات إحدى الطلبة في ثلاثة مواد ومعاملاتها موضحة في الجدول التالي:

المادة	العلامة	المعامل
الإحصاء	12	3
الرياضيات	11	2
المحاسبة	8	4

الحل: لحساب متوسط نقاطه أو المعدل نستخدم المتوسط المرجح نظرا لوجود أوزان أو لكل مادة، نجري الحسابات في الجدول كمايلي:

المادة	x_i	w_i	$w_i \cdot x_i$
الإحصاء	12	3	36
الرياضيات	11	2	22
المحاسبة	8	4	32
المجموع	/	**9**	**90**

$$\bar{X} = \frac{\sum_{i=1}^{s} w_i \cdot x_i}{\sum_{i=1}^{s} w_i} = \frac{90}{9} = 10$$

4-2- للبيانات المبوبة: في حالة توفر التكرارات (بيانات مبوبة) نحسب مراكز الفئات إذا كان التبويب بالفئات لتنوب عن الفئات كما رأينا سابقا ثم، يحسب المتوسط المرجح كمايلي[9]:

$$\bar{X} = \frac{\sum_{i=1}^{k} w_i \cdot n_i \cdot x_i}{\sum_{i=1}^{k} w_i \cdot n_i}$$

- **مثال تطبيقي:** الجدول الموالي يمثل أجر 10 عمال اشتغلوا في ثلاثة ورشات وكل ورشة لها الأجر الخاص بها للساعة الواحدة لساعات إضافية:

أجر الساعة	عدد العمال	عدد الساعات
1200	2	6
1250	5	10
1300	3	5

تريد المؤسسة معرفة متوسط الأجر الذي يتقاضاه العامل للساعة الواحدة

الحل: نعطي الرموز الحقيقية التي تساعدنا في الحل ونستخدم الجدول في الحسابات :

x_i	n_i	w_i	$w_i\, n_i$	$w_i\, n_i\, x_i$
1200	2	6	12	14400
1250	5	10	50	62500
1300	3	5	15	19500
مجاميع حواصل الضرب			**77**	**96400**

$$\bar{X} = \frac{\sum_{i=1}^{k} w_i \cdot n_i \cdot x_i}{\sum_{i=1}^{k} w_i \cdot n_i} = \frac{96400}{77} = 1251.94$$

5- المتوسط الهندسي:Moyenne Géométrique

المتوسط الهندسي له أهمية في أنه يميل للتخفيف من تأثير القيم العالية جدا أو المنخفضة، والتي قد تؤدي إلى التحيز عند حساب مركز البيانات، لكنه يتعامل مع البيانات **الموجبة غير الصفرية**، وفيما يلي طريقة حسابه لمختلف أنواع الجداول التكرارية:

5-1- المتوسط الهندسي للبيانات غير المبوبة: ويعطى بالعلاقة التالية[10]:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

حيث نلاحظ أننا استخدمنا رمز الجداء:

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times x_3 \cdots \times x_n$$

وهذا الأخير أحيانا يكون حسابه صعب، لذا لابد من استخدام طريقة اللوغاريتمات العشرية أو النيبيرية كما يلي:

$$log(G) = log\left(\sqrt[n]{\prod_{i=1}^{n} x_i}\right)$$

$$= log\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

$$= \frac{1}{n} log\left(\prod_{i=1}^{n} x_i\right)$$

$$log\ (G) = \frac{1}{n}\sum_{i=1}^{n} log(x_i)$$

ثم في الأخير نقوم بإدخال الدالة العكسية للوغاريتم العشري (الدالة الآسية ذات الأساس 10) لإيجاد المتوسط الهندسي كما يلي:

$$G = 10^{\log(G)} = 10^{\left[\frac{1}{n}\sum_{i=1}^{n} log(x_i)\right]}$$

أما إذا استخدمنا في البداية اللوغاريتم النيبيري، نقوم بإدخال الدالة الأسية (عكسية الدالة اللوغاريتمية النيبيرية)، كما يلي:

$$ln(G) = \frac{1}{n}\sum_{i=1}^{n} ln(x_i)$$

$$G = e^{ln\,(G)} = e^{\left[\frac{1}{n}\sum_{i=1}^{n} ln(x_i)\right]}$$

- **مثال تطبيقي**: المتوسط الهندسي للبيانات التالية:

5 4 7 4 8 9 6 5

$$ln(G) = \frac{\sum_{i=1}^{n} ln(x_i)}{n}$$

$$= \frac{ln(5) + ln(4) + ln(7) + ln(4) + ln(8) + ln(9) + ln(6) + ln(5)}{8}$$

$$= \frac{14.00}{8} = 1.75$$

$$G = e^{1.75} = 5.75$$

أو باستخدام اللوغاريتم العشري:

$$log(G) = \frac{\sum_{i=1}^{n} log(x_i)}{n}$$

$$= \frac{log(5) + log(4) + log(7) + log(4) + log(8) + log(9) + log(6) + log(5)}{8}$$

$$= \frac{6.08}{8} = 0.76$$

$$G = 10^{(0.76)} = 5.75$$

5-2- المتوسط الهندسي للبيانات المبوبة: ويحسب وفق العلاقة التالية[11]:

$$G = \sqrt[\left(\sum_{i=1}^{k} n_i\right)]{\prod_{i=1}^{k} x_i^{n_i}}$$

وهذا الأخير أحيانا يكون حسابه صعب، لذا لابد من استخدام طريقة اللوغاريتمات العشرية أو النيبيرية كما يلي:

$$\log(G) = \log\left(\sqrt[\left(\sum_{i=1}^{k} n_i\right)]{\prod_{i=1}^{k} x_i^{n_i}}\right)$$

$$= log\left(\prod_{i=1}^{k} x_i^{n_i}\right)^{\left(\frac{1}{\sum_{i=1}^{k} n_i}\right)}$$

$$= \left(\frac{1}{\sum_{i=1}^{k} n_i}\right) log\left(\prod_{i=1}^{k} x_i^{n_i}\right)$$

$$log\ (G) = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} log(x_i)^{n_i}$$

$$log\ (G) = \left(\frac{1}{\sum_{i=1}^{k} n_i}\right) \sum_{i=1}^{k} n_i\ log(x_i)$$

ثم في الأخير نقوم بإدخال الدالة العكسية للوغاريتم العشري (الدالة الآسية ذات الأساس 10) لإيجاد المتوسط الهندسي كما يلي:

$$10^{log\ (G)} = 10^{\left[\frac{\sum_{i=1}^{k} n_i\ log(x_i)}{\sum_{i=1}^{k} n_i}\right]}$$

أما إذا استخدمنا في البداية اللوغاريتم النيبيري، نقوم بإدخال الدالة الأسية (عكسية الدالة اللوغاريتمية النيبيرية)، كما يلي:

$$ln(G) = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} n_i\ ln(x_i)$$

$$e^{ln\ (G)} = e^{\left[\frac{\sum_{i=1}^{n} n_i\ ln(x_i)}{\sum_{i=1}^{k} n_i}\right]}$$

- **مثال تطبيقي:** نستخدم اللوغاريتمات العشرية مثلا لحساب المتوسط الهندسي للتوزيع التالي:

i	فئات المتغيرة	n_i	مراكز الفئات x_i	$log(x_i)$
1	]20-10]	2	15	1.17
2	]30-20]	3	25	1.39
3	]40-30]	3	35	1.54
4	[50-40]	2	45	1.65
/	المجموع	10	/	5.75

$$\frac{1}{\sum_{i=1}^{4} n_i} \sum_{i=1}^{4} n_i \, log(x_i) = \frac{5.75}{10} = 0.575 \Rightarrow \boldsymbol{G = 10^{0.575} = 3.75}$$

- **مزايا المتوسط الهندسي:**

- يستخدم أساسا لإيجاد المتوسط لمعدلات التغير وللأرقام القياسية وهو الأفضل في هذا المجال لأنه يحقق كل الخصائص الرياضية للأرقام القياسية.
- يساعد على خفض التحيز عند وجود قيم كبيرة متطرفة أو صغيرة.
- إذا تساوت جميع قيم المفردات فإن $G = \bar{X}$ وفي غير ذلك $G < \bar{X}$.

- **عيوب المتوسط الهندسي:**

- يستخدم في حالة القيم الموجبة وغير الصفرية.

6- المتوسط التوافقي: Moyenne Harmonique

المتوسط التوافقي هو مقلوب المتوسط الحسابي لمقلوب القيم، ويحسب كما يلي:

6-1- المتوسط التوافقي للبيانات غير المبوبة: ويعطى بالعلاقة التالية[12]:

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

- **مثال تطبيقي (1):** في هذا المثال نظهر المعنى غير الظاهر للمتوسط التوافقي، فمثلا خلال أربع سنوات متتالية إشترى صاحب منزل، مازوت لتدفئة المنزل بتكلفة 2.5، 2.1، 1.8، 1.6 للتر الواحد على الترتيب. فما هو متوسط تكلفة المازوت خلال مدة الأربع سنوات ؟
- **الحل:** لإيجاد متوسط التكلفة قد يلجأ البعض إلى المتوسط الحسابي للأسعار الوحدوية:

$$\frac{1.6+1.8+2.1+2.5}{4} = 2$$

ولكن الطريقة هذه صحيحة في حالة ما إذا إشترى نفس الكمية خلال كل عام، ولكن إذا ما أنفق نفس المبلغ (وليس الكمية) كل سنة سيختلف الأمر، ونعالج هذين الأمرين في الحالتين التاليتين:

الحالة (1): إذا افترضنا أن صاحب المنزل إشترى نفس الكمية في كل عام ولتكن 1000 لتر:

$$\text{متوسط التكلفة} = \frac{\text{التكلفة الكلية}}{\text{الكمية الكلية المشتراة}} = \frac{1600+1800+2100+2500}{4000} = \frac{8000}{4000} = \mathbf{2}\ \ UM/L$$

وهذه النتيجة (2 و. ن./لتر) لا تختلف عن نتيجية المتوسط الحسابي حى لو كانت الكمية المشتراة x من اللتر كل سنة.

الحالة (2): بإفتراض أن صاحب المنزل أنفق نفس المبلغ كل سنة، وليكن 200 وحدة نقدية،

$$\text{متوسط التكلفة} = \frac{\text{التكلفة الكلية}}{\text{الكمية الكلية المشتراة}}$$

$$= \frac{800}{125 + 111.11 + 95.24 + 80} = \frac{800}{411.35} = \mathbf{1.94}\ \ UM/L$$

هذه النتيجة صحيحة حتى من أجل y و. ن. كل سنة، وهذا يساوي المتوسط التوافقي لتكلفة اللتر الواحد، أي:

$$H = \frac{4}{\frac{1}{1.6} + \frac{1}{1.8} + \frac{1}{2.1} + \frac{1}{2.5}} = \frac{4}{0.63 + 0.56 + 0.47 + 0.4} = \frac{4}{2.06} = 1.94$$

وعملية الحصول على المتوسط في الحالتين صحيحة، وقد حسب كل متوسط تحت شروط من الشائع إستخدامها. ويجب ملاحظة أنه في حالة ما إذا إختلف عدد الليترات المستخدمة من سنة إلى أخرى بدلا من بقائها ثابتة، يستبدل المتوسط الحسابي العادي في الحالة (1) بالمتوسط الحسابي المرجح (مرجح بعدد الليترات هنا، ثم تجمع الليترات الكلية لتشكل المقام)[13]. كذلك فإنه إذا تغيرت القيمة الكلية المنفقة من سنة إلى أخرى، يستبدل المتوسط التوافقي العادي في الحالة (2) بالمتوسط التوافقي المرجح (يشبه المتوسط التوافقي في حالة البيانات المبوبة، حيث يكون الترجيح بالمبلغ المنفق كل سنة، ثم تجمع هذه المبالغ لتمثل البسط)

- **مثال (2)**: إذا انتقل شخص من المدينة A إلى المدينة B بسرعة $30\ km/h$ وعاد من المدينة A إلى المدينة B بسرعة $60\ km/h$. أوجد متوسط السرعة للمرحلة كلها.

الحل: بإفتراض أن المسافة بينA وB هي $60\ km$ (على سبيل المثال فقط)، فإن وقت الذهاب من A إلى B ومن B إلى A على التوالي يحسب كما يلي:

$$\frac{60\ km}{30\ km/h} = 2\ h$$

$$\frac{60\ km}{60\ km/h} = 1\ h$$

$$\text{متوسط الرحلة} = \frac{\text{المسافة الكلية}}{\text{الوقت الكلي}} = \frac{120\ km}{3\ h} = \mathbf{40}\ km/h$$

وهذا المتوسط هو المتوسط التوافقي للرقمين 30 و60. بمعنى :

$$H = \frac{2}{\frac{1}{30} + \frac{1}{60}} = \mathbf{40}\ km/h$$

وإذا كانت المسافات المقطوعة ليست متساوية. فإنه يمكن إستخدام المتوسط التوافقي المرجح للسرعات، حيث الأوزان هي المسافات.

أما استخدام المتوسط الحسابي العادي للسرعتين $45\ km/h = \frac{30km/h+60\ km/h}{2}$ يكون خطأ.

2-6- المتوسط التوافقي للبيانات المبوبة: إذا كنا أمام جداول توزيع تكراري سواء بالفئات أو بدون فئات يعطى المتوسط التوافقي بالعلاقة التالية مع الإشارة إلى أنه في حالة الفئات نحسب مراكز الفئات x_i :[14]

$$H = \sum_{i=1}^{k} \frac{\sum_{i=1}^{k} n_i}{\left(\frac{n_i}{x_i}\right)}$$

- **ملاحظة:** إذا كانت التبويب بالفئات يتم حساب مراكز الفئات x_i
- **مثال تطبيقي:** يتم حساب المتوسط التوافقي للتوزيع التكراري التالي:

n_i / x_i	التكرار المطلق n_i	المتغيرة x_i
1.5	3	2
2.33	7	3
1	5	5
2.5	5	2
7.33	20	المجموع

$$H = \sum_{i=1}^{k} \frac{\sum_{i=1}^{k} n_i}{\left(\frac{n_i}{x_i}\right)} = \frac{20}{7.33} = 2.72$$

- **مزايا المتوسط التوافقي:**

– يستخدم في حالة وجود علاقة عكسية بين ظاهرتين (مثلا عدد الوحدات المشتراة بدينار واحد، كلما زاد سعر الوحدة إنخفض عدد الوحدات المشترات، فحساب متوسط السعر للكميات المشتراة بفضل فيه إستخدام المتوسط التوافقي بدلا من المتوسط الحسابي الذي يعطينا نتائج مغلوطة إلا بعد إجراء تعديلات على البيانات)

- **عيوب المتوسط التوافقي:**

– مدلوله غير ظاهر مقارنة بالمقاييس الأخرى.

7- المتوسط التربيعي: Moyenne Quadratique

المتوسط التربيعي هو الجذر التربيعي الموجب لمتوسط مربعات القيم، ويتم حسابه للبيانات غير المبوبة والمبوبة كما يلي[15]:

7-1- المتوسط التربيعي للبيانات غير المبوبة:

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

- **مثال تطبيقي**: المتوسط التربيعي للبيانات التالية:

5 4 7 4 8 9 6 5

$$Q = \sqrt{\frac{5^2 + 4^2 + 7^2 + 4^2 + 8^2 + 9^2 + 6^2 + 5^2}{8}} = \sqrt{\frac{312}{8}} = \sqrt{39} = 6.24$$

7-2- المتوسط التربيعي للبيانات المبوبة:

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

حيث x_i تعبر عن قيم المتغير إذا كان التبويب بدون فئات، وتعبر عن مراكز الفئات في حالة تبويب بالفئات.

- **مثال تطبيقي**: حساب المتوسط التربيعي للتوزيع السابق:

$n_i \ x_i^2$	x_i^2	مراكز الفئات x_i	n_i	فئات المتغيرة
450	225	15	2	]20-10]
1875	625	25	3	]30-20]
42875	1225	35	3	]40-30]
4050	2025	45	2	[50-40]
49250	/	/	10	المجموع

$$Q = \sqrt{\frac{49250}{10}} = \sqrt{4925} = 70.17$$

المبحث الثالث

الوسيط وشبيهات الوسيط

في هذا المبحث نتكلم عن أحد أهم مقاييس النزعة المركزية والذي يسمى الوسيط، ونتطرق أيضا لمقاييس لها نفس فكرة الوسيط نظرا للتشابه الحسابي بينهم وهي ما تسمى بشبيهات الوسيط والتي تدعى بالربيعيات، العشيريات والمئينات، وكما جرت العادة ولكن بإختلاف بسيط نعرض طريقة الحساب في الحالة التي تكون فيها البيانات غير مبوبة أو مبوبة دون فئات، وحالة البيانات المبوبة بالفئات، أي أننا ضمينا حالة التبويب بدون فئات لحالة البيانات غير المبوبة وهذا مع الوسيط وشبيهاته فقط نظرا لكون هذه المقاييس تعتمد على ترتيب البيانات فقط أما في حالة الفئات فتتطلب طريقة نسبية معينة لحساب الوسيط.

1- الوسيـط: Médian

يعرف الوسيط بأنه القيمة التي تقع منتصف القيم بعد ترتيبها تصاعديا أو تنازليا، أي أن الوسيط يقسم العينة أو المجتمع إلى قسمين متساويين، نصف العدد 50 % يكون اكبر من الوسيط والنصف الأخر 50 % أقل منه.

1-1- الوسيط للبيانات غير المبوبة (أو مبوبة بدون فئات): في حالة البيانات غير المبوبة نقوم بترتيب البيانات إما تصاعديا أو تنازليا ثم نقوم بإختيار القسمة التي تتوسط كل القيم ونميز حالتين[16]:

أ. حالة عدد البيانات فردي: في هذه الحالة يكون الأمر بسيطا لأنه توجد قيمة واحدة تتوسط البيانات حيث بعد ترتيب البيانات من 1 الرتبة إلى الرتبة n نختار القيمة ذات الرتبة $\frac{n+1}{2}$ أي:

$$Me = X_{\left(\frac{n+1}{2}\right)}$$

- **مثال**: وسيط البيانات التالية: 2 6 6 9 10 7 4 2 1 هو: Me=6 لأن:

نرتب البيانات تصاعديا مثلا (أو تنازليا) ونلاحظ أن عدد البيانات هو 9 عدد فردي:

الرتبة	1	2	3	4	**5**	6	7	8	**9**
العينة بعد الترتيب التصاعدي	1	2	2	4	**6**	6	7	9	10

ب. حالة عدد البيانات زوجي: هنا توجد قيميتين تتوسطان البيانات حيث في هذه الحالة بعد ترتيب البيانات من 1 الرتبة إلى الرتبة n نختار القيمتين ذاتي الرتبيتين: $\left(\frac{n}{2}+1\right)$ و $\left(\frac{n}{2}\right)$ ثم نقوم بحساب المتوسط الحسابي لهما، أي[17]:

$$Me = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$$

- **مثال:** وسيط البيانات التالية: 2 6 7 9 9 9 7 4 2 1 هو:

نرتب البيانات تصاعديا مثلا (أو تنازليا) ونلاحظ أن عدد البيانات هو 10 عدد زوجي:

الرتبة	1	2	3	4	5	6	7	8	9	**10**
العينة بعد الترتيب التصاعدي	1	2	2	4	**6**	**7**	7	9	9	9

$$Me = \frac{6+7}{2} = 6.5$$

- **ملاحظة:** في حالة البيانات المبوبة بدون فئات، تكون البيانات حتما مرتبة ويكفي الإعتماد على حساب **ت. م. ص.** لإيجاد رتبة الوسيط، ومن ثمة أخذ القيمة المقابلة في عمود المتغير وتكون هي الوسيط (مع الأخذ بعين الإعتبار عدد البيانات أو حجم العينة زوجيا أم فرديا)

1-2- الوسيط للبيانات المبوبة بالفئات: لحساب الوسيط نختار احد الطرق التالية والتي تؤدي إلى نفس النتائج:

- **طريقة 1**

$$Me = A_{Me} + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - n_{Me-1}\nearrow}{n_{Me}} \times C_{Me}$$

- **طريقة 2**

$$Me = A_{Me} + \frac{n_{Me}\swarrow - \frac{\sum_{i=1}^{k} n_i}{2}}{n_{Me}} \times C_{Me}$$

- **طريقة 3**

$$Me = B_{Me} - \frac{\frac{\sum_{i=1}^{k} n_i}{2} - n_{Me+1}\swarrow}{n_{Me}} \times C_{Me}$$

- **طريقة 4**

$$Me = B_{Me} - \frac{n_{Me}\swarrow - \frac{\sum_{i=1}^{k} n_i}{2}}{n_{Me}} \times C_{Me}$$

حيث:

$\frac{\sum_{i=1}^{k} n_i}{2}$: يستخدم هذا المقدار لتحديد الفئة الوسيطية ، حيث نقارن الناتج مع التكرار المتجمع الصاعد حيث يكون الت. م. ص. المساوي له أو أكبر منه هو المقابل **للفئة الوسيطية** (أو نقارن بالتكرار المتجمع النازل وتكون الفئة الوسيطية المقابلة للت. م. ن. الأكبر أيضا من نصف مجموع التكرارات أو يساويه)

A_{Me}: الحد الأدنى للفئة الوسيطية.

B_{Me}: الحد الأعلى للفئة الوسيطية.

C_{Me}: طول الفئة الوسيطية.

↗ n_{Me-1}: التكرار المتجمع الصاعد للفئة التي قبل الفئة الوسيطية.

↙ n_{Me+1}: التكرار المتجمع النازل للفئة التي بعد الفئة الوسيطية.

↙ n_{Me}: التكرار م. ن. للفئة الوسيطية.

↗ n_{Me}: ت م ص للفئة الوسيطية.

n_{Me} : التكرار المطلق للفئة الوسيطية.

- **مثال تطبيقي**: أوجد الوسيط للتوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	10
]30-20]	15
]40-30]	12
]50-40]	8
[60-50]	3

الحل: نختار الطريقة الأولى لحساب الوسيط:

	i	فئات المتغيرة	التكرار المطلق	ت. م. ص	ت. م. ن.	
	1	]20-10]	10	10	48	
الفئة الوسيطية	**2**	**]30-20]**	**15**	**25**	**38**	$\frac{\sum_{i=1}^{5} n_i}{2} = 24 \leq n_2$ ↗↙
	3	]40-30]	12	37	23	
	4	]50-40]	8	45	11	
	5	[60-50]	3	48	3	
		المجموع	48	/	0	

$$\frac{\sum_{i=1}^{5} n_i}{2} = \frac{48}{2} = 24$$

إذن حددنا الآن الفئة الوسيطية (التي تحوي الوسيط) وهي الفئة رقم 2: **]30-20]**، بتطبيق الطريقة 1 نجد:

$$Me = A_{Me} + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - n_{Me-1} \nearrow}{n_{Me}} \times C$$

$$= 20 + \frac{24 - 10}{15} \times 10$$

$$\boldsymbol{Me = 29.33} \in [20 - 30[$$

لو نطبق الطرق 2، 3، 4 الأخرى نحصل على نفس النتيجة:

$$Me = A_{Me} + \frac{n_{Me}\downarrow - \frac{\sum_{i=1}^{k} n_i}{2}}{n_{Me}} \times C = 20 + \frac{38-24}{15} \times 10 = 29.33$$

$$Me = B_{Me} - \frac{\frac{\sum_{i=1}^{k} n_i}{2} - n_{Me+1}\swarrow}{n_{Me}} \times C = 30 - \frac{24-23}{15} \times 10 = 29.33$$

$$Me = B_{Me} - \frac{n_{Me}\swarrow - \frac{\sum_{i=1}^{k} n_i}{2}}{n_{Me}} \times C = 30 - \frac{25-24}{15} \times 10 = 29.33$$

- **مميزات الوسيط:**

✓ سهل في حسابه سواء كانت البيانات مبوبة أو غير مبوبة.

✓ يمكن حسابه في حالة الجداول التكرارية المفتوحة، وكذلك غير المنتظمة.

✓ يمكن حسابه في حالات وجود قيم متطرفة أو شاذة لأننا نعتمد في الحالتين على الفئة الوسطية والتكرار الصاعد والنازل.

✓ سهل الفهم (قيمة الوسيط محددة بـ 50% من الأعلى و 50% من الأسفل بعد ترتيب القيم تصاعديا أو تنازليا).

✓ الوسيط اقل ثباتا من الوسط الحسابي عند حسابه **لعدد** من العينات.

✓ يمكن تقدير الوسيط في حالة الصفات الوصفية التي لا تقاس بأعداد مباشرة أي يستخدم مع الرتب أيضا.

✓ لا يتأثر بالقيم المتطرفة أو الشاذة.

✓ يقسم المدرج (المضلع والمحنى أيضا) التكراري إلى مساحتين متساويتين وهو نقطة تقاطع (الإحدائية الواقعة على محور الفواصل) منحنيي **ت. م. ص.** و **ت. م. ن.**

- **عيوب الوسيط:**

✓ لا تدخل جميع القيم في حسابه بل يعتمد على جزءا منها.

✓ يتأثر بعدد القيم، وتتغير قيمته بتغير أطوال الفئات بالنسبة **لنفس** التوزيع التكراري إذن يتميز بعدم الثبات.

✓ ليس شائعا كالوسط الحسابي في الاستخدام.

✓ نحتاج في حسابه إلى ترتيب البيانات مما يستغرق وقتا طويلا[18].

2- الربيعيات : Quartiles

إذا قسمنا المدى العام للبيانات إلى 4 فئات متساوية من حيث عدد الوحدات(وليس الطول)، نحصل على 3 فواصل، تسمى قيمة المتغيرة عند هذه الفواصل بالربيعيات. أي أنه توجد 3 ربيعيات. ونفرق في حسابه كما فعلنا في الوسيط بين البيانات غير المبوبة أو تلك المبوبة بدون فئات من جهة، والبيانات المبوبة بالفئات من جهة أخرى، حيث أن الربيع الأول هو تلك القيمة التي تسبقها 25 % من البيانات، وتليها 75% منها، أما الربيع الثاني هو القيمة المساوية للوسيط (50 % على يمين الوسيط و50% على يساره من البيانات)، أما الربيع الثالث، فهو القيمة التي تسبقها 75 % من البيانات وتليها 25% من البيانات، ويتم ترتيب البيانات تصاعديا بطبيعة الحال. ويتم حساب مختلف الربيعات كما يلي:

2-1- البيانات غير المبوبة (أو مبوبة بدون فئات):

لإيجاد الربيعي رقم q من أحد الربيعيات الثلاثة: الأول، الثاني، الثالث، نقوم بترتيب البيانات تصاعديا، نأخذ القيمة ذات الرتبة: q مضروبة في $n+1$ من الرتب ومقسومة على 4، لتكون هي الربيع رقم q، أي:

وهذا في حالة الرتبة $r = \left[\frac{q}{4}(n+1)\right]$ قيمة صحيحة، أما إذا كانت عشرية فيتم تحديد قيمة الربيع كالآتي:

نحدد القيمتين التي يق بينهما الربيع:

x_a. القيمة الدنيا

x_b: القيمة العليا

ثم نطبق العلاقة التالية:

$$Q_q = x_a + r_d \times (x_b - x_a)$$

حيث r_d هي الجزء العشري للرتبة r

- **مثال تطبيقي**: لإيجاد الربيعات الثلاثة للبيانات التالية:

8 7 6 7 3 5 3 9 5 7 1 3 2 5 1

نقوم أولا بترتيب البيانات تصاعديا:

الربيع				Q_1				Q_2=Me				Q_3			
الرتبة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
البيانات	1	1	2	3	3	3	5	5	5	6	7	7	7	8	9

الربيع الأول:

$$Q_1 = X_{\left[\frac{1}{4}(15+1)\right]} = X_4 = 3$$

الربيع الثاني:

$$Q_2 = X_{\left[\frac{2}{4}(15+1)\right]} = X_8 = 5$$

الربيع الثالث:

$$Q_3 = X_{\left[\frac{3}{4}(15+1)\right]} = X_{12} = 7$$

2-2- البيانات المبوبة بالفئات: نرمز للربيعي رقم q بالرمز Q_q ويحسب بالعلاقة التالية:

$$Q_q = A_{Q_q} + \frac{\frac{q \cdot \sum_{i=1}^{k} n_i}{4} - n_{Q_q-1}\nearrow}{n_{Q_q}} \cdot C_{Q_q} \;\; ; \;\; q: \overline{1-3}$$

حيث:

$\frac{q\sum_{i=1}^{k} n_i}{4}$: يستخدم هذا المقدار لتحديد الفئة التي تحتوي على الربيع رقم q، حيث نقارن الناتج مع التكرار المتجمع الصاعد حيث يكون الت. م. ص. المساوي له أو أكبر منه هو المقابل للفئة التي تحوي الربيع q

A_{Q_q}: الحد الأدنى للفئة التي تحوي الربيع q

C_{Q_q}: طول الفئة التي تحوي الربيع q

$n_{Q_q-1}\nearrow$: التكرار المتجمع الصاعد للفئة التي قبل الفئة التي تحوي الربيع q

n_{Q_q} : التكرار المطلق للفئة التي تحوي الربيع q

- **مثال تطبيقي:** إيجاد الربيعي الأول والثاني للتوزيع التالي:

فئات المتغيرة	التكرار المطلق	ت. م. ص
]10-20]	10	10
]20-30]	15	25
]30-40]	12	37
]40-50]	8	45
[50-60]	3	48
المجموع	48	/

الربيعي الأول Q_1 ، تحديد فئة الربيعي الأول:

$$\frac{1 \cdot \sum_{i=1}^{5} n_i}{4} = \frac{48}{4} = 12 \;\; \leq 25 \Rightarrow [20-30[$$

$$Q_1 = A_{Q_1} + \frac{\frac{1 \cdot \sum_{i=1}^{5} n_i}{4} - n_{Q_1-1}\nearrow}{n_{Q_1}} \cdot C_{Q_1} = 20 + \frac{12-10}{15}\,10 = \mathbf{21.33}$$

الربيعي الثاني Q_2 ، تحديد فئة الربيعي الثاني:

$$\frac{2 \cdot \sum_{i=1}^{5} n_i}{4} = \frac{96}{4} = 24 \;\; \leq 25 \Rightarrow [20-30[$$

$$\boldsymbol{Q_2} = A_{Q_2} + \frac{\frac{2 \cdot \sum_{i=1}^{5} n_i}{4} - n_{Q_1 - 1\,\nearrow}}{n_{Q_2}} \cdot C_{Q_2} = 20 + \frac{24 - 10}{15}\, 10 = \boldsymbol{29.33} = \boldsymbol{Me}$$

3- العشيريات : Déciles

إذا قسمنا المدى العام للبيانات إلى 10 فئات متساوية من حيث عدد الوحدات (وليس الطول)، نحصل على 9 فواصل، تسمى قيمة المتغيرة عند هذه الفواصل بالعشيريات. أي أنه توجد تسعة عشيريات، فالعشير الأول هو القيمة التي تسبقها 10% من البيانات وتليها 90 % من البيانات، أما العشير الثاني فهو القيمة التي تسبقها 20% من البيانات وتليها 80% من البيانات، وبصفة عامة فالعشير رقم d، هو القيمة التي تسبقها d×10% من البيانات وتليها (10-d)×10% من البيانات. ونفرق في حسابه كما فعلنا في الوسيط بين البيانات غير المبوبة أو تلك المبوبة بدون فئات من جهة، والبيانات المبوبة بالفئات من جهة أخرى، كما يلي:

3-1- البيانات غير المبوبة (أو مبوبة بدون فئات)

لإيجاد العشيري رقم d من أحد العشيريات التسعة: الأول، الثاني... التاسع، نقوم بترتيب البيانات تصاعديا، نأخذ القيمة ذات الرتبة: d مضروبة في $n+1$ من الرتب ومقسومة على 10، لتكون هي العشير رقم d، أي:

$$D_d = X_{\left[\frac{d}{10}(n+1)\right]} \;\; ; \; d = 1; 2; 3; \ldots; 9$$

وهذا في حالة الرتبة $r = \left[\frac{d}{10}(n+1)\right]$ قيمة صحيحة، أما إذا كانت عشرية فيتم تحديد قيمة العشير كالآتي:

نحدد القيمتين التي يق بينهما العشير المعني بالحساب:

x_a: القيمة الدنيا

x_b: القيمة العليا

ثم نطبق العلاقة التالية:

$$D_d = x_a + r_d \times (x_b - x_a)$$

حيث r_d هي الجزء العشري للرتبة r

- **مثال تطبيقي**: لإيجاد بعض العشيريات من البيانات التالية:

9 9 9 8 8 8 7 6 7 3 5 3 9 5 7 1 3 2 5 1

نقوم أولا بترتيب البيانات تصاعديا:

العشير		D_1								D_5							D_7					
الرتبة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
البيانات	1	1	2	3	3	3	5	5	5	6	7	7	7	8	8	8	9	9	9	9

العشير الأول:

$$D_1 = X_{\left[\frac{1}{10}(20+1)\right]} = X_{2.1}$$
$$\Rightarrow D_1 = 1 + 0.1 \times (2-1) = \mathbf{1.1}$$

العشير الخامس:

$$D_5 = X_{\left[\frac{5}{10}(20+1)\right]} = X_{10.5}$$
$$\Rightarrow D_5 = 6 + 0.5 \times (7-6) = \mathbf{6.5} = \boldsymbol{Me}$$

العشير السابع:

$$D_7 = X_{\left[\frac{7}{10}(20+1)\right]} = X_{14.7}$$
$$\Rightarrow D_7 = 8 + 0.7 \times (8-8) = \mathbf{8}$$

3-2- البيانات المبوبة بالفئات: نرمز للعشيري رقم d بالرمز D_d ويحسب بالعلاقة التالية:

$$D_d = A_{D_d} + \frac{\frac{d \cdot \sum_{i=1}^{k} n_i}{10} - n_{D_d-1}\nearrow}{n_{D_d}} \cdot C_{D_d} \quad ; \quad d: \overline{1-9}$$

حيث:

$\frac{\boldsymbol{d\sum_{i=1}^{k} n_i}}{\boldsymbol{10}}$: يستخدم هذا المقدار لتحديد الفئة التي تحتوي على العشير رقم d، حيث نقارن الناتج مع التكرار المتجمع الصاعد حيث يكون الت. م. ص. المساوي له أو أكبر منه هو المقابل للفئة التي تحوي العشير رقم d

A_{D_d}: الحد الأدنى للفئة التي تحوي العشير رقم d

C_{D_d}: طول الفئة التي تحوي العشير رقم d

$n_{D_d-1}\nearrow$: التكرار المتجمع الصاعد للفئة التي قبل الفئة التي تحوي العشير رقم d

n_{D_d} : التكرار المطلق للفئة التي تحوي العشير رقم d

- **مثال تطبيقي:** إيجاد العشير الأول، الخامس، للتوزيع التكراري السابق (عند حساب الوسيط):

فئات المتغيرة	التكرار المطلق	ت. م. ص
]10-20]	10	10
]20-30]	15	25
]30-40]	12	37
]40-50]	8	45
[50-60]	3	48
المجموع	48	/

العشيري الأول D_1 ، تحديد فئة العشيري الأول:

$$\frac{1\cdot\sum_{i=1}^{5}n_i}{10}=\frac{48}{10}=4.8\ \ \leq 10\ \Rightarrow [10-20[$$

$$D_1=A_{D_1}+\frac{\frac{1\cdot\sum_{i=1}^{5}n_i}{10}-n_{D_1-1}\nearrow}{n_{D_1}}\cdot\ C_{D_1}=10+\frac{4.8-0}{10}\ 10=\mathbf{10.48}$$

العشيري الخامس D_5 ، تحديد فئة العشيري الخامس:

$$\frac{5\cdot\sum_{i=1}^{5}n_i}{10}=\frac{5\times 48}{10}=24\ \ \leq 25\ \Rightarrow [20-30[$$

$$D_5=A_{D_5}+\frac{\frac{5\cdot\sum_{i=1}^{5}n_i}{10}-n_{D_5-1}\nearrow}{n_{D_5}}\cdot\ C_{D_5}=20+\frac{24-10}{15}\ 10=\mathbf{22.66}$$

4- المئينات: Percentiles

إذا قسمنا المدى العام للبيانات إلى 100 فئة متساوية من حيث عدد الوحدات (وليس الطول)، نحصل على 99 فاصل، تسمى قيمة المتغيرة عند هذه الفواصل بالمئينات. أي أنه يوجد 99 مئين، فالمئين الأول هو القيمة التي يسبقها 1% من البيانات وتليها 99% من البيانات، والمئين الثاني هو القيمة التي تسبقها 2% من البيانات وتليها 98% من البيانات، وهكذا حتى المئين 99، وبصفة عامة فالمئين رقم p هو القيمة التي تسبقها p% من البينات، وتليها (100-p)% من البيانات. ونفرق في حسابه كما فعلنا في الوسيط بين البيانات غير المبوبة أو تلك المبوبة بدون فئات من جهة، والبيانات المبوبة بالفئات من جهة أخرى، كما يلي:

4-1- البيانات غير المبوبة (أو مبوبة بدون فئات)

لإيجاد المئين رقم p من أحد المئينات التسعة والتسعون: الأول، الثاني... التاسع والتسعون، نقوم بترتيب البيانات تصاعديا، نأخذ القيمة ذات الرتبة: p مضروبة في $n+1$ من الرتب ومقسومة على 100، لتكون هي المئين رقم p، أي:

$$P_p=X_{\left[\frac{p}{100}(n+1)\right]}\ \ ;\ p=1;2;3;\dots;99$$

وهذا في حالة الرتبة $r=\left[\frac{d}{100}(n+1)\right]$ قيمة صحيحة، أما إذا كانت عشرية فيتم تحديد قيمة المئين المعني كالآتي:

نحدد القيمتين التي يق بينهما المئين المعني بالحساب:

x_a: القيمة الدنيا

x_b: القيمة العليا

ثم نطبق العلاقة التالية:

$$P_p = x_a + r_d \times (x_b - x_a)$$

حيث r_d هي الجزء العشري للرتبة r

- **مثال تطبيقي**: لإيجاد بعض المئينات من البيانات التالية :

8	8	8	7	6	7	3	5	3	9	5	7	1	3	2	5	1	
	11	16	11	16	12	15	12	10	13	19	12	18	11	15	18	10	

نقوم أولا بترتيب البيانات تصاعديا:

المئين			P_{10}															
الرتبة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
البيانات	1	1	2	3	3	3	5	5	5	6	7	7	7	8	8	8	9	

المئين					P_{66}												P_{99}
الرتبة	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
البيانات	10	10	11	11	11	12	12	12	13	14	15	15	16	16	18	18	

المئين العاشر:

$$P_{10} = X_{\left[\frac{10}{100}(33+1)\right]} = X_{3.4}$$
$$\Rightarrow P_{10} = 2 + 0.4 \times (3 - 2) = 2.4$$

المئين 66:

$$P_{66} = X_{\left[\frac{66}{100}(33+1)\right]} = X_{22.44}$$
$$\Rightarrow P_{66} = 11 + 0.44 \times (12 - 11) = 11.44$$

المئين 99:

$$P_{99} = X_{\left[\frac{99}{100}(33+1)\right]} = X_{33.66}$$
$$\Rightarrow P_{99} = 18 + 0.66 \times (18 - 18) = \mathbf{18}$$

نلاحظ أن الرتبة 34 غير متوفرة فاعتبرنا القيمة التي اكبر من 18 هي 18 نفسها.

4-2- البيانات المبوبة بالفئات: نرمز للمئين رقم p بالرمز P_p ويحسب بالعلاقة التالية:

$$P_p = A_{P_p} + \frac{\frac{p \cdot \sum_{i=1}^{k} n_i}{100} - n_{P_p - 1}\nearrow}{n_{P_p}} \cdot C_{P_p} \quad ; \quad p: \overline{1 - 99}$$

حيث:

$\frac{\boldsymbol{p \sum_{i=1}^{k} n_i}}{\boldsymbol{100}}$: يستخدم هذا المقدار لتحديد الفئة التي تحتوي على المئين رقم p، حيث نقارن الناتج مع التكرار المتجمع الصاعد حيث يكون الت. م. ص. المساوي له أو أكبر منه هو المقابل للفئة التي تحتوي على المئين رقم p

A_{P_p}: الحد الأدنى للفئة التي تحتوي على المئين رقم p

C_{P_p}: طول الفئة التي تحتوي على المئين رقم p

n_{P_p-1} ↗: التكرار المتجمع الصاعد للفئة التي قبل الفئة التي تحتوي على المئين رقم p

n_{P_p} : التكرار المطلق للفئة التي تحتوي على المئين رقم p

- **مثال تطبيقي**: أوجد المئين 23، والمئين 66، للتوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق	ت. م. ص
]20-10]	10	10
]30-20]	15	25
]40-30]	12	37
]50-40]	8	45
[60-50]	3	48
المجموع	48	/

المئين 23: P_{23} ، تحديد فئة المئين 23:

$$\frac{23 \cdot \sum_{i=1}^{5} n_i}{100} = \frac{23 \times 48}{100} = 11.04 \quad \leq 25 \Rightarrow [20-30[$$

$$P_{23} = A_{P_{23}} + \frac{\frac{23 \cdot \sum_{i=1}^{5} n_i}{100} - n_{P_{23}-1}\nearrow}{n_{P_{23}}} \cdot C_{P_{23}} = 20 + \frac{11.04-10}{10}\ 10 = \mathbf{10.48}$$

المئين 66: P_{66} ، تحديد فئة المئين 66:

$$\frac{66 \cdot \sum_{i=1}^{5} n_i}{100} = \frac{66 \times 48}{100} = 31.68 \quad \leq 37 \Rightarrow [30-40[$$

$$P_{66} = A_{P_{66}} + \frac{\frac{66 \cdot \sum_{i=1}^{5} n_i}{100} - n_{P_{66}-1}\nearrow}{n_{P_{66}}} \cdot C_{P_{66}} = 30 + \frac{31.68-25}{12}\ 10 = \mathbf{35.56}$$

5- متطابقات الوسيط وشبهات الوسيط:

مما سبق وعند التطرق لشبيهات الوسيط والوسيط لا حظنا أن هناك بعض المتطابقات بينهم نظرا لأنهم يعبرون عن فكرة واحدة، والإختلاف يكمن في عدد الفواصل التي يمكن لكل مقياس حسابها، ونذكر منها ما يلي:

$Me = Q_2 = D_5 = P_{50}$

$Q_1 = P_{25}$

$Q_3 = P_{75}$

$D_i = P_{i \cdot 10} \quad ; i = 1; 2; 3; \cdots; 10$

ويمكن وضع هذه المتطابقات في الجدول الموالي:

جدول 3-1 : مقارنة بين الوسيط وشبيهات الوسيط

شبيهات الوسيط التي لها نفس القيمة موضوعة عموديا											
Xmax	*50 %*				***Me***	*50 %*					*Xmin*
	25 %		Q_3	*25 %*		***Q_2***	*25 %*		Q_1	*25 %*	
	D_9	D_8		D_7	D_6	***D_5***	D_4	D_3		D_2	D_1
	P_{90}	P_{80}	P_{75}	P_{70}	P_{60}	***P_{50}***	P_{40}	P_{30}	P_{25}	P_{20}	P_{10}

المصدر: إعداد الباحث إعتمادا على مراجع هذا المبحث.

إن عدم أخذنا لشبيهات الوسيط الصفرية (مثل الربيع الصفري، العشير الصفري، المئين الصفري) هو من منطلق أن هذه القيم من المفترض أن تكون هي القيمة الأولى للبيانات بعد الترتيب التصاعدي، فمن جهة ليس لها أهمية في الحساب بهذه الطرق ومن جهة أخرى أن هناك توزيعات مفتوحة (البيانات المبوبة بالفئات) لا توجد لها حد أدني للفئة الأولى وبالتالي لا يمكن حسابها أصلا، وهنا تكمن أهمية شبيهات الوسيط وهي عدم إمكانية حساب بعض المقاييس الأخرى (مثل المتوسط الحسابي) في التوزيعات المفتوحة، وبالتالي نستعين بشبيهات الوسيط لإيجاد بعض الخواص مثل مقاييس التشتت المعتمدة على شبيهات الوسيط. ونفس الأمر بالنسبة لشبيهات الوسيط العلوية (الربيع الرابع، العشير العاشر، المئين المئة) حيث لا يتم حسابهم لنفس الأسباب المذكورة سلفا.

المبحث الرابع

المنـــــوال: Mode

يعتبر المنوال القيمة الأكثر تكرارا أو الأكثر مشاهدة في مجموعة البيانات ونحتاج بالنسبة لحساب للمنوال إلى التفرقة بين البيانات غير المبوبة والمبوبة بدون فئات من جهة والبيانات المبوبة بالفئات من جهة أخرى، لأن في هذه الأخيرة نستخدم طرقا نسبية لكل من "بيرسون" و"كينغ":

1- المنوال للبيانات غير المبوبة أو مبوبة بدون فئات: في حالة البيانات غير مبوبة أو المبوبة بدون فئات نأخذ القيمة التي لها أكبر تكرار، وتكون هي المنوال، وإن وجدت أكثر من قيمة لها نفس التكرار نقول أن التوزيع التكراري متعدد المنوال[19].

- **مثال 1**: المنوال للبيانات التالية:

1 5 9 6 3 3 3 8 8 6 3 3 2

هو: Mo=3

- **مثال 2** : المنوال للتوزيع التكراري التالي الذي يمثل عدد الأطفال في 11 أسرة، هو Mo=2

عدد الأطفال (المتغيرة)	التكرار المطلق (عدد الاسر)
0	2
1	3
2	4
3	2

- **مثال 3** : البيانات التالية متعددة المنوال 3 و2: حيث نجد كلا من 3 و 2 تتكراران 4 مرات وهو اكبر تكرار.

3 1 0 9 5 2 3 4 6 6 3 2 3 3 2 2

2- المنوال للبيانات المبوبة بالفئات المتساوية (الجداول المنتظمة): يحسب المناول في حالة البيانات المبوبة من الجداول المنتظمة التي لها أطوال فئات متساوية، لكن في حالة الجداول غير المنتظمة يتم تعديل التكرارات بشكل يتناسب من طول كل فئة كما نشرحه بعد حين، ويتم حساب المنوال بطريقتين:

2-1- طريقة الفروقات لـ"بيرسون Pearson" : وتعطى بالعلاقة التالية[20]:

$$\text{Mo} = A_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C_{Mo}$$

$$\begin{cases} \Delta_1 = n_{Mo} - n_{Mo-1} \\ \Delta_2 = n_{Mo} - n_{Mo+1} \end{cases}$$

حيث:

A_{Mo}: الحد الأدنى للفئة المنوالية (الفئة التي لها أكبر تكرار مطلق في حالة الطول الفئات المتساوي، والتي لها أكبر تكرار معدل في حالة أطوال الفئات غير المتساوية)

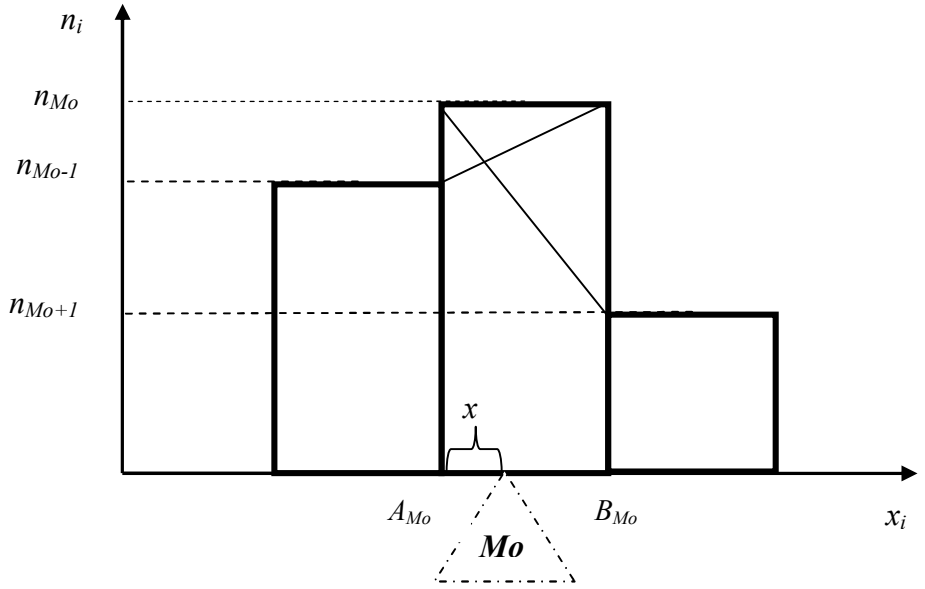
Δ_1: الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها. (نأخذ التكرار المعدل عند التعديل)

Δ_2: الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي بعدها. (نأخذ التكرار المعدل عند التعديل)

C_{Mo}: طول الفئة المنوالية (هنا حالة فئات متساوية).

- **الإستنتاج الرياضي لصيغة المنوال بطريقة الفروقات:**

شكل 3-1 : التمثيل البياني للمنوال بطريقة المدرج التكراري.



المصدر: مبارك أسبر ديب "مبادئ في الاحتمالات والإحصاء" جامعة تشرين، كلية العلوم، 2009/2008، سوريا، ص 57.

تعتمد فكرة حساب المنوال بطريقة الفروقات على التناسب بين الفرق بين تكرار الفئة المنوالية والفئة السابقة لها من جهة والمسافة بين بداية الفئة المنوالية والمنوال، والتناسب بين تكرار الفئة المنوالية وتكرا الفئة اللاحقة لها من جهة والمسافة بين المنوال ونهاية الفئة المنوالية. ثم محاولة مساواة التناسبين. ونعبر رياضيا عن هذا كما يلي:

باستخدام الرموز السابقة:

$$\begin{cases}\Delta_1 = n_{Mo} - n_{Mo-1} \\ \Delta_2 = n_{Mo} - n_{Mo+1}\end{cases}$$

من خلال الرسم نلاحظ أن:

$$Mo = A_{Mo} + x \quad \cdots\cdots (1)$$

لدينا التناسب التالي:

$$\frac{\Delta_1}{x} = \frac{\Delta_2}{C_{Mo} - x} \Leftrightarrow \Delta_1 (C_{Mo} - x) = x\,\Delta_2$$
$$\Leftrightarrow \Delta_1 C_{Mo} - \Delta_1 x = x\,\Delta_2$$
$$\Leftrightarrow \Delta_1 C_{Mo} = (\Delta_1 + \Delta_2)\,x$$
$$\Leftrightarrow x = \frac{\Delta_1}{\Delta_1 + \Delta_2} C_{Mo}$$

نستنتج الآن صيغة المنوال بطريقة الفروقات كما يلي[21]:

$$(1) \Leftrightarrow Mo = A_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} C_{Mo}$$

إذن المنوال من الرسم نلاحظ أنه يميل جهة الفئة المجاووة التي يكون تكرارها كبيرا.

2-2- طريقة الرافعة لـ "كينغ King":

$$\mathrm{Mo} = A_{Mo} + \frac{n_{Mo+1}}{n_{Mo+1} + n_{Mo-1}} \times C_{Mo}$$

حيث:

n_{Mo+1} : تكرار الفئة التي بعد الفئة المنوالية.

n_{Mo-1} : تكرار الفئة التي قبل الفئة المنوالية.

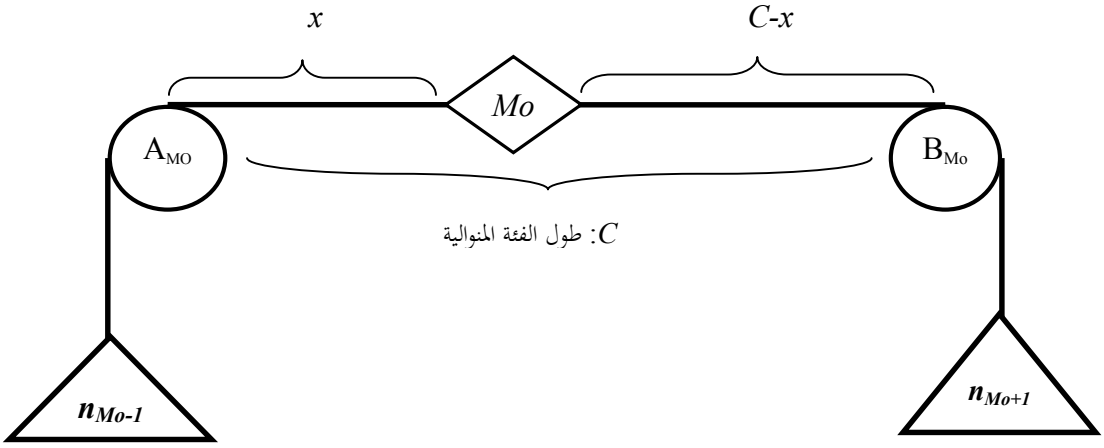
باقي الرموز لها نفس المعنى.

- **ملاحظة:** مما يعاب على طريقة الرافعة أنها لا تأخذ تكرار الفئة المنوالية في الحسبان، إلا عند تحديد الفئة المنوالية.

- **الإستنتاج الرياضي لصيغة المنوال بطريقة الرافعة:**

تقوم فكرة المنوال بطريقة الرافعة (رافعة بالمفهوم الفيزيائي، أي بكرتين بحبل في طرفيه كتلتين) على إفتراض أن تكراري الفئتين السابقة واللاحقة عبارة عن كتلتين وزنهما يقدر بقيمة كل تكرار منهما، تقومان بشد نقطة (المنوال) من طرفين يبعدان عليها بقيمتين معينتين مجموعهما هو طول الفئة المنوالية، نستعين بالرسم لتوضيح ذلك:

شكل 3-2 : التمثيل البياني للمنوال بطريقة الرافعة.



المصدر: علي لزعر "الإحصاء وتوفيق المنحنيات" د. م. ج، 2000، الجزائر، ص 49.

نلاحظ أن:

$$Mo = A_{Mo} + x \quad \cdots\cdots (1)$$

نعلم أن العزوم بالمفهوم الفيزيائي تساوي القوة في الذراع، يمكن إسقاط ذلك كما يلي:

$$\begin{aligned} n_{Mo-1}x = n_{Mo+1}(C_{Mo} - x) &\Leftrightarrow n_{Mo-1}x = n_{Mo+1}C_{Mo} - n_{Mo+1}x \\ &\Leftrightarrow n_{Mo-1}x + n_{Mo+1}x = n_{Mo+1}C_{Mo} \\ &\Leftrightarrow x(n_{Mo-1} + n_{Mo+1}) = n_{Mo+1}C_{Mo} \\ &\Leftrightarrow x = \frac{n_{Mo+1}}{n_{Mo-1} + n_{Mo+1}} C_{Mo} \end{aligned}$$

نستنتج الآن صيغة المنوال بطريقة الرافعة كما يلي[22]:

$$(1) \Leftrightarrow Mo = A_{Mo} + \frac{n_{Mo+1}}{n_{Mo-1} + n_{Mo+1}} C_{Mo}$$

إذن فالمنوال يميل إلى الجهة التي يكون في تكرار الفئة المجاورة كبير (الأكثر وزنا)

- **مثال تطبيقي:** أوجد المنوال بطريقتين من جدول التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	10
]30-20]	15
]40-30]	12
]50-40]	8
[60-50]	3

الحل: نرسم أولا الجدول الذي نحسب فيه كل المقادير ونلاحظ أنه كل أطوال الفئات متساوية إذن لا داعي لتعديل التكرارت:

فئات المتغيرة	التكرار المطلق n_i	طول الفئات C
]20-10]	10	10
]30-20]	15	10
]40-30]	12	10
]50-40]	8	10
[60-50]	3	10
المجموع	48	/

ثم نحدد الفئة المنوالية (التي تحتوي على المنوال) وهي الفئة التي يقابلها أعلى تكرار، ونجد أنها الفئة الثانية التي تقابل 15 قيمة التكرار المطلق.

أ. طريقة الفروقات:

$$Mo = A_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C_{Mo} = 20 + \frac{(15-10)}{(15-10)+(15-12)}\ 10 = 27.14$$

ب. طريقة الرافعة:

$$Mo = A_{Mo} + \frac{n_{Mo+1}}{n_{Mo+1} + n_{Mo-1}} \times C_{Mo} = 20 + \frac{12}{12+10}\ 10 = 25.45$$

3- المنوال في حالة الجداول غير المنتظمة (الفئات غير المتساوية):

إذا كانت الفئات غير متساوية يتم تعديل التكرارات عند حالتين وهما حساب المنوال أو التمثيل البياني (المدرج التكراري، المضلع التكراري، المنحنى التكراري)، ويتم الحساب باستخدام التكرار المعدل عوض التكرار المطلق في حالة التمثيل البياني أو حساب المنوال فقط، وباقي الحسابات تكون بالتكرار المطلق، وتعطى صيغة التكرار المعدل بالعلاقة التالية:

$$n_i^* = \frac{n_i}{C_i} \times h$$

حيث:

n_i^* : التكرار المعدل للفئة i

n_i : التكرار الأصلي للفئة i

C_i: طول الفئة رقم i

h : طول الفئات الشائع (الطول الأكثر تكرار للفئات أو طول الفئات المنتظمة)، في حالة وجود عدد متساوي من تكرار الطول المنتظم نختار الطول الذي يكون في وسط الجدول.

- **مثال تطبيقي**: أوجد المنوال للتوزيع التالي بطريقتين، ثم مثله بيانيا عن طريق المدرج التكراري:

فئات المتغيرة	التكرار المطلق
]20-10]	8
]30-20]	10
]60-30]	12
]70-60]	2
[75-70]	1

- **الحل**: نرسم أولا الجدول الذي نحسب فيه كل المقادير ونلاحظ أن أطوال الفئات غير متساوية إذن لا بد من تعديل التكرارت:

لتعديل التكرارات نأخذ طول الفئة المكرر وهو $h=10$، ثم نقسم كل تكرار مطلق على طول الفئة المقابل له ونضرب الناتج في $h=10$ وهذا حسب القاعدة المذكورة سلفا. ثم نستخدم التكرارات المعدلة بدلا من العادية عند حساب المنوال فقط أو التمثيل البياني.

فئات المتغيرة	التكرار المطلق n_i	أطوال الفئات C_i	التكرارات المعدل n_i^*	
]20-10]	8	10	8	(8/10)10=
]30-20]	10	10	10	(10/10)10=
]60-30]	12	30	4	(12/30)10=
]70-60]	2	10	2	(2/10)10=
[75-70]	1	5	2	(1/5)10=
المجموع	33	/	/	

يتم تحديد الفئة المنوالية أولا وهي الفئة التي تقابل أكبر تكرار معدل 10 وهي]20-30[

أ. **طريقة الفروقات**:

$$Mo = A_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C_{Mo} = A_{Mo} + \frac{(n_{Mo} - n_{Mo-1})}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})} \times C_{Mo}$$

$$= 20 + \frac{(10-8)}{(10-8)+(10-4)} \times 10$$

$$Mo = 22.5$$

ب. طريقة الرافعة:

$$Mo = A_{Mo} + \frac{n^*_{Mo+1}}{n^*_{Mo+1} + n^*_{Mo-1}} \times C_{Mo} = 20 + \frac{4}{4+8}10 = 23.33$$

- **مزايا المنوال:**

✓ لا يتأثر بالقيم المتطرفة أو الشاذة (الكبيرة جدا أو الصغيرة جدا).

✓ يمكن حسابه مع البيانات الكيفية الإسمية، عندما يراد له أن يعبر عن حالة الشيوع.

✓ يمكن حساب المنوال في حالة الجداول المفتوحة.

✓ يسهل تقدير المنوال مجرد النظر إليه وخاصة إذا كانت البيانات قليلة أو ممثلة بيانيا.

✓ يعتبر أكثر المقاييس توفيقا لأنه يعبر عن القيم التي تتجمع عندها البيانات أكثر من غيرها.

- **عيوب المنوال:**

✓ لا يأخذ جميع القيم بالحسبان.

✓ قد يكون للبيانات أكثر من منوال وقد لا يكون لها منوال إذا لم تتكرر البيانات أكثر من مرة.

✓ لا تتغير قيمة المنوال عند حدوث تغير في القيم الأخرى ما دام التكرار لا يتغير .

✓ لا يمثل المنوال القيمه الوسطى في توزيع القيم وبالتالي فانه أقل دقه من المقاييس الأخرى.

✓ تتأثر قيمة المنوال بعدد الفئات في حالة الجداول التكرارية .

✓ يحتاج حسابه إلى تعديل التكرارات في حالة الجداول غير المنتظمة.

المبحث الخامس

العلاقة بين بعض المتوسطات

هناك علاقتين مهمتين نتطرق إليها الآن وهما علاقة الترتيب بين أربعة متوسطات مذكورة سابقا، بالإضافة إلى علاقة أخرى تتطلب توزيعا متماثلا لتحققها ونوردهما كما يلي:

1- علاقة ترتيب ثابتة بين المتوسطات:

توجد بين بعض المتوسطات المذكورة سابقا علاقة ترتيب وتحديدا المتوسط، التوافقي، الهندسي، الحسابي والتربيعي، على الترتيب من الأقل لإلى الأعلى، يمكن أن تكون مساعدة في عملية مراجعة الحسابات في حالة عدم تحققها (لأن ذلك يدل على خطأ)، بالإضافة إلى أنها تعطينا حصرا لبعض المتوسطات بدون حسابها، خاصة عند إيجاد صعوبة في حساب بعضها، مثل المتوسط الهندسي الذي ليس له معنى إذا كانت هناك قيم صفرية، أما إذا حسبناه بطريقة اللوغاريتمات فلا يمكن إيجاد لوغاريتمات الأصفار، وهذه العلاقة نذكرها فيما يلي:

$$H < G < \bar{X} < Q$$

- **مثال تطبيقي:** تحق من العلاقة الترتيبية السابقة للبيانات التالية:

4 3 1 2

الحل: نلخص الحسابات في الجدول التالي:

$log\ (x_i)$	$1/x_i$	x_i^2	x_i
0,30	0,50	4	2
0	1,00	1	1
0,48	0,33	9	3
0,60	0,25	16	4
المجاميع			
1,38	2,08	30,00	10,00

- المتوسط الحسابي: $\bar{X} = \frac{10}{4} = 2.50$
- المتوسط الهندسي: $G = 10^{1.38} = 2.21$
- المتوسط التوافقي: $H = \frac{4}{2.08} = 1.92$
- المتوسط التربيعي: $Q = \sqrt{\frac{30}{4}} = \sqrt{7.5} = 2.73$

إذن نلاحظ:

$$H = 1.92 < G = 2.21 < \bar{X} = 2.5 < Q = 7.5$$

2- علاقة تقريبية مشروطة بين المتوسط، الوسيط والمنوال:

توجد علاقة بين المتوسط الحسابي والمنوال والوسيط، الهدف منها إيجاد أحد المتوسطات الثلاثة الداخلة في حسابها والتي غالبا ما تكون المتوسط الحسابي في حالة الجداول المفتوحة مثلا لأنه في هذه الحالة لا يمكن تحديد مراكز الفئات التي يعتمد عليها المتوسط الحسابي، ولما كان كلا من الوسيط والمنوال لا يتأثران غالبا بالجداول التي تكون مفتوحة، فنظرا لتعدد القوانين التي تعطينا الوسيط حيث منها ما يأخذ الحد الأدنى ومنها ما يأخذ الحد الأعلى ويبقى الإختيار عند الحساب حسب نوع الجدول، أما المنوال فلا يتأثر بالتوزيعات المفتوحة، وتجدر الإشارة إلى أنه نادرا ما تكون الفئة الوسيطية أو المنوالية في أطراف الجداول، نظرا لأن التوزيعات تكون ممركزة في وسط الجداول إذا ما طبقنا النظرية الإحصائية جيدا عند إعدادها، وهذه العلاقة التي هي موضوع حديثنا في هذا البند تعطى بالصيغة الرياضية التالية:

$$\overline{\mathrm{X}} - Mo = 3\,(\bar{X} - Me)$$

غير أن صحة أو مدى دقة هذه العلاقة تبقى مرهونة بشكل التوزيع التكراري من حيث الإلتواء، حيث يشترط أن يكون التوزيع متماثلا (متناظرا) أو يقترب من التماثل على الأقل[23]. ويمكن ملاحظة ذلك من خلال التمثيل البياني أو بحساب مقاييس الشكل التي لا تعتمد على المتوسط الحسابي إذا كان الغرض تقدير المتوسط الحسابي (مجهول)، ونتطرق إلى دراسة شكل التوزيعات التكرارية من حيث الإلتواء في فصل لاحق.

- **مثال تطبيقي**: لتكن البيانات التالية:

4 0 **2** 1 0 **2** 1 **2** 3 3 **2**

- نلاحظ أن المنوال يساوي 2
- ترتيب البيانات من اجل حساب الوسيط:

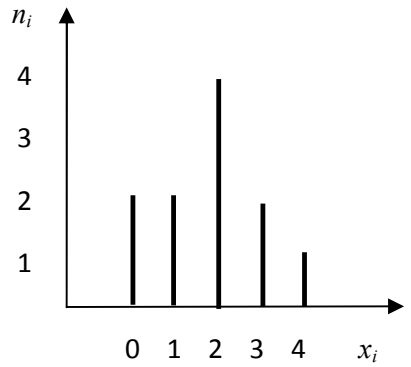
4 3 3 2 2 **2** 2 1 1 0 0

إذن:

$$Mo = 2\,; \quad Me = 2$$

لو نمثل بيانيا البيانات (نستخرج التكرارات):

Σ	4	3	2	1	0	x_i
11	1	2	4	2	2	n_i



نلاحظ أن التوزيع التكراري يقترب من التماثل (التناظر) وبالتالي يمكن الإستعانة بالعلاقة السابقة لتقدير المتوسط الحسابي ومن ثمة مقارنته بالقيمة الحقيقة له:

$$\overline{X} - 2 = 3\,(\bar{X} - 2) \Leftrightarrow \bar{X} - 3\bar{X} = -4 \Leftrightarrow \bar{X} = \frac{-4}{-2} = \mathbf{2}$$

نحسب المتوسط الحسابي بطريقة عادية (نستخدم التكرارات للتسهيل) نجد:

$$\bar{X} = \frac{0 \cdot 2 + 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 2 + 4 \cdot 1}{11} = \mathbf{1.82}$$

وهي قيمة مقاربة لما وجدناه ولكن يزداد قرب القيمتين كلما كبر حجم العينة واقترب التوزيع من التناظر أكثر.

تمارين

- **تمرين 1** : أجب على الأسئلة التالية:

1- ما هو مفهومك لمقاييس النزعة المركزية؟

2- أذكر مقاييس النزعة المركزية التي تعرفها،

3- ما هي أهم العلاقات بين هذه المقاييس (بعضها)

4- ما المقصود بشبيهات الوسيط.

5- ما هي مزايا المتوسط الحسابي؟

6- ما هي عيوب المتوسط الحسابي؟

- **تمرين 2**: لتكن سلسلة البيانات التالية:

4 5 8 9 10 11 6 6 3 2 3 3 2

1- أحسب المتوسطات التالية: المتوسط الحسابي (بطريقتين)، المتوسط التوافقي، المتوسط التربيعي، المتوسط الهندسي

2- قارن بين المتوسطات المحسوبة.

3- أحسب كلا من الوسيط والمنوال.

4- أوجد كل الربيعيات.

5- أحسب العشير الثالث.

- **تمرين 3**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	10
]30-20]	15
]40-30]	12
[50-40]	8

1- احسب المتوسط الحسابي بثلاثة طرق، الوسيط، المنوال (بطريقتين)، المتوسط التوافقي، المتوسط التربيعي، المتوسط الهندسي.

2- أوجد كلا من الربيع الثالث، العشير الخامس، المئين الخمسون، والمئين الخامس والستون.

3- ما هي أهم المتطابقات بين شبيهات الوسيط.

4- أرسم بيانيا هذا التوزيع التكراري.

- **تمرين 4**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق n_i
]10-20]	8
]20-30]	10
]30-60]	12
]60-70]	2
[70-75]	1
المجموع	33

1- أحسب كل مقاييس النزعة المركزية المدروسة

2- مثل بيانيا التوزيع التكراري

- **تمرين 5**: عددان متوسطهما الحسابي 20.5 ومتوسطهما الهندسي 20، أوجد هذين العددين.

• مراجع وإحالات الفصل الثالث:

[1] دومنيك سلفاتور "الاحصاء والاقتصاد القياسي" سلسلة ملخصات شوم، ترجمة: سعدية حافظ منتصر، ديوان المطبوعات الجامعية، الجزائر، ص 17.

[2] عوض منصور، علي قوقزة، عزام صبري "أساسيات علم الإحصاء الوصفي" دار صفاء، ط 1، 1999، الأردن، ص81.

[3] المرجع السابق، ص 82.

[4] شرف الدين خليل "الإحصاء الوصفي" شبكة الأبحاث والدراسات الإقتصادية، [على الخط] 2017/12/06 الرابط:
www.rr4ee.net

[5] عبد الرحمن بن محمد سليمان أبو عمة، محمود محمد ابراهيم هندي "الإحصاء التطبيقي" العبيكان، 1995، م. ع. السعودية، ص 58.

[6] عوض منصور وآخرون، مرجع سابق، ص 86.

[7] المرجع السابق، ص 86.

[8] P. S. MANN «Introductory Statistics » WILEY, 7eme Ed., USA, P 91.

[9] كلية الادارة والاقتصاد، جامعة بابل، العراق [على الخط] 2017/12/02 الرباط:
http://www.uobabylon.edu.iq/uobcoleges/lecture.aspx?fid=9&lcid=38550

[10] مبارك أسبر ديب "مبادئ في الاحتمالات والإحصاء" جامعة تشرين، كلية العلوم، 2009/2008، سوريا، ص 49.

[11] المرجع السابق، ص 50.

[12] موراي شبيجل "نظريات ومسائل في الإحصاء" ترجمة: أحمد حسن الموازيني، الدار الدولية للنشر والتوزيع، 1988، مصر، ص97.

[13] المرجع السابق، ص 97.

[14] المرجع السابق، ص 98.

[15] عوض منصور وآخرون، مرجع سابق، ص 144.

[16] عبد الرحمن بن محمد سليمان أبو عمة، محمود محمد ابراهيم هندي ، مرجع سابق، ص 66.

[17] I. CANNAZ « Introduction a la Statistique » INSA de Lyon, France, P 18.

[18] عبد الحميد عبد المجيد البلداوي "أساليب البحث العلمي والتحليل الإحصائي" دار الشروق، ط1، 2007، الأردن، ص 144.

[19] مجدي الطويل "الإحتمالات" دار النشر للجامعات، 2009، مصر، ص 338.

[20] مبارك أسبر ديب "مبادئ في الاحتمالات والإحصاء" جامعة تشرين، كلية العلوم، 2009/2008، سوريا، ص 55.

[21] علي لزعر "الاحصاء وتوفيق المنحنيات" د. م. ج، 2000، الجزائر، ص 49.

[22] علي لزعر، سبق ذكره، ص 48.

[23] عبد الرحمن بن محمد سليمان أبو عمة، محمود محمد ابراهيم هندي ، مرجع سابق، ص 90.

الفصل الرابع

مقاييس التشتت

تمهيد:

بعدما رأينا في الفصل السابق المقاييس التي تصف مركز البيانات وتحاول إيجاد القيمة التي تنزع البيانات إليها أو التمركز حولها، واختلاف المقاييس ومعانيها وكذا وضوح معاني بعضها وغموض الآخر، ونظرا لمحدودية هذه المقاييس في المقارنة بين التوزيعات التكرارية أو البيانات بصفة عامة. من هذا المنطلق نتطرق في هذا الفصل غلى صنف آخر من المقاييس يساعدنا في المقارنة بين البيانات حتى وإن تساوت في مقاييس النزعة المركزية، وتسمى هذه المقاييس بمقاييس التشتت أو الإختلاف، حيث تصف لنا مدى قرب أو بعد البيانات عن مركزها، ونقسم بدورها إلى مقاييس مطلقة للتشتت ومقاييس نسبية.

المبحث الأول
مفهوم مقاييس التشتت وأنواعها

في هذا المبحث نتطرق إلى مفهوم التشتت وما هي أهميته في دراسة البيانات، بالإضافة إلى طرق قياسه وذلك بالتطرق غلى أهم مقاييس التشتت، وتصنيفها إلى قسمين منها ما هو مطلق ومنها ما هو نسبي، ومعرفة أهمية كل صنف واستعمالاته، فيما تبقى طرق حساب كل مقياس للمباحث الباقية من هذا الفصل.

1- مفهوم مقاييس التشتت: Mesures de Dispersion

1-1- تعريف التشتت: التشتت يعني مدى التقارب أو التباعد في قيم المشاهدات عن بعضها البعض، وان هذا التقارب والتباعد يقاس بدرجة انتشار البيانات حول الوسط أو قيمة معينة. وتسمى درجة الانتشار هذه بالتشتت، التبعثر أو الإختلاف أو تغير البيانات[1]. وكلما كان التشتت أو التبعثر محدوداً كلما اعتبرت البيانات متجانسة وكلما توسع هذا التشتت أو التبعثر كلما كانت البيانات غير متجانسة.

1-2- قياس التشتت وأهميته: يقاس التشتت لأي مجموعة بيانات بواسطة مجموعة من المقاييس تسمى بمقاييس التشتت، وتبرز الحاجة إلى مقاييس التشتت لتقديم وصف دقيق حول التوزيعات التكرارية وخصوصا عند إجراء المقارنات بين المجموعات أو بين الأفراد أو المفردات، فهي تساعد على فهم أدق وأوضح لتوزيع المشاهدات مما يمكننا من إجراء المقارنات بنوع من الثقة. وهذا الأمر قد نستطيع تلمسه بشكل أفضل إذا نظرنا إلى مجموعتين من المشاهدات لهما نفس الوسط لكنهما يختلفان في تشتتهما[2].

- **مثال توضيحي:** فعلى سبيل المثال إذا كانت علامات مجموعة الطلبة من الشعبة (أ) في امتحان الرياضيات هي (2، 1، 3، 10) فان وسط العلامات يكون 4 ويكون وسط علامات الشعبة (ب) والمؤلفة من العلامات (3، 4، 5، 4) أيضا 4 وعلى الرغم من تساوي المعدلات إلا أن معظم علامات طلاب الشعبة (ب) كانت واقعة فعلا حول المعدل فيما ابتعدت في حالة الشعبة (أ) عنه وهذا يعني أن المعَدل في حالة الشعبة (ب) قد يكون معبرا أكثر عن جل العلامات على عكس المعَدل في حالة الشعبة (أ)، إذا فمقاييس التشتت تعطينا بعدا آخر للنظر منه إلى البيانات ومعرفة مدى تجانسها من عدمه وخاصة عند المقارنة بين مجموعتين أو أكثر.

2- أنواع مقاييس التشتت:

هناك عدة مقاييس للتشتت تختلف عن بعضها البعض من حيث الدقة في التعبير والسهولة في العمليات الحسابية وكذلك من حيث الأسس والمبادئ النظرية المبنية عليها وأهمها:

2-1- مقاييس التشتت المطلق: وهي مقاييس التشتت التي تكون وحداتها نفس وحدات القيم الأصلية أو مرفوعة بأس (قوة) معين، ومنها المدى، الانحراف المتوسط، التباين (تربيع الوحدة الأصلية)، الانحراف المعياري

(القياسي)، وتستخدم هذه المقاييس للمقارنة بين البيانات التي لها نفس وحدة القياس، نظرا لأن وحدة هذه المقاييس مرتبطة بالوحدة الأصلية.

2-2- مقاييس التشتت النسبي: وهي مقاييس التشتت التي يتم فيها عزل أو إستبعاد وحدة القياس الأصلية للبيانات وبالتالي تسمح بمقارنة البيانات من ناحية التجانس حتى وإن كانت وحدات القياس الأصلية لكل مجموعة بيانات مختلفة عن المجموعة الأخرى، ويعبر عنها على شكل نسبة مئوية منها معامل الاختلاف *(C. V.)* والمدى الربيعي النسبي والعشري النسبي وهكذا... .

طريقة حساب كل نوع من هذه الأنواع وما تحتويه من مقاييس، نتطرق إليه في المباحث الباقية من هذا الفصل، بالإضافة إلى معرفة مزايا كل نوع وعيوبة.

المبحث الثاني
مقاييس التشتت المطلق
Mesures de Dispersion Absolue

في هذا المبحث نتطرق إلى مقاييس التشتت التي تعطينا قياسا معينا لدرجة تجانس البيانات ولكن بدلالة وحدة القياس الأصلية لهذه البيانات، وهذه المقاييس تساعدنا في مقارنة درجة الإختلاف أو التشتت لمجموعات البيانات التي لها وحدات قياس متشابهة، وهي: المدى، الإنحراف المتوسط، التباين والإنحراف المعياري، وفيما يلي نذكر طريقة حساب كل مقياس في حالة البيانات غير المبوبة وتلك المبوبة سواء بالفئات أو بدون فئات.

1- المدى العام أو المطلق: Range (Etendue)

يعرف المدى أو مدى البيانات بأنه المجال الذي تقع ضمنه البيانات المدروسة ونميز حالتين[3]:

1-1- للبيانات غير المبوبة أو المبوبة بدون فئات: وهو الفرق بين أعلى قيمة وأدنى قيمة ونكتب:

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

- **مثال (1):** المدى العالم للبيانات التالية هو:

7 8 4 5 2 4 6 5 6 5

$$R = x_{max} - x_{min} = 8 - 2 = 6$$

- **مثال (2):** المدى العالم للبيانات التالية هو:

المتغيرة x_i	التكرار المطلق n_i
1	2
3	3
6	5
9	1
المجموع	10

$$R = x_{max} - x_{min} = 9 - 1 = 8$$

1-2- للبيانات المبوبة بالفئات: نرمز له بـ R ويحسب بطريقتين (لا تعطيان نفس النتيجة)، إما الفرق بين مركز الفئة الأخيرة $(x_{i_{max}})$ ومركز الفئة الأولى $(x_{i_{min}})$ ، أو عن طريق الفرق بين الحد الأعلى للفئة الأخيرة(B_k) والحد الأدنى للفئة الأولى(A_1)، ونكتب:

$$R = x_{i_{max}} - x_{i_{min}}$$
$$\dot{R} = B_k - A_1 = B_{max} - A_{min}$$

نفضل الطريقة الثانية لأنها تستغرق (تحوي) جميع البيانات لأن:$\dot{R} > R$.

- **مثال تطبيقي:** المدى العام للبيانات التالية هو:

فئات المتغيرة	التكرار المطلقn_i	مراكز الفئات x_i
]20-10]	2	15
]30-20]	3	25
]40-30]	3	35
[50-40]	2	45
المجموع	10	/

$$R = x_{i_{max}} - x_{i_{min}} = 45 - 15 = 30$$
$$\dot{R} = B_k - A_1 = B_{max} - A_{min} = 50 - 10 = 40$$

- **مزايا المدى العام:**

✓ إذا كانت البيانات جميعها متساوية فإن المدى يساوي 0 أي لا يوجد تشتت.

✓ يعطينا فكرة مبدئية وسريعة عن مجال البيانات، أما عند مقارنة البيانات فالبيانات التي لها مدى أكبر هي الأقل تجانسا، وخاصة عند تساوي مراكز البيانات يعطينا بعدا آخر للمقارنة.

- **عيوب المدى العام:**

✓ لا يمكن حسابه من الجداول المفتوحة.

✓ مقياس مظلل في حالة وجود القيم الشاذة (المتطرفة)[4].

2- المدى الربيعي: Intervalle Interquartile

هو الفرق بين الربيع الثالث والربيع الأول[5]:

$$IQ = Q_3 - Q_1$$

- **المزايا:**

✓ لا يتأثر بالقيم المتطرفة.

✓ يمكن الإستعانة بالطريقة البيانية في حسابه.

✓ يستخدم كمقياس للتشتت في التوزيعات التكرارية شديدة الإلتواء.

✓ يمكن حسابه في حالة الجداول التكرارية المفتوحة.

✓ يتحدد بعدد البيانات وليس بقيمتها.

✓ هو عبارة عن مجال يحوي 50% من البيانات المشتتة حول الوسيط.

- **مثال تطبيقي:** لتكن السلسلة الإحصائية التالية التي تمثل مداخيل (بالوحدات النقدية) 11 وحدة إحصائية، والمطلوب حساب المدى الربيعي:

700، 800، 900، 900، 1000، 1100، 1100، 1200، 1300، 1300، 1400

البيانات مرتبة ترتيبا تصاعديا، رتبة الربيع الأول والثالث على الترتيب: $\frac{1(11+1)}{4} = 3$، $\frac{3(11+1)}{4} = 9$، وبالتالي: $Q_1 = 900, Q_3 = 1300$

$$IQ = Q_3 - Q_1 = 1300 - 900 = 400$$

إذن 50% من الوحدات الإحصائية (حوالي 5 وحدات احصائية) تبتعد من حيث المداخيل عن بعضها بأقل من 400 وحدة نقدية.

3- المدى المئيني:

نرمز له RP ويعني المجال الذي يحوي 80 %من البيانات وهو الفرق بين المئين التسعون والمئين العاشر، ونكتب:

$$RP = P_{90} - P_{10}$$

وله أهمية كبيرة، حيث يقلل من تحيز القيم الشاذة الكبيرة أو الصغيرة، عند حسابه أي لا يتأثر بالقيم المتطرفة نظرا لأنه لا يأخذ بعين الإعتبار طرفي التوزيع، ومن ثمة يمكن حسابه في حالة الجداول المفتوحة[6].

4- الإنحراف الربيعي أو نصف المدى الربيعي: Semi-Interquartile

ونرمز له $\left(\frac{IQ}{2}\right)$ ويحسب كما يلي[7]:

$$IQ/_2 = \frac{Q_3 - Q_1}{2}$$

- **المزايا:**

✓ لا يتأثر بالقيم المطرفة.

✓ يمكن الإستعانة بالطريقة البيانية في حسابه.

✓ يستخدم كمقياس للتشتت في التوزيعات التكرارية شديدة الإلتواء.

✓ يمكن حسابه في حالة الجداول التكرارية المفتوحة.

✓ يتحدد بعدد البيانات وليس بقيمتها.

✓ هو عبارة عن مجال يحوي 25% من البيانات المشتتة حول الوسيط.

- **مثال تطبيقي**: بالعودة إلى مثال المدى الربيعي، نحسب نصف المدى الربيعي (الانحراف الربيعي):

$$\frac{IQ}{2} = \frac{Q_3 - Q_1}{2} = \frac{400}{2} = 200$$

يمكن القول أن 50% من المداخيل تبعد في المتوسط بأقل من 200 و. ن. عن الوسيط.

5- الانحراف المتوسط: Écart Moyen

وهو الغالب يشير إلى متوسط الإنحرافات المطلقة حول المتوسط الحسابي، وأحيانا يعبر عن متوسط الإنحرافات حول الوسيط، ونتطرق إلى كلا الحالتين فيما يلي:

5-1- الإنحراف المتوسط للبيانات غير المبوبة: يحسب في حالة البيانات الخام كما يلي[8]:

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

إذن في الغالب يحسب الإنحراف المتوسط حول المتوسط الحسابي، ويمكن حسابه حول الوسيط، ويكون أقل منه حول المتوسط، وذلك لوجود الخاصية التالية:

إذا كانت لدينا مجموعة من البيانات وa قيمة ما من هذه البيانات، فإن مجموع الإنحرافات المطلقة حول الوسيط يكون أقل ما يمكن:

$$\sum_{i=1}^{n}|x_i - Me| < \sum_{i=1}^{n}|x_i - a| \ ; \ \forall\, a \neq Me$$

وتكون قيمة الإنحراف المتوسط حول الوسيط بذلك أقل منها حول المتوسط لأن حج العينة ثابت، وتعطى صيغة الإنحراف المتوسط حول الوسيط كما يلي:

$$EM_{(Me)} = \frac{\sum_{i=1}^{n}|x_i - Me|}{n}$$

ويمكن إيجاد الإنحراف المتوسط (أو متوسط الإنحرافات) حول مقياس آخر من مقاييس النزعة المركزية وهو المنوال، حيث يتم تعويض الوسيط أو المتوسط بالمنوال فقط[9].

- **مثال تطبيقي** : حساب الإنحراف المتوسط للبيانات التالية:

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

يتطلب حساب المتوسط الحسابي أولا:

$$\bar{X} = \frac{5+4+4+3+2+1}{6} = \frac{19}{8} = 2.375$$

$$EM_{(\bar{X})} = \frac{\sum_{i=1}^{6}|x_i - 2.375|}{n}$$

$$= \frac{|5-2.375| + |4-2.375| + |4-2.375| + |3-2.375| + |2-2.375| + |1-2.375|}{8}$$

$$= \frac{2.625 + 1.625 + 1.625 + 0.625 + 0.375 + 1.375}{8} = \frac{8.25}{8} = 1.03125$$

لو نحسب الإنحراف المتوسط حول الوسيط :

تحديد قيمة الوسيط (البيانات مرتبة): Me=3.5

$$EM_{(Me)} = \frac{\sum_{i=1}^{n}|x_i - Me|}{n}$$

$$= \frac{|5-3.5| + |4-3.5| + |4-3.5| + |3-3.5| + |2-3.5| + |1-3.5|}{8}$$

$$= \frac{0.5 + 0.5 + 0.5 + 0.5 + 1.5 + 2.5}{8} = \frac{6}{8} = 0.76$$

إذن نلاحظ كما سبق أن أشرنا:

$$EM_{(Me)} = 0.76 < EM_{(\bar{X})} = 1.03$$

5-2- الإنحراف المتوسط للبيانات المبوبة: ويتم إدخال التكرارات في الحساب ونحسب مراكز الفئات x_i إذا كانت البيانات مبوبة بالفئات أما إذا كان التبويب بدون فئات فنستخدم البيانات الأصلية[10]:

$$EM_{(\bar{X})} = \frac{\sum_{i=1}^{k} n_i\,|x_i - \bar{X}|}{\sum_{i=1}^{k} n_i}$$

ومثلما أشرنا في البيانات غير المبوبة، إذا كانت لدينا مجموعة من البيانات وa قيمة ما من هذه البيانات، فإن مجموع الإنحرافات المطلقة حول الوسيط يكون أقل ما يمكن:

$$\sum_{i=1}^{n} n_i|x_i - Me| < \sum_{i=1}^{n} n_i|x_i - a| \;;\; \forall\, a \neq Me$$

وتكون قيمة الإنحراف المتوسط حول الوسيط بذلك أقل منها حول المتوسط لأن حجم العينة ثابت، وتعطى صيغة الإنحراف المتوسط حول الوسيط كما يلي:

$$EM_{(Me)} = \frac{\sum_{i=1}^{k} n_i\,|x_i - Me|}{\sum_{i=1}^{k} n_i}$$

- **مثال تطبيقي** : لحساب الانحراف المتوسط حول المتوسط والوسيط، لبيانات مبوبة بفئات نقوم بما يلي:

فئات المتغيرة	التكرار المطلق n_i	مراكز الفئات x_i	$n_i\, x_i$	$\lvert x_i - \bar{X}\rvert$	$n_i \lvert x_i - \bar{X}\rvert$	ت.م.ص
]20-10]	2	15	30	15	30	2
]30-20]	**3**	25	75	5	15	**5**
]40-30]	3	35	105	5	15	8
[50-40]	2	45	90	15	30	10
المجموع	10	/	300	/	90	/

$$\bar{X} = \frac{300}{10} = 30$$

$$E.M._{(\bar{X})} = \frac{\sum_{i=1}^{4} n_i\, |x_i - 30|}{\sum_{i=1}^{4} n_i} = \frac{90}{10} = 9$$

حساب الوسيط:

$$\frac{\sum_{i=1}^{4} n_i}{2} = \frac{10}{2} = 5$$

$$Me = A_{Me} + \frac{\frac{\sum_{i=1}^{k} n_i}{2} - n_{Me-1}\nearrow}{n_{Me}} \times\ C_{Me} = 20 + \frac{5-2}{3}10 = 30$$

بما أن الوسيط يساوي الوسيط، فإن الإنحرافين يكونان متساويان.

$$EM_{(Me)} = 30$$

- **المزايا:**

✓ يعتمد في حسابه على كل القيم.

✓ يمكن حسابه عن طريق الإنحرافات حول الوسيط، مع ملاحظة أن إنحرافه عن الوسيط أقل من إنحرافه عن المتوسط الحسابي أو المنوال.

✓ أدخلت عليه القيمة المطلقة لتفادي خاصية إنعدام مجموع الإنحرافات حول المتوسط الحسابي.

- **العيوب:**

✓ قليل الإستخدام في الإحصاء الإستنتاجي لعدم قابليته للعمليات الجبرية (نظرا لوجود القيمة المطلقة).

✓ يتأثر بالقيم المتطرفة.

✓ لا يمكن حسابه في حالة الجداول التكرارية المفتوحة.

6- التباين: Variance

التباين يعتبر أهم مقياس للتشتت، نظرا لدخوله في حساب مختلف المقاييس الأخرى وسهولة حسابه، وتكون وحدته هي مربع وحدة البيانات الأصلية، وهو متوسط مربعات الانحرافات حول المتوسط الحسابي، ونميز بين البيانات غيرا المبوبة وتلك المبوبة في حسابه وذلك كما يلي[11]:

6-1- التباين للبيانات غير المبوبة:

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

- ويمكن استنتاج صيغة أخرى للتباين[12]:

$$V(X) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{X} + \bar{X}^2\right)}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - 2\bar{X}^2 + \frac{n\,\bar{X}^2}{n}$$

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

- **مثال تطبيقي**: حساب التباين للبيانات السابقة :

5 4 3 3 2 1

$$\bar{X} = 3$$

$$\sigma^2 = \frac{\sum_{i=1}^{6}(x_i - 3)^2}{n}$$

$$= \frac{(5-3)^2 + (4-3)^2 + (3-3)^2 + (3-3)^2 + (2-3)^2 + (1-3)^2}{8}$$

$$= \frac{4+1+0+0+1+4}{8} = \frac{10}{8} = 1.25$$

6-2- التباين للبيانات المبوبة:

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

- ويمكن استنتاج صيغة أخرى للتباين:

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

$$\boxed{V(X) = \frac{\sum_{i=1}^{k} n_i X_i^2}{\sum_{i=1}^{k} n_i} - \bar{X}^2 = \overline{X^2} - \bar{X}^2}$$

- **مثال تطبيقي** : لحساب التباين في حالة بيانات مبوبة بفئات نقوم بمايلي:

$n_i x_i^2$	x_i^2	$n_i(x_i - \bar{X})^2$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})$	$n_i\, x_i$	x_i	n_i	فئات المتغيرة
450	225	450	225	15-	30	15	2	]20-10]
1875	625	75	25	5-	75	25	3	]30-20]
3675	1225	75	25	5	105	35	3	]40-30]
4050	2025	450	225	15	90	45	2	[50-40]
10050	/	1050	/	/	300	/	10	المجموع

$$\bar{X} = \frac{300}{10} = 30$$

$$\sigma^2 = \frac{\sum_{i=1}^{4} n_i\,(x_i - 30)^2}{\sum_{i=1}^{4} n_i} = \frac{1050}{10} = 105$$

أو بالطريقة الثانية، أسهل نوعا ما:

$$\sigma^2 = \frac{\sum_{i=1}^{4} n_i X_i^2}{\sum_{i=1}^{4} n_i} - \bar{X}^2 = \overline{X^2} - \bar{X}^2 = \frac{10050}{10} - 30^2 = 1005 - 900 = 105$$

- **المزايا:**

✓ التباين قابل للعمليات الجبرية، لذلك هو كثير الإستخدام في القوانين والنظريات الإحصائية.

- **العيوب:**

✓ يتأثر بالقيم المتطرفة نتيجة إعتماده على المتوسط الحسابي وهذا الأخير يتأثر بها.

✓ لا يمكن حسابه في حالة الجداول المفتوحة.

✓ تباين قيمة ثابتة يساوي 0.

- **بعض الخواص الرياضية للتباين:** إذا كان X، Y متغيران ما، و a عددا ثابتا فإن[13]:

$$V(a) = 0$$
$$V(a\,X) = a^2\,X$$
$$V(X \mp a) = V(X)$$
$$V(X \mp Y) = V(X) + V(Y) \mp 2\,Cov(X,Y)$$

حيث $Cov(X,Y)$ هو تغاير X و Y (التباين المشترك) ويحسب كما يلي[14]:

$$Cov(X,Y) = \overline{X \cdot Y} - \bar{X} \cdot \bar{Y}$$

أما إذا كان X و Y مستقلان فإنه يكفي وضع[15]:

$$V(X \mp Y) = V(X) + V(Y)$$

- **ملاحظات:**

- إذا كان X و Y مستقلان يستلزم : $Cov(X,Y) = 0$
- إذا كان $Cov(X,Y) = 0$ فليس بالضرووة أن X و Y مستقلان.

- إذا كان X وY متغيرين و a، b عددين ثابتين فإن:

(1) $Cov(X,X) = V(X)$
(2) $Cov(a\,X, b\,Y) = a\,b\,Cov(X,Y)$
(3) $Cov(X,a) = 0$

7- الانحراف المعياري: Déviation Standard

الإنحراف المعياري أهم مقاييس النزعة المركزية نظرا لأن وحدته تكون من نفس طبيعة وحدة البيانات الأصلية، وهو الجذر التربيعي الموجب للتباين،و يأخذ معظم مزايا وعيوب التباين، ويدخل في حساب مقاييس أخرى، وهو سهل الحساب ويحسب كما يلي[16]:

$$\sigma = \sqrt{\sigma^2} = \sqrt{V(X)}$$

- **أهم خصائص الإنحراف المعياري:**

إن المتوسط الحسابي يدخل في حساب الإنحراف المعياري وهو كقيمة وكما ذكرنا سلفا يحافظ على وحدة قياس العينة، وهو يبين مجموع المشاهدات مقسوما على عددها، وبالتالي هو يبن النصيب المتساوي لكل مشاهدة، وتزيد أهميته كونه تتم الاستعانة به في حساب وبالتالي في تفسير الانحراف المعياري، لذا تجدر الإشارة إلى أن الانحراف المعياري يفضل على التباين كوسيلة لقياس التشتت، لأن وحدة الانحراف المعياري هي نفسها وحدة قياس العينة، فالسؤال الجوهري هنا، كيف يمكن تفسير الانحراف المعياري كوسيلة لشرح تشتت البيانات ؟

في هذا الصدد يوجد إحصائي روسي، يدعى "تشيبيشف*" ، توصل إلى اكتشاف هام في هذا الشأن؛ حيث أثبت أنه لأي مجموعة من البيانات أن 75 %، على الأقل من المشاهدات يجب أن تقع داخل إثنين وحدة انحراف معياري (2σ)، بعيدا عن المتوسط سواء أكبر أو أقل، أي $(\bar{x} \pm 2\sigma)$، وأن 89 %، على الأقل من المشاهدات يجب أن تقع داخل ثلاثة وحدات انحراف معياري من المتوسط. بصفة عامة، فقد اثبت أنه على الأقل $100\left\{1 - \left(1/K^2\right)\right\}\%$، من المشاهدات يجب أن تقع داخل K، من الانحرافات المعيارية بعيدا عن المتوسط أي $(\bar{X} \pm K \cdot \sigma)$ أو بتعبير رياضي آخر $[\bar{X} - K\,\sigma\,;\bar{X} + K\,\sigma\,]$ كمجال [17] .

- **مثال تطبيقي:** بالرجوع إلى المثال الأخير، نقوم بحساب الإنحراف المعياري:

$$\sigma = \sqrt{\sigma^2} = \sqrt{V(X)} = \sqrt{105} = 10.24$$

يمكن القول كأحد التفسيرات الممكنة، أنه حسب نظرية "تشيبشيف" أنه على سبيل المثال 89 % من البيانات تقع في المجال التالي:

$$[30 - 3 \times 10.24\,;30 + 3 \times 10.24\,] = [-0.72\,;\ \ 60.72]$$

*بافنوتي تشيبيتشيف (Pafnouti Tchebychev)، رياضي روسي (1821-1894)

ويمكن إيجاد مجال للثقة الذي يحتوي على نسبة معينة للبيانات من أجل أي قيمة تبتعد عن المتوسط الحسابي، ونلاحظ أن مجال الثقة يكون واسعا نوعا ما، حيث وفي مواضيع الإستدلال الإحصائي وعند معرفة توزيعات العينة أو المجتمع يكون مجال الثقة أضيق وموضوع التوزيعات لا يسمح المقام لذكره في هذه المطبوعة نظرا لعلاقته الوطيدة بنظرية الإحتمالات.

8- علاقة إعتبارية بين الإنحرافات:

هناك علاقتين مهمتين تقريبيتين (العلاقة الثالثة مستنتجة من الثانية) بين الانحرافات المدروسة سابقا وهي الانحراف المتوسط والانحراف الربيعي (نصف المدى الربيعي) والإنحراف المعياري[18].

$$\text{الانحراف المتوسط} = \frac{4}{5}\left(\text{الانحراف المعياري}\right)$$

$$\text{الانحراف الربيعي} = \frac{2}{3}\left(\text{الانحراف المعياري}\right)$$

$$\text{المدى الربيعي} = \frac{4}{3}\left(\text{الانحراف المعياري}\right)$$

- **ملاحظة:** وتعتبر هذه العلاقات تجريبية حيث أنها تكون أكثر صحة كلما كان التوزيع اقرب إلى التناظر

المبحث الثالث

مقاييس التشتت النسبي

والقيمة المعيارية

في هذا المبحث نتطرق إلى ذلك النوع من مقاييس التشتت التي تعزل لنا وحدة القياس الأصلية للبيانات لتعبر عن مدى تشتت البيانات عن طريق النسب أو النسب المئوية، وبالتالي تساعد في مقارنة البيانات التي تختلف في وحدة قياسها، وبصفة عامة:

$$\text{التشتت النسبي} = \frac{\text{التشتت المطلق}}{\text{المتوسط}}$$

حيث التشتت المطلق قد يكون المدى أو المدى الربيعي أو الانحراف المعياري ...، أما المتوسط قد يكون الوسيط أو المتوسط الحسابي ...الخ.

بالإضافة إلى التطرق إلى أحد أهم القيم المعروفة في علم الإحصاء وهي القيمة المعيارية.

1- مقاييس التشتت النسبي : Mesures de dispersion relative

تحسب مقاييس التشتت النسبي إعتمادا على مقاييس التشتت المطلق أو بعض مقاييس النزعة المركزية وتكون في شكل نسب ومعزولة عن وحدة القياس:

1-1- المدى النسبي: يستخدم المدى النسبي للمقارنة بين البيانات التي تتساوى في المدى الطلق وتختلف في المتوسط الحسابي، ويحسب كما يلي:

$$R^{\%} = \frac{R}{\overline{X}} \times 100\ \%$$

1-2- النسبة بين المدى الربيعي والمدى العام : $(\mathcal{R})$ يبين هذا المقياس تشتت 50% من الوحدات الإحصائية المركزية (أي التي تقع في مركز التوزيع) حول الوسيط مقارنة بالمدى العام، ونكتب:

$$\mathcal{R} = \frac{(Q_3 - Q_1)}{R} \times 100\ \% = \frac{IQ}{R} \times 100\ \%$$

تأخذ مزاياها وعيوبها من المقاييس الداخلة في حسابها، لتضيف لنا كأحد المزايا إمكانية المقارنة بين البيانات المختلفة في وحدة القياس.

- **مميزات هذه النسبة:**

✓ تستعمل لقياس تشتت 50% من الوحدات الإحصائية التي تقع حول القيمة المركزية بالنسبة لنفس التوزيع.

✓ إذا كانت $\mathcal{R} = 50\ \%$ يكون التوزيع الإحصائي متماثل أو متناظر، أما إذا كانت النسبة أكبر من 50% فإن التوزيع يكون قوي التشتت بالنسبة للقيمة المركزية، أما إذا كانت النسبة اقل من 50% فإن التوزيع التكراري يكون قليل التشتت بالنسبة للقيمة المركزية (الوسيط).

- **مثال تطبيقي**: بالعودة إلى مثال المدى الربيعي (في مقاييس التشتت المطلق)، نحسب النسبة المعنية :

$$\mathcal{R} = \frac{(Q_3 - Q_1)}{R} \times 100\ \% = \frac{IQ}{R} \times 100\ \% = \frac{400}{1400 - 700} 100 = 57.14\ \%$$

إذن النسبة أكبر من 50% ولا تبتعد عنها كثيرا، وبالتالي هناك تشتت نسبي، ولكن أقرب غلى التماثل.

1-3- معامل الاختلاف المعياري: Coefficient de Variation ، يحسب كنسبة مئوية بين الإنحراف المعياري والمتوسط الحسابي، وفق العلاقة التالية[19]:

$$CV = \frac{\sigma}{\bar{X}} \times 100\ \% \quad ;\ \bar{X} \neq 0$$

تظهر أهمية مقياس معامل الإختلاف عند مقارنة مجموعتي بيانات أو أكثر حيث المجموعة التي تظهر بنسبه اقل تدل على تجانس بياناتها أي أن تشتتها أقل، حتى وإن كانت وحدة قياس هذه البيانات لم تكن من نفس الطبيعة، (مثل مقارنة تشتت أوزان وأطوال الأفراد في مجتمع ما) لأنه بعزل وحدة القياس عن طريق حساب النسبة (وحدة القياس على نفسها تختزل)،[20] ولكن إذا لم يكن الغرض هو المقارنة بل معرفة مدى تشتت البيانات فكلما زادت قيمته زاد التشتت ويمكن القول أنه إذا كان هذا المعامل أكبر من 50% يمكن إعتبار قيمته كبيرة (هناك تشتت)، وتجدر الإشارة إلى أنه من أهم عيوبه هو أنه يصبح لا معنى له إذا كان المتوسط الحسابي يساوي الصفر. بالإضافة إلى عيوب المقاييس الداخلة في حسابه مثل أنه لا يمكن حسابه في الجداول المفتوحة لأنه يعتمد على المتوسط الحسابي... الخ.

1-4- معامل الإختلاف الربيعي: Coefficient de Variation Quadratique، وهو النسبة المئوية بين المدى الربيعي والربيع الثاني أو الوسيط، ويعطى بالعلاقة التالية:

$$CVQ = \frac{Q_3 - Q_1}{Q_2} \times 100\ \% = \frac{Q_3 - Q_1}{Me} \times 100\ \%$$

رأينا في معامل الإختلاف العادي السابق أنه لا يمكن حسابه في حالة الجداول المفتوحة على عكس معامل الإختلاف الربيعي حيث يمكن ذلك لأنه يعتمد على الربيعيات فقط في حسابهن وعند مقارنة مجموعة من البيانات تكون البيانات التي نسبتها اكبر عي الأكثر تشتتا (أو الأقل تجانسا)، بالإضافة إلى أنه يصبح لا معنى له -من عيوبه- إذا كان الوسيط (الربيع الثاني) مساويا للصفر.

2- القيمة المعيارية (z) : Valeur Standard، يعطى المتغير المعياري (z_i) أو الدرجات المعيارية بالصيغة التالية:

$$z_i = \frac{x_i - \bar{X}}{\sigma}$$

وتقيس لنا القيمة المعيارية الإنحرافات عن المتوسط الحسابي بوحدات من الانحراف المعياري، وهي كمية لا وحدة قياس لها (مستقلة عن وحدة قياس البيانات الأصلية) قد تكون موجبة (الدرجة المعيارية للقيمة التي أعلى من المتوسط الحسابي) أو سالبة (الدرجة المعيارية للقيمة التي اقل من المتوسط الحسابي للبيانات)، وتسمى هذه القيم بالوحدات المعيارية أيضا (تشكل متغير معياري) ولها أهمية كبيرة في المقارنة بين التوزيعات التكرارية ح، حيث التوزيع الذي له درجة معيارية عند قيمة ما، هو من تكون فيه هذه القيمة مبتعدة بالدرجة المعيارية (بالنسبة إلى الإنحراف المعياري) الموجودة عن المتوسط الحسابي[21].

- **ملاحظة:** يمكن تحويل أي بيانات إلى متغيرة معيارية، يكون متوسط المتغيرة المعيارية هو 0، أما انحرافها المعياري يساوي 1. وهذه القيم المعيارية لها أهمية كبيرة في الإحصاء الإستدلالي[22]. ويتطرق إليها الباحث في مطبوعات أخرى (مطبوعة نظرية الإحتمالات، مطبوعة الإحصاء الرياضي).

- **مثال تطبيقي:** حصل أحد الطلاب على علامة 84 في امتحان الرياضيات حيث كان متوسط العلامات 76 وانحرافها المعياري 10. أما في امتحان العلوم الطبيعية حصل على علامة 90 حيث كان متوسط العلامات فيها 82 وانحرافها المعياري 16. ففي أي مادة كانت درجة استيعابه أعلى؟

الحل: نحسب الدرجة المعيارية للعلامة التي حصل عليها الطالب في كل إمتحان ونستخدم الرموز الموضحة أسفله كما يلي:

z_m: الدرجة المعيارية للرياضيات:

$$z_m = \frac{x_1 - \bar{X}}{\sigma} = \frac{84 - 76}{10} = 0.8$$

z_s : الدرجة المعيارية للعلوم الطبيعية :

$$z_s = \frac{x_2 - \bar{X}}{\sigma} = \frac{90 - 82}{16} = 0.5$$

تفسير النتيجيتين: نلاحظ أن علامة الطالب في الرياضيات بلغت 0.8 درجة معيارية أعلى من المتوسط الحسابي (لأنها موجبة نقول أعلى) أما في العلوم الطبيعية كانت 0.5 درجة معيارية أعلى من المتوسط الحسابي، وبالتالي فغن استيعابه النسبي كان في الرياضيات أعلى منه في العلوم الطبيعية بالرغم من أن علامته في العلوم الطبيعية كانت أعلى.

تمارين

- **تمرين 1**: أجب على الأسئلة التالية:

1- ما هو مفهوم مقاييس التشتت

2- ما هي أنواع مقاييس التشتت.

3- أذكر مقاييس التشتت المدروسة وما هي مزايا وعيوب كل مقياس.

4- ماذا نقصد بالمتغير المعياري وما هي فائدته.

- **تمرين 2**: لتكن سلسلة البيانات التالية:

4 6 6 3 2 3 3 2

✓أحسب مقاييس التشتت التالية: المدى، الانحراف المتوسط، التباين، الانحراف المعياري ومعامل الإختلاف.

- **تمرين 3**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	10
]30-20]	15
]40-30]	12
[50-40]	8

1- احسب المتوسط الحسابي.

2- أحسب مقاييس التشتت المطلق.

3- أحسب مقاييس التشتت النسبي.

4- ما هو الفرق بين مقاييس التشتت النسبي والمطلق.

5- أرسم بيانيا هذا التوزيع التكراري.

- **تمرين 4**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق n_i
]20-.....[	8
]30-20]	10
]60-30]	12
]70-60]	2
.....-70]	1
المجموع	33

1) ما هو نوع جدول التوزيع التكراري

2) أحسب مقاييس النزعة المركزية الممكنة.

2) أحسب مقاييس التشتت المطلق الممكنة.

3) أحسب مقاييس التشتت النسبي الممكنة.

- **تمرين 5**: لتكن البيانات التالية الخاصة بنقاط 56 طالب في مقياس الإقتصاد الجزئي:

12 16 08 08 17 14 06 11 11 05 16 03 05 10 11 07 08 12 10
16 08 08 17 14 06 11 11 05 16 03 05 10 11 07 08 12 10 19
19 12 16 08 08 17 14 06 19 12 16 08 08 17 14 06 19 12

– المطلوب:

1- قم بتبويب البيانات.

2- أحسب المتوسط الحسابي بطريقتين

3- أحسب المدى بطريقتين.

4- أحسب الإنحراف المعياري والتباين.

- **تمرين 6** : ليكن التوزيع التكراري التالي :

الفئات الأجرية	عدد العمال
]10-20]	15
]20-30]	20
]30-40]	30
]40-50]	20
[50-60]	5
المجموع	90

1- أحسب متوسط الأجور بطريقتين مختلفتين.

2- أحسب كل أنواع المدى.

3- أحسب كل أنواع الانحرافات.

4- أحسب كل مقاييس التشتت النسبي الممكنة.

- **تمرين 7**: ليكن التوزيع التكراري التالي :

الفئات	التكرار المطلق
10-0	5
20-10	10
40-20	15
60-40	15
70-60	5
المجموع	50

1- أحسب المتوسط الحسابي، والإنحراف المعياري.

2- ما هو الفرق الأساسي بين التباين والمتوسط الحسابي.

3- احسب جميع الدرجات المعيارية للمتغير، ثم أحسب متوسطها وإنحرافها المعياري

• مراجع وإحالات الفصل الرابع:

[1] موراي شبيجل "نظريات ومسائل في الإحصاء" ترجمة: أحمد حسن الموازيني، الدار الدولية للنشر والتوزيع، 1988، مصر، ص112.

[2] علاء الدين القبانجي، حسام حمامة كرمجي "الإحتمال والإحصاء" منشورات جامعة دمشق، 2012/2011، سوريا، ص 32.

[3] عبد الحميد عبد المجيد البلداوي "أساليب البحث العلمي والتحليل الإحصائي" دار الشروق، ط1، 2007، الأردن، ص 155.

[4] عبد العزيز فهمي هيكل "مبادئ الأساليب الإحصائية" المركز الدولي لتعليم الإحصاء، ط1، 1966، لبنان، ص 266.

[5] P. S. MANN «Introductory Statistics » WILEY, 7eme Ed., USA, P 90.

[6] عوض منصور، علي قوقزة، عزام صبري "أساسيات علم الإحصاء الوصفي" دار صفاء، ط 1، 1999، الأردن، ص 160.

[7] موراي شبيجل، مرجع سابق، ص 113.

[8] المرجع السابق، ص 113.

[9] عبد الرحمن عيسوي "الإحصاء السيكولوجي التطبيقي" دار المعرفة الجامعية، 2000، مصر، ص 39.

[10] A. AZZOUZ « STATISTIQUE : cours , exercices et problème résolus » Tome 1, OPU, 2016, Algérie, P 125.

[11] مبارك أسبر ديب "مبادئ في الاحتمالات والإحصاء" جامعة تشرين، كلية العلوم، 2009/2008، سوريا، ص 68.

[12] K. KHALED « Methodes Statistiques » OPU, 2005, Algérie, P 6.

[13] موراي شبيجل، جون شيلر، ألو سريبيقاسان "الإحصاء والاحتمالات" ملخصات شوم إيزي، ترجمة: محمود علي أبو النصر، مصطفى جلال مصطفى، الدار الدولية للإستثمارات الثقافية، 2003، مصر، ص 39.

[14] A. j. BARLOW « Statistics » WILEY , 1991 , USA , P 15.

[15] F. GUENARD , P. HUG « MATHEMATIQUES : *Algèbre et Probabilités* » DUNOD, 1993, France, P 231.

[16] مبارك أسبر ديب، مرجع سابق، ص 69.

[17] جورج كانافوس، دون ميلر "الإحصاء للتجاريين: مدخل حديث" ترجمة: سلطان محمد عبد الحميد، دار المريخ، 2003، م. ع. السعودية، ص99.

[18] موراي شبيجل، مرجع سابق، ص 137.

[19] B. L. AGARWAL « Basic Statistics » New Age International Publishers, 4th Ed. 2006, India, P 70.

[20] P. S. MANN, Op. Cit., P 98.

[21] موراي شبيجل، مرجع سابق، ص 117.

[22] عبد الله فلاح المنيزل، عايش موسى غرايبة "الإحصاء التربوي" دار المسيرة، الأردن، ص 81.

الفصل الخامس

العزوم ومقاييس الشكل

تمهيد

إذا كانت مقاييس النزعة المركزية ومقاييس التشتت تسمح بتلخيص بيانات أي ظاهرة في صورة أرقام بسيطة تعطي فكرة عن خصائص توزيع هذه البيانات ودرجة تجانسها أو اختلافها فإن هذا الوصف الذي تعطيه يبقى تنقصه الدقة الكافية المطلوبة للتعرف على خواص التوزيع خاصة من حيث انتشار البيانات على المنحنى البياني الممثل لها من حيث التوائه أو تفلطحه عن الوضعية الطبيعية وهي تلك الوضعية التي يكون فيها شكل التوزيع التكراري متناظرا بالنسبة للمتوسط الحسابي، ومعتدلا أي ليس متطاولا وليس منبسطا، وهذا ما نحاول توضيحه في هذا الفصل الذي يرتكز أساسا على الفصول الأولى لما نستخدمه من مقاييس عند الحساب وكذا لفهم التفسيرات المقدمة بعد العمليات الحسابية، لذا نجد أنه تدعونا الحاجة لاستخدام مقاييس أخرى لتحقيق هذا الغرض تسمى عموما مقاييس الشكل وتنقسم إلى مجموعتين حيث تهتم الألوى بشكل التوزيع من حيث الإلتواء، أما المجموعة الثانية فتحدد لنا شكل التوزيع من حيث التفلطح أو الإنبساط، غير أننا في البادية يجب التعريج على أحد المفاهيم الإحصائية المهمة وهي العزوم وأنواعها نظرا لأهميتها الكبيرة في مقاييس الشكل.

المبحث الأول

حسـاب العـزوم

Moments

للعزوم أهمية كبيرة في علم الإحصاء يصفه عامة، حيث أنها تساهم في معرفة شكل التوزيعات التكرارية، بالإضافة إلى أنها تساعد كذلك في استنتاج بعض المقاييس مثل المتوسط الحسابي والتباين والمتوسط التربيعي، نظرا لكون هنا علاقة تربط بين العزوم بأنواعها، فالإضافة إلى ذلك فلها أهمية كبيرة في معرفة شكل التوزيع التكراري من حيث الالتواء والتفلطح، كما نأتي على ذكره لاحقا.

1- العزوم الابتدائية (اللامركزية، البسيطة، الصفرية): Moments Simple

وتسمى بالعزوم حول الصفر، وتحسب في حالة البيانات المبوبة (إذا كانت التبويب بالفئات نحسب مراكز الفئات x_i كالعادة) وغير المبوبة كمايلي:

1-1- حالة البيانات غير المبوبة: يعطى العزم الابتدائي أو غير المركزي من الدرجة r (العزم الرائي حول الصفر) في حالة البيانات غير المبوبة بالعلاقة التالية[1]:

$$M_r = \frac{\sum_{i=1}^{n} {x_i}^r}{n} \quad ; \ r \in \mathbb{N}$$

ونأخذ بعض الرتب على سبيل المثال:

$$M_0 = \frac{\sum_{i=1}^{n} {x_i}^0}{n} = \frac{n}{\mathrm{n}} = 1$$

$$M_1 = \frac{\sum_{i=1}^{n} {x_i}^1}{n} = \overline{X}$$

$$M_2 = \frac{\sum_{i=1}^{n} {x_i}^2}{n} = \sqrt{Q}$$

$$M_3 = \frac{\sum_{i=1}^{n} {x_i}^3}{n}$$

$$\vdots$$

1-2- حالة البيانات المبوبة: يعطى العزم الابتدائي من الدرجة r في حالة البيانات المبوبة، بالفئات أو بدون فئات كما يلي (يتم أخذ مراكز الفئات مكان x_i في حالة تبويب بالفئات)[2]:

$$M_r = \frac{\sum_{i=1}^{k} n_i \, {x_i}^r}{n} \quad ; \ r \in \mathbb{N}$$

ونأخذ بعض الرتب على سبيل المثال:

$$M_0 = \frac{\sum_{i=1}^{k} n_i \, {x_i}^0}{n} = 1$$

$$M_1 = \frac{\sum_{i=1}^{k} n_i \, {x_i}^1}{n} = \bar{X}$$

$$M_2 = \frac{\sum_{i=1}^{k} n_i \, {x_i}^2}{n} = \sqrt{Q}$$: جذر المتوسط التربيعي

$$M_3 = \frac{\sum_{i=1}^{k} n_i \, {x_i}^3}{n}$$

$$\vdots$$

- **مثال تطبيقي شامل:** نأخذ هذا التوزيع التكراري ونحاول تطبيق العزوم بأنواعها علية وكذلك بالنسبة لمقاييس الشكل والتمثيل البياني فيما بعد (نكتفي بالبيانات المبوبة):

الفئات	التكرار المطلق
3-1	1
5-3	3
7-5	5
9-7	2

لنحسب العزوم الابتدائية الخمسة الأولى:

الفئات	x_i	n_i	${x_i}^0$	$n_i {x_i}^0$	x_i	$n_i x_i$	${x_i}^2$	$n_i {x_i}^2$	${x_i}^3$	$n_i {x_i}^3$	${x_i}^4$	$n_i {x_i}^4$
3-1	2	1	1	1	2	2	4	4	8	8	16	16
5-3	4	3	1	3	12	12	16	48	64	192	256	768
7-5	6	4	1	4	24	24	36	144	216	864	1296	5184
9-7	8	2	1	2	16	16	64	128	512	1024	4096	8192
المجموع	/	10	/	10		54		324	/	2088	/	14160

$$M_0 = \frac{10}{10} = 1$$: ثابت

$$M_1 = \frac{54}{10} = 5.4 = \bar{X}$$

$$M_2 = \frac{324}{10} = 32.4$$

$$M_3 = \frac{2088}{10} = 208.8$$

$$M_4 = \frac{14160}{10} = 1416$$

2- العزوم المركزية : Moments Centrés

وتسمى بالعزوم حول المتوسط الحسابي (ويمكن حسابها حول أي نقطة a وفي هذه الحالة لا تسمى عزوما مركزية بل عزوم حول نقطة ثابتة[3])، وتحسب في حالة البيانات المبوبة (إذا كانت التبويب بالفئات نحسب مراكز الفئات x_i كالعادة) وغير المبوبة كمايلي:

2-1- حالة البيانات غير المبوبة: يعطى العزم المركزي من الدرجة r (العزم الرائي حول المتوسط الحسابي) بالعلاقة التالية[4]:

$$\mu_r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^r}{n} \quad ; r \in \mathbb{N}$$

ونأخذ بعض الرتب على سبيل المثال:

$$\mu_0 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^0}{n} = 1$$
$$\mu_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^1}{n} = 0$$
$$\mu_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} = \sigma^2$$
$$\mu_3 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^3}{n} = \sigma^3$$
$$\vdots$$

- **ملاحظة:** إذا عوضنا المتوسط الحسابي بالصفر نتحول إلى العزوم الإبتدائية.

2-2- حالة البيانات المبوبة: إذا كانت البيانات مبوبة سواء بالفئات (يتم حساب مراكز الفئات) أو بدون فئات (توزيع تكراري) يعطى العزم الممركز من الدرجة r بالعلاقة التالية[5]:

$$\mu_r = \frac{\sum_{i=1}^{k} n_i\,(x_i - \bar{X})^r}{n} \quad ; r \in \mathbb{N}$$

ونأخذ بعض الرتب على سبيل المثال:

$$\mu_0 = \frac{\sum_{i=1}^{k} n_i\,(x_i - \bar{X})^0}{n} = 1$$
$$\mu_1 = \frac{\sum_{i=1}^{k} n_i\,(x_i - \bar{X})^1}{n} = 0$$
$$\mu_2 = \frac{\sum_{i=1}^{k} n_i\,(x_i - \bar{X})^2}{n} = \sigma^2$$
$$\mu_3 = \frac{\sum_{i=1}^{k} n_i\,(x_i - \bar{X})^3}{n}$$
$$\vdots$$

- **من المثال السابق:** لنحسب العزوم المركزية الخمسة الأولى: لدينا المتوسط الحسابي 5.4:

$n_i(x_i-5.4)^4$	$n_i(x_i-5.4)^3$	$n_i(x_i-5.4)^2$	$(x_i-5.4)^2$	$n_i(x_i-5.4)$	$n_i(x_i-5.4)^0$	$(x_i-5.4)^0$	$x_i-5.4$	n_i	x_i	الفئات
133.6336	-39.304	11,56	11,56	-3,4	1	1	-3,4	1	2	3-1
11.5248	-8.232	5,88	1,96	-4,2	3	1	-1,4	3	4	5-3
0.5184	0.864	1,44	0,36	2,4	4	1	0,6	4	6	7-5
91.3952	35.152	13,52	6,76	5,2	2	1	2,6	2	8	9-7
237.072	-11.52	32,4	/	0	10	/	/	10	/	المجموع

$\mu_0 = \frac{10}{10} = 1$: ثابت

$\mu_1 = \frac{0}{10} = 0$: ثابت

$\mu_2 = \frac{32.4}{10} = 3.24 = \sigma^2$

$\mu_3 = \frac{-11.52}{10} = -1.152$

$\mu_4 = \frac{237.072}{10} = 23.7072$

3- العلاقة بين العزوم الابتدائية والعزوم المركزية

3-1- دستور نيوتن (ثنائي الحد): إذا كان a وb عددان حقيقيان، وn عددا طبيعيا، يعطى قانون ثنائي الحد لنيوتن بالعلاقة التالية[6]:

$$(a-b)^n = \sum_{x=0}^{n} C_n^x (-1)^x a^{n-x} b^x$$

حيث C_n^x تسمى التوفيقة تحسب كما يلي[7]:

$$C_n^x = \frac{n!}{(n-x)! \cdot x!}$$

و$n!$ يدعى عاملي أو مضروب n يحسب بالصيغة التالية:

$$n! = \begin{cases} \prod_{i=0}^{n-1}(n-i) = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 & si \ n \neq 0 \\ 1 & si \quad n = 0 \end{cases}$$

- **مثال (1) :** حساب C_3^2:

$$C_3^2 = \frac{3!}{(3-2)! \cdot 2!} = \frac{3 \times 2 \times 1}{1 \cdot 2 \times 1} = 3$$

موضوع التوفيقات يتناوله الباحث في مطبوعة أخرى بعنوان "نظرية الإحتمالات" ، وبصفة عامة يمكن حساب أي C_n^k بالإعتماد على مثلث "**باسكال**"، (من أجل قيم أكبر، أنظر **الملحق -3-**):

جدول 5–1 : بناء مثلث باسكال

n \ k	0	1	2	3	...	k-1	k	...
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
...	...	...	...	...	...	...		
n-1	1					C_{n-1}^{k-1}	C_{n-1}^{k}	
n							C_n^k	
⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	...

معادلته: $C_{n-1}^{k-1} + C_{n-1}^{k} = C_n^k$

***Source** : Y. OURAGH « Aide Mémoire de Mathématiques pour l'Ingénieur » Tome 1, OPU, 1994, Algérie, P 28.*

- **مثال (2):** أنشر $(a-2)^3$

$$(a-2)^3 = \sum_{x=0}^{3} C_3^x (-1)^x a^{3-x} 2^x$$
$$= C_3^0(-1)^0 a^{3-0}2^0 + C_3^1(-1)^1 a^{3-1}2^1 + C_3^2(-1)^2 a^{3-2}2^2 + C_3^3(-1)^3 a^{3-3}2^3$$
$$= a^3 - 3a^2 2 + 3a4 - 8$$
$$\boldsymbol{(a-2)^3 = a^3 - 6a^2 + 12a - 8}$$

3–2– استنتاج العلاقة بين العزوم: لإستنتاج العلاقة بين العزوم بكتفي بالبيانات غير المبوبة لسهولة الحساب، ولكن النتيجة صالحة للبيانات المبوبة، حيث نستخدم قانون ثنائي الحد لنيوتن لنشر العزوم المركزية كما يلي:

$$\mu_r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^r}{n}$$
$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - M_1)^r$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=0}^{r} C_r^k (-1)^{r-k}(x_i)^k M_1^{r-k}\right]$$
$$= \sum_{k=0}^{r}\left[C_r^k (-1)^{r-k} M_1^{r-k}\sum_{i=1}^{n}\frac{(x_i)^k}{n}\right]$$
$$\boxed{\mu_r = \sum_{k=0}^{r} C_r^k (-1)^{r-k} M_1^{r-k} M_k}$$

وهذه الأخير هي العلاقة الأساسية بين العزم المركزي الرائي والعزوم الإبتدائية. يكفي تعويض درجة العزم المركزي في العلاقة لإستخراجه، ونكتفي فيما يلي بالعزوم الخمسة الأولى (العزم المركزي من الدرجة 0، 1، 2، 3، 4):

$$\mu_0 = C_0^0 \mathrm{M}_1^0 \mathrm{M}_0 = 1$$
$$\mu_1 = \sum_{k=0}^{1} C_1^k (-1)^{1-k} M_1^{1-k} M_k = M_1 - M_1 = 0$$
$$\mu_2 = \sum_{k=0}^{2} C_2^k (-1)^{2-k} M_1^{2-k} M_k = M_2 - M_1^2 = \sigma^2$$
$$\mu_3 = \sum_{k=0}^{3} C_3^k (-1)^{3-k} M_1^{3-k} M_k = M_3 - 3M_1 M_2 + 2M_1^3$$
$$\mu_4 = \sum_{k=0}^{4} C_4^k (-1)^{4-k} M_1^{4-k} M_k = M_4 - 4M_1 M_3 + 6M_1^2 M_2 - 3M_1^4$$
$$\vdots$$

- **من المثال السابق**: لنحسب العزمين المركزيين من الدرجتين الثالثة والرابعة، بدلالة العزوم الإبتدائية:

$$\mu_3 = M_3 - 3M_1 M_2 + 2M_1^3 = 208.8 - 3(5.4)(32.4) + 2(5.4)^3 = \mathbf{-1.152}$$
$$\mu_4 = M_4 - 4M_1 M_3 + 6M_1^2 M_2 - 3M_1^4$$
$$= 141.6 - 4(5.4)(208.8) + 6(5.4\ )^2(32.4) - 3(5.4\ )^4$$
$$= \mathbf{23.7072}$$

وهما نفس النتيجتين المتحصل عليهما بالطريقة العادية (أنظر بند العزوم المركزية)

المبحث الثاني

مقـاييـــس الإلـتـــــواء

Mesures de Torsion

تدخل مقاييس الشكل Caractéristiques de Forme بصفة عامة، كبعد آخر لتوضيح الفكرة حول الطريقة المثلى لتحليل البيانات واستخلاص النتائج، فمثلا تعتبر مقاييس الإلتواء الكاشف الأكثر مصداقية لوجود القيم الشاذة وذلك عند وجود التواء سواء من ناحية اليسار أو اليمين فهذا يدل على أن هناك فرصة كبيرة لوجود القيم الشاذة، وهذا ما يؤثر على قيمة المتوسط الحسابي والإنحراف المعياري، لذا ففي هذه الحالة يفضل اللجوء إلى مقاييس أخرى للنزعة المركزية والتشتت، بدلا من هذين المقياسين[8]، وفيما يلي نعرض أهم مقاييس الإلتواء.

حيث نقول أن البيانات المأخوذة من دراسة معينة أو أن بيانات المجتمع بأكمله **متماثلة** حول الوسط الحسابي إذا كان شكل المنحنى الذي يمثل هذه القيم متماثلا، بحيث إذا قسمنا هذا المنحنى من منتصفه كان الشكلين الناتجين متطابقين تماما، والشكل المجاور يمثل منحنى متماثل حول الوسط الحسابي:

شكل 5-1 : التوزيع المتماثل



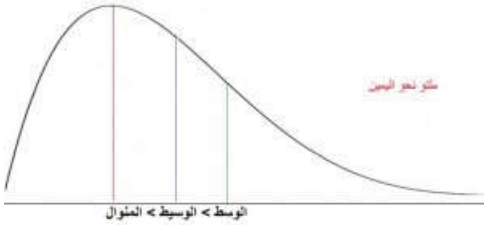
المصدر: دلال القاضي، وآخرون "الإحصاء" دار الحامد، 2005، الأردن، ص 78.

ونلاحظ أنه في حالة التوزيع المتماثل يكون: الوسط = الوسيط = المنوال، وحساب الوسط والوسيط والمنوال أن أمكن أو رسم منحنى القيم هو أفضل الطرق للتأكد من التماثل, إذ أن حساب معامل لالتواء إذا كان الناتج صفرا لا يعني تماما أن المنحنى متماثل.

أما **الالتواء** فيعرف بأنه عدم تماثل القيم حول الوسط الحسابي، بمعنى أنه ليس لدينا نقطة تماثل لتقسم المنحى إلى قسمين متطابقين كما في الحالة السابقة. ويوجد حالتين للالتواء هما:

الأولى : أن يكون المنحنى ملتوي نحو اليمين (موجب الالتواء) وهي الحالة التي يكون ذيل المنحنى ممتد نحو يمين مركز النهاية العظمى عنه إلى يسارها، وفي هذه الحالة يكون الوسط <الوسيط <المنوال , والشكل المجاور يمثل منحنى ملتوي نحو اليمين:

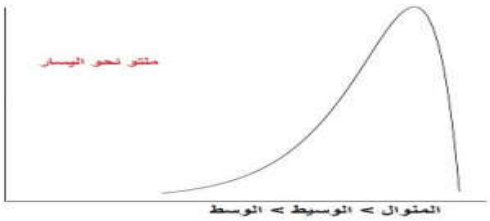
شكل 5-2 : توزيع ملتوي يمينا



المصدر: دلال القاضي، وآخرون "الإحصاء" دار الحامد، 2005، الأردن، ص 79.

الثانية : أن يكون المنحنى ملتوي نحو اليسار (سالب الالتواء) وهي الحالة التي يكون ذيل المنحنى ممتد نحو يسار النهاية العظمة منه إلى يمينها، وفي هذه الحالة يكون الوسط >الوسيط >المنوال، والشكل المجاور يمثل منحنى ملتوي نحو اليسار.

شكل 5-3 : توزيع ملتوي يسارا



المصدر: دلال القاضي، وآخرون "الإحصاء" دار الحامد، 2005، الأردن، ص 79.

وهنالك عدة طرق لحساب معامل الالتواء سنتطرق لها بعد حين، مع التركيز على القاعدة التالية: إذا كان المنحنى متماثل فان معامل الالتواء يساوي صفر، أما إذا كان معامل الالتواء يساوي صفر فهذا لا يعني بالضرورة أن المنحنى متماثل[9].

1- معامل الالتواء لفيشر: Fisher*

يعتبر مقياس الالتواء لـ"فيشر" من بين أهم مقاييس الالتواء، نظرا لأنه يعتمد على العزوم في حسابه، وله قيمة حرجة واحدة، ويعطى بالعلاقة التالية[10]:

$$F = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\mu_3}{\sigma^3}$$

إن القيمة الحرجة لهذا المقياس هي الصفر (0)، وهي القيمة التي يغير فيها التوزيع من سلوكه إما نحو اليسار أو نحو اليمين، ويمكن مناقشة المجالات التي يمكن أن يأخذها هذا المعيار فيما يلي:

* رونالد إيلمر فيشر (Ronald Aylmer FISHER) ، رياضي إنجليزي (1890-1962)

1) F<0: التوزيع التكراري يكون ملتويا نحو اليسار (سالب الالتواء)، وتكون القيم مشتتة بصفة اكبر على يسار المتوسط الحسابي، ومركزة من جهة اليمين.

2) F=0: التوزيع التكراري يكون متناظرا، وتكون القيم مشتتة بصفة متساوية على يسار ويمين المتوسط الحسابي.

3) F>0: التوزيع التكراري يكون ملتويا نحو اليمين (موجب الالتواء)، وتكون القيم مشتتة بصفة اكبر على يمين المتوسط الحسابي، ومركزة من جهة اليسار.

- **ملاحظة:** تبقى نسبة التناظر والالتواء مرتبطة بمدى البعد أو القرب من القيمة صفر "0"، حيث يمكن مقارنة توزيعين أو أكثر والقول أن احدهما أكثر التواء من الأخر أو اقل منه، وذلك بمقارنة معامل فيشر (الأكبر يعني الأكثر التواء).

شكل 5-4 : حالات الإلتواء لمقياس فيشر



F>0: توزيع ملتوي يمينا (موجب الإلتواء) | **F=0:** توزيع متناظر | **F<0:** توزيع ملتوي يسارا (سالب الإلتواء)

المصدر: دلال القاضي، وآخرون "الإحصاء" دار الحامد، 2005، الأردن، ص 44.

- **من المثال السابق:** أوجد شكل التوزيع من حيث الالتواء، بإستخدام معامل فيشر:

$$F = \frac{\mu_3}{\sigma^3} = \frac{-1.152}{\sqrt{3.24}^3} = \frac{-1.152}{5.832} = -0.197 \ < 0$$

ومنه التوزيع ملتوي يسارا.

توجد مقاييس أخرى للإلتواء، نتناولها فيما يلي:

2- معامل بيرسون* 1 للإلتواء: ويحسب وفق العلاقة التالية[11]:

$$\mathcal{P}_1 = \frac{\bar{X} - Mo}{\sigma} \quad \Rightarrow \quad \begin{cases} \mathcal{P}_1 > 0 & : \text{توزيع ملتوي يمينا} \\ \mathcal{P}_1 = 0 & : \text{توزيع متماثل} \\ \mathcal{P}_1 < 0 & : \text{توزيع ملتوي يسارا} \end{cases}$$

يقع هذا المقياس في حدود المجال التالي (حيث يساوي الصفر إذا تطابق المتوسطين الداخلين في حسابه).

$$\mathcal{P}_1 \in [0 \mp 1]$$

* كارل بيرسون (Karl REARSON) ، رياضي إنجليزي (1857- 1936)

يستخدم هذا المقياس إذا كان المتوسط الحسابي يمثل نقطة التمركز في التوزيع، وما يعاب عليه هو أنه يعتمد على المنوال وهو مقياس غير دقيق لا يجب الإعتماد عليه في قياس الإلتواء وعليه يفضل استخدام العلاقة 2 في حالة التوزيعات القريبة من التماثل.

- **من المثال السابق:** أوجد شكل التوزيع، من حيث الالتواء بإستخدام معامل Pearson (1)

الفئات	n_i
3-1	1
5-3	3
7-5	4
9-7	2
المجموع	10

تحديد المنوال:

$$\text{Mo} = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C = 5 + \frac{1}{1+2} 2 = 5.66$$

$$\mathcal{P}_1 = \frac{\bar{X} - Mo}{\sigma} = \frac{5.4 - 5.67}{\sqrt{3.24}} = \frac{-0.27}{1.8} = -0.15 \ < 0$$

ومنه التوزيع ملتوي يسارا، وهذا يساند المقياس السابق.

3- معامل بيرسون 2 للإلتواء: كما أشرنا سابقا يفضل استخدام هذه العلاقة في حالة التوزيعات القريبة من التماثل، ونكتب[12]:

$$\mathcal{P}_2 = \frac{3\,(\bar{X} - Me)}{\sigma} \quad \Rightarrow \quad \begin{cases} \text{توزيع ملتوي يمينا} : \mathcal{P}_2 > 0 \\ \text{توزيع متماثل} \quad : \mathcal{P}_2 = 0 \\ \text{توزيع ملتوي يسارا} : \mathcal{P}_2 < 0 \end{cases}$$

يقع هذا المقياس في حدود المجال التالي (حيث يساوي الصفر إذا تطابق المتوسطين الداخلين في حسابه)[13]:

$$\mathcal{P}_2 \in [0 \mp 1]$$

- **من المثال السابق:** أوجد شكل التوزيع من حيث الالتواء بإستخدام معامل Pearson (2)

الفئات	n_i	$n_i \nearrow$
3-1	1	1
5-3	3	4
7-5	4	8
9-7	2	10
المجموع	10	/

تحديد الوسيط:

$$\frac{\sum_{i=1}^{4} n_i}{2} = \frac{10}{2} = 5$$

$$Me = A + \frac{\frac{\sum_{i=1}^{4} n_i}{2} - n_{Me-1}\nearrow}{n_{Me}} \times C = 5 + \frac{5-4}{4}2 = 5.5$$

حساب المعامل:

$$\mathcal{P}_2 = \frac{3\,(\bar{X} - Me)}{\sigma} = \frac{3(5.4 - 5.5)}{1.8} = -0.16 \ < 0$$

ومنه التوزيع ملتوي يسارا، وهذا يساند المقياسين السابقين.

4- معامل بيرسون 3 للإلتواء:

نظرا لطبيعة هذا المقياس فهو دائما موجب، لأنه حصيلة قسمة العزم المركزي الثالث مربعا، على مكعب التباين وهذا الأخير موجب دائما، لذلك يستخدم هذا المقياس لمعرفة مدى التناظر للتوزيع (عندما يساوي الصفر، نقول أن التوزيع متناظر)، وعند كشفه للإلتواء لا يمكن معرفة إتجاهه (يسارا أو يمينا)، ونكتب[14]:

$$\mathcal{P}_3 = \frac{\mu_3^2}{\mu_2^3} \quad \Rightarrow \quad \left| \begin{array}{l} \text{توزيع ملتوي} : \mathcal{P}_3 > 0 \\ \text{توزيع متماثل} : \mathcal{P}_3 = 0 \end{array} \right.$$

- **من المثال السابق:** أوجد شكل التوزيع من حيث الالتواء بإستخدام معامل (3) Pearson

$$\mathcal{P}_3 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-1.152)^2}{3.24^3} = 0.04 \ > 0$$

إذن التوزيع غير متناظر، ولكنه قليل الالتواء.

5- معامل الالتواء الربيعي لـ "يول[*] و كندال[**]": Yule et Kendall ، يعتمد أساسا في حسابه على الربيعيات وبالتالي فهو يصلح في حالة التوزيعات المفتوحة، ويحسب كمايلي[15]:

$$C_{Y.K.} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Q_1)} \quad \Rightarrow \quad \left| \begin{array}{l} \text{توزيع ملتوي يمينا} : C_{YK} > 0 \\ \text{توزيع متماثل} \quad : C_{YK} = 0 \\ \text{توزيع ملتوي يسارا} : C_{YK} < 0 \end{array} \right.$$

$$C_{Y.K.} \in [0 \mp 1]$$

- **من المثال السابق:** أوجد شكل التوزيع من حيث الالتواء بإستخدام معامل Yule et Kendall

[*] جورج يول (George Yule) ، رياضي بريطاني (1871-1951)

[**] موريس كندال (Maurice Kendall)، رياضي بريطاني (1907-1983)

الوسيط تم حسابه من قيل، حساب كلا من الربيعين الأول والثالث:

$n_i \nearrow$	n_i	الفئات
1	1	3-1
4	3	5-3
8	4	7-5
10	2	9-7
/	10	المجموع

الوسيط تم حسابه سابقا:

$$Me = 5.5$$

تحديد الربيعي الأول:

$$\frac{\sum_{i=1}^{4} n_i}{4} = \frac{10}{4} = 2.5$$

$$Q_1 = A + \frac{\frac{\sum_{i=1}^{4} n_i}{4} - n_{Q_1-1}\nearrow}{n_{Q_1}} \times C = 3 + \frac{2.5-1}{3}2 = 4$$

تحديد الربيعي الثالث:

$$\frac{3\sum_{i=1}^{4} n_i}{4} = 3\frac{10}{4} = 7.5$$

$$Q_3 = A + \frac{\frac{3\sum_{i=1}^{4} n_i}{4} - n_{Q_3-1}\nearrow}{n_{Q_3}} \times C = 5 + \frac{7.5-4}{4}2 = 6.75$$

التعويض:

$$C_{Y.K.} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Q_1)} = \frac{(6.75 - 5.5) - (5.5 - 4)}{(6.75 - 4)} = \frac{-0.25}{2.75} = -0.09 \ < 0$$

ومنه التوزيع ملتوي يسارا، وهذا يساند المقاييس السابقة.

6- معامل الالتواء المئيني: ويعتمد في حسابه على المئينات وله أهمية كبيرة خاصة في التوزيعات المفتوحة كما يلي[16]:

$$P = \frac{(P_{90} - Me) - (Me - P_{10})}{(P_{90} - P_{10})} \quad \Rightarrow \quad \begin{cases} P > 0 : \text{توزيع ملتوي يمينا} \\ P = 0 : \text{توزيع متماثل} \\ P < 0 : \text{توزيع ملتوي يسارا} \end{cases}$$

- **من المثال السابق:** أوجد شكل التوزيع من حيث التطاول بإستخدام معامل الالتواء المئيني:

الفئات	n_i	$n_{\text{i}}\nearrow$
3-1	1	1
5-3	3	4
7-5	4	8
9-7	2	10
المجموع	10	/

استخراج المئين العاشر:

$$\frac{10\sum_{i=1}^{4} n_i}{100} = 10\frac{10}{100} = 1$$

$$P_{10} = A + \frac{\frac{10\sum_{i=1}^{4} n_i}{100} - n_{P_{10}-1}\nearrow}{n_{P_{10}}} \times C = 1 + \frac{1-0}{1}2 = 3$$

استخراج المئين التسعون:

$$\frac{90\sum_{i=1}^{4} n_i}{100} = 90\frac{10}{100} = 9$$

$$P_{90} = A + \frac{\frac{90\sum_{i=1}^{4} n_i}{100} - n_{P_{90}-1}\nearrow}{n_{P_{90}}} \times C = 7 + \frac{9-8}{2}2 = 8$$

التعويض:

$$P = \frac{(Q_3 - Me) - (Me - Q_1)}{(P_{90} - P_{10})} = \frac{(7-5.5)-(5.5-3)}{8-3} = \frac{-1}{5} = -0.2 \quad < 0$$

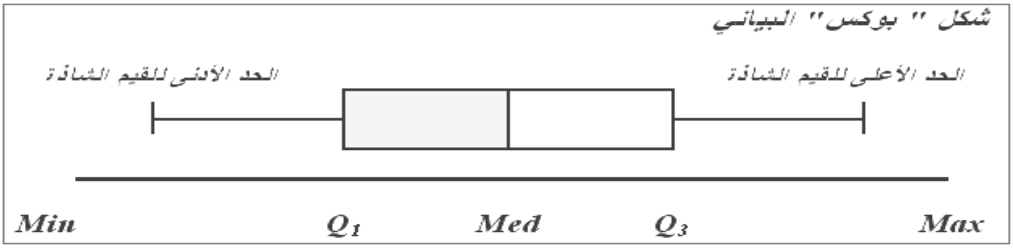
ومنه التوزيع ملتوي يسارا، وهذا يساند المقاييس السابقة.

7- التمثيل الصندوقي لـ "توكي: Tuky" : Boîtes à Moustaches [*]

رأينا فيما سبق مختلف التمثيلات البيانية وذلك في الفصل الثاني، غير أننا أجلنا هذا النوع من التمثيلات نظرا لإعتماده على الوسيط شبيهاته المشروحين في الفصل الثالث، وإعتماده أيضا على مفهوم المدى بأنواعه (المطلق والربيعي) وهذا ما تناولناه في الفصل الرابع، هذا من جهة ومن جهة أخرى لأهميته في معرفة إلتواء أو تماثل البيانات وهو موضوع هذا الفصل عموما والمبحث الثاني خصوصا.

[*] بالإنجليزية: Box Plots، وصاحب هذا التمثيل "جون توكي": TUKEY John (1915-2000)

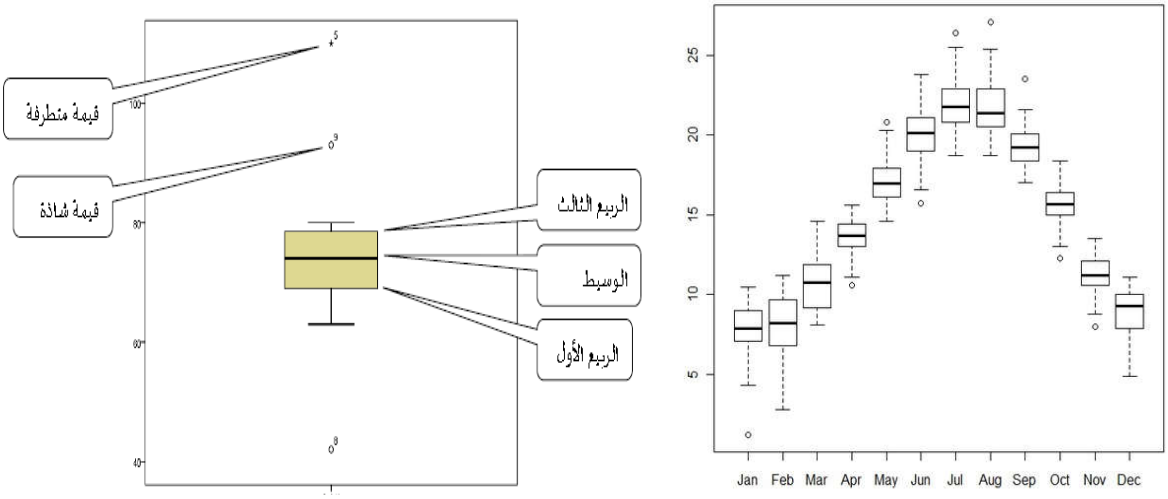
شكل 5-5 : الشكل الصندوقي أفقيا



المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 24.

تكون الاتجاهات واضحة عند التمثيل بالطريقة الأفقية لكن غالبا ما يتم رسم التمثيل الصندوقي في معلم خاصة عند مقارنة مجموعات من البيانات، وتظهر بطريقة عمودية بالشكل التالي:

شكل 5-6 : الشكل الصندوقي عموديا

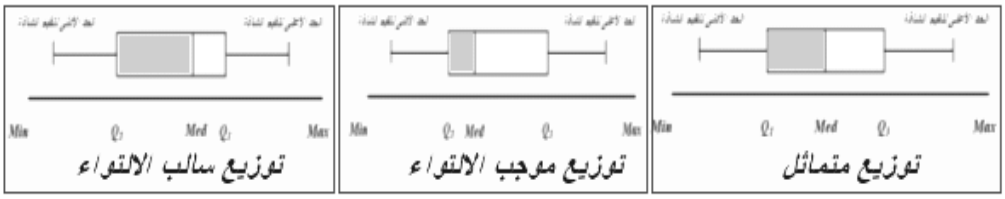


المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 24.

ويمكن استخدام الشكل البياني، أعلاه في وصف البيانات من حيث الآتي:

أ. **من حيث التماثل**: إذا كان الوسيط Med يقع في المنتصف على بعد متساوي من Q_3، Q_1 كان التوزيع متماثلا، وإذا كان الوسيط أقرب إلى الرباعي الأول من الرباعي الثالث كان التوزيع موجب الالتواء، وأما إذا كان الوسيط أقرب إلى الرباعي الثالث من الرباعي الأول كان التوزيع سالب الالتواء. ويظهر ذلك كما في الشكل التالي:

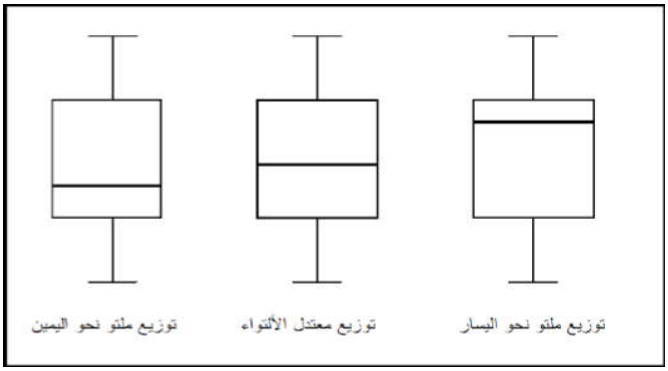
شكل 5-7 : حالات الالتواء في الشكل الصندوقي (أفقيا)



المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 24.

وهذا بالنسبة للتمثيل الأفقي للتوضيح فقط، وإذا كنا بصدد شكل عمودي فيتم التفسير كما يلي:

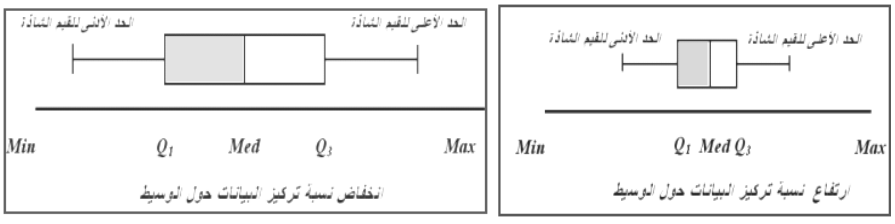
شكل 5-8 : حالات الالتواء في الشكل الصندوقي (عموديا)



المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 24.

ب. **من حيث تركز البيانات:** إذا كان الصندوق* بالنسبو للإستطالات الخارجة منه ضيق، دل ذلك على تركز نسبة كبيرة من البيانات حول الوسيط، وإذا كان الصندوق واسع دل ذلك على انخفاض نسبة تركز البيانات حول الوسيط، والشكل التالي يبين ذلك:

شكل 5-9 : وصف تركز البيانات في الشكل الصندوقي

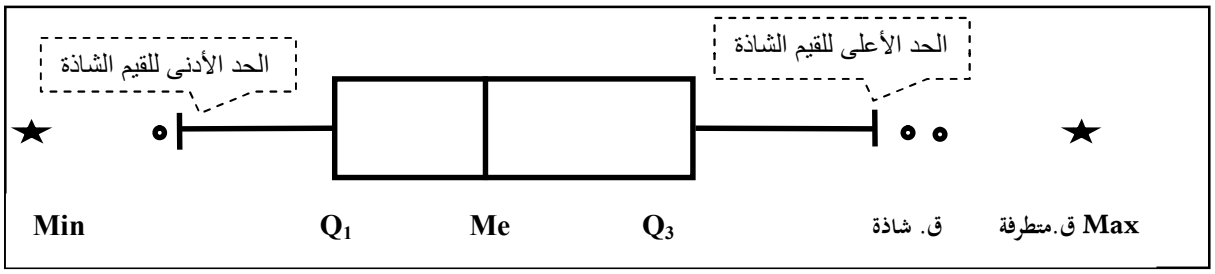


المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 22.

* بالإنجليزية: Box

ج. **من حيث وجود قيم شاذة ومتطرفة:** إذا وقعت قيم بعض المشاهدات خارج الحدين الأدنى والأعلى الشاذ (خارج الاستطالتين) أي انها تقع خارج حدي الصندوق بمسافة أكثر من مرة ونصف المدى الربيعي ولا تزيد عن ثلاثة أضعاف المدى الربيعي، كانت هذه القيم شاذة، وتظهر هذه القيم على الرسم في شكل دوائر صغيرة (°) وتمثل على استقامة واحدة مع الاستطالتين، والقيم الواقعة على مساقة تزيد عن ثلاثة أضعاف المدى الربيعي تكون قيما متطرفة ويشار إليها بنجمة (*)، كذلك تمثل على استقامة واحدة مع الاستطالتين والشكل التالي يبين طريقة عرض القيم الشاذة الدنيا والعليا وكذا القيم المتطرفة من الجهتين إن وجدت، على الرسم [17].

شكل 5-10 : تحديد القيم الشاذة والمتطرفة باستخدام التمثيل الصندوقي



المصدر: غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام **IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 23.

ونلاحظ أن القيم الشاذة والمتطرفة ليس بينهما أي حاجز بل يتم التفرقة بينهما بالشكل فقط، خاصة وإن القيم المتطرفة نادرا ما توجد.

- **طريقة حساب حدي القيم الشاذة (مسافة الإستطالتين)**: لحساب الحدين الأعلى والأدنى للقيم الشاذة، أو ما يعرف بالإستطالتين للصندوق، حيث يكون نهاية الاستطالة هو بداية وجود القيم الشاذة، نتبع الخطوات:

1- حساب المدى الربيعي: $IQ = Q_3 - Q_1$

2- حساب الاستطالة اليسرى (السفلى) أو الحد الأدنى للقيم الشاذة: $Low = Q_1 - 1.5 \times IQ$

3- حساب الإستطالة اليمنى (العلوية) أو الحد الأعلى للقيم الشاذة: $Upp = Q_3 + 1.5 \times IQ$

وإذا وقعت قيم خارج الحدين تعتبر هذه القيم من القيم الشاذة، بحيث لا تتجاوز هذه القسم الحدين التاليين:

- بالنسبة للاستطالة اليسرى: $Q_1 - 3 \times IQ$
- بالنسبة للإستطالة اليمنى: $Q_3 + 3 \times IQ$

مع مراعاة عند الرسم أن الإستطالتين لا يمكن أن تخرج على مجال البيانات حيث إذا كانت تخرجان على مجال البيانات سواء القيمة الكبرى أو الصغرى، يتم التوقف عند القيمة التي تتوقف عندها البيانات الأصلية سواء (x_{max} أو x_{min})

4- إذا تجاوزت القيم المجالين السابقين تعتبر قيما متطرفة، وغالبا ما تكون هذه القيم قيما غير صحيحة يطلب من الباحث مراجعة صحتها قبل أي عمل تحليل إحصائي.

إذن القيم x الشاذة هي:

$$\{Q_1 - 3 \times IQ < x < Q_1 - 1.5 \times IQ\} \wedge \{Q_3 + 1.5 \times IQ < x < Q_3 + 3 \times IQ\}$$

والقيم x المتطرفة تكون بنهاية مجال القيم الشاذة، وهي:

$$\{x < Q_1 - 3 \times IQ\} \wedge \{x > Q_3 + 3 \times IQ\}$$

- **مثال تطبيقي:**

فيما يلي الإنفاق الاستهلاكي بالألف دينار خلال الشهر لعينة حجمها 12 أسرة:

6 10 25 3 9 10 5 6 18 8 2 7

والمطلوب:

1- رسم الشكل الصندوقي البياني

2- اكتب تحليل وصفي لهذه البيانات.

الحـل

1- رسم الشكل الصندوقي البياني

✓ ترتيب القيم تصاعديا .

2 3 5 6 6 7 8 9 10 10 **18 25**

✓ تحديد أقل وأعلى إنفاق استهلاكي، وحساب الرباعيات:

$x_{min}=2$; $x_{max}=25$

الربيعي الأدنى Q_1 :

موقع الربيعي: $\frac{n+1}{4} = \frac{13}{4} = 3.25$

$$Q_1 = 5 + 0.25(6 - 5) = 5.25$$

الوسيط Med:

$$\frac{7 + 8}{2} = 7.5$$

الربيعي الأعلى Q_3:

موقع الرباعي: $\frac{3(n+1)}{4} = \frac{3(13)}{4} = 9.75$

$$Q_3 = 10 + 0.75(10 - 10) = 10$$

✓ حساب الحدين الأعلى والأدنى الشاذين:

الحد الأدنى للقيم الشاذة:

$$Low = Q_1 - 1.5 \times IQ = 5.25 - 1.5(10 - 5.25) = -1.875$$

الحد الأعلى للقيم الشاذة :

$$Upp = Q_3 + 1.5 \times IQ = 10 + 1.5(10 - 5.25) = 17.125$$

ومنه توجد قيمة شاذة هي 18.

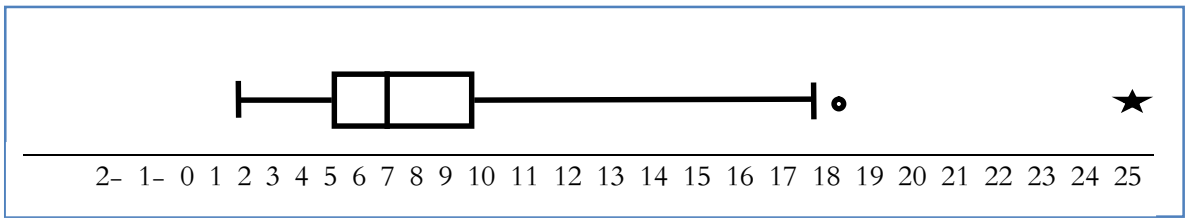
✓ تحديد القيم المتطرفة إن وجدت:

$$5.25 - 3(10 - 5.25) = -9$$
$$10 + 3(10 - 5.25) = 24.25$$

ومنه توجد قيمة متطرفة هي 25.

2- رسم التمثيل الصندوقي:

العنوان: التمثيل الصندوقي للإنفاق الاستهلاكي لعينة من الأسر



المصدر: يكتب مصدر البيانات والبرمجية التي اعتمدت في الرسم إن وجدت

ونلاحظ أنه الإستطالتين لا يمكن أن تخرج على مجال البيانات، حيث وبالرغم من وجود المجال الأيسر للاستطالة بالسالب إلا أننا توقفنا عند القيمة الدنيا (2).

3- تحليل وصفي من خلال الشكل أعلاه:

✓ درجة التماثل : التوزيع قريب جدا من التماثل لوقوع الوسيط في المنتصف .

✓ تركز البيانات: حوالي 60% من القيم تتركز حول الوسيط.

✓ القيم الشاذة: توجد قيمة شاذة عليا هي القيمة **18**.

✓ القيم المتطرفة: توجد قيمة متطرفة عليا هي 25.

ويمكن استخدام شكل "بوكس" البياني لمقارنة مجموعتين أو أكثر .

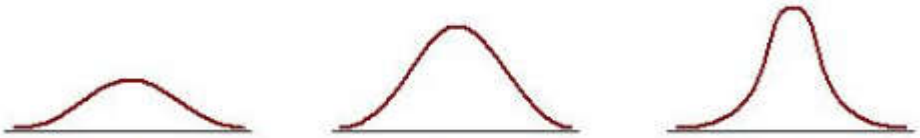
المبحث الثالث

مقاييس التفلطـح (التطاول، الإنبساط)

Mesures d'Aplatissement

عند تمثيل التوزيع التكراري في شكل منحنى تكراري، قد يكون هذا المنحنى منبسط، أو مدبب، فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى، ويقل في طرفيه، يكون المنحنى مدببا، وعندما يتركز عدد أكبر على طرفي المنحنى، ويقل بالقرب من المنتصف يكون المنحنى مفرطحا، أو منبسطا، ويظهر ذلك من الشكل التالي:

شكل 5-11 : حالات التطاول في التوزيع التكراري



توزيع متطاول (مدبب) | توزيع معتدل | توزيع مفلطح (مفلطح، منبسط، مسطح)

المصدر: دلال القاضي، وآخرون "الإحصاء" دار الحامد، 2005، الأردن، ص 44.

ويمكن قياس التفرطح باستخدام عدد من الطرق، كل منها له ايجابياته وسلبياته، نسردها فيما يلي:

1- معامل التفلطح لكارل بيرسون Karl Pearson

يعتبر مقياس التفلطح لـ"بيرسون" من بين أهم مقاييس الالتواء، نظرا لأنه يعتمد على العزوم في حسابه، ويعطى بالعلاقة التالية[18]:

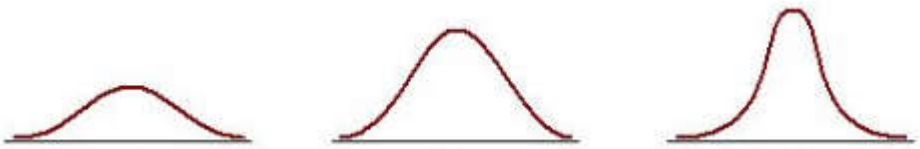
$$K = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$$

إن القيمة الحرجة لهذا المقياس هي ثلاثة (3)، وهي القيمة التي يغير فيها التوزيع من سلوكه إما تطاولا أو إنبساطا، ويمكن مناقشة المجالات التي يمكن أن يأخذها هذا المعيار فيما يلي:

1) K<3: التوزيع التكراري يكون منبسطا، أي تشتت كبير، وغير مركز حول المتوسط الحسابي.

2) K=3: يكون التوزيع التكراري في هذه الحالة معتدلا.

3) K>3: التوزيع التكراري يأخذ شكلا متطاولا، وبالتالي تكون التكرارات مركزة حول المتوسط الحسابي، مما يقلل من التشتت.



K>3: توزيع متطاول | **K=3:** توزيع معتدل | **K<3:** توزيع منبسط

- **من المثال السابق (بيانات مبوبة)**: أوجد شكل التوزيع من حيث التطاول بإستخدام معامل بيرسون:

$$K = \frac{\mu_4}{\mu_2^2} = \frac{23.7072}{(3.24)^2} = 2.25 < 3$$

ومنه التوزيع منبسط بشكل بسيط جدا (لا يبتعد كثيرا عن 3).

هنالك مقاييس أخرى للتفلطح، نذكرها فيمايلي:

2- معامل فيشر للتفلطح:

له نفس مبدأ الحساب بالنسبة للمعامل السابق (K)، يتم تركيز القيمة الحرجة فقط بإنقاص 3 من المعامل لتصبح المقارنة حول الصفر[19]:

$$F_2 = K - 3 \quad ; \begin{cases} \text{توزيع متطاول} : F_2 > 0 \\ \text{توزيع معتدل} : F_2 = 0 \\ \text{توزيع منبسط} : F_2 < 0 \end{cases}$$

- **من المثال السابق**: أوجد شكل التوزيع من حيث التطاول بإستخدام معامل فيشر (2):

$$K = F - 3 = 2.25 - 3 = -0.75 < 0$$

ومنه التوزيع منبسط.

3- معامل التفلطح المئيني:

$$\beta = \frac{(Q_3 - Q_1)}{2\,(P_{90} - P_{10})} \quad \Rightarrow \quad \begin{cases} \text{توزيع متطاول} : \beta > 0.263 \\ \text{توزيع معتدل} : \beta = 0.263 \\ \text{توزيع منبسط} : \beta < 0.263 \end{cases}$$

هذا المقياس له أهمية كبيرة في التوزيعات المفتوحة، لأنه يعتمد على شبيهات الوسيط (لا تتأثر بالقيم المتطرفة)[20].

- **من المثال السابق (بيانات مبوبة):** أوجد شكل التوزيع من حيث التطاول بإستخدام معامل التفلطح المئوي، ثم مثل بيانيا التوزيع التكراري:

التعويض:

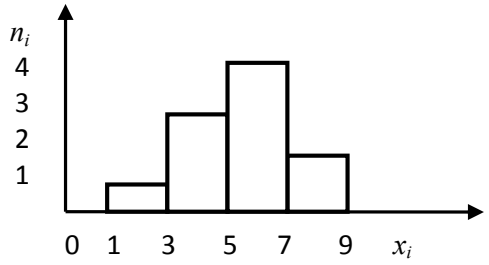
$$\beta = \frac{(Q_3 - Q_1)}{2\,(P_{90} - P_{10})} = \frac{6.75 - 4}{2(8-3)} = \frac{2.75}{10} = 0.275 \;\; > 0.263$$

نلاحظ أن النتيجة تظهر ا، التوزيع يقترب من الإعتدال مع تدبب بسيط جدا، هي نتيجة مقاربة للعلاقات السابقة لكن أظهرت لنا الإعتدال باتجاه التدبب وليس الانبساط.

- **التمثيل البياني:**

بعد حساب مختلف المعاملات الخاصة بالشكل، يمكن التحقق من التواء وتفرطح التوزيع التكراري من خلال التمثيل البياني عن طريق المدرج التكراري (لأنه متغير كمي مستمر) لنفس المثال:

n_i	الفئات
1	3-1
3	5-3
4	7-5
2	9-7



نلاحظ من خلال الشكل أن التوزيع ملتوي يسارا (غير متناظر من ناحية اليسار) أي أن المتغيرة مركزة على يمين المتوسط الحسابي وأقرب إليه، أما من ناحية اليسار فهي مشتتة بطريقة أبعد من الوسط الحسابي، كما أنه منبسط، أي أن قيم المتغيرة غير مركزة عند المتوسط الحسابي في العموم بل موزعة على الجانبين ولكن بصفة اقل وللملاحظة شكل التوزيع ننوه إلى ضرورة الإنتباه لمقياس الرسم على المحورين، حيث أن إختيار المقياس غير المناسب يظللنا لمعرفة شكل التوزيع من حيث التفلطح. ويكون دراسة شكل التوزيع أكثر وضوحا وفعالية عند مقارنة أكثر من توزيع، خاصة وأن مقاييس الشكل معزولة عن وحدة القياس (لا وحدة لها).

تمارين

- **تمرين 1** : أجب على الأسئلة التالية:

1- ما هو مفهوم مقاييس الشكل وما هي أنواعا.

2- ما هو دور العزوم في مقارنة البيانات.

- **تمرين 2**: لتكن سلسلة البيانات التالية:

4 5 8 9 10 11 6 6 3 2 3 3 2

1- أوجد العزوم الإيتدائية الثلاثة الأولى، والعزم الإبتدائي الصفري.

2- أحسب العزوم اللامركزية الأربعة الأولى بالإضافة غلى العزم اللامركزي الصفري.

3- أوجد شكل التوزيع من حيث الإلتواء والتفلطح.

4- ارسم بيانيا هذه البيانات.

- **تمرين 3**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق
]20-10]	10
]30-20]	15
]40-30]	12
[50-40]	8

1- أوجد العزوم الإيتدائية (اللامركزية) الثلاثة الأولى، والعزم الإبتدائي الصفري.

2- أحسب العزوم المركزية الأربعة الأولى بالإضافة غلى العزم اللامركزي الصفري.

3- أوجد شكل التوزيع من حيث الإلتواء والتفلطح.

4- ارسم بيانيا هذه البيانات.

- **تمرين 4**: ليكن التوزيع التكراري التالي:

فئات المتغيرة	التكرار المطلق n_i
]20-10]	8
]30-20]	10
]60-30]	12
]70-60]	2
[75-70]	1
المجموع	33

1- أوجد العزوم الإيتدائية الثلاثة الأولى، والعزم الإبتدائي الصفري.

2- أحسب العزوم اللامركزية الأربعة الأولى بالإضافة غلى العزم اللامركزي الصفري.

3- أوجد شكل التوزيع من حيث الإلتواء والتفلطح.

4- ارسم بيانيا هذه البيانات.

- **تمرين 5**: لتكن البيانات التالية الخاصة بنقاط 56 طالب في مقياس الإقتصاد الجزئي:

12 16 08 08 17 14 06 11 11 05 16 03 05 10 11 07 08 12 10
16 08 08 17 14 06 11 11 05 16 03 05 10 11 07 08 12 10 19
19 12 16 08 08 17 14 06 19 12 16 08 08 17 14 06 19 12

- المطلوب:

1- قم بتبويب البيانات بأحد الطرق (يول، ستورج)

2- أحسب جميع مقاييس النزعة المركزية

3- أحسب جميع مقاييس التشتت المدروسة.

4- حدد شكل التوزيع، مستخدما الطرق الممكنة لذلك.

5- مثل بيانيا التوزيع التكراري، ثم قارنه بمقاييس الشكل المحسوبة.

- **مراجع وإحالات الفصل الخامس:**

[1] موراي شبيجل "نظريات ومسائل في الإحصاء" ترجمة: أحمد حسن الموازيني، الدار الدولية للنشر والتوزيع، 1988، مصر، ص139.
[2] المرجع السابق، ص 140.
[3] مديحة السيد محمد موسى "أساسيات الإحصاء الرياضي وتطبيقاتها" دار الكتاب الحديث، 2009، مصر، ص 66.
[4] موراي شبيجل، ص 139.
[5] المرجع السابق، ص 140.
[6] M. R. SPIEGEL, S. LIPSCUTZ, J. LIU « Mathematical Handbook of Formulas and Tables » SCHUM'S Outlines, Mc Grow Hill, 3rd Ed., USA, P 8.
[7] S. TOURI « Algèbre » EDICEF, 1991, France, P 44.
[8] موراي شبيجل، جون شيلر، ألو سريبيقاسان "الإحصاء والاحتمالات" ملخصات شوم إيزي، ترجمة: محمود علي أبو النصر، مصطفى جلال مصطفى، الدار الدولية للإستثمارات الثقافية، 2003، مصر، ص 29.
[9] عبد الله الجديد "ماذا نعني بالإلتواء" [على الخط] يوم: 2017/12/26، الرابط:
https://www.statinkitchen.com/?p=105
[10] مجدي الطويل "الإحتمالات" دار النشر للجامعات، 2009، مصر، ص 347.
[11] H. HAMDANI « Statistique Descriptive » OPU, 6eme éd., 2010, Algérie, P 141.
[12] مجدي الطويل، سبق ذكرة، ص 345.
[13] عبد الحميد عبد المجيد البلداوي "أساليب البحث العلمي والتحليل الإحصائي" دار الشروق، ط1، 2007، الأردن، ص 159.
[14] D. FREDON, M. MAUMY-BERTRAND, F. BERTRAND « Mathématiques : Statistique et Probabilités » DUNOD, 2009, France, P 118.
[15] H. HAMDANI, Op. Cit. P 141, P 142.
[16] موراي شبيجل، مرجع سابق، ص 142.
[17] غيث البحر، معن التنجي "التحليل الإحصائي للإستبيانات بإستخدام IBM SPSS Statistics" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا، ص 23.
[18] موراي شبيجل، مرجع سابق، ص 143.
[19] عبد الحفيظ مصطفى "نظرية الاحتمالات" ج 1، د. م. ج.، ط2، 2004، الجزائر، ص 256.
[20] موراي شبيجل، مرجع سابق، ص 134.

المراجع

المراجع

- **باللغة العربية:**

1- أحمد شيبات "**الإحصاء الوصفي**" دار الهدى، 2001، الجزائر.

2- أحمد عبد السميع طبية "**مبادئ الإحصاء**" دار البداية، ط 1، 2008، الأردن.

3- أماني موسى محمد "**التحليل الإحصائي للبيانات**" مركز تطوير الدراسات العليا والبحوث في العلوم الهندسية، ط 1، 2008، مصر.

4- شرف الدين خليل "**الإحصاء الوصفي**" شبكة الأبحاث والدراسات الإقتصادية، بدون سنة، بدون بلد.

5- دلال القاضي، وآخرون "**الإحصاء –للإدرايين والإقتصاديين–**" دار الحامد، 2005، الأردن

6- دومنيك سلفاتور "**الإحصاء والإقتصاد القياسي**" ترجمة : سعدية حافظ منتصر، الدار الدولية للنشر والتوزيع، مصر.

7- جلاطو جيلالي "**الإحصاء مع تمارين ومسائل محلولة**" د. م. ج، 1998، الجزائر

8- جورج كانافوس، دون ميلر "**الإحصاء للتجاريين: مدخل حديث**" ترجمة: سلطان محمد عبد الحميد، دار المريخ، 2003، م. ع. السعودية،

9- غيث البحر، معن التنجي "**التحليل الإحصائي للإستبيانات بإستخدام IBM SPSS Statistics**" مركز سبر للدراسات الإحصائية والسياسات العامة، 2014، تركيا،

10- عبد الحفيظ مصطفى "**نظرية الاحتمالات**" ج 1، د. م. ج.، ط2، 2004، الجزائر.

11- عبد اللطيف يوسف الصديقي "**تاريخ الإحصاء**" المؤسسة الجامعية للدراسات والنشر والتوزيع، ط 1، 2006، لبنان.

12- عبد الرحمن بن محمد سليمان أبو عمة، محمود محمد إبراهيم هندي "**الإحصاء التطبيقي**" العبيكان، 1995، م. ع. السعودية.

13- عبد الرحمن عيسوي "**الإحصاء السيكولوجي التطبيقي**" دار المعرفة الجامعية، 2000، مصر.

14- عبد الله فلاح المنيزل، عايش موسى غرايبة "**الإحصاء التربوي**" دار المسيرة، الأردن.

15- علاء الدين القبانجي، حسام حمامة كربجي "**الإحتمال والإحصاء**" منشورات جامعة دمشق، 2012/2011، سوريا.

16- عبد العزيز فهمي هيكل "**مبادئ الأساليب الإحصائية**" المركز الدولي لتعليم الإحصاء، ط1، 1966، لبنان.

17- عوض منصور، علي قوقزة، عزام صبري "**أساسيات علم الإحصاء الوصفي**" دار صفاء، ط 1، 1999، الأردن.

18- لورانس سي هاملتون "**الإحصاء مع برنامج STATA**" ترجمة: رمضان مفتاح الفيتوري، الفيتوري مفتاح الفيتوري، دار المريخ، 2015، م. ع. السعودية.

19- مبارك أسبر ديب "**مبادئ في الاحتمالات والإحصاء**" جامعة تشرين، كلية العلوم، 2009/2008، سوريا.

20- موراي شبيجل، جون شيلر، ألو سريبيقاسان "**الإحصاء والاحتمالات**" ملخصات شوم إيزي، ترجمة: محمود علي أبو النصر، مصطفى جلال مصطفى، الدار الدولية للإستثمارات الثقافية، 2003، مصر.

21- موراي شبيجل "**نظريات ومسائل في الإحصاء**" ترجمة: أحمد حسن الموازيني، الدار الدولية للنشر والتوزيع، 1988، مصر.

22- مديحة السيد محمد موسى "**أساسيات الإحصاء الرياضي وتطبيقاتها**" دار الكتاب الحديث، 2009، مصر.

23- مجدي الطويل "**الإحتمالات**" دار النشر للجامعات، 2009، مصر.

24- عبد الحميد عبد المجيد البلداوي "**أساليب البحث العلمي والتحليل الإحصائي**" دار الشروق، ط1، 2007، الأردن.

25- نبيل جمعة صالح النجار " **الإحصاء**" دار الحامد، ط1، 2010، الأردن.

- **باللغات الأجنبية:**

1- A. AZZOUZ « **STATISTIQUE*: cours, exercices et problème résolus*** » Tome 1, OPU, 2016, Algérie.

2- A. J. BARLOW « **Statistics** » WILEY , 1991 , USA.

3- B. L. AGARWAL **« Basic Statistics** » New Age International Publishers, 4th Ed. 2006, India.

4- D. FREDON, M. MAUMY-BERTRAND, F. BERTRAND **« Mathématiques *: Statistique et Probabilités*** » DUNOD, 2009, France.

5- F. GUENARD , P. HUG « **MATHEMATIQUES : *Algèbre et Probabilités*** » DUNOD, 1993, France.

6- H. HAMDANI « **Statistique Descriptive** » OPU, 6eme éd., 2010, Algérie.
7- I. CANNAZ « **Introduction a la Statistique** » INSA de Lyon, France.
8- K. KHALED « **Méthodes Statistiques** » OPU, 2005, Algérie.
9- M. R. SPIEGEL, S. LIPSCUTZ, J. LIU « **Mathematical Handbook of Formulas and Tables** » SCHUM'S Outlines, Mc Grow Hill, 3rd Ed., USA.
10- P. S. MANN «**Introductory Statistics** » WILEY, 7eme Ed., USA
11- S. TOURI « **Algèbre** » EDICEF, 1991, France.
12- Y. OURAGH « **Aide Mémoire de Mathématiques pour l'Ingénieur** » Tome 1, OPU, 1994, Algérie.

- **مواقع الأنترنت:**

1- جامعة الملك فيصل:

https://www.kfu.edu.sa/sites/Home

2- موسوعة البحوث والمواضيع المدرسية :

http://bohotti.blogspot.com/2014/12/blog-post_444.html

3- عبد الله الجديد "ماذا نعني بالإلتواء" [على الخط] يوم: 2017/12/26، الرابط:

https://www.statinkitchen.com/?p=105

الملاحق

- **ملحق –1– الأحرف الإغريقية**

المرتبة	الحرف الكبير	الحرف الصغير	باللاتنية	بالعربية
1	A	α	*Alpha*	ألفا
2	B	β	*Bêta*	بيتا
3	Γ	γ	*Gamma*	قاما
4	Δ	δ	*Delta*	ديلتا
5	E	ε	*Epsilon*	إيبسيلون
6	Z	ζ	*Zeta*	زيتا
7	H	η	*Eta*	إيتا
8	Θ	θ	*Thêta*	ثيتا
9	I	ι	*Iota*	يوتا
10	K	κ	*Kappa*	كبا
11	Λ	λ	*Lambda*	لمدا
12	M	μ	*Mu*	ميو
13	N	ν	*Nu*	نيو
14	Ξ	ξ	*Xi*	كسي
15	O	ο	*Omicron*	أوميكرون
16	Π	π	*Pi*	پي
17	P	ρ	*Rho*	رو
18	Σ	σ	*Sigma*	سيقما
19	T	τ	*Tau*	تو
20	Y	υ	*Upsilon*	يوبسيلون
21	Φ	ø	*Phi*	في
22	X	χ	*Chi ou Khi*	كي
23	Ψ	ψ	*Psi*	پسي
24	Ω	ω	*Omega*	أوميقا
	ϑ	φ	*Formes archaïques de "thêta" et "phi".*	*الأشكال القديمة لِ"ثيتا" و"في".*
	ς	ϖ	*ς Est la forme de "sigma" comme lettre finale.*	ς *شكل "سيقما"كحرف أخير.*

- **ملحق –2– جدول الأرقام العشوائية**

Random Number Table

13962 70992 65172 28053 02190 83634 66012 70305 66761 88344
43905 46941 72300 11641 43548 30455 07686 31840 03261 89139
00504 48658 38051 59408 16508 82979 92002 63606 41078 86326
61274 57238 47267 35303 29066 02140 60867 39847 50968 96719
43753 21159 16239 50595 62509 61207 86816 29902 23395 72640
83503 51662 21636 68192 84294 38754 84755 34053 94582 29215
36807 71420 35804 44862 23577 79551 42003 58684 09271 68396
19110 55680 18792 41487 16614 83053 00812 16749 45347 88199
82615 86984 93290 87971 60022 35415 20852 02909 99476 45568
05621 26584 36493 63013 68181 57702 49510 75304 38724 15712
06936 37293 55875 71213 83025 46063 74665 12178 10741 58362
84981 60458 16194 92403 80951 80068 47076 23310 74899 87929
66354 88441 96191 04794 14714 64749 43097 83976 83281 72038
49602 94109 36460 62353 00721 66980 82554 90270 12312 56299
78430 72391 96973 70437 97803 78683 04670 70667 58912 21883
33331 51803 15934 75807 46561 80188 78984 29317 27971 16440
62843 84445 56652 91797 45284 25842 96246 73504 21631 81223
19528 15445 77764 33446 41204 70067 33354 70680 66664 75486
16737 01887 50934 43306 75190 86997 56561 79018 34273 25196
99389 06685 45945 62000 76228 60645 87750 46329 46544 95665
36160 38196 77705 28891 12106 56281 86222 66116 39626 06080
05505 45420 44016 79662 92069 27628 50002 32540 19848 27319
85962 19758 92795 00458 71289 05884 37963 23322 73243 98185
28763 04900 54460 22083 89279 43492 00066 40857 86568 49336
42222 40446 82240 79159 44168 38213 46839 26598 29983 67645
43626 40039 51492 36488 70280 24218 14596 04744 89336 35630
97761 43444 95895 24102 07006 71923 04800 32062 41425 66862
49275 44270 52512 03951 21651 53867 73531 70073 45542 22831
15797 75134 39856 73527 78417 36208 59510 76913 22499 68467
04497 24853 43879 07613 26400 17180 18880 66083 02196 10638
95468 87411 30647 88711 01765 57688 60665 57636 36070 37285
01420 74218 71047 14401 74537 14820 45248 78007 65911 38583
74633 40171 97092 79137 30698 97915 36305 42613 87251 75608
46662 99688 59576 04887 02310 35508 69481 30300 94047 57096
10853 10393 03013 90372 89639 65800 88532 71789 59964 50681
68583 01032 67938 29733 71176 35699 10551 15091 52947 20134
75818 78982 24258 93051 02081 83890 66944 99856 87950 13952
16395 16837 00538 57133 89398 78205 72122 99655 25294 20941
53892 15105 40963 69267 85534 00533 27130 90420 72584 84576
66009 26869 91829 65078 89616 49016 14200 97469 88307 92282
45292 93427 92326 70206 15847 14302 60043 30530 57149 08642
34033 45008 41621 79437 98745 84455 66769 94729 17975 50963
13364 09937 00535 88122 47278 90758 23542 35273 67912 97670
03343 62593 93332 09921 25306 57483 98115 33460 55304 43572
46145 24476 62507 19530 41257 97919 02290 40357 38408 50031
37703 51658 17420 30593 39637 64220 45486 03698 80220 12139
12622 98083 17689 59677 56603 93316 79858 52548 67367 72416
56043 00251 70085 28067 78135 53000 18138 40564 77086 49557
43401 35924 28308 55140 07515 53854 23023 70268 80435 24269
18053 53460 32125 81357 26935 67234 78460 47833 20496 35645

- **ملحق –3– مثلث باسكال (Pascal)**

$$C_n^k = \frac{n!}{(n-k)!\,k!}\ ;\ k \le n$$

ci-dessous.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1																
1	1	1															
2	1	2	1														
3	1	3	3	1													
4	1	4	6	4	1												
5	1	5	10	10	5	1											
6	1	6	15	20	15	6	1										
7	1	7	21	35	35	21	7	1									
8	1	8	28	56	70	56	28	8	1								
9	1	9	36	84	126	126	84	36	9	1							
10	1	10	45	120	210	252	210	120	45	10	1						
11	1	11	55	165	330	462	462	330	165	55	11	1					
12	1	12	66	220	495	792	924	792	495	220	66	12	1				
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1			
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1		
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1	
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008	4368	1820	560	120	16	1