جامعة الشهيد حمة لخضر

كلية العلوم الاقتصادية وعلوم التسيير والعلوم التجارية

قسم علوم التسيير

مطبوعة دروس مقدمة لطلبة السنة الأولى جذع مشترك

بعنوان :

الإحصاء 01

من اعداد الأستاذ : ضو نصر

الموسم الجامعي 2024/2023

جامعة الشهيد حمه لخضر

كلية العلوم الاقتصادية وعلوم التسيير والعلوم التجارية

قسم علوم التسيير

مطبوعة دروس مقدمة لطلبة السنة الاولى جذع مشترك

بعنوان :

الإحصاء 01

من اعداد الأستاذ : ضو نصر

الموسم الجامعي 2024/2023

قائمة المحتويات

مختصر معلومات المقياس

السداسي : الاول

وحدة التعليم : منهجية

المادة : احصاء 1

الرصيد: 05

المعامل: 03

نمط التعليم: حضوري

أهداف التعليم

تهدف هذا المادة التعليمية إلى تزويد الطالب ببعض التقنيات الرياضية التي تستخدم في مادة الإحصاء الوصفي ففي نهاية هذا المقياس يصبح الطالب قادر على ضبط ماهية الإحصاء الوصفي ، والتعلم على مختلف المفاهيم الأساسية لمقياس الإحصاء الوصفي والقدرة على توظيف الأساليب الإحصائية المناسبة لوصف البيانات

المعارف المسبقة المطلوبة

حتى يتمكن الطالب من دراسة محتوى هذه المادة لابد أن يكون متحكما في مادة الرياضيات بالإضافة الى أساسيات الإعلام الآلي والمفاهيم الأساسية لتكنولوجيا المعلومات .

الفئة المستهدفة :

المقياس موجه بالتحديّد إلى طالب السنة الأولى جذع مشترك علوم اقتصادية وعلوم تسيير والعلوم التجارية ، ويمكن ان يكون مفيدا لاي شخص مطالب باستخدام البيانات الاحصائية في اعداد التقارير او في اطار اعداد مذكرة التخرج .

محتوى المادة:

- الفصل الاول : مدخل للإحصاء وعرض البيانات
- الفصل الثاني : مقاييس النزعة المركزية
- الفصل الثالث : مقاييس التشتت والشكل
- الفصل الرابع : الارتباط والانحدار البسيط
- الفصل الخامس : المتغيرات العشوائية ذات التوزيع الاحتمالي
- الفصل السادس : السلاسل وحلولها

طريقة التقييم: تقيم مستمر + إمتحان نهائي ويقاس معدل المادة بالوزن الترجيحي للدروس(60%) والأعمال الموجهة (40%)

قائمة المحتويات

المقدمة

مقدمة

يعدّ الاحصاء الوصفي أحد أهم الأدوات التحليلية في البحث العلمي، حيث يهدف إلى وصف وتلخيص البيانات المتاحة بطرق كمية وموضوعية. يعتبر الإحصاء الوصفي جزءًا أساسيًا من عملية التحليل البياني والاستنتاج العلمي، حيث يساعد الباحث على فهم الظواهر وتفاصيلها بشكل دقيق ومفصل.

تهدف هذه المطبوعة إلى استكشاف الاحصاء الوصفي وأهميته في البحث العلمي. سيتم تناول تعريف الاحصاء الوصفي وشرح أهم الأساليب المستخدمة فيه، بالإضافة إلى الاستخدامات الشائعة والفوائد التي يمكن أن يوفرها في سياق البحث العلمي.

سيتم التطرق في هذه المطبوعة إلى مفهوم الاحصاء الوصفي ودوره في عرض البيانات العلمية بطريقة مبسطة ومفهومة. سنتناول أيضًا الأدوات الإحصائية المستخدمة في تحليل البيانات الوصفية، مثل المتوسط والانحراف المعياري والترددات والتوزيعات الاحتمالية.

بالإضافة إلى ذلك، سنتناول الاستخدامات الشائعة للإحصاء الوصفي في البحث العلمي، مثل وصف العينة الاستكشافية وتلخيص البيانات المتعددة الأبعاد ورصد التغيرات عبر الزمن. سنسلط الضوء أيضًا على الفوائد التي يمكن أن يوفرها الاحصاء الوصفي في توجيه الأبحاث المستقبلية واتخاذ القرارات العلمية السليمة.

باختصار، يعتبر الاحصاء الوصفي أداة حاسمة في البحث العلمي لتحليل وتلخيص البيانات بطريقة دقيقة ومنهجية. يوفر الاحصاء الوصفي نظرة شاملة على الظواهر العلمية ويساعد الباحثين على فهم العلاقات والاتجاهات والتغيرات في البيانات.

فالإحصاء الوصفي (Descriptive Statistics) هو فرع من فروع الإحصاء يهتم بجمع وتلخيص البيانات ووصفها بطرق كمية ورسوم بيانية ومقاييس أخرى. يستخدم الإحصاء الوصفي لتحليل وتفسير البيانات المتاحة بغرض فهمها والكشف عن الأنماط والمعلومات الهامة المتعلقة بها.

يعمل الإحصاء الوصفي على تلخيص البيانات وتقديمها بطريقة سهلة ومفهومة. يتضمن هذا العمل تجميع البيانات، وحساب المقاييس المركزية مثل الوسط الحسابي والمتوسط الحسابي، والمقاييس

التشتتية مثل الانحراف المعياري والانحراف المتوسط المطلق، والعرض الدنيا والعرض الأقصى. كما يمكن استخدام الرسوم البيانية مثل الرسوم البيانية الشريطية والهيستوغرامات والرسوم الدائرية لتوضيح البيانات وتحليلها.

تهدف الإحصاءات الوصفية إلى توفير صورة شاملة وموجزة للبيانات، والتعرف على الخصائص الأساسية والتركيب العام للمجموعة المدروسة. يمكن استخدام الإحصاء الوصفي في العديد من المجالات مثل الأعمال التجارية، والعلوم الاجتماعية، والطب، وعلوم البيئة، والعديد من المجالات الأخرى التي تتطلب تحليل البيانات ووصفها بشكل دقيق.

يجب أن يتم تفسير الإحصاء الوصفي بحذر، حيث أنه يقدم فقط وصفًا للبيانات المتاحة ولا يسمح بالاستدلال الإحصائي بشأن العلاقات السببية أو التنبؤ بالأحداث المستقبلية.

الإحصاء الوصفي هو فرع من فروع الإحصاء الذي يهتم بتجميع وتلخيص البيانات الإحصائية ووصفها بشكل موجز ومفهوم. يعتبر الإحصاء الوصفي أساسيًا في العديد من المجالات وله أهمية كبيرة لعدة أسباب منها يساعد الإحصاء الوصفي في تلخيص البيانات الضخمة والمعقدة إلى مقاييس ومؤشرات أساسية يمكن فهمها وتفسيرها بسهولة. يعطينا الإحصاء الوصفي فكرة واضحة عن الخصائص الأساسية للبيانات مثل المتوسط، الانحراف المعياري، التوزيع الاحتمالي والتباين ، وتوفير معلومات أساسية وأرقام قياسية تعكس الخصائص المهمة للبيانات. هذه المعلومات الأساسية تساعد في فهم البيانات واتخاذ القرارات المستندة إلى الأدلة ، و يمكن استخدام الإحصاء الوصفي لوصف ومقارنة مجموعات مختلفة من البيانات. يساعد في تحليل الفروق والتشابهات بين المجموعات المختلفة وفهم العلاقات بين المتغيرات المختلفة و للكشف عن العلاقات والأنماط غير المتوقعة في البيانات. يمكن استخدام أدوات الإحصاء الوصفي مثل الرسوم البيانية والتوصيفات الإحصائية لاكتشاف الاتجاهات والتغيرات المفاجئة والقيم الشاذة.

وبشكل عام، يساهم الإحصاء الوصفي في فهم البيانات واكتشاف النماذج وتلخيص المعلومات الكمية، مما يسهم في اتخاذ القرارات الأفضل وفهم الظواهر والعمليات المختلفة في العديد من المجالات مثل العلوم الاجتماعية، والاقتصاد، والطب، والعلوم البيئية، والتسويق، وغيرها.

و يعد الإحصاء الوصفي أداة قوية لتلخيص البيانات والوصول إلى رؤى قيمة. يساعد في فهم الخصائص الرئيسية للبيانات وتوضيح العلاقات والأنماط الموجودة فيها. يمكن استخدام الإحصاء الوصفي بشكل واسع في مجالات مثل الأعمال التجارية والاقتصاد والعلوم الاجتماعية والطب والعديد من المجالات الأخرى التي تتعامل مع البيانات الكمية.

تتضمن مبادئ الإحصاء الوصفي العديد من المفاهيم والتقنيات، ومن بينها تجميع البيانات و يشمل جمع البيانات النوعية والكمية المتعلقة بالمتغيرات المدروسة. يمكن جمع البيانات من خلال مسوح أو استبيانات أو تجارب أو مصادر أخرى وتلخيص البيانات باستخدام المقاييس الإحصائية مثل المتوسط والوسيط والوضع والانحراف المعياري. يتم استخدام هذه المقاييس لفهم القيم المركزية وتوزيع البيانات و يشمل كذلك تصور البيانات استخدام الرسومات والمخططات البيانية مثل الرسوم البيانية والهيستوغرامات والدوائر الدائرية لتوضيح النماذج والاتجاهات والتباين في البيانات.

ويستخدم التحليل الوصفي لفهم العلاقات والترابطات بين المتغيرات واستنتاج المعلومات النوعية حول البيانات المدروسة. يمكن استخدام تقنيات الإحصاء الوصفي مثل الاختبارات الفرضية والانحدار البسيط والتحليل التبايني لهذا الغرض.

وبدورنا نقدم للقارئ والطالب الجامعي بصفة عامة ولطلبة العلوم الاجتماعية والاقتصادية بصفة خاصة هذا الإصدار والذي جاء بأسلوب سهل وقد احتوى كل فصل أمثلة عدة توضيحية حتى يتمكن الطالب من فهم مقياس الإحصاء الوصفي بأيسر الطرق.

وتنقسم هذا المطبوعة إلى جزئين فالأول يحتوي على خمس فصول تشرح الجانب النظري الإحصاء الوصفي مع أمثلة توضيحية كما سبق، وهي كما يلي:

الفصل الأول : مدخل للإحصاء وعرض البيانات، جاء فيه المفاهيم الأساسية لعلم الإحصاء .

الفصل الثاني : مقاييس النزعة المركزية، تناول كيفية حساب الوسط الحسابي والمنوال والوسيط والرباعيات .

الفصل الثالث: مقاييس التشتت والتشكل، ومقاييس أخرى لوصف تشتت البيانات .

الفصل الرابع : الارتباط والانحدار الخطي البسيطين .

الفصل الخامس :المتغيرات العشوائية ذات التوزيع الاحتمالي .

أما الجزء الثاني يحتوي على أكثر من عشر سلاسل من التمارين منها ما هو محلول بالتفصيل ومنها ما هو مختصر.

الفصل الأول :

مدخل للإحصاء وعرض البيانات

الفصل الأول : مدخل للإحصاء وعرض البيانات

1) تعريف علم الإحصاء:

يهتم علم الإحصاء بجمع البيانات (حول ظاهرة ما) وتلخيصها وتبويبها للاستفادة منها في وصف الظواهر وتحليلها وقياس واستقراء الوقائع للتنبؤ بسلوك الظاهرة حاضرا ومستقبلا لاتخاذ قرارات سليمة في ظروف عدم التأكد

ونستفيد من هذا التعريف أن:

- البيانات تكون متجانسة أي من نفس الصنف.
- توجد بين البيانات علاقات عددية مستقلة عن الصدفة.
- قياس الظواهر لاستقراء الواقع واستخراج العلاقات.
- تحليل البيانات للتنبؤ بالقيم المستقبلية للظاهرة المدروسة وفق شروط معينة.
- اتخاذ القرارات المناسبة من خلال تحليل سلوك الظواهر.

2) وظائف علم الإحصاء:

2-1) وصف البيانات:

تكون البيانات خاما ولا يمكن الاستفادة منها إلا إذا تم جمعها وعرضها إما في شكل جدولي أو بياني وحساب بعض المؤشرات الإحصائية (الوسط، المنوال، الانحراف، التباين.....الخ) التي تدلنا على طبيعة البيانات.

2-2) الاستدلال الإحصائي:

ويستند إلى فكرة اختيار جزء من المجتمع يسمى عينة بطريقة علمية مناسبة، واستخدام بيانات هذه العينة في التوصل إلى نتائج يمكن تعميمها على مجتمع الدراسة، ويمر الاستدلال الإحصائي بمرحلتين:

2-2-1) التقدير:

وهو معرفة معالم المجتمع (الخصائص التي تميزه كالمتوسط والتباين والتوزيع) عن طريق حساب مؤشرات تقديرية من بيانات العينة تسمى إحصائيات، نستخدم إحصائيات العينة لتقدير معالم

المجتمع. أي تقدير متوسط المجتمع مثلا من خلال معرفة متوسط العينة، وهذا ما يسمى التقدير بنقطة أما إذا أردنا أن نقدر من خلال العينة المجال الذي يمكن أن تقع فيه معلمة المجتمع (متوسط المجتمع) فنسمي ذلك التقدير بفترة.

2-2-2) اختبار الفرضيات(الفروض): يتم تحديد فروض معينة حول خصائص ظاهرة مدروسة (معالم المجتمع) واستخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفروض المحددة حول معالم المجتمع.

(فرضية: مثلا تتأثر ظاهرة التسرب المدرسي بالدخل الشهري للعائلة)

2-2-3) التنبؤ: نستخدم نتائج الاستدلال الإحصائي التي تدلنا على سلوك الظاهرة في الماضي لمعرفة ما يمكن أن يحدث لها في الحاضر والمستقبل.

3) مفاهيم أساسية:

3-1) الظاهرة: وهي الخاصية أو المتغير المدروس في المجتمع الإحصائي (الطول، السن، علامة الامتحان، الادخار، الاستهلاك، الدخل....)، وقد تتكون الظاهرة من متغير واحد أو عدة متغيرات فظاهرة التسرب المدرسي يمكن أن نجد فيها عدة متغيرات: (سن الطفل، بعد المدرسة عن المنزل، دخل الأسرة، عدد الأفراد في الأسرة، معاملة المعلم للطفل).

3-2) المجتمع الإحصائي: وهو مجموعة من الوحدات الإحصائية تتعلق بظاهرة قابلة للقياس وقد يكون المجتمع متجانسا (الأسر، المؤسسات) أو غير متجانس (كالديانات، والجنسيات).

3-3) الوحدة الإحصائية: وهي أصغر وحدة أساسية لتكوين المجتمع المدروس وتسمى أيضا المفردة أو المشاهدة، المجتمع الإحصائي (مجتمع الأسر، مجتمع المؤسسات)، والوحدات الإحصائية بالترتيب (الأسرة، المؤسسة).

4) أنواع البيانات:

4-1) بيانات وصفية: (نوعية أو كيفية) وهي بيانات غير رقمية مرتبة في شكل مستويات أو شكل فئات رقمية، ، ويتم قياسها بمعيارين هما:

4-1-1) معيار اسمي: تقاس به البيانات غير الرقمية التي تتكون من مجموعات (حالات) متنافية، وكل مجموعة لها خصائص تميزها عن المجموعات الأخرى حيث لا يمكن مقارنة المجموعات

(الحالات) أو ترتيبها، وقد يتم ترميزها أي إعطاؤها أرقاما لتسهيل معالجتها بالحواسيب حيث لا تمثل الأرقام قيمتها ولكن تمثل نوع الحالة فقط.

أمثلة: متغير الجنس: وحالاته (ذكر 1، أنثى2).

متغير الحالة الاجتماعية: وحالاته (أعزب1، متزوج2، أرمل3، مطلق4).

متغير أصناف التمر: وحالاته (......).

متغير الجنسية: وحالاته (جزائري1، مصري2، فرنسي3....).

4-1-2) معيار ترتيبي: تقاس به البيانات غير الرقمية التي تتكون من مستويات أو فئات يمكن ترتيبها تنازليا أو تصاعديا: أمثلة: متغير علامات الطلبة في المنهجية الإنجليزية: وحالاته (A A^+ B B^+ C C^+ D D^+).

متغير المستوى التعليمي: وحالاته (باحث، جامعي، ثانوي، أساسي، ابتدائي).

متغير الرأي: وحالاته (موافق تماما، موافق، غير موافق، غير موافق تماما).

4-2) بيانات كمية: هي بيانات يعبر عنها بأرقام عددية تمثل القيمة الفعلية للظاهرة وتنقسم لقسمين هما.

4-2-1) بيانات فترة (متقطعة أو منفصلة): هي بيانات رقمية تقاس بمقدار بعدها عن الصفر، ومتقطعة لأنها تأخذ قيما صحيحة ثابتة لا يمكن تجزئتها فعدد الأطفال يأخذ القيم1-2-3-4... ولكن لا يمكن أن يأخذ القيم 2,5 3,256....

4-2-1) بيانات نسبة (مستمرة أو متصلة): وتأخذ قيما غير متناهية في مجال صغير حيث يمكن للمتغير أن يأخذ قيما لا متناهية بين قيمتين مثل أطوال الطلبة التي قد تأخذ أي قيمة بين 175سم و176سم مثل 175,12 سم أو 175,47 سم...الخ.

قاعدة: كل متغير كمي يقاس قياسا هو متغير مستمر، وكل متغير كمي يعد ويحسب حسابا فهو متغير متقطع.

5) مصادر البيانات:

1-5) مصادر أولية: تمدنا بالبيانات بشكل مباشر من مصدرها حيث يقوم الباحث بجمعها من مفردات محل البحث مباشرة، إذا كان البحث حول الأسرة فجمع البيانات يتم من رب الأسرة مباشرة أو من ينوب عنه.

- تتميز بيانات هذه المصادر بالشمولية، التعميم، الدقة والثقة، لأن الباحث يجمعها بنفسه.
- ومن عيوبها أنها تحتاج للوقت والجهد وتكلفتها المادية الكبيرة نسبيا.

2-5) مصادر ثانوية: تمدنا بالبيانات بشكل غير مباشر (من أشخاص آخرين غير معنيين بالبحث)، كالأجهزة والهيئات الرسمية المتخصصة (نشرات الوزارات، وتقارير المعهد الوطني للإحصاء والتخطيط....). وتسميتها بالثانوية لا يعني التقليل من أهميتها ولكنها تعبر ثانوية بالنسبة للأولى.

- تتميز بيانات هذه المصادر بتوفير الجهد والوقت والمال.
- ومن عيوبها أنها تنقصها الثقة والدقة وليست بنفس الدرجة كالمصادر الأولية لأن هدف الباحث من جمعها يختلف عن هدف الهيئات التي نشرتها.

6) طرق جمع البيانات:

وتعتبر من أهم مراحل البحث الإحصائي حيث أن جمعها بأسلوب علمي يفيدنا بالوصول إلى نتائج دقيقة في التحليل، وتجد طريقتين لجمع البيانات وذلك حسب كبر المجتمع والهدف من البحث:

1-6/ أسلوب الحصر الشامل: أي حصر وعد كامل المجتمع ويتميز بالشمول وعدم تحيز البيانات ودقة النتائج، ومن عيوبه أنه يحتاج للوقت والمجهود والتكلفة العالية.

2-6/ أسلوب المعاينة: اختيار جزء من المجتمع محل الدراسة بطريقة علمية سليمة عن طريق مجموعة من العمليات بهدف تكوين عينة، ودراستها ثم تعميم النتائج على المجتمع، حيث يفضل هذا الأسلوب في حالة كبر حجم المجتمع ويصعب إجراء حصر شامل لمفرداته مثل دراسة المصابيح الكهربائية المنتجة في مصنع ما، تعداد أسماك البحر، معاينة نوع فصيلة الدم للسكان في بلد ما، ويتميز هذا الأسلوب بتقليل الجهد والوقت والتكلفة وبياناته أكثر تفصيلا، ومن عيوبه نقص الدقة والثقة (خاصة إذا كانت العينة لا تمثل المجتمع جيدا). ويتوقف نجاح استخدام هذا الأسلوب على:

كيفية تحديد حجم العينة n،

- طريقة اختيار مفردات العينة $X_1, X_2, X_3 \ldots X_n$،
- نوع العينة المختارة.

6-2-1/ العينة:

حالات كثيرة يكون من الصعب أو المستحيل دراسة جميع مفردات مجتمع البحث موضوع الإهتمام لضخامة حجم هذا المجتمع، لذا يلجأ الباحث إلى استخدام أسلوب العينات في دراسة خصائص مجتمع البحث"، لذا يمكن تعريف العينة بأنها جزء من المجتمع الذي يتم اختياره بطريقة علمية محددة ليستخدم في الحكم على الكل. وتسمى عملية الاختيار بالعينة ويفترض أن تكون العينة المختارة ممثلة للمجتمع وخواصه أصدق تمثيل بما في ذلك الاختلاف بين وحداته وذلك بأحسن ما يسمح به.

- **حجم العينة:**

ويقصد بحجم المجتمع جميع وحدات المعاينة التي يتكون منها المجتمع ويرمز له بالرمز n. أما حجم العينة فهو وحدات المعاينة التي تم اختيارها ويرمز له بالرمز.

- **وحدة المعاينة:**

ويقصد بوحدة المعاينة المفردة الأساسية التي تتكون منها عناصر مجتمع البحث، أو هي الجزء الصغير الذي تجمع منه البيانات وهي كل وحد من وحدات المجتمع وقد تكون متشابه من حيث الحجم أو مختلفة وعند تنفيذ البحوث الميدانية يجب تعريف وتحديد وحدة المعاينة تعريفا واضحا لجمع البيانات من الوحدات التي يشملها البحث عدم تدخل تلك الوحدات مع تلك التي لا يشملها البحث.

- **المعاينة:**

هي عملية اختيار جزء من المجتمع الاحصائي للاستدلال على خواص المجمع بأكمله عن طريق تعميم نتائج العينة وتقوم على علم وفن التحكم وقياس دقة المعلومات الإحصائية عن طريق استخدام بعض النظريات الرياضية وليست مجرد استخدام جزء من المجتمع بدلا من كله تتطلب عملية المعاينة وضع إطار يحتوي على وحدات المعاينة حتى يمكن اختيار العينة. إذ بدون هذا الإطار لا يمكن أن تتم تغطية كاملة للمجتمع أو إجراء معاينة احتمالية، ويجب أن يوضح الإطار الموقع والعنوان والحدود ومجموعة القواعد التي يمكن بموجبها الوصول إلى أي وحدة معاينة تختار في العينة. ويجدر التنويه بان عملية المعاينة ليست اقل كفاية أو دقة من عملية التعداد الشامل كما يتبادر إلى الذهن، بل إن نتائج العينة قد تكون أدق من نتائج التعدادات الشاملة بنفس الظروف.

يتميز البحث عن طريق العينة باختصار الوقت والجهد اللازمين وبالتالي بتخفيض التكاليف كما يمكن الحصول بسهولة على الردود الكاملة والدقيقة باستخدام جزء من المجتمع الكلي، وكذالك على بيانات أكثر تفصيلا ودقة من أفراد العينة، وتلخيصها وتحليلها على وجه السرعة، إضافة إلى سهولة تتبع غير المجيبين بينما يكون ذلك صعبا في حالة الحصر الشامل. وتساعد بحوث العينات على دقة معرفة الحصر الشامل، حيث يتم اختيار عينة ودراستها بدقة، وبمقارنة نتائجها مع نتائج التعداد يمكن معرفة مدى دقة نتائج الحصر الشامل يتضح مما سبق أهمية استخدام العينات والدور الذي تلعبه في الدراسات في مختلف الميادين حتى إن استخدام الحصر الشامل أصبح لا يغني عن استخدام العينة كما إن تحليل نتائج التعداد الشامل تحتاج إلى وقت طويل، بحيث يمكن أن تضيع الحكمة من التعديد أو تقل الاستفادة منه، وفي هذه الحالة يتحتم أخذ عينة وتحليل نتائجها لتعطي فكرة عن النتائج النهائية.

- **أهداف المعاينة**

يفترض تحديد الهدف الرئيسي والأهداف التفصيلية للمعاينة أو المشكلة المراد دراستها تحديدا واضحا، وذلك لتحديد البيانات المطلوب جمعها واستخدامها وبعد ذلك توضع التصميمات المختلفة والممكنة عن طريق الأسئلة المراد الحصول على إجابات عليها.

إن الغرض الأول من إجراء بحث أو تجربة هو إيجاد إجابات لأسئلة معينة لوضع أساس سليم للتنمو ولإتخاذ إجراءات معينة ولذلك لابد من تفسير نتائج المعاينة بطريقة تعطي اقصى الفوائد لا وضع التقديرات الإحصائية المختلفة المعالم المجتمع، ولابد أيضا من قيام دقة هذه التقديرات، من أهم المسائل وتصميم العينات هو الانتهاء إلى معادلة أو معادلات الحساب التقديرات من بيانات العينة وهذه المعادلة أو المعادلات المختارة لا بد أن تحتفظ بكل المعلومات الخاصة بالمجتمع التي تم الحصول عليها من العينة.

إن التقديرات هي قيم تقريبية لمعالم المجتمع الحقيقية، ويفترض أن يكون الفرق بين التقدير المحسوب من العينة والقيم الحقيقية للمجتمع ضئيلا بدرجة كافية تسمح بالاعتماد على التقدير دراسة المجتمع، ويغير ذلك فإن الباحث يعاني بعض الخسائر إذا ما استخلص نتائجه على أساس هذا التقدير، وإذا تم اختيار العينة والحصول على التقدير بطرق تعتمد على نظرية الاحتمالات فيمكن معرفة دقة هذا التقدير.

إن تقديرات المجتمع التي يمكن الحصول عليها من العينة كثيرة وأبسطها المتوسط الحسابي لعينة عشوائية، فمن المعروف بأن هذا المتوسط يعطى تقديرا لمتوسط المجتمع الذي سحبت منه العينة، غير أنه لن يكون مساويا تماما لمتوسط المجتمع وذلك يرجع إلى أخطاء المعاينة. ومن التقديرات الأخرى المعالم المجتمع التي تحصل عليها من المعاينة هي التباين والتفرطح والالتواء.

7) أنواع المعاينات:

هنالك نوعان أساسيان من العينات هما العينات الاحتمالية Probability Samples والعينات غير الاحتمالية Non-Probability Samples، فالعينات الاحتمالية هي تلك العينات التي تكون فيها كل الوحدات في المجتمع ذات فرصة معلومة في الاختيار، أي أن العينة الاحتمالية التي تم اختيار الوحدات الإحصائية فيها بطريقة الصدفة by chance، وأن هنالك فرصة معروفة لكل وحدة تم اختيارها.

أما العينات غير الاحتمالية فهي تلك العينات التي لا يكون لكل مفردة من مفردات مجتمع البحث فرصة أو درجة احتمال متساوية للاختيار في العينات، أي أن احتمالات اختيار أي مفردة من مفردات مجتمع البحث غير معروفة، ومما هو جدير بالإشارة أن العينات الاحتمالية تعتبر الأكثر دقة وبحوث التسويق التي تركز على التقدير الإحصائي لظاهرة أو خاصية في مجتمع البحث.

7-1/ العينات الاحتمالية

يدخل ضمن مفهوم العينات الاحتمالية الأنواع الآتية:

أ- **العينة العشوائية البسيطة: Simple Random Sample**

يتلخص مفهوم العينة العشوائية البسيطة بان لكل مفردة من مفردات المجتمع فرص متكافئة في الاختيار والتمثيل، أي أنه ليس هنالك تحيز ينتج من الاختيار، فيتم اختيار هذه العينة بالقرعة مثلا وذلك بكتابة الأسماء ووضعها في صندوق صغير ثم يتم سحب القطع الورقية حتى تعرف الأسماء أو بإعطاء كل فرد من أفراد المجتمع رقما معينا ويتم سحب ذلك عشوائيا دون تحديد. فالأرقام التي تم سحبها تمثل المجتمع كله.

ويمكن أيضا كتابة الأرقام على كرات عديدة وتدور هذه الكرات بسرعة ثم يتم تنزيل كل من هذه الكرات التي تحمل رقما من الأرقام وهذه الحالة غالبا ما تستخدم في حالة القرعة في توزيع الألعاب في الدورات الرياضية مثلا. وهناك جداول خاصة بالأرقام العشوائية اعددها مختصون في هذا المجال.

ب-العينة المتعددة المراتب: Multiage Sample

وغالبا ما تستخدم هذه العينات في المجتمعات الكبيرة نسبيا وهي تمتاز بعدم تجانسها، وهذه العينات قد تنجز على مدى خطوتين أو أكثر لابتكار العينة المطلوبة، والملاحظ أن المجتمع الواحد يقسم إلى عدة مجاميع (أسر، أفراد، طلبة...الخ)، وكل مجموعة بدورها تنقسم إلي مجموعات فرعية أصغر (أسر كبيرة العدد صغيرة العدد، أو طلبة جامعة طلبة ثانوية) وهكذا، والأمثلة السائدة هي ما ينطبق على العينة المسيحية التي يتم إجراؤها على مراحل متعددة.

وعلى فرض أنه طلب أخذ عينة من مقاطعات القطر حول عدد الأفراد الموظفين في الأسر العراقية فالخطوة الأولى هي ابتكار عينة عنقودية من المقاطعات المختلفة في القطر، كل مقاطعة لها احتمالية بأن تكوين عينة عنقودية نسبة إلى مجتمعها.

والخطوة الثانية هي الحصول على عينة عنقودية للدور من كل مقاطعة مختارة الخطوة الثالثة اختيار عينة عنقودية من صفوف الدور من كل مدينة. مرة ثانية أن يتم وزن كل صف من الدور بعدد الساكنين فيه. وأخيرا فإن عينة منتظمة من الساكنين من كل قطع يتم اختيارها وأن عينة عشوائية من الساكنين يتم الحصول عليها، والنتيجة هي عينة عشوائية للمنطقة وإذ يكون لكل أسرة فرصة مساوية لأن تكون في العينة.

ومن مزايا هذه الطريقة أن كلفة العمل الميداني تكون واطئة خاصة وأنها لا تحتاج إلى عمل ميداني واسع.

ج- العينة الطبقية: stratified Sample

وفقا للعينة الطبقية فإن المجتمع الواحد يقسم الي مجاميع متجانسة، بصفات معينة خاصة عندما يكون مجتمع البحث غير متجانس في الصفات والخصائص، كالتباين الحاصل دخول الأفراد والمستوى الثقافي ومحل السكن وإلى غير ذلك من الخصائص المتباينة في المجتمع الواحد.

وهنالك شكلان من العينة الطبقية: العينة الحصصية والعينة المسحية. ففي مثل هذه الحالة يتم تقسيم مفردات المجتمع إلى مجاميع متجانسة على شكل فئات أو قطاعات أو طبقات وبعد ذلك يتم أخذ عينات عشوائية من كل فئة أو طبقة من هذه الطبقات وبحث أن حجم العينة العشوائية في كل فئة أو رتبة يكون مساويا للحجم النسبي للرتبة في مجتمع للدراسة.

د- العينة المنتظمة(المرتبة): Systemtic Sample

والعينة المنتظمة يتطلب أن تكون عينة الدراسة محددة طبيعيا وهذه غالبا ما تأخذ شكل القوائم التي تضم أسماء(فراد معينين مع أرقام لهم كما هي الحال بالنسبة لدليل التلفون وعلي وفق هذه العينة فإن المجتمع الأصلي يقسم إلي مجاميع متساوية العدد (فئات) فإذا كان الهدف هو اختيار عينة من 500 طالب تم سحبها من 5000 مفردة، فان المجتمع في هذه الحالة يقسم إلى 10 مجاميع متساوية بتقسيم (500/5000)= 10 مجاميع والمرحلة الثانية وهي نقطة الانطلاق بين النقطة 1 والنقطة 10 وهي ما تسمى بمدى المعاينة (من 1 إلى 10)، فيتم اختيار أية مفردة من المفردات الواقعة بين هذين الرقمين وليكن 3، عليه فان الطلبة الواجب اختيارهم سيمثلون رقم 23، 13، 3،.....483.

والعينة المنتظمة لها العديد من المزايا تكمن في السهولة وأنها قليلة، ومن جهة أخرى لأن العمل الميداني يمكن إجراؤه على مساحات واسعة، وفضلا عن ذلك فإن هذا الأسلوب، بعد أكثر فعالية لبعض الأنواع من الدراسات، مثال ذلك لدراسة مواقف الأفراد تجاه خدمات التلفون.

هـ - العينة العنقودية: Cluster sample

على وفق هذا المفهوم فإن المجتمع يقسم مرة ثانية إلى مجاميع فرعية، ويتم اختيار المجاميع على شكل عناقيد بدلا من المجاميع الفرعية، أي أن اختيار الوحدات من المجتمع غير المرتب(غير المنطقي) يتم في مجاميع أو عناقيد.

والملاحظ أن هذه الفئات الممثلة على شكل عنقود تشترك فيما بينها بصفات معينة جعلتها تظهر شكل عنقود، ولهذا السبب فان هذه العينات تعد من أفضل أنواع العينات وذلك لأن عملية الاختيار مبنية على أسس معينة وعلمية.

فمثلا إن كان لدينا 1000 فلاح ويراد معرفة الاستهلاك الأسبوعي من الخبز لكل فلاح، فإنه يمكن تقسيم 1000 فلاح إلى 20 قسما، كل قسم يضم 50 فلاحا، فعينة من 100 فلاح يمكن أخذها من قسمين فقط ومن أية مجمعة من المجاميع وبشكل عشوائي من العشرين قسما، ويتم تجميع المعلومات من كل الفلاحين بهاتين المجموعتين فقط تمثلان الفئة كلا والعينة المسحية (المكانية): Areal sample تعد العينة المسحية شكلا من أشكال العينة العنقودية، ويفترض في هذه العينة أن كل الوحدات والمجتمع لها محل للإقامة فيه، ووفقا لذلك فإن المناطق الجغرافية يتم تقسيمها إلى تقسيمات عديدة، ابتداءا من تقسيم المنطقة الجغرافية إلى بلدان عديدة أو أن يقسم البلد إلى مناطق عديدة كالمحافظات ثم تقسيم ذلك إلى مدن والمدينة العديد من الأحياء السكنية وهكذا.

فلو أراد الباحث معرفة موقف استهلاك الأسر من الطحين وفي مدينة من المدن الواقعة في بلد معين. وعلى وفق التقسيم المذكور أنفا، فان يتم اختيار عينة عشوائية من صفوف الدور في حي من الأحياء في المدينة المذكورة، وان عملية الاختيار لهذه الصفوف من الدور كانت عشوائية وعن طريق الأفراد الباحثين الطوافين على هذه الدور يكون بالإمكان تحديد مقدار الاستهلاك الأسبوعي أو الشهري للطحين.

ويلاحظ أن إجراء الدراسة على وفق هذه العينات هو من المسائل السهلة نسبيا خاصة وان العينة جاهزة في أي وقت كان ومن أي منطقة يتم اختيارها وفضلا عن ذلك فان تكاليف إجراءات العينة واطئة نسبيا، عدا أن المقابلات كلما طالت وزاد عددها فان ذلك سيزيد من المكلفة.

7-2/ العينات غير الأحتمالية:

هناك أنواع من العينات غير الاحتمالية

أ- العينة المستندة على حكم الباحث: Judgement Sample

في مثل هذه العينات يقوم الباحث باختيار الوحدات التي يعتقد أنها مناسبة في تمثيل مجتمع الدراسة، وهذه مسألة تعتمد على خبرة الباحث وقدرته على الاختيار، ولعل هذا الاختيار لم يأت بنحو اعتباطي أو عاطفي وإنما غالبا ما يستند الباحث على خبرته وتجربته.

والملاحظ أن هذا النوع من العينات يتجنب استخدامه العديد من الباحثين وذلك لان التقدير الشخصي قد لا يكون حاسما في العديد من الحالات، الذي غالبا ما تتحكم فيه عوامل أخرى، كذلك فإن اختيار التمثيل للعينة يعد أمرا غير ممكنا بسبب أن الحكم الشخصي هو المعتمد في هذا الجانب

ت- العينة الحصصية: Quota sample

تعد العينات الحصصية أكثر أنواع العينات غير الاحتمالية استخداما، إذ أنها تضم تقديرا سابقا للرقم الناسب أو المطلق للوحدات الواجب أخذها ومعاينتها كل طبقة من المجتمع فقبل الشروع بالبحث فإنه من الضروري أن يكون الباحث قد راجع العديد من الحقائق والأرقام المراد دراستها وضوء ذلك باختيار العينة الحصصية هذه، إذ يتم اختيار أفراد العينة من بين أفراد الجماعات أو الفئات التي تمتاز بخصائص معينة وذلك بنسبة الحجم العددي، فإذا كان عدد الطلبة في المجتمع 40 طالبا من مجموع 200 فرد والمجتمع فإن نسبة من يختارهم من الطلبة يجب لا يتجاوز 20% فإذا تم اختيار عشرين فردا من المجتمع فإن عدد الطلبة يجب أن لا يتجاوز 5 طلبة من مجموع العشرين فردا، وقد تبدو العينة المختارة بطريقة الحصة مماثلة للعينة الطبقية التي سبق الإشارة إليها،

ولكن الاختلاف يكمن في أن اختيار المفردات في العينة الطبقية لا يترك للشخص الذي يقوم بالمقابلة بل يتم ذلك عشوائيا، أما والعينة الحصصية فإن الشخص القائم بالبحث تترك له حرية اختيار الأشخاص حتى يحصل على الحصة المطلوبة من كل طبقة أو فئة مما يؤدي ذلك إلى بعض التحيز، وفضلا عن ذلك فان دقة المعاينة الحصصية لا يمكن حسابها بالمعادلات الرياضية.

ج- العينة الميسرة: Convenience Sampling

العينة الميسرة العملية هي تلك العينة التي تكون فيها عملية اختيار وحدات المجتمع على أساس السهولة أو الملائمة، أي التيسير توفير المراد مقابلتهم في محل أو مكان محدد ". حيث أن الافتراض الخاص بالعينة الميسرة هو أن عناصر المجتمع الواحد يعدون متجانسين في صفاتهم وخصائصهم، إلا أن النقد الموجه لها يكمن في أن العناصر المناسبة أو الميسرة للمجتمع تختلف اختلافا جوهريا من تلك التي تكون أقل تيسيرا.

وتتميز العينة الميسرة بقلة تكاليف اختيار العينة وجمع البيانات لاسيما وأنه يمكن للمقابل إجراء عدد كبير من المقابلات خلال وقت موجز. ولكن يؤخذ عليها أنها تعتمد على الراي الشخصي والاختيار الذي قد لا يكون موفقا في بعض الحالات الدراسية.

كما يمكن الإشارة إلى أن الفرق بين قيمة مجتمع البحث محل الدراسة وبين قيمة العينة له غير معروف، سواء فيما تعلق بحجم الفرق أو اتجاهه. كما أننا لا نستطيع قياس خطا المعاينة، ولا تستطيع استخلاص بيانات استنتاجية عن نتائج مثل هذه العينة. ومع ذلك فإن العينة الميسرة يمكن أن تكون مبررة لاستخدامها في مرحلة البحث الاستطلاعي، كأساس لتحديد الفروض، وكذلك في الدراسات الاستنتاجية عندما يرغب المدير أن يقبل المخاطرة بأن نتائج الدراسة قد تكون غير دقيقة إلى حد بعيد.

8) عرض البيانات:

تمر العملية الإحصائية بمراحل متعددة تبدأ بمرحلة التصميم ثم تليها مرحلة جمع البيانات ومراجعتها ميدانيا، وأخيرا مرحلة التجهيز بما تشمله من مراجعة مكتبية وترميز وإدخال البيانات إلى الحاسب .

ولكي نضع البيانات في جداول إحصائية يجب أولاً تقسيم البيانات إلى مجموعات متشابهة تسمى فئات ونضع في كل فئة المفردات التي تنتمي إليها (أو بمعنى آخر نوجد عدد مرات تكرار الفئات) ثم نضع هذه الفئات وتكرارها في جداول ويطلق على الفئات لفظ (الفئات التكرارية) وكل جدول يحتوي على عدد من هذه الفئات التكرارية يسمى " جدولا تكراريا "، ويختلف شكل الجدول طبقا لنوع البيانات، وحسب عدد المتغيرات، وفيما يلي عرض بيانات متغير (وصفي أو كمي) في شكل جدول تكراري بسيط.

8-1) جدوليا: في حالة دراسة ظاهرة تحوي متغير وصفي واحد فإننا نعرض البيانات في جدول تكراري حيث:

- العمود الأول نضع فيه (X_i) المجموعات المكونة للظاهرة أو الحالات.
- العمود الثاني نضع فيه f_i تكرار الظاهرة في كل مجموعة أو حالة (fréquences).
- العمود الثالث وفيه **التكرار النسبي** $f\% = f_i$ تكرار الحالة / مجموع التكرارات n **(حجم العينة)**.

$$\left(f\% = \frac{f_i}{n}\right)$$

وفي حالة بيانات أو متغيرات كمية فإننا نضعها في فئات متساوية، ونكون جدول التوزيع التكراري حيث: وقبل تكوين التوزيع التكراري يتم تحديد المدى الذي تتراوح فيه البيانات، تحديد عدد الفئات، تحديد طول الفئة، تحديد حدود الفئات (الحد الأدنى والأعلى لكل فئة)، مراكز الفئات:

- حساب المدى Range الذي تتراوح فيه البيانات: وهو الفرق بين أكبر قيمة وأصغر قيمة للبيانات في العينة.
- تحديد عدد الفئات Classes : توجد عدة قوانين إلا أن أغلب الباحثين يرون أنها يجب أن لا تقل عن 5 فئات وأن لا تزيد عن 15 فئة وأكثر عدد الفئات استعمالا هو 8 فئات .
- تحديد طول الفئة L ويساوي حاصل قسمة المدى على عدد الفئات، range / nombre de classes = L.
- حدود الفئات:

. الحد الأدنى للفئة الأولى هو أقل قيمة للبيانات.

. الحد الأعلى للفئة الأولى هو الحد الأدنى لها زائدا طول الفئة.

. **(وهكذا الحد الأعلى لكل فئة = حدها الأدنى + طول الفئة)**

. الحد الأدنى للفئة الثانية هو الحد الأعلى للفئة الأولى.

. **(الحد الأدنى لكل فئة = الحد الأعلى للفئة التي قبلها)** إلا الفئة الأولى فإن حدها الأدنى هو أقل قيمة في البيانات.

- مراكز الفئات: وهي القيمة التي تقع في منتصف الفئة وتحسب كما يلي:

مركز الفئة = (الحد الأدنى للفئة + حدها الأعلى) / 2

تكوين جدول التوزيع التكراري:

- العمود الأول نضع فيه (X_i) المجموعات المكونة للظاهرة أو الحالات.
- العمود الثاني نضع فيه f تكرار الظاهرة في كل مجموعة.
- العمود الثالث ونضع فيه التكرار المتجمع الصاعد $f\uparrow$.
- العمود الرابع ونضع فيه التكرار المتجمع النازل $f\downarrow$.
- العمود الخامس ونضع فيه التكرار النسبي $f\%$ = تكرار كل مجموعة f_i / مجموع التكرارات n ($f\% = \frac{f_i}{n}$)
- العمود السادس ونضع فيه التكرار المتجمع الصاعد النسبي $f\uparrow\%$ ($f\uparrow\% = \frac{f\uparrow}{n}$).
- العمود السابع ونضع فيه التكرار المتجمع النازل النسبي $f\downarrow\%$ ($f\uparrow\% = \frac{f\downarrow}{n}$).

إذا كان التكرار البسيط والتكرار النسبي يعبران بالترتيب عن حجم الحالة ونسبة الحالة في العينة فإننا:

نستعمل التكرارات **المتجمعة <u>الصاعدة</u>** البسيطة أو النسبية لمعرفة عدد أو نسبة المشاهدات **<u>التي تقل</u>** عن قيمة معينة.

نستعمل التكرارات **المتجمعة <u>النازلة</u>** البسيطة أو النسبية لمعرفة عدد أو نسبة المشاهدات **<u>التي تزيد</u>** عن قيمة معينة.

مثال : ليكن لديك التوزيع التكراري لعينة حجمها 40 مزرعة حسب نوع التمر الذي تنتجه

التوزيع التكراري النسبي المئوي	التوزيع التكراري النسبي	عدد المزارع (التكرارات) (f)	نوع التمر
0.125*100=12.5	(5/40)=0.125	5	سكري
0.25*100=25	(10/40)=0.25	10	خلاص
0.325*100=32.5	(13/40)=0.325	13	برحي
0.20*100=20	(8/40)=0.20	8	صقعي
0.10*100=10	(4/40)=0.10	4	نبوت سيف
100	1.00	40	المجموع

من الجدول السابق يلاحظ أن نسبة المزارع التي تنتج النوع "برحي" في العينة هي 32.5% وهي أكبر نسبة مما يدل على أن النمط الشائع في إنتاج التمور هو ذلك النوع، بينما نجد أن نسبة المزارع التي تنتج النوع "نبوت سيف" حوالي 10.0% وهي أقل نسبة.

مثال 2:

فيما يلي بيانات عن المستوى التعليمي لعينة من 50 فرد.

متوسط	يقرأ ويكتب	ثانوي	متوسط	ثانوي	أعلى من جامعي	متوسط	ابتدائي
يقرا ويكتب	متوسط	ثانوي	ثانوي	متوسط	ثانوي	ابتدائي	متوسط
ابتدائي	ثانوي	يقرا ويكتب	جامعي	ثانوي	ابتدائي	يقرا ويكتب	ثانوي
متوسط	ابتدائي	متوسط	ثانوي	ابتدائي	متوسط	جامعي	متوسط
ثانوي	متوسط	ابتدائي	ثانوي	يقرا ويكتب	ابتدائي	ثانوي	ابتدائي
جامعي	ثانوي	جامعي	ابتدائي	جامعي	أعلى من جامعي	ثانوي	ثانوي
متوسط	يقرا ويكتب						

والمطلوب:

1- اعرض البيانات في شكل جدول تكراري.

2- كون التوزيع التكراري النسبي، ثم علق على النتائج.

1/ عرض البيانات في شكل جدول تكراري:

المستوى التعليمي (يقرأ ويكتب- ابتدائي_ متوسط- ثانوي- جامعي- أعلى من جامعي) متغير وصفي ترتيبي، ويمكن عرض البيانات أعلاه في شكل جدول تكراري بإتباع الآتي:

تكوين جدول تفريغ البيانات:

المستوى التعليمي	عدد الأفراد (التكرارات)
يقرأ ويكتب	6
ابتدائي	10
متوسط	12
ثانوي	15
جامعي	5
أعلى من جامعي	2
المجموع	**50**

التوزيع التكراري لعينة حجمها 50 فرد حسب المستوى التعليمي

المستوى التعليمي	عدد الأفراد (التكرارات) (f)	التوزيع التكراري النسبي	التوزيع التكراري النسبي المئوي
يقرأ ويكتب	6	0.12	12
ابتدائي	10	0.20	20
متوسط	12	0.24	24
ثانوي	15	0.30	30
جامعي	5	0.10	10
أعلى من جامعي	2	0.04	4
المجموع	**50**	**1.00**	**100**

ومن التوزيع النسبي يلاحظ أن حوالي 30% من أفراد العينة ممن لديهم مؤهل ثانوي، بينما يكون نسبة الأفراد ممن لديهم مؤهل اقل من الثانوي (متوسط، ابتدائي، يقرأ ويكتب) أكثر من 5%، أما نسبة الأفراد الحاصلين على مؤهل أعلى من جامعي حوالي 4% وهي أقل نسبة.

ملاحظات على الجدول

عند تكوين جدول ما لعرض البيانات، يجب مراعاة الآتي:

1- كتابة رقم للجدول.

2- كتابة عنوان للجدول.

3- لكل عمود من أعمدة الجدول عنوان يدل على محتواه.

4- يجب كتابة مصدر البيانات في الجدول.

عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط

بنفس الأسلوب السابق المتبع في تكوين جدول تكراري، يمكن أيضا عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط، ويتكون هذا الجدول من عمودين، الأول يحتوي على فئات تصاعدية للقراءات التي يأخذها المتغير، والثاني يشمل التكرارات أو عدد المفردات التي تنتمي قراءاتها للفئة المناسبة لها، والمثال التالي يبين كيف يمكن عرض البيانات الكمية.

مثال :

فيما يلي بيانات درجات 70 طالب في الاختبار النهائي لمقرر مادة الإحصاء.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

والمطلوب:

1- كون التوزيع التكراري لدرجات الطلاب.

2- كون التوزيع التكراري النسبي.

3- ما هو نسبة الطلاب الحاصلين على درجة ما بين 70إلى أقل من 80؟

4- ما هو نسبة الطلاب الحاصلين على درجة أقل من 70 درجة؟

5- ما هو نسبة الطلاب الحاصلين على درجة 80 أو أكثر ؟

الحل :

1- **تكوين التوزيع التكراري:**

درجة الطالب في الاختبار متغير كمي مستمر، ولكي يتم تبويب البيانات في شكل جدول تكراري، يتم اتباع الآتي:

- **حساب المدى Range(R)**

Range = Maximum – Minimum

$$R = 94 - 55 = 39$$

- تحديد عدد الفئات Classes(C):

تتحدد عدد الفئات وفقا لاعتبارات منها: رأي الباحث، والهدف من البحث، وحجم البيانات، ويرى كثيرا من الباحثين أن أفضل عدد للفئات يجب أن يتراوح بين 5 إلى 15 ، بفرض أن عدد الفئات هو 8 فئات، أي أن: (C=8).

- حساب طول الفئة Length(L):
- **تحديد الفئات:**

الفئة تبدأ بقيمة تسمي الحد الأدنى، وتنتهي بقيمة تسمي الحد الأعلى، ومن ثم نجد أن :

– الحد الأدنى للفئة الأولى هو أقل قراءة (درجة) أي أن الحد الأدنى للفئة الأولى = 55

الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة = L + 55 = 5+55=60

إذا الفئة الأولى هي: "55 to les than 60" وتقرأ " من 55 إلى أقل من 60 "

_ الحد الأدنى للفئة الثانية = الحد الأعلى للفئة الأولى = 60

الحد الأعلى للفئة الثانية = الحد الأدنى للفئة + طول الفئة = 5 + 60 = 65

إذا الفئة الثانية هي: "60 to les than 65" وتقرأ "من 60 إلى أقل من 65"

- وبنفس الطريقة يتم تكوين حدود الفئات الأخرى، وهي:

الفئة الثالثة : 65 to les than 70 الفئة الرابعة : 70 to les than 75

الفئة الخامسة: 75 to les than 80 الفئة السادسة: 80 to les than 85

الفئة السابعة: 85 to les than 90 الفئة الثامنة: 90 to les than 95

ويمكن كتابة الفئات بأشكال مختلفة كما هو مبين بجدول تفريغ البيانات

جدول تفريغ البيانات

الدرجة			عدد الطلاب (التكرارات)
فئات	فئات	فئات	
to les than 60 55	60 – 55	-55	10
to les than 65 60	65 – 60	-60	12
to les than 70 65	70 – 65	-65	13
to les than 75 70	75 – 70	-70	16
to les than 80 75	80 – 75	-75	10
to les than 85 80	85 – 80	-80	4
to les than 90 85	90 – 85	-85	3
to les than 95 90	95 – 90	-95	2
المجموع			70

تكوين الجدول التكراري

التوزيع التكراري لعدد 70 طالب حسب درجاتهم في اختبار مقرر الإحصاء

فئات الدرجة	عدد الطلاب (التكرارات) (f)	التكرار النسبي
60 – 55	10	0.143
65 – 60	12	0.171
70 – 65	13	0.186
75 – 70	16	0.229
80 – 75	10	0.143
85 – 80	4	0.057
90 – 85	3	0.043
95 – 90	2	0.028
المجموع	70	1.00

نسبة الطلاب الحاصلين على درجات ما بين 70 إلى أقل من 80 هو مجموع التكرارين النسبيين للفئتين الرابعة والخامسة:

= نسبة الطلاب الحاصلين على درجات ما بين (80 ، 70)

أي حوالي 37.2% من الطلاب حصلوا على درجات ما بين (80 ، 70) .

نسبة الطلاب الحاصلين على درجات أقل من 70، هو مجموع التكرارات النسبية للفئات الأولى والثانية، والثالثة:

- نسبة الطلاب الحاصلين على درجة أقل من 70

أي أن حوالي 50% من الطلاب حصلو على درجة أقل من 70 درجة

نسبة الطلاب الحاصلين على درجة 80 أو أكثر، هو مجموع التكرارات النسبية للفئات الثلاث الأخيرة:

- نسبة الطلاب الحاصلين على درجات 80أو أكثر

أي أن حوالي 12.8% من الطلاب حصلوا على درجة 80 أو أكثر.

8-2) بيانيا:

8-2-1/ العرض البياني للبيانات الوصفية

يمكن عرض البيانات الخاصة بمتغير وصفي في شكل دائرة بيانية أو أعمدة بيانية، يمكن من خلاله وصف ومقارنة مجموعات أو مستويات هذا المتغير

- **الدائرة النسبية :**

هي عبارة عن دائرة تقسم الى قطاعات زواياها المركزية تتناسب مع القراءات لعرض بيانات المتغير الوصفي، ويتم توزيع الـ 360^{o} درجة حسب التكرار النسبي لمجموعات المتغير، حيث تحدد مقدار الزاوية الخاصة بالمجموعة بتطبيق المعادلة التالية:

مقدار زاوية القطاع = التكرار النسبي للمجموعة ×360

مثال :

الجدول التكراري التالي يبين توزيع عينة حجمها 500 أسرة حسب المنطقة التي تنتمي إليها.

المنطقة	الرياض	الشرقية	القصيم	الغربية	المجموع
عدد الأسر	150	130	50	170	500

المطلوب :مثل البيانات أعلاه في شكل دائرة بيانية.

1- تحديد مقدار الزاوية المخصصة لكل منطقة، بتطبيق المعادلة:

التكرار النسبي للمنطقة ×360^{o}= مقدار الزاوية المخصص للمنطقة

المنطقة	عدد الأسر	التكرار النسبي	مقدار الزاوية
الرياض	150	0.30	$360 \times 0.30 = 108^{o}$
الشرقية	130	0.26	$360 \times 0.26 = 93.6^{o}$
القصيم	50	0.10	$360 \times 0.10 = 36^{o}$
الغربية	170	0.34	$360 \times 0.30 = 122.4^{o}$
المجموع	500	1.00	360^{o}

يتم رسم دائرة وتقسيمها إلى أربع أجزاء لكل منطقة جزء يتناسب مع مقدار الزاوية المخصصة له، كما هو مبين في الشكل التالي:

ومن الشكل أعلاه يلاحظ أن نسبة الأسر التي تنتمي للمنطقة الغربية حوالي 34% وهي أكبر نسبة في العينة، بينما يكون نسبة الأسر في منطقة القصيم حوالي 10% وهي أقل نسبة في العينة.

- **الاعمدة البيانية :**

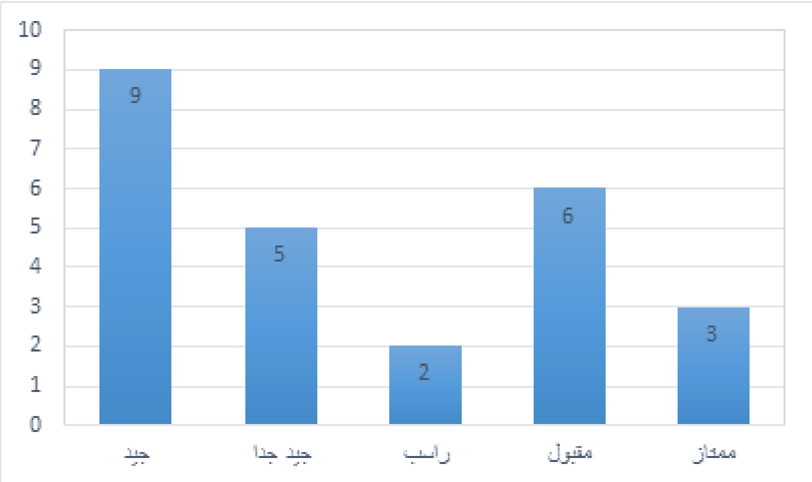
هي عبارة عن مجموعة من الأعمدة الرأسية أو المستطيلات المتساوية القاعدة والتي يتناسب ارتفاعها مع البيانات التي تمثلها، وعادة يأخذ المحور الرأسي لتمثيل قيم الظاهرة، والمحور الأفقي لتمثيل الفئة

مثال : البيانات التالية تبين التقديرات التي حصل عليها (25) طالبا في اختبار مادة الاحصاء.

التكرار	التقدير
9	جيد
5	جيد جدا
2	راسب
6	مقبول
3	ممتاز
25	المجموع

المطلوب تمثيل البيانات بالأعمدة البيانية.

نستخدم المحور الرأسي لتمثيل التكرار، والمحور الأفقي لتمثيل التقدير ومن خلال البيانات الموجودة بالجدول السابق نحصل على الشكل التالي:



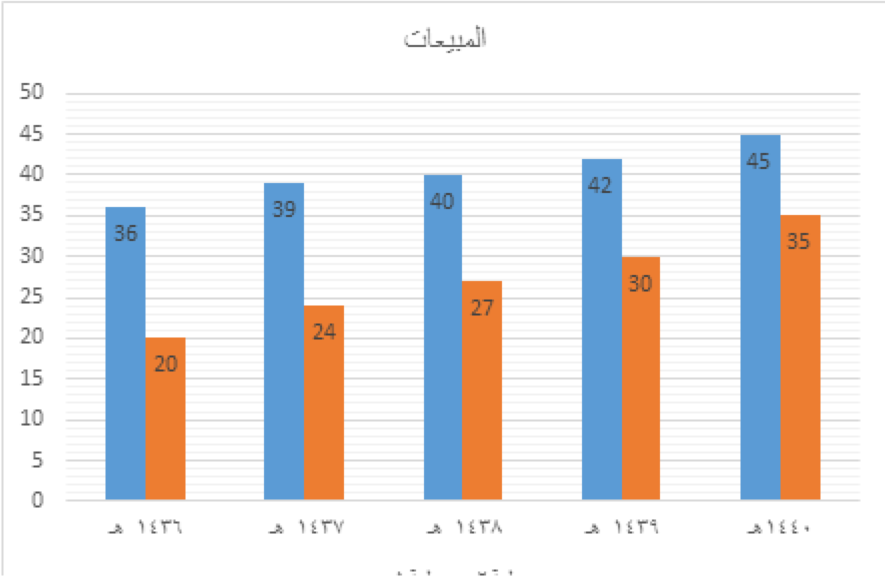
- **الاعمدة البيانية المزدوجة :**

ستخدم الأعمدة البيانية المزدوجة اذا كان الهدف هو مقارنة ظاهرتين أو أكثر لعدة سنوات أو اذا كان لدينا بيانات مزدوجة لخواص مختلفة. ونحصل عليها برسم عمودين متلاصقين يمثلان قيم الظاهرتينمحل الدراسة في كل سنة يحيث يتناسب طول العمود مع العدد الذي يمثله ونفرق بين الأعمدة بالألوان المختلفة، ومن الضروري أن تكون قواعد المستطيلات متساوية وعلى أبعاد متساوية.

مثال : البيانات التالية تمثل الكميات المباعة لسلعتين خلال الفترة من 1436 حتى 1440 هـ

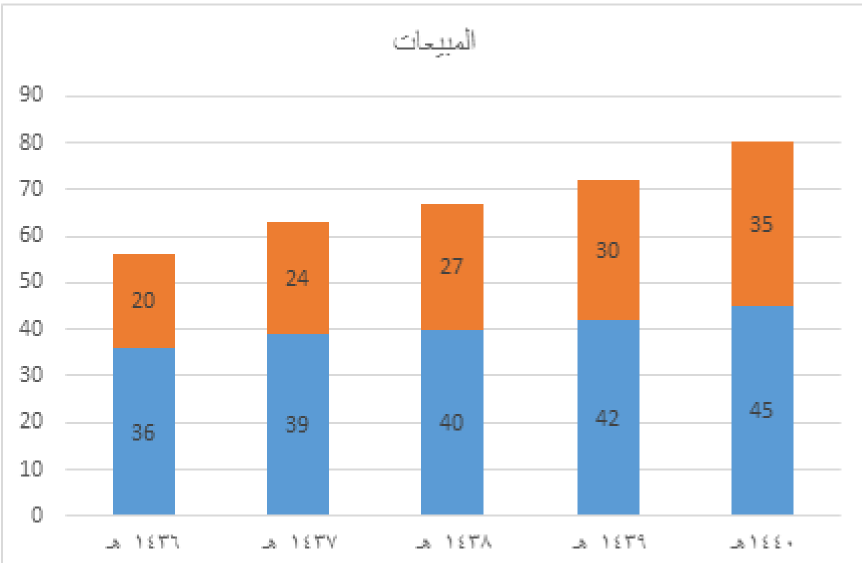
السنة المالية		1436 هـ	1437 هـ	1438 هـ	1439 هـ	1440هـ
كمية المبيعات	سلعة 1	36	39	40	42	45
	سلعة 2	20	24	27	30	35

وبتمثيل البيانات نحصل على الشكل التالي:



الأعمدة البيانية المجزأة :

تستخدم الأعمدة البيانية المجزأة اذا كان الهدف هو مقارنة الجزء بالكل لظاهرتين أو أكثر ويتم الحصول عليها برسم عمود واحد يمثل جملة الظواهر محل الدراسة في كل سنة، ثم نقسم كل عمود الى مكوناته بحيث يتناسب كل جزء مع العدد الذي يمثله، ونميز هذه الأجزاء بالألوان المختلفة.



8-2-2/ العرض البياني للبيانات الكمية

العرض البياني للبيانات، هو أحد طرق التي يمكن استخدامها في وصف البيانات، من حيث شكل التوزيع ومدى تمركز البيانات، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة، وتختلف طرق عرض البيانات بيانيا حسب نوع البيانات المبوبة في شكل جدول تكراري، وفيما يلي عرض للأشكال البيانية المختلفة.

- **المدرج التكراري :**

نحتاج لرسم كل من المدرج التكراري (الأعمدة البيانية التكرارية)، المضلع التكراري والمنحنى البياني للتوزيع التكراري البسيط؟، نحتاج لرسمها إلى:

- لرسم المدرج البياني أو الأعمدة البيانية يحدد طول الخط الأفقي أو العمودي من خلال أصغر وأكبر قيمة:

. على الخط الأفقي نكتب أرقام الفئات (1–2…7–8) أو نكتب قيم حدودها (18–22…46–50).

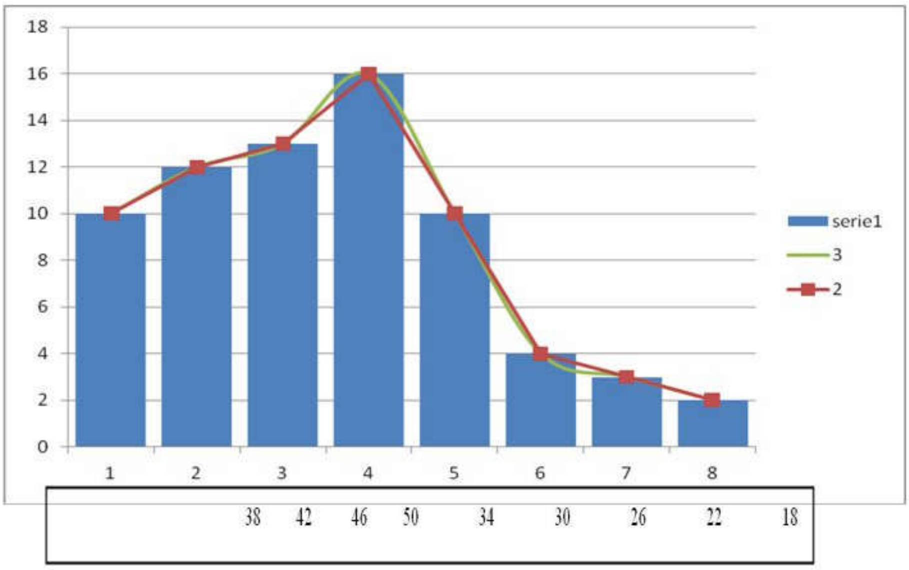
. وعلى الخط العمودي نبحث عن أعلى قيمة للتكرار وحسب تمرين مثال أعمار الطلبة فإن: أكبر قيمة للتكرار هي 16 وأدنى قيمة هي 2 وتقترب من 0، لذا نقسم الخط العمودي إلى أجزاء متساوية بين0 و16 حسبما يفيد الطالب في الرسم (ولا يشترط أن تكون أرقام هذه الأجزاء موجودة في التكرارات).

ونرسم الأعمدة البيانية حيث: عرض العمود = طول الفئة وطول المدرج = تكرار الفئة

- ولرسم المضلع أو المنحنى نحدد كل الثنائيات (مركز الفئة، تكرار الفئة) ونصل بالمضلع أو المنحنى بينها.

الشكل في حالة التوزيع التكراري البسيط المدرج والمضلع والمنحنى.

الشكل (1 –1): التوزيع التكراري البسيط المدرج والمضلع والمنحنى

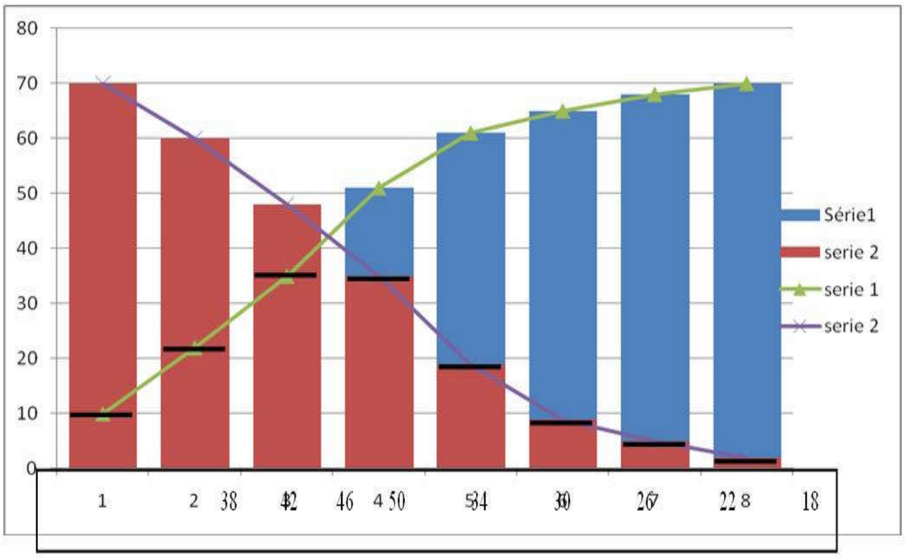


ثم يتم تحديد شكل المنحنى (ملتوي جهة اليمين أي موجب الالتواء ، متماثل، ملتوي جهة اليسار أي سالب الالتواء).

أما في حالة التوزيع التكراري المتجمع الصاعد أو النازل فيكون الشكل كالتالي:

وبنفس الخطوات السابقة للخطين الأفقي والعمودي فقط تتبدل القيمة العليا للخط العمودي فتصبح 70 لأنه تكرار صاعد إلى 70 أو نازل من 70، فيقسم الخط حيث يستحسن أن يبدأ من 0 إذا كانت القيمة الدنيا تقترب من 0.

الشكل (1− 2): التوزيع التكراري للمتجمع الصاعد والنازل



أما لرسم التكرارات المتجمعة النسبية (البسيطة، الصاعدة والنازلة) في عرضها البياني (مدرج، مضلع ومنحنى)، نتبع نفس الخطوات من بداية هذه الصفحة، حيث تبقى قيم الخط الأفقي كما هي وتتبدل القيمة العليا للخط العمودي في كل التوزيعات النسبية لتصبح نفسها قيمة التكرار الأعلى في الجدول (حيث أن القيمة العليا للتكرار النسبي البسيط هي قسمة أكبر تكرار بسيط على مجموع التكرارات، أما التكرارات النسبية الصاعدة والنازلة فإن أعلى قيمة لها هي 1)، وكل قيمة لتكرار متجمع **صاعد** تقابله **قيمة الحد الأدنى** لفئته، وكل قيمة لتكرار متجمع **نازل** تقابله **قيمة الحد الأعلى** لفئته.

الفصل الثاني :

مقاييس النزعة المركزية

الفصل الثاني : مقاييس النزعة المركزية

لا يكفى أسلوب العرض الجدولي والبياني لوصف ظاهرة ما وتحليل معطياتها إذ تميل البيانات إلى التمركز حول قيمة معينة وكلما ابتعدنا عن تلك القيمة تبدأ البيانات في التناقص (تسمى هذه الظاهرة بالنزعة المركزية)، لذا وضعت مقاييس إحصائية يمكن من خلالها التعرف على خصائص الظاهرة محل البحث وكذلك إمكانية مقارنة ظاهرتين أو أكثر، وتحليل المعطيات والتنبؤ واتخاذ القرار، وتسمى هذه المقاييس مقاييس النزعة المركزية، وتسمى كلها المتوسطات أو مقاييس الموضع وهي القيم التي تتركز جميع القيم حولها أو تنزع إليها:

1) الوسط الحسابي (arithmetic mean) $\overline{X}$:

يعرَّف الوسط الحسابي لمجموعة من القيم بأنه **مجموع** هذه القيم مقسوما على **عدد هذه القيم**:

حيث وإذا كانت لدينا قيم المجتمع فإن وسطها الحسابي هو $\boxed{u = \frac{\sum X_i}{N}}$، أما عينة فيها مجموعة من القيم في شكل سلسلة **غير مبوبة (ليست في فئات)** فإن الوسط الحسابي لهذه القيم هو:

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

مثال : فلو كان لدينا درجات 12 طالبا في مقياس الإحصاء، كما يلي: 10 15 13 17 12 14 11 16 14 18 18 10 فإن الوسط الحسابي سيكون:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{10+15+13+17+12+14+11+16+14+18+18+10}{12} = 14$$

أي أن الوسط الحسابي لدرجة الطالب (كمتغير) في مقياس الإحصاء هو 14 نقطة.

أما قيم التوزيع التكراري أي **الموضوعة في فئات** فلا يمكننا معرفة القيم الأصلية لها، لذا يتم التعبير عن كل قيمة تقع ضمن حدود أي فئة بمركز هذه الفئة (أي مركز هذه الفئة هو قيمة تقديرية لكل مفردة تقع ضمن هذه الفئة).

مثال : فيما يلي أوزان 40 طالب والمطلوب إيجاد الوسط الحسابي:

وزن الطالب	57-51	63-57	69-63	75-69	81-75	87-81
عدد الطلبة	3	7	10	9	9	2

ويحسب الوسط الحسابي بالخطوات التالية:

- إيجاد مجموع التكرارات (والذي هو غالبا يساوي حجم العينة).
- حساب مراكز الفئات x_i.
- ضرب مركز كل فئة في التكرار المناظر لتلك الفئة (xf)، وحساب المجموع ($\sum xf$).
- حساب الوسط الحسابي بتطبيق المعادلة التالية:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i = n} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k = n}$$

حيث: (k هو عدد الفئات)، ($x_1\ x_2 \dots x_k$ هي مراكز هذه الفئات)، ($f_1\ f_2 \dots f_k$ هي تكرارات القيم في هذه الفئات). ويتم حساب الوسط الحسابي كما يلي:

فئات الوزن	التكرارات f_i	مراكز الفئات x_i	$x_i f_i$	$(x_i - \overline{X})$
57-51	3	54=2/(57+51)	0112	-15.00
63-57	7	60=2/(63+57)	0420	-9.00
69-63	10	66	0660	-3.00
75-69	9	72	0648	3.00
81-75	9	78	0702	3.00
87-81	2	84	0168	15.00
المجموع	$\sum_{i=1}^{6} f_i = n = 40$		$\sum x_i f_i = 2760$	$00 \sum (x_i - \overline{X}) =$

إذن الوسط الحسابي $\overline{X}$ لوزن الطالب هو: $\overline{X} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i = n} = \frac{2760}{40} = 69$

1-1) خصائص الوسط الحسابي:

- الوسط الحسابي للقيمة الثابتة يساوي القيمة نفسها: فلو اخترنا مجموعة من 6 طلاب عشوائيا ووجدنا أن وزن كل واحد منهم يساوي 63 فإن:

$$\overline{\mathrm{X}} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{63+63+63+63+63+63+63}{6} = \frac{378}{6} = 63$$

- مجموع انحرافات القيم عن وسطها الحسابي معدوم: $\sum (x_i - \overline{\mathrm{X}}) = 0$، فلو راجعنا مثال درجات الطلبة أعلاه لوجدنا أن:　حيث لدينا الوسط الحسابي هو 14:

درجات الطالب x_i	10	15	13	17	12	14	11	16	14	18	18	10
الانحرافات $(x_i - \overline{\mathrm{X}})$	10-14=-4	15-14=1	1-	3	2-	0	3-	2	0	4	4	4-

ونجد أن: $\sum (x_i - \overline{\mathrm{X}}) = 0$، وكذلك في مثال أوزان الطلبة.

- إذا عدّلنا جميع القيم حيث نضيف مقدارا ثابتا a لكل منها فإن الوسط الحسابي للقيم المعدلة $\overline{\mathrm{Y}}$ يساوي الوسط الحسابي للقيم الأصلية قبل التعديل $\overline{\mathrm{X}}$ مضافا إليها ذلك المقدار الثابت $\boxed{\overline{\mathrm{Y}} = \overline{\mathrm{X}} + a}$
- إذا عدّلنا جميع القيم حيث ضربنا مقدارا ثابتا a في كل منها فإن الوسط الحسابي للقيم المعدلة $\overline{\mathrm{Y}}$ يساوي الوسط الحسابي للقيم الأصلية قبل التعديل $\overline{\mathrm{X}}$ مضروبا في ذلك المقدار الثابت $\boxed{\overline{\mathrm{Y}} = \overline{\mathrm{X}} \times a}$.

2-1) الوسط الحسابي المرجح:

غالبا ما تختلف أهمية قياسات المتغير الإحصائي من قيمة لأخرى ولكل منها أهمية نسبية أو وزن نسبي، لذا أدخل ما يسمى بالترجيح في علاقة الوسط الحسابي وأحسن مثال على ذلك متغير نقاط الطلبة في 5 مقاييس مثلا: 12 16 13 14 17، وكانت معاملاتها على التوالي: 2 1 3 1 2.
فإن وسطها الحسابي دون الترجيح سيكون: (12+16+13+14+17)/5 = 14.4 وبإدخال الترجيح فإننا سنضرب قيمة كل مقياس في وزنه أو أهميته النسبية (المعاملات) ونقسم مجموعها على مجموع

الأوزان النسبية (المعاملات)، فإن الوسط الحسابي المرجح $\overline{W}$ هو:

$14.11 = (2+1+3+1+2) \div ((2\times17)+(1\times14)+(3\times13)+(1\times16)+(2\times12))$

ونلاحظ أن الوسط الحسابي المرجح أكثر دقة ويحسب بالمعادلة التالية:

$$\boxed{\overline{W} = \frac{\sum x_i w_i}{\sum w_i}}.$$

1-3) مزايا وعيوب الوسط الحسابي:

يتميز الوسط الحسابي بسهولة الحساب، يأخذ كل القيم في الاعتبار، أنه أكثر المقاييس استخداما وفهما. ومن عيوبه:

- يتأثر بالقيم المتطرفة: كالدرجات 3 4 5 **20** أو الأعمار 10 11 13 **49** **55** . إذ لا يبين القيمة الحقيقية لها.
- يصعب حسابه في حالة البيانات الوصفية.
- يصعب حسابه في حالة الجداول التكرارية المفتوحة.

يطلب من الطالب البحث عن الوسط الهندسي والوسط التوافقي (والمرجح لكل منهما).

2) المنوال (mode):

يعرف المنوال بأنه القيمة الأكثر تكرارا في البيانات ويكثر استخدامه في حالة البيانات الوصفية لمعرفة الحالة أو المجموعة أو النمط أو المستوى الأكثر تكرارا أو شيوعا، ويتم حسابه كما يلي:

1-2) المنوال للبيانات غير المبوبة (Mod):

المنوال (Mod) = القيمة (المستوى أو الحالة) الأكثر تكرارا.

مثال : اختيرت عينة عشوائية بسيطة من طلبة كلية الآداب والعلوم الإنسانية، وتم رصد درجات هؤلاء الطلاب في مقياس معين، وكانت النتائج كالتالي: حيث طلب منكم إيجاد منوال الدرجات.

القسم	الدرجات										المنوال= القيمة الأكثر تكرارا
قسم الأدب العربي	67	58	70	65	77	77	77	75	77	80	المنوال هو 77 وتكرر 4 مرات
قسم علم الاجتماع	90	95	85	77	65	93	75	60	68	88	لا يوجد منوال
قسم الإنجليزية	80	86	65	76	88	65	80	69	65	80	المنوال هو 65 وتكرر 3 مرات المنوال هو 80 وتكرر 3 مرات
قسم الفرنسية	85	72	73	69	69	73	85	69	73	85	المنوال هو 69 وتكرر 3 مرات المنوال هو 73 وتكرر 3 مرات المنوال هو 85 وتكرر 3 مرات

2-2) المنوال للبيانات المبوبة (Mod):

ويحسب بطريقة الفروقات: (حيث أن فئة المنوال هي الفئة الأكثر أو الأعلى تكرارا) ومعادلة المنوال هي:

$$\boxed{\text{Mod} = A + \frac{d_1}{d_1 + d_2} \times L}$$

حيث : A هي الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار).

d_1 هو الفرق الأول = (تكرار فئة المنوال – التكرار السابق).

d_2 هو الفرق الثاني = (تكرار فئة المنوال – التكرار اللاحق).

L هو طول الفئة.

مثال: فيما يلي توزيع 30 أسرة حسب الإنفاق الشهري لها **بالدينار**: أحسب منوال الإنفاق الشهري للأسرة:

فئات الإنفاق الشهري دج	5000-2000	8000-5000	11000-8000	14000-11000	17000-14000
عدد الأسر لكل فئة	04	07	10	05	04

الحل: لحساب المنوال نتبع طريقة الفروقات ومعادلة المنوال حيث يتم أولا تحديد ما يلي:

- تحديد الفئة المنوالية وهي الفئة المناظرة لأكبر تكرار (أكبر تكرار 10 إذن فئة المنوال 8000-11000).
- حساب الفروق:

الفرق الأول = d_1= (تكرار فئة المنوال – التكرار السابق) = 10-7 = **3**

الفرق الثاني = d_2= (تكرار فئة المنوال – التكرار اللاحق) = 10-5 = **5**

- تحديد الحد الأدنى للفئة المنوالية (A=8)، وتحديد طول الفئة (3 = 8-11 =L).
- وبتطبيق معادلة المنوال في حالة البيانات المبوبة حيث : d_1=3، d_2=5، A=8 و L=3، فنجد:

$$Mod = A+\frac{d_1}{d_1+d_2}\times L = 8+\frac{3}{3+5}\times 3 = 8+\frac{3}{8}\times 3 = 8+\frac{9}{8} = 9.125$$ إذن القيمة الأكثر إنفاقا من طرف الأسر

وهي المنوال: $\boxed{9125}$ دج.

3) الوسيط (median):

هو القيمة التي تقسم بيانات المجتمع الإحصائي إلى قسمين متساويين بحيث تُرَتَّب قيم المتغير الإحصائي تصاعديا أو تنازليا، (أي القيمة التي يقل عنها نصف عدد القيم $(\frac{n}{2})$ ويزيد عنها النصف الآخر $(\frac{n}{2})$، وبتعبير آخر 50% من القيم أقل من الوسيط و 50% الأخرى من القيم أكبر من الوسيط).

3-1) الوسيط للبيانات غير المبوبة (Med):

- تُرَتَّب السلسلة تصاعديا أو تنازليا.
- تُحَدَّد رتبة أو ترتيب الوسيط وهي:

. إذا كان n عددا فرديا فإن قيمة الوسيط هي القيمة ذات الرتبة رقم

. $\boxed{\frac{n+1}{2}}$: $\boxed{\text{رتبة الوسيط}= (\frac{n+1}{2})}$.

. إذا كان n عددا زوجيا فإن قيمة الوسيط في الأصل هي $\boxed{\frac{n+1}{2}}$ وبما أن n عدد زوجي فإن قيمة الوسيط غير محددة لذا نجدها بتحديد الوسط الحسابي للقيمتين اللتان تحملان الرتبتين ($\boxed{\frac{n}{2}}$) و ($\boxed{\frac{n}{2}+1}$).

مثال : قسمت أرض زراعية إلى 17 وحدة، وتمت زراعتها بمحصول القمح باستخدام نوعين من السماد (حيث جرِّب النوع أ على 7 وحدات، والنوع ب على 10 وحدات) وبعد انتهاء الموسم تم تسجيل الإنتاج التالي (بالطن):

إنتاجية الوحدات المزروعة بالنوع أ: 12 27.5 32.5 20 30 23 15.

إنتاجية الوحدات المزروعة بالنوع ب: 45 18 35 37.5 20 25 15 40 25 30

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم.

الحل: لحساب وسيط الإنتاج من النوع أ:

- ترتيب القيم تصاعديا (12 15 20 23 27.5 30 32.5)، وعددها ($n_1=7$) عدد فردي.
- رتبة الوسيط هي $4=\frac{7+1}{2}=\frac{n+1}{2}$. وقيمته هي القيمة التي تحمل الرتبة رقم 4 وهي: Med_1 = 23.

لحساب وسيط الإنتاج من النوع ب:

- ترتيب القيم تصاعديا (15 18 20 25 25 30 35 37.5 40 45)، وعددها ($n_2=10$) عدد زوجي.
- رتبة الوسيط هي $5.5=\frac{10+1}{2}=\frac{n+1}{2}$، ولتحديده فهو يساوي **وسط** القيمتين ذات الرتبتين ($\boxed{\frac{n}{2}}$) و($\boxed{\frac{n}{2}+1}$) أي (الرتبتين 5 و6) وقيمتيهما هما (25 و30) ووسطهما هو $\frac{25+30}{2}=27.5$ إذن الوسيط: $Med_2 = 27.5$.

3-2) الوسيط للبيانات المبوبة (Med):

ولحساب الوسيط للبيانات المبوبة أي الموضوعة في فئات يتم إتباع ما يلي:

- تكوين الجدول التكراري المتجمع الصاعد أو النازل.

- تحديد رتبة أو ترتيب الوسيط: وهي نصف مجموع التكرارات $\boxed{\frac{\sum f_i}{2}=\frac{n}{2}}$.
- تحديد فئة الوسيط وهي الفئة التي يقع فيها الوسيط، أو هي التي تقابل التكرار التجميعي الذي يساوي ترتيب الوسيط أو التكرار الأكبر منه مباشرة.
- حساب الوسيط بالمعادلة التالية: $\boxed{Med = \mathrm{A} + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L}$

حيث:

A هو الحد الأدنى لفئة الوسيط.

f_2, f_1 التكرارين المتجمعين الذين تقع بينهما رتبة الوسيط.

L هو طول الفئة.

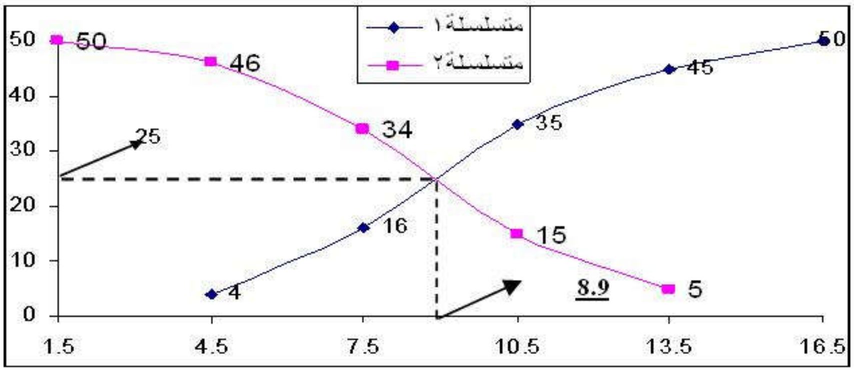
<u>مثال</u> : فيما يلي توزيع 50 فرد حسب الساعات التي يقضونها أمام التلفاز، أحسب الوسيط: حسابيا وبيانيا.

16.5-13.5	13.5-10.5	10.5-7.5	7.5-4.5	4.5-1.5	فئة الحجم الساعي
5	10	19	12	4	عدد الأشخاص

<u>الحل</u>: تكوين الجدول التكراري المتجمع الصاعد.

$f\downarrow$	$f\uparrow$	عدد الأشخاص f	فئة الحجم الساعي
50	04	04	4.5-1.5
46	16	12	7.5-4.5
34	35	19	10.5-7.5
15	45	10	13.5-10.5
05	50	05	16.5-13.5
		50	المجموع

- رتبة الوسيط هي: $\frac{\sum f_i}{2} = \frac{n}{2} = \frac{50}{2} = 25$.
- الفئة التي تشمل قيمة الوسيط، هي التي تقابل تكرار متجمع المساوي لرتبة الوسيط 25 أو أكبر منها مباشرة وهو 35 وفئة هذا التكرار هي فئة الوسيط وهي 7.5–10.5.
- الحد الأدنى لفئة الوسيط هو $A = 7.5$.
- التكرارين الصاعدين الذين تقع بينهما رتبة الوسيط هما: $f_1 = 16, f_2 = 35$. وطول الفئة: $L = 10.5 - 7.5 = 3$.
- إذن قيمة الوسيط هي: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{\frac{50}{2} - 16}{35 - 16} \times 3 = 7.5 + \frac{9}{19} \times 3 = 8.92$.
- تحدد قيمة الوسيط بيانيا برسم منحنى بياني للتكرار المتجمع الصاعد أو النازل، وتحدد الرتبة على المحور العمودي ورسم خط أفقي منها حتى يلمس المنحنى في نقطة ومن هذه النقطة نسقط خط عمودي ليقطع المحور الأفقي في نقطة معينة هي قيمة الوسيط.



3-3) مزايا وعيوب الوسيط: يتميز الوسيط بأنه لا يتأثر بالقيم المتطرفة، كما أنه سهل الحساب.

ومن عيوبه أنه يعتمد على القيمة الوسطية أو القيمتين الوسطيتين ولا يأخذ كل القيم في الاعتبار.

4) الرباعيات العشيريات والميئيات (quartiles):

عند تقسيم قيم العينة بعد ترتيبها إلى 4 أجزاء متساوية تنتج لنا 3 إحصاءات ترتيبية تسمى الرباعيات Q_i.

عند تقسيم قيم العينة بعد ترتيبها إلى 10 أجزاء متساوية تنتج لنا 9 إحصاءات ترتيبية تسمى العشيريات D_i.

عند تقسيم قيم العينة بعد ترتيبها إلى 100 جزء متساوي تنتج لنا 99 إحصائية ترتيبية تسمى الميئيات P_i .

- الربيع الأول Q_1 هي القيمة التي تقل عنها 25% من القيم.
- الربيع الثاني Q_2 هي القيمة التي تقل عنها 50% من القيم.
- الربيع الثالث Q_3 هي القيمة التي تقل عنها 75% من القيم.

4-1) حساب الربيعيات، العشيريات والميئيات في حالة البيانات غير المبوبة كما يلي:

- ترتيب القيم تصاعديا أو تنازليا.
- تحدد رتبة الربيع رقم i (Q_i) $\boxed{R=(n+1)\times(\frac{i}{4})}$. حيث تمثل i رقم الربيع (الأول أو الثاني أو الثالث).
- إذا كان R عددا صحيحا فإن قيمة الربيع هي: $Q_i = X_R$.
- إذا كان R عددا كسريا فإن قيمة الربيع هي: $\boxed{Q_i = x_K + (R-K)\times(x_U - x_K)}$

حيث: **Q_i** هو الربيع حيث **i** هي رقم الربيع.

R هي رتبة الربيع.

K هي الرتبة التي تسبق رتبة الربيع، (مثلا إذا كان R=3.25 فإن K=3).

U هي الرتبة التي تلي رتبة الربيع، (مثلا إذا كان R=3.25 فإن U=4).

X_U X_K هما قيمتين تقع **(R)** رتبة الربيع بين رتبتيهما **(K U)** وبالتالي فإن قيمة الربيع Q_i تقع بين القيمتين **(X_U و X_K)**.

وبنفس كيفية حساب الربيعيات نحسب العشيريات حيث نستبدل Q بـ: D، و i تأخذ القيمة من 1 إلى 9 وتتبدل فقط رتبة العشير وتصبح: $\boxed{R=(n+1)\times(\frac{i}{10})}$ وقيمته هي: $\boxed{D_i = x_K + (R-K)\times(x_U - x_K)}$.

وبنفس كيفية حساب الربيعيات نحسب الميئيات حيث نستبدل Q بـ: P، و i تأخذ القيمة من 1 إلى 99 وتتبدل فقط رتبة الميئين وتصبح: $\boxed{R=(n+1)\times(\frac{i}{100})}$ وقيمته هي: $\boxed{P_i = x_K + (R-K)\times(x_U - x_K)}$.

<u>مثال للبيانات غير المبوبة</u>:فيما يلي كمية إنتاج عشر بقرات من الحليب باللتر يوميا: حيث اختيرت هذه الأبقار من مزرعة معينة: 25 23 29 32 34 29 20 18 27 30

أحسب الوسيط الربيعي الثالث (ما هي القيمة التي تقل عنها 75% من كمية الإنتاج؟)، والوسيط العشيري السابع (ما هي القيمة التي تقل عنها 70% من كمية الإنتاج؟)، والوسيط الميئي الثالث والستون (القيمة التي تقل عنها 63% من كمية الإنتاج؟).

<u>الحل</u>: أولا نرتب القيم تصاعديا:18 20 23 25 27 29 29 30 32 34

- رتبة الوسيط الربيعي الثالث: $R=(n+1)\times(\frac{i}{4})=(10+1)\times\frac{3}{4}=8.25$ ،

إذن الرتبة K التي تسبق رتبة الربيع R=8.25 هي: K=8. والقيمة التي تسبق قيمته هي $X_K= X_8=30$.

إذن الرتبة U التي تلي رتبة الربيع R=8.25 هي: U=9. والقيمة التي تلي قيمته $X_U= X_9=32$.

إذن قيمة الربيع: $Q_3=x_K+(R-K)\times(x_U-x_K)=30+(8.25-8)\times(32-30)=30.5$

- رتبة الوسيط العشيري السابع: $R=(n+1)\times(\frac{i}{10})=(10+1)\times\frac{7}{10}=7.7$ ،

وقيمته: $D_i=x_K+(R-K)\times(x_U-x_K)=29+(7.7-7)\times(30-29)=29.7$

- رتبة الوسيط الميئي الثالث والستون:

$R=(n+1)\times(\frac{i}{100})=(10+1)\times\frac{63}{100}=6.3$ ،

وقيمته: $P_i=x_K+(R-K)\times(x_U-x_K)=29+(6.3-6)\times(29-29)=29$

4-2). حساب الرباعيات، العشيريات والميئيات في حالة البيانات المبوبة كما يلي:

- بعد تكوين الجدول التكراري المتجمع الصاعد والذي يمثل رتب القيم X_i.
- تحديد رتبة الربيع أو العشير أو الميئين ووضعها في المعادلة كما يلي:

– والمعادلة هي نفسها معادلة الوسيط و نفس الخصائص الأخرى:

$$\boxed{Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L}$$ ، إلا أننا:

- لحساب الربيع نستبدل $\left(\frac{n}{2}\right)$ بـ: $\left(\frac{n}{4}\times i\right)$ وهي الرتبة، ونستبدل Med بـ: Q_i حيث: i يتغير من 1 إلى 3.
- ولحساب العشير نستبدل $\left(\frac{n}{2}\right)$ بـ: $\left(\frac{n}{10}\times i\right)$ وهي الرتبة، ونستبدل Med بـ: D_i حيث: i يتغير من 1 إلى 9.
- ولحساب الميئين نستبدل $\left(\frac{n}{2}\right)$ بـ: $\left(\frac{n}{100}\times i\right)$ وهي الرتبة، ونستبدل Med بـ: P_i حيث: i يتغير من 1 إلى 99.

– وفئة (Q_i، D_i، P_i) : هي التي تقابل التكرار المتجمع المساوي لرتبة(Q_i، D_i، P_i)، أو التكرار الأكبر منه مباشرة

– حيث: A : هو الحد الأدنى لفئة الربيع أو العشير أو الميئين.

f_1, f_2 : هما التكرارين المتجمعين الذين تقع بينهما رتبة الربيع أو العشير أو الميئين.

L : هو طول الفئة.

ملاحظة: ستجدون أن:

الوسيط العادي = الربيع الثاني = الوسيط العشيري الخامس = الوسيط الميئي الخمسون.

مثال للبيانات غير المبوبة: ليكن التوزيع التكراري التالي:

$x_i f_i$	$f\uparrow$	f	مركز الفئة		
6	3	3	2	4-	-0
36	9	6	6	8-	-4
72	16	7	9	10-	-8
99	25	9	11	12-	-10
224	41	16	14	16-	-12
234	54	13	18	20-	-16
184	62	8	23	26-	-20
140	67	5	28	30-	-26
105	70	3	35	40-	-30
1091		70			

- عين القيمة التي تقسم المجتمع الإحصائي إلى نصفين.
- عين الربيع: الأول والثالث، والعشير: الرابع والسادس.
- حدد الوسيط المبئي: السادس والثلاثون(36) والثالث والثمانون(83).

الحل: قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى نصفين هي قيمة الوسيط وهي:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 12 + \frac{35 - 25}{41 - 25} \times 4 = 14.5$$

- القيمة التي تقل عنها 25% من القيم هي الربيع الأول:

$$Q_1 = A + \frac{\frac{n}{4} \times 1 - f_1}{f_2 - f_1} \times L = 10 + \frac{17.5 - 16}{25 - 16} \times 2 = 10.3 .$$

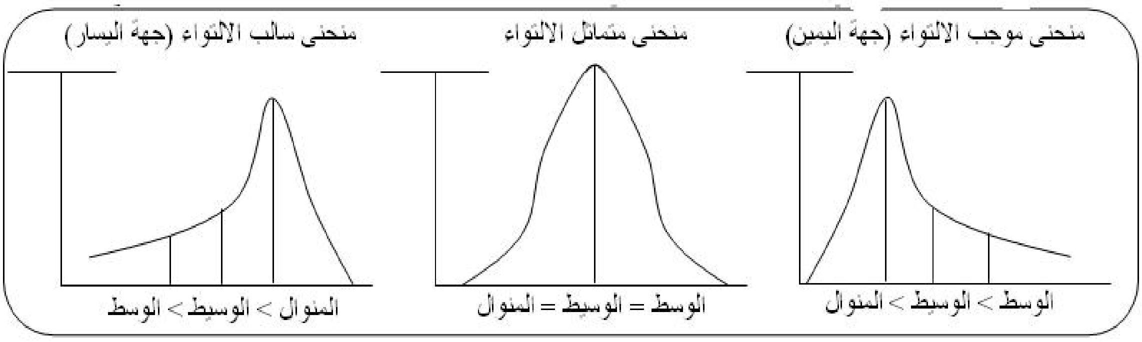
- القيمة التي تقل عنها 75% من القيم هي الربيع الثالث:

$$Q_3 = A + \frac{\frac{n}{4} \times 3 - f_1}{f_2 - f_1} \times L = 16 + \frac{52.5 - 41}{54 - 41} \times 4 = 19.54$$

- العشير الرابع: $D_4 = A + \frac{\frac{n}{10} \times 4 - f_1}{f_2 - f_1} \times L = 12 + \frac{28 - 25}{41 - 25} \times 4 = 12.75$
- العشير السادس: $D_6 = A + \frac{\frac{n}{10} \times 6 - f_1}{f_2 - f_1} \times L = 16 + \frac{42 - 41}{54 - 41} \times 4 = 16.3$
- المبئي السادس والثلاثون: $P_1 = A + \frac{\frac{n}{100} \times 36 - f_1}{f_2 - f_1} \times L = 12 + \frac{25.2 - 25}{41 - 25} \times 4 = 12.05$
- المبئي الثالث والثمانون: $P_1 = A + \frac{\frac{n}{100} \times 83 - f_1}{f_2 - f_1} \times L = 20 + \frac{58.1 - 54}{54 - 41} \times 6 = 21.89$

استخدام مقاييس النزعة المركزية لتحديد شكل البيانات: يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري ويعبر عن شكل توزيع البيانات كما يلي:

توزيع البيانات



أسباب اختيار أحد مقاييس النزعة المركزية دون غيرها:

نختار أحد مقاييس النزعة حسب طبيعة وهدف الدراسة ونوع العلاقة الموجودة بين الخصائص المدروسة:

- عند استيراد مادة استهلاكية أساسية فإننا نبحث عن **الوسط الحسابي** للاستهلاك لتحديد الكمية المستوردة.
- لمعرفة الكتل البرلمانية ذات الأغلبية المطلقة أو النسبية في نتائج الانتخابات البرلمانية وذلك باستعمال المنوال أي القيمة الأكثر تكرارا أو انتشارا، وكذلك لإنتاج نوع معين من الأحذية ننتج النوع الأكثر استعمالا(**المنوال**).
- لتقسيم المجتمع إلى طبقتين أو أكثر حسب الدخل نستعمل الوسيط العادي أو أحد أنواعه (ربيع، عشير ميئي).
- ويكثر استعمال الوسيط في حالة التوزيعات التكرارية المفتوحة لعدم تأثره بالقيم المتطرفة.
- لرسم التوزيع يفضل استعمال الأعمدة البيانية عند عدم تساوي أطوال الفئات.

أمثلة محلولة

مثال (1) :

استغرقت مفاوضات السلام بين بلدين خمس جولات، وكانت كل جولة تستغرق عدة أيام كما يلي :

7, 10, 12, 8, 9

أحسب الوسط الحسابي لعدد الأيام في هذه الجولات.

الحل :

1 – لدينا خمس حولات أو خمس قيم، أي أن عدد القيم = 5 جولات.

2 – مجموع الأيام أو مجموع القيم هو :

7 + 10 + 12 + 8 + 9 = 46 يوماً

3 – الوسط الحسابي = $\frac{\text{مجموع هذه القيم}}{\text{عددها}} = \frac{46}{5} = 9.2$ يوماً

أي أن متوسط عدد الأيام في هذه الجولات من المفاوضات هو 9.2 يوماً وللباحث السياسي بعدها حرية إعطاء التفسير لطول أو قصر هذه المدة.

مثال (2) :

أخذت عينة عشوائية لعدد من سكان أحد الأحياء الفقيرة في دولة نامية حجمها 10 أشخاص، وكانت دخولهم اليومية بالدولار هي :

3.6 , 4.2 , 2.9 , 3.7 , 4.8 , 2.5 , 3.1 , 3.9 , 3.4 , 4.5

أحسب الوسط الحسابي لدخول هؤلاء الأشخاص.

الحل :

1 – عدد الأشخاص يساوي 10 أي أن

$n = 10$

2 – مجموع القيم (مجموع دخولهم اليومية) هو :

$$\sum X = 3.6 + 4.2 + 2.9 + 3.7 + 4.8 + 2.5 + 3.1 + 3.9 + 4.5$$
$$\sum X = 36.6$$

3 – الوسط الحسابي لدخول هؤلاء الأشخاص هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{36.6}{10} = 3.66$$

وهو معدل يعكس بلا شك مجموعة من الحقائق قد يكون أهمها الصعوبات الاقتصادية التي تواجه هذه الدولة. ومن الشرح والأمثلة السابقة يتضح ما يلي :

مثال (3) :

إذا كانت أعداد الطلعات الجوية التدريبية الشهرية للمقاتلات الحربية في بلد ما لست من مقاتلاتها هي :

10 , 81 , 84 , 83 , 82 , 80

أحسب الوسط الحسابي الشهري لأعداد هذه الطلعات.

الحل :

قبل الشروع في الحل نلاحظ أن عدد هذه الطلعات متقاربة جداً وكلها بأعداد عالية (80 فأكثر) باستثناء طائرة واحدة بلغ عدد طلعاتها (10) طلعات فقط. وهي تمثل قيمة شاذة (أو متطرفة) بالنسبة لباقي طلعات الطائرات الأخرى (حيث أنها صغيرة جداً بالنسبة للطلعات الأخرى). وسوف نرى فيما يلي تأثيرها على قيمة الوسط الحسابي.

1 – عدد القيم (عدد الطلعات) هو 6 أي أن :

$$n = 6$$

2 – مجموع القيمة (أي مجموع الطلعات) هو :

$$\sum X = 80 + 82 + 83 + 84 + 10$$

$$\sum X = 420$$

3 – الوسط الحسابي لعدد الطلعات هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{420}{6} = 70$$

فكأن الوسط الحسابي لعدد الطلعات قد أنخفض إلى 70 طلعة على الرغم من أن كل الطائرات (باستثناء طائرة واحدة شاذة) كانت **80** فاكثر.

مثال (4) :

إذا كانت لدينا مجموعتين من الطلاب تدرسان المقرر نفسه. وكان عدد الطلاب في المجموعتين هو :

$$n_1 = 40 \qquad n_2 = 25$$

وكان الوسط الحسابي لدرجات الطلاب في المجموعتين هو :

$\overline{X}_1 = 75 . \overline{X}_2 = 80$ فما هو الوسط الحسابي لدرجات الطلاب في المجموعتين معاً ؟

الحل :

المتوسط المرجح $\overline{X}$ يحسب باستخدام العلاقة رقم (2) :

الهاتف: 0021332120740 | 0021332120720 webmaster@univ-eloued.dz

$$\overline{X} = \frac{n_1 \,.\, \overline{x}_1 + n_2 \,.\, \overline{x}_2}{n_1 + n_2}$$
$$= \frac{40 \times 75 + 25 \times 80}{40 + 25}$$
$$\frac{3000 + 2000}{65} =$$
$$\frac{5000}{65} = 76.9$$

مثال (5) :

في المثال السابق إذا كان لدينا المتوسطان نفسهما، ولكن عدد الطلاب في كل من المجموعتين متساوي وليكن يساوي **40** في كل منهما :

$$\overline{X}_1 = 75 \ , \ \overline{X}_2 = 80$$
$$n_1 = n_2 = 40$$

أحسب الوسط المرجح للمجموعتين معاً ؟

الحل :

نلاحظ أن حجمي العينتين (أو المجموعتين) متساويان وبالتالي فإن المتوسط المرجح – في هذه الحالة الخاصة هو :

$$\overline{X} = \frac{\overline{x}_1 + \overline{x}_2}{2}$$
$$= \frac{75 + 80}{2} = 77.5$$

ولو استخدمنا العلاقة رقم (2) وهي الحالة العامة لحصلنا على النتيجة نفسها وذلك كما يلي :

$$\overline{X} = \frac{n_1 \,.\, \overline{x}_1 + n_2 \,.\, \overline{x}_2}{n_1 + n_2}$$
$$= \frac{40 \text{ x } 74 + 40 \, x \, 80}{40 + 40}$$
$$= \frac{3000 + 3200}{80}$$
$$= \frac{6200}{80} = 77.5$$

وهذه النتيجة نفسها بطبيعة الحال. أي أنه في حالة تساوي أحجام العينات فقط تجمع المتوسطات وتقسم على عددها. أما في حالة إختلاف احجام العينات فنحسب المتوسط المرجح (باستخدام العلاقة (2) أو (3) أو الحالة العامة لهما وذلك على حسب عدد العينات).

مثال (6) :

الجدول التالي يعطي عينة من الأسر تمثل نمط الإنجاب في واحدة من دول العالم الثالث، والمطلوب حساب الوسط الحسابي لعدد الأطفال في الأسرة في هذه العينة؟

عدد الأطفال بالأسر X	أعداد الأسر (التكرارات) F
4	2
5	6
6	10
7	15
8	9
9	5
10	3
المجموع	**50**

الحل :

أرقام الجدول أعلاه تقول أن العينة تحوي 50 أسرة منهم أسرتان بكل منهما 4 أطفال، 6 أسر بكل منهم 5 أطفال، 10 أسر لدى كل واحداً منهم 6 أطفال... وهكذا. ولكي نحسب الوسط الحسابي نحصل أولاً على مجموع الأطفال ثم نقسم على إجمالي عدد أفراد العينة (والذي يساوي في هذا المثال 50 أسرة). ولكي نحصل على مجموع عدد الأطفال نضرب عدد الأطفال في عدد الأسر ثم نجمع لكل الأسر. فمثلاً الأسرتان اللتان لدى كل منهما 4 أطفال مجموع أطفالهما = 2 × 4 = 8 والأسر الست التي لدى كل منهم 5 أطفال مجموع عدد أطفالهم 30 = 6 × 5 وهكذا بالنسبة لباقي الجدول. ويمكن تنظيم ذلك في عمود جديد يضاف إلى الجدول السابق كما يلي:

عدد الأطفال بالأسر	أعداد الأسر (التكرارات)	حاصل الضرب (مجموع الأطفال)
X	f	Xf
4	2	$4 \times 2 = 8$
5	6	$5 \times 6 = 30$
6	10	$6 \times 10 = 60$
7	15	$7 \times 15 = 105$
8	9	$8 \times 9 = 72$
9	5	$9 \times 5 = 45$
10	3	$10 \times 3 = 30$
المجموع	$\sum f = 50$	$\sum xf = 350$

ونعلم أن الوسط الحسابي لعدد الأطفال بالأسرة يساوي مجموع الأطفال على عدد الأسر، أي مجموع القيم على عددها. وبالتالي فإن المعادلة تصبح :

$$\boxed{\overline{X} = \frac{\sum xf}{\sum f}} \quad (4)$$

وبالتعويض في قانون الوسط الحسابي رقم (4) نحصل على الوسط الحسابي لحجم الأسرة كما يلي :

$$\overline{X} = \frac{350}{50} = 7$$

أي أن الوسط الحسابي لحجم الأسرة بعينة هذا البلد يساوي 7 أطفال، وهو متوسط مرتفع دون شك.

مثال (7) :

الجدول التالي يمثل توزيع مجموعة من الطلاب حسب فئات الدرجات كما يلي :

أعداد الطلاب F	فئات الدرجات Classes
3	2 – 4
9	4 – 6
10	6 – 8
5	8 – 10
$\sum f = 27$	**المجموع**

والمطلوب حساب الوسط الحسابي لدرجات الطلاب.

الحل :

الجدول يقول أن 3 طلاب حصل كل منهم على درجة تتراوح بين 2 **وأقل من** 4 (لكن لا نعلم ما درجة كل منهم بالتحديد)، 9 طلاب حصل كل منهم على درجة تتراوح بين 4 **وأقل من** 6 (لكن لا نعلم ما درجة كل منهم بالتحديد)، وهكذا بالنسبة لباقي الفئات. وفي هذه الحالة نحسب مراكز الفئات كأحسن قيم تمثل هذه الفئات. ومركز الفئة هو القيمة التي تقع في منتصف الفئة، أي أن :

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

أي أننا نستعيض عن الفئات بمراكزها وهي التي تمثل القيم (كما في الأمثلة السابقة) وسوف نرمز لها بالرمز X، ثم نكمل الحل كما في المثال السابق :

حاصل الضرب (مجموع الدرجات)	مراكز الفئات	أعداد الطلاب	فئات الدرجات
x. f	x	f	Classes
3 × 3 = 9	$\frac{2+4}{2} = 3$	3	2 – 4

$4-6$	9	$\frac{4+6}{2}=5$	$5 \times 9 = 45$
$6-8$	10	$\frac{6+8}{2}=7$	$7 \times 10 = 70$
$8-10$	5	$\frac{8+10}{2}=9$	$9 \times 5 = 45$
المجموع	$\sum f = 27$		$\sum xf = 169$

وبالتعويض في قانون الوسط الحسابي رقم (4) نحصل على :

$$\overline{X} = \frac{\sum xf}{\sum f} = \frac{169}{27} = 6.26$$

أي أن الوسط الحسابي يساوي 6.26 درجة.

مثال (8) :

البيانات التالية تمثل أعمار مجموعة من الناخبين :

32 24 20 35 29 فما هو وسيط العمر ؟

الحل :

أولاً : نرتب هذه الأعمار تصاعدياً كما يلي :

35 32 **29** 24 20

ثانياً : نلاحظ أن عدد القيم فردي (يساوي 5) وأنه توجد قيمة واحدة في المنتصف هي 29 وبالتالي فإن قيمة الوسيط تساوي 29 سنة.

مثال (9) :

البيانات التالية تمثل دخول بعض الأفراد اليومية بالدولار الأمريكي في إحدى الدول.

11 19 14 18 12 15 أحسب وسيط هذه الدخول ؟

الحل :

أولاً : نرتب هذه الدخول تصاعدياً كما يلي :

19 18 **15** **14** 12 11

ثانياً : نلاحظ أن عدد القيم زوجي (يساوي 6) وأنه توجد قيمتان في المنتصف هما 15، 14 لذلك نجمعهما ونقسم على 2. أي أن الوسيط يساوي :

$$\frac{14+15}{2}=14.5$$ دولاراً

مثال (10) :

4	6	**8**	9	10
4	6	**8**	9	100
4	6	**8**	9	1000

ونلاحظ أن قيمة الوسيط في الحالات الثلاث تساوي 8 (سواء كانت أكبر قيمة تساوي 10 أو 100 أو 1000) أي لم تتأثر قيمة الوسيط بوجود قيمة شاذة أو متطرفة.

مثال (11) :

البيانات التالية تمثل تقديرات بعض عينة مختارة من الناخبين لاحتمال فوز أحد المرشحين في أحد الانتخابات :

good , v.good , fair , good , excellent , fair , good

ولحساب وسيط هذه التقديرات نتبع الخطوات التالية :

الحل :

رغم أن البيانات غير كمية إلا أنها ترتيبية أي يمكن ترتيبها (تصاعدياً أو تنازلياً). وترتيبها تصاعدياً يكون كما يلي :

fair , fair , good , $\boxed{\text{good}}$, goo , v.good , excellent

وحيث أن التقدير good هو الذي يقع في منتصف التقديرات بعد ترتيبها تصاعدياً فإن وسيط التقديرات هو good أو جيد.

مثال (12) :

البيانات التالية تمثل أعمار مجموعة من الناخبين :

25 , 29 , 34 , 29 , 36 , 42 , 29 , 50 , 29 , 36

فما هو منوال هذه الأعمار ؟

الحل :

بما أن العمر 29 سنة هو العمر الذي تكرر أكثر من غيره من الأعمار (تكرر 4 مرات) فإن : منوال العمر = 29 سنة

(لاحظ أن البيانات في هذا المثال كمية)

مثال (13) :

البيانات التالية تمثل تقديرات مجموعة من الطلاب في أحد المقررات.

good, excellent , good , v.good , fair , good, fair

فما هو منوال هذه التقديرات ؟

الحل :

منوال التقديرات هو التقدير "good" لأنه تكرر أكثر من غيره **(تكرر ثلاث مرات)**.

(لاحظ أن البيانات في هذا المثال وصفية ترتيبية)

مثال (14) :

البيانات التالية تمثل توزيع فوج من السائحين لإحدى الدول حسب جنسياتهم :

الجنسية	عدد السائحين
ألمانية	50
فرنسية	80
أمريكية	120
إيطالية	90

من هذا الجدول نجد أن منوال الجنسية (أي الجنسية الشائعة أو التي تكررت أكثر من غيرها) **هي الجنسية الأمريكية** (120 سائحاً).

(لاحظ أن البيانات في هذا المثال وصفية اسمية Nominal).

بعض الملاحظات على المنوال :

1 – لاحظنا من الأمثلة السابقة أنه يمكن إيجاد المنوال لكل انواع البيانات (كمية أو ترتيبية أو أسمية).

2 – حسب تعريف المنوال قد **لا** تتكرر قيمة أكثر من غيرها، وبالتالي قد لا يوجد منوال لبعض البيانات.

مثال (15) :

فإذا كانت البيانات التالية تمثل أعمار مجموعة من الناخبين :

40 55 39 48 32 25

فإنه **لا** يوجد منوال لهذه الأعمار .

3 – وحسب تعريف المنوال أيضاً قد يوجد أكثر من منوال واحد للبيانات.

مثال (16) :

البيانات التالية تمثل توزيع مجموعة من الناخبين حسب أعمارهم.

	الأعمار	أعداد الناخبين	
	25	3	
	30	5	
المنوال الأول =	35	9	أكبر تكرار (التكرار المنوالي)
	40	4	
المنوال الثاني =	45	9	أكبر تكرار (التكرار المنوالي)
	50	2	

في هذا الجدول نلاحظ أن العمر 35 تكرر 9 مرات (وهو أكبر تكرار) وكذلك العمر 45 تكرر أيضاً 9 مرات (وهو أكبر تكرار) لذلك فإن :

المنوال الأول = 35 سنة والمنوال الثاني = 45 سنة.

مثال (17) شامل على المتوسطات :

الجدول التالي يمثل أهم الحروب التي شهدها العالم من عام 1945 وحتى عام 1980م. أحسب المتوسطات الثلاثة لهذه البيانات.

أهم الحروب التي شهدها العالم من عام 1945م وحتى عام 1980م

الرقم	العام	العدد	المكان
1	1945	4	سوريا – لبنان، أندونيسا، الصين، ماليزيا
2	1946	2	الهند الصينية، اليونان
3	1947	3	مدغشقر ، الهند والباكستان، كشمير
4	1948	4	الفلبين، الحرب العربية الإسرائيلية الأولى. حيدر اباد، بورما
5	1949	0	
6	1950	3	كوريا، فرموزا، التبت
7	1951	0	
8	1952	1	كينيا
9	1953	0	

الرقم	العام	العدد	المكان
10	1954	2	جواتيمالا، الجزائر
11	1955	0	السودان، قبرص
12	1956	3	سينا . هنغاريا، السويس
13	1957	0	
14	1958	2	لبنان، كوبا
15	1959	4	فيتنام، هملايا، راوندا، لاوس
16	1960	2	الكونغو ، كولومبيا
17	1961	3	كوبا (خليج الخنازير) جيو ، انغولا
18	1962	3	غرب غينيا الجديدة، اليمن، غينيا الأسبانية
19	1963	4	الجزائر – المغرب، قبرص، ماليزيا، الصومال – كينيا
20	1964	3	جنزيار ، تايلند، موزنبيق
21	1965	3	الهند – الباكستان، جمهورية الدومينيكان، أندونيسا
22	1966	1	بيافرا
23	1967	1	الحرب العربية الإسرائيلية الثانية
24	1968	1	تشيوكوسلفاكيا
25	1969	4	ماليزيا، السلفادور ، تشاد، شمال إيرلندا
26	1970	1	أثيوبيا (أرثيريا)
27	1971	2	كمبوديا – بانجلاديش / كشمير
28	1972	1	بروندا‌ي
29	1973	1	الحرب العربية الإسرائيلية الثالثة (حرب أكتوبر)
30	1974	2	العراق (الأكراد) قبرص
31	1975	3	أنغولا، تايمور ، لبنان
32	1976	1	أسبانيا / المغرب
33	1977	4	الصومال، أثيوبيا، أثيوبيا (ارثيريا)، سوريا – لبنان، ليبيا – مصر
34	1978	6	إيران، نيكاراغو ، فيتنام – لاوس، تشاد، زائير ، روديسيا (زمبابوي)
35	1979	6	اليمن الشمالية – اليمن الجنوبية، أوغندا – تنزانيا، الصين – فيتنام، فيتنام – كمبوديا، نيكاراغوا، جنوب أفريقيا – انغولا
36	1980	3	روسيا – أفغانستان، العراق – إيران، السلفادور
	المجموع	**85**	

الحل :

أولاً : حساب الوسط الحسابي :

1 – عدد القيم (عدد السنوات) $n=36$

2 – مجموع الحروب خلال تلك الفترة $\sum x = 85$

3 – الوسط الحسابي لعدد الحروب خلال تلك الفترة : $X = \frac{\sum x}{n} = \frac{85}{36} = 2.36$

أي أن الوسط الحسابي لعدد الحروب يساوي **2.36** حرباً في السنة الواحدة.

ثانياً : حساب الوسيط :

1 – ترتيب البيانات تصاعدياً (لاحظ أن العدد هو 36) أي ترتب عدد الحروب تصاعدياً :

0000 11111111 2222222 333333333 444444 66

2 – ترتيب الوسيط $\frac{n+1}{2}$

$$\frac{n+1}{2} = \frac{36+1}{2} = \frac{37}{2} = 18.5$$

أي أن الوسيط يقع بين القيمتين الثامنة عشرة والتاسعة عشرة.

3 – قيمة الوسيط :

(نجمع القيمتين الثامنة عشرة والتاسعة عشرة ونقسم على اثنين)

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

أي أن وسيط الحروب يساوي (2) حرباً في السنة.

ثالثاً : حساب المنوال :

يعرف المنوال بأنه القيمة التي تكررت أكثر من غيرها وحيث أن القيمة 3 هي التي تكررت أكثر من غيرها (تكررت تسع مرات) فإن منوال الحروب يساوي (3) حروب في السنة. لاحظ أن المتوسطات الثلاثة ليست بالضرورة متساوية.

ملاحظة :

أبحث عن المكان الذي تكرر أكثر من غيره خلال تلك الفترة. أي منوال المكان (أو الدولة أو المنطقة).

الفصل الثالث

مقاييس التشتت والشكل

الفصل الثالث : مقاييس التشتت والشكل

1- مقدمة:

لقد ذكرنا في الفصل السابق بعضًا من مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما. وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة. وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتت البيانات.

لتكن قيم توزيعين كما يلي:

التوزيع الأول: 65 70 71 73 76 70 70 65

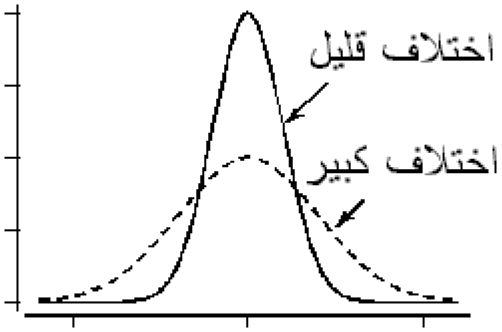
(المدى = 76-65=11)

التوزيع الثاني: 46 54 70 91 110 89 70 55 45

(المدى = 110-45 = 65)

عند حساب الوسط الحسابي والوسيط والمنوال لكلا التوزيعين نلاحظ تساوي الخصائص الثلاثة للنزعة المركزية في كلا التوزيعين $\overline{X}$ =Med=Mod= 70، إلا أننا نلاحظ اختلافا كبيرا في انتشار القيم وتوزيعها على مجال الدراسة واختلاف طوله في التوزيعين (المدى العام 65 و 11)، أي أن التوزيع الثاني أكثر تشتتا من التوزيع الأول بالنسبة للقيمة المركزية، ولقياس هذا التشتت واستكمال دراسة التوزيع الإحصائي نتطرق في هذا الدرس إلى أكثر **المقاييس** استعمالا في تبيان كيفية توزيع وانتشار قيم المتغير الإحصائي حول القيمة المركزية **مقاييس التشتت**، ونتطرق إلى **مقاييس التشكل** التي تبين تناظر وعدم تناظر التوزيع الإحصائي بالنسبة للقيمة المركزية.

الشكل التالي يوضح المضلعين التكراريين لمجموعتين من البيانات لهما نفس مقاييس النزعة المركزية ولكنهما مختلفتان في طبيعة التشتت.



2- مقاييس التشت:

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

1. المدى: Range
2. نصف المدى الربيعي Semi-Inter-quartile Range
3. الانحراف المتوسط (MD) (Mean Deviation):
4. التباين: Variance والانحراف المعياري: Standard Deviation
5. معامل الاختلاف (أو التغير): Coefficient of Variation
6. نظرية (متراجحة) تشيبيشيف Chebychev Inequality
7. الدرجات (القيم) المعيارية : Standard Scores (Values)

(2-1) المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات.

تعريف (1):

نعرف المدى لمجموعة من البيانات على أنه الفرق بين أكبر قيمة وأصغر قيمة، أي أن المدى هو:

$$Range = X_{max} - X_{min} \quad (1)$$

حيث نعرف X_{max} و X_{max} كما يلي:

(أ) للبيانات المفردة:

X_{max} = أكبر قيمة

X_{min} = أصغر قيمة

(ب) للبيانات المبوبة:

X_{max} = مركز الفترة العليا

X_{min} = مركز الفترة الدنيا

ملاحظة (1):

تعرف بعض الكتب أكبر قيمة وأصغر قيمة للبيانات المبوبة كما يلي:

X_{max} = الحد الأعلى للفترة العليا

X_{min} = الحد الأدنى للفترة الدنيا

مثال (1):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 50, 55, 35, 45, 40, 30, 25

<u>الحل:</u>

$$X_{max} = 55$$

$$X_{min} = 25$$

$$Range = X_{max} - X_{min} = 55 - 25 = 30$$ (كيلوجرامًا)

مثال (2):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

التكرار F	مركز الفترة x	مستوى الهيموجلوبين
3	**13.45**	**12.95 – 13.95**
5	14.45	13.9 5– 14.95
15	15.45	14. 95– 15.95
16	16.45	15. 95– 16.95
10	17.45	16. 95– 17.95
1	**18.45**	**17. 95– 18.95**

X_{max} = مركز الفترة العليا = 18.45

X_{min} = مركز الفترة الدنيا = 13.45

$$Range = X_{max} - X_{min} = 18.45 - 13.45 = 5.00$$

بعض مميزات وعيوب المدى:

أ- يتميز المدى بسهولة التعريف والحساب

ب- يعيب المدى العيوب التالية:

1. يتأثر المدى بالقيم الشاذة أو المتطرفة.
2. لا يأخذ المدى في الاعتبار جميع البيانات.

ملاحظة (2):

1. وحدة المدى هي نفس وحدة البيانات الأصلية.
2. نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.

(2-2) نصف المدى الربيعي Semi-Inter-quartile Range:

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو نصف المدى الربيعي. وحيث أن القيم المتطرفة هي

تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%). الشكل التالي يوضح موضع القيم الشاذة والمتطرفة.

```
 ¼ n = 25%                                                    ¼ n = 25%
-/-/-/-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-/-/-/-
ربع البيانات الصغيرة   Q1              Q2              Q3  ربع البيانات الكبيرة
```

يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

$$Q = \frac{Q_3 - Q_1}{2} \qquad (2)$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (3):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني باستخدام:

(أ) الطريقة الحسابية.

(ب) الطريقة البيانية.

الحل:

(أ) الطريقة الحسابية:

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5$$

$$A = 14.95, \quad L = 1.00, \quad F_1 = 8 \ , \quad F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

$$R^* = \frac{3n}{4} = \frac{3 \times 50}{4} = 37.5$$

$$A^* = 15.95, \quad L^* = 1.00, \quad F_1^* = 23 \ , \ F_2^* = 39$$

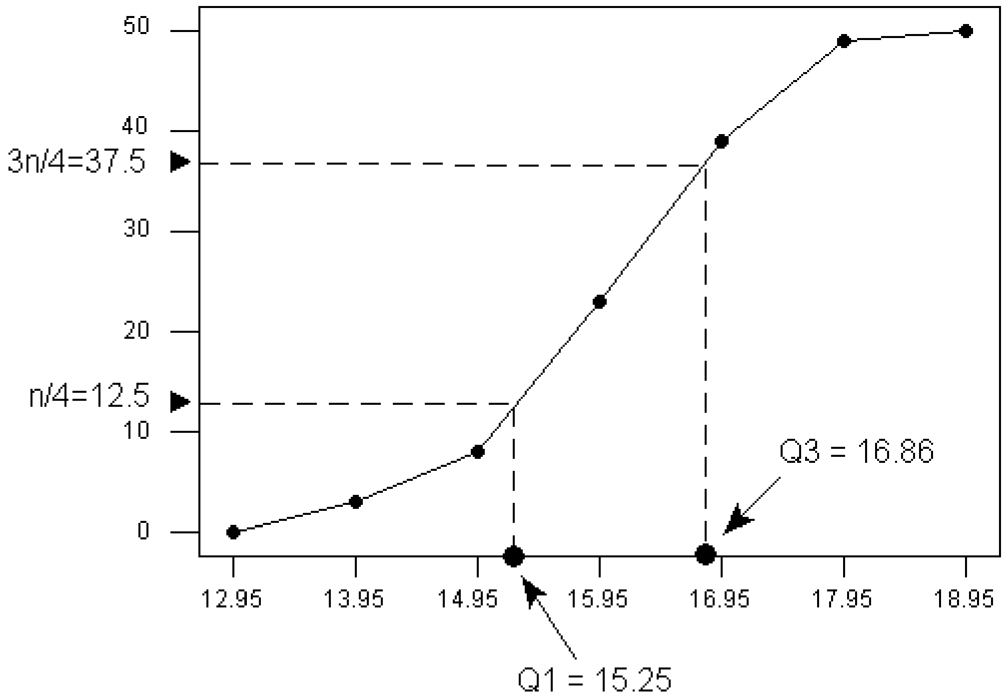
$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right) \times L^* = 15.95 + \left(\frac{37.5-23}{39-23}\right) \times 1.00 = 16.86$$

وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
	14. 95= A	أقل من	8 = F_1	$\leftarrow R = \frac{n}{4} = 12.5$
$Q_1 \Rightarrow$	15. 95= A^*	أقل من	23 = F_2 = F^*_1	$\leftarrow R^* = \frac{3n}{4} = 37.5$
	16. 95	أقل من	39 = F^*_2	
$Q_3 \Rightarrow$	17. 95	أقل من	49	
	18.95	أقل من	50	

(ب) الطريقة البيانية:



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

بعض مميزات وعيوب نصف المدى الربيعي:

1. من المميزات أنه لا يتأثر بالقيم الشاذة أو المتطرفة.
2. من العيوب أنه لا يأخذ في الاعتبار جميع البيانات.

ملاحظة (3):

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

3-2) الانحراف المتوسط (MD) (Mean Deviation):

هو متوسط الانحرافات المطلقة للقيم عن وسطها الحسابي، والانحراف المطلق هو الفرق بين قيم المتغير الإحصائي والوسط الحسابي (نحولها إلى مطلقة لأن مجموعها في الحالة العادية معدوم حسب خصائص الوسط الحسابي).

في حالة البيانات غير المبوبة: $\boxed{MD = \frac{\sum \left| x_i - \overline{X} \right|}{n}}$

وفي حالة البيانات المبوبة: $MD=\frac{\sum\left|x_i-\overline{X}\right|\times f}{n}$

حيث: $\overline{X}$ هو الوسط الحسابي.

x_i في المعادلة الأولى هو قيم المتغير X و i يتغير من 1 إلى n حيث n عدد قيم المتغير.

x_i في المعادلة الثانية هو مركز الفئة i حيث i يتغير من 1 إلى 6 أو 7 أو 8.... حسب عدد الفئات.

f هو تكرار الفئة.

مزايا وعيوب الانحراف المتوسط:

- يتميز بأخذه كل القيم بعين الاعتبار
- ويعاب عليه تأثره بالقيم الشاذة وصعوبة التعامل معه رياضيا.

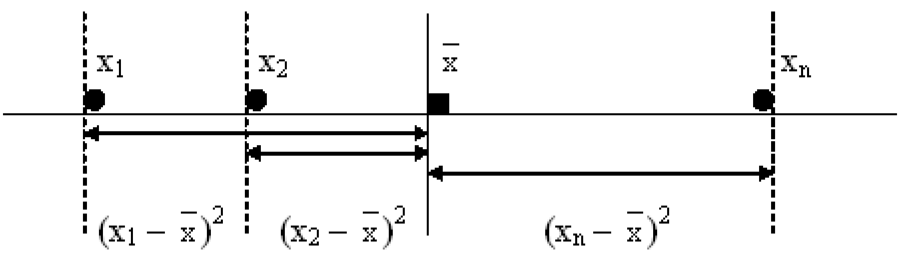
(2-4) التباين (Variance) والانحراف المعياري (Standard Deviation):

يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

أ/ التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2. الشكل التالي يبين مربعات انحرافات القيم عن الوسط الحسابي.



والجدول التالي يلخص طريقة حساب انحرافات ومربعات انحرافات القيم عن المتوسط.

القيم (البيانات)	x_n	…	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	…	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	…	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

تستخدم مربعات انحرافات القيم عن الوسط الحسابي لحساب التباين. فمن الواضح أن تشتت البيانات يزداد بزيادة مربعات الانحرافات والعكس بالعكس. والسؤال الذي يتبادر للأذهان هو أي واحد من هذه المربعات ينبغي أن يستخدم لقياس التشتت؟ إن من المنطقي أن نبحث عن قيمة نموذجية تمثل هذه المربعات لاستخدامها لقياس التباين. وبناءً على دراستنا السابقة لمقاييس النزعة المركزية فإن المتوسط من أفضل المقاييس التي تمثل مجموعة البيانات. وعليه فإن متوسط مربعات انحرافات القيم عن الوسط الحسابي $\sum_{i=1}^{n}(x_i - \bar{x})^2 / n$ يعتبر أحد المقاييس التي يمكن استخدامها لقياس التباين.

ب/ الانحراف المعياري: Standard Deviation

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز s.

حساب التباين والانحراف المعياري:

سنستعرض طرق حساب التباين والانحراف المعياري في حالة البيانات المفردة وفي حالة البيانات المبوبة.

التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (3)$$

لاحظ أننا قسمنا على المقدار $(n-1)$، وهو ما يسمى بدرجات الحرية، بدلاً من القسمة على عدد البيانات n في الصيغة السابقة وذلك لكي نحصل على مقياس يتمتع بصفات إحصائية جيدة.

وأما الانحراف المعياري فإنه يعرف بالصيغة:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad (4)$$

ملاحظة (4):

1. $s^2 \geq 0$ (دائمًا) وكذلك $s \geq 0$ (دائمًا).
2. $s^2 = 0 \Leftrightarrow s = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).
3. وحدة التباين، s^2، هي وحدة البيانات الأصلية المربعة.
4. وحدة الانحراف المعياري، s، هي نفس وحدة البيانات الأصلية.
5. يمكن حساب التباين بإحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right\} \qquad (5)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right\} \qquad (6)$$

ولحساب تباين العينة باستخدام الصيغتين الحسابيتين السابقتين فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

1. حجم العينة = n.
2. مجموع البيانات = $\sum_{i=1}^{n} x_i$.
3. مجموع مربعات البيانات = $\sum_{i=1}^{n} x_i^2$.

والصيغتين الحسابيتين السابقتين تستخدمان لحساب تباين العينة وذلك لسببين هما:

1. لأنهما أكثر سهولة من صيغة التعريف.
2. لأنها أكثر دقة في الحساب من صيغة التعريف عندما يكون هناك تقريب في حساب متوسط العينة.

مثال (6):

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية:

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	الانحراف $(x_i - \bar{x}) = (x_i - 5.16)$	مربع الانحراف $(x_i - \bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	−2.66	7.0756	6.25
2.5	−2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.8$	$\sum_{i=1}^{n} x_i^2 = 160.9$

من هذا الجدول نوجد الكميات التالية:

$$n = 5, \sum_{i=1}^{n} x_i = 25.8, \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832, \quad \sum_{i=1}^{n} x_i^2 = 160.96$$

إن متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16$$ (كيلوجرامًا)

حساب تباين العينة:

(أ) باستخدام التعريف:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958$$ (كيلوجرامًا مربعًا)

(ب) باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n\right\} = \frac{1}{5-1}\left\{160.96 - (25.8)^2 / 5\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

(ج) باستخدام الصيغة الحسابية الثانية:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - n\ (\bar{x})^2\right\} = \frac{1}{5-1}\left\{160.96 - (5)(5.16)^2\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

الانحراف المعياري هو:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = s = \sqrt{6.958} = 2.6378$$ (كيلوجرامًا)

بعض خصائص التباين والانحراف المعياري:

1. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية. فإذا كان s^2 و s هما على الترتيب التباين والانحراف المعياري للبيانات $x_1, x_2, \ldots, x_n$ وكان a و b مقدارين ثابتين، فإن:

- ☒ تباين البيانات $x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$ هو s^2 وأما انحرافها المعياري فهو S. لذلك فإن التباين والانحراف المعياري لا يتأثران بإضافة أو طرح مقدار ثابت من جميع المشاهدات.
- ☒ تباين البيانات $ax_1, ax_2, \cdots, ax_n$ هو $a^2 s^2$ وأما انحرافها المعياري فهو $|a|s$. حيث $|a|$ هي القيمة المطلقة للقيمة a.
- ☒ تباين البيانات $ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$ هو $a^2 s^2$ وأما انحرافها المعياري فهو $|a|s$.

☒ تباين المقدار الثابت يساوي الصفر.

ويمكن تلخيص هذه الخاصية في الجدول التالي:

الانحراف المعياري	التباين	المشاهدات
s	s^2	$x_1, x_2, \ldots, x_n$
s	s^2	$x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$
$\lvert a \rvert s$	$a^2 s^2$	$ax_1, ax_2, \cdots, ax_n$
$\lvert a \rvert s$	$a^2 s^2$	$ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$

مثال (7):

التباين	الانحراف المعياري	المشاهدات	
$s^2 = 2.5$	$s = 1.581$	2, 6, 4, 3, 5	x :
2.5	1.581	7, 11, 9, 8, 10	$x+5$:
$9\times2.5=22.5$	$\lvert3\rvert\times1.581=4.743$	6, 18, 12, 9, 15	$3x$:
$9\times2.5=22.5$	$\lvert3\rvert\times1.581=4.743$	11, 23, 17, 14, 20	$3x+5$:

مثال (8):

إذا كان التباين للمشاهدات $x_1, x_2, \ldots, x_n$ هو 36 فإن التباين للمشاهدات $\frac{x_1-10}{2}, \frac{x_2-10}{2}, \ldots, \frac{x_n-10}{2}$ هو $\left(\frac{1}{2}\right)^2 36 = \frac{36}{4} = 9$ وأما الانحراف المعياري فهو $\sqrt{9} = 3$.

2. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها s_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها s_2^2 وإذا كان $\bar{x}_1 = \bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة

من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

مثال (9):

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$s_1^2 = 3$	$s_2^2 = 3.5$

الحل:

أولاً نلاحظ أن متوسطي المجموعتين متساويان.

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$
$$= \frac{(4-1)(3) + (6-1)(3.5)}{4+6-1} = 2.944$$

التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

أ- عدد الفترات هو k.

ب- مراكز الفترات هي $x_1, x_2, \ldots, x_k$.

ج- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$.

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n - 1} \qquad (7)$$

حيث أن:

$$n=\sum_{i=1}^{k} f_i\ , \quad \bar{x}=\frac{\sum xf}{\sum f}=\frac{\sum xf}{n}\ ,$$

كما يمكن استخدام إحدى الصيغتين الحسابيتين التاليتين:

$$s^2=\frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i-\left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\} \qquad (8)$$

$$s^2=\frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i-n\left(\bar{x}\right)^2\right\} \qquad (9)$$

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	$x\,f$	$x^2\,f$	$f\,(x-\bar{x})^2$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$	$x_1^2\,f_1$	$f_1\,(x_1-\bar{x})^2$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$	$x_2^2\,f_2$	$f_2\,(x_2-\bar{x})^2$
:	:	:	:	:	:
:	:	:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$	$x_k^2\,f_k$	$f_k\,(x_k-\text{ر})^2$
المجموع		$\sum f=n$	$\sum xf$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i\,(x_i-\bar{x})^2$

مثال (10):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

نلخص إيجاد التباين باستخدام الجدول التالي:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf	x^2f	$f(x-\bar{x})^2$ $f(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f = 50$	$\sum xf = 800.5$	$\sum x^2 f = 12882.33$	$\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 66.320$

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800.5}{50} = 16.01$$

حساب التباين باستخدام صيغة التعريف:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\}$$

$$= \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005)$$

$$= \frac{66.325}{49} = 1.3536$$

حساب التباين باستخدام الصيغة الحسابية الثانية:

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

$$= \frac{1}{50-1}\{12882.33 - (50)(16.01)^2\}$$

$$=\frac{1}{49}\{12882.33-12816.005\}$$

$$=\frac{66.325}{49}=1.3536$$

حساب الانحراف المعياري:

$$s=\sqrt{s^2}=\sqrt{1.3536}=1.163$$

(5-2) معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

1. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة ببعض.
2. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيرًا والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\bar{x}$ وانحرافها المعياري s بالصيغة التالية:

$$C.V.=\frac{s}{\bar{x}} \qquad (10)$$

مثال (11):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا)؛ بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\bar{x}$ والانحراف المعياري s لكل من بيانات الأوزان وبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\bar{x}$	الانحراف المعياري s	معامل الاختلاف $C.V.=\frac{s}{\bar{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانسًا من بيانات الأطوال.

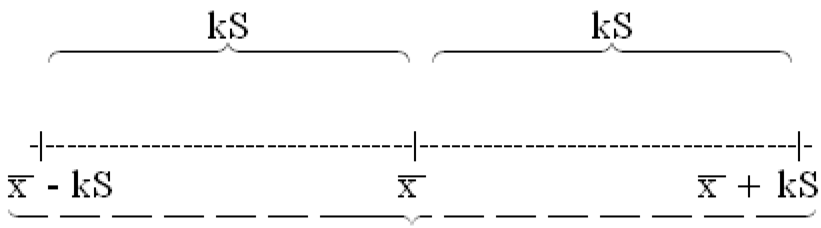
(2-6) نظرية (متراجحة) تشيبيشيف Chebychev Inequality:

إن نظرية تشيبيشيف من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة.

نظرية (1):

إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري s فإن نسبة البيانات الواقعة في الفترة $(\bar{x}-ks\,,\bar{x}+ks)$ لايقل عن $\left(1-\frac{1}{k^2}\right)$ حيث أن $k>1$.

الشكل التالي يوضح فكرة نظرية تشيبيشيف.



نسبة البيانات الواقعة في هذه الفترة لا يقل عن $\left(1-\frac{1}{k^2}\right)$

ملاحظة (5):

1. في بعض الأحيان نكتب الفترة $(\bar{x}-ks\,,\bar{x}+ks)$ على الصورة $\bar{x}\pm ks$.
2. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط $\bar{x}$.
3. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):
   - أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.
   - ب- تحديد فترة يقع فيها ما لا يقل عن نسبة معينة.

مثال (12):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري s=5 فما هي نسبة البيانات الواقعة في الفترة $(-4\ ,\ 18)$؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة $(-4,18)$ هو المتوسط $\bar{x}=7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$(\bar{x}-ks\,,\bar{x}+ks)=(-4\ ,\ 18)\quad\Rightarrow\quad \bar{x}+ks=18$$

$$\Leftrightarrow\quad 7+k(5)=18$$

$$\Leftrightarrow \quad 5k = 11$$

$$\Leftrightarrow \quad k = 11/5$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right)$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934$$

لذلك فإن نسبة البيانات الواقعة في الفترة $(-4,18)$ لا تقل عن 79.34%.

مثال (13):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $s = 5$ فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1 - \frac{1}{k^2}\right) = 0.75 \quad \Leftrightarrow \quad \frac{1}{k^2} = 1 - 0.75$$

$$\Leftrightarrow \quad \frac{1}{k^2} = 0.25$$

$$\Leftrightarrow \quad k = \sqrt{\frac{1}{0.25}}$$

$$\Leftrightarrow \quad k = 2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$(\bar{x} - ks, \bar{x} + ks) = (7 - 2\times5 \,,\, 7 + 2\times5)$$

$$= (7 - 10 \,,\, 7 + 10)$$

$$= (-3 \,,\, 17)$$

(2-7) الدرجات (القيم) المعيارية : Standard Scores (Values)

نستطيع بكل يسر وسهولة استخدام قيمتي مشاهدتين في نفس المجموعة لمقارنتهما ببعض. فمثلاً، نستطيع أن نقول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما أفضل من أداء الطالب الحائز على الدرجة 80 في نفس الاختبار إذا كان الطالبان في نفس الشعبة. وفي المقابل، لا نستطيع القول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما في الشعبة التي يدرسها المدرس

(أ) أفضل من أداء طالب آخر حائز على الدرجة 80 في نفس الاختبار ولكنه في شعبة أخرى يدرسها المدرس (ب). من هنا، نرى أنه من الضروري إيجاد قيم لا تعتمد على الوحدات ويمكن استخدامها لمقارنة البيانات في المجموعات المختلفة. هذه القيم التي لا تعتمد على الوحدات نسميها بالقيم (أو الدرجات) المعيارية.

تعريف (2):

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\bar{x}$ وانحرافها المعياري s. نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$z_i = \frac{x_i - \bar{x}}{s}; i = 1, 2, \ldots, n \qquad (11)$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \ldots, x_n$ هي:

$$z_1 = \frac{x_1 - \bar{x}}{s}, \; z_2 = \frac{x_2 - \bar{x}}{s}, \; \ldots, \; z_n = \frac{x_n - \bar{x}}{s}$$

ملاحظة (6):

1. القيمة $z_i = \frac{x_i - \bar{x}}{s}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.
2. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i = \bar{x} + S\, z_i$.
3. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بين المشاهدات المختلفة في المجموعات المختلفة للبيانات.
4. متوسط الدرجات المعيارية =0.
5. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (14):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $s = 5$ فأوجد:

1. الدرجة المعيارية للقيمة $x = 9$.
2. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

1. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z = \frac{x - \bar{x}}{s} = \frac{9 - 7}{5} = \frac{2}{5} = 0.4$$

2. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \bar{x} + s\,z = 7 + 5\times 0.1 = 7 + 0.5 = 7.5$$

مثال (15):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

نلخص إيجاد الدرجات المعيارية في الجدول التالي:

المقرر	المتوسط $\bar{x}$	الانحراف المعياري s	الدرجة x	الدرجة المعيارية $z = \frac{x - \bar{x}}{s}$
الإحصاء	75	10	82	$z = \frac{82 - 75}{10} = 0.7$
الرياضيات	81	16	89	$z = \frac{89 - 81}{16} = 0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.

3- مقاييس التشكل:

عند تمثيل بيانات الظاهرة في شكل منحنى تكراري فإن هذا المنحنى يأخذ أشكالا مختلفة فعند كون الشكل متماثل فإن قيم الوسط والوسيط والمنوال متساوية، وفي كثير من الحالات توجد قيم شاذة كبيرة تجذب الوسط الحسابي إليها فيكون للمنحنى التكراري ذيل جهة اليمين مشيرا لوجود التواء موجب، أو توجد قيم شاذة صغيرة تجذب الوسط الحسابي إليها فيكون للمنحنى التكراري ذيل جهة اليسار مشيرا لوجود التواء سالب، وتسمى هذه الظاهرة **بالالتواء**، وإذا كان شكل منحنى التوزيع منبسطا أو مدببا فهذا يسمى **بالتفرطح**. بالإضافة لمقاييس أخرى.

3-1) مقاييس الالتواء:

معامل بيرسون α: يحسب معامل الالتواء لبيرسون كما يلي:

$$\text{معامل بيرسون } \alpha = \frac{3(\text{الوسط الحسابي} - \text{الوسيط})}{\text{الانحراف المعياري}} = \frac{3(\bar{X} - Med)}{s} = \frac{3(\text{Mean-Median})}{\text{Standard deviation}} = \alpha$$

- إذا كان $\alpha = 0$ فإن منحنى التوزيع التكراري متماثل.
- إذا كان $\alpha > 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).
- إذا كان $\alpha < 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

معامل الميئين $\alpha_{i,\,100-i}$: يفضل استخدامه في حالة وجود قيم شاذة

وتعتمد فكرة الميئين لقياس الالتواء على مدى قرب الميئين P_i والميئين P_{100-i} من الميئين P_{50}، من فمثلا إذا أردنا أن نعرف مدى قرب قرب الميئين P_{20} والميئين P_{100-80} من الميئين P_{50} فإننا نتبع المعادلة التالية:

$$\boxed{\alpha_{i,\,100-i} = \frac{(P_{100-i} - P_{50}) - (P_{50} - Pi)}{P_{100-i} - Pi}}$$ حيث: $P_{100-i} < P50 < Pi$

- إذا كان $\alpha_{i,\,100-i} = 0$ فإن منحنى التوزيع التكراري متماثل.
- إذا كان $\alpha_{i,\,100-i} > 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).
- إذا كان $\alpha_{i,\,100-i} < 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

3-2) مقاييس التفرطح:

عندما يتركز عدد كبير من القيم بالقرب من منتصف المنحنى ويقل في طرفيه يكون المنحنى مدببا، وعندما يتركز عدد كبير من القيم في طرفي المنحنى ويقل في وسطه يكون المنحنى مفرطحا، ويقاس التفرطح باستخدام العديد من المقاييس أهمها **معامل التفرطح** (k) باستخدام العزوم بتطبيق المعادلة التالية في حالة بيانات غير مبوبة: $\boxed{k=\frac{\frac{1}{n}\sum(x_i-\overline{X})^4}{S^4}}$، أو المعادلة التالية في حالة بيانات مبوبة $\boxed{k=\frac{\frac{1}{n}\sum f(x_i-\overline{X})^4}{S^4}}$.

ويعتمد هذا المعامل k على العزم المركزي الرابع حول الوسط والذي يساوي k=3 في حالة توزيع طبيعي أي يكون المنحنى على شكل جرسي، ومن ثم يمكن وصف منحنى التوزيع من حيث التفرطح والتدبب كما يلي:

- إذا كان k=3 كان منحنى التوزيع معتدلا (طبيعيا على شكل جرس).
- إذا كان k>3 كان منحنى التوزيع التكراري مدببا.
- إذا كان k<3 كان منحنى التوزيع التكراري منبسطا (مفرطحا).

4) مقاييس أخرى لوصف تشتت البيانات:

4-1) الدرجة المعيارية:

تقيس الدرجة المعيارية لقيمة معينة عدد الوحدات الانحراف المعياري التي تزيد بها أو تقل بها هذه القيمة عن الوسط الحسابي، فإذا كان لدينا n عدد المشاهدات وهي: $x_1 \; x_2 \ldots x_n$، ووسطها الحسابي $\overline{X}$ وانحرافها المعياري s، فإن الدرجة المعيارية للقيمة x والتي يرمز لها z تحسب بالمعادلة التالية: $\boxed{Z=\frac{x_i-\overline{X}}{S}}$

ويمكن استعمال هذه القيمة في مقارنة قيمتين أو أكثر تكون مختلفة من حيث وحدات القياس.

فإذا كان لدينا مجموعتين من البيانات الوسط الحسابي للمجموعة الأولى هو 173 وانحرافها المعياري 23، والمجموعة الثانية وسطها الحسابي 198 وانحرافها المعياري 24، وعند قياس قيمة

واحدة لا على التعيين من كل مجموعة وجدنا أن القيمة الأولى هي x=178 والثانية x=180، فإن الدرجة المعيارية للقيمتين بعد تطبيق القاعدة السابقة هما z=0.22 و z=0.75.

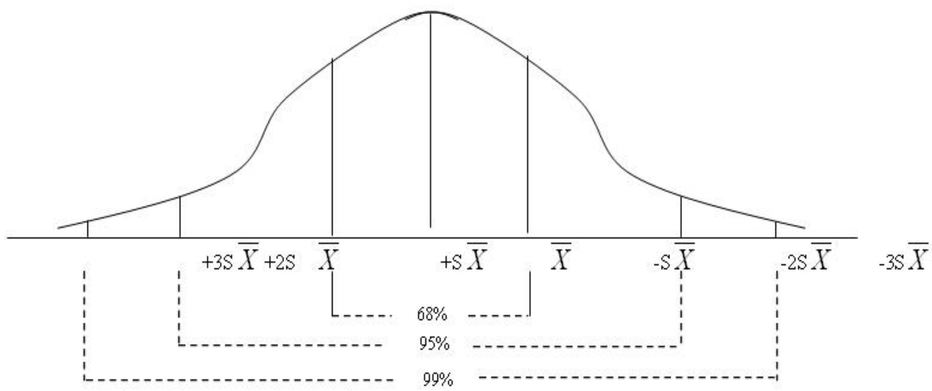
نلاحظ أن القيمة الأولى تزيد عن الوسط الحسابي بـ:0.22 انحراف معياري، وأن القيمة الثانية تزيد عن الوسط الحسابي بـ:0.75 انحراف معياري أي أن الأهمية النسبية للقيمة الأولى أعلى من القيمة الثانية، أي أن المجموعة الأولى أقل تشتتا من المجموعة الثانية.

2-4) القاعدة العملية:

إذا كان لدينا n عدد المشاهدات وهي: $x_1\ x_2.....x_n$، ووسطها الحسابي $\overline{X}$ وانحرافها المعياري s، فيكون منحنى توزيع هذه المشاهدات متماثل إذا تحقق ما يلي:

- 68% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm S$
- 95% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm 2S$
- 99% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm 3S$

ويمكن بيان ذلك من الشكل التالي:



حيث المساحة تحت المنحنى تساوي (1) الواحد الصحيح، أي أن جميع القيم (100%) تقع تحت المنحنى.

الفصل الرابع :

الارتباط والانحدار البسيط

الفصل الرابع : الارتباط والانحدار الخطي البسيطين

عرضنا المقاييس التي تصف شكل المنحنى مثل النزعة المركزية والتشتت والتشكل من خلال وصف وعرض البيانات التي نجمعها من متغير واحد، وننتقل الآن للتعامل مع متغيرين (أو أكثر) باستخدام بعض طرق التحليل الإحصائي: مثل تحليل الارتباط الخطي البسيط لدراسة العلاقة بين المتغيرين ونوعها وقوتها، أو دراسة وتحليل تأثير أحدهما في الآخر بتحليل الانحدار الخطي البسيط، ويعتمد الانحدار والارتباط على دراسة العلاقة بين متغيرين أحدهما يتغير تبعا لتغير الأخر أي أحدهما يؤثر في الآخر(فيسمى الذي يؤثر متغير **مستقل X**، والمتأثر يسمى متغير **تابع Y**)، والغرض من كل ذلك صياغة نموذج (معادلة) للعلاقة بينهما من أجل إمكانية التنبؤ مستقبلا بقيم المتغير التابع وفق قيم المتغير المستقل (إذا بقيت نفس الظروف على حالها)

1. معامل الارتباط

ان الهدف من دراسة الارتباط Correlation هو الكشف عن قوة أو درجة العلاقة بين متغيرين أو أكثر، وتتراوح درجة العلاقة بين أي متغيرين والتي يعبر عنها باصطلاح معامل الارتباط Correlation Coefficient بين + 1 ، −1. فكلما كانت درجة الارتباط قريبة من 1 فإن ذلك يعني ان الارتباط قوياً بين المتغيرين، وكلما قلت درجة الارتباط كلما ضعفت العلاقة بين المتغيرين.

وقد تتخذ العلاقة الارتباطية بين المتغيرين أحد شكلين:-

1. **علاقة طردية:** زيادة قيمة أحد المتغيرين تؤدي الى زيادة قيمة المتغير الآخر وكذلك نقصان قيمة أحد المتغيرين تؤدي الى نقصان قيمة المتغير الآخر كالعلاقة بين المصروف على الاعلان والمبيعات.
2. **علاقة عكسية:** زيادة قيمة أحد المتغيرين تؤدي الى نقصان قيمة المتغير الآخر، مثل العلاقة بين معدل دوران العمل والانتاجية. ويمكن ان تكون العلاقة بالعكس، فنقصان قيمة أحد المتغيرين قد يؤدي الى زيادة قيمة المتغير الآخر .

بشكل عام فإنه يمكن اعتبار ان العلاقة ضعيفة اذا كانت قيمة معامل الارتباط أقل من 0.30 ، ويمكن اعتبارها متوسطة اذا تراوحت قيمة معامل الارتباط بين 0.30 الى 0.70 أما اذا كانت قيمة معامل الارتباط أكثر من 0.70 فتعتبر العلاقة قوية بين المتغيرين.

ومن الجدير بالذكر ان الارتباط يدل على وجود علاقة ما بين متغير وآخر، الا انه يجب أن ندرك بأن هذه العلاقة لا تدل على السببية أو العلية، فهي لا تدل على وجود أثر لمتغير على أخر، فقد تكون هناك علاقة طردية بين شرب القهوة ومعدلات الوفيات الا أن شرب القهوة لا يعتبر سبباً في زيادة معدلات الوفيات بين الناس، فقد يكون هناك عامل آخر كالتدخين مثلاً ينتج عن زيادة معدلات شرب القهوة ويؤثر في معدلات الوفيات فزيادة معدلات شرب القهوة تؤدي الى زيادة استهلاك السجائر مما يؤثر في زيادة معدلات الوفيات.

1-1/ معامل ارتباط بيرسون Pearson

يستخدم معامل ارتباط بيرسون لقياس قوة العلاقة بين قيم متغيرين كالعلاقة بين مصروف الاعلان وحجم المبيعات أو العلاقة بين التدريب وانتاجية العاملين.

ويمكن استخراج معامل الارتباط من خلال تطبيق المعادلة التالية:

$$R=\frac{N(\sum x y)-\sum x \sum y}{\sqrt{\left[N\sum x^2-(\sum x)^2\right]\left[N\sum y^2-(\sum y)^2\right]}}$$

حيث R = معامل ارتباط بيرسون

X = المتغير الأول

Y = المتغير الثاني

1-2/ معامل ارتباط الرتب Spearman

قد يضطر الباحث الى التعامل مع ترتيب البيانات بدلاً من التعامل مع قيمها. وفي هذه الحالة بإمكانه استخدام معامل ارتباط بيرسون والذي يعتمد على اساس اعطاء كل مفردة في كل متغير ترتيباً معيناً وليس قيماً محددة، فإذا قمنا بترتيب مفردات المتغير X وكذلك مفردات المتغير Y ووجدنا

ان ترتيب هذه المفردات في كلا المتغيرين متوافقة ومنسجمة فإن ذلك يعني ان هناك ارتباطاً بين المتغرين.

ويمكن قياس معامل الارتباط بين مفردات أي متغيرين بترتيب كل من هذه المفردات في المتغير، ثم حساب الفرق بين رتبتي كل مفردة وتربيع هذه الفروق. ولكي نتمكن من استخراج معامل ارتباط الرتب Spearman نستخدم المعادلة التالية:

$$R = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

حيث D = الفرق بين رتبتي كل مفردة.

يتميز معامل سبيرمان بسهولة طريقة حسابه الا انه يعطي قيمة تقريبية أقل دقة من معامل ارتباط بيرسون، حيث انه يعتمد على ترتيب القيم وبدون اعتبار لتساوي المسافات بين كل ترتيب وآخر. فلو أدخلنا نفس القيم الموجودة في المثال السابق واخترنا معامل ارتباط سبيرمان بدلاً من معامل ارتباط بيرسون، حيث ستجد أن قوة الارتباط في مثالنا 849. ، بينما بلغت حسب معامل ارتباط بيرسون845. عند اختيارك لمعامل ارتباط سبيرمان، فقد قام البرنامج بترتيب القيم ثم قام بعدها بحساب معامل الارتباط.

<u>مثال</u>: الملاحظات غير مكررة حيث عدد الملاحظات يساوي عدد الطلبة وهي غير مكررة

أوجد R_S معامل ارتباط الرتب لسبيرمان بين ملاحظات الطلبة في مقياسي الإحصاء وعلم الاجتماع لعينة عشوائية من 6 طلاب:

رقم الطالب	ملاحظات الاقتصاد X	ملاحظات الإحصاء Y	رتب الاقتصاد	رتب الإحصاء	d	d^2
1	ضعيف	مقبول	5	4	1	1
2	ممتاز	جيد جدا	1	2	1-	1
3	جيد	جيد	3	3	0	0
4	ضعيف جدا	ضعيف	6	5	1	1
5	مقبول	ضعيف جدا	4	6	2	4
6	جيد جدا	ممتاز	2	1	1	1
المجموع						8

- نرتب الملاحظات تصاعديا (أو تنازليا): بداية من: ممتاز1، جيد جدا2، جيد3، مقبول4، ضعيف5 ثم ضعيف جدا6.
- نضع في خانتي الرتب الأرقام التي تقابل كل ملاحظة.

وبتطبيق المعادلة السابقة نجد: $R_S = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6\times 8}{6(36-1)} = 1 - 0.2286 = 0.7714$

ويمكن أن نقول أن العلاقة بين ملاحظات الإحصاء وملاحظات الاقتصاد **<u>طردية</u> <u>وقوية</u>**.

<u>مثال</u> : (الملاحظات مكررة) فلو كانت ملاحظات الطالب في الإحصاء والاقتصاد بالصيغة التالية:

d^2	d	رتب Y	رتب X	Y	X	n
1	1	1	2	A+	A	1
6.25	2.5	10	7.5	D	C+	2
4	2	8	10	C	D	3
1	1	8	9	C	D+	4
4	2	2	4	A	B+	5
6.25	2.5	5	7.5	B	C+	6
4	2	3	1	B+	A+	7
1	1	5	6	B	B	8
16	4	8	4	C	B+	9
1	1	5	4	B	B+	10
44.5						

- نرتب الملاحظات من A+ إلى C كما يلي:

10	9	8	7	6	5	4	3	2	1	الرتب
D	D+	C+	C+	B	B+	B+	B+	A	A+	X
10	9	7.5		6	4=3/(5+4+3)			2	1	رتب X
D	C	C	C	B	B	B	B+	A	A+	Y
10	8			5			3	2	1	رتب Y

- في عمودي رتب X و Y في الجدول نضع الرتبة المقابلة لكل ملاحظة في عمودي X و Y.
- ثم نحسب الفروق بين رتب الملاحظات ومربعات الفروق.
- ونحسب المعامل كما يلي:

$$R_S = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 44.5}{10(100-1)} = 0.7303$$

فنقول أن العلاقة بين ملاحظات الإحصاء والاقتصاد قوية.

2-الارتباط الجزئي Partial Correlations

يقيس الارتباط الجزئي العلاقة بين متغيرين Y,X بعد ثبات أثر أي متغيرات أخرى، أي تحييد أثر المتغيرات الاخرى التي قد تؤثر على أحد المتغيرين Y,X واللذان نريد قياس العلاقة بينهما.

السؤال الآن.....لماذا نستخدم الارتباط الجزئي؟؟؟؟

إن العلاقة بين المتغيرين Y,X قد تكون علاقة كاذبة أو غير حقيقية Spurious وذلك عندما يكون هناك متغير ثالث خارجي يؤثر في كل منهما وفي نفس الوقت لا يؤثر أحدهما في الآخر. فقد تكون هناك علاقة بين زيادة مبيعات الآيس كريم وانخفاض الانتاجية ولكن هذه العلاقة غير ناتجة عن تأثير الايس كريم في الانتاجية أو تأثير الانتاجية في الآيس كريم. انها علاقة غير حقيقية لأنه قد يكون هناك متغيراُ ثالثاً مثل درجة الحرارة المرتفعة يؤثر في كلا المتغيرين، وهذا المتغير الثالث يسمى المتغير العرضي Extraneous .

ويستخدم تحليل الارتباط الجزئي لاختبار قوة واتجاه العلاقة الخطية بين متغيرين بعد تثبيت أثر متغير آخر غير حقيقي قد يتسبب وجوده في نتائج غير دقيقة.

3- الانحدار الخطي البسيط :

الهدف الاساسي من تحليل الانحدار Regression Analysis هو تقدير الصورة الرياضية للعلاقة بين متغير مستقل ومتغير تابع. ويستخدم تحليل الانحدار لدراسة مدى تأثير متغير مستقل واحد أو

أكثر على متغير تابع محدد بحيث نستطيع التنبؤ بقيم المتغير التابع اذا علمنا قيم المتغير المستقل أو المتغيرات المستقلة.

ويجب ان تتوفر شروط أساسية لاجراء تحليل الإنحدار حتى تكون النتائج دقيقة ويمكن الوثوق بها، حيث ينبغي ان يكون توزيع المتغيرين المستقل والتابع توزيعاً طبيعياً، كما ينبغي ان تكون العينة مختارة بشكل عشوائي.

وهناك نوعين من الانحدار الخطي:

أ. *الانحدار الخطي البسيط Simple Regression*: يبحث في تأثير متغير مستقل واحد في متغير تابع واحد.

ب. *الانحدار الخطي المتعددMultiple Regression*: يبحث في تأثير أكثر من متغير مستقل في متغير تابع واحد

معادلة الانحدار الخطي البسيط:

يعد الانحدار الخطي البسيط من أكثر الموضوعات استخداماً في العمليات الإحصائية . أن عملية الإنحدار الخطي في أبسط صورها تبدأ بوجود متغير واحد مستقل Independent ومتغير آخر تابع Dependent ، فإذا توفرت بيانات للمتغيرين يكون المطلوب الحصول على أحسن خط يمثل العلاقة بين المتغيرين باستخدام هذه البيانات.

ويمكن تمثيل العلاقة بين المتغير المستقل والتابع على شكل معادلة كما يلي:-

$$Y = a + bx + e$$

حيث : Y = المتغير التابع

a= قيمة ثابتة Constant وهي تمثل البعد بين تقاطع الخط المستقيم مع المحور Y وبين نقطة الاصل.

b= ميل الانحدار (ميل الخط المستقيم) Slope

x= المتغير المستقل.

e= الأخطاء العشوائية

هناك إجمالاً حالتين لتجمع النقاط على الخط :

أ- الحالة الأولى تجمع النقاط بالضبط فوق الخط المستقيم مما يشير إلى أن العلاقة بين المتغيرين Exact

ب- الحالة الثانية تجمع النقاط حول الخط مما يستدعي ضرورة إنشاء الخط الأكثر ملاءمة Best-of-fit والذي يمر بأكثر النقاط

ان من المهم معرفة كيفية الوصول الى المعادلة التي تعين لنا مسار الخط الذي يعبر عن العلاقة الخطية بين المتغيرين. وينبغي أن نراعي أن يمر الخط المستقيم أو الخط الأكثر ملاءمة بأكبر عدد من النقاط بحيث يكون مجموع مربع انحرافات هذه النقاط عن الخط المستقيم أقل ما يمكن. هذه هي الفكرة الاساسية لما يسمى بطريقة المربعات الصغرى Method of Least Squares

1. *ولايجاد كل من قيمتي Ъ, ả فإننا نستخدم المعادلتان التاليتان:*

$$Ъ = \frac{\sum xy - N\overline{xy}}{\sum x^2 - N(\overline{x})^2}$$

$$ả = \overline{y} + b\overline{x}$$

حيث: $\overline{x}$ = تمثل المتوسط الحسابي للمتغير المستقل

$\overline{y}$ = تمثل المتوسط الحسابي للمتغير التابع

العلاقة الخطية تعني أن نسبة الزيادة في المتغير X تساوي نسبة الزيادة في المتغير Y

وكلمة البسيطة تعني العلاقة بين متغيرين فقط.

والأمثلة في مجال العلوم الاجتماعية كثيرة لا يمكن حصرها، فإذا جمعنا بيانات عن متغيرين (X Y) وتم تمثيل قيمهما بيانيا (فيما يسمى شكل الانتشار)، فإن العلاقة بينهما غالبا ما تكون على أحد الأشكال التالية:

- علاقة خطية طردية.
- علاقة خطية عكسية.
- لا توجد علاقة.
- علاقة غير خطية.

الفصل الخامس

المتغيرات العشوائية ذات التوزيع الاحتمالي

الفصل الخامس : المتغيرات العشوائية ذات التوزيع الاحتمالي

- **التجربة العشوائية:** أي عملية أو تجربة لا يمكن تحديد نتائجها.
- **فراغ العينة:** مجموع النتائج أو الحودث الممكنة للتجربة.(قطعة نقد مرتين (n(s)=4))، قطعة نرد مرتين (n(s)=36).
- **الحادث:** هو كل مفردة في فراغ العينة (ويكون بسيطا أو مركبا) (وله حالات الاتحاد، التقاطع، التكامل، التنافي).
- **الإمكانية:** تعبر عن فرصة أو نسبة وقوع حادث معين(<u>التكرار</u>).
- **الاحتمال التجريبي:** هو **<u>التكرار النسبي</u>** أي تكرار الحادث مقسوما على فراغ العينة.

يُدْرَسُ المتغير المتقطع بعدة توزيعات احتمالية منها (توزيع ذي الحدين، وتوزيع Poisson). و يُدْرَسُ المتغير المستمر بعدة توزيعات منها (التوزيع المنتظم، التوزيع الأسي السالب، توزيع Student والتوزيع الطبيعي).

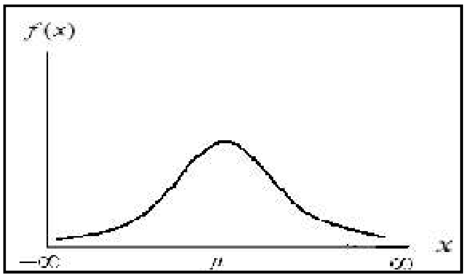
1) التوزيع الاحتمالي للمتغير الكمي المستمر (باستخدام التوزيع الطبيعي):

يعتبر هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما في النواحي التطبيقية كالاستدلال الإحصائي (التقدير، التنبؤ واختبار الفروض)، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع.

1-1) شكل دالة الكثافة الاحتمالية للتوزيع الطبيعي:

إذا كان x_i متغير عشوائي يتوزع توزيعا طبيعيا مداه هو $-\infty < x_i < +\infty$ فإن دالة كثافة احتماله هي:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \overline{X}}{\sigma}\right)^2}}$$



حيث $\boxed{-\infty < x_i < +\infty}$ ، $\boxed{\pi = \frac{22}{7}}$

الشكل (5- 1): تماثل على جانبي الوسط الحسابي

ولهذا التوزيع منحنى متماثل على جانبي الوسط الحسابي كما في الشكل المقابل:

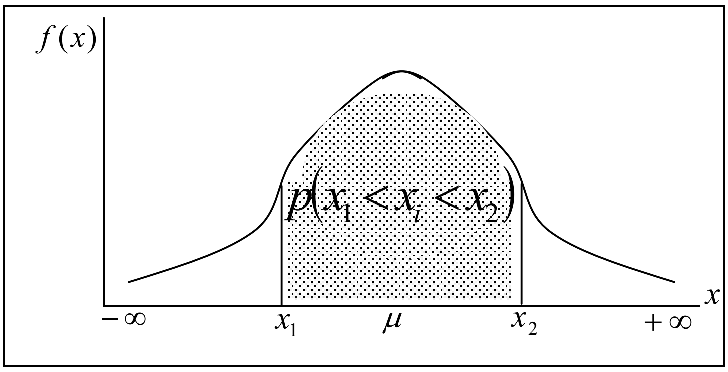
1-2) معالم وخصائص التوزيع الطبيعي:

لهذا التوزيع معلمتين هما: الوسط الحسابي $E(x) = \mu$ والتباين $\mathrm{var}(x) = \sigma^2$ ويعبر عن هذا التوزيع كما يلي: $\boxed{x \sim N(\mu, \sigma^2)}$

ويعني أن المتغير العشوائي x يتبع قانون التوزيع الطبيعي (loi de **N**ormalité) بمتوسط قدره μ وتباين قدره σ^2. ومن خصائص هذا التوزيع أنه أكثر التوزيعات الاحتمالية استعمالا، وتشتق منه كل التوزيعات الاحتمالية المستعملة في الاستدلال الإحصائي: ووسطه μ وتباينه σ^2 وهو متماثل على جانبي الوسط الحسابي.

1-3) كيفية حساب الاحتمالات (التكرارات النسبية $f(x)$):

نفرض أن الاحتمال المراد حسابه هو: ما هو احتمال أن تقع القيمة x_i مثلا بين قيمتين هما x_1 و x_2. أي $p(x_1 < x_i < x_2)$ وهذا الاحتمال يحدد بالمساحة الموجودة تحت المنحنى حيث أن المساحة الإجمالية تحت المنحنى تساوي الواحد الصحيح (أي مجموع التكرارات النسبية).



وحيث أن هذا التوزيع من التوزيعات المستمرة فإن هذه المساحة تحت المنحنى تحسب بتكامل معادلة دالة الكثافة الاحتمالية السابقة (الطلبة غير مطالبين بذلك). ونظرا لصعوبته لجأ الإحصائيون إلى عملية تحويل رياضية وبتعويض متغير جديد (Z) بدل (X) حيث: $\boxed{Z = \frac{x_i - \overline{X}}{\sigma}}$.

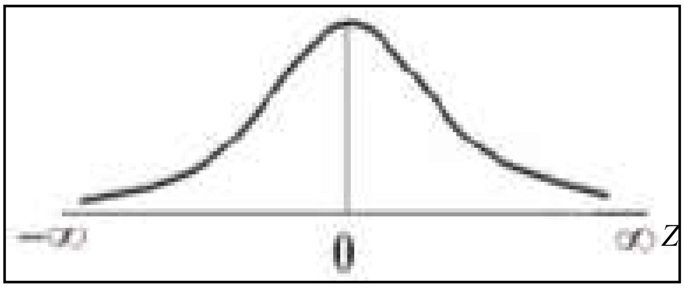
وحيث أن المتغير x يعرف بالمتغير الطبيعي. فالمتغير الجديد z يعرف بالمتغير الطبيعي القياسي($Standard\ Normal\ Variable$)، وهذا المتغير دالة كثافته الاحتمالية:

$$\boxed{f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{Z^2}{2}}}$$

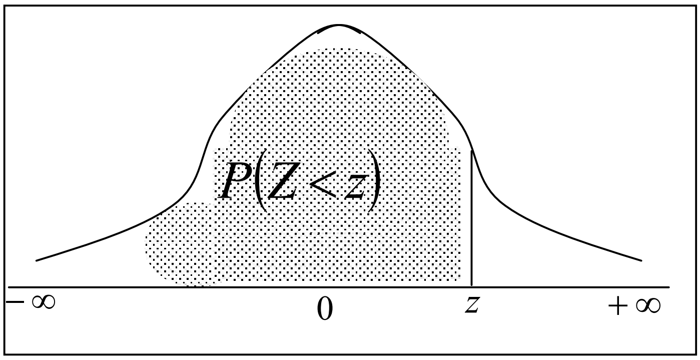
حيث $\boxed{-\infty<z_i<+\infty}$، $\boxed{\pi=\frac{22}{7}}$

لهذا التوزيع معلمتين هما: الوسط الحسابي $E(z)=0$ والتباين $\text{var}(z)=1$.

ويعبر عن هذا التوزيع كما يلي: $z\sim N(0,1)$ ويعني ذلك أن المتغير العشوائي z يتبع قانون التوزيع الطبيعي (loi de Normalité) بمتوسط حسابي قدره 0 وتباين قدره 1.



ويأخذ هذا المنحنى الشكل الناقوس أو الجرسي المتماثل على جانبي الوسط الحسابي 0.



وصمم الإحصائيون جداول إحصائية لحساب دالة التوزيع التجميعي F حيث أن $F(z)=P(Z<z)$ كما يبين الشكل التالي: وحيث أن المساحة تحت المنحنى تساوي 1 والمنحنى متماثل فإن $P(Z<0)=0.5$ أي أن نصف مساحة المنحنى الأقل من الوسط الحسابي 0 تساوي 0.5.

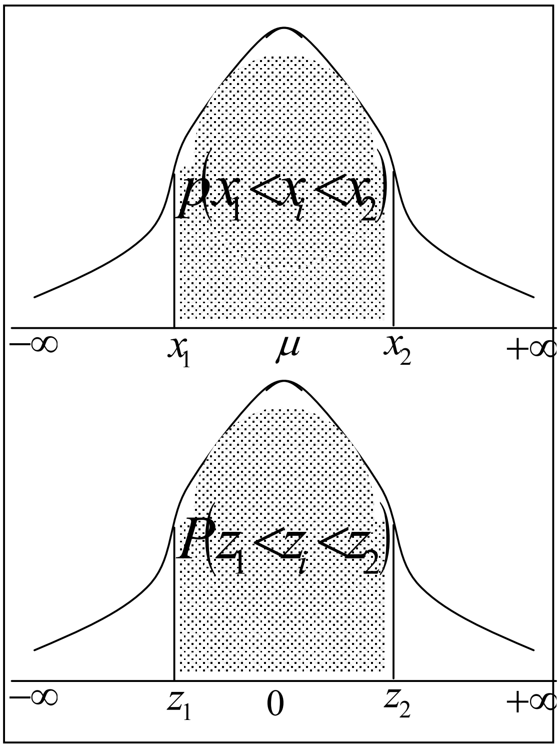
- أوجد الاحتمالات التالية باستخدام جدول التوزيع الطبيعي:

$P(z < 2.57)$ $P(z < 1.57)$ $P(z < 0.57)$

$P(z > 1.96)$ $P(z > 1.21)$ $P(z > 0.96)$

$P(z > -2.68)$ $P(z > -1.68)$ $P(z > -0.68)$

$P(z < -2.33)$ $P(z < -1.33)$ $P(z < -0.33)$



ونفس الشيء بالنسبة لاحتمال وقوع قيمة x_i بين قيمتين x_1 و x_2. أي

$$p(x_1 < x_i < x_2) = p(z_1 < z_i < z_2)$$

- نحول القيم الطبيعة (x_1,x_2) إلى قيم طبيعية قياسية (z_1, z_2):

$$Z_2 = \frac{x_2 - \overline{X}}{\sigma} \text{ و } Z_1 = \frac{x_1 - \overline{X}}{\sigma}$$

- ومن ثم يكون الاحتمال: $p(x_1 < x_i < x_2) = p(z_1 < z_i < z_2)$
- نستخدم جداول التوزيع الطبيعي القياسي لحساب الاحتمالات.
- ثم أوجد الاحتمالات التالية: $P(2.01 < z < 1.28)$ $P(-2.01 < z < 1.28)$

$P(-2.01 < z < -1.28)$

ونفس الطريقة عند استعمال جدول توزيع ستيودنت (وذلك عند كون حجم العينة أقل من 30)

الفصل السادس :

السلاسل وحلولها

الفصل السادس : سلاسل تمارين

السلسلة الأولى (عرض البيانات):

التمرين الأول:

سجلت مصلحة طب العيون بمستشفى، حضور 80 مريضا يوم: 2008/11/16، حيث تم قياس ضغط دمهم (وهو المتغير المدروس ونعبر عنه بـ X_i).

المطلوب:

1- يتكون المجتمع المدروس من:

أ- كل المرضى في المستشفى.

ب- كل المرضى المصابين بمرض العيون.

د- كل المرضى المسجلين بهذه المصلحة.

ج- 80 مريضا الذين حضروا يوم 2008/11/16.

2- المتغير X_i المدروس هو متغير:

أ- كمي مستمر. ب- كمي متقطع.

ج- كيفي اسمي. د- كيفي ترتيبي.

التمرين الثاني:

أجري استطلاع للرأي على 200 مواطن جزائري لمعرفة رأيهم حول تعديل الدستور بصيغته المقترحة، وكانت الإجابة المطلوبة من المواطنين هي إما: "موافق أو غير موافق" على تعديل الدستور، أما الإجابة المطلوبة من طلبة العلوم الاقتصادية سنة أولى "ل م د" هي:

1- يتكون المجتمع الإحصائي من:

أ- كل المواطنين.

ب- المواطنين الجزائريين.

ج- من المواطنين المستجوبين.

د- من المواطنين الذين اختير منهم 200 مستجوب.

2- المتغير المدروس هو:

أ- كمي مستمر. ب- كمي متقطع.

ج- وصفي اسمي. د- وصفي ترتيبي.

التمرين الثالث:

يتكون مجتمع ما من 4 فئات اجتماعية (حسب المهنة) $N=5000$، وحجم كل فئة هو كالتالي: $N1 = 1000 ; N2 = 1800 ; N3 = 1600 ; N4 = 600$، ونريد سحب عينة حجمها $n = 180$.

المطلوب:

1- ما هي طبيعة المجتمع المدروس؟.

2- حدد عدد الوحدات الإحصائية التي يجب سحبها من كل فئة.

السلسلة الثانية (عرض البيانات):

التمرين الأول:

فيما يلي بيانات عن المستوى التعليمي لعينة من 50 فرد.

أميّ	جامعي	أساسي	ثانوي	أساسي	ابتدائي	جامعي	ثانوي	أساسي	ابتدائي
أساسي	ثانوي	جامعي	أميّ	ثانوي	أساسي	ثانوي	جامعي	أميّ	ثانوي
ثانوي	ابتدائي	ثانوي	أساسي	ثانوي	جامعي	ثانوي	أساسي	ثانوي	أساسي
ابتدائي	أميّ	ابتدائي	ثانوي	جامعي	أساسي	ثانوي	ثانوي	أساسي	ثانوي
جامعي	أساسي	ثانوي	أساسي	أميّ	جامعي	ابتدائي	ثانوي	ابتدائي	جامعي

المطلوب:

1- اعرض البيانات في شكل جدول تكراري وكون التوزيع التكراري النسبي.

2- علق على النتائج.

التمرين الثاني:

فيما يلي بيانات درجات 70 طالب في اختبار مقياس الإحصاء الوصفي.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

المطلوب:

1- كون جميع أنواع التوزيعات التكرارية، بما فيها النسبية.

2- ما هى نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من 80 ؟، نسبة الطلاب الحاصلين على درجة أقل من 70 درجة ؟، وما هى نسبة الطلاب الحاصلين على درجة 80 أو أكثر

3- أرسم المدرج البياني والمضلع البياني، والمنحنى البياني لجميع التوزيعات السابقة.

حل السلسلة الثانية (عرض البيانات):

حل التمرين الأول:

1- عرض البيانات في جدول تكراري وتكوين التوزيع التكراري النسبي:

نلاحظ من البيانات وجود 5 مستويات أو حالات لمتغير المستوى التعليمي:

المتغير X_i	التكرار	التكرار النسبي
أميّ	5	0.10=
ابتدائي	7	0.14
أساسي	12	0.24
ثانوي	16	0.32
جامعي	10	0.20
المجموع	50	1.00

2- التعليق على النتائج: (تعميم النتائج على المجتمع)

- درجة الأمية بلغت 10 % من أفراد المجتمع.
- 14% من أفراد المجتمع لا يتعدى مستواهم التعليمي الابتدائي.
- 24% من أفراد المجتمع لا يتعدى مستواهم التعليمي الأساسي.
- 32% من أفراد المجتمع يملكون المستوى الثانوي.
- 20% فقط من أفراد المجتمع يملكون المستوى الجامعي.

حل التمرين الثاني:

1- تكوين جميع أنواع التوزيعات التكرارية، بما فيها النسبية:

فئات في مثال الدرس		فئات في السلسلة		تكرار بسيط	تكرار صاعد	تكرار نازل	نسبي بسيط	نسبي صاعد أقل من الحد الأعلى للفئة	نسبي نازل أكبر من الحد الأدنى للفئة
18	22	55	60	10	10	70	0.14	0.14	1.00
22	26	60	65	12	22	60	0.17	0.31	0.86
26	30	65	70	13	35	48	0.19	0.50	0.69
30	34	70	75	16	51	35	0.23	0.73	0.50
34	38	75	80	10	61	19	0.14	0.87	0.27
38	42	80	85	4	65	9	0.06	0.93	0.13
42	46	85	90	3	68	5	0.04	0.97	0.07
46	50	90	95	2	70	2	0.03	1.00	0.03
				70			1.00		

2-نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من 80: هي مجموع نسبة الفئة 70-75 و75-80 وهي: (0.14+0.23 = 0.37 أي 37%)، نسبة الطلاب الحاصلين على درجة أقل من 70 درجة هي مجموع نسب الفئات الأقل من 70 وهي (في التكرار الصاعد النسبي 0.50 أي 50%)، ونسبة الطلاب الحاصلين على درجة 80 أو أكثر هي مجموع نسب الفئات الأكثر من 80 وهي (في التكرار النازل النسبي 0.13 أي 13%).

3- أرسم المدرج البياني والمضلع البياني، والمنحنى البياني لجميع التوزيعات السابقة.

السلسلة الثالثة (النزعة المركزية):

التمرين الأول:

قرر مدير إحدى المؤسسات بمكافأة أكثر العمال كفاءة بسحب عينة عشوائية من عمال مؤسسته فوجد كمية إنتاجهم الشهرية كالتالي: 115 123 119 123 124 119 123 121 123 121.

المطلوب:

1- ما هو متوسط الإنتاج الشهري لعمال العينة المختارة؟، وما هي القيمة التي تزيد عنها نصف كمية الإنتاج؟، وما هي قيمة الإنتاج الأكثر تكرارا؟، ثم حدد شكل الالتواء لهذه البيانات.

2- أحسب الوسط الحسابي المعدل إذا أضفنا قيمة = 5 لكل القيم العشرة السابقة.

3- أحسب الوسط الحسابي المعدل إذا ضربنا قيمة = 3 في كل القيم العشرة السابقة.

4- أحسب مجموع مربعات انحرافات القيم عن وسطها الحسابي.

5- أحسب الربيع الأول والثاني والثالث، ماذا تلاحظ؟.

التمرين الثاني:

الجدول التالي يبين توزيع 1000 ساكن حسب المساحة بالكلم^2:

المساحة بالكلم^2	70-50	90-70	110-90	130-110	150-130	170-150	190-170
عدد السكان	80	150	280	200	150	80	60

المطلوب:

1- أحسب المتوسط، والوسيط والمنوال؟، وماذا تمثل هذه القيم؟.

2- بين شكل التواء المنحنى المتعلق بتوزيع البيانات.

3- أحسب العشير الخامس والسابع، ماذا تلاحظ؟.

4- أحسب الميئي الخمسين والسبعين، والثمانين، ماذا تلاحظ؟

التمرين الثالث:

يبين الجدول التالي مستوى مبالغ كراء 2000 منزل في مدينة ما لسنة 2007:

حجم مبالغ الكراء بـ: دج	-4000 5000	-5000 6000	-6000 7000	-7000 8000	-8000 9000	-9000 10000
عدد البنايات	184	306	436	410	376	288

المطلوب:

1- بين شكل توزيع مبالغ كراء المنازل.

2- أحسب العشير السادس والثامن، ثم الميئي الثمانين والتسعين، ماذا تلاحظ.

حل السلسلة الثالثة (النزعة المركزية):

حل التمرين الأول:

1- **متوسط** الإنتاج الشهري لعمال العينة المختارة هو :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{115 + 123 + 119 + 123 + 124 + 119 + 123 + 121 + 123 + 121}{10} = \frac{1211}{10} = 121.1$$

القيمة التي تزيد عنها **نصف** كمية الإنتاج: بما أن البيانات غير مبوبة أي في شكل سلسلة فإن القيمة هي:

- نرتب القيم تصاعديا: 115 119 119 121 121 123 123 123 123 124.
- نحدد رتبة الوسيط وهي: بما أن عدد المشاهدات 10 عدد زوجي فإن قيمة الوسيط هي الوسط الحسابي للقيمتين ذات الرتبتين 10/2 و (10+1)/2 أي الرتبتين 5 و6، قيمتاهما هما: 121 و123 ووسطهما الحسابي هو $\boxed{122}$، وهو يساوي وسيط البيانات أي القيمة التي تزيد عنها نصف كمية الإنتاج.

المنوال وبما أنها بيانات في شكل سلسلة فإن المنوال هو القيمة الأكثر تكرارا وهي بالملاحظة $\boxed{123}$.

نلاحظ أن 123>122>121.1 أي: المنوال > الوسيط > الوسط، ومنه فإن المنحنى ملتوي جهة اليسار.

2- إذا أضفنا قيمة = 5 لكل القيم السابقة تصبح: 120 128 124 128 129 124 128 126 128 126، والوسط الحسابي المعدل يصبح

$$\bar{Y}=\frac{120+128+124+128+129+124+128+126+128+126}{10}=\frac{1261}{10}=126.1$$

نلاحظ أن: $\bar{Y} = \bar{X} + 5 \Leftarrow 126.1 = 121.1 + 5$

3- إذا ضربنا قيمة = 3 في كل القيم السابقة تصبح 345 369 357 369 372 357 369 363 369 363، والوسط الحسابي يصبح:

$$\bar{Y}=\frac{345+369+357+369+372+357+369+363+369+363}{10}=\frac{3633}{10}=3633$$

نلاحظ أن: $\bar{Y} = 3\times\bar{X} \Leftarrow 363.3 = 121.1 \times 3$

4- حساب مجموع مربعات انحراف القيم عن وسطها الحسابي:

$\sum(x_i-\bar{X})^2=(115-121.1)^2+.....+(121-121.1)^2=68.9$ ونلاحظ أن: $\sum(x_i-\bar{X})\cong 0$

5- حساب الربيع الأول والثاني والثالث:

- تحديد رتبة الربيع الأول: $R=(n+1)\times(\frac{i}{4})=11\times\frac{1}{4}=2.75$
- الرتبة K التي تسبق رتبة الربيع الأول R=2.75 هي: K=2. والقيمة التي تسبق قيمته هي $X_K= X_2=119$.
- الرتبة U التي تلي رتبة الربيع الأول R=2.75 هي: U=3. والقيمة التي تلي قيمته $X_U= X_3=119$.
- قيمة **الربيع الأول** هي: $Q_3=x_K+(R-K)\times(x_U-x_K)=119+(2.75-2)\times(119-119)=119$
- تحديد رتبة الربيع الثاني: $R=(n+1)\times(\frac{i}{4})=11\times\frac{2}{4}=5.5$
- الرتبة K التي تسبق رتبة الربيع الثاني R=5.25 هي: K=5. والقيمة التي تسبق قيمته هي $X_K= X_5=121$.
- الرتبة U التي تلي رتبة الربيع الثاني R=5.25 هي: U=6. والقيمة التي تلي قيمته $X_U= X_6=123$.

- قيمة الربيع **الثاني** هي: $Q_3 = x_K + (R-K)\times(x_U - x_K) = 121+(5.5-5)\times(123-121)=122$
- تحديد رتبة الربيع الثالث: $R=(n+1)\times(\frac{i}{4})=11\times\frac{3}{4}=8.25$
- الرتبة K التي تسبق رتبة الربيع الثالث R=8.25 هي: K=8. والقيمة التي تسبق قيمته هي $X_K= X_8=123$.
- الرتبة U التي تلي رتبة الربيع الثالث R=8.25 هي: U=9. والقيمة التي تلي قيمته $X_U= X_9=123$.
- قيمة الربيع **الثالث** هي: $Q_3 = x_K + (R-K)\times(x_U - x_K) = 123+(8.25-8)\times(123-123)=123$

نلاحظ أن الربيع الثاني يساوي الوسيط.

حل التمرين الثاني:

1- حساب مقاييس النزعة المركزية الثلاثة.

فئات المساحة بالكلم^2	عدد السكان (التكرارات f)	مراكز الفئات X_i	$f_i x_i$	صاعد	نازل
70-50	80	60	4800	80	1000
90-70	150	80	12000	230	920
110-90	280	100	28000	**510**	770
130-110	200	120	24000	710	490
150-130	150	140	21000	860	290
170-150	80	160	12800	940	140
190-170	60	180	10800	1000	60
المجموع	1000		113400		

$\overline{X}=\frac{\sum_{i=1}^{7} x_i f_i}{n}=\frac{113400}{1000}=113.4$ إذن متوسط المساحة هو 113.4 كلم^2 لكل 1000 ساكن.

- المساحة التي يسكن فيها نصف عدد السكان تمثل وسيط المساحة ويحسب كما يلي:

بعد تكوين التكرار الصاعد، نحدد رتبة الوسيط وهي (n/2 = 2/1000= 500) والتكرار الصاعد الأكبر من هذه القيمة مباشرة هو تكرار الفئة الثالثة (90-110) وطولها L=20، ويتم حساب الوسيط بالمعادلة التالية:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.3$$

نصف عدد السكان (500) يسكنون في مساحة قدرها 109.3 كلم^2،

- حساب المنوال بالمعادلة التالية: (طريقة الفروقات):

$$\text{Mod} = A + \frac{d_1}{d_1 + d_2} \times L = 90 + \frac{280 - 150}{(280-150) + (280-200)} \times 20 = 1024$$

معظم السكان يسكنون في مساحة قدرها: 102.4 كلم^2.

2- نلاحظ أن 102.4<109.3<113.4 أي: الوسط > الوسيط > المنوال ، ومنه فإن المنحنى ملتوي جهة اليمين.

3- حساب العشير الخامس والسابع:

- العشير الخامس: $D_5 = A + \frac{\frac{n}{10} \times 5 - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.29$
- العشير السابع:
- $D_7 = A + \frac{\frac{n}{10} \times 7 - f_1}{f_2 - f_1} \times L = 110 + \frac{700 - 510}{710 - 510} \times 20 = 129$
- الميئي الخمسين: $P_{50} = A + \frac{\frac{n}{100} \times 50 - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.29$
- الميئي السبعين: $P_{70} = A + \frac{\frac{n}{100} \times 70 - f_1}{f_2 - f_1} \times L = 110 + \frac{700 - 510}{710 - 510} \times 20 = 129$
- الميئي الثمانين: $P_{80} = A + \frac{\frac{n}{100} \times 80 - f_1}{f_2 - f_1} \times L = 130 + \frac{800 - 710}{860 - 710} \times 20 = 142$

- نلاحظ أن الميئي الخمسين يساوي العشير الخامس ويساوي الوسيط؛ وأن الميئي السبعين يساوي العشير السابع.

حل التمرين الثالث:

1- لتبيان شكل التوزيع نحسب الوسط، الوسيط والمنوال، ونحتاج لحساب مراكز الفئات والتكرار المتجمع الصاعد:

نازل	صاعد	$f_i x_i$	مراكز الفئات X_i	عدد السكان (التكرارات f)	فئات المبالغ دج
2000	184	828000	4500	184	5000-4000
1816	490	1683000	5500	306	6000-5000
1510	926	2834000	6500	436	7000-6000
1074	1336	3075000	7500	410	8000-7000
664	1712	3196000	8500	376	9000-8000
288	**2000**	2736000	9500	288	10000-9000
		14352000		2000	المجموع

- الوسط الحسابي: $\overline{X} = \frac{\sum_{i=1}^{6} x_i f_i}{n} = \frac{14352000}{2000} = 7176$

- الوسيط: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7000 + \frac{1000 - 926}{1336 - 926} \times 1000 = 7180.5$

- المنوال: $Mod = A + \frac{d_1}{d_1 + d_2} \times L = 6000 + \frac{436-306}{(436-306)+(436-410)} \times 1000 = 6833.3$

نلاحظ أن $7180.5 > 7176 > 6833.3$ أي: الوسيط > الوسط > المنوال ، ومنه فإن المنحنى ملتوي جهة اليمين.

2- حساب العشير السادس والثامن، والميئي الثمانين والتسعين: ونلاحظ أن العشير الثامن يساوي الميئي الثمانين

- العشير السادس: $D_6 = A + \frac{\frac{n}{10} \times 6 - f_1}{f_2 - f_1} \times L = 7000 + \frac{1200 - 926}{1336 - 926} \times 1000 = 7668.3$

- العشير الثامن: $D_8 = A + \dfrac{\frac{n}{10}\times 8 - f_1}{f_2 - f_1}\times L = 8000 + \dfrac{1600-1336}{1712-1336}\times 1000 = 8702.13$
- الميئي الثمانين: $P_{80} = A + \dfrac{\frac{n}{100}\times 80 - f_1}{f_2 - f_1}\times L = 8000 + \dfrac{1600-1336}{1712-1336}\times 1000 = 8702.13$
- الميئي السبعين: $P_{90} = A + \dfrac{\frac{n}{100}\times 90 - f_1}{f_2 - f_1}\times L = 9000 + \dfrac{1800-1712}{2000-1712}\times 1000 = 9305.6$

السلسلة الرابعة (التشتت والتشكل):

التمرين الأول:

إذا كانت الطاقة التصديرية لعشر محطات لتحلية المياه بالمليون متر مكعب كما يلي:

17 10 20 19 12 16 18 14 13 11 15

المطلوب:

1- أوجد المدى العام الذي تتراوح فيه قيم الطاقة التصديرية، ثم أوجد المدى الربيعي ثم الانحراف الربيعي.

2- أحسب الانحراف المتوسط ، والتباين ثم الانحراف المعياري.

3- حدد شكل التواء وتفرطح هذه البيانات (بدون حساب المنوال).

التمرين الثاني:

فيما يلي الأجور الشهري بالألف دينار لـ:40 عامل في مؤسسة ما.

فئات الأجر	5-3	7-5	9-7	11-9	13-11	15-13	17-15	19-17
عدد العمال	3	4	6	7	8	5	4	3

المطلوب:

1- أوجد المدى والمدى الربيعي والانحراف الربيعي، ثم الانحراف المتوسط.

2- أحسب التباين والانحراف المعياري.

3- حدد شكل التواء البيانات بطريقة بيرسون ثم بطريقة الميئين.

4- هل شكل البيانات مدبب أم مفرطح.

التمرين الثالث:

فيما يلي 3 أنواع من المصابيح وسحبنا من كل نوع 50 مصباحا وقمنا بقياس مدة اشتعالها قبل أن تحترق:

العينة الأولى	فئات الساعات	3-1	5-3	7-5	9-7	11-9	13-11
	عدد المصابيح	2	4	6	18	15	5
العينة الثانية	فئات الساعات	3-1	5-3	7-5	9-7	11-9	13-11
	عدد المصابيح	7	9	10	9	8	7
العينة الثالثة	فئات الساعات	3-1	5-3	7-5	9-7	11-9	13-11
	عدد المصابيح	2	7	9	14	13	5

المطلوب:

1- أوجد المدى العام لكل نوع، ثم المدى الربيعي والانحراف الربيعي، ثم الانحراف المتوسط.

2- ما هو أحسن نوع يجب استعماله من المصابيح؟(النوع الذي تشتته صغير وأحسن مقياس للتشتت هو الانحراف المعياري).

3- بالنسبة للنوع الأحسن:

- حدد شكل التواء توزيع البيانات بطريقتين ؟.
- هل شكل توزيع البيانات مدبب أم مفرطح؟.

حل السلسلة الرابعة (التشتت والتشكل):

حل التمرين الأول:

1- المدى العام الذي تتراوح فيه قيم الطاقة التصديرية يساوي الفرق بين أكبر قيمة وأصغر قيمة

المدى = 20-10 =10

ولحساب المدى الربيعي والانحراف الربيعي نحتاج لحساب الربيع الأول والربيع الثالث:

أولا نرتب القيم تصاعديا: 10 11 12 13 14 15 16 17 18 19 20

- تحديد رتبة الربيع الأول: $R = (n+1) \times (\frac{i}{4}) = 12 \times \frac{1}{4} = 3$

وبما أن R عدد صحيح فإن قيمة الربيع الأول هي: $Q_1 = X_R = X_3 = 12$.

- تحديد رتبة الربيع الثالث: $R = (n+1)\times(\frac{i}{4}) = 12\times\frac{3}{4} = 9$

وبما أن R عدد صحيح فإن قيمة الربيع الأول هي: $Q_3 = X_R = X_9 = 18$.

إذن المدى الربيعي هو: $IQ = Q_3 - Q_1 = 17 - 11 = 6$ أي أن 50% من قيم الطاقة التصديرية تتراوح في مجال قدره 6 مليون متر مكعب. والانحراف الربيعي هو: $Q = \frac{IQ}{2} = \frac{Q_3 - Q_1}{2} = \frac{6}{2} = 3$ أي أن نصف عدد قيم الطاقة التصديرية لا تبعد عن القيمة الوسيطية إلا بمقدار 3 مليون متر مكعب.

2- حساب الانحراف المتوسط والتباين والانحراف المعياري:

- تكوين الجدول التكراري: حيث أن الوسط الحسابي = 15

المجموع	15	11	13	14	18	16	12	19	20	10	17	القيم
0	0	4-	2-	1-	3	1	3-	4	5	5-	2	$x_i - \overline{X}$
30	0	4	2	1	3	1	3	4	5	5	2	$\lvert x_i - \overline{X}\rvert$
110	0	16	4	1	9	1	9	16	25	25	4	$(x_i - \overline{X})^2$
1958	0	256	16	1	81	1	81	256	625	625	16	$(x_i - \overline{X})^4$

- حساب الانحراف المتوسط $MD = \frac{\sum\lvert x_i - \overline{X}\rvert}{n} = \frac{30}{11} = 2.73$.
- حساب التباين: $s^2 = \frac{\sum(x_i - \overline{X})^2}{n} = \frac{110}{11} = 10$ والانحـــــــــــراف المعيـــــــــــاري:

$$s = \sqrt{s^2} = \sqrt{10} = 3.16$$

3- يتحدد شكل التواء البيانات بمعامل بيرسون أو معامل الميئين، أما شكل تفرطح البيانات فيتحدد بمعامل التفرطح:

الوسط=15، والوسيط رتبته $\frac{n+1}{2} = 6$ (لأن n فردي) وبالتالي قيمته هي قيمة الرتبة 6 بعد ترتيب القيم: Med =15.

الهاتف: 0021332120740 | 0021332120720 webmaster@univ-eloued.dz

- معامل بيرسون α معدوم: $a=\frac{3(\overline{X}-Med)}{s}=\frac{3(15-15)}{3.16}=0$ وبالتالي فإن منحنى التوزيع التكراري متماثل.

- معامل التفرطح k: $k=\frac{\frac{1}{n}\sum(x_i-\overline{X})^4}{S^4}=\frac{\frac{1}{11}\times 1958}{100}=1.78$

- نلاحظ أن k<3 وبالتالي فإن منحنى التوزيع التكراري منبسط أي مفرطح.

حل التمرين الثاني:

1- تحديد المدى، المدى الربيعي، الانحراف الربيعي والانحراف المتوسط، وقبل ذلك نضع جدول لتوزيع التكراري:

فئات الأجر	f	x_i	$f\uparrow$	$x_i f_i$	$\lvert x_i-\overline{X}\rvert$	$\lvert x_i-\overline{X}\rvert\times f$	$(x_i-\overline{X})^2$	$f\times(x_i-\overline{X})^2$	$f\times(x_i-\overline{X})^4$
5-3	3	4	2	12	7	21	49	147	7203
7-5	4	6	7	24	5	20	25	100	2500
9-7	6	8	13	48	3	18	9	54	486
11-9	7	10	20	70	1	7	1	7	7
13-11	8	12	28	96	1	8	1	8	8
15-13	5	14	33	70	3	15	9	45	405
17-15	4	16	37	64	5	20	25	100	2500
19-17	3	18	40	54	7	21	49	147	7203
المجموع	40			438	32	130	168	608	20312

- المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى = 115-35 = 80.

- الانحراف الربيعي يحسب بحساب الربيع الأول والثالث:

$$Q_1=\mathrm{A}+\frac{\frac{n}{4}\times 1-f_1}{f_2-f_1}\times L=7+\frac{10-7}{13-7}\times 2=8$$

$$Q_3=\mathrm{A}+\frac{\frac{n}{4}\times 3-f_1}{f_2-f_1}\times L=13+\frac{30-28}{33-28}\times 2=13.8$$

المدى الربيعي: $IQ = Q_3 - Q_1 = 5.8$ والانحراف الربيعي: $Q = \frac{IQ}{2} = 2.9$

الانحراف المتوسط للبيانات المبوبة هو $\boxed{MD = \frac{\sum |x_i - \overline{X}| \times f}{n}}$ حيث $\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i = n} = \frac{438}{40} \cong 11$

وبالتالي الانحراف المتوسط: $MD = \frac{\sum |x_i - \overline{X}| \times f}{n} = \frac{130}{40} = 3.25$

2- حساب التباين: $s^2 = \frac{\sum f \times (x_i - \overline{X})^2}{n} = \frac{608}{40} = 15.2$ والانحراف المعياري:

$$s = \sqrt{s^2} = \sqrt{15.2} = 3.9$$

3- تحديد شكل التواء البيانات بطريقة الميئين: (طريقة بيرسون رأيناها في التمرين الأول):

$\alpha_{i\,,\,100-i} = \frac{(P_{100-i} - P_{50}) - (P_{50} - Pi)}{P_{100-i} - Pi}$ فنحسب الميئين P_{20} والميئين P_{80} مثلا:

- الميئي العشرين: $P_{20} = A + \frac{\frac{n}{100} \times 20 - f_1}{f_2 - f_1} \times L = 7 + \frac{8-7}{13-7} \times 2 = 7.33$
- الميئي الثمانين: $P_{80} = A + \frac{\frac{n}{100} \times 80 - f_1}{f_2 - f_1} \times L = 13 + \frac{32-28}{33-28} \times 2 = 14.6$

4- نحدد شكل تفرطح البيانات بحساب معامل التفرطح بالمعادلة التالية:

$$k = \frac{\frac{1}{n}\sum (x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{40} \times 20312}{3.899^4} = 2.2$$

نلاحظ أن k>3 وبالتالي فإن منحنى التوزيع التكراري مدبب.

حل التمرين الثالث:

1- تحديد المدى العام لكل نوع:

- نلاحظ أن المدى لكل نوع نفسه = الحد الأعلى للفئة الأخيرة – الحد الأعلى للفئة الأولى = 1200=100-1300
- لحساب المدى الربيعي والانحراف الربيعي نكون جدول التوزيع التكراري:

فئات الساعات	مركز	$f1$	$f2$	ب$f3$	ت ص 1	ت ص 2	ت ص 3	النوع الأول $(x_i-\overline{X})^2$	النوع الثاني $(x_i-\overline{X})^2$	النوع الثالث $(x_i-\overline{X})^2$
3-1	2	2	7	2	2	7	2	38.44	24.21	0.58
5-3	4	4	9	7	6	16	9	17.64	8.53	1.54
7-5	6	6	10	9	12	26	18	4.84	0.85	5.02
9-7	8	18	9	14	30	35	32	0.04	122.77	1.54
11-9	10	15	8	13	45	43	45	3.24	65.29	0.06
13-11	12	5	7	5	50	50	50	14.44	3.69	0.58
المجموع		50	50	50						
الوسط الحسابي لكل نوع		8.2	6.9	7.8						

الفئات	النوع الأول $f(x_i-\overline{X})^2$	النوع الثاني $f(x_i-\overline{X})^2$	النوع الثالث $f(x_i-\overline{X})^2$	النوع الأول $f(x_i-\overline{X})^4$	النوع الثاني $f(x_i-\overline{X})^4$	النوع الثالث $f(x_i-\overline{X})^4$
1	76.88	169.44	1.16	1477.63	585.95	0.33
2	70.56	76.74	10.76	311.17	72.70	2.36
3	29.04	8.46	45.16	23.43	0.72	25.18
4	0.72	1104.90	21.53	0.00	15071.59	2.36
5	48.60	522.29	0.75	10.50	4262.31	0.00
6	72.20	25.80	2.89	208.51	13.59	0.33
المجموع	298.00	1907.64	82.24	2031.24	20006.86	30.58

المدى الربيعي والانحراف الربيعي للنوع الأول:

- نحسب الربيع الأول: $Q_1 = A + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 7 + \dfrac{12.5-12}{30-12}\times 2 = 7.06$
- نحسب الربيع الثالث: $Q_3 = A + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-30}{45-30}\times 2 = 10$

بالنسبة للنوع الأول: المدى الربيعي: $IQ = Q_3 - Q_1 = 2.94$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 1.47$

المدى الربيعي والانحراف الربيعي للنوع الثاني :

- نحسب الربيع الأول: $Q_1 = \text{A} + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 3 + \dfrac{12.5-7}{16-7}\times 2 = 4.22$
- نحسب الربيع الثالث: $Q_3 = \text{A} + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-35}{43-35}\times 2 = 9.63$

بالنسبة للنوع الثاني: المدى الربيعي: $IQ = Q_3 - Q_1 = 5.41$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 2.75$

<u>المدى الربيعي والانحراف الربيعي للنوع الثالث:</u>

- نحسب الربيع الأول: $Q_1 = \text{A} + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 5 + \dfrac{12.5-9}{18-9}\times 2 = 5.8$
- نحسب الربيع الثالث: $Q_3 = \text{A} + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-32}{45-32}\times 2 = 9.8$

بالنسبة للنوع الثالث: المدى الربيعي: $IQ = Q_3 - Q_1 = 4$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 2$

2- حساب الانحراف المعياري:

بالنسبة للنوع الأول: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{298}{50}} = 2.44$

بالنسبة للنوع الثاني: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{1907.64}{50}} = 6.18$

بالنسبة للنوع الثالث: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{82.24}{50}} = 1.28$

ونعرف أن الانحراف المعياري من عيوبه أنه يتأثر بالقيم الشاذة ومن مزاياه أنه يعتمد على جميع القيم، بينما الانحراف الربيعي يمتاز بعدم تأثره بالقيم الشاذة ويعاب عليه اعتماده على قيمتين فقط هما الربيع الأول والثالث واللتان تعتمدان بدورهما على قيمتين فقط

رغم أن مقياس التشتت: الانحراف الربيعي كان في النوع الأول أصغر ويبين أن أحسن نوع هو النوع الأول وبما أنه لا توجد قيم شاذة واضحة فمقياس الانحراف المعياري يبين أن أحسن نوع هو النوع الثالث.

3- شكل الالتواء بطريقة بيرسون:

بالنسبة للنوع الأول فإن: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7 + \frac{25-12}{30-12} \times 2 = 8.44$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(8.2 - 8.44)}{2.44} = -0.3$$

$\alpha < 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

بالنسبة للنوع الثاني فإن: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 5 + \frac{25-16}{26-16} \times 2 = 6.8$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(6.9 - 6.8)}{6.18} = 0.05$$

$\alpha > 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).

بالنسبة للنوع الثالث فإن: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7 + \frac{25-18}{32-18} \times 2 = 8$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(8 - 7.8)}{1.28} = 0.47$$

$\alpha > 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).

شكل تفرطح البيانات:

- معامل التفرطح k بالنسبة للنوع الأول: $k = \frac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{50} \times 2031.24}{2.44^4} = 1.14$
- معامل التفرطح k بالنسبة للنوع الثاني: $k = \frac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{50} \times 20006.86}{6.18^4} = 0.27$
- معامل التفرطح k بالنسبة للنوع الثالث: $k = \frac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{50} \times 30.58}{1.28^4} = 0.23$

نلاحظ أن k>3 بالنسبة للتوزيعات التكرارية للأنواع الثلاثة، وبالتالي فإن المنحنيات كلها منبسطة أي مفرطحة.

السلسلة الخامسة (الانحدار):

التمرين الأول:

إذا كان انحدار الرضا الوظيفي لدى العامل على أجره الشهري ممثلا بالمعادلة التقديرية التالية:

$$Y = B_0 + B_1 X + e$$

المطلوب:

1- فسر كل متغير ومعلمة في المعادلة السابقة.

2- مالفائدة من استخدام المعادلات التقديرية لانحدار العلاقات بين المتغيرات.

3- إذا كان: $B_1 = 0.8$ $B_0 = 2$ أرسم خط الانحدار.

التمرين الثاني:

فيما يلي عدد ساعات المذاكرة يوميا لـ 8 طلاب وعلاماتهم في مقياس معين:

الطلبة	1	2	3	4	5	6	7	8
عدد ساعات المذاكرة	2	3	4	5	6	8	9	7
علامة الاختبار	12	13	14	15	16	17	19	20

المطلوب:

1- حدد المتغير المستقل والمتغير التابع.

2- أوجد المعادلة التقديرية لانحدار x على y.

3- ماذا تمثل كل معلمة(B_1B_2).

4- إذا كان إذا كان عدد ساعات المذاكرة للطالب هو 10، فقدر علامة اختبار الطالب.

التمرين الثالث:

فيما يلي المساحة المزروعة بالأعلاف بالألف هكتار وكمية إنتاج اللحوم بالألف طن في الفترة (1995-2002):

السنوات	1995	1996	1997	1998	1999	2000	2001	2002
المساحة المزروعة	305	313	297	289	233	214	240	217
كمية اللحوم المنتجة	592	603	662	607	635	699	719	747

المطلوب:

1. حدد المتغير المستقل والمتغير التابع.
2. أوجد المعادلة التقديرية لانحدار x على y.
3. ماذا تمثل كل معلمة(B_1B_2).
4. إذا كانت المساحة المزروعة في سنة 2010 تساوي 400 ألف هكتار، فقدر كمية اللحوم المنتجة سنة 2010.

السلسلة السادسة (الارتباط):

التمرين الأول:

فيما يلي عدد ساعات المذاكرة يوميا لـ 8 طلاب وعلاماتهم في مقياس معين:

الطلبة	1	2	3	4	5	6	7	8
عدد ساعات المذاكرة	2	3	4	5	6	8	9	7
علامة الاختبار	12	13	14	15	16	17	19	20

المطلوب:

1. أوجد قيمة معامل الارتباط البسيط بين عدد ساعات المذاكرة وعلامة الاختبار؟.
2. ما هي درجة قوة العلاقة بينهما وما نوعها؟.

التمرين الثاني:

فيما يلي المساحة المزروعة بالأعلاف بالألف هكتار وكمية إنتاج اللحوم بالألف طن في الفترة (1995-2002):

السنوات	1995	1996	1997	1998	1999	2000	2001	2002
المساحة المزروعة	305	313	297	289	233	214	240	217
كمية اللحوم المنتجة	592	603	662	607	635	699	719	747

المطلوب:

1. أوجد قيمة معامل الارتباط البسيط بين المساحة المزروعة وكمية اللحوم المنتجة؟.
2. ما هي درجة قوة العلاقة بينهما وما نوعها؟.

التمرين الثالث:

فيما يلي إجابات 12 طالب حول الأداء التعليمي لأستاذين (نوعين من الإجابات: ملاحظات وعلامات):

لما كان السؤال عن **إعطاء علامة** لكل أستاذ من طرف الطلبة كانت **العلامات** كما في الجدول الأول.

ولما كان السؤال عن **رأي الطلبة** حول أداء الأستاذين كانت **الملاحظات** كما في الجدول الثاني.

الجدول الأول

الطلبة	أستاذ 1	أستاذ 2
1	F	B
2	A	H
3	L	J
4	H	L
5	C	G
6	J	A
7	B	K
8	I	F
9	G	I
10	D	C
11	E	E
12	K	D

الجدول الثاني

الطلبة	أستاذ 1	أستاذ 2
1	ضعيف	حسن
2	جيد	ضعيف
3	ممتاز	ممتاز
4	متوسط	جيد
5	حسن	ضعيف
6	جيد	ممتاز
7	حسن	ممتاز
8	ممتاز	جيد
9	جيد	ضعيف
10	ضعيف	جيد
11	ضعيف	حسن
12	جيد	حسن

أحسب معامل الارتباط وحدد نوع العلاقة وقوتها؟. (في كل من الجدولين)

حل السلسلتين الخامسة والسادسة (الارتباط والانحدار):

حل التمرين الأول:

1. المتغير المستقل هو المساحة المزروعة والتي تؤثر في كمية إنتاج اللحوم وهو المتغير التابع:

$\sum y^2$	$\sum x^2$	$\sum x_i y_i$	الكمية Y	المساحة X	السنوات
350464.00	93025.00	180560.00	592.00	305.00	1995
363609.00	97969.00	188739.00	603.00	313.00	1996
438244.00	88209.00	196614.00	662.00	297.00	1997
368449.00	83521.00	175423.00	607.00	289.00	1998
403225.00	54289.00	147955.00	635.00	233.00	1999
488601.00	45796.00	149586.00	699.00	214.00	2000
516961.00	57600.00	172560.00	719.00	240.00	2001
558009.00	47089.00	162099.00	747.00	217.00	2002
3487562.00	567498.00	1373536.00	5264.00	2108.00	المجموع

لدينا الوسط الحسابي لـx: $\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n}=\frac{2108}{8}=263.5$

لدينا الوسط الحسابي لـy: $\overline{y}=\frac{\sum_{i=1}^{n} y_i}{n}=\frac{5264}{8}=658$

- نستطيع حساب معامل الارتباط بطريقة بيرسون بدون حساب الوسط الحسابي:

$$R_p=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-\sum(x_i)^2}\sqrt{n\sum y_i^2-\sum(y_i)^2}}=\frac{8(1373536)-(11096512)}{135564.95}=-0.789$$

- نلاحظ أن معامل الارتباط سالب وبالتالي فالعلاقة عكسية بين المساحة المزروعة كمية إنتاج اللحوم.
- نلاحظ من قيمة معامل الارتباط أن العلاقة قوية بين المساحة المزروعة وكمية إنتاج اللحوم.

2. إيجاد المعادلة التقديرية لانحدار y على x: وذلك بإيجاد المعلمات.

$$\beta_1=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2}=\frac{-108224}{190800}=-0.57$$

$$\beta_0=\frac{\sum y_i-\beta_1\sum x_i}{n}=\frac{5264+0.57(2108)}{8}=807.46$$

إذن المعادلة التقديرية لانحدار x على y هي: $y=\beta_0+\beta_1 x=807.46-0.57x$

3. إذا كانت المساحة المزروعة سنة 2003 تساوي 400 ألف هكتار، فإن:

$$y=807.46-0.57(400)=580.58$$

كمية اللحوم المنتجة سنة 2010 هي: 580.58 ألف طن.

حل التمرين الثاني:

1. المتغير المستقل هو عدد ساعات المذاكرة والتي تؤثر في علامة الطالب وهو المتغير التابع:

الطالب	الساعات x	العلامة y	$\sum x_i y_i$	$\sum x^2$	$\sum y^2$
1	2	12	24.00	4.00	144.00
2	3	13	39.00	9.00	169.00
3	4	14	56.00	16.00	196.00
4	5	15	75.00	25.00	225.00
5	6	16	96.00	36.00	256.00
6	8	17	112.00	64.00	196.00
7	9	19	171.00	81.00	361.00
8	7	20	140.00	49.00	400.00
المجموع	44.00	123.00	713.00	284.00	1947.00

لدينا الوسط الحسابي لـx: $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{44}{8} = 5.5$

لدينا الوسط الحسابي لـy: $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{123}{8} = 15.38$

- نستطيع حساب معامل الارتباط بطريقة بيرسون بدون حساب الوسط الحسابي:

$$R_p = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \sum (x_i)^2}\sqrt{n\sum y_i^2 - \sum (y_i)^2}} =$$

$$\frac{8(713) - (5412)}{387.55} = 0.7535$$

- نلاحظ أن معامل الارتباط موجب وبالتالي فالعلاقة طردية بين ساعات الاستذكار وعلامة الاختبار.
- نلاحظ من قيمة معامل الارتباط أن العلاقة قوية بين المساحة المزروعة وكمية إنتاج اللحوم.

2. إيجاد المعادلة التقديرية لانحدار x على y: وذلك بإيجاد المعالم.

$$\beta_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \sum (x_i)^2} = \frac{292}{447} = 0.65$$

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n} = \frac{123 - 0.65(44)}{8} = 11.78$$

إذن المعادلة التقديرية لانحدار x على y هي: $y = \beta_0 + \beta_1 x = 11.78 - 0.65x$

3. إذا كانت المساحة المزروعة سنة 2003 تساوي 400 ألف هكتار، فإن

$$y = 11.78 - 0.65(10) = 18.31$$

علامة الطالب المقدرة لو ذاكر 10 ساعات هي: 18.31.

حل التمرين الثالث:

1. حساب معامل الارتباط وتحدد نوع العلاقة وقوتها في الجدول الأول:

- أولا ترتيب العلامات كما يلي:

رتب العلامات	ترتيب العلامات لأستاذ 2	رتب العلامات	ترتيب العلامات للأستاذ 1
1	A	1	A
2	B	2	B
3	C	3	C
4	D	4	D
5	E	5	E
6	F	6	F
7	G	7	G
8	H	8	H
9	I	9	I
10	J	10	J
11	K	11	K
12	L	12	L

- تكوين جدول الرتب لحساب مجموع مربعات الفروق بين رتب الملاحظات: $\sum d^2$

d^2	d	رتب2	رتب 1	أستاذ 2	أستاذ 1	الطلبة
16	4	2	6	B	F	1
49	7	8	1	H	A	2
4	2	10	12	J	L	3
16	4	12	8	L	H	4
16	4	7	3	G	C	5
81	9	1	10	A	J	6
81	9	11	2	K	B	7
9	3	6	9	F	I	8
4	2	9	7	I	G	9
1	1	3	4	C	D	10
0	0	5	5	E	E	11
49	7	4	11	D	K	12
326						مجموع

-ونحسب المعامل كما يلي: $R_S = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6 \times 326}{12(144-1)} = -0.1399$

فنقول أن العلاقة بين الملاحظات للأستاذين عكسية وضعيفة جدا.

2. حساب معامل الارتباط وتحدد نوع العلاقة وقوتها في الجدول الثاني:

- أولا ترتيب الملاحظات كما يلي:

الطلبة	ترتيب الملاحظات للأستاذ 1	رتب الملاحظات	ترتيب الملاحظات لأستاذ 2	رتب الملاحظات
1	ضعيف	2	ضعيف	1.5
2	ضعيف		ضعيف	
3	ضعيف		حسن	3.5
4	حسن	4.5	حسن	
5	حسن		متوسط	6
6	متوسط	6	متوسط	
7	جيد	8.5	متوسط	
8	جيد		جيد	8.5
9	جيد		جيد	
10	جيد		ممتاز	11
11	ممتاز	11.5	ممتاز	
12	ممتاز		ممتاز	

- تكوين جدول الرتب لحساب مجموع مربعات الفروق بين رتب الملاحظات: $\sum d^2$

الطلبة	أستاذ 1	أستاذ 2	رتب 1	رتب 2	d	d^2
1	ضعيف	حسن	2	3.5	1.5	2.25
2	جيد	متوسط	8.5	6	2.5	6.25
3	ممتاز	ممتاز	11.5	11	0.5	0.25
4	متوسط	متوسط	6	6	0	0
5	حسن	ضعيف	4.5	1.5	3	9
6	جيد	ممتاز	8.5	11	2.5	6.25
7	حسن	ممتاز	4.5	11	6.5	42.25
8	ممتاز	جيد	11.5	8.5	3	9
9	جيد	ضعيف	8.5	1.5	7	49
10	ضعيف	جيد	2	8.5	6.5	42.25
11	ضعيف	متوسط	2	6	4	16
12	جيد	حسن	8.5	3.5	5	25
مجموع						207.50

-ونحسب المعامل كما يلي: $R_S = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6 \times 207.5}{12(144-1)} = 0.2745$

فنقول أن العلاقة بين الملاحظات للأستاذين طردية وضعيفة جدا.

السلسلة السابعة (الانحدار والارتباط):

التمرين الأول:

تقدير علامة الإحصاء	تقدير علامة الاقتصاد
A	A+
C+	D
D	C
D+	C
B+	A
C+	B
A+	B+
B	B
B+	C
B+	B

المطلوب:

أوجد نوع وقوة العلاقة بين تقدير العلامات في المقياسين؟.

(0.703)

التمرين الثاني:

أراد طالب أن يدرس العلاقة الارتباطية بين عدد المتخرجين من قسم العلوم الاقتصادية في جامعة ما والذين تمكنوا من العمل مباشرة فرصد الطالب البيانات التالية خلال 7 سنوات:

السنوات	عدد المتخرجين	عدد العاملين
2000	74	10
2001	65	9
2002	79	12
2003	77	15
2004	69	12
2005	82	16
2006	80	17

المطلوب:

اوجد نوع وقوة العلاقة بين عدد المتخرجين ومن تمكنوا من العمل مباشرة ؟ (0.79).

التمرين الثالث:

1- قدر علامة أحمد في اختبار ما إذا كانت نسبة توتره هي 94، حيث عند دراسة العلاقة بين (X) نسبة التوتر و(Y) علامات 8 طلاب حصلنا على:

$\sum x = 240$ ، $\sum xy = 510$ ، $\sum x^2 = 750$ ،

$\sum y^2 = 300$ ، $\sum y = 195$.

2- أوجد نوع العلاقة ودرجة قوتها بين التوتر وعلامة الاختبار، وأوجد نسبة تأثير X في Y.

(0.99-77.4)

التمرين الرابع:

1- هل يوجد ارتباط بين علامات تحصيل 8 طلبة في مقياسي الإحصاء والمحاسبة . إذا كانت نتائجهم كما يلي:

إحصاء	13	9	19	15	11	8	16	11
المحاسبة	15	7	17	15	10	9	14	10

2- هل يوجد ارتباط بين علامات تحصيل 8 طلبة في مقياسي الإحصاء والمحاسبة. إذا كانت نتائجهم كما يلي:

إحصاء	D	F	A	C	E	G	B	E
المحاسبة	B	F	A	B	D	E	C	D

الحلول (0.94 0.89)

تابع للسلسلة السابعة (الانحدار والارتباط):

التمرين الأول:

يمثل الجدول التالي توزيع عدد سنوات الخدمة لعينة من 7 عمال، وتقابل عدد سنوات الخدمة لكل عامل نسبة رضاه عن منصبه:

عدد سنوات الخدمة	نسبة الرضا الوظيفي
11	93
4	80
9	88
5	85
8	83
6	80
7	91

المطلوب:

1- أوجد المعادلة التقديرية لانحدار الرضا على سنوات الخدمة؟.

2- أرسم خط الانحدار من خلال معادلة الانحدار؟.

3- اشرح كلا من B_1 و B_0؟.

4- حدد نوع العلاقة بين المتغيرين؟.

5- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل الارتباط المناسب؟.

6- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل ارتباط سبيرمان للرتب؟. (ماذا تلاحظ علما أن معامل سبيرمان حالة خاصة من معامل بيرسون).

7- حدد نسبة تأثير عدد سنوات الخدمة في الرضا الوظيفي ونسبة تأثير باقي المتغيرات غير المدروسة؟.

التمرين الثاني: أراد أحد طلبة العلوم الاقتصادية سنة أولى LMD بالجامعة أن يدرس العلاقة بين عدد أو حجم الطلبة في الفوج ونسبة الذين يواصلون دراساتهم العليا فحصل على البيانات في الجدول المقابل:

عدد الطلبة	نسبة المواصلين للدراسات العليا
30	3
25	4
20	1
15	7
10	5
34	2
29	4
26	8

المطلوب:

1- أوجد المعادلة التقديرية للانحدار.

2- أرسم خط الانحدار من خلال معادلة الانحدار.

3- اشرح كلا من B_1 و B_0.

4- حدد نوع العلاقة بين المتغيرين.

5- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل الارتباط المناسب.

6- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل ارتباط سبيرمان للرتب (ماذا تلاحظ علما أن معامل سبيرمان حالة خاصة من معامل بيرسون).

7- حدد نسبة تأثير حجم أو عدد طلبة الفوج في نسبة المواصلين لدراساتهم العليا ونسبة تأثير باقي المتغيرات غير المدروسة.

السلسلة الثامنة

(الارتباط بين المتغيرات الوصفية ذات النوع الإسمي 2*2):

التمرين الأول:

لتحديد مدى علاقة العمل بالتدخين أي هل يرتبط تدخين الأفراد بالعمل أم لا وإلى أي درجة تم رصد النتائج التالية.

العمل \ التدخين	يدخن	لا يدخن	$\sum$
يعمل	20	20	40
لا يعمل	50	10	60
$\sum$	70	30	100

(الحل 0.34)

التمرين الثاني:

أراد طبيب أن يحدد مدى وجود علاقة بين التطعيم ومقاومة مرض معين في عينة من 60 مريضا، حيث وجد:

-من كل المرضى في العينة تم تطعيم 40 منهم فقط.

-من كل المرضى 20 فقط قاوم الأمراض.

مع العلم أن من بين الذين تم تطعيمهم يوجد 15 منهم قاوم الأمراض.

العمل \ التدخين	تم تطعيمهم	لم يتم تطعيم	$\sum$
قاوم	15	5	20
لم يقاوم	25	15	40
$\sum$	40	20	60

هل توجد علاقة بين التطعيم ومقاومة الأمراض في هذه العينة؟.

(الحل 0.12)

التمرين الثالث:

في عينة من 100 فرد وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 40 رجلا و 60 امرأة.
- حسب التوجه السياسي: 70 وطنيا.

علما أن عدد الرجال الذين توجههم وطني هو 30.

المطلوب:

2- أوجد قوة العلاقة بين المتغيرين؟.

(الحل 0.09)

التمرين الرابع:

اخترنا 50 طالبا من الجامعة وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 15 ذكور.
- حسب متغير فرع التخصص: 20 علميا.
- حسب متغير الحضور والغياب: 10 يتغيبون.

علما أن عدد الإناث في الفروع العلمية هو 10.

علما أن معدل الإناث اللواتي يتغيبن هو 6.

علما معدل الغيابات في الفروع العلمية هو 8.

المطلوب:

- أوجد قوة العلاقة بين متغيري الجنس وفرع التخصص؟.
- أوجد قوة العلاقة بين متغيري الجنس والغياب؟.
- أوجد قوة العلاقة بين متغيري الفرع والغياب؟.

(الحلول 0.16 0.00 0.20)

السلسلة التاسعة

(الارتباط بين المتغيرات ذات النوع الوصفي الإسمي أكثر من 2*2)

التمرين الأول:

في الجدول التالي بيانات لعينة من 80 أسرة حسب متغيري عمل الأم والتحصيل العلمي للأبناء.

التحصيل / عمل الأم	جيد	متوسط	ضعيف	$\sum$
تعمل	6	5	4	15
لا تعمل	30	15	20	65
$\sum$	31	25	24	80

(ملاحظة هذه البيانات وهمية ولا تمد للواقع بصلة)

المطلوب:

أوجد العلاقة بين عمل الأم والتحصيل العلمي للأبناء؟

التمرين الثاني:

أوجد العلاقة بين التدخين والمستوى التعليمي؟. في البيانات التالية لـ 200 فرد:

التدخين / المستوى	يدخن	لا يدخن	$\sum$
جامعي	30	40	70
ثانوي	40	20	60
أساسي	10	40	50
ابتدائي	15	5	20
$\sum$	95	105	200

التمرين الثالث:

أوجد العلاقة بين لون الزهور وقوة الرائحة؟. لـ 30 زهرة حسب بيانات الجدول التالي:

الرائحة \ لون الزهرة	أصفر	أبيض	أحمر	$\sum$
قوية	3	5	3	11
متوسطة	4	2	4	10
ضعيفة	4	1	4	9
	11	8	11	30

التمرين الرابع:

اخترنا 100 فرد من عنابة ووهران وعند تصنيفهم وجدنا:

- حسب متغير (الهجرة غير الشرعية): 20 ينوون الهجرة غير الشرعية.
- حسب المستوى المعيشي: 60 منخفض، 30 متوسط، 10 مرتفع.
- عدد من مستواهم المعيشي متوسط وينوون الهجرة غير الشرعية هو 3.
- عدد من مستواهم المعيشي منخفض وينوون الهجرة غير الشرعية هو16.

المطلوب:

1- أوجد قوة العلاقة بين متغيري المستوى المعيشي و الهجرة غير الشرعية؟.

(الحل 0.20)

التمرين الخامس:

درسنا 80 أسرة في مدينة ورقلة من خلال متغيرين هما المستوى المعيشي والتحصيل العلمي للأبناء (قيم افتراضية) فوجدنا:

- حسب التحصيل العلمي: 30 جيد، 20 متوسط، 30 ضعيف.
- حسب المستوى المعيشي: 50 منخفض، 20 متوسط، 10 مرتفع.
  - عدد ذوي التحصيل الجيد والمستوى المعيشي المنخفض هو17.
  - عدد ذوي التحصيل المتوسط والمستوى المعيشي المتوسط هو04.
  - عدد ذوي التحصيل الجيد والمستوى المعيشي المرتفع هو 01.
  - عدد ذوي التحصيل المتوسط والمستوى المعيشي المرتفع هو03

1- أوجد قوة العلاقة بين متغيري المستوى المعيشي والتحصيل العلمي؟. (الحل 0.31).

السلسلة العاشرة

(الارتباط بين المتغيرات ذات النوع الوصفي

الإسمي أكثر من 2*2)

المسألة:

أخذنا عينة من 50 عاملا في الجامعة وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 20 ذكور.
- حسب متغير المهنة: 5 رؤساء أقسام و15 أستاذا و30 إداريا.
- حسب المستوى التعليمي: 4 دكاترة 16 ماجستير و30 ليسانس.
- مع العلم أن عدد الأساتذة الذكور هو 5.
- عدد رؤساء الأقسام الذكور هو 4.
- عدد الذكور ذوي المستوى ماجستير هو 6.
- عدد الإناث ذوات المستوى ليسانس هو 19.
- عدد رؤساء الأقسام ذوي المستوى دكتور هو 4.
- عدد رؤساء الأقسام ذوي المستوى ماجستير هو 1.
- عدد الأساتذة ذوي المستوى ماجستير هو 13.
- عدد الإداريين ذوي المستوى ليسانس هو 16.

المطلوب:

1- أوجد قوة العلاقة بين متغيري الجنس والمهنة؟.

2- أوجد قوة العلاقة بين متغيري الجنس والمستوى التعليمي؟.

3- أوجد قوة العلاقة بين متغيري المهنة والمستوى التعليمي؟.

(الحلول 0.26 0.21 0.76)

سلسلة التوزيع الطبيعي (المعياري):

تمرين رقم 01:

أوجد المساحة تحت التوزيع الطبيعي المعياري (القياسي)؟. في القيم التالية:

$(-1<Z<+1)$ $(-2<Z<+2)$ $(-3<Z<+3)$

$(0<Z<0.88)$ $(1.6<Z<2.55)$ $(Z<1.6)$

$(2.55<Z)$ **خارج المجال** $(-1.6<Z<2.55)$.

تمرين رقم 02: (استعمل قاعدة التحويل)

إذا كان المتوسط الحسابي لأوزان الطلبة هو 70 بانحراف معياري 07:

- ما هو احتمال أن يكون وزن الطالب أقل من 80.
- ما هو احتمال أن يكون وزن الطالب أقل من 84.
- ما هو احتمال أن يكون وزن الطالب أقل من 63.
- ما هي نسبة الطلبة الذين تتراوح أوزانهم بين 55 و85.

تمرين رقم 03: (استعمل قاعدة التحويل)

يتبع دخل الأسر توزيعا طبيعيا بوسط حسابي 30.000دج وانحراف معياري 9.000دج.

حددت قيمة 12.000دج كعتبة للفقر، وأن دخل الفئة المتوسطة يتراوح بين 15.000دج و60.000. وأكثر من 60.0000 هم فئة الأغنياء.

- فما هي نسبة كل فئة في المجتمع؟.

تمرين رقم 04: (استعمل قاعدة التحويل)

كان لدينا متوسط أطوال الطلبة في كلية الرياضة 165 سم وانحراف معياري 5سم.

- الطول المطلوب في رياضة كرة القدم بين 157سم و172سم.
- الطول المطلوب في كرة السلة هو أكثر من 172سم.
- باقي الطلبة ينضمون لرياضة كرة اليد.

إذا علمت أن عدد الطلبة في الكلية هو 4000 طالب فما هو عدد الطلبة الذين سيشاركون في الرياضات التالية (كرة القدم، كرة السلة، كرة اليد، الشطرنج).

قائمة المراجع

قائمة المراجع

أولا: الكتب باللغة العربية

-أبو صالح محمد صبحي ، **الطرق الإحصائية** ، دار اليازوري العلمية للنشر والتوزيع ، الطبعة الاولى ، 2000.

- ادوارد مينيكا وزوريانا كورزيج ، **الاحصاء في الادارة مع التطبيق على الحاسب الالي** ، تعريب سرور علي ابراهيم سرور ، الطبعة الثانية ، دار المريخ ، الرياض ، بدون سنة النشر .
- إمتثال محمد حسن ومحمد علي محمد احمد ، **مبادئ الاستدلال الاحصائي** ، الدار الجامعية ، الاسكندرية ،2000 .
- جيلالي جلاطو ، **الاحصاء التطبيقي مع تمارين ومسائل محلولة** ، دار الخلدونية ، الجزائر ، 2007
- حسين علي بخيت وسحر فتح الله ، **مقدمة في الاقتصاد القياسي** ، الدار الجامعية للطباعة والنشر ، بغداد ، 2002.
- دلال القاضي واخرون ، **الاحصاء للاداريين والاقتصاديين** ، دار الحامد للنشر والتوزيع ، عمان ، 2003.
- دومينيك سالفادور ، **الاحصاء والاقتصاد القياسي** ، ديوان المطبوعات الجامعية ، الجزائر ، بدون سنة نشر .
- سعد الدين ابة الفتوح الشرنوبي ، **المفاهيم والمعالجات الاساسية في الاحصاء** ، مكتبة الاشعاع ، الاسكندرية ، الطبعة الاولى ، 2001 .
- صلاح الدين الهيتي ، **الاساليب الاحصائية في العلوم والادارة** ، الطبعة الاولى ، دار وائل للطباعة والنشر ، عمان ،2004.
- عامر احمد عامر ، **محاضرات في الاحصاء 2** ، دار الغرب للنشر ، وهران ، الجزائر ، 2004 .
- عبد الرحمن الاحمد العبيد ، **مبادئ التنبؤ الاداري** ، النشر العلمي والمطابع جامعة الملك سعود ، الرياض ، 2004.

- عبد القادر محمد عبد القادر عطية ، **الاقتصاد القياسي بين النظرية والتطبيق**،الطبعة الثانية ، الدار الجامعية ، الاسكندرية ،مصر ،2000.
- عدنان ماجد عبد الرحمن بري ، **طرق التنبؤ الاحصائي** ، الجزء الاول ، جامعة الملك سعود ، 2002.
- علي لزعر ، **الاحصاء وتوفيق المنحنيات** ، ديوان المطبوعات الجامعية ، الجزائر ، 2000.

2- الكتب باللغة الاجنبية

- Allan G.Bluman , **elmentary statistices a step by step approach**, McGraw-hillcompanies, 2004.
- Allen, Bruce T. **Managerial Economics: Theory**, Applications and cases, W Norton & Company, 2001.
- B Coutrot et F Droesbeke, **les méthodes de prévision,** ED PUF , Paris.1984.
- Backam, R. H., **Applied Statistical Time Series Analysis**, Englewood Cliffs,NJ: Prentice Hall, 1988.
- Bendib. R, Econométrie : **théorie et Applications**, OPU, Alger, 2000 .
- Cadoret. I et Benjamin. C et autre, **Econométrie appliquée**: **méthodes, application corrigés,De Boeck**, Bruxelles, 1 ère édition,2004.
- Christian Marmuse ,Les aides à la décision –**techniques quantitatives de gestion-**, 2ème édition FERNAND NATHAN 1983.
- David Anderson, Dennis Sweeney and Thomas William, **Quantitative Methods for Business,** South Western college Publishing , Ohio,2001.