مقدمة في الإحصاء

يتمثل الهدف الرئيس من هذا الفصل في تقديم لمحة سريعة للقارئ عن أهم المفاهيم الأساسية في علم الإحصاء بالإضافة إلى تزويد القارئ بمقدمة أساسية عن الإحصاء التطبيقي، وسيتم التطرق في هذا الفصل إلى كل من التعاريف الخاصة بجوانب ومكونات علم الإحصاء بالإضافة إلى استعراض سريع لأهم طرق حساب مقاييس النزعة المركزية ومقاييس التشتت والتي تمثل أساسيات المدخل إلى الإحصاء التطبيقي

1-1 مفاهيم أساسية في علم الإحصاء

يعتبر علم الإحصاء في الوقت الحالي واحد من أهم العلوم الحديثة التي تلعب دوراً حيوياً في كثير من العلوم والدراسات المختلفة. كما يعتبر الإحصاء من أقدم العلوم حيث ظهر مع حاجة الإنسان الأولى للتعامل مع القيم والأعداد لتسيير الحياة اليومية. فالتاجر يسعى إلى حصر وحفظ البيانات المتعلقة بتجارته والمزارع يقوم دوماً بإحصاء الإنتاج والمعلومات الأخرى المتعلقة كعدد الأشجار وأوقات الحصاد والبذر وغيرها من المعلومات والبيانات ذات العلاقة.

ومع التطور الهائل في العلوم كافة في أواخر القرن العشرين تطور علم الإحصاء ليستفيد من تقنيات الحاسب الآلي بشكل يجعله العلم الأكثر تداخلاً مع العلوم الأخرى المختلفة، حيث أصبح يستخدم علم الإحصاء في العلوم التجارية وعلوم الطب والهندسة والأدب وجميع العلوم الأخرى دون استثناء. كما ساهم عصر المعلومات والانفتاح العالمي الحديث في إبراز أهمية تفعيل عملية التعامل مع البيانات بأسلوب يضمن السيطرة عليها وقراءتها، مما كان له الأثر الواضح على تطور علم الإحصاء كونه العلم الذي يحقق تلك الغاية. كما اتجهت كثير من العلوم والدراسات الأكاديمية والبحثية لاسيما التطبيقية إلى استخدام علم الإحصاء من

خلال حصر بيانات مشكلة البحث والتعامل معها إحصائيا للوصول إلى فهم أفضل وحلول موضوعية.

يتم الاستفادة من علم الإحصاء في مجالات متنوعة تشمل ميادين عديدة كالصناعة والزراعة والطب والبحوث وغيرها من مجالات الإدارة والأعمال والعلم بشكل عام. ويتم تطبيق الأساليب الإحصائية في الجوانب المختلفة للصناعة كمراقبة جودة المنتجات وتسويقها والتخزين وتشغيل خطوط الإنتاج. كما يتم استخدام علم الإحصاء في المجال الطبي لدراسة الأمراض المختلفة والبحث في مسبباتها وطرق علاجها. وفي مجال الزراعة يتم بحث إحصاءات الثروة الحيوانية والنباتية ودراسة العلاقة بين أنواع الأسمدة والأساليب الزراعية المختلفة وزيادة الإنتاج. كما يتم دراسة السكان والمساكن من خلال الإحصاء الديموجرافي، حيث يتم التركيز على القوى العاملة وخصائصها والأجور والدخل والإنفاق. أما في مجال الأعمال والتجارة فان الإحصاء يلعب دوراً حيوياً يتمثل في دراسة السوق واتجاهات المستهلكين ودراسات الأسعار وكميات لإنتاج.

يختلف مفهوم كلمة إحصاء عند العامة، فهي تعني البيانات عند البعض بينما يتم استخدامها عند البعض الآخر للدلالة على عملية جمع البيانات وعملية حفظها. ويتجه البعض إلى فهم الإحصاء بصفته العلم الذي يقوم بجمع ووصف البيانات بهدف إعادة تشكيلها بأسلوب يسهل معه قراءتها ومن ثم يتم تهيئتها للمساندة في اتخاذ القرار أو الحصول على المعلومة ذات العلاقة بالمشكلة محل الدراسة. وحقيقةً يمثل علم الإحصاء الأداة العلمية التي يتم من خلالها جمع البيانات ومن ثم وصفها باستخدام الجداول والرسوم البيانية وذلك بهدف إبراز المعلومة المحتواة في البيانات والتي يصعب قراءتها من خلال البيانات مباشرة. وبالطبع لا يتوقف الأمر عند حد وصف البيانات بل يتجاوزه ليدخل مرحلة مهمة تعتمد على تقنية الحاسب الحديثة حيث يتم تحليل البيانات بطرق علمية متطورة

يمكن من خلالها قراءة المعلومات الموجودة في البيانات بدقة ومصداقية عالية. ويمكن القول بان علم الإحصاء الحديث يتكون من قسمين هما قسم الإحصاء الوصفي (Descriptive Statistics) وقسم الإحصاء الاستدلالي (Inferential Statistic)، حيث يتم في القسم الأول إبراز البيانات الإحصائية من خلال جداول وأشكال بيانية وحساب إحصائيات مختلفة لغرض تسهل قراءة البيانات الخام، بينما يتم في القسم الثاني استخدام أساليب احصائية مختلفة لتحليل البيانات والاستدلال على مجتمع الدراسة من المعلومات المتوفرة لدى الباحث.

الإحصاء (Statistic)

الإحصاء هو العلم الذي يهتم بطرائق جمع البيانات وأساليب وصفها وتحليلها بهدف استخراج المعلومات والحقائق واتخاذ القرارات.

1-1-1 أنواع البيانات الإحصائية

تمثل البيانات الإحصائية مؤشرات حول ناتج كمي أو وصفي لحالة أو تساؤل محدد. وتتقسم البيانات الإحصائية إلى أنواع يمكن وصفها بشكل عام من خلال نوعين أساسيين هما:

النوع الأول: بيانات وصفية والتي تتقسم البيانات الوصفية بدورها إلى قسمين رئيسيين هما بيانات وصفية اسمية وبيانات وصفية ترتيبية. فالبيانات الاسمية تعمل على تصنيف حالات الدراسة إلى مجموعات، في حين أن البيانات الوصفية الترتيبية تتميز بصفة إضافية وهي أن تصنف المجموعات لها ترتيب طبيعي. فالحالة الاجتماعية يمكن تصنيفها إلى أعزب ومتزوج ومطلق و أرمل دون أن يكون هنالك حاجة إلى ترتيبها بشكل محدد، مما يشير إلى أن بيان الحالة

الاجتماعية يمثل هنا بياناً وصفياً اسمياً. وفي المقابل فإن المستوى التعليمي يعكس ترتيباً محدداً حيث يتم ذكر مثلا حقل "غير متعلم" يليه " اقل من ثانوي" ثم "اقل من جامعي" يليه "جامعي فأعلى".

النوع الثاني: البيانات الكمية وهي تلك البيانات الإحصائية التي تقيس صفة لحالات الدراسة، وتأخذ قيمها أرقاماً ضمن فترة قياس هذه الصفة. وتختلف البيانات الكمية تبعاً لاختلاف الصفة، فالبيانات التي تأخذ أرقاماً محددة تمثل بيانات كمية متقطعة مثل عدد أطفال الاسرة وعدد المخالفات المرورية وذلك لأن عدد الأطفال وعدد المخالفات المرورية لابد وان يكون عدداً صحيحاً ولا تقبل أن تكون أعداد كسرية. وتسمى البيانات الكمية التي تأخذ قيماً محصورة في فترة بالبيانات الكمية المتصلة، فعلى سبيل المثال فإن وزن الطفل أو التكاليف المادية للحوادث تمثل بيانات كمية متصلة. ونظريا يمثل عدد القيم التي يمكن أن يأخذها المتغير الكمي المتصل عدداً لا نهائياً بحكم أن القيم التي تكون في أي فترة بغض النظر عن طولها تكون أيضا غير محدودة.

في الواقع يمكن التوصل إلى آلية لتحديد هوية المتغير الكمي، من حيث كونه متقطعا أو متصلا، وذلك من خلال تحديد، أولا، ما إذا كان من الممكن حصر وتحديد جميع القيم التي يأخذها البيان، ومن ثم ، ثانياً، تحديد ما إذا كانت القيمة التي يأخذها المتغير الكمي تمثل فعلا القيمة الحقيقية دون تقريب حسابي من أي نوع.

2-1-1 المجتمع والعينة

تختلف وتتنوع أساليب جمع المعلومات والبيانات وذلك تبعاً لتنوع أهداف الدراسات الإحصائية المطبقة في المجالات المختلفة. وبحكم كون الإحصاء علم يهتم في الأساس بدراسة ظواهر ومعلومات تتعلق بمجتمعات، لذا فان التركيز يكون على المجتمعات شاملة جميع مفرداته. تتحدد المجتمعات تبعاً لصفة وحدة الدراسة الممثلة للمصدر الأساسي للمعلومة المطلوبة. فمثلا، الدراسة التي تهتم بصفات المؤسسات التجارية في مدينة الرياض يتم تحديد المجتمع لها من خلال حصر جميع المؤسسات التجارية في مدينة الرياض دون استثناء، والدراسة التي تهتم بحوادث السيارات في منطقة جغرافية محددة يتم تحديد مجتمعها بجميع الحوادث التي وقعت في المنطقة.

تختلف المجتمعات بصفاتها من عدة محاور. فهنالك المجتمعات المحدودة والمجتمعات غير المحدودة، حيث يعتبر المجتمع محدودا إذا كان بالإمكان حصر جميع وحدات الدراسة فيه. فمجتمع طلاب المدارس الثانوية في المملكة العربية السعودية يعتبر مجتمع محدود، في حين أن مجتمع نوع معين من الأسماك يمثل مجتمع غير محدود لا يمكن بحال من الأحول حصر جميع وحداته. كذلك يتم تصنيف المجتمعات تبعا للحجم، فهناك مجتمعات كبيرة ومجتمعات صغيرة. فالمجتمعات الكبيرة تحتوى عدد كبير من الوحدات التي تتطلب جهدا بدنيا هائلا لحصرها كما تتطلب وقتاً طويلاً لدراستها بالإضافة إلى ارتفاع تكاليف حصرها ودراستها. في حين أن المجتمعات الصغيرة تضم عدداً قليلاً من الوحدات مقارنة بالمجتمعات الكبيرة، ومن ثم يرافقها تكاليف ووقت وجهد اقل نسبيا.

عندما تكون المجتمعات غير محدودة فان أسلوب دراسة جميع وحدات المجتمع والذي يطلق عليه أسلوب الحصر الشامل يصبح مستحيلا. كذلك الحال في بعض المجتمعات المحدودة والتي لا يقبل المنطق تطبيق أسلوب الحصر الشامل، مثل فحص دم شخص، حيث لا يمكن سحب جميع دمه، أو اختبار قوة تحمل مصابيح كهربائية لجهد كهربائي، حيث لا يمكن إحراق جميع المصابيح الكهربائية المتوفرة لمعرفة الجهد الأعلى المطلوب لاحتراقها. لذا فان الأسلوب الأمثل هنا يكمن في تبني أسلوب المعاينة حيث يتم الاعتماد على عينة يتم سحبها من المجتمع المستهدف بشرط أن تكون ممثلة له. وفي حالة إجراء دراسة إحصائية لا ترقى إلى مستوى تحمل تكاليف تغطية المجتمع، أو في حالة عدم توفر الجهد الكافي أو الإنفاق المادي اللازم لتغطية المجتمع فان الحل يكون في إجراء دراسة على عينة من المجتمع تتناسب والإمكانيات المتاحة.

تختلف طرق سحب العينات من المجتمعات، بيد أنها تتفق جميعا في كونها جزء من المجتمع وممثلاً له. وعند الاعتماد على عينة عشوائية ممثلة لمجتمع الدراسة يتم الحصول على قيم إحصائية تسمى بالإحصائيات (Statistics) والتي تعتبر قيم عشوائية متغيرة تعتمد على القيم الواردة في العينة العشوائية. وفي المقابل فإنه يطلق على القيم والمؤشرات الإحصائية التي يتم الحصول عليها من خلال تغطية مجتمع الدراسة بالمعالم (Parameters). وتمتاز قيم معالم المجتمعات بأنها قيم ثابتة لا تتغير تعبا لمن يقوم بحسابها. فمتوسط معدل الطلاب التراكمي لطلاب جامعة الملك سعود (مجتمع الدراسة) في وقت محدد يعتبر قيمة ثابتة يحصل عليها كل من قام بحسابها. في حين أن العينات العشوائية المسحوبة من مجتمع محدد لا تتوافق قيمها في الغالب مما ينتج عنه قيم إحصائية مختلفة تمثل جميعها تقديرات مقبولة لمعلمة المجتمع المجهولة.

2-1 تبويب البيانات

تعتبر عملية التعامل مع البيانات من المراحل المهمة جدا والحساسة في الدراسات الإحصائية. وتتبع أهميتها من كونها الأداة التي يتم فيها إنشاء القاعدة التي يبنى عليها المراحل اللاحقة في التحليل الإحصائي. كما تتبع الأهمية الكبيرة لمرحلة التعامل مع البيانات في كون معظم الدراسات الإحصائية التطبيقية تتعامل مع كم كبير من البيانات والتي بدورها تتطلب معالجة إحصائية ليتم تحويلها إلى شكل يتم من خلاله تطبيق التحليل الإحصائي بفعالية وسهولة.

فعند توفر عدد كبير من البيانات يتطلب الأمر في كثير من الأحيان وضع القيم في جدول تكراري يلخص البيانات الإحصائية المدروسة بشكل يمكن من خلاله التعامل مع البيانات بقدرة وكفاءة أعلى. وذلك بدوره يتيح للباحث القدرة على التعمق في فهم البيانات الإحصائية بالإضافة إلى إمكانية إجراء تحليل إحصائي استدلالي. وفي الواقع يمكن تبويب كل من البيانات الكمية والبيانات الوصفية على حد سواء. كذلك يمكن تبويب البيانات الكمية المتقطعة والبيانات الكمية المتصلة، وفي كل الحالات تتم عملية التبويب من خلال إنشاء جدول تكراري.

يتم تبويب البيانات من خلال تفريغها في جداول تكرارية تتكون من عامودين أساسيين. يمثل العامود الأول قيم المتغير بينما يمثل العامود الثاني التكرارات المرافقة لكل قيمة من قيم المتغير. كما يمكن تضمين أعمدة إضافية تحتوى معلومات تفصيلية عند الحاجة، مثل بيان التوزيع النسبي أو التوزيع التراكمي. يطلق على الجدول التكراري المكون من العامودين الأساسين فقط بالجدول التكراري البسيط. وفي حال كون المتغير اسمي أو متقطع بحيث تكون القيم التي يأخذها المتغير محدودة نوعا ما، فان العملية تتم من خلال حصر جميع

القيم المختلفة التي يأخذها المتغير ومن ثم اعتمادها كمسميات لحقول الجدول التكراري (العامود الأول). وبما أن الهدف الأساسي من تبويب البيانات هو تلخيص وإيضاح أهم معالم المتغير، فان عدد حقول الجدول التكراري يجب أن تكون محدودة ليكون ذا فائدة إحصائيا، وإذا كان عدد القيم كبيراً يصبح الجدول ذا حقول كثيرة ومن ثم يصعب الحصول على معلومة منه.

وفي حال التعامل مع متغير كمي متقطع يأخذ قيم محدودة يتم سرد القيم التي يأخذها المتغير في العامود الأول والتكرارات المصاحبة لتلك القيم في العامود الثاني. وبالطبع يتم استخدام القيم الكمية المختلفة التي يأخذها المتغير الكمي المتقطع لتسمية حقول الجدول التكراري المستهدف.

أما في حال التعامل مع متغير عشوائي كمي متقطع يأخذ قيم كثيرة بحيث يتعذر إنشاء جدول تكراري عملي بحقول محدودة، فان الأسلوب الأمثل هنا يكمن في إنشاء جدول تكراري بحيث يكون كل حقل مخصص لمجموعة أو فترة من القيم. في هذه الحالة يقوم الباحث بالتحديد مسبقا لعدد الحقول المستهدفة في الجدول المطلوب، ومن ثم يتم اعتماد مجموع عدد مرات ظهور القيم المحددة للحقل كتكرار للحقل نفسه.

3-1 عرض البيانات

هنالك مثل يقول "رب صورة عادلت ألف كلمة" وتلك مقولة يمكن الجزم بأنها صحيحة تمامًا في المجال الإحصائي على الأقل. فالمعلومة المتوفرة بمتغير كمي يتكون من مئات أو آلاف القيم لا يمكن بحال من الأحوال التعرف عليها أو استكشافها من خلال النظر فقط إلى القيم. كذلك قد لا يتسع الوقت لمتلقي المعلومة، خاصة عندما يكون من الإدارة العليا، للتمعن في القيم الغير مبوبة للمتغير الخام والتي تتطلب قدر من المهارة العلمية والخبرة ليتم قراءتها علميا مباشرة. لذا دعت الحاجة إلى تبني وسيلة لعرض المؤشرات والحقائق حول المتغيرات الكمية والاسمية. ويتم العرض من خلال أسلوبيين رئيسيين هما العرض الجدولي والعرض البياني. وحيث انه قد تم التطرق إلى موضوع العرض الجدولي في الجزئية السابقة فانه سيتم التطرق في هذه الجزئية إلى مسائلة العرض البياني للمتغيرات الكمية والاسمية على حد سواء.

تختلف أساليب العرض الجدولي والبياني للمتغيرات الإحصائية وتتنوع وذلك تبعا للهدف منها. ويمثل أهم هدف يرجى من الجداول والرسوم البيانية في إيصال المعلومة الإحصائية بأسرع وابسط طريقة ممكنة إلى متخذ القرار أو القارئ والذي قد لا يكون ملماً بالعمليات الرياضية الإحصائية. لذا فان جودة عرض البيانات تتشكل من خلال عدة عوامل أهمها العامل الإحصائي والعامل الفني. يمثل العامل الفني القدرة الشخصية لاختيار الأحجام والألوان وأسلوب العرض والخطوط وغيرها من التفاصيل الفنية بينما يمثل العامل الإحصائي المهارة في عملية إيجاد التكرارات أو الأرقام والمؤشرات الإحصائية المطلوبة. ويمكن أن يتم ذلك من خلال التعامل مع التكرارات مباشرة أو التعامل مع نسب التكرارات في

الحقـول المختلفـة إلـى إجمـالي التكـرارات. كـذلك يمكـن التعبيـر عـن التكـرارات فـي الجـداول التكراريـة بالصـيغ التجميعيـة أو الصـيغ النسـبية العاديـة أو التجميعيـة. وفيمـا يتعلـق بـالعرض البيـاني، فانـه يمكـن تبنـي أسـلوب يتميـز بالبسـاطة والوضـوح، كمـا انـه يمكن التعامل مع التكرارات أو النسب مباشرة أو التعامل مع الصيغ التجميعية.

يتكـون العـرض البيـاني مـن عـدة أجـزاء تتكامـل لتخـرج جملـة بيانيـة مفهومـة غيـر منقوصـة المعنـى. ويمثـل أهـم تلـك الأجـزاء عنـوان الرسـم البيـاني والـذي يكـون فـي العـادة نقطـة الانطـلاق للعـين المتفحصـة للعـرض أو الرسـم البيـاني. يلـي ذلـك فـي الأهميـة عنـاوين المحـاور وقيمهـا، وتلـك معلومـة ضـرورية لا يمكـن لمـتفحص الرسـم البيـاني معرفـة أسـاس القـيم المرسـومة ووحـدة قياسـها مـا لـم تكـن تلـك المحـاور معرفـة تمامـا. تمثـل مفـاتيح الرسـم الجـزء الثالـث المهـم فـي عمليـة إيصـال المعلومـة الإحصـائية بيانيـا. ويهـتم مفتـاح الرسـم بوصـف مكونـات الرسـم الداخليـة وعلاقتهـا بـالمتغيرات المختلفـة. ويمكـن اسـتخدام الألـوان أو الأشـكال للتعبيـر عـن مفـاتيح الرسـم، بيـد أن قـوة التعبيـر هنـا تعتمـد بشـكل أساسـي علـى الوضـوح والبسـاطة دون الحاجـة إلـى المبالغـة باختيـار الألـوان أو الأشـكال. يـأتي فـي المرتبـة الرابعـة ألـوان الخلفيـة وحجـم الرسـم والخطـوط ونـوع الرسـم ومـدى مناسـبته للمعلومـة المبـرزة وموضـع مفـاتيح الرسـم والإيضـاحات والعنـوان وغيرهـا مـن التفاصـيل الشـكلية ولكـن المهمـة فـي وضـوح المعنى المستهدف في الرسم البياني.

تختلـف أنـواع الرسـوم البيانيـة بـاختلاف أهـدافها. ومـن الأنـواع الأكثـر اسـتخداما الأعمـدة البيانيـة والخـط البيـاني والـدوائر. وفـي الواقـع لا يمكـن بـأي حـال مـن الأحـوال حصـر جميـع الأشـكال التـي يمكـن أن يـتم إيضـاح المعلومـة الإحصـائية مـن خلالهـا خاصـة مـع التطـور التقنـي الحـديث ومـا صـاحبه مـن تطـور فـي البـرامج الإحصـائية مثـل برنـامج SAS و برنـامج SPSS وغيرهـا والتـي أعطـت اهتمـام كبيـر للرسـوم البيانيـة والإشـكال الفنيـة لهـا. لـذا فانـه سـيتم التطـرق إلـى بعـض الأمثلـة بهـدف

إيضاح جانب من الإمكانات المتاحة دون الادعاء بان تلك الأمثلة ستكون شاملة لجميع جوانب موضوع العرض البياني.

يعتمد اختيار نوع الرسم البياني المناسب على عدة عوامل من أهمها طبيعة القيم الكمية المراد رسمها. فرسم المتغير الكمي المتقطع يختلف عن رسم المتغير الكمي المتصل، كما أن المتغير المرتبط بالزمن كالإنتاج الشهري أو الأرباح السنوية يمكن رسمها بأسلوب مختلف عن القيم الغير مرتبطة بالزمن كالحقول غير الترتيبية مثل الحالة الاجتماعية والمنطقة السكنية والحالة التعليمية. وفي حال التعامل مع أكثر من متغير فانه يمكن رسم المتغيرات في بيان واحد. في هذه الحالة يأتي دور مفتاح الرسم ليدل على موقع كل متغير في الرسم البياني. كما يمكن اللجوء في بعض الحالات إلى رسم النسب بدل من القيم الفعلية خاصة عند التعامل مع أكثر من متغير، حيث يمكن لهذا الأسلوب إبراز المعلومة بشكل أوضح خاصة عندما يكون هنالك تباين كبير بين قيم المتغير المستهدف.

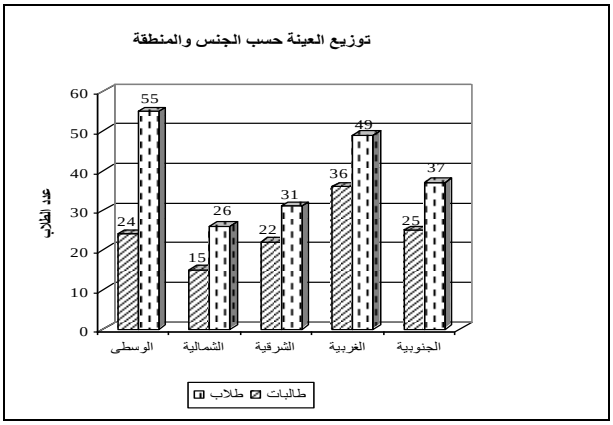
الجدول التالي يوضح التوزيع التكراري لعينة عشوائية حجمها 320 طالب في التعليم العالي في المملكة العربية السعودية، حيث تم تصنيفهم حسب الجنس (طالب، طالبة) وحسب المنطقة (وسطى، شمالية، شرقية، غربية، جنوبية). المطلوب إبراز البيانات في الجدول بيانيا بعدة طرق مختلفة.

توزيع 320 طالب تعليم عالي حسب الجنس والمنطقة

المنطقة	طلاب	طالبات
الوسطى	55	24
الشمالية	26	15
الشرقية	31	22
الغربية	49	36
الجنوبية	37	25

في البداية يمكن استخدام الأعمدة البيانية المزدوجة لتمثيل أعداد الطلاب والطالبات في العينة العشوائية. الشكل التالي يبين إحدى الإمكانات التي يمكن فيها إبراز المعلومات في الجدول بيانيا، حيث تم تمثيل كل منطقة بعاموديين: أحداهما للطلاب والآخر للطالبات. تجدر الإشارة إلى إن الرسم البياني قد احتوى على جميع العناصر المطلوبة وهي عنوان الرسم وعناوين المحاور ومفتاح الرسم بالإضافة إلى القيم المرافقة للأعمدة.

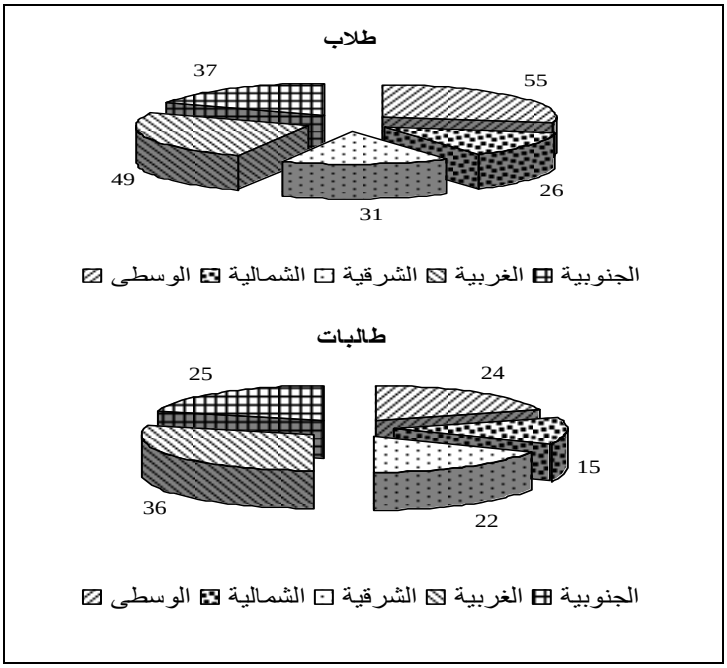
الأعمدة البيانية المزدوجة



وكطريقة أخرى في عرض المعلومة المبينة في الجدول يمكن استخدام التوزيع النسبي للتكرارات. فالرسم البياني التالي بوضح توزيع العينة حسب المنطقة موزعة نسبياً حسب الجنس، مع ملاحظة أن كل منطقة تم تمثيلها بعامود واحد يبين مجموع الطلاب ومجموع الطالبات.

ويمكن دائما استخدام رسوم بيانية أكثر جمالا وجاذبية، إلا أنه يجب أن لا ننسى الهدف الأساسي من الرسم البياني وهو إيصال المعلومة بشكل واضح وتجنب تعقيدات قد تلغي الفائدة المرجوة من الرسوم البيانية

رسوم دائرية لتوزيع الطلاب والطالبات حسب المنطقة



4-1 مقاييس إحصائية

تتكون العملية الإحصائية من عدة خطوات يمثل فيها الوصف البياني الخطوة الأولى بينما تأتى عملية وصف البيانات كمياً كخطوة ثانية مهمة ومكملة للوصول إلى فهم أعمق ورؤية أوضح للمعلومة الموفرة في القيم الكمية محل الدراسة. وتتلخص عملية وصف البيانات كمياً في محاولة الحصول على قيم تشير بشيء من التفصيل إلى توجهات المتغيرات الكمية. وفي الواقع تتنوع المقاييس الكمية الإحصائية التي يمكن تطبيقها على المتغيرات الكمية بيد انه يمكن تقسيمها إلى نوعين رئيسيين هما مقاييس نزعة مركزية ومقاييس تشتت.

تسعى الدراسات الإحصائية إلى كشف معالم المجتمعات المرتبطة بالظاهرة المدروسة. ويتطلب ذلك بالطبع حصر جميع البيانات الكمية المتصلة بمشكلة الدراسة والمتوفرة في المجتمع. وفي حال حصر جميع قيم المجتمع فانه يتم الحصول على المعالم المطلوبة والتي تكون بالطبع قيم ثابتة. في المقابل عندما يكون المجتمع غير محدود أو كبير جدا لدرجة تعذر عملية حصر جميع القيم الموجودة فيه فان الأسلوب الإحصائي الأمثل هنا يكمن في استخدام عينة عشوائية ممثلة للمجتمع ومن ثم الحصول على تقديرات للمعالم المطلوبة. وبحكم كون العينة عشوائية فان جميع المؤشرات الإحصائية التي يتم الحصول عليها من خلال العينة العشوائية والتي يطلق عليها بإحصائيات العينة تعتبر متغيرات عشوائية تختلف قيمها تبعا لاختلاف العينة العشوائية المستخدمة. لذا فان عملية حساب المؤشرات الإحصائية سواء المتعلقة بالنزعة المركزية أو بالتشتت ستختلف تبعا لنوع وعاء البيانات الإحصائية المستخدم، مجتمع أم عينة عشوائية.

يمكن حساب مقاييس النزعة المركزية ومقاييس التشتت للبيانات الكمية عموما، بيد أن الطريقة المستخدمة في عملية تقدير قيم تلك المقاييس تختلف كذلك باختلاف أنواع البيانات المدروسة. فيتم حساب مقاييس النزعة المركزية ومقاييس التشتت لبيانات خام بأسلوب مختلف عن البيانات المبوبة (الجداول التكرارية). في البداية سيتم التعريف بآلية حساب وتقدير المقاييس الإحصائية لبيانات كمية خام، غير مبوبة، يلي ذلك التطرق إلى عملية التقدير في حال توفر الصيغة المبوبة فقط للمتغيرات الكمية المدروسة.

تمثل البيانات الخام المصدر الأدق للمعلومة، حيث يجب الاعتماد عليها دوما إذا كانت متوفرة. ولكن في كثير من الأحيان لا تتوفر البيانات الخام وإنما يتوفر جداول تكرارية لمتغيرات عشوائية تم الحصول عليها من مصادر تاريخية مثل التقارير الحكومية والدراسات السابقة والنشرات الإحصائية. لذا فانه في هذه الحالة يتم اللجوء إلى البيانات المبوبة لتقدير قيم المقاييس الإحصائية المطلوبة مع الانتباه إلى أن القيمة المقدرة لا تمثل القيم الحقيقة وإنما قيمة قريبة ومقبولة إحصائيا.

تعتبر المقاييس الكمية على اختلاف أنواعها مؤشرات مهمة جدا حيث تلعب دورا حيويا في معظم مراحل العملية الإحصائية الاستدلالية. وتتنوع تلك المقاييس مع توحد الهدف من وجودها، لتعطي الباحث إمكانية التعامل مع جميع أنواع البيانات الكمية ومن ثم الوصول إلى وصف كمي دقيق لها. سيتم التطرق إلى كل من أهم مقاييس النزعة المركزية واهم مقاييس التشتت، مع الإشارة إلى أن الهدف من الباب الحالي هو إعطاء القارئ مقدمة أساسية ومعلومة سريعة أساسية ومطلوبة لفهم الأبواب المطروحة في الكتاب ككل.

1-4-1 مقاييس النزعة المركزية

تهتم مقاييس النزعة المركزية بتوفير مؤشرات كمية تمثل التوجه العام لقيم المتغير الكمي المدروس، حيث يتم الحصول على مؤشر يفيد عن توجه القيم، دون الحاجة إلى التعامل مع جميع القيم المختلفة للمتغير الكمي. لذا فان مقاييس النزعة المركزية تنتج في النهاية أرقام محدودة تمثل التقديرات لتلك المقاييس وذلك بغض النظر عن عدد القيم الأصلي، سواء كان صغيراً أم كبيراً. كما أن الاستعاضة عن القيم كلها برقم واحد يفقدنا كثير من المعلومات حول البيانات، إلا أنه لا يمكن حجب الأهمية الكبيرة والدور الهام لمقاييس النزعة المركزية في مجال الاستدلال الإحصائي، حيث تعتبر هذه المقاييس حجر الأساس ونقطة البداية لأي دراسة تحليلية إحصائية، كما يتم الانطلاق منها إلى مستويات متقدمة في عمليات التحليل ومن ثم الاستدلال حول توجه البيانات والصفات المميزة للمتغيرات الكمية التابعة لها. تتنوع استخدامات مقاييس النزعة المركزية في عمليات الاستدلال الإحصائي مما ينتج عنه تنوع في طبيعة تلك المقاييس المختلفة، لذا سيتم التطرق في هذا الفصل إلى ثلاث أنواع من مقاييس النزعة المركزية والتي تمثل الأهم، وهي كل من الوسط الحسابي والوسيط والمنوال.

المنوال:

يمثل المنوال (Mode) القيمة الأكثر شيوعا من بين القيم المختلفة للمتغير العشوائي محل الدراسة. ويتم تحديد قيمة المنوال من خلال تحديد تكرار جميع القيم المختلفة للمتغير العشوائي محل الدراسة إذا كانت البيانات غير مبوبة (بيانات خام)، بينما يتم الاستعانة بقاعدة رياضية إذا كانت قيم المتغير العشوائي متوفرة في جدول تكراري (بيانات مبوبة).

بالنسبة للبيانات الخام، يتم تحديد قيمة وحيدة للمنوال إذا وجدت قيمة واحدة تكررت أكثر من باقي القيم المختلفة للمتغير العشوائي. كذلك يمكن أن يكون المنوال متمثل بأكثر من قيمة إذا كان هنالك أكثر من قيمة واحدة لها نفس التكرار الأكثر من بين جميع التكرارات المتوفرة. وفي حال عدم تكرر أي قيمة من قيم المتغير العشوائي المختلفة فانه في هذه الحالة لا يكون هنالك منوال بين قيم المتغير العشوائي.

المنوال: بيانات خام (غير مبوبة)

المنوال لبيانات خام هو القيمة أو القيم الأكثر شيوعا أو تكرارا.

الوسيط:

يعتبر الوسيط (Median) مقياس آخر للنزعة المركزية، حيث يتم من خلال الوسيط الوصول إلى رقم كمي يمثل القيمة التي تقع في منتصف قيم المتغير الكمي المدروس. لذا فإن الوسيط يمثل القيمة الكمية التي تكون نصف قراءات المتغير الكمي اقل منها بينما النصف الآخر أعلى منها. ولحساب الوسيط لا بد أولا من أن يتم ترتيب القيم تصاعديا، حيث يتم ذلك من خلال الترتيب التصاعدي (أو الهابط) العادي، ثم تحديد ترتيب الوسيط، وأخيرا مرحلة تحديد قيمة الوسيط.

الوسيط:

الوسيط، Q_2، هو القيمة التي تقع في منتصف القيم، بحيث تكون نصف القيم اقل منها والنصف الآخر أعلى منها.

الوسط الحسابي:

يعتبر الوسط الحسابي أو المتوسط (Arithmetic Mean) المقياس الأكثر شهرة من مقاييس النزعة المركزية وكذلك الأكثر أهمية. وتمثل قيمة الوسط الحسابي القيمة التي تتمركز حولها جميع القيم المختلفة للمتغير الكمي. يمكن الحصول على القيمة الحقيقية لمتوسط متغير عشوائي في مجتمع محدود إذا تم التعامل مع كافة القيم في المجتمع. في هذه الحالة يرمز لقيمة الوسط الحسابي المحصل بالرمز μ والتي تمثل معلمة المجتمع.

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$$

أما في حال التعامل مع عينة عشوائية ممثلة لمجتمع الدراسة حجمها n قراءة، فان معادلة تقدير قيمة الوسط الحسابي يتم من خلال الدالة،

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

حيث تم الإشارة رياضيا لقيمة الوسط الحسابي باستخدام الشرطة على رمز المتغير الكمي. وبالطبع تعتبر القيمة $\bar{X}$ إحصائية عينة تمثل تقدير مقبول إحصائيا لقيمة معلمة المجتمع المجهولة μ.

يضم الجدول عدد الموظفين لجميع فروع شركة كبيرة والبالغ عددها 22 فرع منتشرة في جميع أنحاء المملكة العربية السعودية. أوجد كل من المنوال والوسيط والوسط الحسابي لعدد الموظفين في فروع الشركة.

عدد الموظفين في فروع الشركة

12	8	7	8	15	9	6	15	12	11	8
3	11	8	17	12	11	8	14	21	7	9

لإيجاد قيمة المنوال المطلوبة يتم أولا حساب عدد مرات ظهور القيم المختلفة للمتغير العشوائي الممثل لعدد الموظفين في فروع الشركة، ويبين الجدول التالي التوزيع التكراري لفروع الشركة حسب عدد الموظفين.

جدول 1-4-1-3: توزيع فروع الشركة حسب عدد الموظفين

عدد الموظفين	21	17	15	14	12	11	9	8	7	6	3
عدد الفروع	1	1	2	1	3	3	2	5	2	1	1

وبما أن المنوال هو القيمة الأكثر تكرارا لذا فان عدد الموظفين 8 هو العدد الأكثر تكرارا في فروع الشركة (تكرر 5 مرات)، وبذلك فإن المنوال يساوي 8

لحساب الوسيط يتم أولا ترتيب القيم تصاعديا، ثم تحديد رتبة الوسيط، ومن ثم حساب قيمة الوسيط. وبذلك فإن الوسيط يساوي 10 وذلك لأن نصف فروع الشركة لديها موظفين اقل من 10 بينما النصف الأخر لديه موظفين أكثر.

أما بالنسبة إلى حساب قيمة الوسط الحسابي لعدد الموظفين في فروع الشركة، فإنه تجدر الإشارة هنا إلى أنه إذا كانت القيم المحصلة تمثل مجتمع الدراسة، أي جميع فروع الشركة، لذا فان المطلوب هنا الحصول على قيمة معلمة المجتمع μ_X الممثلة لمتوسط عدد الموظفين في الفرع الواحد للشركة.

$$\mu_X = \frac{\sum_{i=1}^{22} X_i}{22} = \frac{232}{22} = 10.5$$

2-4-1 مقاييس التشتت

تمثل مقاييس التشتت الجانب الآخر من المقاييس الإحصائية الأساسية بجانب مقاييس النزعة المركزية، حيث تستخدم تلك المقاييس في وصف البيانات والتعرف على خصائصها. كما تعمل مقاييس التشتت كجزئية مكملة ومهمة جدا بجانب مقاييس النزعة المركزية في عمليات الاستدلال الإحصائي المبنية على عملية التعامل مع البيانات. وينصب الاهتمام عند التعامل مع مقاييس التشتت حول قياس درجة الاختلاف بين القيم المختلفة للمتغير الكمي المدروس، ويتم ذلك من خلال عدة مقاييس مختلفة يهتم كل واحد منها بقياس درجة الاختلاف من زاوية مختلفة. يمثل التباين والانحراف المعياري والانحراف المتوسط والانحراف الربيعي بالإضافة إلى المدى مقاييس مختلفة لقياس تشتت المتغيرات الكمية.

يتم الحصول على تصور دقيق عن خصائص المتغير الكمي في حال توفر كل من مقياس النزعة المركزية ومقياس التشتت، حيث تعطي مقاييس النزعة المركزية تصور عن تمركز القيم بينما تعطي مقاييس التشتت تصور عن درجة اختلاف تلك القيم عن بعضها البعض. لذا يمكن القول بأن الاعتماد على مقياس واحد قد لا يغني عن الآخر في عملية الاستدلال الإحصائي، حيث ينتج عنه دوما قصور في المعلومة المعتمد عليها ومن ثم عدم القدرة على قراءة البيانات إحصائيا بشكل سليم.

يستخدم المدى الممثل للفرق بين اكبر قيمة واصغر قيمة كمقياس بسيط وسطحي عن درجة تشتت قيم المتغير الكمي. ولكن لا يجب الأخذ بهذا المقياس والاعتماد عليه في العمليات الاستدلالية الإحصائية حيث انه يتأثر بشدة بالقيم المتطرفة بالإضافة إلى عدم استخدامه لباقي قيم المتغير الكمي. وكمقياس أدق يعمل في حال وجود قيم متطرفة، يمكن استخدام الانحراف الربيعي والذي يعتمد على ترتيب القيم كما هو معمول به في عملية حساب الوسيط. في المقابل عندما لا تكون مشكلة القيم المتطرفة حاضرة فانه يمكن استخدام الانحراف المتوسط أو

الانحراف المعياري كمقياس للتشتت، حيث يتم استخدام كافة القيم في عملية حساب المقاييس السابقة.

يتم في الواقع حساب مقاييس التشتت للبيانات الكمية بصيغتيها الخام والمبوبة، ولكن الطريقة المتبعة في عمليات الحساب تختلف باختلاف طبيعة البيانات المدروسة. وبالطبع، كما تم الإشارة إليه سابقا، تمثل التقديرات المحصلة من البيانات الخام معلومة أدق وأكثر صحة من المعلومة المحصلة من البيانات المبوبة، لذا فانه يجب الاعتماد على البيانات الخام في حال توفرها في عملية حساب مقاييس التشتت، كما يجب قصر الاعتماد على البيانات المبوبة ليتم فقط في حالة عدم توفر الصيغة الخام للبيانات حيث أنها تعطي قيم تقريبية لا ترقى إلى دقة التقديرات المحصلة من خلال استخدام البيانات الغير مبوبة.

عند التعامل مع بيانات خام فانه يمكن أن يتم تغطية مجتمع الدراسة إذا كان المجتمع محدود الحجم وكانت الدراسة موجهة للتعامل مع المجتمع كاملا مع ما يصاحب تلك العملية من جهد ووقت وتكاليف مادية عالية. وإذا كان مجتمع الدراسة غير محدود أو كانت الدراسة لا تستطيع تحمل الوقت أو الجهد أو التكلفة العالية المطلوبة لتغطية مجتمع الدراسة، فانه يمكن الاعتماد على عينة عشوائية مسحوبة من المجتمع وممثلة له للحصول على تقديرات لمقاييس التشتت المطلوبة. تختلف طريقة حساب التباين والانحراف المعياري جوهريا عند التعامل مع مجتمع كامل عنها عند التعامل مع عينة عشوائية. لذلك فسيتم إيضاح الفرق في طريقة التقدير عند التطرق إلى كل من الانحراف المعياري والتباين لاحقا. تجدر الإشارة إلى أن جميع مقاييس التشتت هي قيم موجبة، وذلك شرط أساسي يجب توفره في جميع مقاييس التشتت. لذا فان مقاييس التشتت لا تأخذ قيم سالبة أبدا بل تكون قيمها موجبة دوما أو مساوية للصفر فقط وذلك إذا كانت جميع قيم المتغير الكمي محل الدراسة متساوية، أي انه لا يوجد تباين أو تشتت أصلا.

الانحراف المعياري:

يعتبر الانحراف المعياري والتباين من أهم مقاييس التشتت الإحصائية. ويرتبط المقياسين بعلاقة رياضية قوية، حيث يمكن دوما الحصول على المقياس الآخر في حال معرفة قيمة احدهما. يرمز للتباين بالرمز σ^2 في حال الحصول على قيمته من خلال تغطية مجتمع الدراسة، بينما يتم استخدام الرمز S^2 للدلالة على مقدر التباين المحصل من خلال بيانات عينة عشوائية مسحوبة من مجتمع الدراسة. وبأخذ الجذر التربيعي للتباين يتم الحصول على قيمة الانحراف المعياري وذلك في الحالتين، حالة المجتمع وحالة العينة،

$$\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad S = \sqrt{S^2}$$

وبحكم العلاقة الرياضية القوية بين كل من التباين والانحراف المعياري فإنه يمكن اعتبارهما وجهين لعملة واحدة لهما نفس الأهمية.

يعتمد الانحراف المعياري والتباين على فكرة تربيع الفروق بين قيم المتغير الكمي X ووسطها الحسابي، وفي حال التعامل مع مجتمع بحجم N يتم الحصول على الانحراف المعياري من خلال الدالة التالية،

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \mu_X\right)^2}{N}}$$

أما في حالة التعامل مع عينة عشوائية بحجم n مسحوبة من مجتمع الدراسة وممثلة له، فان صيغة حساب الانحراف المعياري هي،

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}}$$

وبالتبسيط الرياضي تصبح،

$$S_X = \sqrt{\frac{\sum_{i=1}^{n} X_i^{\,2} - n\bar{X}^{\,2}}{n-1}}$$

بالرجوع إلى بيانات المثال السابق فإن يمكن إيجاد الانحراف المعياري والتباين لعدد الموظفين في فروع الشركة.

$$\sum_{i=1}^{22}(X_i - \mu_X)^2 = (8-10.545)^2 + (11-10.545)^2 + ...$$

$$....+(11-10.545)^2 + (3-10.545)^2 = 349.4545$$

وعليه فان قيمة كل من الانحراف المعياري والتباين لمتغير عدد الموظفين في فروع الشركة تصبح،

$$\sigma_X = \sqrt{\frac{\sum(X-\mu_X)^2}{N}} = \sqrt{\frac{349.4545}{22}} = \sqrt{15.88429545} = 3.98551$$

$$\sigma_X^2 = \frac{\sum(X_i-\mu_X)^2}{N} = \frac{349.4545}{22} = 15.88429545$$

تختلف مقاييس النزعة المركزية من ناحية التفسير عن مقاييس التشتت. ويمكن الاستدلال مباشرة عن القيمة التي تتمركز حولها جميع القيم المشمولة في دراسة ما من خلال حساب إحدى مقاييس النزعة المركزية، بينما لا يمكن تفسير القيمة الوحيدة المحصلة من خلال حساب إحدى مقاييس التشتت. يتم استخدام مقاييس التشتت في الأصل في عمليات المقارنة بين مجموعتين من البيانات، ففي حال توفر مجموعة أخرى من البيانات يمكن حساب مقاييس التشتت للمجموعتين ومن ثم الحكم على المجموعة التي لها مقياس تشتت اكبر في القيمة بأنها المجموعة الأكثر تشتتا. لتقريب الصورة، افترض أننا نتعامل مع مجموعتين من القيم لهما نفس وحدة القياس (درجة مثلا) في كل مجموعة ثلاثة قيم. تضم المجموعة الأولى (مجموعة A) القيم 12 و 48 و 90 بينما تضم المجموعة الثانية

(المجموعة B) القيم 45 و 52 و 53، عند إيجاد الوسط الحسابي للمجموعتين يتبين أن لهما نفس المتوسط،

$$\mu_A = \mu_B = 50$$

مما يشير للوهلة الأولى بان المجموعتين لهما نفس الصفات الإحصائية، بينما عند حساب إحدى مقاييس التشتت، التباين مثلا،

$$\sigma_A^2 = 1016 \quad \& \quad \sigma_B^2 = 12.7$$

يتبين الفرق الكبير جدا بين المجموعتين. فالمجموعة A لها تشتت اكبر بكثير من المجموعة B، مما يشير إلى أن قيم المجموعة A تتباعد عن بعضها بشكل كبير مقارنة بتباعد قيم المجموعة B. لذلك، فان مقاييس النزعة المركزية ومقاييس التشتت تمثلان عند اجتماعهما أداة قوية تعطي تصور واضح عن صفات القيم إحصائيا.

وكمثال آخر فإن عند دراسة مستوى الأداء الأكاديمي لطلاب وطالبات الكليات الطبية في الجامعات السعودية تم اخذ عينتين عشوائيتين من مجتمع الطلاب (X) ومجتمع الطالبات (Y)، بواقع عينة لكل مجتمع. وبحساب متوسط وتباين المعدل التراكمي (يتم حساب المعدل التراكمي من 5 نقاط) الطلاب تبين

$$\bar{X} = 3.8 \quad \& \quad S_X^2 = 5.6$$

في حين تم الحصول على نتائج تتعلق بالمعدلات التراكمية للطالبات وهي

$$\bar{Y} = 2.9 \quad \& \quad S_Y^2 = 3.2$$

وبمقارنة أداء الطلاب والطالبات يتبين تفوق الطلاب حيث حققت عينة الطلاب متوسط معدل تراكمي 3.8 بينما كان متوسط المعدل التراكمي للطالبات 2.9، وذلك يشير إلى أن أداء الطلاب في المتوسط أفضل من أداء الطالبات. ولكن بالنظر إلى تباين معدلات الطلاب $S_X^2 = 5.6$ وتباين معدلات الطالبات $S_Y^2 = 3.2$ يتبين أن تشتت معدلات الطلاب

اكبر من تشتت معدلات الطالبات، وهذا يشير إلى أن المشكلة الأكاديمية هي اكبر في الطلاب منها في الطالبات. فالاختلاف بين مستويات الطلاب يقارب لضعف الاختلاف بين مستويات الطالبات، مما يشير إلى أن التشتت في معدلات الطلاب قد يكون أكثر من الطبيعي، وقد يحتاج إلى معالجة أكاديمية تهدف إلى تقليل ذلك الاختلاف.

وكمثال آخر، تسعى وزارة التربية والتعليم إلى دراسة حال التعليم العام من جميع الجوانب. وبدراسة أعمار كل من المدرسين X وأعمار المدرسات Y تم حصر 300 مدرس و 200 مدرسة وتم تحديد أعمارهم بالسنوات فتبين التالي

$$\bar{X} = 35 \quad \& \quad S_X^2 = 52$$
$$\bar{Y} = 33 \quad \& \quad S_Y^2 = 12$$

وتبعا للمؤشرات الإحصائية المتعلقة بأعمار المدرسين والمدرسات يتبين أن كل من المدرسين والمدرسات لهما نفس متوسط العمر تقريبا، ولكن الاختلاف يكمن في أن تشتت أعمار المدرسات $S_Y^2 = 12$ هو اقل بكثير من تباين أعمار المدرسين $S_X^2 = 52$. وتلك النتيجة تشير إلى أن أعمار المدرسات متقاربة كثيرا بدرجة قد تكون غير طبيعية، مما يثير التخوف بان تأتي مرحلة في العمر تصبح فيه جميع المدرسات متقادمات في العمر ولا يتوفر لهن إحلال إذا لم يتم تغذية الكادر التعليمي النسائي بمدرسات جدد كل فترة من الزمن.

معامل الاختلاف:

يتم مقارنة تشتت مجموعتين إذا كانت وحدة القياس واحدة، حيث يمثل ذلك شرط أساسي للحكم على تشتت المجموعتين مباشرة من خلال مقارنة قيم مقاييس التشتت المحسوبة للمجموعتين. فمثلا، يمكن مقارنة تشتت أعمار العمالة السعودية مع تشتت أعمار العمالة المقيمة، كما يمكن مقارنة تشتت الدخل الشهري

بالريال السعودي للموظفين الحكوميين وموظفي القطاع الخاص أو مقارنة تشتت درجات الطلاب مع تشتت درجات الطالبات للاستدلال دوما على المجموعة الأكثر اختلافا وتشتتا. أما إذا اختلفت وحدة القياس في المجموعتين، كأن تكون المجموعة الأولى ممثلة لساعات وقت العمل بينما المجموعة الثانية ممثلة لعمر العمال بالسنة، فانه لا يمكن بحال من الأحول مقارنة مؤشر إحصائي تم حسابه من ساعات بمؤشر إحصائي آخر تم حسابه بسنوات.

عندما تكون وحدة القياس مختلفة في المجتمعين المدروسين، فان أسلوب مقارنة التشتت يمكن أن تتم من خلال مقياس تشتت يسمى معامل الاختلاف. يهدف معامل الاختلاف إلى كشف المجموعة الأكثر تشتتا من بين المجموعات التي تكون وحدة قياسها مختلفة، حيث تعتبر المجموعة التي لها معامل اختلاف اكبر في القيمة هي المجموعة الأكثر تشتتا، ويتم حساب معامل الاختلاف من خلال قيم كل من المتوسط الحسابي والانحراف المعياري كما يلي:

$$CV_p = \frac{\sigma_X}{\mu_X} \times 100$$

حيث يمثل القيمة الحقيقة لمعامل الاختلاف المبني على قيم معالم المجتمع. أما إذا لم تتوفر قيم معالم المجتمع وتم الحصول على تقديرات لها بواسطة قيم عينة عشوائية ممثلة لمجتمع الدراسة، فانه يمكن تقدير قيمة معامل الاختلاف من خلال العلاقة،

$$CV = \frac{S_X}{\bar{X}} \times 100$$

مع الإشارة إلى أن معامل الاختلاف يكون بدون تمييز (وحدة قياس) حيث انه نسبة مئوية.

معامل الاختلاف

لمجتمع محدود

$$CV_p = \frac{\sigma_X}{\mu_X} \times 100$$

لعينة عشوائية

$$CV = \frac{S_X}{\bar{X}} \times 100$$

فعلى سبيل المثال، يعتمد مصنع كبير على مواد خام يتم استيرادها من مصادر مختلفة وكثيرة. وتختلف المصادر في كل من المدة المستغرقة لوصول الطلبيات والتي يتم حسابها باليوم وتكاليف الشحن التي يتم حسابها بالريال السعودي. وحيث أن الوقت المهدر في انتظار الطلبيات لا يقل أهمية عن الخسائر المادية الناتجة عن زيادة تكاليف الشحن، فان إدارة المخزون في المصنع ترغب في دراسة ومعرفة العامل الأهم من ناحية التكلفة هل هو مدة وصول الطلبيات أم تكاليف الشحن.

وبهدف الوصول إلى إجابة على التساؤل السابق تم حساب كل من الوسط الحسابي والتباين لمدة وصول الطلبيات X باليوم وتكلفة شحن الطلبية Y بالريال لعينة عشوائية من طلبيات سابقة فتم الحصول على الإحصائيات التالية:

$$\bar{X} = 7.8 \quad \& \quad S_X^2 = 10.5$$
$$\bar{Y} = 3200 \quad \& \quad S_Y^2 = 250000$$

وتبعا لاختلاف وحدة القياس في المتغيرين محل الدراسة، حيث يتم قياس مدة وصول الطلبية باليوم بينما يتم قياس تكلفة الشحن بالريال، لذا فان عملية مقارنة مقاييس التشتت العادية للمتغيرين غير مقبولة. وبهدف مقارنة تشتت المتغيرين، يتم حساب معامل الاختلاف لهما. وباستخدام إحصائيات العينة العشوائية المدروسة يتم الحصول على القيمة

$$CV\left(X\right)=\frac{S_X}{\overline{X}}\times 100=\frac{\sqrt{10.5}}{7.8}\times 100=41.5\%$$

والتي تمثل تقدير لقيمة معامل الاختلاف لمتغير مدة وصول الطلبيات.

كذلك يمكن الحصول على تقدير لقيمة معامل الاختلاف لمتغير تكلفة الشحن كما يلي:

$$CV\left(Y\right)=\frac{S_Y}{\overline{Y}}\times 100=\frac{\sqrt{250000}}{3200}\times 100=15.6\%$$

وعليه فان متغير مدة وصول الطلبية هو المتغير الأكثر تشتتاً، حيث يتضح من خلال الفرق الكبير بين معامل الاختلاف له المساوي لـ 41.5% ومعامل الاختلاف لمتغير تكلفة الشحن والمساوي لـ 15.6%. لذلك فان الاهتمام يجب أن يوجه إلى دراسة الاختلاف بين مدة وصول الطلبيات المختلفة والتي يمكن عند التحكم بها تقليل درجة الاختلاف ومن ثم تثبيت التكاليف المتحققة من خلال الانتظار لوصول الطلبيات أو على الأقل تثبيت تلك التكلفة.