مدخل إلى علم الإحصاء

د. علي بن محمد الجمعه

مادة الإحصاء العام

دبلوم التسويق والمحاسبة

الفصل الدراسي الثاني

للعام

١٤٢٧-١٤٢٨هـ

الباب الأول

الإحصاء الوصفي

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 والتي تمثل معلومة أساسية لاستيعاب محتوى الكتاب. سيتم 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. كذلك سيتم التطرق بشكل سريع إلى أهم قواعد الاحتمالات والمطلوبة كأساس لاستيعاب كثير من المعلومات المطروحة في فصول الكتاب المختلفة.

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 أشكال بيانية سهلة القراءة، بينما يتم في القسم الثاني الغوص داخل أعماق البيانات والقراءة بين السطور فيها للوصول إلى معلومات يصعب الحصول عليها بدون علم الإحصاء الحديث.

الإحصاء (Statistic)

الإحصاء هو العلم الذي يهتم بطرق جمع البيانات وأساليب وصفها وتحليلها بهدف استخراج المعلومات والحقائق التي لا يمكن الحصول عليها بطرق أخرى.

1-1-1 أنواع البيانات الإحصائية

تمثل البيانات الإحصائية مؤشرات حول ناتج كمي أو وصفي لحالة أو تساؤل محدد. وتنقسم البيانات الإحصائية إلى أنواع يمكن وصفها بشكل عام من خلال نوعين أساسيين هما بيانات وصفية و بيانات كمية. تنقسم البيانات الوصفية بدورها إلى قسمين رئيسيين هما بيانات وصفية اسمية وبيانات وصفية ترتيبية، حيث يتركز الفرق بين النوعين في كون البيانات الوصفية الاسمية تشير في الواقع إلى حقول مختلفة لا تمثل ترتيباً محدداً، بينما يكون للترتيب معناً محدداً في البيانات الوصفية الترتيبية. 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. في الجهة الأخرى نجد أن البيانات الوصفية النوعية تمثل حقول تشير إلى ترتيب محدد لا يمكن تجاهله. فالمستوى التعليمي يعكس ترتيباً محدداً حيث يتم ذكر مثلا حقل "غير متعلم" يليه " اقل من ثانوي" ثم "اقل من جامعي" يليه "جامعي فأعلى".

يمثل النوع الآخر للبيانات الإحصائية بيانات تأخذ أرقاماً ويطلق عليها بيانات كمية. تمثل البيانات الكمية عصب الحياة للعملية الإحصائية. فمعظم العمليات الإحصائية تتم من خلال التعامل مع أرقاماً أكثر من التعامل مع كلمات أو صفات. وحتى في عملية التعامل مع الصفات فان الهدف الرئيسي يكون غالباً في دراسة التكرارات أو عدد مرات ظهور الصفة المحددة والتي يمكن تسجيلها كمياً.

تختلف البيانات الكمية تبعاً لاختلاف القراءات للبيان، فالبيانات التي تأخذ أرقاماً محددة تمثل بيانات كمية متقطعة بينما البيانات الكمية التي تأخذ قيماً محصورة في فترة محددة تمثل بيانات كمية متصلة. فمثلاً، يمثل عدد الأطفال في البيت بياناً كمياً متقطعاً بحكم أن عدد الأطفال لا بد وان يكون عدداً صحيحاً. كذلك يمثل عدد الحوادث المرورية بياناً كمياً متقطعاً يأخذ قيماً بصيغ أعداد صحيحة لا تقبل أن تكون كسور، فعندما يقال أن عدد الحوادث

أربعة فانه يقصد فعلا أنها أربعة بالتحديد دون تقريب حسابي. في الجهة الأخرى يمثل وزن الطفل أو التكاليف المادية للحوادث بياناً كمياً متصلاً، غالبا ما يتم فيه التقريب لدرجة معينة للحصول على قيمة يمكن التعامل معها. 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.

وبالرجوع إلى مثال وزن الطفل نجد انه لا يمكن بحال من الأحوال وضع تصور لعدد القيم التي يمكن أن يأخذها البيان حتى مع وجود حد أعلى للأوزان. افترض أن الحد الأعلى لأوزان الأطفال لا يتجاوز 20 كيلوجرام مما ينتج عنه الفترة (20 , 0) كنطاق لقيم البيان. في هذه الحالة يوجد عدد لا نهائي من القيم للبيان الكمي والذي يقع في تلك الفترة، ولا يمكن التعبير عن تلك القيم بشكل مختلف عن الفترة المحددة سابقا. كذلك عندما يتم تحديد قيمة في تلك الفترة، مثل 13.5 كيلوجرام، كقيمة للمتغير العشوائي فان تلك القيمة تكون تقريبية في الغالب ولا تشير إلى القيمة الحقيقية الدقيقة والتي قد تكون قيمة دقيقة جداً يحددها درجة دقة الميزان المستخدم، بالإضافة إلى أن عملية الحصول على القيمة الحقيقية دون تقريب تعتبر عملية مستحيلة حيث يمكن للوزن أن يأخذ عدداً لانهائياً من الأرقام بعد الفاصلة العشرية.

يأخذ البيان المتقطع قيماً صحيحة في الغالب، مع إمكانية أن يأخذ قيم كسرية ولكن يظل بياناً عشوائياً متقطعاً. كمثال، افترض انه يوجد لدينا آلة للكشف عن التركيز الكيميائي لمحلول محدد، وافترض أن ذلك التركيز الكيميائي يأخذ إحدى ثلاث حالات فقط وهي: تركيز 0.17، تركيز o.32 وتركيز 0.60. في هذه الحالة نجد أن القيم التي يأخذها البيان محددة تحديداً تاماً وتعني القيمة الحقيقة بالإضافة إلى إمكانية حصرها في ثلاث حالات فقط. لذا فان البيان السابق يمثل بياناً متقطعاً وليس متصلاً.

يمكن للبيانات الكمية المتقطعة أن تكون ترتيبية أو نوعية. ويقصد هنا بالترتيب كون القيم التي يأخذها البيان المتقطع توحي أو تأخذ ترتيباً معيناً. فمثلاً، عدد الأطفال في البيت يمثل بياناً متقطعاً يحوي ترتيباً في القيم. فرقم 5 يأتي بعد الرقم 4 وهكذا. كمثال آخر لبيان متقطع نوعي، نجد أن رقم الآلة المنتجة لوحدة إنتاجية في مصنع يحتوي 6 آلات يمثل بياناً متقطعاً لا يحتوي ترتيباً في القيم. فالآلة رقم 3 لا يشترط أن تأتي بعد الآلة رقم 2 بحكم انه يمكن دوما إعادة توزيع الأرقام على الآلات من جديد مما قد ينتج عنه أرقاماً مختلفة للآلات. وفي الواقع يتم تحويل كثير من البيانات الاسمية إلى بيانات كمية متقطعة نوعية وذلك فقط ليتم التعامل معها كمياً. فالحالة الاجتماعية والجنس والتخصص في الجامعة والحالة العملية ونوع السيارة تمثل بيانات اسمية يمكن الإشارة إليها كمياً لتصبح بيانات كمية متقطعة نوعية.

بالرجوع إلى كثير من الكتب الإحصائية يتبين كثرة استخدام مصطلح متغير للإشارة إلى البيان الكمي، وقد يكون لذلك ارتباط لترجمة المصطلح الإنجليزي Variable والمستخدم عالمياً للإشارة إلى البيان الإحصائي. وفي الواقع يعطي مصطلح متغير معنى أقوى من المعنى المحصل من كلمة بيان، حيث أن كلمة متغير تشير إلى وجود تغير للقيمة المحصلة من البيان وتعددها مما ينتج عنه وجود حاجة للتعامل كمياً أو إحصائيا مع تلك القيم

المختلفة. وعليه فإننا سنعتمد مصطلح متغير في الفصول اللاحقة للإشارة إلى البيان الإحصائي، حيث سيتم استخدام كل من مصطلح متغير وصفي ومتغير كمي متصل ومتغير كمي متقطع وبنفس التفصيل الوصفي المعمول به مع مصطلح البيان الإحصائي.

يمثل المتغير الكمي المتصل أو ما يطلق عليه أحيانا المستمر، المتغير الأكثر استخداما في العمليات الإحصائية عموما. فمعظم المتغيرات التي يتم استخدامها في عملية الاستدلال الإحصائي تمثل متغيرات كمية متصلة، خاصة في مجال الأموال والإنتاج والدراسات الإدارية عموماً. فالربح والخسارة والوقت المستغرق لوصول طلبية أو لإنتاج وحدة ونسبة المعيب في الإنتاج تمثل متغيرات كمية مستمرة. وبحكم كون المتغير الكمي المتصل يأخذ قيماً عددها لانهائي فان هذا المتغير يمثل المناخ المناسب لتفعيل النظريات الإحصائية القائمة على التقريب والاستدلال حول القيم الحقيقة سواء لمتوسطات أو لقيم متنبأ بها.

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، فمنها العينات العشوائية، كما سيرد تفصيل لها في الفصل الثالث، ومنها العينات الغير عشوائية، بيد أنها تتفق جميعا في كونها جزء من المجتمع وممثلاً له. 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 تبعا 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 المقابل عند 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. وبالطبع فان 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 ينبني 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.

في السابق وقبل توفر التقنية الحديثة كانت عملية التعامل مع البيانات وتبويبها وحصرها وإبرازها تستنزف الجهد الكبير والوقت الكثير من الباحثين مما شكل عائقاً كبيراً أمام العمليات التطبيقية الإحصائية. بيد انه ومع التطور التقني وتوفر برامج حاسوبية، تمنح الباحث قدرات هائلة في التعامل مع البيانات مهما كان حجمها ودقتها، أصبحت العمليات الإحصائية التطبيقية يتم تنفيذها بوقت قياسي وفعالية اكبر بكثير مما كان معمول به في السابق.

سيتم التطرق في هذا الجزء إلى آلية عرض البيانات جدوليا (تبويبها) وذلك من خلال بيان آلية بناء الجداول التكرارية وطرق الحصول على التكرارات النسبية والتكرارات التراكمية. كما سيتم التطرق إلى عملية عرض البيانات بيانيا والأساليب المختلفة فيها، وسيتم التركيز بشكل اكبر على الأساليب المختلفة لقراءة الرسوم البيانية دون التفصيل في أداء الرسم بشكل يدوي حيث تتوفر كثير من البرامج الإحصائية والتي يمكن من خلالها الحصول على الكثير من الرسومات البيانية والمعدة بشكل كامل ومتقن.

عند 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 (في حدود العشرة حقول كحد أعلى ليصبح مفيداً إحصائيا)، وإذا كان عدد القيم كبيراً يصبح الجدول ذا حقول كثيرة ومن ثم يصعب الحصول على معلومة منه.

مثال 1-2-1

الجدول 1-2-1 يمثل التخصص الأكاديمي لـ 80 طالب في كلية العلوم الإدارية والتي تضم سبعة تخصصات هي الأساليب الكمية وإدارة الأعمال والإدارة العامة والمحاسبة والاقتصاد والعلوم السياسية والأنظمة. المطلوب إنشاء جدول تكراري بسيط لتوزيع الطلاب حسب التخصص الأكاديمي.

جدول 1-2-1: التخصص الأكاديمي لـ 80 طالب في كلية العلوم الإدارية

أعمال	محاسبة	عامة	أنظمة	عامة	أساليب	أساليب	أعمال
عامة	أساليب	عامة	أعمال	اقتصاد	أنظمة	سياسة	أنظمة
أعمال	أعمال	سياسة	محاسبة	أنظمة	أعمال	أساليب	أنظمة
محاسبة	أساليب	اقتصاد	أساليب	عامة	سياسة	عامة	محاسبة
عامة	سياسة	عامة	أعمال	أنظمة	عامة	أساليب	أنظمة
أساليب	أنظمة	اقتصاد	محاسبة	اقتصاد	أساليب	اقتصاد	اقتصاد
أنظمة	عامة	أعمال	عامة	أعمال	محاسبة	أساليب	اقتصاد
أساليب	أنظمة	سياسة	سياسة	اقتصاد	أساليب	اقتصاد	سياسة
محاسبة	سياسة	أساليب	أساليب	عامة	أساليب	أعمال	اقتصاد
أنظمة	اقتصاد	اقتصاد	أنظمة	محاسبة	سياسة	اقتصاد	أعمال

الحـل

بما أن المتغير العشوائي الاسمي يتكون من سبع مسميات أقسام مختلفة لذا فان الجدول التكراري المطلوب يتكون من سبعة حقول كما هو مبين في الجدول (1-2-2).

يتم تفريغ بيانات المتغير العشوائي بالبحث عن عدد مرات ظهور مسمى التخصص المحدد في البيانات الخام، فمثلا بالبحث عن عدد مرات ظهور مسمى "أساليب" تبين ورودها 15 مرة، لذا فان التكرار المرافق لمسمى "أساليب كمية" في الجدول التكراري هو 15. وبتطبيق نفس الأسلوب على باقي مسميات التخصصات المختلفة يتم الحصول على جميع التكرارات المصاحبة لحقول الجدول التكراري المطلوب.

جدول 1-2-2: التوزيع التكراري للطلاب حسب التخصص

التكرار	التخصص
15	أساليب كمية
8	محاسبة
11	إدارة الأعمال
12	أنظمة
13	اقتصاد
12	إدارة عامة
9	علوم سياسية
80	المجموع

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.

مثال 1-2-2

الجدول 1-2-3 يمثل عدد السيارات المملوكة لـ 120 عائلة تتكون من عشرة أفراد يعيشون في منزل واحد. قم بإنشاء جدول تكراري بسيط لتوزيع العائلات حسب عدد السيارات المملوكة لها.

جدول 1-2-3: عدد السيارات المملوكة بواسطة أفراد العائلة

3	6	1	6	1	1	7	5	1	1	2	3
2	3	7	2	6	3	7	4	7	1	2	6
4	4	5	3	7	2	6	1	7	3	7	4
2	2	1	6	3	6	3	5	4	4	3	3
7	5	5	1	5	7	3	5	7	3	1	1
5	6	4	5	3	2	6	7	6	3	3	5
2	6	3	2	2	2	2	5	1	1	3	7
6	3	7	5	6	2	4	3	5	1	1	4
3	7	1	6	3	7	6	3	6	1	4	5
5	1	3	6	3	6	3	5	2	7	5	7

الحـــل

بالبحث في مدى المتغير الكمي المتقطع يتبين أن اصغر قيمة هي 1 واكبر قيمة هي 7 حيث تفيد بإمكانية اعتماد سبعة حقول لإنشاء الجدول التكراري المطلوب، الجدول 1-2-4. وبتفريغ بيانات المتغير العشوائي من خلال البحث عن عدد مرات ظهور قيمة محددة من بين القيم المحددة بين

الرقم 1 والرقم 7 يتم الحصول على التكرارات المرافقة لحقول الجدول التكراري. فمثلا بالبحث عن عدد مرات ظهور الرقم 4 يتبين وروده 10 مرات، لذا فان التكرار المرافق لحقل القيمة 4 في الجدول التكراري هو 10.

جدول 1-2-4: التوزيع التكراري لعدد السيارات المملوكة بواسطة أفراد العائلة

عدد السيارات	التكرار
1	18
2	15
3	25
4	10
5	17
6	18
7	17
المجموع	120

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.

مثال 1-2-3

الجدول 1-2-5 يمثل عدد الموظفين في 144 مؤسسة تجارية. المطلوب إنشاء جدول تكراري بسيط مكون من خمس حقول لتوزيع عدد الموظفين في المؤسسات التجارية.

جدول 1-2-5: عدد الموظفين في 144 مؤسسة تجارية

20	8	6	7	9	16	14	13	11	6	8	10
19	2	9	16	13	5	1	12	13	17	17	12
8	11	9	7	1	20	16	7	18	13	10	11
16	10	9	16	17	16	16	8	10	20	6	4
1	19	20	8	9	7	11	4	11	4	14	20
4	17	1	14	15	10	14	11	18	17	15	8
1	8	20	4	11	12	20	17	6	16	17	15
9	16	5	3	11	6	2	14	3	11	5	20
13	10	15	1	18	4	19	19	13	19	10	17
6	14	19	6	8	7	13	9	5	2	7	14
7	16	5	10	9	18	6	19	13	5	3	5
4	2	11	10	9	2	3	9	10	20	6	5

الحــل

تبعا لمدى المتغير العشوائي محل الدراسة يتبين أن القيم تتــراوح بــين الــرقم 1 والــرقم 20 ممــا يشــير إلى عدم إمكانية إفراد حقل مستقل في الجدول التكــراري المطلــوب لكــل قيمــة حيــث ســينتج جــدول بعشرين حقل تتلاشى معه الفائدة الأساسية من الجــدول وهــي تلخــيص وإبــراز المتغيــر فــي وعــاء سهل القراءة. وبما أن المطلوب هنا هو إنشاء جدول تكــراري مكــون مــن خمــس حقــول فانــه ســيتم تقسيم مدى المتغير العشوائي المتمثل بعدد الموظفين إلــى خمــس فئــات بحيــث يــتم الحصــول علــى تحديد كامل للقيم الداخلة في حقول الجدول كما هو مبين في الجدول 1-2-6.

جدول 1-2-6: التوزيع التكراري لعدد الموظفين

التكرار	عدد السيارات
22	1,2,3,4
32	5,6,7,8
33	9,10,11,12
29	13,14,15,16
28	17,18,19,20
144	المجموع

وبمعرفة القيم المحددة لحقول الجدول التكراري يــتم عــد مــرات ظهــور القــيم المحــددة للحقــل فــي البيانات ليتم اعتماد المجموع كتكرار للحقل.

أما بالنسبة لتبويب البيانات الكمية المتصلة فانه يتم من خلال تفريغها فــي جــداول تكراريــة مكونــة مــن فئــات مختلفة ومتسلسلة بحيث تغطي الفئات مدى المتغير المستهدف إفراغه في الجــدول التكــراري. ويــتم بنــاء الفئــة بالشــكل a - b حيث يشير a إلى الحد الأدنى للفئة ويشير b إلى الحد الأعلى للفئة، وتقرأ الفترة بالشكل التالي:

"a - b" = "من القيمة a إلى اقل من القيمة b"

وعليه فان تلك الفئة تحوي جميع قيم المتغير الكمي المفرغ في الجــدول التكــراري والتــي تكــون اكبــر مــن أو تســاوي القيمة a وفي نفس الوقت تكون اقل من b. يمثل العامود الثاني التكرارات المصــاحبة لكــل فئــة والتــي تشــير إلــى عــدد قيم المتغير الكمي التي تقع في الفئة. تجدر الإشارة إلى انه يجب أن يكــون مجمــوع التكــرارات لجميــع الفئــات مســاوي لعدد قيم المتغير الإجمالي.

وبافتراض توفر متغير كمي متصل ويرغب في إفراغه في جــدول تكــراري. لتحقيــق هــذا الهــدف يــتم تنفيــذ الخطوات التالية:

- تحديد مدى المتغير الكمي المستهدف
- تحديد عدد الفئات أو طول الفئة المطلوب

- تحديد حدود الفئات
- إفراغ البيانات

عندما تتوفر قيم لتغير عشوائي كمي متصل يتم في البداية الكشف عن مدى المتغير، حيث يمكن تحديد المدى R من خلال الدالة التالية:

$$R = Max - Min \qquad (1)$$

حيث

- Max = اكبر قيمة مشاهدة للمتغير
- Min = اصغر قيمة مشاهدة للمتغير

يجب التنبيه إلى حقيقة تأثر المدى سلبيا بالقيم المتطرفة أو الشاذة للمتغير العشوائي، حيث يمكن لها أن تضخم المدى بشكل مبالغ فيه مما ينتج عنه فئات لا تمثل التوزيع التكراري المناسب للمتغير العشوائي. ولكن يمكن استبعاد القيم المتطرفة، إذا وجدت، عند حساب المدى ومن ثم إدراجها في الفئة الأولى أو الفئة الأخيرة حسب موقع تطرفها. في هذه الحالة يطلق على الجدول التكراري صفة الجدول المفتوح بحكم كون الفئة الأولى أو الأخيرة أو كلاهما ستكون بدون إحدى الحدود، ففي حال الفئة الأولى يتم كتابتها بالشكل التالي

"< b"

مشيرة إلى احتواء الفئة الأولى على جميع القيم التي تقل عن القيمة b. أما في حالة الفئة الأخيرة فانه يتم التعبير عنها بالشكل التالي:

"≥ a"

والتي تعني احتواء الفئة الأخيرة على جميع القيم التي تكون اكبر من أو مساوية للحد الأدنى للفئة.

وفي الوضع الطبيعي عندما لا يكون هناك قيم متطرفة فانه يمكن تحديد عدد الفئات أو طول الفئة من خلال العلاقة التالية:

$$K = \frac{R+1}{L} \qquad (2)$$

أو

$$L = \frac{R+1}{K} \qquad (3)$$

حيث

- K = عدد الفئات
- R = المدى
- L = طول الفئة

وحيث أن عملية تبويب البيانات هي في الأصل عملية تقريبية تلخيصية فانه يمكن دوما تلافي الكسور لاسيما عند التعامل مع متغير يأخذ قيماً كبيرةً نسبيا. ويتم التقريب هنا إلى الأعلى وذلك ضمانا لتغطية مدى الفئات

للبيانات، لذا يمكن تحديد طول الفئة (L) بأقرب عدد صحيح من الأعلى. في المقابل يمكن تحديد عدد الفئات (K) الواجب اعتماده إذا كان محدد مسبقا طول الفئة، ويتم ذلك باستخدام المعادلة (3) حيث يتم أيضا تقريب الناتج إلى اقرب عدد صحيح للأعلى.

وبتحديد عدد الفئات وطول الفئة يتم الحصول على حدود الفئات للجدول التكراري المستهدف. وباعتماد القيمة الصغرى من بين قيم المتغير الكمي كحد أدنى للفئة الأولى، وبإضافة طول الفئة للحد الأدنى لها للحصول على الحد الأعلى للفئة،

$$a_1 \text{ - } (a_1 + L)$$

حيث تضم هذه الفئة جميع القيم التي تكون مساوية للحد الأدنى للفئة أو اكبر إلى اقل من الحد الأعلى. يتم تحديد الفئة الثانية بنفس الأسلوب مع الأخذ بالاعتبار أن الحد الأدنى لها سيكون مساوياً للحد الأعلى للفئة السابقة،

$$a_2 - \ (a_2 + L)$$

حيث

$$a_2 = a_1 + L$$

لذا فان فئات الجدول التكراري المحتوي على K فئة بطول ثابت لكل الفئات L يمكن كتابتها رياضيا بالشكل التالي:

$$a_i \text{ - } a_{i+1} \qquad \forall \quad i = 1, 2, ..., K \tag{4}$$

وبتحديد حدود فئات الجدول التكراري المطلوب يتم تفريغ قيم المتغير العشوائي في خانات التكرار (F) في الجدول التكراري، بحيث يمثل كل تكرار مرافق لفئة محددة (F_i) عدد مرات ظهور قيم المتغير في تلك الفئة. الجدول 1-2-7 يبين هيكل الجدول التكراري العادي حيث يمثل مجموع التكرارات $\sum F_i$ إجمالي عدد القيم المشاهدة والتي يرمز لها في حال البيانات غير المبوبة بالرمز n. يمثل الحصول على التكرار النسبي $(F\%)$ واحد من المعلومات المهمة إحصائيا والتي يمكن استنتاجها من خلال الجداول التكرارية. ويتم في الواقع حساب التكرار النسبي بواسطة قسمة التكرار للفئة على مجموع التكرارات الكلي والضرب في 100،

$$F_i\% = \frac{F_i}{\sum F_i} \times 100 \tag{5}$$

جدول 1-2-7: هيكل الجدول التكراري والتكراري النسبي

حدود الفئات	F	$F\%$
a_1 - a_2	F_1	$F_1\%$
a_2 - a_3	F_2	$F_2\%$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
a_K - a_{K+1}	F_K	$F_K\%$
المجموع	$\sum_{i=1}^{K} F_i$	100

يشير التكرار النسبي إلى حجم تركيز تواجد قيم المتغير في كل فئة مما يعطى قدرة سريعة على عملية مقارنة ذلك التركيز بين الفئات. تنبع تلك الفائدة عندما تكون التكرارات كبيرة ويصعب من خلالها الحصول على تصور ذهني مباشر وسريع لتركيز القراءات في الفئات. وحيث أن الهدف الأساسي من إنشاء جدول تكراري يكمن في عملية تلخيص البيانات والحصول على مؤشرات حول اتجاهات قيم المتغير فان التكرار النسبي يكون من المؤشرات المهمة في هذا المجال.

مثال 1-2-5

الجدول 1-2-8 يمثل حجم المبيعات اليومية بآلاف الريالات لإحدى المحلات التجارية الكبرى وذلك للثلاثة شهور الماضية. المطلوب إنشاء جدول تكراري للمبيعات اليومية بحيث يكون عدد الفئات خمسة. كذلك أوجد التكرارات النسبية المرافقة لحقول الجدول التكراري الناتج.

جدول 1-2-8: المبيعات اليومية بآلاف الريالات لإحدى المحلات التجارية

26.9	77.8	14.9	33	46.3	71.5	46	7.9	64.5	87.8
33.7	27	71.8	33.8	59.1	28.8	80.2	81.3	43.7	41.6
7.1	64	58.3	37.9	72.2	79.9	54.3	89.7	61.3	21.5
70.1	67	15.5	2.2	28.6	31.7	4.9	26.8	99.2	0.3
10.5	28.3	99.3	14.1	47	98.1	20.3	44.2	21.7	66.3
3.2	6.2	67	39.6	39.2	62.3	21.9	72.2	67.5	64
93.7	51	40	20.2	73.9	18.2	25.4	24.5	67.3	16.4
25.4	35.9	26.5	83.1	88	37.8	19.1	80	45.4	73.8
73.9	63.8	54.2	23.2	53	52.3	63.6	46.2	24.9	55.6

الحـل

كخطوة أولى يتم تحديد المدى

$$R = Max - Min = 99.3 - .3 = 99$$

وحيث أن المطلوب إنشاء جدول تكراري من خمس فئات $(K = 5)$، لذا فان طول الفئة يجب أن يكون مساوياً لنتيجة الدالة التالية،

$$L = \frac{R+1}{K}$$
$$= \frac{99+1}{5} = 20$$

وباستخدام القيمة الصغرى في قيم المتغير والمساوية تقريبا للصفر كحد أدنى للفئة الأولى يتم تحديد حدود الفئات الخمس المطلوبة بالإضافة إلى تحديد عدد القيم المضمنة في كل فئة والتي تمثل التكرارات في الجدول المطلوب. الجدول 1-2-9 يمثل النتيجة المحصلة من العملية السابقة.

جدول 1-2-9: الجدول التكراري والتكراري النسبي للمبيعات اليومية

حدود الفئات (آلف ريال)	F	$F\%$
0 – 20	14	15.56
20 – 40	26	28.89
40 – 60	17	18.89
60 – 80	22	24.44
80 – 100	11	12.22
المجموع	90	100

وبتوفر كل من التكرارات العادية والتكرارات النسبية للمتغير الكمي في الفئات المحددة يمكن حساب كل من التكرار التراكمي الصاعد والتكرار التراكمي الهابط بالإضافة إلى التكرار التراكمي النسبي الصاعد والهابط. يقصد هنا بكلمة تراكمي إجمالي التكرارات إلى فئة معينة وابتداء من إحدى أطراف الجدول. لذا فإنه من المتوقع أن تكون بداية الجداول التراكمية الصاعدة مرتبطة بالتكرار صفر بينما تكون نهايتها مرتبطة بالتكرار التراكمي المساوي لإجمالي التكرارات في الجدول التكراري الأصلي. وبالنسبة للجداول التكرارية التراكمية الهابطة يكون الوضع مغايراً للوضع السابق، حيث يكون تكرار البداية للفئة الأولى مساوياً لإجمالي التكرارات وينتهي بتكرار مساوي للصفر للفئة الأخيرة.

لإنشاء جدول تكراري تراكمي (متجمع) صاعد أو هابط يتم أولا تحديد الفئات التراكمية الصاعدة أو الهابطة ومن ثم يتم إيجاد التكرارات التراكمية المصاحبة لها. يتم صياغة حدود الفئات الصاعدة باستخدام الحدود الدنيا لفئات الجدول التكراري بالإضافة إلى الحد العالي للفئة الأخيرة، حيث تسبق تلك الحدود إشارة "اقل من" (>)، بينما يتم استخدام إشارة "اكبر من أو يساوي" (≥) لبناء حدود الجدول التكراري الهابط. لذا فان عدد فئات الجداول التكرارية التجميعية يكون مساوياً لعدد فئات الجدول التكراري العادي زائد واحد صحيح، حيث ترجع الزيادة إلى حقيقة استخدام الحد الأعلى للفئة الأخيرة بالإضافة إلى استخدام جميع الحدود الدنيا.

ويتم تحديد التكرارات المصاحبة للفئات التراكمية الصاعدة من خلال القاعدة التالية:

$$CF_r = \sum_{i=0}^{r-1} F_i \qquad \forall \quad r = 1,2,....,K,(K+1) \tag{6}$$

حيث $F_0 = 0$ بينما يتم استخدام الدليل $(K+1)$ للإشارة إلى الحد الأخير في الجدول التجميعي الصاعد أو الهابط. وبأخذ نسبة تكرار الفئة التراكمية الصاعدة إلى إجمالي التكرارات يتم الحصول على التكرار التراكمي الصاعد النسبي، وبنفس الأسلوب يتم الحصول على التكرارات التراكمية الهابطة النسبية.

$$CF_r\% = \frac{CF_r}{\sum_{i=1}^{K} F_i} \qquad \forall \quad r = 1,2,....,K,(K+1) \tag{7}$$

حيث تمتاز النسب التراكمية الصاعدة أو الهابطة بأنها تعطي مؤشراً مباشراً عن مواقع التركيز للقراءات بنسب مختلفة. الجدول 1-2-10 يمثل الهيكل للجدول التكراري الصاعد، بينما الجدول 1-2-11 يمثل الهيكل للجدول التكراري الهابط.

جدول 1-2-10: هيكل الجدول التكراري المتجمع الصاعد

حدود صاعدة	تكرار صاعد CF
$< a_1$	$CF_1 = 0$
$< a_2$	CF_2
$< a_3$	CF_3
$\vdots$	$\vdots$
$< a_K$	CF_K
$< a_{K+1} = a_K + L$	$CF_{K+1} = n$

جدول 1-2-11: هيكل الجدول التكراري المتجمع الهابط

حدود هابطة	تكرار هابط
$\geq a_1$	$\sum_{i=1}^{K} F_i - CF_1 = \sum_{i=1}^{K} F_i$
$\geq a_2$	$\sum_{i=1}^{K} F_i - CF_2$
$\geq a_3$	$\sum_{i=1}^{K} F_i - CF_3$
$\vdots$	$\vdots$
$\geq a_K$	$\sum_{i=1}^{K} F_i - CF_K$
$\geq a_{K+1} = a_K + L$	$\sum_{i=1}^{K} F_i - CF_{K+1} = 0$

تمتاز الجداول التكرارية التراكمية بأنها جداول تكرارية منتظمة اتجاه التغير (Monotonic) بمعنى أن تغير التكرارات فيها يكون تصاعديا أو تنازليا فقط ولا يمكن أن يحوي الصفتين معا. وتلك الميزة تعطي الباحث القدرة على تحديد قيمة المتغير الكمي والتي تكون نسبة القيم التي تقل أو تزيد عنها محددة مسبقا.

مثال 1-2-6

باستخدام الجدول التكراري المحصل في المثال 1-2-5، أوجد كل من الجدول التكراري المتجمع الصاعد والنسبي الصاعد والجدول التكراري المتجمع الهابط والنسبي الهابط. ومن ثم أوجد التالي:

أ) عدد الأيام التي تكون فيها المبيعات 40 آلف ريال فأكثر

ب) عدد الأيام التي تكون فيها المبيعات اقل من 60 آلف ريال

ج) عدد الأيام التي تكون فيها المبيعات بين 40 آلف و 80 آلف ريال

د) نسبة الأيام التي تقل فيها المبيعات عن 60 آلف ريال

هـ) نسبة الأيام التي لا تقل فيها المبيعات عن 80 آلف ريال

و) نسبة الأيام التي تكون فيها المبيعات بين 20 آلف و 60 آلف ريال

الحل

باستخدام بيانات الجدول 1-2-9 يتم تحديد كل من الحدود الصاعدة والحدود الهابطة للجداول المطلوبة، ومن ثم يتم حساب التكرارات الصاعدة والنسبية الصاعدة والتكرارات الهابطة والنسبية الهابطة باستخدام كل من المعادلة (1.6) والمعادلة (1.7) على التوالي. الجداول 1-2-12 و 1-2-13 يمثلان النتائج المحصلة.

باستخدام الجداول المحصلة يتم الحصول على الأجوبة للاستفسارات المبينة في السؤال.

أ) عدد الأيام التي تكون فيها المبيعات 40 آلف ريال فأكثر بلغ 50 يوم، حيث يمكن ملاحظة النتيجة في الجدول التجميعي الهابط، الجدول 1-2-13، في الفئة الثالثة.

ب) عدد الأيام التي تكون فيها المبيعات اقل من 60 آلف ريال بلغ 57 يوم (الفئة الرابعة في الجدول 1-2-1-12)

ج) عدد الأيام التي تكون فيها المبيعات بين 40 آلف و 80 آلف ريال بلغ 39 يوم والذي يمثل الفرق بين عدد الأيام التي تكون فيها المبيعات اقل من 40 آلف (40) وعدد الأيام التي تكون فيها المبيعات اقل من 80 آلف (79) من الجدول الصاعد، أو الفرق بين عدد الأيام التي تكون فيها المبيعات أكثر من 40 آلف (50) وعدد الأيام التي تكون فيها المبيعات أكثر من 80 آلف (11) من الجدول الهابط.

جدول 1-2-12: الجدول التكراري التراكمي الصاعد والنسبي الصاعد

حدود صاعدة	تكرار صاعد	تكرار نسبي صاعد
<0	0	0
<20	14	15.56
<40	40	44.45
<60	57	63.34
<80	79	87.78
<100	90	100

جدول 1-2-13: الجدول التكراري التراكمي الهابط والنسبي الصاعد

حدود هابطة	تكرار هابط	تكرار نسبي هابط
≥ 0	90	100
≥ 20	76	84.44
≥ 40	50	55.55
≥ 60	33	36.66
≥ 80	11	12.22
≥ 100	0	0

د) نسبة الأيام التي تقل فيه المبيعات عن 60 آلف ريال هي 63.34% من إجمالي عدد الأيام، من الجدول الصاعد.

هـ) نسبة الأيام التي لا تقل فيه المبيعات عن 80 آلف ريال هي 12.22% من إجمالي عدد الأيام، من الجدول الصاعد.

و) نسبة الأيام التي تكون فيها المبيعات بين 20 آلف و 60 آلف ريال هي الفرق بين النسبتين نسبة الأيام التي تكون فيه المبيعات اقل من 20 آلف ونسبة الأيام التي تكون فيه المبيعات اقل من 60 آلف والتي تساوي $63.34 - 15.56 = 47.78$ ، من الجدول الصاعد

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 الحقول المختلفة إلى إجمالي التكرارات. كذلك يمكن التعبير عن التكرارات في الجداول التكرارية بالصيغ التجميعية أو الصيغ النسبية العادية أو التجميعية. وفيما يتعلق بالعرض البياني، فانه يمكن تبني أسلوب يتميز بالبساطة والوضوح، كما انه يمكن التعامل مع التكرارات أو النسب مباشرة أو التعامل مع الصيغ التجميعية.

يتكون العرض البياني من عدة أجزاء تتكامل لتخرج جملة بيانية مفهومة غير منقوصة المعنى. ويمثل أهم تلك الأجزاء عنوان الرسم البياني والذي يكون في العادة نقطة الانطلاق للعين المتفحصة للعرض أو الرسم البياني. يلي ذلك في الأهمية عناوين المحاور وقيمها، وتلك معلومة ضرورية لا يمكن لمتفحص الرسم البياني معرفة أساس القيم المرسومة ووحدة قياسها ما لم تكن تلك المحاور معرفة تماما. تمثل مفاتيح الرسم الجزء الثالث المهم في عملية إيصال المعلومة الإحصائية بيانيا. ويهتم مفتاح الرسم بوصف مكونات الرسم الداخلية وعلاقتها بالمتغيرات المختلفة. ويمكن استخدام الألوان أو الأشكال للتعبير عن مفاتيح الرسم، بيد أن قوة التعبير هنا تعتمد بشكل أساسي على الوضوح والبساطة دون الحاجة إلى المبالغة باختيار الألوان أو الأشكال. يأتي في المرتبة الرابعة ألوان الخلفية وحجم الرسم والخطوط ونوع الرسم ومدى مناسبته للمعلومة المبرزة وموضع مفاتيح الرسم والإيضاحات والعنوان وغيرها من التفاصيل الشكلية ولكن المهمة في وضوح المعنى المستهدف في الرسم البياني.

تختلف أنواع الرسوم البيانية بــاختلاف أهــدافها. ومــن الأنــواع الأكثــر اســتخداما الأعمــدة البيانيــة والخــط البياني والدوائر. وفي الواقع لا يمكن بأي حــال مــن الأحــوال حصــر جميــع الأشــكال التــي يمكــن أن يــتم إيضــاح المعلومة الإحصائية من خلالها خاصة مع التطور التقني الحــديث ومــا صــاحبه مــن تطــور فــي البــرامج الإحصــائية مثل برنامج SAS و برنامج SPSS وغيرها والتي أعطت اهتمام كبير للرســوم البيانيــة والإشــكال الفنيــة لهــا. لــذا فانــه سيتم التطرق إلى بعض الأمثلة بهدف إيضاح جانب مــن الإمكانــات المتاحــة دون الادعــاء بــان تلــك الأمثلــة ســتكون شاملة لجميع جوانب موضوع العرض البياني.

يعتمد اختيار نوع الرسم البياني المناسب على عدة عوامــل مــن أهمهــا طبيعــة القــيم الكميــة المــراد رســمها. فرسم المتغير الكمي المتقطع يختلف عن رسم المتغيــر الكمــي المتصــل، كمــا أن المتغيــر المــرتبط بــالزمن كالإنتــاج الشهري أو الأربح السنوية يمكن رسمها بأسلوب مختلف عــن القــيم الغيــر مرتبطــة بــالزمن كــالحقول غيــر الترتيبيــة مثل الحالة الاجتماعية والمنطقة السكنية والحالة التعليمية. وفي حال التعامــل مــع أكثــر مــن متغيــر فانــه يمكــن رســم المتغيرات في بيان واحد. في هذه الحالة يأتي دور مفتاح الرسم ليدل على موقــع كــل متغيــر فــي الرســم البيــاني. كمــا يمكن اللجوء في بعض الحالات إلى رسم النسب بدل من القيم الفعلية خاصة عند التعامــل مــع أكثــر مــن متغيــر، حيــث يمكن لهذا الأسلوب إبراز المعلومة بشكل أوضح خاصة عندما يكون هنالك تباين كبير بين قيم المتغير المستهدف.

مثال 1-3-1

الجدول 1-3-1 يبين التوزيع التكراري لعينــة عشــوائية حجمهــا 320 طالــب فــي التعلــيم العــالي فــي المملكة العربية السعودية، حيــث تــم تصــنيفهم حســب الجــنس (طالــب، طالبــة) وحســب المنطقــة (وسطى، شمالية، شرقية، غربية، جنوبية). المطلوب إبراز البيانــات فــي الجــدول بيانيــا بعــدة طــرق مختلفة.

جدول 1-3-1: توزيع 320 طالب تعليم عالي حسب الجنس والمنطقة

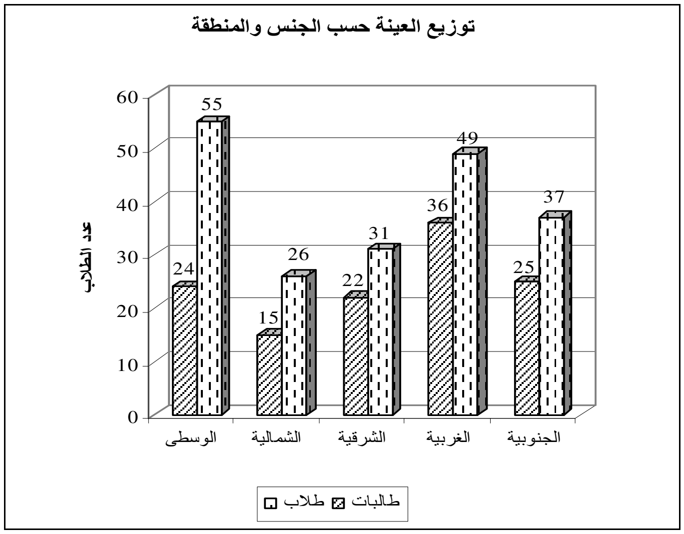
المنطقة	طلاب	طالبات
الوسطى	55	24
الشمالية	26	15
الشرقية	31	22
الغربية	49	36
الجنوبية	37	25

الحــل

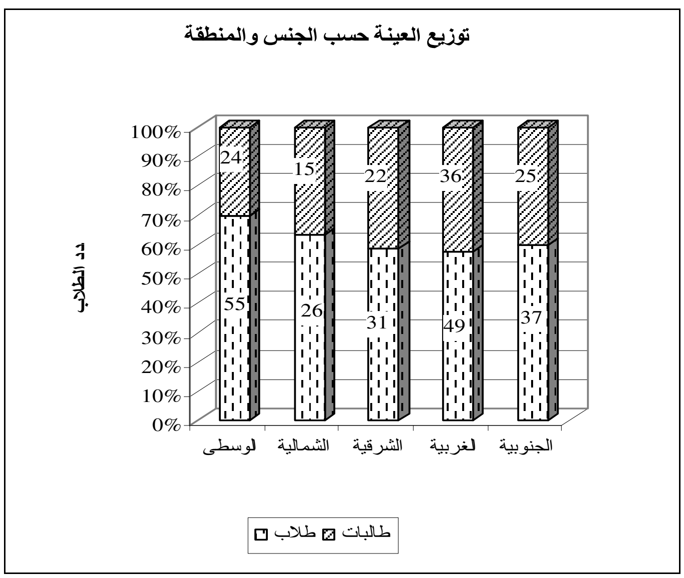
في البداية يمكن استخدام الأعمدة البيانيــة المزدوجــة لتمثيــل أعــداد الطــلاب والطالبــات فــي العينــة العشوائية. الشكل 1-3-1 يبين إحدى الإمكانات التي يمكــن فيهــا إبــراز المعلومــات فــي الجــدول 1-3-1 بيانيا، حيث تم تمثيل كل منطقة بعامودين: أحداهما للطــلاب والآخــر للطالبــات. تجــدر الإشــارة إلــى إن الرسم البياني قد احتوى على جميع العناصــر المطلوبــة وهــي عنــوان الرســم وعنــاوين المحــاور ومفتاح الرسم بالإضافة إلى القيم المرافقة للأعمدة.

كطريقة أخرى في عرض المعلومة المبينة في السؤال يمكن استخدام التوزيع النسبي للتكرارات. الشكل 1-3-2 يوضح توزيع العينة حسب المنطقة موزعة نسبياً حسب الجنس، مع ملاحظة أن كل منطقة تم تمثيلها بعامود واحد يبين مجموع الطلاب ومجموع الطالبات.

شكل 1-3-1 الأعمدة البيانية المزدوجة

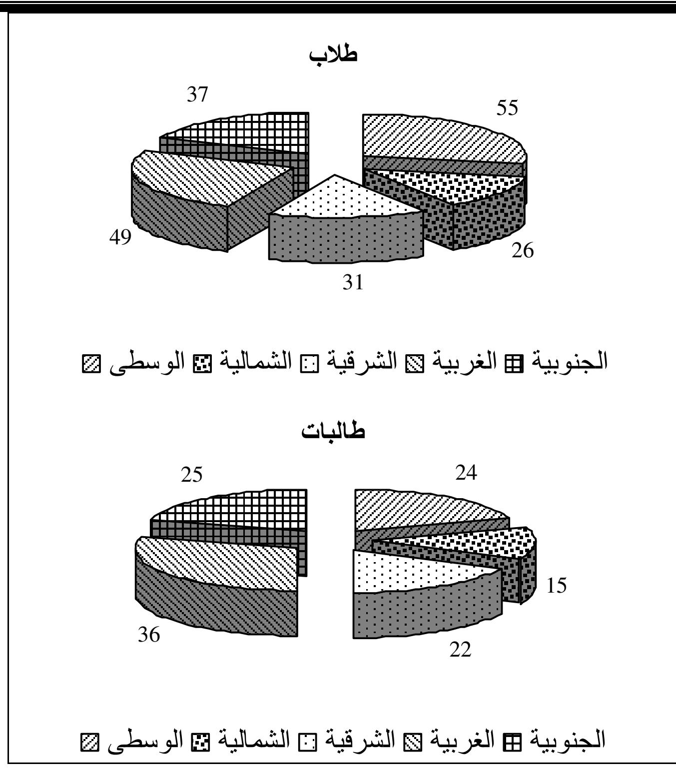


شكل 1-3-2 الأعمدة البيانية المزدوجة النسبية



يمكن كذلك استخدام رسوم بيانية أكثر جمالا وجاذبية مثل رسوم الدوائر، بيد أن مثل هذه الرسوم تكون أكثر فعالية في حال التعامل مع متغير واحد فقط، في حين أن التعامل مع أكثر من متغير يضيف إلى الرسم الواحد تعقيدات قد تلغي الفائدة المرجوة من مثل تلك الرسوم ألا وهي إيصال المعلومة بشكل واضح. الشكل 1-3-3 يبين رسوم دوائر للتوزيع التكراري لمناطق كل من الطلاب والطالبات كل على رسم مستقل.

شكل 1-3-3
رسوم دائرية لتوزيع الطلاب والطالبات حسب المنطقة



مثال 2-3-1

الجدول 2-3-1 يبين كل من المبيعات والتكاليف والإرباح الشهرية (بآلاف الريالات) لإحدى المحلات التجارية الكبرى في المنطقة وذلك لفترة السنة الماضية، 1424هـ، قم بعرض البيانات بأسلوب يوضح أهم المعلومات فيها.

جدول 2-3-1: البيانات المالية الشهرية بآلاف الريالات لإحدى المحلات التجارية

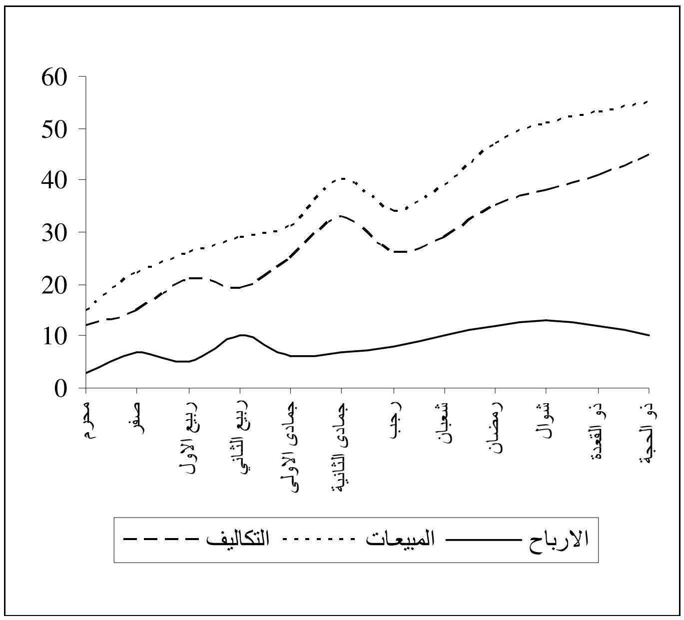
الشهر	المبيعات	التكاليف	الإرباح
محرم	15	12	3
صفر	22	15	7
ربيع الأول	26	21	5
ربيع الثاني	29	19	10
جمادى الأولى	31	25	6
جمادى الثانية	40	33	7
رجب	34	26	8
شعبان	39	29	10
رمضان	47	35	12
شوال	51	38	13
ذو القعدة	53	41	12
ذو الحجة	55	45	10

الحــل

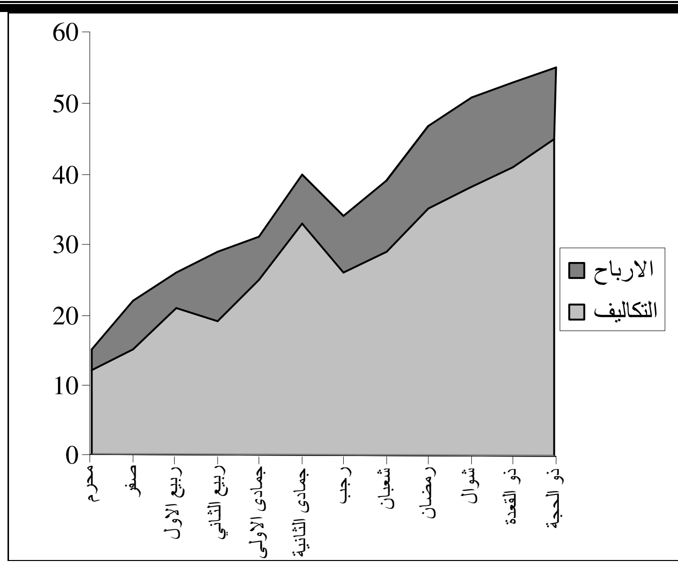
بما أن البيانات ذات العلاقة مرتبطة بسلسلة زمنيــة هــي الأشــهر، لــذا فــان الأســلوب الأفضــل هنــا يصبح الرسم البياني عن طريق المنحنى. يتم في رسوم المنحنيات تمثيـل قـيم المتغيـر المتتاليـة زمنيـا بخط يبين كل من قيمة المتغير في فترة محددة بالإضافة إلى نهـج التغيـر الحاصـل فـي بـين الفتـرات الزمنية في البيانات. الشكل 1-3-4 يبين المنحنيات البيانيــة لكـل مـن التكـاليف والمبيعـات والإربــاح الشهرية، حيث يتضح التزايد في كل من المبيعات والتكــاليف والإربــاح مــع الــزمن ممــا يشــير إلــى وجود علاقة طردية بين الزمن ومتغيرات الدراسة.

شكــل 1-3-4

البيانــات الماليــة الشــهرية بــآلاف الريالات لإحــدى المحلات التجارية



وباستخدام منحنيات متراكبة يــتم عــرض كــل مــن التكــاليف والإربــاح ليمثــل مجمــوع المبيعــات، وبالطبع قد يكون لمثل هذا الأسلوب اثر فنــي علــى شــكل المعلومــة المنقولــة والمقدمــة فــي الرســم البياني مما يؤدى في اغلب الأحيــان إلــى جــذب انتبــاه المشــاهد ومــن ثــم إلــى إيصــال المعلومــة المطلوبة. الشــكل 1-3-5 يبــين المنحنيــات البيانيــة لكــل مــن التكــاليف والإربــاح الشــهرية للمحــل التجاري.



شكل 1-3-5 البيانات المالية الشهرية بآلاف الريالات لإحدى المحلات التجارية

مثال 1-3-3

الرسوم البيانية التالية (الشكل 1-3-6، الشكل 1-3-7) تمثل كل من عدد السيارات المباعة حسب مصدر صنع السيارة للسنوات 1415–1420هـ و نسبة السيارات المباعة المستعملة و الجديدة إلى إجمالي السيارات المباعة في نفس السنوات،

استخدم الرسوم البيانات للإجابة على الأسئلة التالية:

◊ ما نوع هذه الرسوم؟

◊ تقريبا كم عدد السيارات المباعة في عام 1417هـ؟

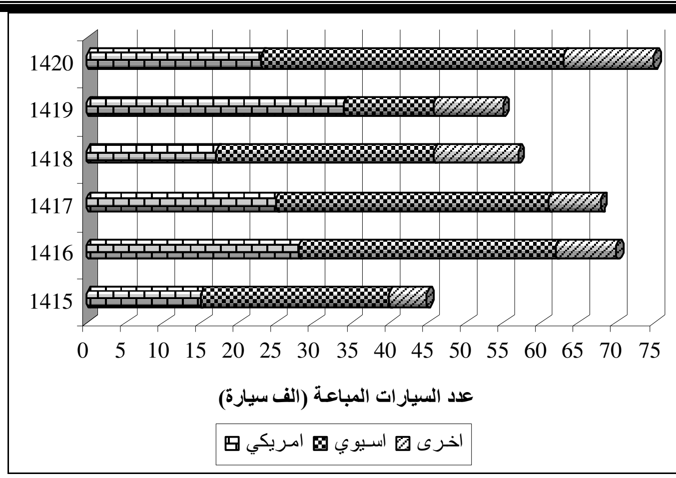
◊ تقريبا كم عدد السيارات المباعة من النوع الأمريكي والأسيوي خلال السنوات 1418–1420هـ؟

◊ في أي سنة كانت مبيعات السيارات الأسيوية الأقل مقارنة بالمبيعات في السنوات الأخرى؟

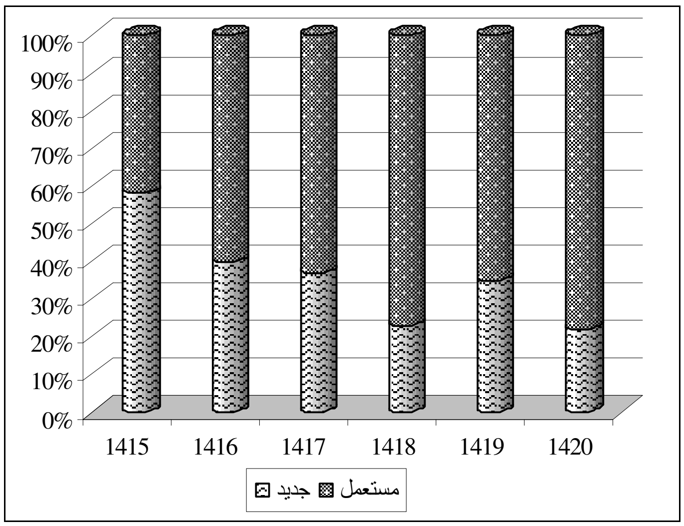
◊ في أي سنة زادت نسبة السيارات المباعة الجديدة عن المباعة المستعملة، وكم كانت النسبة؟

◊ تقريبا كم نسبة الزيادة في عدد السيارات الأمريكية المباعة سنة 1420 عنها في سنة 1415هـ؟

شكــل 1-3-6
التوزيع التكراري للسيارات المباعة في السنوات من 1415هـ إلى 1420هـ مصنفة حسب مصدر الصنع



شكــل 1-3-7
التوزيع النسبي للسيارات المباعة في السنوات من 1415هـ إلى 1420هـ مصنفة حسب حالة السيارة



الحــل

◊ تبعا لمعطيات السؤال وباستخدام الأشكال البيانية المرفقة يتبين التالي: يمثل الشكل 1-3-6 رسم بياني لأعمدة متراصة تمثل الثلاثة متغيرات وهي عدد السيارات الأمريكية المباعة وعدد السيارات الأسيوية المباعة وعدد السيارات الأخرى المباعة. كما يمثل الشكل 1-3-7 رسم بياني لأعمدة نسبية لمتغيرين هما عدد السيارات الجديدة المباعة وعدد السيارات المستعملة المباعة.

◊ وباستخدام الشكل 1-3-6 يتبين أن عدد السيارات المباعة في عام 1417هـ بلغ تقريبا 67 آلف سيارة.

◊ كذلك بلغ تقريبا عدد السيارات المباعة من النوع الأمريكي والأسيوي خلال السنوات 1418-1420هـ التالي: من النوع الأمريكي، بيع 16 آلف و 33 آلف و 22 آلف للأعوام 1418 و 1419 و 1420هـ على التوالي. أما من النوع الأسيوي فقد بيع 29 آلف و 12 ألف و 40 آلف للأعوام 1418 و 1419 و 1420هـ على التوالي.

◊ بلغت مبيعات السيارات الأسيوية في عام 1419هـ الأقل مقارنة بالمبيعات في السنوات الأخرى.

◊ زادت نسبة السيارات المباعة الجديدة عن المباعة المستعملة في العام 1415هـ، وكانت النسبة تقريبا 55%

◊ تقريبا، بلغت نسبة الزيادة في عدد السيارات الأمريكية المباعة سنة 1420 عنها في سنة 1415هـ 57% حيث تم الحصول عليها من خلال قسمة الفرق بين مبيعات السنتين على مبيعات السنة الأساس 14/(14–22)

4-1 مقاييس إحصائية

تتكون العملية الإحصائية من عدة خطوات يمثل فيها الوصف البياني الخطوة الأولى بينما تأتي عملية وصف البيانات كمياً كخطوة ثانية مهمة ومكملة للوصول إلى فهم أعمق ورؤية أوضح للمعلومة المحتواة 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.

بالنسبة للبيانات المبوبة والتي تكون موجودة في جدول تكراري مركب من عدة فئات يرافقها تكرارات، يتم حساب المنوال بإحدى طريقتين هما طريقة الفروق وطريقة الرافعة. تعتمد كلا الطريقتين على معطيات الجدول التكراري حيث يتم في البداية تحديد الفئة المنوالية. تعتبر الفئة المنوالية الفئة المقابلة لأكبر تكرار إذا كان الجدول منتظم (طول الفئة ثابت)، في حين يتم تحديد الفئة المنوالية من خلال التكرار المعدل الأعلى إذا كان الجدول غير منتظم. لذا فان الخطوة الأولى في عملية حساب المنوال لجدول تكراري تكمن في إيجاد التكرارات المعدلة إذا كان الجدول غير منتظم. ويتم حساب التكرارات المعدلة F^* من خلال الدالة الرياضية التالية:

$$F_i^* = \frac{F_i}{L_i} \qquad \forall \quad i = 1,2,....,K \tag{8}$$

حيث يتكون الجدول من K فئة، ويمثل الرمزان F_i و L_i تكرار الفئة i وطولها، على التوالي. تجدر الشارة إلى أن التكرار المعدل سوف يودي إلى نتيجة متطابقة مع النتيجة المحصلة من خلال استخدام التكرار الأصلي إذا كان الجدول التكراري المستخدم منتظما. ومن ثم فانه إذا كان الجدول التكراري غير منتظم بمعنى أن تكون إحدى أو بعض فئات الجدول لها طول مختلف عن باقي الفئات الأخرى فان التكرارات المعدلة تحل محل التكرارات الأصلية ويتم استخدامها في جميع مراحل إيجاد المنوال لبيانات مبوبة.

وبتحديد الفئة المنوالية والممثلة للفئة المقابلة لأكبر تكرار (أو تكرار معدل في حال الجداول غير المنتظمة) يتم الحصول على المعطيات المطلوبة لحساب المنوال وهي تكرار الفئة المنوالية، $F_\circ$، وتكرار الفئة السابقة للفئة المنوالية، $F_\leftarrow$، وتكرار الفئة اللاحقة، $F_\rightarrow$، والحد الأدنى للفئة المنوالية، $A_\circ$، وطول الفئة المنوالية، $L_\circ$. وباستخدام طريقة الفروق،

$$M = A_\circ + \frac{F_\circ - F_\leftarrow}{2F_\circ - F_\leftarrow - F_\rightarrow}(L_\circ) \tag{9}$$

يتم الحصول على تقدير لقيمة المنوال. كما يمكن استخدام طريقة الرافعة،

$$M = A_\circ + \frac{F_\rightarrow}{F_\leftarrow + F_\rightarrow}(L_\circ) \tag{10}$$

للحصول على تقدير لقيمة المنوال. وبالطبع قد تودي الطريقتين إلى تقديرين مختلفين، حيث يرجع الاختلاف إلى اختلاف الأسلوب المبني عليه فكرة التقدير والتي لا يتسع المجال هنا للخوض فيها.

المنوال: بيانات مبوبة

طريقة الفروق

$$M = A_\circ + \frac{F_\circ - F_\leftarrow}{2F_\circ - F_\leftarrow - F_\rightarrow}(L_\circ)$$

أو طريقة الرافعة

$$M = A_\circ + \frac{F_\rightarrow}{F_\leftarrow + F_\rightarrow}(L_\circ)$$

حيث: $F_\circ$ تكرار الفئة المنوالية

$F_{\leftarrow}$ التكرار السابق لتكرار الفئة المنوالية

$F_{\rightarrow}$ التكرار اللاحق لتكرار الفئة المنوالية

$A_{\circ}$ الحد الأدنى للفئة المنوالية

$L_{\circ}$ طول الفئة المنوالية

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 في حال البيانات الخام، أو إيجاد الجدول التكراري المتجمع الصاعد (أو الهابط) في حال البيانات المبوبة، وسيتم التركيز هنا على الترتيب التصاعدي في عملية إيجاد الوسيط.

يلي ترتيب القيم تحديد ترتيب الوسيط والذي يشير إلى ترتيب القيمة التي تمثل قيمة الوسيط المطلوبة. ويتم تحديد ترتيب الوسيط في حال البيانات الخام من خلال القيمة

$$\left({[n+1]}\big/{2} \right)$$

حيث تم استخدام n بهدف التبسيط للدلالة على حجم المجتمع أو حجم العينة. ومن ثم فان قيمة الوسيط تصبح

$$Q_2 = X_{\frac{n+1}{2}} \qquad (11)$$

إذا كان عدد قيم المتغير العشوائي مفردا. أما إذا كان عدد القيم زوجياً فانه لا يوجد قيمة وحيدة تقع في منتصف القيم، بل يتوفر قيمتين تقعا في نفس الوقت في منتصف القيم، ويمكن تحديد ترتيب القيمتين بالتالي:

$$\left(\frac{n}{2} \quad \& \quad \frac{n}{2}+1 \right)$$

وباستخدام الترتيبين يتم تحديد القيمتين الداخلتين في حساب قيمة الوسيط وهما،

$$\left(X_{\frac{n}{2}} \quad \& \quad X_{\frac{n}{2}+1} \right)$$

يلي ذلك استخدام تلك القيمتين وبالتحديد وسطهما الحسابي،

$$Q_2 = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \qquad (12)$$

للحصول على تقدير جيد لقيمة الوسيط المطلوبة.

لذا فان عملية حساب الوسيط لبيانات خام تمر بثلاث مراحل هي، مرحلة ترتيب القيم تصاعديا، ومرحلة تحديد ترتيب الوسيط، وأخيرا مرحلة تحديد قيمة الوسيط.

الوسيط: بيانات خام (غير مبوبة)

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

عندما يكون حجم العينة (عدد القيم) n فرديا،

$$Q_2 = X_{\frac{n+1}{2}}$$

عندما يكون حجم العينة n زوجيا،

$$Q_2 = \frac{X_{(n/2)} + X_{[(n/2)+1]}}{2}$$

حيث

X_i تمثل القيمة المرافقة للترتيب i بعد ترتيب قيم المتغير X تصاعديا.

عند التعامل مع بيانات مبوبة وبحيث يتعذر الحصول على البيانات الخام الأصلية فان عملية الحصول على تقدير لقيمة الوسيط ستختلف. في البداية يتم ترتيب البيانات تصاعديا وذلك من خلال إنشاء جدول تكراري متجمع صاعد. يلي ذلك كخطوة ثانية تحديد ترتيب الوسيط المتمثل في الدالة

$$\frac{1+\sum_{i=1}^{K} F_i}{2}$$

حيث يشير الرمز K إلى عدد الفئات في الجدول التكراري. وبالحصول على ترتيب الوسيط يتم تحديد موقع ذلك الترتيب في الجدول التكراري المتجمع الصاعد ليتم الحصول على المعطيات المطلوبة في دالة حساب الوسيط،

$$Q_2 = A_\circ + \frac{\left(\frac{1+\sum F}{2}\right) - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_\circ) \qquad (13)$$

حيث يشير الرمز $A_\circ$ إلى الحد الصاعد للفئة التي تسبق ترتيب الوسيط، والتي يطلق عليها بالفئة الوسيطية. كما يشير الرمز $L_\circ$ إلى طول الفئة الوسيطية. كذلك يشير كل من الرمز $CF_{\leftarrow}$ والرمز $CF_{\rightarrow}$ إلى التكرار الصاعد السابق لترتيب الوسيط والتكرار الصاعد اللاحق لترتيب الوسيط، على التوالي.

الوسيط: بيانات مبوبة

$$Q_2 = A_\circ + \frac{\left(\frac{1+\sum F}{2}\right) - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_\circ)$$

حيث: $CF_{\leftarrow}$ التكرار السابق لترتيب الوسيط

$CF_{\rightarrow}$ التكرار اللاحق لترتيب الوسيط

$A_\circ$ الحد الأدنى للفئة الوسيطية $L_\circ$ طول الفئة الوسيطية

الوسط الحسابي:

يمثل الوسط الحسابي أو المتوسط (Arithmetic Average or Mean) مقياس النزعة المركزية الأكثر شهرة والأكثر أهمية في المقاييس المختلفة. 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. وبافتراض التعامل مع متغير عشوائي X لمجتمع محدود حجمه N فانه يمكن حساب قيمة الوسط الحسابي من خلال الدالة التالية (لبيانات خام أو غير مبوبة)،

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N} \tag{14}$$

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

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{15}$$

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.

عندما لا تتوفر قيم المتغير العشوائي محل الدراسة بشكل خام فانه لا يمكن الحصول على القيمة الحقيقية لمعلمة المجتمع μ. ولكن يمكن الحصول على تقدير لقيمة متوسط المجتمع إذا كانت البيانات متوفرة بشكل جدول تكراري (جدول 1-2-7)، حيث يتم تقدير قيمة متوسط المجتمع μ بأسلوب مختلف عنه في حال التعامل مع بيانات خام. وتتم عملية التقدير من خلال استخدام مراكز الفئات كقيم ممثلة للقيم المحتواة في الفئات المحددة. ويتم حساب مركز الفئة من خلال الدالة

$$X_i = \frac{a_i + a_{i+1}}{2} \quad \forall \; i = 1,2,...,K \tag{16}$$

حيث تم الرمز للحد الأدنى للفئة i بالرمز a_i بينما مثل الرمز a_{i+1} الحد الأعلى للفئة.

الوسط الحسابي: بيانات خام (غير مبوبة)

الوسط الحسابي هو القيمة التي تتمركز حولها القيم.

لمجتمع محدود بحجم N،

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

لعينة عشوائية بحجم n

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

وبضرب مراكز الفئات بتكراراتها يتم الحصول على تقدير لمجموع القيم الداخلة في دراسة المتغير العشوائي X، حيث يتم قسمة المجموع على عدد القيم المساوي لمجموع التكرارات للحصول على التقدير المطلوب للوسط الحسابي،

$$\bar{X} = \frac{\sum_{i=1}^{K} X_i F_i}{\sum_{i=1}^{K} F_i} \qquad (17)$$

الجدول 1-4-1-1 يبين الهيكل المطلوب لتقدير قيمة متوسط مجتمع لمتغير عشوائي X من خلال استخدام البيانات المتوفرة في جدول تكراري (بيانات مبوبة).

جدول 1-4-1-1: هيكل حساب المتوسط لبيانات مبوبة

حدود الفئات	F	X	XF
a_1 - a_2	F_1	X_1	X_1F_1
a_2 - a_3	F_2	X_2	X_2F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_K - a_{K+1}	F_K	X_K	X_KF_K
المجموع	$\sum_{i=1}^{K} F_i$		$\sum_{i=1}^{K} X_i F_i$

الوسط الحسابي: بيانات مبوبة

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

حيث:

X_i مركز الفئة i

F_i تكرار الفئة i

مثال 1-4-1-1

الجدول 1-4-1-2 يضم عدد الموظفين لجميع فروع شركة كبيرة والبالغ عددها 22 فرع منتشرة في جميع أنحاء المملكة العربية السعودية. أوجد كل من المنوال لعدد الموظفين في فروع الشركة، والوسيط والوسط الحسابي.

جدول 1-4-1-2: 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

الحــل

لإيجاد قيمة المنوال المطلوبة يتم أولا حساب عــدد مــرات ظهــور القــيم المختلفــة للمتغيــر العشــوائي الممثل لعدد الموظفين في فروع الشركة، الجــدول 1-4-1-3 يبــين التوزيــع التكــراري لفــروع الشــركة 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 تكــرارا لــذا فــان الــرقم 8 يصــبح المنــوال مشــير إلـــى أن عــدد الموظفين 8 هو العدد الأكثر تكرارا في فروع الشركة (تكرر 5 مرات)،

$$M = 8$$

لحساب الوسيط يتم أولا ترتيب القيم تصاعديا كما هو مبين في الجدول 1-4-1-4.

جدول 1-4-1-4: ترتيب عدد الموظفين في فروع الشركة تصاعديا

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

يلي ذلك تحديد ترتيب الوسيط، حيث يتوفر هنا ترتيبين لان عدد القيم زوجي،

$$\left(\frac{N}{2} = \frac{22}{2} = 11 \quad \& \quad \frac{N}{2} + 1 = 11 + 1 = 12\right)$$

لذا فان قيمة الوسيط، بافتراض استخدام الرمز X للإشارة إلى عدد الموظفين في الفرع، تصبح،

$$Q_2 = \frac{X_{11} + X_{12}}{2}$$
$$= \frac{9 + 11}{2} = 10$$

أي أن نصف فروع الشركة لديها موظفين اقل من 10 بينما النصف الأخر لديه موظفين أكثر.

بالنسبة إلى حساب قيمة الوسط الحسابي لعــدد المــوظفين فــي فــروع الشــركة، تجــدر الإشــارة فــي البداية إلى أن القيم المحصلة تمثل مجتمع الدراســة حيــث احتــوت جميــع فــروع الشــركة، لــذا فــان المطلوب هنا الحصول على قيمة معلمة المجتمــع μ_X الممثلــة لمتوســط عــدد المــوظفين فــي الفــرع الواحد للشركة. وبتطبيق المعادلة (14) يتم الحصول على المطلوب،

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

مثال 1-4-1-2

الجدول 1-4-1-5 يبين قيم تمثل عدد السيارات المباعة في اليوم الواحد لعينة عشوائية حجمها 13 معرض سيارات،

جدول 1-4-1-5: عدد السيارات المباعة في اليوم الواحد

10	39	14	26	35	18	25	29	23	15	19	22	37

أوجد كل من المنوال والوسيط والوسط الحسابي بعدد السيارات المباعة باليوم الواحد في معارض السيارات.

الحــل

بما أن القيم المدرجة في الجدول 1-4-1-4 لا تحتوي على أي تكرار لذا فانه لا يوجد منوال للبيانات المتوفرة. أما بالنسبة للوسيط فان الخطوة الأولى تتمثل في ترتيب القيم تصاعديا، كما هو مبين في الجدول 1-4-1-5

جدول 1-4-1-6: عدد السيارات المباعة في اليوم الواحد مرتبة تصاعديا

10	14	15	18	19	22	23	25	26	29	35	37	39

وحيث أن عدد القيم فردي فان ترتيب الوسيط يصبح

$$\frac{n+1}{2} = \frac{13+1}{2} = \frac{14}{2} = 7$$

أي أن الوسيط يساوي القيمة السابعة في القيم المرتبة،

$$Q_2 = X_7 = 23$$

وذلك باستخدام الرمز X للإشارة إلى عدد السيارات المباعة في اليوم الواحد. وعليه فان قيمة الوسيط تفيد بأن نصف معارض السيارات يبيع اقل من 23 سيارة في اليوم الواحد.

بالنسبة لإيجاد قيمة الوسط الحسابي، تجدر الإشارة في البداية إلى أن القيم المعطاة في السؤال تمثل عينة عشوائية من مجتمع جميع معارض السيارات، لذا فان المطلوب هنا يصبح الحصول على تقدير لمتوسط المجتمع وذلك عن طريق حساب متوسط العينة،

$$\bar{X} = \frac{\sum_{i=1}^{13} X_i}{13} = \frac{312}{13} = 24$$

والذي يفيد بان عدد السيارات المباعة في اليوم الواحد للمعــرض الواحــد فــي المتوســط مســاوي لـــ 24 سيارة.

مثال 1-4-1-3

الجدول 1-4-1-7 يمثل التوزيع التكراري لعينـــة عشـــوائية حجمهـــا 850 رحلـــة بحريـــة ســـياحية حســـب وقت الرحلة بالساعة،

جدول 1-4-1-7: التوزيع التكراري لــ 850 رحلة بحرية سياحية حسب الوقت بالساعة

حدود الفئات (ساعة)	F
0 -	60
10 -	160
30 -	240
50 -	270
80 - 100	120
$\sum F$	850

أوجد كل من المنوال بطريقة الفروق وبطريقـــة الرافعـــة والوســـيط والوســـط الحســـابي لوقـــت الرحلـــة البحرية.

الحـــل

لحساب المنوال يتم في البداية إيجاد التكرار المعدل F^* وذلـــك بســـبب كـــون الجـــدول التكـــراري غيـــر منتظم، ويتم الحصول على التكرار المعدل من خلال قســـمة التكـــرار العـــادي F علـــى طـــول الفئـــة. الجدول 1-4-1-8 يبين التكرارات المعدلة المطلوبة،

جدول 1-4-1-8: التكرارات المعدلة

حدود الفئات (ساعة)	F	F^*
0 -	60	6
10 -	160	8
30 -	240	12
50 -	270	9
80 - 100	120	6
$\sum F$	850	

ومن خلال التكرارات المعدلة يتم تحديد الفئـــة المنواليـــة وهـــي الفئـــة الثالثـــة. لاحـــظ هنـــا انـــه ســـيتم التعامل مع الفئة الرابعة باعتبارها الفئة المنوالية في حـــال الاعتمـــاد علـــى التكـــرارات العاديـــة لتحديـــد الفئة المنوالية، وذلك بلا شك سيؤدي إلى نتائج خاطئة في النهاية.

ومن خلال المعادلة (9) وبتحديد الفئة الثالثة كفئة منوالية يتم الحصول على قيمة المنوال بطريقة الفروق، حيث يمثل تكرار الفئة المنوالية القيمة $F_{\circ} = 12$ بينما يمثل التكرار اللاحق والسابق القيمتين

$$F_{\leftarrow} = 8 \quad \& \quad F_{\rightarrow} = 9$$

كما تحدده بداية الفئة المنوالية بالرقم 30 ومن ثم أصبح $L = 20$. وباستخدام المعطيات السابقة يتم الحصول على قيمة المنوال بطريقة الفروق،

$$\begin{aligned} M &= A_{\circ} + \frac{F_{\circ} - F_{\leftarrow}}{2F_{\circ} - F_{\leftarrow} - F_{\rightarrow}}(L_{\circ}) \\ &= 30 + \frac{12-8}{2(12)-8-9}(20) = 41.4 \end{aligned}$$

كما يمكن استخدام المعادلة (10) للحصول على المنوال بطريقة الرافعة

$$\begin{aligned} M &= A_{\circ} + \frac{F_{\rightarrow}}{F_{\leftarrow} + F_{\rightarrow}}(L_{\circ}) \\ &= 30 + \frac{9}{8+9}(20) = 40.6 \end{aligned}$$

وعليه فانه يمكن القول بان المدة 40.6 أو 41.4 ساعة تمثل الوقت الأكثر حدوثا في الرحلات البحرية السياحية.

لحساب قيمة الوسيط يتم في البداية ترتيب القيم تصاعديا، ويتم ذلك من خلال إيجاد الجدول التكراري المتجمع الصاعد (جدول 1-4-1-9).

جدول 1-4-1-9: الجدول التكراري المتجمع الصاعد

حدود صاعدة	CF
< 0	0
< 10	60
< 30	220
< 50	460
< 80	730
< 100	850

وبتحديد ترتيب الوسيط الناتج من خلال قسمة مجموع التكرارات على 2،

$$\frac{1+\sum F}{2} = \frac{851}{2} = 425.5$$

يتم تحديد موقع الترتيب في الجدول التكراري الصاعد ليتم بعد ذلك تحديد المعطيات المطلوبة لحساب قيمة الوسيط والتي تساوي

$$\begin{aligned} Q_2 &= A_{\circ} + \frac{\left(\frac{1+\sum F}{2}\right) - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_{\circ}) \\ &= 30 + \frac{(425.5)-220}{460-220}(20) = 47.1 \end{aligned}$$

ساعة. وعليه فان نصف الرحلات البحرية السياحية تستغرق اقل من 47.1 ساعة.
أما بالنسبة لحساب قيمة الوسط الحسابي لوقت الرحلة البحرية السياحية بالساعة، فانه يمكن استخدام المجاميع المعطاة في الجدول 1-4-1-10

جدول 1-4-1-10: جدول حساب متوسط وقت الرحلة البحرية

حدود الفئات (ساعة)	F	X	XF
0 -	60	5	300
10 -	160	20	3200
30 -	240	40	9600
50 -	270	65	17550
80 - 100	120	90	10800
$\sum$	850		41450

بالإضافة إلى المعادلة (17)، ليتم الحصول على القيمة

$$\bar{X} = \frac{\sum_{i=1}^{K} X_i F_i}{\sum_{i=1}^{K} F_i}$$
$$= \frac{41450}{850} = 48.8$$

والتي تشير إلى أن متوسط وقت الرحلات البحرية السياحية يبلغ 48.8 ساعة.

1-4-2 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 الربيعي:

يمثل الانحراف الربيعي (Quartile Deviation) احد مقاييس التشتت المهمة ويرمز له بالرمز Q. ويمكن الاعتماد على الانحراف الربيعي كمقياس للتشتت إذا كان من بين قيم المتغير العشوائي قيماً كبيرة جدا أو صغيرة جدا عند مقارنتها بباقي القيم، أي قيم متطرفة. لذا فان الانحراف الربيعي والذي يعتمد على ترتيب القيم لا يتأثر بالقيم المتطرفة.

ويعتمد الانحراف الربيعي على ترتيب القيم وبالتحديد على كل من قيمة الربيع الأول (Q_1) وقيمة الربيع الثالث (Q_3)، حيث يتم حساب قيمة الانحراف الربيعي من خلال المعادلة

$$Q = \frac{Q_3 - Q_1}{2} \qquad (18)$$

حيث تشير قيمة الربيع الأول إلى القيمة التي يكون ربع القيم اقل منها في حين يكون ثلاث أرباع القيم أعلى منها. كما تشير قيمة الربيع الثالث إلى القيمة التي تكون ثلاث أرباع القيم اقل منها بينما ربع القيم يكون أعلى منها.

تختلف طريقة الحصول على تقدير لقيمة كل من الربيع الأول والربيع الثالث تبعاً لنوع البيانات، خام أم مبوبة. وفي حال كانت البيانات خام فان الأسلوب هنا يشابه أسلوب حساب الوسيط المبين في السابق، فيتم تقدير قيمة الربيع الأول والثالث من خلال ثلاث خطوات. تمثل الخطوة الأولى ترتيب جميع قيم المتغير العشوائي المستخدمة في الدراسة تصاعديا، يلي ذلك كخطوة ثانية تحديد ترتيب كل من الربيع الأول والربيع الثالث، ومن ثم يتم في الخطوة الأخيرة الحصول على تقدير لقيمتهما. عند التعامل مع بيانات خام، فانه يتحدد ترتيب الربيع الأول من خلال العلاقة الرياضية التالية:

$$\frac{(n+1)}{4}$$

كما يتحدد ترتيب الربيع الثالث من خلال العلاقة الرياضية التالية:

$$\frac{3(n+1)}{4}$$

وبتحديد ترتيب الربيع الأول والربيع الثالث يتم الحصول على تقدير لقيمتيهما من خلال اعتماد القيمة المرافقة لترتيب الربيع المطلوب كتقدير لقيمة الربيع، أي

$$Q_1 = X_{\frac{n+1}{4}} \tag{19}$$

و

$$Q_3 = X_{\frac{3(n+1)}{4}} \tag{20}$$

عندما يكون الترتيب عدداً صحيحاً فانه توجد قيمة تتوافق مع الترتيب، بينما لا توجد قيمة محددة تتوافق مع ترتيب لعدد كسري. لذا فانه كأسلوب تقريبي يمكن استخدام الوسط الحسابي للقيمتين التي يقع بينهما الترتيب المساوي لعدد كسري وذلك كتقدير تقريبي لقيمة الربيع المطلوبة. فمثلا عندما يكون الترتيب مساوي لـ 21 فان القيمة المطلوبة تصبح X_{21}، أي القيمة التي يكون ترتيبها الواحد والعشرون من بين القيم المرتبة تصاعديا. ولكن عندما يكون الترتيب مساوي، مثلا، لـ 35.7 فان القيمة المطلوب تصبح عبارة عن قيمة متوسط القيمتين الخامسة والثلاثون والسادسة والثلاثون من بين القيم المرتبة تصاعديا،

$$X_{35.7} = \frac{X_{35} + X_{36}}{2}$$

الانحراف الربيعي: بيانات خام (غير مبوبة)

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

حيث

Q_1 قيمة الربيع الأول:

$$Q_1 = X_{\frac{n+1}{4}}$$

Q_3 قيمة الربيع الثالث:

$$Q_3 = X_{\frac{3(n+1)}{4}}$$

يتم حساب الانحراف الربيعي لبيانات مبوبة من خلال الدالة (18)، ولكن باختلاف أسلوب عملية حساب كل من الربيع الأول والربيع الثالث. يتم حساب الربيع الأول الثالث لبيانات مبوبة بأسلوب مشابه لحساب الوسيط لبيانات مبوبة، حيث يتم استخدام ترتيب الربيع لتحديد معطيات دالة حساب الربيع. يتم حساب ترتيب الربيع الأول من خلال الدالة

$$\frac{\left(1+\sum F\right)}{4}$$

كما يتم حساب ترتيب الربيع الثالث باستخدام الدالة

$$\frac{3\left(1+\sum F\right)}{4}$$

وباستخدام الجدول التكراري المتجمع الصاعد يتم تحديد موقع ترتيب الربيع ومن ثم يتم تحديد كل من الحد الأدنى للفئة الربيعية $A_{\circ}$، وطول الفئة الربيعية $L_{\circ}$، والتكرار المتجمع الصاعد السابق $CF_{\leftarrow}$ بالإضافة إلى والتكرار المتجمع الصاعد اللاحق $CF_{\rightarrow}$.

تأخذ معادلة الربيع الأول الشكل التالي،

$$Q_1 = A_{\circ} + \frac{\frac{\left(1+\sum F\right)}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}\left(L_{\circ}\right) \qquad (21)$$

بينما يتم حساب الربيع الثالث باستخدام الدالة

$$Q_3 = A_{\circ} + \frac{\frac{3\left(1+\sum F\right)}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}\left(L_{\circ}\right) \qquad (22)$$

الانحراف الربيعي: بيانات مبوبة

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

حيث

$$Q_1 = A_{\circ} + \frac{\frac{1+\sum F}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}\left(L_{\circ}\right)$$

$$Q_3 = A_\circ + \frac{\frac{3\left(1+\sum F\right)}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}\left(L_\circ\right)$$

الانحراف المتوسط:

يمثل المقياس الثاني من مقاييس التشتت والمطروح في هذا الفصل الانحراف المتوسط، Mean Deviation، حيث يأخذ الرمز MD. وتستند فكرة الانحراف المتوسط على قياس متوسط الانحرافات المطلقة (بدون إشارة) للقيم عن وسطها الحسابي. وفي الواقع تم الاعتماد على القيمة المطلقة لتلافي الحصول على قيمة ثابتة مساوية للصفر والناتجة من الحقيقة بأن إجمالي الفروقات لقيم المتغير الكمي عن وسطها الحسابي سيساوي دوما الصفر،

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = 0$$

ونتيجة لاستخدام القيم المطلقة فقط فإن قيمة الانحراف المتوسط ستكون دوما موجبة، حيث تشير القيم الكبيرة إلى تشتت اكبر وكلما اتجه المجموع إلى الصفر كلما قل تشتت قيم المتغير الكمي محل الدراسة.

وبافتراض أن المتغير العشوائي محل الدراسة يأخذ الرمز X ومتوفر في شكل بيانات خام تمثل جميع القيم في مجتمع الدراسة المحدود الحجم، N، فانه يمكن حساب قيمة الانحراف المتوسط، حيث يتم كخطوة أولى حساب متوسط المجتمع،

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

يلي ذلك إيجاد مجموع القيم المطلقة لانحرافات القيم عن الوسط الحسابي، ومن ثم قسمة المجموع المحصل على عدد القيم، N، لينتج الانحراف المتوسط للمجتمع MD_P،

$$MD_P = \frac{\sum_{i=1}^{N}\left|X_i - \mu_X\right|}{N} \tag{23}$$

حيث يشير الرمز | | إلى القيمة المطلقة.

في الجهة الأخرى عندما يتعذر تغطية جميع قيم مجتمع الدراسة لكبر حجمه أو عدم محدوديته، فانه يمكن التعامل مع عينة عشوائية بحجم n مسحوبة من مجتمع الدراسة وتكون ممثلة له. وبتوفر قيم المتغير العشوائي في العينة يتم تقدير قيمة الانحراف المتوسط للعينة، MD، وذلك من خلال قسمة مجموع انحرافات قيم العينة العشوائية عن وسطها الحسابي، $\bar{X}$ ، على عددها n،

$$MD = \frac{\sum_{i=1}^{n}\left|X_i - \bar{X}\right|}{n} \tag{24}$$

الانحراف المتوسط: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

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

لعينة عشوائية بحجم n

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

حيث

μ_X متوسط المجتمع $\bar{X}$ متوسط العينة

يمكن كذلك حساب الانحراف المتوسط لبيانات مبوبة وذلك من خلال تفعيل السياسة السابقة المتبعة في عملية تقدير الانحراف المتوسط لبيانات غير مبوبة. ويتم التعامل مع البيانات المبوبة من خلال جدول تكراري يمثل قيم المتغير العشوائي. وبحكم كون الجدول التكراري مرتبط بفئات مما يجعل القيم الحقيقة مجهولة فانه لا يمكن التعامل مع معلمة المجتمع المجهولة، لذا فان البيانات المبوبة تؤدي فقط إلى تقدير لقيمة معلمة المجتمع. وبالتعامل مع تكرارات الفئات ومراكزها الممثلة للقيم فيها يتم تقدير قيمة الانحراف المتوسط من خلال الدالة التالية،

$$MD = \frac{\sum_{i=1}^{K} \left(|X_i - \bar{X}| F_i \right)}{\sum_{i=1}^{K} F_i} \tag{25}$$

حيث تم افتراض توفر عدد مساوي لتكرار الفئة المحددة من قيمة مركزها، ومن ثم يتم ضرب أو ما نسميه ترجيح العمليات الحسابية على مراكز الفئات بتكراراتها.

الانحراف المتوسط: بيانات مبوبة

$$MD = \frac{\sum_{i=1}^{K} \left(|X_i - \bar{X}| F_i \right)}{\sum_{i=1}^{K} F_i}$$

حيث

K عدد الفئات

$\bar{X}$ متوسط العينة

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،

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

حيث يمكن تبسيطها رياضيا لتصبح،

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_X^2} \qquad (26)$$

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

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

وبالتبسيط الرياضي تصبح،

$$S_X = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}} \qquad (27)$$

الانحراف المعياري: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

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

لعينة عشوائية بحجم n

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

كذلك يتم الحصول على التباين للمجتمع وللعينة من خلال تربيع الانحراف المعياري المحصل في المعادلتين (26) و (27)، على التوالي،

$$\sigma_X^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_X^2 \tag{28}$$

$$S_X^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} \tag{29}$$

التباين: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

$$\sigma_X^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_X^2$$

لعينة عشوائية بحجم n

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

وكما هو الحال في عملية تقدير الوسط الحسابي لبيانات مبوبة (جدول تكراري)، يتم حساب الانحراف المعياري و التباين لبيانات مبوبة من خلال إيجاد مراكز الفئات والتي يتم معاملتها كتقدير لقيم المتغير الكمي المجهولة. ومن ثم يتم ترجيح تلك القيم بتكرار الفئات بالإضافة إلى معاملة مجموع التكرارات معاملة حجم العينة في البيانات غير المبوبة. تجدر الإشارة إلى انه لا يمكن حساب قيمة معلمة المجتمع الحقيقية σ أو σ^2 من خلال البيانات المبوبة، ويجع السبب في ذلك إلى أن القيم الفعلية للمتغير العشوائي غير معلومة في الأصل، ويتم تقديرها من خلال مراكز الفئات. لذا فان صيغة حساب الانحراف المعياري لبيانات مبوبة تصبح،

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

وبالتبسيط الرياضي تنتج الصيغة الرياضية التالية:

$$S_X = \sqrt{\frac{\sum_{i=1}^{K}\left(F_i X_i^2\right) - \left(\sum_{i=1}^{K} F_i\right)\bar{X}^2}{\left(\sum_{i=1}^{K} F_i\right) - 1}} \tag{30}$$

الانحراف المعياري: بيانات مبوبة

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

حيث

K عدد الفئات

كذلك يتم الحصول على التباين من خلال تربيع الانحراف المعياري المحصل،

$$S_X^2 = \frac{\sum_{i=1}^{K}\left(F_i X_i^2\right) - \left(\sum_{i=1}^{K} F_i\right)\bar{X}^2}{\left(\sum_{i=1}^{K} F_i\right) - 1} \qquad (31)$$

التباين: بيانات مبوبة

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

حيث

K عدد الفئات

مثال 1-2-4-1

بالرجوع إلى بيانات المثال 1-1-4-1 أوجد كل من الانحراف الربيعي والانحراف المتوسط والانحراف المعياري والتباين لعدد الموظفين في فروع الشركة.

الحــل

بما أن بيانات المتغير العشوائي الممثل لعدد الموظفين والمبينة في الجدول 1-4-1-2 تغطي جميع فروع الشركة، لذا فان القيم تمثل مجتمع دراسة. كذلك يتضح من صفة البيانات بأنها بيانات خام، مما يحتم استخدام طريقة محددة لحساب معالم المجتمع المطلوبة.

لحساب الانحراف الربيعي يتم أولا حساب الربيع الأول، Q_1، والربيع الثالث، Q_3، ويتم ذلك من خلال ترتيب القيم تصاعديا يلي ذلك تحديد قيمة كل من الربيع الأول والربيع الثالث. وباستخدام القيم بعد ترتيبها (جدول 1-4-1-4) يتم إيجاد القيم المطلوبة،

$$Q_1 = X_{\frac{N+1}{4}}$$

$$= X_{\frac{22+1}{4}} = X_{5.75}$$

$$= \frac{X_5 + X_6}{2} = \frac{8+8}{2} = 8$$

و

$$Q_3 = X_{\frac{3(N+1)}{4}}$$
$$= X_{\frac{3(22+1)}{4}} = X_{17.25}$$
$$= \frac{X_{17} + X_{18}}{2} = \frac{12+14}{2} = 13$$

لذا فان قيمة الانحراف المعياري تساوي،

$$Q = \frac{Q_3 - Q_1}{2}$$
$$= \frac{13-8}{2} = 2.5$$

لحساب قيمة الانحراف المتوسط يتم أولا إيجاد قيمة الوسط الحسابي

$$\mu_X = \frac{\sum X}{N} = \frac{232}{22} = 10.545$$

وبطرح قيمة المتوسط من كل قيمة للمتغير العشوائي يتم الحصول على مجموع الانحرافات،

$$\sum_{i=1}^{22} |X_i - \mu_X| = |8-10.545| + |11-10.545| + ...$$
$$.... + |11-10.545| + |3-10.545| = 70$$

وعليه فان قيمة الانحراف المتوسط تصبح

$$MD_p = \frac{\sum_{i=1}^{N} |X_i - \mu_x|}{N} = \frac{70}{22} = 3.182$$

يتم كذلك استخدام قيمة المتوسط لحساب كل من الانحـراف المعيـاري والتبـاين، ويـتم ذلـك بحسـاب مجموع مربعات انحرافات القيم عن وسطها الحسابي،

$$\sum_{i=1}^{22} (X_i - \mu_X)^2 = (8-10.545)^2 + (11-10.545)^2 + ...$$
$$.... + (11-10.545)^2 + (3-10.545)^2 = 349.4545$$

وعليه فان قيمة كل من الانحراف المعيـاري والتبـاين لمتغيـر عـدد المـوظفين فـي فـروع الشـركة تصبح،

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

و

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

مثال 1-4-2-2

بالرجوع إلى بيانــات المثــال 1-4-1-2 أوجــد كــل مــن الانحــراف الربيعــي والانحــراف المتوســط والانحراف المعياري والتبــاين لعــدد الســيارات المباعــة فــي اليــوم الواحــد للعينــة العشــوائية مــن المعارض.

الحــل

في البداية يجب التنبيه إلى أن البيانات محل الدراسة تمثـل عينــة عشــوائية ولــيس مجتمعــاً، لــذا فــان أسلوب حساب بعض مقاييس التشتت تأخذ شكلاً مختلفاً عنه فــي حــال التعامــل مــع مجتمــع. وحيــث أن البيانات أتت بصفة البيانات الخام لذا فان الانحراف الربيعي يساوي

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

حيث

$$\begin{aligned} Q_1 &= X_{\frac{n+1}{4}} \\ &= X_{\frac{13+1}{4}} \\ &= X_{3.5} \\ &= \frac{X_3 + X_4}{2} \\ &= \frac{15+18}{2} = 16.5 \end{aligned}$$

و

$$\begin{aligned} Q_3 &= X_{\frac{3(n+1)}{4}} \\ &= X_{\frac{3(13+1)}{4}} \\ &= X_{10.5} \\ &= \frac{X_{10} + X_{11}}{2} \\ &= \frac{29+35}{2} = 32 \end{aligned}$$

وعليه

$$Q = \frac{32-16.5}{2} = 7.75$$

لحساب الانحراف المتوسط يتم أولا إيجاد الوسط الحسابي لبيانات العينة العشوائية،

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^{13} X_i}{13} \\ &= \frac{312}{13} = 24 \end{aligned}$$

ومن ثم نوجد مجموع القيم المطلقة لانحرافات القيم عن وسطها الحسابي،

$$\begin{aligned} \sum_{i=1}^{13} \left| X_i - \bar{X} \right| &= |10-24| + |14-24| + \ldots.. \\ &\ldots. + |37-24| + |39-24| = 94 \end{aligned}$$

وعليه فان قيمة الانحراف المتوسط تصبح

$$MD=\frac{\sum_{i=1}^{n}\left|X_i-\bar{X}\right|}{n}=\frac{94}{13}=7.231$$

وبحساب مجموع مربعات انحرافات القيم عن وسطها الحسابي،

$$\sum_{i=1}^{13}\left(X_i-\bar{X}\right)^2=(37-24)^2+(22-24)^2+....$$
$$....+(39-24)^2+(10-24)^2=988$$

يتم الحصول على قيمة كل من الانحراف المعياري والتباين لمتغير عدد السيارات المباعة،

$$S_X=\sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}}$$
$$=\sqrt{\frac{988}{13-1}}=\sqrt{82.33}=9.074$$

و

$$S_X^2=\frac{\sum\left(X_i-\bar{X}\right)^2}{n-1}$$
$$=\frac{988}{12}=82.33$$

مثال 1-4-2-3

بالرجوع إلى بيانات المثال 1-4-1-3 أوجد كل من الانحراف الربيعي والانحراف المتوسط والانحراف المعياري والتباين لبيانات الوقت المستغرق للرحلات البحرية السياحية والمبينة في الجدول التكراري 1-4-1-7.

الحــل

بهدف حساب الانحراف الربيعي يتم استخدام الجدول التكراري الصاعد (جدول 1-4-1-9) لحساب كل من الربيع الأول Q_1 والربيع الثالث Q_3.

وبتحديد ترتيب الربيع الأول،

$$\frac{\left(1+\sum F\right)}{4}=\frac{1+850}{4}=212.75$$

يتم استخدام المعادلة (21) للحصول على قيمة الربيع الأول،

$$Q_1=A_\circ+\frac{\frac{\left(1+\sum F\right)}{4}-CF_{\leftarrow}}{CF_{\rightarrow}-CF_{\leftarrow}}\left(L_\circ\right)$$
$$=10+\frac{212.75-60}{220-60}(20)=29.1$$

جدول 1-4-1-9: الجدول التكراري المتجمع الصاعد

حدود صاعدة	CF

< 0	0
< 10	60
< 30	220
< 50	460
< 80	730
< 100	850

كذلك بتحديد ترتيب الربيع الثالث،

$$\frac{3\left(1+\sum F\right)}{4}=\frac{3\left(1+850\right)}{4}=638.25$$

يتم استخدام المعادلة (22) للحصول على قيمة الربيع الثالث،

$$Q_3=A_\circ+\frac{\frac{3\left(1+\sum F\right)}{4}-CF_{\leftarrow}}{CF_{\rightarrow}-CF_{\leftarrow}}\left(L_\circ\right)$$

$$=50+\frac{638.25-460}{730-460}\left(30\right)=69.8$$

وعليه فان قيمة الانحراف الربيعي تصبح،

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

$$=\frac{69.8-29.1}{2}=20.35$$

لحساب قيمة الانحراف المتوسط يتم إنشاء جدول تكراراي يحتوي على المجاميع المطلوبة (الجدول 1-2-4-1). وباعتماد قيمة الوسط الحسابي للوقت المستغرق في الرحلات البحرية السياحية والمساوي لـــ

$$\bar{X}=\frac{\sum XF}{\sum F}=48.8$$

يتم الحصول على مجموع فروقات القيم عن الوسط الحسابي.

جدول 1-2-4-1: جدول حساب الانحراف المتوسط

حدود الفئات (ساعة)	F	X	XF	$\left(X-\bar{X}\right)$	$\left\|X-\bar{X}\right\|$	$\left\|X-\bar{X}\right\|F$
0 -	60	5	300	-43.8	43.8	2628
10 -	160	20	3200	-28.8	28.8	4608
30 -	240	40	9600	-8.8	8.8	2112
50 -	270	65	17550	16.2	16.2	4374
80 - 100	120	90	10800	41.2	41.2	4944
$\sum$	850		41450			18666

وعلية فان قيمة الانحراف المتوسط يتم حسابها بالتالي،

$$MD=\frac{\sum\left(\left|X-\bar{X}\right|F\right)}{\sum F}$$

$$= \frac{18666}{850} = 21.96$$

وبهدف إيجاد تقديرات كل من الانحراف المعياري والتباين لمتغير الوقت المستغرق في الرحلات البحرية السياحية، يتم إنشاء جدول يضم المجاميع المطلوبة لحساب كل من الانحراف المعياري و التباين (جدول 1-4-2-2).

جدول 1-4-2-2: جدول حساب الانحراف المعياري والتباين

حدود الفئات (ساعة)	F	X	XF	X^2F	$(X-\bar{X})^2$	$(X-\bar{X})^2F$
0 -	60	5	300	1500	1918.44	115106.4
10 -	160	20	3200	64000	829.44	132710.4
30 -	240	40	9600	384000	77.44	18585.6
50 -	270	65	17550	1140750	262.44	70858.8
80 - 100	120	90	10800	972000	1697.44	203692.8
$\sum$	850		41450	2562250		540954

وباعتماد قيمة المتوسط 48.8 يتم إيجاد مجموع مربعات انحرافات القيم عن وسطها الحسابي. وكما تم الإشارة إليه سابقا، فان عملية حساب كل من الانحراف المعياري والتباين يمكن أن تتم من خلال طريقتين، طريقة مختصرة

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

$$= \sqrt{\frac{2562250 - (850)\left(\frac{41450}{850}\right)^2}{850-1}}$$

$$= \sqrt{637.16} = 25.242$$

وطريقة مطولة،

$$S_X = \sqrt{\frac{\sum\left[(X-\bar{X})^2 F\right]}{\left(\sum F\right) - 1}}$$

$$= \sqrt{\frac{540954}{849}} = \sqrt{637.1661} = 25.242$$

والتي يفترض أن تعطي نتائج متماثلة قد تختلف قليلا بسبب التقريب في بعض الحالات.

وبتربيع الانحراف المعياري يتم الحصول على التباين،

$$S_X^2 = 637.16$$

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.

مثال 4-2-4-1

لدراسة 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

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

علق على النتائج، مع بيان أي المجمعتين الطلاب أم الطالبات أكثر تشتتا في المعدلات التراكمية.

الحـل

بمقارنة 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.

مثال 5-2-4-1

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

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

هل يمكن الإشارة إلى خلل في العملية التعليمة من خلال الإحصائيات السابقة.

الحــل

تبعا 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:

$$CV_p = \frac{\sigma_X}{\mu_X} \times 100 \tag{31}$$

حيث يمثل القيمة الحقيقية 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،

$$CV = \frac{S_X}{\bar{X}} \times 100 \qquad (32)$$

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

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

مثال 6-2-4-1

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:

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

إلى أي عامل يجب أن يوجه الاهتمام، مدة وصول الطلبة أم تكلفة الشحن، ليتم محاولة تقليل التكلفة بشكل عام، ولماذا؟

الحـل

تبعا 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

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

$$=\frac{\sqrt{10.5}}{7.8}\times 100=41.5\%$$

والتي تمثل تقدير لقيمة معامل الاختلاف لمتغير مدة وصول الطلبيات.

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

$$CV\ (Y\ )=\frac{S_Y}{\bar{Y}}\times 100$$

$$=\frac{\sqrt{250000}}{3200}\times 100=15.6\%$$

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.

5-1 الاحتمالات

تمثل الاحتمالات ركن أساسي في علم الإحصاء يدخل بشكل مباشر في كثير من النظريات والطرق الإحصائية المستخدمة في جميع جوانب الإحصاء بشكل عام. ويمكن القول أن ارتباط الإحصاء بالاحتمالات هي الصفة القوية التي تميز عام الإحصاء عن علم الرياضيات والمرتبط بحقائق وعلاقات رياضية حتمية على عكس علم الإحصاء المرتبط بنتائج احتمالية متعددة يطغى عليها في الغالب ظروف عدم التأكد. كما أن جانب الاستدلال الإحصائي الحديث والذي صاحب التطور التقني العالمي والتطور العلمي الحديث والمستخدم في جميع العلوم الأخرى سواء التجريدية أو التنظيرية، قد استفاد بشكل كبير من النظريات الاحتمالية ليخرج بوجه علمي جديد يجعله ركن أساسي في جميع العلوم دون تحديد.

يفترض في متلقي علم الإحصاء الإلمام بأساسيات الاحتمالات، كما ترتفع الجرعة المطلوبة من الاحتمالات كلما اتجه التركيز إلى الجانب الرياضي للأدوات الإحصائية و الأسس الرياضية لها، في حين أن عملية استخدام الإحصاء كأداة تطبيقية في علم ما يظل يتطلب حد أدنى من المعلومة الرياضية في جانب الاحتمالات والتي تعطي الباحث القدرة على معرفة الطرق المتعددة الممكنة لأداء الاستدلال الإحصائي. سيتم في هذا الجزء من الفصل التطرق إلى أساسيات الاحتمالات وذلك بشكل سريع دون الدخول بتفاصيل دقيقة تخرج عن نطاق الكتاب.

1-5-1 التجارب والحوادث

تنقسم التجارب من ناحية نوعية النتائج إلى قسمين أساسين. يمثل القسم الأول التجارب ذات الناتج الوحيد الثابت المتوفر من جراء تنفيذ التجربة في ظروف ومحيط ثابت، مثل تجربة إضافة الصوديوم إلى الماء فينتج عنها غاز الهيدروجين كناتج أكيد واقع دون محالة بإذن الله، أو تجربة خفض درجة حرارة ماء إلى الصفر فينتج عنها

تجمد الماء أو تحوله إلى الحالة الصلبة. وعليه فان احتمال وقوع الناتج في مثل هذه التجارب يكون دوما مساوياً للواحد الصحيح مما يدل على أن الناتج واقع باحتمال 100% دون شك.

يمثل القسم الثاني التجارب العشوائية والتي تمثل التجارب التي يختلف الناتج منها من مرة إلى أخرى ليتنوع من خلال عدد من النتائج الممكنة الحدوث مع إبقاء الظروف واحدة عند إجراء التجربة، وترتبط كلمة عشوائية بحقيقة تنوع الناتج في كل مرة تجرى فيها التجربة العشوائية دون إمكانية التدخل من قبل معد التجربة لفرض ناتج محدد بعينه. يطلق على النتائج الممكنة الحدوث لتجربة عشوائية بفراغ العينة، S، فمثلا عند إلقاء حجر نرد متزن فان النتائج الممكنة الحدوث هي ستة،

$$S = \{1,2,3,4,5,6\}$$

تحدث بشكل عشوائي دون إمكانية منا للتدخل لكي يحدث رقم محدد. ويطلق على النتائج المختلفة لفراغ العينة بعناصر الفراغ، E، حيث يرتبط كل عنصر باحتمال،

$$0 \le \Pr(E_i) \le 1 \quad \forall \quad i = 1,2,...,n \qquad (33)$$

حيث n تمثل عدد عناصر فراغ العينة. وبالطبع يمثل مجموع احتمالات حدوث عناصر فراغ العينة 100% كدلالة على حتمية حدوث إحدى العناصر عند إجراء التجربة،

$$\Pr(S) = \sum_{i=1}^{n} \Pr(E_i) = 1 \qquad (34)$$

عندما يكون لتجربة عشوائية عدد من النتائج الممكنة الحدوث (عناصر) فان الحدث A قد يمثل هنا وقوع إحدى النتائج أو وقوع مجموعة من النتائج، لذا فانه يمكن القول بان الحدث A يمثل فعلا مجموعة جزئية من المجموعة S الممثلة لفراغ العينة التابع للتجربة،

$$A \subseteq S \qquad (35)$$

فمثلاً في تجربة إلقاء حجر نرد يمكن تعريف الحدث A على انه الحصول على رقم زوجي، مما يحدد العناصر الداخلة في مجموعة الحدث لتكون العناصر من المجوعة S والمستوفية لشرط الحدث وهو أن يكون العنصر رقماً زوجياً،

$$A = \{2,4,6\}$$

كما يرتبط احتمال وقوع الحدث A بشرط الاحتمالات الأساسي وهو أن يكون محصور بين الصفر والواحد الصحيح،

$$0 \le \Pr(A) \le 1 \qquad (36)$$

2-5-1 احتمالات الحوادث

يتم تحديد قيمة الاحتمال المرافق للحدث A من خلال أحد أسلوبين. يمثل الأسلوب الأول الاحتمال النظري، بينما يطلق على الأسلوب الثاني بالاحتمال التجريبي. يعتمد الأسلوب النظري في إيجاد الاحتمالات على

حقيقة أساسية مفادها أن تكون جميع الاحتمالات المرافقة لجميع عناصر فراغ العينة معروفة ومحددة. وعليه، فان احتمال حدوث الحدث A سيساوي مجموع الاحتمالات المرافقة لجميع العناصر الموجودة في مجموعة الحدث،

$$\Pr(A) = \sum \Pr(E_j) \quad \forall \quad E_j \in A \tag{37}$$

والتي تمثل مجموعة جزئية من مجموعة فراغ العينة (المعادلة 1.35)، لذا

$$\Pr(A) \leq \Pr(S) = 1 \tag{38}$$

وفي الحالة الخاصة عندما تكون جميع عناصر فراغ العينة مرتبطة باحتمال متساوي، أي أن تكون لها نفس فرصة الظهور، فان احتمال حدوث الحدث A يكون،

$$\Pr(A) = \frac{n(A)}{n(S)} \tag{39}$$

حيث يمثل الرمز $n(A)$ عدد عناصر المجموعة A بينما تم الرمز لعدد عناصر فراغ العينة بالرمز $n(S)$.

ويتم حساب الاحتمال المكمل لاحتمال وقوع الحدث A من خلال إيجاد احتمال عدم وقوع الحدث، ويتم الرمز لمتمم الحدث A بالرمز A' والذي يضم جميع العناصر الأخرى غير المضمنة في الحدث A والموجودة في فراغ العينة S. ويتم حساب احتمال وقوع متمم الحدث A من خلال الدالة التالية:

$$\Pr(A') = 1 - \Pr(A) \tag{40}$$

حيث

$$\Pr(A) + \Pr(A') = \Pr(S) = 1 \tag{41}$$

مثال 1-2-5-1

افترض انه تم إلقاء حجر نرد متزن ذو ستة أوجه مرقمة من رقم 1 إلى رقم 6، وان الحدث A يمثل ظهور عدد زوجي. أوجد كل من فراغ العينة S واحتمال ظهور الحدث A.

الحـــل

بما أن حجر النرد متزن وله ستة وجوه مرقمة لذا فان فراغ العينة يتكون من الأرقام الموجودة في أوجه حجر النرد،

$$S = \{1,2,3,4,5,6\}$$

وحيث أن احتمال ظهور أي وجه من الوجوه الستة لحجر النرد ثابت ومساوي لسدس

$$\Pr(E) = \frac{1}{6} \quad \forall \quad E = 1,2,3,4,5,6$$

كما أن الحدث A يضم ثلاثة عناصر من أصل الست عناصر في فراغ العينة

$$A = \{2,4,6\}$$

لذا،

$$\Pr(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = 0.50$$

مثال 2-2-5-1

يتوفر لدى إحدى الشركات الكبيرة أربعة فروع يضم كل فرع عدد من الموظفين، ومن بين موظفي الفروع يتوفر 3 موظفين في الفرع الأول و 5 موظفين في الفرع الثاني و 4 موظفين في الفرع الثالث و 2 موظف في الفرع الرابع والأخير، جميعهم يمكن لهم العمل على نوع محدد من المعاملات الإدارية الهامة للشركة. وبهدف إشراك جميع الفروع في تحمل المسؤولية بالإضافة إلى مبدأ توزيع العمل فقد تم تبني آلية لإرسال المعاملات إلى موظفي الفروع بشكل عشوائي دون أن يكون لرقم الفرع أو الموظف دور في عملية التوزيع. إذا توفرت معاملة جديدة فأوجد احتمال أن يتم إرسالها إلى الفرع الثالث.

الحل

في البداية يتم تحديد فراغ العينة المتمثل في جميع الموظفين القادرين على أداء العمل المطلوب في الفروع المختلفة، أي أن فراغ العينة يتكون من 14 عنصر، وهو إجمالي عدد الموظفين في الفروع الأربعة،

$$S=\{1.1,1.2,1.3,2.1,2.2,2.3,2.4,2.5,3.1,3.2,3.3,3.4,4.1,4.2\}$$

حيث تم الرمز لعناصر فراغ العينة، E، والتي تمثل الموظفين برقم الفرع متبوعاً برقم الموظف في الفرع، فمثلاً الرقم 3.2 يشير إلى الموظف رقم 2 في الفرع الثالث. وحيث أن عملية توزيع العمل تتم بشكل عشوائي بحت فان جميع عناصر فراغ العينة متساوية الاحتمال وان فرصة ظهور أي منها ثابتة. وعليه فان احتمال ظهور أي عنصر من بين العناصر الأربعة عشر الموجودة في فراغ العينة يصبح

$$\Pr(E)=\frac{1}{14}$$

وتبعا لتعريف الحدث A في السؤال والمتمثل في أن يكون الموظف المكلف يعمل في الفرع الثالث، فان عناصر الحدث A والتي عددها 4 هي،

$$A=\{3.1,3.2,3.3,3.4\}$$

لذا فان الاحتمال المطلوب يصبح

$$\Pr(A)=\frac{n(A)}{n(S)}=\frac{4}{14}$$

يتم حساب احتمال حدوث حدث ما في الأسلوب التجريبي في حال عدم توفر الاحتمالات لعناصر فراغ العينة، حيث تكون في الأصل مجهولة وغير محددة في التجربة. فمثلاً احتمال أن يكون الوضع الاجتماعي المتقدم لوظيفة جديدة أعزب أو متزوج أو مطلق يكون مجهولاً دوماً ولا يمكن معرفته إلا بعد إجراء التجربة أو بالاعتماد على تجارب سابقة. كمثال آخر، احتمال أن تكون الصفقة القادمة لإحدى الشركات الكبرى التي تعمل في مدن المملكة العربية السعودية الكبيرة (الرياض، جدة، الدمام، المدينة المنورة) في إحدى المدن غير معروفة ولا يمكن تحديدها من خلال معلومات التجربة.

لذا فان عملية حساب الاحتمالات التجريبية تتم من خلال إحدى طريقين، إما أن يتم الاعتماد على معلومات سابقة تعطي انطباع عن قيم الاحتمالات المرافقة لعناصر الحدث المطلوب، أو أن يتم إجراء التجربة فعليا لعدد كبير من المرات ومن ثم يتم حساب احتمالات عناصر الحدث من خلال قسمة عدد مرات ظهور العنصر أو الحدث على إجمالي عدد المحاولات. وبالطبع تتطلب هذه الطريقة إجراء التجربة عدداً كبيراً جداً من المرات ليتم الحصول على تقدير بمصداقية عالية.

مثال 1-5-2-3

في دراسة عن الحوادث المرورية التي تقع في إحدى المناطق، يتجه الاهتمام إلى تقدير احتمال وقوع الحوادث في يوم محدد من أيام الأسبوع. لتقدير الاحتمالات المرافقة لأيام الأسبوع تم حصر جميع الحوادث السابقة في المنطقة لفترة السنة الماضية وتم إفراغها في جدول يصنفها حسب يوم وقوع الحادث، أي يوم الأسبوع، الجدول 1-5-2-1

جدول 1-5-2-1: توزيع 4762 حادث مروري حسب يوم وقوع الحادث

عدد الحوادث	اليوم
501	السبت
356	الأحد
406	الاثنين
235	الثلاثاء
623	الأربعاء
1256	الخميس
1385	الجمعة

أوجد التوزيع الاحتمال لحدوث الحوادث على أيام الأسبوع.

الحل

لحساب احتمالات حدوث الحوادث في أيام الأسبوع المختلفة والتي تمثل عناصر الحدث المطلوب، يتم قسمة عدد مرات حدوث الحوادث في اليوم المحدد على مجموع الحوادث الكلي لينتج تقدير للاحتمال المطلوب. الجدول 1-5-2-2 يبين الاحتمالات المرافقة لأيام الأسبوع السبعة. لاحظ هنا أن مجموع الاحتمالات يساوي الواحد الصحيح إشارةً إلى تمثيل أيام الأسبوع لفراغ العينة. لاحظ كذلك انه قد تم تقريب الاحتمالات المحصلة تحت شرط واحد أن يكون مجموع الاحتمالات السبعة مساوياً للواحد الصحيح.

جدول 1-5-2-2: التوزيع الاحتمالي للحوادث المرورية حسب أيام الأسبوع

Pr	اليوم
0.11	السبت
0.07	الأحد
0.09	الاثنين

الثلاثاء	0.05
الأربعاء	0.13
الخميس	0.26
الجمعة	0.29
$\sum Pr$	1

لذا فان احتمال وقوع حادث في يوم السبت يساوي 11% (نسبة الحوادث يوم السبت) بينما احتمال وقوع حادث في يوم الأحد من الأسبوع يساوي 7% (نسبة الحوادث يوم الأحد) وهكذا يتم تفسير باقي الاحتمالات المحصلة في الجدول السابق.

3-5-1 احتمالات حوادث متجمعة

عندما يكون هنالك حدثان محتمل وقوعهما فانه يمكن تحديد عدد من أوجه العلاقة الرياضية التي يمكن أن تربطهما معا. وبالطبع يفترض أن يكون كلا الحدثين في مجال فضاء العينة للتجربة، أي أن يكون كل من الحدث الأول A والحدث الثاني B مجموعات جزئية من المجموعة الممثلة لفراغ العينة S،

$$A \subseteq S \quad \& \quad B \subseteq S$$

يمكن تمثيل اتحاد حدثين بحدث C يمثل مجموع العناصر الموجودة في الحدثين. ويرمز للاتحاد بالرمز $\cup$،

$$C = A \cup B$$

كما يتم الإشارة إلى الحدث D والذي يمثل تقاطع مجموعتين حيث يعبر عن عملية التقاطع بالرمز $\cap$

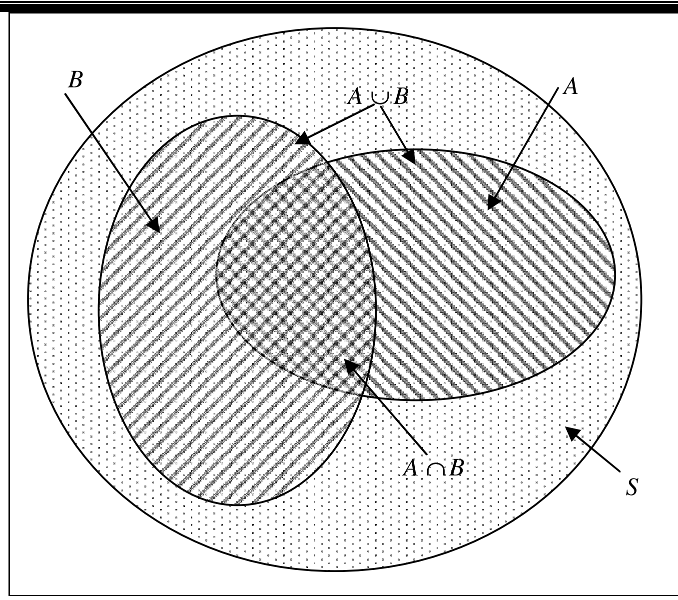
$$D = A \cap B$$

والتي تضم فقط العناصر الموجودة في المجموعة A وفي نفس الوقت موجودة في المجموعة B. الشكل 1-3-5-1 يبين أشكال فن لكل من الاتحاد والتقاطع.

يتم الرمز للمجموعة الخالية بالرمز $\varnothing$ والذي يمثل المجموعة الجزئية من فراغ العينة والتي لا يوجد بها عناصر. وتعتبر المجموعة الخالية المجموعة المتممة لمجموعة فراغ العينة $S^{'}$، أي

$$S^{'} = \varnothing$$

شكـــل 1-3-5-1
اتحــــاد وتقــــاطع المجموعات



وعندما يكون تقاطع مجموعتين مساوياً للمجموعة الخالية

$$A \cap B = \varnothing$$

فان ذلك يشير إلى أن المجموعتين هما في الواقع متنافيتان، أي لا يمكن حدوثهما معا (Mutually Exclusive). كما أن تقاطع مجموعة مع مكملتها تساوي المجموعة الخالية

$$A \cap A' = \varnothing$$

مشيرا إلى أن المجموعة A ومكملتها هما في الواقع حوادث متنافية الوقوع معا.

مثال 1-3-5-1

افترض أن فراغ العينة معطى بالمجموعة التالية:

$$S = \{1,3,5,7,9,11,13,15,17\}$$

كذلك افترض الحوادث التالية الممثلة بالمجموعات التالية:

$$A = \{5,9,13,17\}$$

$$B = \{3,5,7,15,17\}$$

$$C = \{1,3,7\}$$

اوجد،

a) A' b) $A \cup C$ c) $A \cap C$

d) $A' \cup B$ e) $(A \cup B)'$ f) $A' \cup B'$

g) $A \cup B \cup C$ h) $A \cap B \cap C$

الحــل

a) $A' = \{1,3,7,11,15\}$

b) $A \cup C = \{1,3,5,7,9,13,17\}$

c) $A \cap C = \varnothing$

d) $A' \cup B = \{1,3,5,7,11,15,17\}$

e) $(A \cup B)' = A' \cap B' = \{1,11\}$

f) $A' \cup B' = (A \cap B)' = \{1,3,7,9,11,13,15\}$

g) $A \cup B \cup C = \{1,3,5,7,9,13,15,17\}$

h) $A \cap B \cap C = \varnothing$

وكما تبين من المثال السابق، فانه يمكن إدخال إشارة المتمم على الأقواس بشرط تغير نوع العلاقة،

$$(A \cap B)' = A' \cup B'$$

و

$$(A \cup B)' = A' \cap B'$$

وباعتماد العلاقات الرياضية بين المجموعات يمكن سرد بعض قوانين الاحتمالات المهمة والتي تعنى بالعلاقات بين المجموعات. وبحكم العلاقة بين مجموعة ومكملتها فان مجموع احتمالاتهما يجب أن يساوي للواحد الصحيح،

$$\Pr(A) + \Pr(A') = 1 \tag{42}$$

ويتم حساب احتمال المجموعة الخالية من خلال العلاقة التالية:

$$\Pr(\varnothing) = 0 \tag{43}$$

حيث

$$\begin{aligned} \Pr(\varnothing) &= \Pr(S') \\ &= 1 - \Pr(S) \\ &= 1 - 1 = 0 \end{aligned}$$

وفي حال كان فراغ العينة S مكوناً بشكل كامل من حدثين A و B، فانه يمكن تقسيم احتمال حدوث حدث محدد تبعا لتقاطعه مع الحدث الآخر، فاحتمال حدوث الحدث A، مثلا، يساوي مجموع احتمال تقاطعه مع كل من الحدث B وتقاطعه مع متمم الحدث B،

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B') \tag{44}$$

كذلك يمكن حساب قيمة احتمال حدوث الحدث B من خلال علاقته بالحدث A بنفس الأسلوب السابق،

$$\Pr(B) = \Pr(B \cap A) + \Pr(B \cap A') \tag{45}$$

وبشكل عام فان قيمة احتمال وقوع أحد حدثين أو وقوعهما معا يتم التعبير عنهما من خلال عملية الاتحاد ليتم الحصول على الاحتمال المطلوب،

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \tag{46}$$

وفي حال كان الحدثين متنافيين الوقوع معا،

$$\Pr(A \cap B) = 0 \tag{47}$$

فان احتمال اتحاد حدثين يصبح

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{48}$$

مثال 1-5-3-2

افترض أن فراغ العينة لتجربة ما ممثل بالمجموعة التالية:

$$S = \{1,3,5,7,9,11,13,15,17\}$$

حيث يتوزع الاحتمال على العناصر بالتساوي، (عناصر متساوية فرصة الوقوع). افترض الحوادث التالية:

$$A = \{5,9,13,17\}$$

$$B = \{3,5,7,15,17\}$$

$$C = \{1,3,7\}$$

اوجد،

a) $\Pr(A')$ b) $\Pr(A \cup C)$ c) $\Pr(A \cap C)$

d) $\Pr(A' \cup B)$ e) $\Pr(A \cup B)'$ f) $\Pr(A' \cup B')$

g) $\Pr(A \cup B \cup C)$ h) $\Pr(A \cap B \cap C)$

الحـــل

a) $\Pr(A') = \frac{5}{9}$ b) $\Pr(A \cup C) = \frac{7}{9}$

c) $\Pr(A \cap C) = 0$

d) $\Pr(A' \cup B) = \frac{7}{9}$

e) $\Pr(A \cup B)' = \Pr(A' \cap B') = \frac{2}{9}$

f) $\Pr(A' \cup B') = \Pr(A \cap B)' = \frac{7}{9}$

g) $\Pr(A \cup B \cup C) = \frac{8}{9}$

h) $\Pr(A \cap B \cap C) = 0$

مثال 1-5-3-3

افترض أن A و B حادثتان غير فارغتين وممثلتين في فراغ العينة S حيث،

$$\Pr(A) = 0.6$$

$$\Pr(B) = 0.7$$

$$\Pr(A \cap B) = 0.3$$

اوجد،

a) $\Pr\left(B^{'}\right)$ b) $\Pr\left(A\cap B\right)^{'}$

c) $\Pr\left(A\cup B\right)$ d) $\Pr\left(A^{'}\cup B\right)$

e) $\Pr\left(A\cup B\right)^{'}$ f) $\Pr\left(A^{'}\cup B^{'}\right)$

الحـــل

a) $$\Pr\left(B^{'}\right)=1-\Pr\left(B\right)=1-0.7=0.3$$

b) $$\Pr\left(A\cap B\right)^{'}=1-\Pr\left(A\cap B\right)$$
$$=1-0.3=0.7$$

c) $$\Pr\left(A\cup B\right)=\Pr\left(A\right)+\Pr\left(B\right)-\Pr\left(A\cap B\right)$$
$$=0.6+0.7-0.3=1.0$$

d) $$\begin{aligned}\Pr\left(A^{'}\cup B\right)&=\Pr\left(A^{'}\right)+\Pr\left(B\right)-\Pr\left(A^{'}\cap B\right)\\&=\Pr\left(A^{'}\right)+\Pr\left(B\right)-\left[\Pr\left(B\right)-\Pr\left(A\cap B\right)\right]\\&=\Pr\left(A^{'}\right)+\Pr\left(A\cap B\right)\\&=\Pr\left(B\right)=0.7\end{aligned}$$

e) $$\Pr\left(A\cup B\right)^{'}=1-\Pr\left(A\cap B\right)=0$$

f) $$\Pr\left(A^{'}\cup B^{'}\right)=\Pr\left(A\cap B\right)^{'}$$
$$=1-\Pr\left(A\cap B\right)=1-0.3=0.7$$

4-5-1 الاحتمالات الشرطية ونظرية بيز

يمكن للحوادث المختلفة أن تكون مستقلة بمعنى أن لا يتأثر احتمـــال وقـــوع احـــدها بمـــا يجـــرى علـــى الآخـــر، كما يمكن أن يكون العكس تماماً حيث يتغير احتمال وقوع حدث مـــا تبعـــا لنتيجـــة الحـــدث الآخـــر. ويـــتم هنـــا اســـتخدام الاحتمالات الشرطية للتعبير عن العلاقة بين متغيرين غير مســـتقلين، حيـــث يـــتم التعبيـــر رياضـــيا عـــن احتمـــال وقـــوع الحدث A إذا علم نتيجة وقوع الحدث B بالشكل التالي:

$$\Pr\left(A\mid B\right)$$

وإذا كان الحدثان مستقلين تماما فان الاحتمال الشرطي يصبح غير ضروري،

$$\Pr\left(A\mid B\right)=\Pr\left(B\right) \qquad (49)$$

وعليه، فان قانون حساب احتمال تقاطع حدثين مستقلين يصبح،

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{50}$$

أما الحوادث الغير مستقلة فيتم حساب احتمال تقاطعهما بالصيغة،

$$\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) \tag{51}$$

أو بالصيغة

$$\Pr(A \cap B) = \Pr(B)\Pr(A \mid B) \tag{52}$$

تمثل نظرية بيز للاحتمالات واحدة من أهم استخدامات الاحتمالات الشرطية، حيث تعنى بدراسة العلاقة بين حدوث حدث محدد في ظل حدوث حدث واحد من عدة حوادث متنافية للوقوع معا، وممثلة لفراغ العينة. لإيضاح الفكرة، افترض أن فراغ العينة S مكون من عدد r من الحوادث المتنافية للوقوع معا، حيث

$$S = A_1 \cup A_2 \cup \cdots \cup A_r$$

و

$$A_i \cap A_j = \varnothing \qquad \forall \quad i \neq j$$

لذا

$$\sum_{i=1}^{r} \Pr(A_i) = 1 \tag{53}$$

افترض ألان أن هنالك حدث D مرتبط بالحوادث السابقة من خلال معلومية العلاقات الشرطية التالية

$$\Pr(D \mid A_i) \qquad \forall \quad i = 1, 2, \ldots, r \tag{54}$$

لذا فان العلاقة بين الحدثين A و D يمكن التعبير عنها رياضيا بالعلاقة التالية،

$$\Pr(A_i \cap D) = \Pr(A_i)\Pr(D \mid A_i) \qquad \forall \quad i = 1, 2, \ldots, r \tag{55}$$

وبمعلومية كل من $\Pr(A_i)$ و $\Pr(D \mid A_i)$ يتم الحصول على احتمال حدوث الحدث D من خلال الدالة التالية،

$$\begin{aligned} \Pr(D) &= \sum_{i=1}^{r} \Pr(A_i \cap D) \\ &= \sum_{i=1}^{r} \left[\Pr(A_i)\Pr(D \mid A_i)\right] \end{aligned} \tag{56}$$

كذلك يتم الحصول على الاحتمال الشرطي لحدوث الحدث A_i بمعلومية حدوث الحدث D من خلال تطبيق قاعدة بيز في الاحتمالات،

$$\Pr(A_i \mid D) = \frac{\Pr(A_i \cap D)}{\Pr(D)} \tag{57}$$

ويمكن تفصيل المعادلة (1.57) لتصبح بالشكل التالي،

$$\Pr(A_i \mid D) = \frac{\Pr(A_i)\Pr(D \mid A_i)}{\sum_{i=1}^{r}\left[\Pr(A_i)\Pr(D \mid A_i)\right]} \qquad (58)$$

مثال 1-4-5-1

يعتمد مصنع ما على أربعة خطوط إنتاج هي A_1، A_2، A_3 و A_4. إذا كان إنتاج المصنع ككل يتوزع على خطوط الإنتاج الأربعة حسب النسب التالية:

$$\Pr(A_1) = 0.20 \quad \& \quad \Pr(A_2) = 0.30$$
$$\Pr(A_3) = 0.40 \quad \& \quad \Pr(A_4) = 0.10$$

إذا كانت خطوط الإنتاج تنتج وحدات معيبة بنسب 0.05، 0.07، 0.06 و 0.09 على التوالي، وتم اختيار وحدة منتجة عشوائيا وتبين أنها معيبة فاوجد احتمال أن تكون من إنتاج خط الإنتاج A_3.

الحل

تبعا لمعطيات السؤال يتم تحديد الاحتمالات الشرطية التالية

$$\Pr(D \mid A_1) = 0.05 \quad \& \quad \Pr(D \mid A_2) = 0.07$$
$$\Pr(D \mid A_3) = 0.06 \quad \& \quad \Pr(D \mid A_4) = 0.09$$

وحيث أن المطلوب هو

$$\Pr(A_3 \mid D)$$

فانه يتوجب الحصول على احتمال إنتاج وحدة معيبة $\Pr(D)$ واحتمال إنتاج وحدة معيبة ومن خط الإنتاج الثالث $\Pr(A_3 \cap D)$

$$\Pr(D) = \sum_{i=1}^{4}\Pr(A_i \cap D)$$

حيث

$$\Pr(A_1 \cap D) = \Pr(A_1)\Pr(D \mid A_1) = 0.20(0.05) = 0.01$$

و

$$\Pr(A_2 \cap D) = \Pr(A_2)\Pr(D \mid A_2) = 0.30(0.07) = 0.021$$

و

$$\Pr(A_3 \cap D) = \Pr(A_3)\Pr(D \mid A_3) = 0.40(0.06) = 0.024$$

و

$$\Pr(A_4 \cap D) = \Pr(A_4)\Pr(D \mid A_4) = 0.10(0.09) = 0.009$$

لذا

$$\Pr(D) = 0.01 + 0.021 + 0.024 + 0.009 = 0.064$$

وعليه يمكن إيجاد الاحتمال المطلوب،

$$\Pr(A_3 \mid D) = \frac{\Pr(A_3)\Pr(D \mid A_3)}{\sum_{i=1}^{4}\left[\Pr(A_i)\Pr(D \mid A_i)\right]}$$

$$= \frac{0.024}{0.064} = 0.375$$

مثال 2-4-5-1

تمثل نسبة المبرمجين السعوديين من بين إجمالي المبرمجين في المملكة العربية السعودية فقط 20% . إذا كان 30% من المبرمجين السعوديين حاصلين على شهادة $MSCE$ بينما 55% من المبرمجين غير السعوديين حاصلين على الشهادة ذاتها، وإذا تم اختيار مبرمج حاصل على الشهادة بشكل عشوائي فما احتمال أن يكون سعودياً؟

الحل

للحصول على الاحتمال المطلوب يتم في البداية كتابة الاحتمالات المعطاة والاحتمالات المطلوبة بشكل رياضي، وبافتراض أن الجنسية السعودية يرمز لها بالرمز A والحصول على الشهادة المحددة بالرمز D يتم التالي:

$$\Pr(A) = 0.20$$

و

$$\Pr(A') = 0.80$$

حيث يشير الرمز A' إلى الجنسية غير السعودية، كذلك

$$\Pr(D \mid A) = 0.30$$

و

$$\Pr(D \mid A') = 0.55$$

وعليه فان احتمال أن يكون مبرمج حاصلاً على الشهادة يمكن الحصول عليه من خلال الدالة التالية:

$$\Pr(D) = \Pr(D \cap A) + \Pr(D \cap A')$$
$$= \Pr(A)\Pr(D \mid A) + \Pr(A')\Pr(D \mid A')$$
$$= 0.20(0.3) + 0.80(0.55)$$
$$= 0.06 + 0.44 = 0.50$$

ومن ثم فان الاحتمال الشرطي المطلوب في السؤال يتم الحصول عليه باستخدام نظرية بيز كما يلي

$$\Pr(A \mid D) = \frac{\Pr(A \cap D)}{\Pr(D)}$$
$$= \frac{\Pr(A)\Pr(D \mid A)}{\Pr(A)\Pr(D \mid A) + \Pr(A')\Pr(D \mid A')}$$
$$= \frac{0.06}{0.50} = 0.12$$

معادلات مهمة

مدى المتغير العشوائي

$$R = Max - Min$$

المنوال لبيانات مبوبة (طريقة الفروق)

$$M = A_{\circ} + \frac{F_{\circ} - F_{\leftarrow}}{2F_{\circ} - F_{\leftarrow} - F_{\rightarrow}}(L_{\circ})$$

المنوال لبيانات مبوبة (طريقة الرافعة)

$$M = A_{\circ} + \frac{F_{\rightarrow}}{F_{\leftarrow} + F_{\rightarrow}}(L_{\circ})$$

الوسيط: بيانات خام (غير مبوبة)

عندما يكون حجم العينة (عدد القيم) n فرديا،

$$Q_2 = X_{\frac{n+1}{2}}$$

عندما يكون حجم العينة n زوجيا،

$$Q_2 = \frac{X_{(n/2)} + X_{[(n/2)+1]}}{2}$$

الوسيط: بيانات مبوبة

$$Q_2 = A_{\circ} + \frac{\left(\frac{1+\sum F}{2}\right) - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_{\circ})$$

الوسط الحسابي: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N،

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

لعينة عشوائية بحجم n

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

الوسط الحسابي: بيانات مبوبة

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

الانحراف الربيعي: بيانات خام (غير مبوبة)

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

$$Q_3 = X_{3\frac{(n+1)}{4}} \quad و \quad Q_1 = X_{\frac{n+1}{4}}$$

الانحراف الربيعي: بيانات مبوبة

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

$$Q_1 = A_\circ + \frac{\frac{1+\sum F}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_\circ)$$

$$Q_3 = A_\circ + \frac{\frac{3\left(1+\sum F\right)}{4} - CF_{\leftarrow}}{CF_{\rightarrow} - CF_{\leftarrow}}(L_\circ)$$

الانحراف المتوسط: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

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

لعينة عشوائية بحجم n

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

الانحراف المتوسط: بيانات مبوبة

$$MD = \frac{\sum_{i=1}^{K}\left(\left|X_i - \bar{X}\right| F_i\right)}{\sum_{i=1}^{K} F_i}$$

الانحراف المعياري: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

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

لعينة عشوائية بحجم n

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

التباين: بيانات خام (غير مبوبة)

لمجتمع محدود بحجم N

$$\sigma_X^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_X^2$$

لعينة عشوائية بحجم n

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

الانحراف المعياري: بيانات مبوبة

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

التباين: بيانات مبوبة

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

معامل الاختلاف

لمجتمع محدود

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

لعينة عشوائية

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

تمارين

(1

الجدول التالي يمثل التوزيع التكراري لـــ 54 مطعم وجبات خفيفة حسب عدد طلبات السيارة خلال ساعة واحدة.

حدود الفئات (عدد الطلبيات خلال ساعة)	التكرار
2 -	12
8 -	14
12 -	16
20 - 30	12

أوجد التالي:

- التكرارات المعدلة ومنها أوجد الفئة المنوالية
- الانحراف الربيعي (نصف المدى الربيعي)
- الوسط الحسابي

(2

- الجدول التالي يمثل التوزيع التكراري لـــ 50 شركة حسب عدد الفروع،

حدود الفئات (عدد الفروع)	التكرار
1 -	8
5 -	20
9 -	12
13 - 17	10

المطلوب الحصول على التالي:

- التباين
- المنوال
- الجدول النسبي المتجمع الصاعد و منه أوجد نسبة الشركات التي يبلغ عدد فروعها 5 فأكثر.

(3

الجدول التالي يمثل الآجر الشهري ل 22 عامل بمئات الريالات

فئات الأجور	التكرار
10 -	2
14 -	8
18 -	9
22 - 26	3
مجموع التكرارات	22

أوجد:-

- الانحراف المعياري
- المنوال
- الوسيط

(4

تقوم شـركه باسـتيراد بضـائعها مـن أربعـه مصـادر (A، B ، C ، D) بنسـب 15% ، 20% ، 5% ، 60%، علـى التوالي. فإذا كان احتمال أن يتأخر إرسال طلبيـه مـن المصـدر هـو علـى التـوالي 20% ، 7% ، 13% ، 5% مـن المصادر A ، B ، C ، D ، على التوالي. إذا اختيرت طلبيه معينه و كانـت متـأخرة فأوجـد احتمـال أن تكـون مرسـله من المصدر الرابع (D).

(5

القيم التالية تمثل المصاريف الشهري لمؤسسه ما بآلاف الريالات

17	12	10	14	19	26	18	10

أوجد:-

- الانحراف المتوسط
- التباين
- معامل الاختلاف

(6

إذا كان

$$\Pr(A) = 0.50$$
$$\Pr(B^{'} | A) = 0.60$$

فأوجد

$$\Pr(A \cap B)$$

(7

إذا كان احتمال أن يكون الطقس الجوي ليوم ما مشـابه لليـوم السـابق يسـاوي 0.8 فأوجـد احتمـال أن يكـون الطقـس الجوي ليوم بعد غد ممطراً إذا كان الطقس لليوم الحالي ممطراً.

(8

توفر لدى باحث 100 قيمة لمتغير ما و يرغب هذا الباحث فـي وضـعها فـي جـدول تكـراري بحيـث يكـون هنـاك 4 فئات فقط. إذا كانت اكبر قيمة من بـين الـــ 100 قيمـة تسـاوي 38 و اصـغر قيمـة تسـاوي 26 فأوجـد طـول الفئـة المناسب على افتراض أن الفئات الأربع منتظمة.

---

9)

الجدول التالي يمثل المصاريف الشهري لمؤسسه تجارية بآلاف الريالات

الشهر	1	2	3	4	5	6	7	8	9
المصروف	23	17	12	10	14	19	26	18	10

أوجد:-

- الوسط الحسابي
- الانحراف المتوسط
- التباين
- نصف المدى الربيعي

10)

الجدول التالي يمثل الآجر الشهري ل 38 عامل بمئات الريالات

حدود الفئات	التكرار
8 -	3
12 -	6
16 -	14
20 -	10
24 - 28	5
مجموع التكرارات	38

أوجد:-

- متوسط أجور العمال
- الوسيط
- الانحراف المعياري
- المنوال

11)

إذا كان المتغير X يرمز للدخل الشهري و المتغير Y يرمز للأنفاق الشهري لعشر عائلات بمئات الريالات، و كان من المعطيات القيم التالية:

$$\bar{X} = 380 \quad \& \quad S_x = 84$$
$$\bar{Y} = 89 \quad \& \quad S_y = 25$$

فأي المتغيرين (الأنفاق أو الدخل) أكثر تشتتا، لماذا؟

12)

البيانات التالية تمثل الربح الشهري (بآلاف الريالات) لإحدى الشركات الكبيرة وذلك لفترة الثلاثة عشر شهراً الماضية

8 25 17 21 29 25 21

26 29 29 35 29 34

أوجد كل من

- الوسيط
- التباين

(13

الجدول التالي يمثل توزيع 50 طرداً بريدياً حسب الوزن بالكيلوجرام

التكرار	حدود الفئات
2	6 -
14	8 -
18	12 -
10	20 -
6	30 - 50

أوجد التالي:

- الوسيط
- المنوال
- التباين

(14

الجدول التالي يمثل توزيع 145 عاملاً مهنياً حسب الجنسية والمستوى التعليمي

أعلى من ثانوي	ثانوي	اقل من ثانوي	المؤهل التعليمي الجنسية
5	24	33	سعودي
14	42	27	غير سعودي

إذا تم اختيار أحد العمال عشوائيا فأوجد احتمال أن

- يكون سعودياً إذا علم أن مؤهله ثانوي
- يكون غير سعودي أو ذو مؤهل أعلى من ثانوي

(15

البيانات التالية تمثل قيم متغير في مجتمع ما

12 23 11 14 10 18 16 10

أوجد التالي:

- الانحراف المعياري
- معامل الاختلاف

(16

إذا كانت درجات الطلاب النهائية في إحدى الشعب تتبع التوزيع الطبيعي بمتوسط 74 درجة و انحراف معياري 8 درجات، إذا تم اختيار أحد الطلاب عشوائياً فكم احتمال أن تكون درجته

- بين 80 و 90 درجة
- بين 80 و 70 درجة
- بين 70 و 64 درجة
- اكبر من 60

(17

الجدول التالي يمثل التكلفة الإجمالية لــ 40 طريق ترابي قامت به شركة مقاولات في السنة الماضية

التكرار	التكلفة (بآلاف الريالات)
12	0 -
20	40 -
11	100 -
7	200 - 300

أوجد التالي:

- الوسط الحسابي
- الوسيط

(18

الجدول التالي يبين توزيع 100 طالب تم قبولهم في جامعة الملك سعود حسب الجنس و الشهادة الثانوية

ذكر	أنثى	
22	11	علمي
36	19	أدبي
12	0	تجاري

إذا تم اختيار طالب عشوائياً، أوجد التالي:

- احتمال أن يكون ذكراً و يحمل مؤهل ثانوي علمي
- إذا علمنا أن مؤهله في الثانوي أدبي فأوجد احتمال أن يكون ذكراً

(19

القيم التالية تمثل عدد الرحلات المتأخرة يومياً القادمة لأحد المطارات

23	14	20	1	26	14	14	18	22	6	15
17	7	14	20	22	34	9	28	24	12	7
15	16	21	5	13	14	24	11	9	27	4
19	23	16	7	9	25	22				

أوجد التالي

- الوسط الحسابي
- الوسيط

- المنوال
- ضع هذه البيانات في جدول منتظم مكون من خمسة فئات

(20

الجدول التالي يمثل التوزيع التكراري لـ 71 رحلة سياحية مقدمة عبر احد المكاتب السياحية وموزعة حسب تكلفة الرحلة للفرد الواحد

عدد الرحلات	التكلفة بآلاف الريالات
14	0 -
9	10 -
12	15 -
6	20 -
30	25 - 40

أوجد التالي

- الانحراف المعياري
- نصف المدى الربيعي
- معامل الاختلاف

(21

الجدول التالي يمثل توزيع طلاب شعبة معينة تحتوي على 100 طالب مصنفين حسب الحالة الاجتماعية و المستوي الأكاديمي

ضعيف	جيد	ممتاز	المستوى الأكاديمي الحالة الاجتماعية
17	37	11	أعزب
5	10	20	متزوج

إذا تم اختيار طالب عشوائيا فأوجد احتمال

- أن يكون أعزباً أو ضعيفاً
- أن يكون أعزباً و غير ممتاز
- أن يكون جيداً إذا كنا نعلم انه متزوج

إذا تم اختيار طالب آخر (بدون إرجاع الطالب الأول)، فأوجد احتمال

- أن يكون الطالب الثاني أعزباً إذا كان الطالب الأول متزوجاً

(22

الجدول التالي يمثل توزيع 42 عملية جراحية، حسب الوقت اللازم لإنجازها في الساعة، والتي أنجزت في إحدى المستشفيات الخاصة خلال الشهر الماضي

عدد العمليات	الوقت المستغرق

2 -	13
6 -	18
10 -	7
14 - 18	4

أوجد التالي:

- المنوال للوقت المستغرق لإنجاز العمليات في المستشفى وفسره
- الوسط الحسابي للوقت المستغرق لإنجاز العمليات في المستشفى
- الربيع الأول للوقت المستغرق لإنجاز العمليات في المستشفى وفسره

(23

الجدول التالي يمثل المبيعات بآلاف الريالات في أحد المحلات التجارية في الأسبوع الماضي

اليوم	المبيعات
السبت	7
الأحد	10
الاثنين	12
الثلاثاء	3
الأربعاء	5
الخميس	4
الجمعة	5

أوجد التالي:

- الانحراف المعياري للمبيعات اليومية
- الوسيط للمبيعات اليومية
- معامل الاختلاف

(24

باستخدام البيانات غير المبوبة التالية

7 ، 9 ، 6 ، 3 ، 10 ، 17 ، 11 ، 4 ، 8 ، 5 ، 20

أوجد

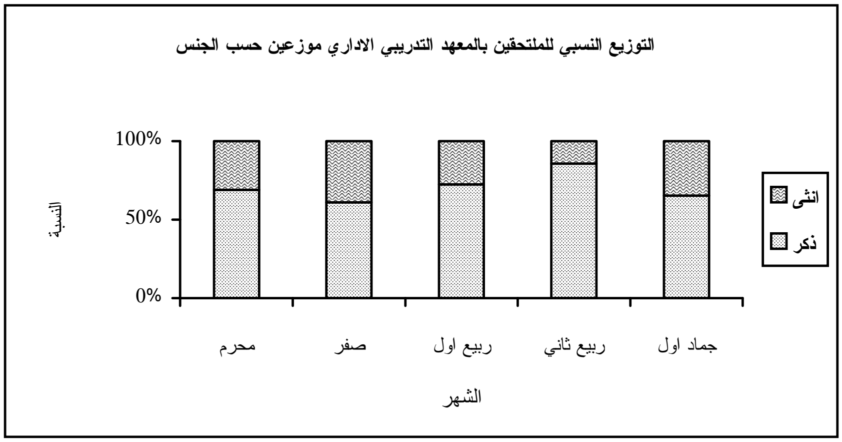
- الانحراف المعياري
- الانحراف المتوسط
- معامل الاختلاف النسبي

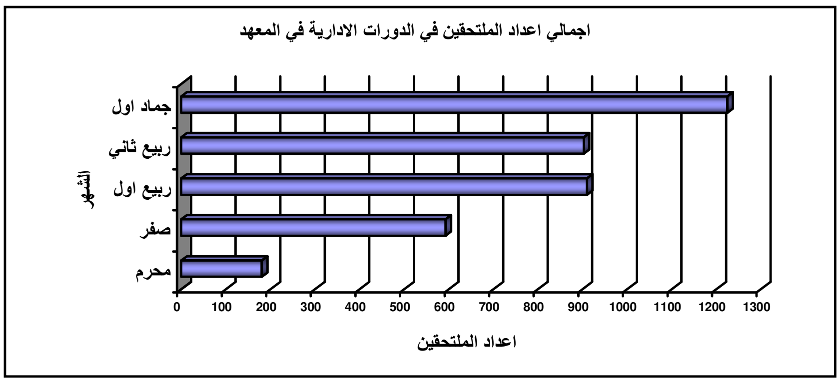
(25

الجدول التالي يحوي أعداد الملتحقين بالدورات التدريبية في مجال الإدارة لأحد المعاهد التدريبية. أكمل بيانات الجدول مستعينا بمعطيات الرسوم البيانية التالية. استخدم أرقام تقريبية.

جماد أول	ربيع ثاني	ربيع أول	صفر	محرم	

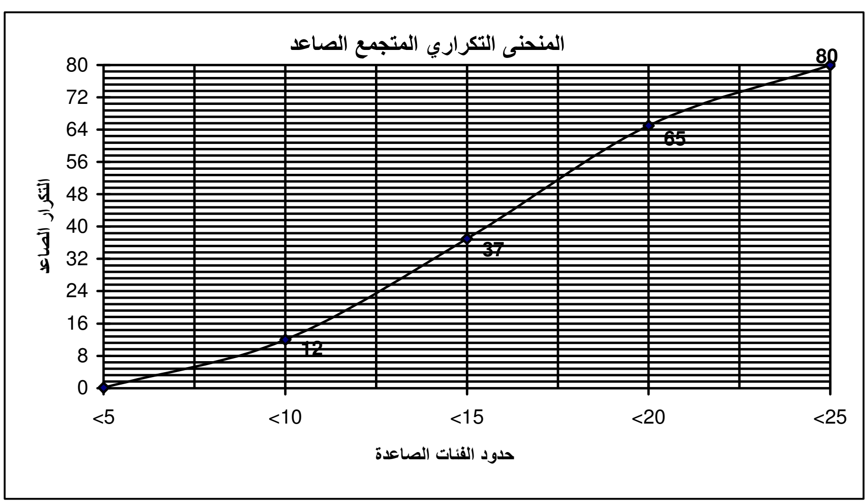
ذكر
أنثى





(26

الرسم البياني التالي يبين المنحنى المتجمع الصاعد لتوزيع أعداد 80 طالب حسب الدرجة النهائية في امتحان من 25 درجة.



استخدم الرسم البياني أعلاه للإجابة على التالي:

- ارسم المنحنى المتجمع الهابط على نفس الرسم السابق، مع وضع علامات وقيم التكرار المتجمع الهابط وعناوين حدود الفئات الهابطة.
- قدر قيمة الوسيط وفسرها
- قدر قيمة الربيع الأول وفسره
- قدر قيمة الربيع الثالث وفسره
- قدر الدرجة التي حصل 40% من الطلاب على اقل منها
- قدر عدد الطلاب الحاصلين على درجة اقل من 18
- قدر عدد الطلاب الحاصلين على درجة أكثر من 12
- قدر عدد الطلاب الحاصلين على درجات بين 17 و 22

27)

الجدول التالي يبين توزيع 48 شركة سعودية متوسطة الحجم حسب الربح الشهري بالريال السعودي،

التكرار	حدود الفئات (بآلاف الريالات)
5	10 -
8	10 - 30
16	30 - 70
12	70 - 100
7	100 - 150

أوجد المقاييس التالية:

- الوسط الحسابي
- التباين
- نصف المدى الربيعي
- المنوال (طريقة الفروق)
- في دراسة أخرى لنفس الـ 48 شركة تبين أن متوسط أعمارها مساوي لـ 6.2 سنة بينما كان الانحراف المعياري لأعمارها مساوي لـ 3.4 سنة. بين أيهما أكثر تشتتاً، أعمار الشركات أو الربح الشهري لها؟ لماذا؟

28)

يعتمد مصنع على ثلاث خطوط إنتاج حيث ينتج الخط الأول 30% من إنتاج المصنع بينما ينتج الخط الثاني 20% من إنتاج المصنع. إذا كان نسب الإنتاج المعيب في الخطوط الثلاثة هي على التوالي: 10% للخط الأول، 15% للخط الثاني، 12% للخط الثالث. إذا تم اختيار وحدة منتجة ووجد أنها معيبة فأوجد احتمال أن تكون من إنتاج الخط الثالث؟

29)

البيانات التالية تمثل أوزان وتكاليف إرسال 14 طرداً بريدياً تم إرساله مؤخرا بواسطة أحد المكاتب البريدية الخاصة،

الوزن (كلجم)	التكلفة (ريال)	رقم الطرد
2.3	12	1

2	20	5.6
3	36	9.4
4	20	1.8
5	15	3.2
6	20	5.6
7	20	4.9
8	14	3.2
9	25	6
10	32	7.5
11	12	3.2
12	26	6.1
13	45	9
14	9	2.6

باستخدام بيانات الجدول السابق أوجد التالي:

- الوسط الحسابي لكل من أوزان وتكاليف الطرود البريدية
- المنوال لكل من أوزان وتكاليف الطرود البريدية
- الوسيط لأوزان وتكاليف الطرود
- الانحراف المعياري لتكلفة الطرود البريدية
- التباين لوزن الطرود البريدية
- أيهما أكثر تشتتا الأوزان أم التكاليف؟ لماذا؟
- ضع الأوزان في جدول تكراري من ثلاث فئات

(30

في دراسة عن توزيع المضاربات في سوق الأسهم السعودية على كل من القطاع التجاري و الصناعي والزراعي وجد أن 0.30 من المضاربات تكون في أسهم شركات تجارية بينما 0.20 من المضاربات تكون في أسهم شركات صناعية و 0.50 في أسهم شركات زراعية. إذا علمت أن احتمال أن تكون أي صفقة في أسهم أي من القطاعات الثلاثة فاشلة (خاسرة) وذلك تحت افتراض أن نوع الأسهم معروفة مسبقا، هي كالتالي: 0.10 للأسهم الصناعية، 0.26 للأسهم الزراعية و 0.14 للأسهم التجارية. إذا تم سحب بطريقة عشوائية صفقة تمت في السابق فأوجد التالي:

- احتمال أن تكون الصفقة خاسرة.
- إذا علمت أن الصفقة خاسرة فأوجد احتمال أن تكون الصفقة في أسهم شركة زراعية.

(31

باستخدام الجدول التكراري التالي:

التكرار	حدود الفئات
11	6.5 -
23	12.5 -
16	15.5 -
7	18.5 -
3	20.5 - 21.5
60	$\sum$

أوجد كل من

- الوسط الحسابي

- الوسيط

32)

الجدول التالي يمثل توزيع 20 شقة حسب الإيجار السنوي في أحد الأحياء

الإيجار السنوي (بآلاف الريالات)	التكرار
10 -	3
20 -	11
30 -	5
40 - 50	1

أوجد

- المنوال
- الوسيط
- الانحراف المعياري
- معامل الاختلاف

33)

الجدول التالي يمثل المساحة بالمتر المربع لـــ 800 محل تجاري في مدينة الرياض

حدود الفئات	التكرار
0 -	244
50 -	132
70 -	84
100 - 200	340

أوجد التالي

- المنوال
- الانحراف المعياري

34)

القيم التالية تمثل درجات الحرارة المئوية للعشرين يوم الماضي في احد المناطق

21 20 19 19 18 20 19 20 23 23 24 22 23
20 18 19 19 20 23 20

أوجد التالي الوسط الحسابي

- التباين
- معامل الاختلاف
- الوسيط
- ضع البيانات في جدول تكراري من ٤ فئات

35)

افترض أن

$\Pr(A) = 0.20 \quad \& \quad \Pr(B|A) = 0.50 \quad \& \quad \Pr(A|B) = 0.4$ أوجد

$\Pr(A \cup B)$

(36

إذا علمت أن

$$\Pr(A) = 0.20 \quad \& \quad \Pr(B) = 0.30 \quad \& \quad \Pr(A \cap B) = 0.15$$

فأوجد التالي:

- $\Pr(A \cup B)$
- $\Pr(B | A)$

(37

إذا علم أن

$$\Pr(A \cup B) = 0.65 \quad \& \quad \Pr(B) = 0.30 \quad \& \quad \Pr(A) = 0.50$$

فأوجد $\Pr(A \cap B)$

(38

فيما يلي أعمار مجموعة من الطلاب في الصف الثاني الابتدائي بالإضافة آلي ترتيب الطفل بين إخوانه في المنزل من حيث الكبر في العمر:

عمر الطفل (X)	ترتيب العمر (Y)
6	1
7	3
6	2
7	2
7	3
8	6
9	4
8	5
7	3
6	1
6	2
9	7
8	4
7	1

أوجد التالي:

- متوسط أعمار الطلاب
- أيهما أكثر تشتتاً X أم Y
- ضع بيانات العمر (X) في جدول تكراري مكون من ٣ فئات
- تباين أعمار الطلاب
- الوسيط لكل من X و Y

(39

إذا كان

$$\Pr(A) = 0.40 \quad \& \quad \Pr(B) = 0.30 \quad \& \quad \Pr(A \cup B) = 0.58$$

فهل الحادثتين A و B مستقلتان أم لا؟ لماذا؟

(40

الجدول التالي يمثل توزيع 34 سيارة مستعملة حسب السعر بآلاف الريالات

التكرار	فئات السعر
8	10 -
12	20 -
9	30 -
5	40 - 50

أوجد التالي

- الوسط الحسابي
- الوسيط
- المنوال

(41

افترض أن البيانات التالية تمثل أعداد الطلاب في 16 شعبة في كلية العلوم الإدارية:

66	48	59	62	57	45	41	66	45	42	55	65	45
								53	40	67	45	65

أوجد التالي:

- الوسيط
- طول الفئة إذا كنا نرغب في إنشاء جدول تكراري للبيانات بحيث يتكون الجدول من أربع فئات منتظمة

(42

الجدول التالي يمثل توزيع 78 سيارة مستعملة حسب أسعارها

التكرار	حدود الفئات (بآلاف الريالات)
6	20 -
15	30 -
36	40 -
17	50 -
4	60 - 70
78	$\sum$

أوجد

- الربيع الأول
- الربيع الثالث
- السعر الذي لا تتجاوزه 60% من السيارات

- نسبة السيارات التي يزيد سعرها عن 45 آلف ريال

43)

إذا كان

$$\Pr(A \cap B) = 0.15 \quad \& \quad \Pr(A) = 0.50$$

فأوجد

$$\Pr(B \mid A)$$

الباب الثاني

المتغيرات العشوائية والتوزيعات الاحتمالية

1-2 مقدمـــــــــــــة

تدور كثير من الأمور الاحتمالية التي يهتم بها عادة رجال أعمـــال الإدارة حـــول نـــاتج رقمـــي عشـــوائي، مثـــل عدد الأشخاص الذين يقومون بإلغاء حجوزاتهم في أحد الفنادق خلال يوم واحد، وهـــذا العـــدد قـــد يختلـــف مـــن يـــوم إلـــى آخر ومن فندق إلى آخر ولكنه في النهاية يعتبر متغير عشوائي يأخذ قيماً لا يمكن تحديدها مسبقا.

ومما لاشك فيه فإن تقدم العلوم في كل المجالات أدى بدوره إلـــى تقـــدم علـــم الإدارة الحديثـــة فأصـــبحت عمليـــة اتخاذ القرار، والتي تعتبر البنية الأساســـية فـــي عمليـــة الإدارة، تقـــوم علـــى أســـاس علمـــي ولـــيس علـــى أســـاس آراء شخصية أو خبرة فقط كما كانت في السابق. ولبيان مدى مساهمة تقـــدم العلـــوم وبـــالأخص علـــم الإحصـــاء فـــي عمليـــة اتخاذ القرارات الإدارية نورد المثال التالي:

يواجه مدير تمويل في إحدى الشركات مشكلة اتخاذ قرار يتعلـــق بعـــدد الآلات التـــي يجـــب عليـــه شـــحنها إلـــى منطقة نائية تابعة للشركة، حيث قام الفرع هناك بطلب 6 وحدات مـــن الآلات. ولكـــن المشـــكلة تقـــع فـــي حقيقـــة أن هـــذه الآلات حساسة وزجاجية، أي أنها معرضة للتلف السريع خاصة وان الطريق إلـــى المنطقـــة النائيـــة غيـــر ممهـــد ووعـــر، لذلك فإن احتمال أن تصل الآلة سليمة إلى تلك المنطقة قد تم تقديره بواســـطة خبـــراء الإحصـــاء فـــي الشـــركة بــــــ 0.75. لا يرغب مدير التمويل في إرسال عدد اكبر من الذي يحتاجه ذلـــك الفـــرع فـــي المنطقـــة النائيـــة وفـــي نفـــس الوقـــت لا يريد المخاطرة بإرسال عدد قليل قد يصل منها اقل مـــن 6 وحـــدات ســـليمة ممـــا ينـــتج عنهـــا إعـــادة للرحلـــة مـــع مـــا يصاحبها من تكاليف مادية باهظة بالإضافة إلى تأخير الأعمـــال. فـــي هـــذه الحالـــة يمكـــن لمـــدير التمويـــل إتبـــاع أحـــد أسلوبين لحل هذه المشكلة. يمثل الأسلوب التقليدي الخيار الأول والذي يعتمـــد علـــى الحـــدس والـــرأي الشخصـــي، فمـــثلا قد يقرر مدير التمويل أن عدد الوحدات التـــي يجـــب إرســـالها هـــي 8 وحـــدات معتقـــداً أن فرصـــة وصـــول 6 وحـــدات سليمة ستكون فرصة معقولة ومقبولة في تقديره الشخصي. يلاحـــظ هنـــا أن هـــذا الأســـلوب يعتمـــد بشـــكل كبيـــر علـــى الحدس الشخصي والتوقعات الغير مبنية على أي أساس احتمـــالي رياضـــي ممـــا يقلـــل مـــن مصـــداقية نتائجـــه ويضـــعف فرصة الاعتماد عليه في إيجاد حلول لمثل هذه الأنـــواع مـــن المشـــاكل. يعتمـــد الأســـلوب الآخـــر، والـــذي يطلـــق عليـــه أحيانا مصطلح الأسلوب الحديث في الإدارة، على الاحتمالات والإحصـــاء فـــي عمليـــات اتخـــاذ القـــرارات. وبأتبـــاع هـــذا الأسلوب يستطيع مدير التمويل حساب احتمال وصول 6 وحدات على الأقل ســـليمة مـــن اصـــل 8 وحـــدات يـــتم إرســـالها معتمدا على توزيع احتمالي مثل توزيع ذي الحدين أو توزيع بواســـون. وباســـتخدام توزيـــع ذي الحـــدين يكـــون الاحتمـــال السابق مساوياً لــ 0.68، أي إن احتمال أن يصـــل اقـــل مـــن 6 وحـــدات ســـليمة يســـاوي 0.32 أي الثلـــث تقريبـــا، وهـــذا يعتبر احتمال كبير نوعا ما خاصة إذا كانت تكلفة إرسال الوحدات عالية. لـــذلك قـــد يقـــرر مـــدير التمويـــل إرســـال عـــدد أكثر من 8 وحدات رغبة في رفع احتمال النجاح (وهو وصول 6 وحـــدات علـــى الأقـــل ســـليمة)، ولكـــن مـــا هـــو العـــدد الذي يمكن القول معه أن احتمال المخاطرة مقبول؟

باستخدام نفس الأسلوب السابق يمكن حساب احتمـــال وصـــول 6 وحـــدات علـــى الأقـــل ســـليمة مـــن أصـــل 11 وحدة يتم إرسالها والذي سوف يكون مساوياً لــ 0.97، يعتبـــر هـــذا الاحتمـــال كبيـــر جـــداً مقارنـــة بالاحتمـــال الســـابق، وبالتالي يجب على المدير الناجح تجنب المخاطرة بأقل تكلفة ممكنة. لـــذلك نـــرى مـــدى الفائـــدة التـــي يمكـــن أن يجنيهـــا الإداري باستخدام الإحصاء والاحتمالات، وان العمليات الإدارية الحديثة ليست كلهـــا آراء وخبـــرة وحـــدس بـــل إنهـــا مـــع

تطور العلوم الإدارية الحديثة وتطور العلم بشكل عام أصبحت مثلها مثل سائر العلوم الأخرى تتطلب الدراسة العلمية والحسابات الدقيقة الاحتمالية والإحصائية للوصول إلى قرارات سليمة.

2-2 المتغير العشوائي

كما شاهدنا في الجزء السابق، يكون هناك قيمة رقمية في أي تجربة، تمثل فيها هذه القيمة متغير عشوائي يأخذ قيم مختلفة باحتمالات محددة. كمثال، يعتبر عدد الوحدات السليمة في المثال السابق متغير عشوائي يمكن أن يأخذ أي عدد صحيح من صفر إلى عدد الوحدات المرسلة، ولا يمكن الجزم بعدد الوحدات التي سوف تصل سليمة لان هذا العدد سوف تتحكم فيه العشوائية وإلا ما أصبح هنالك مشكلة أساساً.

المتغير العشوائي

المتغير العشوائي هو دالة تعطي قيماً حقيقية محددة لكل صفة معّرفة في فراغ العينة

افترض أننا قمنا برمي حجري نرد متزنين مرة واحدة و أردنا أن نرى القيم الممكنة الحدوث واحتمالاتها للمتغير العشوائي والذي نفترض انه يمثل القيمة المطلقة للفرق بين القراءتين، أي بين عدد النقط من الحجريين. عند رمي حجري النرد المتزنين، يكون هنالك 36 ناتج محتمل من جراء إجراء هذه التجربة حيث أن كل حجر نرد يحتوي 6 أوجه وأن هنالك حجري نرد. الجدول (2-2-1) يمثل جميع النتائج الممكنة الحدوث.

جدول 2-2-1: جميع الحالات الممكنة الحدوث من جراء إلقاء حجري نرد متزنين

الحجر الثاني / الحجر الأول	1	2	3	4	5	6
1	1,1	2,1	3,1	4,1	5,1	6,1
2	1,2	2,2	3,2	4,2	5,2	6,2
3	1,3	2,3	3,3	4,3	5,3	6,3
4	1,4	2,4	3,4	4,4	5,4	6,4
5	1,5	2,5	3,5	4,5	5,5	6,5
6	1,6	2,6	3,6	4,6	5,6	6,6

يطلق على مجموعة الحوادث الممكنة الوقوع بفراغ العينة، حيث أن كل حادثة أو عنصر في فراغ العينة يكون له نفس احتمال الحدوث، أي أن الحوادث يكون لها نفس فرصة الحدوث. بالنسبة لتجربة إلقاء حجري نرد متزنين نجد أن احتمال حدوث أي عنصر في فراغ العينة المبين في الجدول (2-2-1) مساوي لـ $1/36$ وذلك لوجود 36 فرصة متساوية احتمال الحدوث. دعنا الآن نحدد متغير ما يعتمد على فراغ العينة والذي سوف يمثل المتغير العشوائي في التجربة. بافتراض أن اهتمامنا منصب على القيمة المطلقة للفرق بين العدديين الناتجين من جراء إلقاء حجري النرد المتزنين، نجد أن القيم التي يمكن أن يأخذها هذا المتغير العشوائي، والذي نرمز له بالرمز X، ستكون محصورة في ستة قيم مختلفة هي:

$$X = 0,1,2,3,4,5$$

حيث أن اصغر فرق بين عدد النقاط في الحجرين هو صفر (وذلك عندما يكون الرقمين في الحجرين متساويين) واكبر فرق هو 5 (وذلك عندما يكون إحدى الحجرين 6 والآخر 1)

الجدول (2-2-2) يبين العلاقة بين المتغير العشوائي وفراغ العينة للتجربة والتي يتمثل دورها فقط في حصر الرقمين الناتجين. بالنسبة لفراغ العينة للمتغير العشوائي المتمثل في القيمة المطلقة للفرق بين الرقمين X، نجد أن الأمر يتطلب إيجاد ستة قيم مختلفة تمثل القيم التي يمكن أن يأخذها المتغير العشوائي X والتي ترافقها في هذه الحالة احتمالات غير متساوية.

جدول 2-2-2: قيم المتغير العشوائي X (القيم المطلقة للفروق)

الحجر الثاني / الحجر الأول	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

باستخدام الجدول (2-2-2) يمكن حساب الاحتمالات الستة المرافقة لقيم المتغير العشوائي X. فمثلا، احتمال أن تكون قيمة المتغير العشوائي، X، مساوية للصفر يتم حسابها كالتالي:

$$\Pr(X = 0) = \frac{6}{36}$$

حيث يمثل هذا الاحتمال عدد الحالات التي يكون فيها الناتج صفر نسبة إلى عدد الحالات الكلية. بنفس الطريقة يمكن إيجاد باقي الاحتمالات والتي يتم حسابها كالتالي:

$$\Pr(X = 1) = \frac{10}{36}$$

$$\Pr(X = 2) = \frac{8}{36}$$

$$\Pr(X = 3) = \frac{6}{36}$$

$$\Pr(X = 4) = \frac{4}{36}$$

$$\Pr(X = 5) = \frac{2}{36}$$

لاحظ هنا أن كل من القيم التي يأخذها المتغير العشوائي والاحتمالات المرافقة لهذه القيم تتحدد بتحديد ماهية المتغير العشوائي، فمثلا إذا افترضنا أن المتغير العشوائي X يمثل مجموع الرقمين الناتجين من التجربة، فإننا نجد أن القيم التي يمكن أن يأخذها المتغير العشوائي ستكون كالتالي:

$$X = 2,3,4,5,6,7,8,9,10,11,12$$

حيث يمثل الرقم 2 اصغر مجموع والرقم 12 اكبر مجموع. ويرافق هذه القيم احتمالات مختلفة يتم حسابها بنفس الطريقة السابقة التي تم استخدامها في حساب الاحتمالات السابقة. (يترك حساب هذه الاحتمالات المرافقة كتمرين)

من المميزات المهمة للمتغيرات العشوائية هي كونها تأخذ في اغلب الأحيان الصورة الكمية، وبالتالي تكون على صورة رقمية. لإعطاء تصور عن نوعية المتغيرات العشوائية المقصودة نورد الأمثلة التالية:

- الدخل الشهري لشخص ما في السنة القادمة.
- درجة الحرارة العليا ليوم الغد.
- كمية الخسارة المتوقعة في الصفقة الحالية لشركة ما.
- عدد الطلاب الذين سيسجلون في شعبة ما في الفصل القادم.
- درجة الامتحان الشهري التي يتوقعها طالب ما.
- عدد الحوادث المرورية التي يتوقع حدوثها ليوم غد في أحد المدن.
- نسبة النجاح في الشعب المطروحة في إحدى الكليات.

1-2-2 أنواع المتغيرات العشوائية

يمكن تقسيم المتغيرات العشوائية إلى نوعين رئيسين، هما متغيرات عشوائية منفصلة أو متقطعة ومتغيرات عشوائية متصلة أو مستمرة. المتغير العشوائي المتقطع هو المتغير الذي يمكن أن يأخذ قيماً محددة ومعلومة، في العادة تكون قيم الأعداد صحيحة غير كسرية بينما يأخذ المتغير العشوائي المستمر عدد لانهائي من القيم غير المحددة والتي تكون محصورة بين حدي فترة معينة. فمثلا، عدد الصفقات الناجحة التي تقوم بها شركة معينة يعتبر متغير عشوائي متقطع بينما تكلفة القيام بهذه الصفقات بالريال تعتبر متغير عشوائي مستمر. عدد الوحدات المعيبة والمنتجة بواسطة آلة معينة يعتبر، كمثال آخر، متغير عشوائي منفصل، بينما يمثل طول شخص ما بالسنتيمتر متغير عشوائي مستمر، حيث انه لا يمكن تحديد الطول الحقيقي بالدقة المتناهية ولو أردنا أن نحدد الطول الحقيقي لما أمكن ذلك حتى لو تم قياس الطول إلى جزء من البليون أو حتى مقياس اصغر. بعبارة أخرى، لا يأخذ المتغير العشوائي المستمر قيماً محددة ولكن يأخذ قيماً تقريبية يمكن أن تختلف تبعاً لدرجة الدقة المطلوبة. فمثلاً، إذا تم قياس طول شخص ما فوجد انه مساوي لـ 170 سنتيمتر، فإن الطول الحقيقي لن يكون بالضبط مساوي لـ 170 سنتيمتر بل قد يكون أطول أو اقصر من هذا الرقم ولو بجزء صغير جدا، وفي الواقع يعتبر الطول أيضاً من المتغيرات المتأثرة بالزمن، أي انه مع مرور الوقت قد يختلف الطول الحقيقي لشخص ما سواء بالزيادة أو بالنقصان. من الأمثلة الأخرى على المتغيرات العشوائية المستمرة، وزن شخص ما، الدرجة في امتحان، وقت صعود الرحلة، كمية العصير من برتقالة واحدة، طول المسافة بين بلدين .... وهكذا.

2-2-2 التوزيعات الاحتمالية

تعتبر المتغيرات العشوائية لبنة أساسية في معظم النظريات الإحصائية، حيث يتم تحديد العلاقة بين قيم المتغير العشوائي والاحتمالات المرافقة لها بواسطة التوزيعات الاحتمالية للمتغيرات العشوائية.

التوزيع الاحتمالي

يقوم التوزيع الاحتمالي بتوفير الاحتمالات المرافقة للقيم التي يمكن أن يأخذها المتغير العشوائي المتقطع أو للفترات التي يمكن أن يقع فيها المتغير العشوائي المستمر.

وعند التعامل مع متغير عشوائي سواء كان متصلاً أم منفصلاً لا بد أولاً من إيجاد التوزيع الاحتمالي والذي من خلاله نستطيع إيجاد الاحتمالات المرافقة لكل القيم أو الفترات التي يمكن أن يأخذها المتغير العشوائي. يطلق على الدالة المرافقة للتوزيع الاحتمالي للمتغير العشوائي بدالة الاحتمال للمتغير العشوائي المتقطع، وبدالة كثافة الاحتمال للمتغير العشوائي المستمر. بالنسبة للمتغير العشوائي المتقطع، يجب أن يتوفر شرطان أساسيان في دالة الاحتمال الممثلة للمتغير العشوائي المتقطع. ينص الشرط الأول على أن يكون الاحتمال المرافق لأي قيمة يأخذها المتغير العشوائي المتقطع X موجباً (غير سالب) وتكون قيمته بين الصفر والواحد صحيح،

$$0 \le \Pr(X = x) \le 1$$

حيث ترمز Pr إلى قيمة دالة الاحتمال للمتغير X والتي تمثل الاحتمال في نفس الوقت، بينما ترمز X للمتغير العشوائي و x للقيمة التي يأخذها المتغير العشوائي. الشرط الثاني الذي يجب أن يتحقق في الدالة الاحتمالية لتصبح دالة توزيع احتمالي هو أن يكون مجموع الاحتمالات لجميع القيم الممكنة الحدوث مساوي للواحد الصحيح،

$$\sum_{x} \Pr(X = x) = 1$$

إذا كان المتغير العشوائي متصلاً أو مستمراً فإن احتمال أن يأخذ المتغير العشوائي قيمة محددة يساوي صفراً، وذلك لان عدد القيم التي يمكن أن يأخذها المتغير العشوائي المتصل في أي فترة مهما صغرت سيكون عدداً لا نهائياً. لهذا السبب يطلق على دالة الاحتمال للمتغير العشوائي المتصل بدالة كثافة الاحتمال وذلك تحت افتراض توفر شرطين أساسين هما:

الشرط الأول:

أن تكون قيمة الدالة موجبة،

$$f(X = x) \ge 0$$

حيث f تمثل قيمة دالة كثافة الاحتمال أو قيمة الدالة الاحتمالية، كما انه يمكن لهذه القيمة أن تتجاوز الواحد الصحيح حيث إنها لا تمثل الاحتمال مباشرة وإنما تمثل بعد منحنى الدالة عن المحور الأفقي (X).

الشرط الثاني:

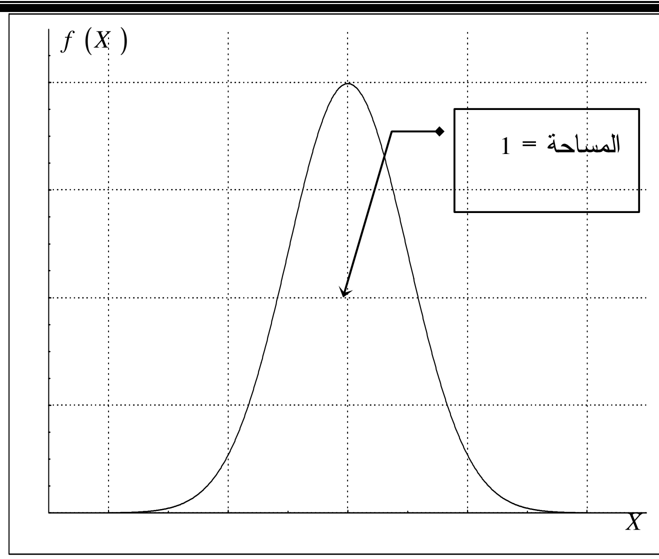
أن تكون المساحة تحت منحني دالة كثافة الاحتمال للفترة كاملة والتي يمكن أن يقع فيها المتغير العشوائي المستمر تساوي واحد صحيح.

الشكل (2-2-1) يوضح بيانيا الشرطيين السابقين، حيث تكون قيمة الدالة دائماً موجبة والمساحة تحت المنحني مساوية للواحد الصحيح،

$$\int_{-\infty}^{\infty} f(X = x) = 1$$

في الفقرة القادمة سوف نتطرق إلى بعض التوزيعات الاحتمالية المتقطعة، ثم بعد ذلك سيتم الإشارة إلى بعض التوزيعات الاحتمالية المتصلة في الفقرة التي تلي، مع مراعاة سرد أكثر أمثلة ممكنة على النوعين لتغطية موضوع المتغير العشوائي وتوزيعاته المهمة تغطية وافية، وذلك نظراً لأهمية هذا الموضوع في عملية فهم كثير من النظريات الإحصائية والمواضيع التي ستكون محور الاهتمام في الأبواب القادمة.

شكل 2-2-1
منحنى دالة كثافة الاحتمال لمتغير عشوائي مستمر



2-3 توزيعات احتمالية لمتغير عشوائي متقطع

كما سبق الإشارة إليه، يأخذ المتغير العشوائي المتقطع قيماً محددة، كما وان دالة الاحتمال لهذا المتغير تعطي مباشرة الاحتمالات المرافقة للقيم المختلفة التي يأخذها المتغير العشوائي المتقطع X. كذلك يمكن أيضا حساب مقاييس إحصائية للمتغير العشوائي X والتي تلعب دورا حساساً في كثير من النظريات الإحصائية، بالإضافة إلى إنها تعطي تصوراً عن شكل المتغير العشوائي أو بالأصح توزيعه. تتوفر كثير من المقاييس الإحصائية المهمة المستخدمة في العمليات الإحصائية، نذكر منها القيمة المتوقعة والتباين والانحراف المعياري للمتغير العشوائي المتقطع X.

يتم حساب القيمة المتوقعة للمتغير العشوائي المتقطع X عن طريق المعادلة التالية:

$$E(X) = \sum_{x}\left[x \Pr(X = x)\right] \tag{1}$$

حيث يمثل الرمز $E(X)$ القيمة المتوقعة للمتغير العشوائي X. تتم عملية الجمع على جميع قيم x والتي تمثل القيم الممكنة الحدوث. ففي مثال حجري النرد المتزنين، يمكن حساب القيمة المتوقعة للمتغير العشوائي X الذي يمثل القيمة المطلقة للفرق بين الرقمين الناتجين كالتالي:

$$\begin{aligned}
E(X) &= \sum_{x=0}^{5}\left[x \Pr(X = x)\right] \\
&= 0\left(\frac{6}{36}\right)+1\left(\frac{10}{36}\right)+2\left(\frac{8}{36}\right) \\
&= 3\left(\frac{6}{36}\right)+4\left(\frac{4}{36}\right)+5\left(\frac{8}{36}\right) \\
&= 1.94
\end{aligned}$$

يمثل هذا الناتج متوسط القيم الناتجة عند القيام بإجراء التجربة عدد كبير جدا من المرات، في كل مرة يتم قياس القيمة المطلقة للفرق بين العدديين الناتجين. في الواقع يمكن القول بأنه كلما اتجهت عدد المرات التي نقوم بها بإجراء التجربة إلى مالا نهاية، كلما اتجه متوسط القيم الناتجة إلى القيمة المتوقعة والتي تساوي 1.94 والمحسوبة عن طريق الاعتماد على دالة الاحتمال للمتغير العشوائي X.

وبما أن القيمة المتوقعة للمتغير العشوائي تعتمد على الأساس الاحتمالي الرياضي للتوزيع وليس على نتائج تجربة محددة لذلك فإن هذه القيم تمثل معلمات مجتمع، وبالتالي يرمز لها برموز المجتمع المتعارف عليها عالميا والمستخدمة في جميع البرامج الإحصائية في الحاسب الآلي وهي:

- μ (ميو) للقيمة المتوقعة أو التوقع الرياضي للمتغير العشوائي (والتي تمثل الوسط الحسابي للمتغير العشوائي).
- σ^2 (سيقما تربيع) لتباين المتغير *العشوائي*.

الدوال التالية تمثل عمليات حساب كل من القيمة المتوقعة (أو التوقع) والتباين والانحراف المعياري للمتغير العشوائي، X، على التوالي.

القيمة المتوقعة أو التوقع

$$\mu = E(X) = \sum_x \left[x \Pr(X = x) \right] \qquad (2)$$

التباين

$$\sigma^2 = \sum_x \left[(x-\mu)^2 \Pr(X = x) \right] \qquad (3)$$

يمكن تبسيط هذه الصيغة رياضيا لتصبح بالصورة المختصرة التالية:

$$\sigma^2 = \sum_x \left[x^2 \Pr(X = x) \right] - \mu^2 \qquad (4)$$

وحيث أن الانحراف المعياري هو الجذر التربيعي الموجب للتباين فإن الانحراف المعياري يمكن التعبير عنه من خلال القاعدة التالية:

الانحراف المعياري

$$\begin{aligned} \sigma &= \sqrt{\sigma^2} \\ &= \sqrt{\sum_x \left[x^2 \Pr(X - x) \right] - \mu^2} \end{aligned} \qquad (5)$$

مثال 1-3-2

عند رمي قطعة نقود متزنة مرة واحدة فإنه يتوقع أن يكون الناتج صورة أو رقم. أفترض أن المتغير العشوائي Xيمثل الحدث ظهور صورة. لذلك، فإن المتغير العشوائي X يأخذ القيمة 0 للدلالة على ظهور رقم بينما يأخذ القيمة 1 للدلالة على ظهور صورة. هذين الحدثين لهما نفس احتمال الحدوث. أوجد كل من التوزيع الاحتمالي والتوقع والتباين والانحراف المعياري للمتغير العشوائي X.

الحـــل

بما أن القطعة متزنة، فإن احتمال أن تظهر صورة من رمية واحدة يساوي احتمال أن لا تظهر صورة. وبالتالي فإن التوزيع الاحتمالي للمتغير العشوائي المتقطع X، والذي يمثل عدد الصور في رمية واحدة، يمكن الحصول عليه كما في الجدول التالي:

X	0	1	المجموع
$\Pr(X)$	0.5	0.5	1

وعليه فإن التوقع والتباين والانحراف المعياري للمتغير العشوائي X يمكن حسابها كما يلي:

$$\mu = E(X) = \sum_{x=0}^{1} x \Pr(X = x)$$
$$= 0(0.5) + 1(0.5) = 0.5$$

و

$$\sigma^2 = \sum_{x=0}^{1}\left(x^2 \Pr(X = x)\right) - \mu^2$$
$$= \left[0^2(0.5) + 1^2(0.5)\right] - (0.5)^2$$
$$= 0.25$$

كذلك

$$\sigma = \sqrt{\sigma^2}$$
$$= \sqrt{0.25}$$
$$= 0.5$$

مثال 2-3-2

افترض أن هناك صندوق فيه 10 كرات حمراء و 6 كرات بيضاء و 8 كرات صفراء. إذا تم سحب 3 كرات عشوائيا بدون إرجاع وكان المتغير العشوائي يمثل عدد الكرات الحمراء في الثلاث كرات المسحوبة. أوجد دالة الاحتمال للمتغير العشوائي وكل من التوقع والتباين والانحراف المعياري للمتغير العشوائي.

الحـــل

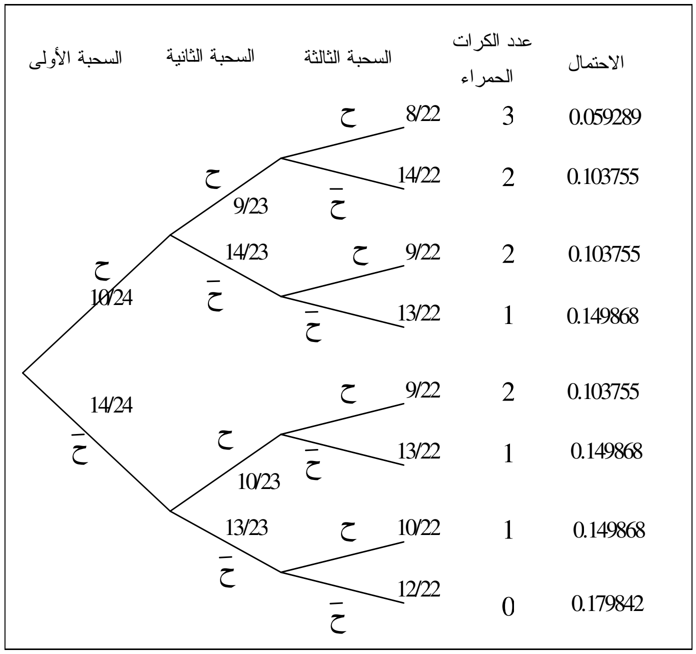
سوف نرمز للمتغير العشوائي والذي يمثل عدد الكرات الحمراء ضمن الكرات الثلاث المسحوبة بالرمز X حيث يأخذ هذا المتغير العشوائي قيم معلومة ومحددة وهي:

$$X = 0, 1, 2, 3$$

يلاحظ هنا أن المتغير العشوائي X هو متغير عشوائي متقطع. ولأجل إيجاد دالة الاحتمال للمتغير العشوائي X، يجب أولا حساب الاحتمالات المرافقة لكل القيم التي يمكن أن يأخذها المتغير العشوائي X. في الواقع هناك أكثر من أسلوب أو طريقة لإيجاد هذه الاحتمالات، حيث تعتمد على خلفية الشخص في موضوع الاحتمالات، ولعل ابسطها و أوضحها هو أسلوب الشجرة الاحتمالية. لإنشاء شجرة احتمالية سوف نحتاج إلى ثلاثة فروع أو حقول يكون فيها الحقل الأول مخصص للسحبة الأولى والحقل الثاني للسحبة الثانية والحقل الثالث للسحبة الثالثة. في كل سحبة سوف يكون هناك احتمالات ينصب عليها تركيزنا. في الواقع سيكون هناك احتمالان في كل

سحبة سيتم التركيز عليهما وهما احتمال أن تكون الكرة المسحوبة حمراء و الاحتمال المضاد، أي أن تكون السحبة بيضاء أو صفراء (غير حمراء). الشكل (2-3-1) يمثل الشجرة الاحتمالية لعملية سحب ثلاث كرات بدون إرجاع.

شكل 2-3-1 شجرة احتمالية لاختيار ثلاث كرات عشوائيا بدون إرجاع من بين 10 كرات حمراء و 8 صفراء و 6 بيضاء.



ح =الكرة المسحوبة حمراء. ح̄ =الكرة المسحوبة صفراء أو بيضاء (غير حمراء)

لاحظ هنا أن مجموع الاحتمالات يساوي واحد صحيح وهذا شرط أساسي لتكوين توزيع احتمالي للمتغير العشوائي، والذي يمكن كتابته كما في الجدول التالي:

المتغير العشوائي X	الاحتمال $\Pr(X=x)$
0	0.179842
1	0.449604
2	0.311265
3	0.059289
المجموع	1.000000

وعليه فإنه يمكن حساب كل من التوقع والتباين والانحراف المعياري للمتغير العشوائي X كما يلي:

$$\mu = E(X) = \sum_{x=0}^{3} x \Pr(X=x)$$
$$= 0(0.179842) + 1(0.449604)$$
$$+ 2(0.311265) + 3(0.059289)$$
$$= 1.250001$$
$$\sigma^2 = \sum_{x=0}^{3}\left[x^2 \Pr(X=x)\right] - \mu^2$$

$$= \left[0^2(0.179842) + 1^2(0.449604) + 2^2(0.311265) + 3^2(0.059289)\right] - 1.250001^2$$
$$= 2.228265 - 1.5625025 = 0.6657625$$
$$\sigma = \sqrt{\sigma^2}$$
$$= \sqrt{0.6657625} = 0.8159427$$

لاحظ أن الفكرة الرئيسية في حل المثال 2-3-2 هي في كيفية الوصول إلى التوزيع الاحتمالي للمتغير العشوائي X، أي في كيفية إيجاد الاحتمالات المرافقة للقيم التي يمكن يأخذها المتغير العشوائي X والتي يجب أن تستوفي شروط التوزيع الاحتمالي، وذلك بأن تكون جميع الاحتمالات موجبة، أي قيمتها بين الصفر والواحد، وان يكون مجموعها مساوي للواحد الصحيح. في حال توفر التوزيع الاحتمالي الذي يستوفي هذه الشروط، يمكن حينئذ حساب التوقع والتباين والانحراف المعياري للمتغير العشوائي المتقطع باستخدام القواعد المبينة في الدوال (2) ، (4) و (5)، على التوالي.

ذكرنا أن أهم خطوة للتعرف على أي متغير عشوائي متقطع هي عملية إيجاد أو معرفة التوزيع الاحتمالي له، والتي تمثل القيم التي يأخذها المتغير العشوائي والاحتمالات المرافقة لها. في بعض الحالات، كما في المثال 2-3-1، يتم إيجاد التوزيع الاحتمالي بسهولة نوعاً ما، خاصة إذا كان عدد القيم التي يأخذها المتغير العشوائي قليلاً. ولكن في الواقع غالبا ما تكون عملية إيجاد الاحتمالات، سواء بالملاحظة أو عن طريق شجرة الاحتمالات صعبة للغاية و تستغرق وقتاً طويلاً جدا.

كمثال، افترض انه تم سحب 8 كرات عشوائياً في المثال (2-3-2) والمطلوب إيجاد التوزيع الاحتمالي للمتغير العشوائي المتقطع X والذي يمثل عدد الكرات الحمراء ضمن الثماني كرات المسحوبة. في حالة استخدام شجرة الاحتمالات، يكون هناك ثماني قطاعات تضم $2^8 = 256$ احتمال نهائي. لذلك، فإنه من غير المجدي رسم شجرة احتمالات في هذه الحالة، حيث سوف تحوي الشجرة على 256 غصن، وبالتالي تصبح العملية مضيعة للجهد والوقت دون الحصول على فائدة تذكر. هذه الصعوبة تتزايد بتزايد عدد الكرات المسحوبة، فمثلا إذا ارتفع عدد الكرات المسحوبة إلى 15 كرة مثلاً، فإن عدد أغصان شجرة الاحتمالات سوف يكون مساوي لـ $2^{15} = 32768$، وبلا شك تصبح عملية رسم مثل هذه الشجرة محاولة يائسة لإيجاد الاحتمالات.

لذلك دعت الحاجة إلى إيجاد نموذج رياضي احتمالي يقوم بعملية إيجاد الاحتمالات المرافقة لكل قيم المتغير العشوائي، وهذا النموذج الرياضي الاحتمالي يطلق عليه التوزيع الاحتمالي للمتغير العشوائي.

2-3-1 توزيع ذي الحدين

يعتبر توزيع ذي الحدين واحد من أهم التوزيعات الاحتمالية للمتغير العشوائي المتقطع. يستخدم توزيع ذي الحدين لإيجاد الاحتمال المرافق لكل قيمة يأخذها المتغير العشوائي إذا تحقق في هذا المتغير العشوائي شروط ثلاثة معينة هي:

الشرط الأول:

أن يكون هناك عدة محاولات، مثل إلقاء قطعة نقود عدة مرات أو سحب عدة كرات بإرجاع من صندوق به كرات مختلفة الألوان.

الشرط الثاني:

أن يكون هناك احتمالين فقط لكل محاولة، يطلق على أحد الاحتمالين عادة احتمال النجاح ويطلق على الأخر احتمال الفشل. كمثال افترض أننا قمنا بإجراء تجربة متمثلة في إلقاء قطعة نقود عدة مرات ومن ثم يتم حساب عدد المرات التي تكون فيها النتيجة صورة. في هذه الحالة نعتبر حدوث صورة هو حدث النجاح بينما حدوث غير الصورة هو حدث الفشل. كمثال آخر، إذا كنا نهتم بعدد الكرات الحمراء المسحوبة من صندوق به كرات حمراء وغير حمراء فإن الحالات الممكنة الحدوث تكون حالتان، هما إما أن تكون الكرة المسحوبة حمراء أو أن لا تكون حمراء، أي إما أن تكون المحاولة ناجحة أو أن تكون فاشلة، على التوالي.

الشرط الثالث:

أن تكون المحاولات مستقلة، أي أن احتمال حدوث الحدث في محاولة ما لا يتأثر بنتيجة محاولة سابقة أو أخرى. في مثال 2-3-2 نجد أن المحاولات غير مستقلة وذلك لان الكرات يتم سحبها جميعا مرة واحدة. هذه العملية هي نفس عملية سحب كرة واحدة في كل مرة ولكن بدون إرجاع، وبالتالي فإنه إذا كانت السحبة الأولى، مثلا، حمراء فإن احتمال أن تكون السحبة الثانية أيضا حمراء سيكون مختلفا عنه لو كانت السحبة الأولى غير حمراء. لذلك فإنه في هذه الحالة يمكن القول بان المحاولات غير مستقلة وبالتالي لا يمكن استخدام توزيع ذي الحدين لإيجاد الاحتمالات المرافقة لقيم المتغير العشوائي. من زاوية أخرى، لو كان سحب الكرات مع الإرجاع لكان احتمال أن تكون الكرة المسحوبة حمراء ثابت وبالتالي تكون المحاولات مستقلة. في هذه الحالة يمكن استخدام توزيع ذي الحدين حيث أن كل من الشرطين المطلوبين للاستقلال متوفرين.

إذا توفرت الشروط الثلاثة السابقة في متغير عشوائي متقطع فإنه يمكن استخدام توزيع ذي الحدين لإيجاد الاحتمالات المرافقة لقيم التي يأخذها المتغير العشوائي. افترض أن المتغير العشوائي المتقطع X يمثل عدد مرات النجاح في عدد n من المحاولات المستقلة، يكون في كل محاولة احتمال النجاح والفشل ثابت. يلاحظ هنا أن المتغير العشوائي يأخذ القيم التالية:

$$X = 0,1,2,3,.....,n$$

يتم حساب الاحتمالات المرافقة لهذه القيم باستخدام دالة توزيع ذي الحدين والتي تأخذ الصورة المبينة في القاعدة التالية:

توزيع ذي الحدين

$$\Pr(X = x) = \binom{n}{x} P^{x} (1-P)^{(n-x)} \tag{6}$$

حيث

n عدد المحاولات.

x عدد مرات النجاح المطلوب إيجاد الاحتمال لها.

P احتمال النجاح و $(1-p)$ احتمال الفشل.

يدل الرمز $\binom{n}{x}$ على عدد الطرق الممكنة لاختيار x محاولة من n محاولة ويطلق عليها طرق العدد بأسلوب التوافيق. يتم حساب التوافيق أو عدد الطرق باستخدام المعادلة التالية:

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} \tag{7}$$

حيث يمثل $n!$ مضروب n ويساوي

$$n! = n\,(n-1)(n-2)........1$$

يشير الرمز $\Pr(X = x)$ إلى احتمال أن يكون عدد المحاولات الناجحة x من بين n محاولة. لذلك فإنه إذا كان المتغير العشوائي المتقطع يتبع توزيع ذي الحدين فلا بد من معرفة معلمتين لكي نتمكن من استخدام توزيع ذي الحدين. هاتين المعلمتين هما عدد المحاولات n واحتمال النجاح P. يستخدم الرمز الرياضي التالي:

$$X \sim Bin\,(n, P)$$

للدلالة على أن المتغير العشوائي X يتبع توزيع ذي الحدين بمعلمتين هما n و P حيث يمثلان عدد المحاولات واحتمال النجاح.

مثال 2-3-1-1

إذا كان من المعلوم أن 0.1 من إنتاج مصنع مصابيح معين يكون معيب، أي لا يعمل. فأوجد الاحتمالات التالية إذا تم اختيار 11 مصباح عشوائيا من إنتاج المصنع:

1) احتمال أن تكون جميع المصابيح تالفة أو معيبة.

2) احتمال أن يكون هناك على الأقل مصباح واحد سليم.

الحـــل

في البداية يلاحظ أن هذه الحالة تستوفي جميع شروط توزيع ذي الحدين حيث:

أ) عدد المصابيح يمثل عدد المحاولات

ب) في كل محاولة (مصباح) يحتمل وقوع واحد من حدثين هما: إما أن يكون المصباح معيب (حالة نجاح) أو أن يكون سليم (حالة فشل)

ج) احتمال أن يكون المصباح معيب (0.10) أو سليم (0.90) هو احتمال ثابت لا يتغير من محاولة إلى أخرى. كذلك المحاولات تعتبر مستقلة لعدم وجود علاقة تربط الوحدات مع بعضها البعض خاصة فيما يتعلق بكون الوحدات معيبة أو سليمة.

يفترض هنا أن المتغير العشوائي المتقطع X يمثل عدد الوحدات المعيبة وبالتالي فإن احتمال النجاح P سيمثل احتمال وجود مصباح معيب والمساوي لـ 0.10 كما سبق الإشارة إليه. وبما انه يتم سحب 11 مصباح من إنتاج المصنع فإن المتغير العشوائي يأخذ القيم التالية:

$$X = 0,1,2,3,4,5,6,7,8,9,10,11$$

وبالتالي يمكن التعبير عن المتغير العشوائي كما يلي:

$$X \sim Bin(11, 0.10)$$

وعليه فإن الاحتمالات المطلوبة يمكن حسابها كما يلي:

1) احتمال أن تكون جميع المصابيح معيبة مساوي لاحتمال أن يكون عدد المصابيح المعيبة مساوي لـ 11 مصباح،

$$\Pr(X=11)=\binom{11}{11}(0.10)^{11}(1-0.10)^{11-11}$$
$$=\frac{11!}{11!\,(11-11)!}(0.10)^{11}(0.90)^{0}$$
$$=1(0.10)^{11}(1)\approx 0$$

2) احتمال أن يكون على الأقل مصباح واحد صالح مساوي لاحتمال أن يكون على الأكثر مصباح واحد معيب،

$$\Pr(X\le 10)=\Pr(X=0)+\Pr(X=1)+\Pr(X=2)+$$
$$\Pr(X=3)+\Pr(X=4)+\Pr(X=5)+$$
$$\Pr(X=6)+\Pr(X=7)+\Pr(X=8)+$$
$$\Pr(X=9)+\Pr(X=10)$$

وحيث أن مجموع الاحتمالات يساوي واحد صحيح

$$\sum_{x=0}^{11}\Pr(X=x)=1$$

لذلك يمكن حساب الاحتمال المطلوب بالطريقة التالية:

$$\Pr(X\le 10)=1-\Pr(X=11)$$
$$=1-\left[\binom{11}{11}(0.10)^{11}(1-0.10)^{11-11}\right]$$
$$\approx 1-0=1$$

ومن نتائج الفقرتين السابقتين نستنتج أن العينة المكونة من 11 مصباح سوف تحوي على الأقل مصباح واحد سليم باحتمال مساوي تقريبا للواحد الصحيح، أي أن احتمال أن تكون كلها معيبة سيكون مساوياً للصفر.

عندما يتحدد التوزيع الاحتمالي لأي متغير فإنه يمكن حينئذ حساب كل من التوقع والتباين والانحراف المعياري للمتغير العشوائي، كما سبق الإشارة إليه في الفصل السابق. وعليه فإن التوقع والتباين والانحراف المعياري للمتغير العشوائي X والذي يتبع توزيع ذي الحدين، يمكن حسابها باستخدام الصيغ التالية:

توقع توزيع ذي الحدين

$$\mu=E(X)=np \qquad (8)$$

تباين توزيع ذي الحدين

$$\sigma^2 = np(1-p) \qquad (9)$$

الانحراف المعياري لتوزيع ذي الحدين

$$\sigma = \sqrt{nP(1-P)} \qquad (10)$$

في المثال السابق كانت معالم التوزيع تساوي $n = 11$ و $P = 0.10$، لذلك فإنه يمكن حساب كل من التوقع والتباين والانحراف المعياري لمتغير العشوائي حيث يتم الحصول على التقديرات التالية:

$$\mu = E(X) = np = 11(0.10) = 1.1$$
$$\sigma^2 = nP(1-P) = 11(0.10)(0.90) = 0.99$$
$$\sigma = \sqrt{nP(1-P)} = \sqrt{0.99} = 0.995$$

مثال 2-1-3-2

قامت شركة بدراسة مدى التزام السفن التجارية بمواعيد الوصول إلى ميناء مدينة معينة، فوجدت انه من بين كل 100 سفينة قادمة يصل 75 منها في الموعد المحدد مسبقاً والباقي يصل متأخراً. فإذا كانت الشركة في انتظار 8 سفن تجارية تحمل لها بضائع من مصادر مختلفة، فأوجد التالي:

أ) احتمال أن تصل جميعها في موعدها.

ب) احتمال أن يكون عدد السفن المتأخرة 2 أو اقل.

ج) احتمال أن يكون عدد السفن المتأخرة أكثر من 6.

د) التوقع والتباين والانحراف المعياري لعدد السفن المتأخرة من بين الثماني سفن المتوقع وصولها.

الحـل

في البداية يلاحظ التالي:

1) هناك ثمان محاولات من واقع أن عدد السفن المتوقع وصولها مساوي لثمانية.

2) هناك احتمالان لكل محاولة، وهما إما أن تصل السفينة في موعدها أو أن تكون متأخرة.

3) المحاولات مستقلة، أي انه لا توجد علاقة بين مواعيد وصول الرحلات، وبالأخص كون الرحلة متأخرة أم لا، بالإضافة إلى أن احتمال أن تصل أي رحلة في موعدها ثابت ويساوي 0.75.

افترض أن المتغير العشوائي X يمثل عدد السفن التي تصل في موعدها. وعليه فإن احتمال النجاح يساوي احتمال أن تصل السفينة في موعدها،

$$P = 0.75$$

وبالتالي فإن عدد السفن التي تصل في موعدها X يمثل متغير عشوائي يتبع توزيع ذي الحدين بمعالم $n=8$ و $P=0.75$،

$$X \sim Bin(8,0.75)$$

هذه المعلومة ستكون حجر الأساس في عملية إيجاد الاحتمالات المطلوبة.

أ) احتمال أن تصل جميعها في موعدها مساوي لاحتمال أن يكون عدد السفن التي تصل في موعدها مساوي لـ 8،

$$\Pr(X=8)=\binom{8}{8}(0.75)^8(1-0.75)^{8-8}=0.10011$$

ب) احتمال أن يكون عدد السفن المتأخرة 2 أو اقل مساوي لاحتمال أن يكون عدد السفن التي تصل في موعدها 6 أو أكثر،

$$\Pr(X\geq 6)=\Pr(X=6)+\Pr(X=7)+\Pr(X=8)$$
$$=\binom{8}{6}(0.75)^6(1-0.75)^{8-6}+\binom{8}{7}(0.75)^7(1-0.75)^{8-7}$$
$$+\binom{8}{8}(0.75)^8(1-0.75)^{8-8}$$
$$=0.12458+0.26697+0.10011=0.49166$$

ج) احتمال أن يكون عدد السفن المتأخرة أكثر من 6 مساوي لاحتمال أن يكون عدد السفن التي تصل في موعدها اقل من 2،

$$\Pr(X<2)=\Pr(X=1)+\Pr(X=0)$$
$$=\binom{8}{1}(0.75)^1(1-0.75)^7+\binom{8}{0}(0.75)^0(1-0.75)^8$$
$$=0.00037+0.00002=0.00039$$

د) توقع المتغير العشوائي X،

$$\mu=E(X)=np=8(0.75)=6$$

والتباين والانحراف المعياري للمتغير العشوائي X،

$$\sigma^2=np(1-p)$$
$$8(0.75)(1-0.75)=1.5$$
$$\sigma=\sqrt{np(1-p)}=\sqrt{1.5}=1.22$$

2-3-2 توزيع بواسون

يعتبر توزيع بواسون واحد من أهم التوزيعات المتقطعة بجانب توزيع ذي الحدين، حيث يكثر استخدام توزيع بواسون في مجالات العلوم الإدارية والإنتاج. وكما هي الحال في التوزيعات، يستخدم توزيع بواسون لإيجاد الاحتمالات المرافقة للقيم التي يمكن أن يأخذها المتغير العشوائي المتقطع X وذلك بافتراض توفر بعض الشروط وهي:

(1) أن يكون هناك احتمالين مرافقين لكل حالة أو محاولة، يمثل الاحتمال الأول منهما احتمال النجاح بينما يمثل الاحتمال الأخر احتمال الفشل.

(2) أن يمثل المتغير العشوائي المتقطع X عدد الحوادث (نجاح أو فشل) التي تحدث خلال فترة زمنية معينة أو خلال محيط معين، مثل المكان أو الحجم أو الوقت أو الطول ... الخ.

(3) أن يأخذ المتغير العشوائي قيم موجبة صحيحة تبدأ من صفر إلى ما لانهاية.

وفي حالة تحقق هذه الشروط فإنه يمكن أن يستخدم توزيع بواسون لإيجاد الاحتمالات المرافقة لقيم المتغير العشوائي X، والتي يمكن التعبير عنها رياضيا كما يلي:

$$X \sim Poi\,(\lambda)$$

يستخدم هذا التعبير للدلالة على أن المتغير العشوائي المتقطع X يتبع توزيع بواسون وان معلمة التوزيع الوحيدة هي λ، حيث تمثل توقع المتغير العشوائي X. وبهدف إيضاح فكرة توزيع بواسون نستعرض الأمثلة التالية:

افترض أننا نرغب بدراسة عدد الحوادث التي تحدث في طريق معين. يلاحظ هنا أن المتغير العشوائي الذي يمثل عدد الحوادث يعتبر متغير متقطع، وذلك لان عدد الحوادث لا بد أن يكون عدداً صحيحاً وبالتحديد صفر أو واحد أو اثنين .... الخ. نلاحظ كذلك انه توجد حالتين لكل رحلة أو محاولة (رحلة سيارة) وهما أن يكون هناك حادث أو أن لا يكون. لاحظ كذلك انه لا يوجد حد أعلى يقيد القيم التي يأخذها المتغير العشوائي المتقطع X، وعليه فإنه يمكن القول بان المتغير العشوائي X يتبع توزيع بواسون. في هذه الحالة يتوجب علينا تحديد قيمة معلمة التوزيع λ والتي تمثل توقع أو متوسط عدد الحوادث اليومية على الطريق المعين. لاحظ أيضا انه يوجد تقيد آخر على المتغير العشوائي X، وهو تقيد زمني، حيث أن المتغير X محصور في فترة زمنية محددة متمثلة في يوم واحد. وبالنسبة لقيمة معلمة التوزيع قد تكون معلومة مسبقا من خلال دراسات سابقة أو قد يتم تقديرها من خلال الدراسة نفسها.

في بعض الدراسات الطبية يتطلب الأمر معرفة عدد كريات الدم الحمراء في حجم مليمتر مكعب من الدم. يلاحظ هنا أن التقييد هو تقييد في الحجم وليس في عدد الكريات التي تمثل المتغير العشوائي X، والذي يأخذ القيم الصحيحة من صفر إلى ما لانهاية.

عدد الوحدات المعيبة المنتجة بواسطة آلة معينة في خلال ساعة واحدة من عملية الإنتاج يمثل متغير عشوائي متقطع يتبع توزيع بواسون. لاحظ هنا أن المتغير العشوائي يأخذ قيماً من صفر إلى غير محدد، بشرط أن تكون أعداد صحيحة. لاحظ كذلك أن احتمال أن تكون أي وحدة تم إنتاجها بواسطة الآلة معيبة ثابت ومتساوي لكل الوحدات المنتجة، لذلك يمكن القول بان الوحدات المنتجة مستقلة. في كل محاولة، والتي تمثل وحدة منتجة، يوجد هناك احتمالين، إما أن تكون الوحدة المنتجة معيبة أو أن تكون سليمة. لذا فإن المتغير العشوائي المتقطع والذي يمثل عدد الوحدات المعيبة يتبع توزيع بواسون،

$$X \sim Poi\,(\lambda)$$

حيث تمثل λ متوسط عدد الوحدات المعيبة المنتجة من الآلة خلال ساعة إنتاج واحدة. يمثل الزمن عنصر التقييد في هذه الحالة حيث حدد ساعة إنتاج واحدة للآلة يتم خلالها إحصاء عدد الوحدات المعيبة المنتجة. تلعب هذه

النقطة المهمة دور فعال في عملية التميز بين تطبيقات توزيع بواسون وتطبيقات توزيع ذي الحدين. فمثلا لو كان التقييد في عدد الوحدات المنتجة وليس الزمن، أي أنه تم اختيار عدد n من الوحدات المنتجة ويراد معرفة التوزيع الاحتمالي لعدد الوحدات المعيبة في الوحدات المنتجة المختارة، فإنه في هذه الحالة يعتبر المتغير العشوائي المتقطع X والذي يمثل عدد الوحدات المعيبة تابع لتوزيع ذي الحدين، أي

$$X \sim Bin(n,p)$$

حيث يمثل كل من n و P معالم التوزيع.

يمكن إيجاد الاحتمالات المرافقة لقيم المتغير العشوائي X الذي يتبع توزيع ذي الحدين، ولكن بافتراض أن قيمتي المعلمتين معلومتين. في بعض الحالات تكون قيمة n كبيرة جدا، مما يصعب عملية إيجاد الاحتمالات المرغوبة باستخدام توزيع ذي الحدين، لذلك فإنه يمكن استخدام توزيع بواسون كتقريب لتوزيع ذي الحدين، ولكن تحت شروط معينة سوف نتطرق إليها لاحقاً.

توزيع بواسون

$$\Pr(X=x)=\frac{\lambda^x e^{-\lambda}}{x!} \tag{11}$$

حيث:

x = 0 ، 1 ، 2 ، ....

λ = متوسط أو توقع المتغير العشوائي X

e = العدد الطبيعي ويساوي تقريبا 2.71828183

$x!$ = مضروب x $\left[x!=x(x-1)(x-2)\cdots 1\right]$

مثال 1-2-3-2

افترض أن متوسط عدد المراجعين لبنك معين خلال فترة نصف ساعة يساوي 5 مراجعين. أوجد الاحتمالات التالية:

أ) احتمال أن يكون عدد المراجعين خلال نصف ساعة 6 مراجعين.

ب) احتمال أن يكون عدد المراجعين خلال نصف ساعة اقل من 3 مراجعين.

الحل

في البداية نلاحظ أن المتغير العشوائي الذي يمثل عدد المراجعين للبنك خلال نصف ساعة يمثل متغير عشوائي متقطع يأخذ قيماً موجبة صحيحة تتراوح قيمها بين الصفر والحد الأعلى لاستيعاب البنك للمراجعين خلال نصف ساعة، والذي يفترض أن يكون كبيراً نوعاً ما أو بالأصح مجهول القيمة ولا يمكن الجزم بمساواته لرقم محدد ثابت. يلاحظ كذلك أن التقييد زمني ويمثل نصف ساعة، وأن الأحداث المتوقعة هي إما أن يأتي مراجع أو أن لا يأتي، لذلك فإنه يمكن افتراض أن

المتغير العشوائي المتقطع X يتبع توزيع بواسون. قبل إيجاد الاحتمالات المطلوبة، لا بد أولا من إيجاد قيمة معلمة توزيع بواسون λ، والتي غالباً ما تكون معطاة كما في هذا المثال حيث تم تحديد متوسط عدد المراجعين للبنك خلال نصف ساعة بـ 5 مراجعين،

$$\lambda = 5$$

بما أن التوزيع محدد وقيمة معلمة التوزيع معروفة فإنه يمكن إيجاد الاحتمالات المطلوبة باستخدام دالة توزيع بواسون. ولحساب الاحتمال المطلوب في الفقر الأولى (فقرة أ) نجري الحسابات التالية:

$$\begin{aligned}\Pr(X = 6) &= \frac{\lambda^6 e^{-\lambda}}{6!} \\ &= \frac{(5)^6 e^{-5}}{6\ !} \\ &= \frac{(5)^6 (2.71828183)^{-5}}{6\times5\times4\times3\times2\times1} = 0.146223\end{aligned}$$

ملاحظة: يمكن استخدام الآلات الحاسبة العلمية البسيطة للقيام بالعمليات الحسابية السابقة حيث يرمز عادة للعدد الطبيعي بـ $\boxed{e}$ أو بالرمز $\boxed{Exp}$ ويرمز لمضروب العدد بالرمز $\boxed{!}$.

بالنسبة للفقرة الثانية من السؤال (فقرة ب) يتم حساب الاحتمال المطلوب بصورة تجميعية، حيث المطلوب هو أن يكون عدد المراجعين اقل من ثلاثة، أي صفر أو واحد أو اثنين. وبما أن المحاولات التي تمثل المراجعين مستقلة عن بعض فإنه يمكن جمع الاحتمالات الثلاثة للحصول على الاحتمال المطلوب.

$$\begin{aligned}\Pr(x < 3) &= \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) \\ &= \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} \\ &= \frac{5^0 e^{-5}}{0!}\left(1 + \frac{5}{1} + \frac{5^2}{2}\right) = 0.124652\end{aligned}$$

مثال 2-2-3-2

افترض أن متوسط عدد الحفر الصغيرة والناتجة عن تأخر الصيانة في شوارع مدينة كبرى يقدر بـ 2 حفرة لكل كيلومتر. إذا تم اختيار جزء طوله كيلومتر من شارع تم اختياره عشوائيا من بين شوارع المدينة، فأوجد التالي:

أ) احتمال أن يكون الشارع خالي من الحفر.

ب) احتمال أن يكون عدد الحفر أكثر من أو يساوي 3.

ج) احتمال أن يكون هناك حفرتين فقط.

الحـــل

من الواضح أن المتغير العشوائي المتقطع، والذي يمثل عدد الحفر في الكيلومتر الواحد في شوارع المدينة يمثل متغير عشوائي متقطع يتبع توزيع بواسون بمعلمة $\lambda = 2$، وعليه فإن الاحتمالات المطلوبة يمكن حسابها كالتالي:

أ) احتمال أن يكون الشارع خالي من الحفر مساوي لاحتمال أن يكون عدد الحفر صفر،

$$\Pr(X = 0) = \frac{2^0 e^{-2}}{0!} = 0.135335$$

ب) احتمال أن يكون عدد الحفر أكثر من أو يساوي ثلاثة مساوي لاحتمال أن يكون عدد الحفر أربعة أو خمسة أو ستة أو ..... الخ. وبما أن مجموع الاحتمالات الكلية يساوي واحد صحيح، إذا يمكن حساب الاحتمال السابق عن طريق طرح الاحتمالات المكملة من الواحد صحيح، أي إن الاحتمال المطلوب يساوي واحد صحيح مطروح منه احتمال أن يكون عدد الحفر 0 أو 1 أو 2.

$$\begin{aligned}
\Pr(X \geq 3) &= 1 - \Pr(X < 3) \\
&= 1 - \left[\Pr(X = 2) + \Pr(X = 1) + \Pr(X = 0)\right] \\
&= 1 - \left[\frac{2^2 e^{-2}}{2!} + \frac{2^1 e^{-2}}{1!} + \frac{2^0 e^{-2}}{0!}\right] \\
&= 1 - \left[\frac{2^0 e^{-2}}{0!}\left(\frac{2^2}{2} + \frac{2}{1} + 1\right)\right] \\
&= 1 - 0.676676 = 0.323324
\end{aligned}$$

ج) احتمال أن يكون عدد الحفر 2 فقط

$$\Pr(X = 2) = \frac{2^2 e^{-2}}{2!} = 0.270671$$

يتم حساب التوقع، المتوسط، وكل من التباين والانحراف المعياري للمتغير العشوائي المتقطع X والذي يتبع توزيع بواسون باستخدام الدوال المبينة في القواعد التالية:

توقع توزيع بواسون

$$\mu = E(X) = \lambda \qquad (12)$$

تباين توزيع بواسون

$$\sigma^2 = \lambda \qquad (13)$$

الانحراف المعياري لتوزيع بواسون

$$\sigma = \sqrt{\lambda} \qquad (14)$$

من النقاط المهمة التي يتميز بها توزيع بواسون هي إمكانية استخدامه كتقريب لإيجاد الاحتمالات المرافقة لقيم متغير عشوائي متقطع يتبع توزيع ذي الحدين وذلك في حالة كون عدد المحاولات n كبير $(n > 100)$ وكون

الاحتمال (احتمال النجاح) صغير $(P < 0.05)$. في هذه الحالة يمكن استخدام توزيع بواسون عوضا عن توزيع ذي الحدين وذلك بهدف تسهيل العمليات الحسابية المرافقة لإيجاد الاحتمالات المطلوبة. ولإجراء عملية التقريب المتمثلة باستخدام توزيع بواسون بدلا من توزيع ذي الحدين، يلزم في البداية تقدير معلمة توزيع بواسون λ والتي ستساوي حاصل ضرب معلمتي توزيع ذي الحدين، أو بالأحرى توقع توزيع ذي الحدين،

$$\lambda = E(X) = np \tag{15}$$

مثال 3-2-3-2

تقوم شركة أدوية باستخدام أجهزة لإنتاج أدويتها وتعبئتها في علب صغيرة، ونظرا لخطورة هذا الدواء فإن الشركة تشترط أن يكون عدد العبوات المعيبة والتي يكون فيها تركيز الدواء أكثر من المسموح به اقل من 5 عبوات من بين كل 1000 عبوة. فإذا كان احتمال أن تنتج الآلة عبوة معيبة يساوي 0.0002 وتم اختيار عينة من 1000 عبوة تم إنتاجها بواسطة الآلة، فكم احتمال أن تقبل العينة؟

الحـــل

بما أن المتغير العشوائي X والذي يمثل عدد العبوات المعيبة يمثل متغير متقطع، وبما أن هناك عينة تتكون من 1000 عبوة، تمثل عدد المحاولات، وفي كل محاولة يحتمل حدوث إحدى حدثين: إما أن تكون العبوة صالحة أو معيبة، والعبوات أو المحاولات مستقلة. بالإضافة إلى أن احتمال أن تكون أي عبوة أو محاولة معيبة ثابت. لذلك فإننا نستطيع القول بأن المتغير العشوائي المتقطع X والذي يمثل عدد العبوات المعيبة يتبع توزيع ذي الحدين،

$$X \sim Bin(1000.0.0002)$$

وحيث أن المطلوب هو احتمال أن تقبل العينة، أي احتمال أن يكون عدد العبوات المعيبة اقل من 5 من بين الألف عبوه، إذا نوجد الاحتمال كما يلي:

$$\begin{aligned}
\Pr(X<5) &= \Pr(X=4)+\Pr(X=3)+\Pr(X=2)\\
&\quad+\Pr(X=1)+\Pr(X=0)\\
&=\binom{1000}{4}(0.0002)^4(0.9998)^{1000-4}+\\
&\quad\binom{1000}{3}(0.0002)^3(0.9998)^{1000-3}+\\
&\quad\binom{1000}{2}(0.0002)^2(0.9998)^{1000-2}+\\
&\quad\binom{1000}{1}(0.0002)^1(0.9998)^{1000-1}+\\
&\quad\binom{1000}{0}(0.0002)^0(0.9998)^{1000-0}\\
&=0.0000543+0.0010890+0.0163645\\
&\quad+0.1637756+0.8187144=0.9999978
\end{aligned}$$

يلاحظ هنا صعوبة إجراء العمليات الحسابية لكبر الأرقام، خاصة عند استخدام الآلة الحاسبة اليدوية، ولتلافي هذه المشكلة يمكننا استخدام توزيع بواسون حيث أن $n > 100$ و $p < 0.05$. في هذه الحالة نستطيع القول أن المتغير العشوائي المتقطع X يتبع تقريبا توزيع بواسون بمعلمة يتم تحديد قيمتها من خلال الدالة التالية:

$$\lambda = nP = 1000(0.0002) = 0.2$$

ومن ثم يتم إيجاد الاحتمال المطلوب السابق كما يلي:

$$\begin{aligned}
\Pr(X < 5) &= \sum_{x=0}^{4} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{(0.2)^0 e^{-0.2}}{0!} + \frac{(0.2)^1 e^{-0.2}}{1!} + \frac{(0.2)^2 e^{-0.2}}{2!} \\
&\quad + \frac{(0.2)^3 e^{-0.2}}{3!} + \frac{(0.2)^4 e^{-0.2}}{4!} \\
&= \frac{(0.2)^0 e^{-0.2}}{0!}\left(1 + \frac{0.2}{1} + \frac{0.2^2}{2!} + \frac{0.2^3}{3!} + \frac{0.2^4}{4!}\right) \\
&= 0.8187308(1.2214) = 0.9999978
\end{aligned}$$

يلاحظ انه في المثال السابق تم الحصول على نفس الناتج باستخدام كل من توزيع ذي الحدين وتوزيع بواسون، مع أن الجهد والعمليات الحسابية المطلوبة لاستخدام توزيع بواسون أسهل واقل بكثير من الجهد والعمليات الحسابية المطلوبة عند استخدام توزيع ذي الحدين. لذلك يفضل دائما إذا تحقق الشرطين $(n > 100)$ و $(P < 0.05)$ أن نستخدم توزيع بواسون كتقريب لتوزيع ذي الحدين. أيضا في بعض الأحيان قد يحتم المطلوب استخدام توزيع بواسون، مثل أن يكون المطلوب إيجاد احتمال أن يكون عدد العبوات التالفة أو المعيبة مساوي لـ 60 عبوة من بين 100 عبوة تم اختيارها عشوائيا. في هذه الحالة إذا أردنا استخدام توزيع ذي الحدين لإيجاد الاحتمال المطلوب فإنه يتوجب إجراء الحسابات التالية:

$$\begin{aligned}
\Pr(X = 60) &= \binom{1000}{60}(0.0002)^{60}(0.9998)^{1000-60} \\
&= \frac{1000 \times 999 \times 998 \times \cdots \times 941}{60 \times 59 \times 58 \times \cdots \times 1}(0.0002)^{60}(0.9998)^{1000-60}
\end{aligned}$$

بينما تكون العملية في حالة استخدام توزيع بواسون كما يلي:

$$\Pr(X = 60) = \frac{(0.2)^{60} e^{-0.2}}{60!}$$

ملاحظة : يمكن استخدام الآلات الحاسبة العلمية العادية لإيجاد مضروب 60، ولكن معظم الآلات الحاسبة العلمية العادية المتوفرة حاليا لا تستطيع إيجاد مضروب الأرقام الكبيرة مثل الرقم 1000، حيث لا تعطى الناتج المطلوب لمثل هذه العمليات، حيث أن معظم الآلات الحاسبة تستطيع فقط التعامل مع الأرقام التي لا تتجاوز 10^{99}.

4-2 توزيعات احتمالية لمتغير عشوائي متصل

عندما يكون المتغير العشوائي متقطعاً فإنه يأخذ قيماً محـــدده، بمعنـــى أن قيمـــة المتغيـــر العشـــوائي لا يمكـــن أن تأخذ قيماً تقريبية أو قيماً قريبة من القيم المحـــددة أو حتـــى قـــيم مختلفـــة عـــن القـــيم المحـــددة. فمـــثلا عنـــدما نقـــول أن المتغير العشوائي المتقطع يمثل عدد أطفال الأسرة فإن قيم هـــذا المتغيـــر العشـــوائي ســـوف تكـــون صـــفر أو 1 أو 2 أو 3 ...الخ.

ولأننا نطلق على المتغير صفة عشوائي، فإنه يتوقـــع أن يكـــون هنـــاك احتمـــال مرافـــق لكـــل قيمـــة يمكـــن أن يأخذها المتغير العشوائي، وهذه الاحتمالات غالباً ما يمكن التعبير عنهـــا بواســـطة دالـــة التوزيـــع الاحتمـــالي. فمـــثلا فـــي الفقرة السابقة تطرقنا إلى كل من دالة التوزيع الاحتمـــالي لتوزيـــع ذي الحـــدين وتوزيـــع بواســـون، وكـــل منهمـــا يعتبـــر توزيع احتمالي لمتغير عشوائي متقطع. من ناحيـــة أخـــرى، قـــد يكـــون المتغيـــر العشـــوائي متصـــل، أي أن القـــيم التـــي يمكن أن يأخذها المتغير العشوائي يتم تحديدها من خلال فترات أو فترة تمثـــل النطـــاق الـــذي يمكـــن أن تقـــع فيهـــا هـــذه القيم. في هذه الحالة يكون عدد القيم التي يمكن أن يأخذها المتغيـــر العشـــوائي المتصـــل عـــدد كبيـــر جـــداً وغيـــر محـــدد (مالا نهاية)، وذلك بغض النظر عن طول أو قصر الفترة المحددة للمتغير العشوائي المتصل.

من الأمثلة على المتغيرات العشوائية المتصلة، طول شـــخص مـــا بالســـنتيمتر. فمـــثلا قـــد يقـــال أن طـــول هـــذا الشخص 170 سم بينما في الحقيقة قد يكون الطول الحقيقي قيمة قريبة جدا مـــن 170 ســـم. وفـــي الواقـــع، تعتبـــر محاولـــة معرفة الطول الحقيقي عملية مستحيلة لأن الشخص يتغير طولـــه مـــع الـــزمن، ولـــو بجـــزء صـــغير جـــدا، كجـــزء مـــن البليون من السنتيمتر. تعتبر درجة طالب فـــي امتحـــان معـــين، كمثـــال آخـــر، متغيـــر عشـــوائي متصـــل، حيـــث تكـــون الدرجة غالباً محدده بأن تقع بين صفر و 100 درجـــة. كـــذلك جـــرى العـــرف أيضـــا علـــى أن تكـــون الدرجـــة المعطـــاة للطالب عدداً صحيحاً، مثلا 76 من 100، بينما نظرياً يمكن إعطاء الطالـــب أي درجـــة حتـــى ولـــو كانـــت تحتـــوى علـــى جزء صغير جدا مما يؤدى إلى أن عدد القيم التي يحتمـــل أن يأخـــذها الطالـــب، وبالتـــالي المتغيـــر العشـــوائي، نظريـــا لا يساوي فقط 100 قيمة بل عدد لانهائي من القيم.

على عكس المتغيـــر العشـــوائي المتقطـــع، لا يمكـــن أن توجـــد معادلـــة أو دالـــة احتماليـــة للمتغيـــر العشـــوائي المتصل تقوم بإعطاء الاحتمالات المرافقة لكل القيم التي يمكن أن يأخذها المتغيـــر العشـــوائي. ويرجـــع الســـبب فـــي ذلـــك إلى أن عدد قيم المتغير العشوائي المتصل يكون دوما ما لانهاية، وبالتـــالي يصـــبح احتمـــال أن يأخـــذ المتغيـــر العشـــوائي المتصل أي من هذه القيم مساوياً للصفر. لهذا السبب دعـــت الحاجـــة إلـــى إيجـــاد أســـلوب آخـــر يختلـــف عـــن أســـلوب المتغير العشوائي المتقطع لتحديد الاحتمالات المرافقة لقيم المتغير العشوائي المتصـــل. هـــذا الأســـلوب يتمثـــل فـــي عمليـــة إيجاد احتمال أن يقع المتغير العشوائي في فترة جزئية من الفتـــرة الكليـــة. فمـــثلا، نعلـــم أن احتمـــال أن يأخـــذ طالبـــاً مـــا درجة بين الصفر و 100 في امتحان ما يساوي واحد صحيح،

$$\Pr(0 < X < 100) = 1$$

وهذا افتراض مسلّم به ومقبول منطقياً. لذلك فإنه يمكن إيجـــاد احتمـــال أن يأخـــذ الطالـــب درجـــة تقـــع بـــين أي قيمتـــين تنحصران بين صفر و 100 درجة. فمثلا يمكن إيجاد احتمال أن يأخـــذ الطالـــب درجـــة بـــين 20 و 60 أو درجـــة بـــين 80 و 82، وهذه الاحتمالات يمكن التعبير عنها رياضياً كما يلي:

$$\Pr(20 < X < 60)$$

و

$$\Pr(80 < X < 82)$$

لتوضيح هذه الفكرة نورد المثال التالي:

مثال 2-4-1

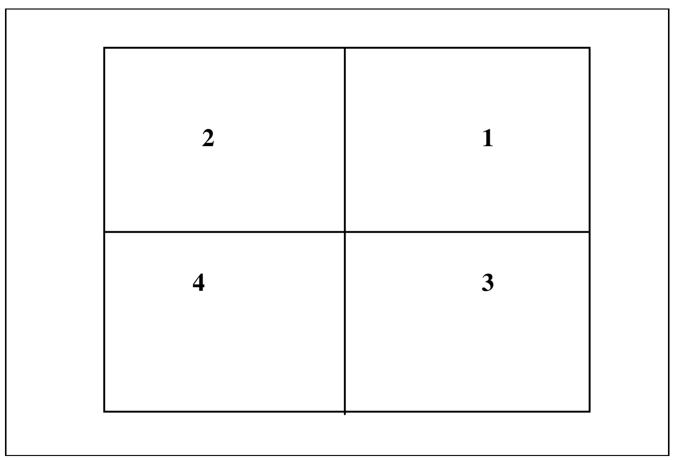
طُلب من شخص معصوب العينين القيام بالتصويب باستخدام مسدس ذو طلقة واحدة على لوحة مقسمة إلى أربع أجزاء متساوية كما هو مبين في شكل (2-4-1). فإذا كنا نفترض أن الشخص سوف يصيب اللوحة بالتأكيد، فأوجد الاحتمالات التالية:

أ) احتمال أن يصيب الجزء الرابع.

ب) احتمال أن يصيب الخط الواقع بين الجزء الأول والثاني.

ج) احتمال أن يصيب الجزء الأول أو الثاني.

شكل 2-4-1
لوحة تصويب مقسمة إلى أربعة أجزاء متساوية



الحل

أ) بما أن الشخص الذي يقوم بالتصويب معصوب العينين، ويفترض انه سوف يصيب اللوحة، فإن احتمال أن يصيب اللوحة سوف يكون واحد صحيح. وبالتالي، فإن احتمال أن يصيب الجزء الرابع سوف يكون مساوي لـ 0.25، وذلك لأن اللوحة مقسمة إلى أربعة أجزاء متساوية، وفرصة أن يصاب أي جزء متساوية. هذا الاحتمال مقبول نظرياً، لان مجموع الاحتمالات للأجزاء الأربعة يساوي الواحد صحيح.

ب) بما أن احتمال أن يصيب الشخص أي جزء وارد، فإن احتمال أن يصيب الشخص الخط الواقع بين الجزئيين الأول والثاني سوف يكون مساوي لنسبة المساحة التي يشغلها هذا الخط إلى المساحة للوحة كاملة. ولأن هذه النسبة تكون صغيرة جدا، على افترض أن الخط رفيع جدا، فإن الاحتمال سوف يكون مساوي للصفر أو قيمة قريبة جدا من الصفر.

ج) لإيجاد احتمال أن يصيب الشخص الجزء الأعلى من اللوحة والذي يمثل الجزء الأول والثاني، نوجد نسبة المساحة التي يشغلها الجزء الأعلى إلى المساحة الكلية والتي تساوي $\frac{2}{4}$ أو 0.50. وعليه فإن احتمال أن يصيب الشخص الجزء الأعلى سيساوي 0.50.

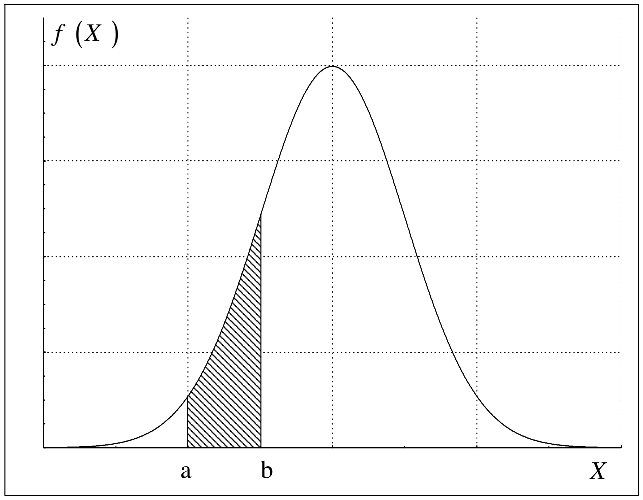
وبناء على الفكرة الواردة في المثال السابق، نستنتج أن الاحتمالات المتعلقة بفترات يقع فيها المتغير العشوائي المتصل، يتم حسابها عن طريق حساب المساحات لهذه الفترات من ثم حساب نسبتها إلى المساحة الإجمالية للفترة الكلية، بالإضافة إلى أن الاحتمال المرافقة للخطوط والنقاط ستكون قريبة من الصفر، والتي سيفترض فعلاً أنها تساوي الصفر. هذه الافتراضات سوف تكون صحيحة دائما طالما أن المتغير العشوائي X يمثل متغير عشوائي متصل وليس متقطع.

عندما يكون المتغير العشوائي مستمراً (متصلاً)، فإن الأمر يتطلب معادلة رياضية لتمثيل هذا المتغير العشوائي، $f\ (X)$. وفي حالة التمكن من تقدير أو إيجاد صيغة هذه الدالة، فإنه يمكن تمثيلها بيانياً، ومن ثم استخدامها في إيجاد الاحتمالات المرافقة للفترات التي يقع فيها المتغير العشوائي المتصل، ولكن على شرط أن تكون قيمة هذه الدالة موجه دائماً وان تكون المساحة تحت منحنى الدالة مساوي للواحد الصحيح، كما في الشكل (2-4-2).

وباستخدام منحنى دالة المتغير العشوائي المستمر يتم إيجاد أي احتمال لوقوع المتغير في فترة محددة. فمثلا، لإيجاد احتمال أن يقع المتغير العشوائي X بين القيمتين $(a\ ,\ b)$ نوجد المساحة تحت منحنى الدالة، $f\ (X)$، للمتغير العشوائي X، وذلك بين القيمتين a و b. وبما أن المساحة الكلية تساوي واحد صحيح، فإن المساحة المحسوبة تحت المنحنى بين القيمتين a و b ستمثل الاحتمال المطلوب. تتم عملية حساب المساحات تحت المنحنيات باستخدام التكامل المعروف رياضيا. وعليه فإن عملية حساب احتمال أن تقع قيمة المتغير العشوائي X بين القيمتين a و b يمكن تمثيلها رياضياً كما يلي:

$$\Pr(a < X < b) = \int_a^b f\ (x)\ dx$$

وللتمييز بين دالتي الاحتمال للمتغير العشوائي المستمر والمتقطع، فإننا سوف نطلق على دالة الاحتمال للمتغير العشوائي المتصل أو المستمر بدالة كثافة الاحتمال. وهذه الدالة لن تعطي الاحتمالات المرافقة للقيم التي يمكن أن يأخذها المتغير العشوائي X مباشرة، كما هو الحال للمتغير العشوائي المتقطع، بل أنها سوف تمثل النقاط على منحنى دالة المتغير العشوائي حيث يتم حساب المساحة تحت المنحنى ليمثل الاحتمال المرافق للفترة التي قد يقع فيها المتغير العشوائي المتصل.



شـــكل 2-4-2
منحنى دالة التوزيع الاحتمالي للمتغيـــر العشوائي المستمر أو المتصل (دالـــة التوزيع الطبيعي)

1-4-2 التوزيع الطبيعي

واحد من أهم التوزيعات للمتغير العشوائي المتصل هو التوزيع الطبيعـــي والـــذي يعتبـــر، فـــي الواقـــع، الدعامـــة الأساسية لمعظم النظريات الإحصائية وأساليب التحليل الإحصائي. ونظـــرا لأهميـــة هـــذا التوزيـــع خاصـــة فـــي الفصـــول القادمة فإننا سوف نحاول الإسهاب في تغطية جوانب هذا التوزيـــع. يعتمـــد التوزيـــع الطبيعـــي علـــى معلمتـــين، متوســـط التوزيع μ وتباينه σ^2.

دالة كثافة الاحتمال للتوزيع الطبيعي

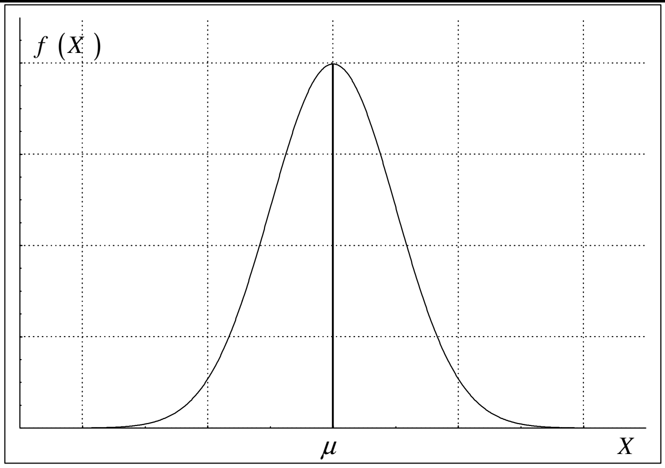
$$f\left(x:\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (16)$$

حيث :

$$\sigma^2>0 \quad ، \quad -\infty<\mu<\infty \quad ، \quad -\infty<X<\infty$$

$$\pi\approx 3.1415926 \quad ، \quad e\approx 2.7182818285$$

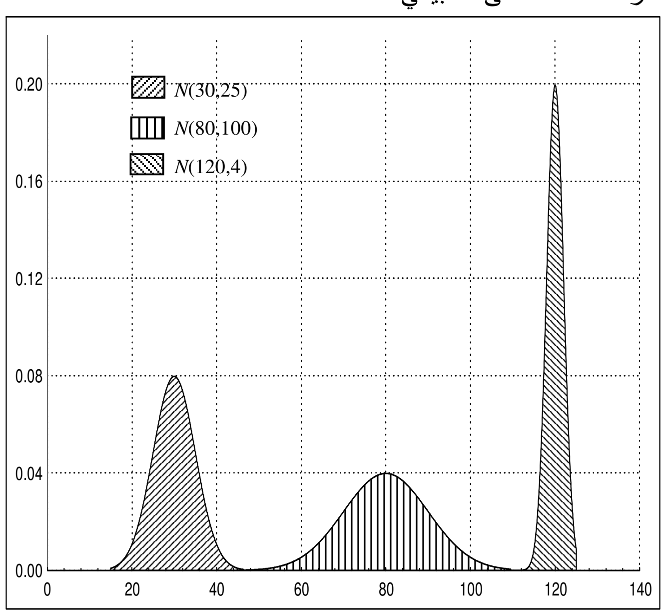
ويتميز التوزيع الطبيعي بمزايا خاصة، من أهمها انه عند رســـم دالـــة كثافـــة الاحتمـــال للتوزيـــع الطبيعـــي فإننـــا نجد أن شكل المنحنى يكون ناقوسي ومتماثل حول الوسط الحســـابي. أي أن الجهـــة التـــي علـــى يمـــين الوســـط أو التوقـــع μ تنطبق على الجهة التي على يساره، كما هو موضح في شكل (2-4-1-1). بالإضـــافة إلـــى ذلـــك فـــإن المســـاحة تحـــت المنحنى دائما سوف تساوي واحد صحيح. وبالتـــالي، فـــإن المســـاحة علـــى يمـــين التوقـــع أو الوســـط μ ستســـاوي 0.5، وذلك لأن التوزيع متماثل حول الوسط.



شكل 2-4-1-1
منحنى التوزيع الطبيعي

من المميزات المهمة أيضا للتوزيع الطبيعي هي أن التوقع أو الوسط الحسابي μ يكون دائماً مساوياً لكل من قيمتي الوسيط والمنوال للمتغير العشوائي X الذي يتبع التوزيع الطبيعي. ويمثل الوسيط القيمة التي تقع في المنتصف، أي القيمة التي يكون نصف القراءات اقل منها والنصف الآخر أكثر منها، بينما يمثل المنوال القيمة الأكثر شيوعا أو تكرارا. من النقاط المهمة التي يجدر الإشارة إليها أيضاً هي أن طرفي منحنى التوزيع الطبيعي لا يجب أن تلامس المحور الأفقي وإنما تمتد إلى ما لانهاية، مع مراعاة أن إجمالي المساحة تحت المنحنى دائما تكون مساوية للواحد الصحيح، وان الانحراف المعياري σ يحدد شكل تفلطح التوزيع أو المنحنى، حيث انه كلما زادت قيمة σ كلما زاد المنحنى تفلطحاً، كما هو موضح في الشكل (2-4-1-2).

إن الهدف الأساسي من عملية إيجاد التوزيع الاحتمالي، وبالتالي إيجاد المنحنى، يتمثل في إيجاد الاحتمالات المرافقة للمتغير العشوائي. وحيث أن المتغير العشوائي متصل، فإن احتمال أن يأخذ المتغير العشوائي المتصل قيمة محدده سوف يكون دائما مساوي للصفر كما تم الإشارة إليه سابقا. لذلك فإن التركيز سينصب على عملية إيجاد احتمال أن يقع المتغير العشوائي المتصل X في فترة معينة والذي، بالطبع، سيتم تقديره بالمساحة المحصورة في تلك الفترة تحت المنحنى الطبيعي.



شكل 2-4-1-2
علاقة شكل المنحنى الطبيعي بالانحراف المعياري والوسط الحسابي

تمثل ميزة التماثل واحدة من أهم المزايا للمنحنى الطبيعي، حيث أن المساحات تحت المنحنى الطبيعي لفترات متماثلة تكون متساوية، وأن الوسط الحسابي μ يقع في منتصف المنحنى ويقسمه إلى نصفين متماثلين. من المزايا المهمة أيضا لمنحنى التوزيع الطبيعي، هي كون المساحة تحت المنحنى الطبيعي لفترات مقاسه بوحدة الانحراف المعياري σ معلومة. فمثلا 68.26% من القيم التي يمكن أن يأخذها المتغير العشوائي X والذي يتبع التوزيع الطبيعي سوف تقع بين $(\mu+\sigma)$ و $(\mu-\sigma)$ وذلك بغض النظر عن قيمتي μ و σ.

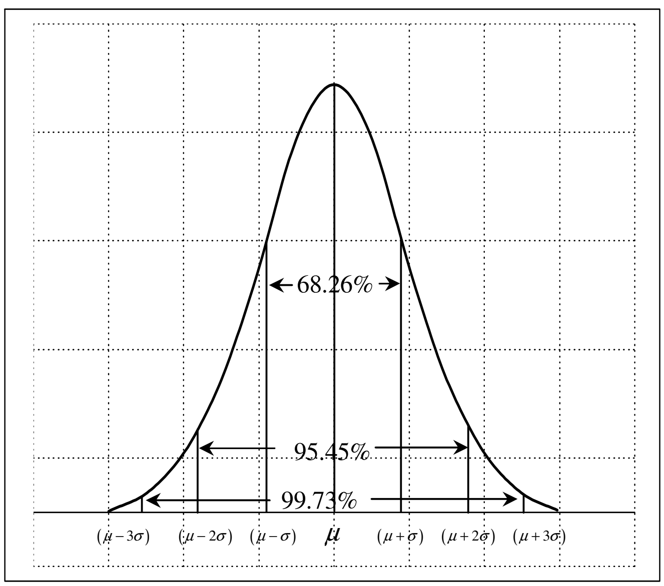
يوضح الشكل (2-4-1-3) الفكرة السابقة مع بيان الاحتمالات أو المساحات المرافقة لثلاث فترات هي:

1) المساحة المحصورة بين $(\mu+\sigma)$ و $(\mu-\sigma)$ تساوي 68.26%
2) المساحة المحصورة بين $(\mu+2\sigma)$ و $(\mu-2\sigma)$ تساوي 95.45%
3) المساحة المحصورة بين $(\mu+3\sigma)$ و $(\mu-3\sigma)$ تساوي 99.73%

غالبا ما يرمز للمتغير العشوائي المتصل X، والذي يتبع التوزيع الطبيعي بوسط حسابي أو توقع μ وتباين σ^2، بالرمز التالي:

$$X \sim N\left(\mu,\sigma^2\right)$$

حيث μ و σ^2 تمثلان معالم التوزيع الطبيعي، والتي يجب أن تكون قيمتيهما معلومتين لكي نتمكن من استخدام دالة كثافة الاحتمال لإيجاد احتمالات مرافقة لفترات معينة.



شكل 2-4-1-3
علاقة المنحنى الطبيعي بوسط μ وتباين σ^2 بالمساحات لفترات محددة

عندما تكون قيمة الوسط الحسابي μ، للمتغير العشوائي الطبيعي، مساوية للصفر والانحراف المعياري مساوياً لواحد صحيح، فإن التوزيع يأخذ صورة خاصة مهمة جداً، حيث يطلق على دالة الاحتمال للمتغير العشوائي بدالة كثافة الاحتمال المعياري. كما أن المتغير العشوائي يستخدم له حرف آخر مختلف عن X وهو Z، لتمييز

هذه الحالة الخاصة المهمة عن غيرها من الحالات الأخرى. ويسـمى المتغير العشـوائي Z فـي هـذه الحالـة بـالمتغير العشوائي المعياري حيث يتبع التوزيع الطبيعي المعياري،

$$Z \sim N\,(0,1)$$

دالة كثافة الاحتمال للتوزيع الطبيعي المعياري

$$f\,(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \qquad (17)$$

حيث :

$$-\infty < Z < \infty$$

$$\pi \approx 3.1415926 \quad \text{،} \quad e \approx 2.7182818285$$

مثال 2-4-1-1

إذا كان Z يمثل متغير عشوائي طبيعي معياري،

$$Z \sim N\,(0,1)$$

فأوجد التالي:

أ) $\Pr(-1 < Z < 1)$

ب) $\Pr(-2 < Z < 2)$

ج) $\Pr(-3 < Z < 3)$

الحـل

باسـتخدام الخاصـية السـابقة للتوزيـع الطبيعـي والمبينـة فـي الشـكل (2-4-1-3) نسـتطيع إيجـاد الاحتمالات المطلوبة كما يلي:

أ)

$$\begin{aligned}\Pr(-1 < Z < 1) &= \Pr\big((0-1) < Z < (0+1)\big)\\ &= \Pr\big((\mu-\sigma) < Z < (\mu-\sigma)\big)\\ &= 0.6826\end{aligned}$$

ب)

$$\begin{aligned}\Pr(-2 < Z < 2) &= P\,r\big((0-2) < Z < (0+2)\big)\\ &= \Pr\big((\mu-2\sigma) < Z < (\mu+2\sigma)\big)\\ &= 0.9545\end{aligned}$$

ج)

$$\Pr(-3 < Z < 3) = \Pr\big((0-3) < Z < (0+3)\big)$$

$$= \Pr\left(\left(\mu - 3\sigma\right) < Z < \left(\mu + 3\sigma\right)\right)$$
$$= 0.9973$$

ونظرا لأن كل من المتغير العشوائي الطبيعي X والمتغير العشوائي الطبيعي المعياري Z يعتبران متغيران متصلان، فإن عملية إيجاد احتمال أن يقع المتغير العشوائي المتصل في فترة محددة تتم عن طريق إيجاد تكامل دالة كثافة الاحتمال للفترة والتي تمثل المساحة تحت المنحنى والمحصورة بين حدي الفترة،

$$\Pr\left(a < X < b\right) = \int_a^b f\left(x\right)dx \qquad (18)$$

بالنسبة للتوزيع الطبيعي المعياري، يتم حساب التكامل لأي فترة باستخدام جدول خاص وضع خصيصاً لإيجاد الاحتمالات المرافقة للفترات التي يقع فيها المتغير العشوائي الطبيعي المعياري Z، جدول 1 في الملحقات. ولأن التوزيع الطبيعي المعياري هو أيضاً توزيع متماثل فإن الجدول يهتم بإيجاد المساحة على يسار أي قيمة موجبة للمتغير العشوائي Z. وبهدف التبسيط والإيضاح سيتم الرمز للمساحة على يسار القيمة المعيارية تحت المنحنى الطبيعي المعياري بالرمز ϕ والذي يمكن التعبير عنه رياضياً بالتالي:

$$\phi(a) = \Pr\left(-\infty < Z < a\right) \qquad \forall \; a \geq 0$$

مثال 2-1-4-2

أوجد احتمال أن تكون قيمة المتغير العشوائي الطبيعي المعياري Z اقل من 1.39، أي أوجد الاحتمال التالي

$$\Pr\left(Z < 1.39\right)$$

الحـل

يمثل الاحتمال المطلوب هنا مباشرة القيمة $\phi(1.39)$ والتي يتم الحصول عليها من جدول 1 في الملحقات (جدول التوزيع الطبيعي المعياري). تبداء عملية إيجاد الاحتمال المطلوب بتحديد العدد الصحيح و أول رقم عشري وهو 1.3 على العامود الأول في الجدول والمخصص لقيم Z إلى الرقم العشري الأول فقط، ثم نتحرك أفقياً، على اليمين، إلى أن نجد القيمة الموجودة في العامود المرافقة للقيمة العشرية الثانية في قيمة المتغير Z، وهي 0.09. هذه القيمة سوف تمثل المساحة، تحت منحنى التوزيع الطبيعي المعياري على يسار القيمة 1.39 والتي تساوي 0.9177. لذلك، يمكن القول بأن الاحتمال المطلوب يساوي:

$$\phi(1.39) = \Pr\left(Z < 1.39\right) = 0.9177$$

وبحكم كون التوزيع الطبيعي المعياري متماثل حول الصفر (الوسط الحسابي) بالإضافة إلى أن جدول التوزيع الطبيعي المعياري يعطي فقط المساحات على يمين قيم المتغير العشوائي المعياري والممثلة لاحتمال أن تكون قيمة المتغير العشوائي المعياري اقل من تلك القيمة المحددة،

$$\phi(a) = \int_{-\infty}^{a} f\left(Z\right)dz$$

لذلك توجد قواعد يمكن الاعتماد عليهــا للوصــول إلــى الاحتمــالات المختلفــة المتعلقــة بــالتوزيع الطبيعــي المعيــاري. ويمكن إبراز القواعد المعنية من خلال اســتخدام قيمتــين مــوجبتين ($a \geq 0$ و $b \geq a$) فــي الــدوال التاليــة والتــي ســيتم إيضاحها من خلال الأمثلة والتطبيقات اللاحقة.

$$\Pr(Z > a) = 1 - \phi(a) \tag{19}$$

$$\Pr(a < Z < b) = \phi(b) - \phi(a) \tag{20}$$

$$\Pr(Z < -a) = 1 - \phi(a) \tag{21}$$

$$\Pr(Z > -a) = \phi(a) \tag{22}$$

$$\Pr(-a < Z < b) = \phi(a) + \phi(b) - 1 \tag{23}$$

$$\Pr(-b < Z < -a) = \phi(b) - \phi(a) \tag{24}$$

وفيما يلي يتم إبراز الإثبات الرياضي لكل قاعدة من القواعد السابقة.

أولا: إثبات القاعدة المبينة في المعادلة (19)،

$$\begin{aligned}
\Pr(Z > a) &= \Pr(a < Z < \infty) \\
&= 1 - \Pr(-\infty < Z < a) \\
&= 1 - \phi(a)
\end{aligned}$$

ثانياً: إثبات القاعدة المبينة في المعادلة (20)،

$$\Pr(a < Z < b) = \Pr(-\infty < Z < b) - \Pr(-\infty < Z < a)$$

$$= \phi(b) - \phi(a)$$

ثالثاً: إثبات القاعدة المبينة في المعادلة (21)،

$$\begin{aligned}
\Pr(Z < -a) &= \Pr(-\infty < Z < -a) \\
&= \Pr(a < Z < \infty) \\
&= 1 - \Pr(-\infty < Z < a) \\
&= 1 - \phi(a)
\end{aligned}$$

رابعاً: إثبات القاعدة المبينة في المعادلة (22)،

$$\begin{aligned}
\Pr(Z > -a) &= \Pr(-a < Z < \infty) \\
&= 1 - \Pr(-\infty < Z < -a) \\
&= 1 - \Pr(a < Z < \infty) \\
&= 1 - \left[1 - \Pr(-\infty < Z < a)\right] \\
&= 1 - 1 + \Pr(-\infty < Z < a) \\
&= \Pr(-\infty < Z < a) \\
&= \phi(a)
\end{aligned}$$

خامساً: إثبات القاعدة المبينة في المعادلة (23)،

$$\begin{aligned}
\Pr(-a<Z<b) &= \Pr(-\infty<Z<b)-\Pr(-\infty<Z<-a)\\
&= \Pr(-\infty<Z<b)-\Pr(a<Z<\infty)\\
&= \Pr(-\infty<Z<b)-\left[1-\Pr(-\infty<Z<a)\right]\\
&= \Pr(-\infty<Z<b)-1+\Pr(-\infty<Z<a)\\
&= \Pr(-\infty<Z<b)+\Pr(-\infty<Z<a)-1\\
&= \phi(b)+\phi(a)-1
\end{aligned}$$

سادساً: إثبات القاعدة المبينة في المعادلة (24)،

$$\begin{aligned}
\Pr(-b<Z<-a) &= \Pr(-\infty<Z<-a)-\Pr(-\infty<Z<-b)\\
&= \Pr(a<Z<\infty)-\Pr(b<Z<\infty)\\
&= \left[1-\Pr(-\infty<Z<a)\right]-\left[1-\Pr(-\infty<Z<b)\right]\\
&= 1-\Pr(-\infty<Z<a)-1+\Pr(-\infty<Z<b)\\
&= \Pr(-\infty<Z<b)-\Pr(-\infty<Z<a)\\
&= \phi(b)-\phi(a)
\end{aligned}$$

مثال 2-4-1-3

إذا كان Z يمثل متغير عشوائي طبيعي معياري، فأوجد الاحتمال التالي:

$$\Pr(-1.39<Z<0)$$

الحـــل

يمكن إيجاد الاحتمال المطلوب كما يلي،

$$\begin{aligned}
\Pr(-1.39<Z<0) &= \Pr(Z<0)-\Pr(Z<-1.39)\\
&= \phi(0)-\left(1-\phi(1.39)\right)
\end{aligned}$$

مثال 2-4-1-4

أوجد احتمال أن تكون قيمة المتغير العشوائي الطبيعي المعياري Z بين القيمتين 1.12 و 0.98–، أي أوجد الاحتمال

$$\Pr(-0.98<Z<1.12)$$

الحـــل

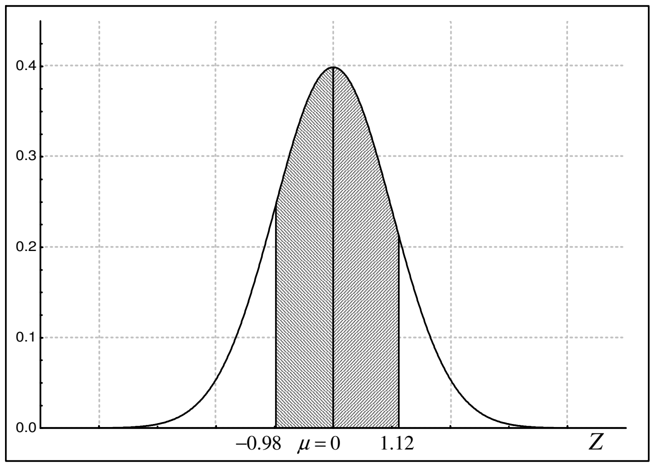
لأننا نتعامل مع متغير عشوائي طبيعي معياري، Z، فإنه يمكن استخدام الجدول 1 في الملحقات لإيجاد الاحتمال المطلوب. ولكن في هذه الحالة يتطلب الأمر فصل الاحتمال إلى جزئيين، يمثل كل جزء فترة تكون إحدى أطرافها قيمة الوسط الحسابي والتي تساوي صفر، كما هو مبين في الشكل (2-4-1-4). وعليه فإنه يمكن حساب الاحتمال المطلوب باستخدام القاعدة في المعادلة (21) كما يلي:

$$\begin{aligned}\Pr(-0.98 < Z < 1.12) &= \Pr(Z < 1.12) - \Pr(Z < -0.98)\\ &= \phi(1.12) - \phi(-0.98)\\ &= \phi(1.12) - (1 - \phi(0.98))\\ &= 0.8686 - (1 - 0.8365)\\ &= 0.8686 - 0.1635 = 0.7051\end{aligned}$$

أي انه باحتمال 70.51% فإن قيمة المتغير العشوائي Z الطبيعي المعياري سوف تقع بين 1.12 و -0.98.

شكل 2-4-1-4

المساحة المحصورة تحت منحنى التوزيع الطبيعي المعياري للفترة (-0.98 ، 1.12)



مثال 2-4-1-

بافتراض أن

$$Z \sim N(0,1)$$

أوجد الاحتمالات التالية:

أ) $\Pr(0.07 < Z < 2.08)$

ب) $\Pr(-1.75 < Z < -0.75)$

الحل

بنفس الأسلوب المتبع في حل المثال السابق، يمكن تجزئة الاحتمال المطلوب إلى فترتين أو جزئيين، وباستخدام القاعدتين المبينتين في المعادلتين (20) و (24) ويتم حساب الاحتمالات المطلوبة كما يلي:

أ) لحساب احتمال أن يقع المتغير العشوائي Z في الفترة $(0.07, 2.08)$، نقوم بالتقسيم التالي:

$$\Pr(0.07 < Z < 2.08) = \Pr(Z < 2.08) - \Pr(Z < 0.07)$$

والذي يمثل الاحتمال المطلوب، الشكل (2-4-1-5)، لاحظ هنا أن قيمة الطرف الأيسر للفترة لا يساوي الصفر ولو انه قريب من قيمة الصفر. ولإيجاد المساحة بين نقطتين موجبتين يتوجب إيجاد الفرق بين المساحتين المرافقتين للقيمتين الموجبتين [القاعدة (20)]،

$$\Pr(0.07 < Z < 2.08) = \Pr(Z < 2.08) - \Pr(Z < 0.07)$$

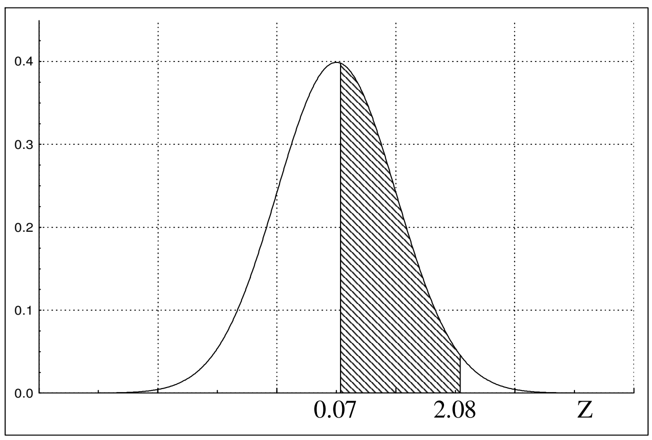
$$= \phi(2.08) - \phi(0.07)$$
$$= 0.9812 - 0.5279$$
$$= 0.4533$$

ب) الشكل (2-4-1-6) يوضح المساحة المطلوبة لحساب احتمال أن يقع المتغير العشوائي Z في الفترة (0.75- و 1.75-). ولحساب هذا الاحتمال نستخدم القاعدة (24) والتي تهتم بإيجاد المساحة بين قيمتين سالبتين. لذلك فإن الاحتمال المطلوب يمكن حسابه كما يلي:

$$\Pr(-1.75 < Z < -0.75) = \phi(1.75) - \phi(0.75)$$
$$= 0.9599 - 0.7734 = 0.1865$$

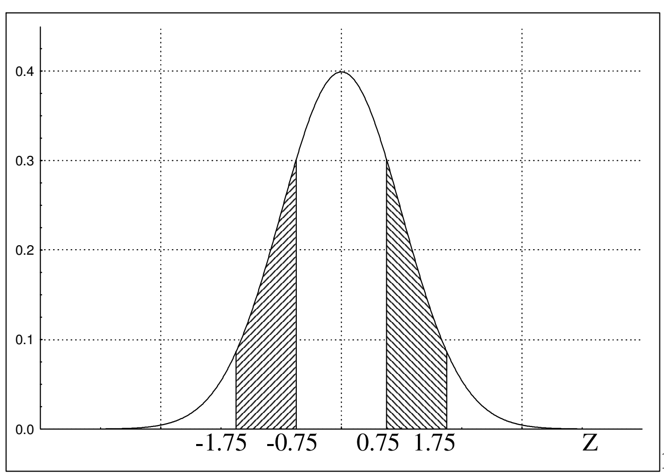
شكل 2-4-1-5

المساحة المحصورة تحت منحنى التوزيع الطبيعي المعياري للفترة (2.05 ، 0.07)



شكل 2-4-1-6

المساحة المحصورة تحت منحنى التوزيع الطبيعي المعياري بين الرقمين 0.75 ، 1.75 والرقمين – 0.75 ، -1.75



مثال 2-4-1

بافتراض أن

$$Z \sim N(0,1)$$

أوجد الاحتمالات التالية:

1) $\Pr(Z < 1.8)$

2) $\Pr(Z > 1.8)$

3) $\Pr(Z < -1.8)$

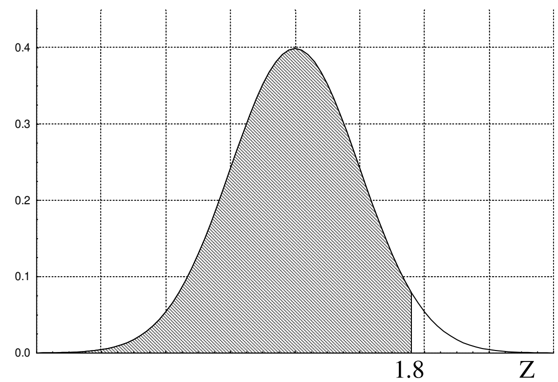
4) $\Pr(Z > -1.8)$

الحــل

(1

$$\begin{aligned}\Pr(Z < 1.8) &= \phi(1.8)\\ &= 0.9641\end{aligned}$$

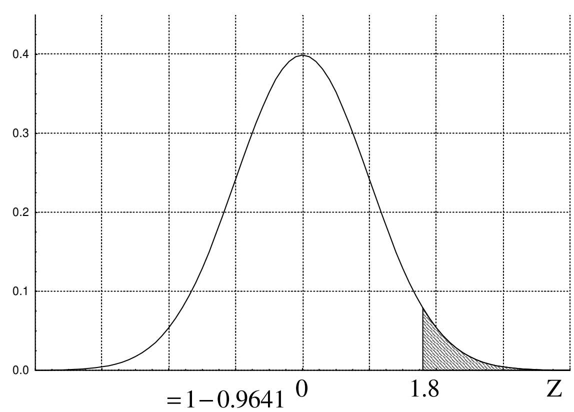
الشكل التالي يبين الاحتمال الناتج،



(2

$$\begin{aligned}\Pr(Z > 1.8) &= 1-\phi(1.8)\\ &= 1-0.9641\\ &= 0.0359\end{aligned}$$

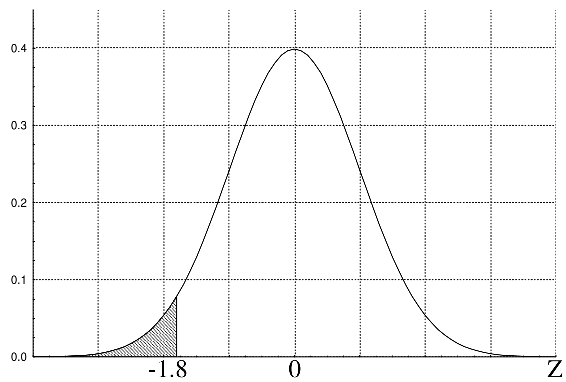
الشكل التالي يبين الاحتمال الناتج،



(3

$$\begin{aligned}&= 1-0.9641\\ &= 0.0359\end{aligned}$$

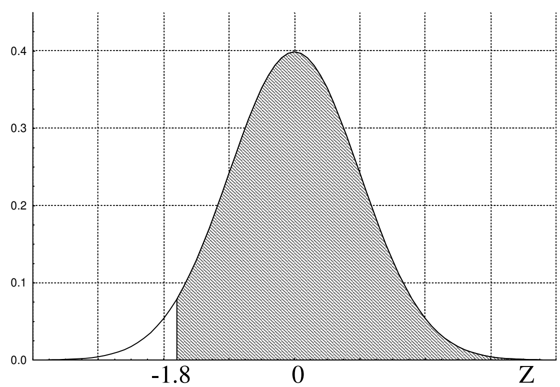
الشكل التالي يبين الاحتمال الناتج،



(4

$$\Pr(Z > -1.8) = \phi(1.8)$$
$$= 0.9641$$

الشكل التالي يبين الاحتمال الناتج،



في الأمثلة السابقة تم تغطية بعض الحالات التي تتعلق بالمتغير العشوائي المستمر Z، والذي يتبع التوزيع الطبيعي المعياري ويشكل حالة خاصة من التوزيع الطبيعي. وحيث أن قيم كل من الوسط الحسابي والتباين للمتغير العشوائي الطبيعي المعياري، Z، تكون محددة ومعروفة القيم، وهما صفر كوسط حسابي أو توقع و 1 كانحراف معياري. ولكن في الحياة العملية، توجد كثير من المتغيرات التي يفترض أنها تتبع التوزيع الطبيعي، مثل الطول والوزن والمعدل التراكمي وغيرها، وهذه المتغيرات غالبا لا تتميز بمزايا المتغير العشوائي الطبيعي المعياري Z، أي أن الوسط الحسابي أو التوقع لا يساوي صفر والتباين لا يساوي واحد، بل تساوي أرقام أخرى على خط الأعداد الحقيقية.

ولإيجاد احتمال أن يقع متغير عشوائي طبيعي X، بمتوسط μ وتباين σ^2، في فترة معينة، نستخدم إحدى طريقتين: يتم في الطريقة الأولى إيجاد الاحتمال عن طريق إيجاد تكامل دالة كثافة الاحتمال للتوزيع الطبيعي. فمثلا لإيجاد احتمال أن يقع المتغير العشوائي X في الفترة (a,b) نوجد التكامل التالي (المعادلة 18):

$$\Pr(a < X < b) = \int_a^b f(X)\,dx$$

$$= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

حيث تمثل كل من μ و σ معالم التوزيع الطبيعي وتكون قيمتيهما معلومتين.

تعتمد الطريقة الثانية على فكرة الاستفادة من جدول التوزيع الطبيعي المعياري، والذي يعتبر أسهل وأسرع من عملية إيجاد الاحتمال بطريقة التكامل. وحيث أن جدول التوزيع الطبيعي المعياري قد صمم خصيصا لإيجاد الاحتمال للمتغير العشوائي الطبيعي المعياري Z والذي يتميز بأن توقعه أو وسطه الحسابي مساوي للصفر وتباينه مساوي للواحد الصحيح، فإنه يتوجب، أولا، تحويل قيم المتغير العشوائي X، والذي يتبع التوزيع الطبيعي بمتوسط μ وتباين σ^2، إلى قيم معيارية، تمثل المتغير العشوائي المعياري Z، وذلك لكي يمكن استخدام جدول 1 في الملحقات لإيجاد الاحتمالات المطلوبة.

تتم عملية التحويل من متغير عشوائي طبيعي X إلى متغير عشوائي طبيعي معياري Z (الدرجة المعيارية) من خلال العلاقة أو التحويل التالي:

$$Z = \frac{X - \mu}{\sigma} \tag{25}$$

أي انه يتم طرح التوقع، μ، من قيمة المتغير العشوائي X، ومن ثم يتم قسمة الناتج على الانحراف المعياري σ، وبذلك تتحول قيمة المتغير العشوائي الطبيعي X إلى صورة متغير عشوائي طبيعي معياري Z. لإيضاح هذه الفكرة، افترض إننا نتعامل مع متغير عشوائي طبيعي X له توقع μ وتباين σ^2، وأننا نريد إيجاد احتمال أن يقع المتغير العشوائي X في الفترة (a,b) ولإيجاد الاحتمال المطلوب، لابد أولا من إجراء التحويل، ومن ثم يتم استخدام جدول 1 في الملحقات لإيجاد الاحتمال المطلوب. الشكل (2-4-1-7) يبين العلاقة بين المتغيرين Z و X.

$$\Pr(a < X < b) = \Pr\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$
$$= \Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

لاحظ هنا أن كل من a، b، μ و σ قيم معلومة، وبالتالي فإن القيم

$$\frac{a-\mu}{\sigma} \quad , \quad \frac{b-\mu}{\sigma}$$

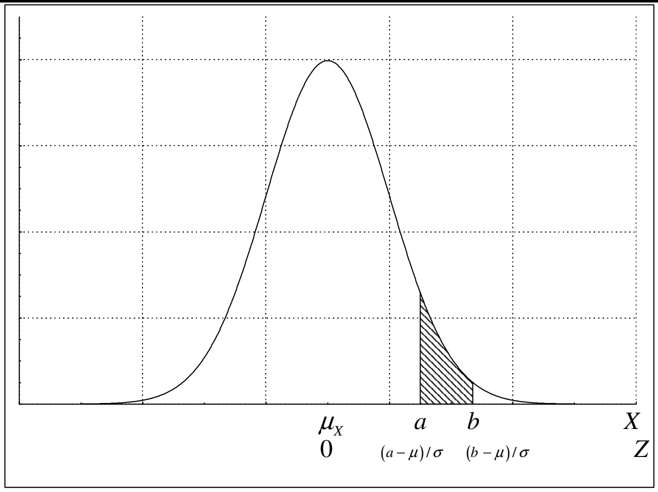
ستكون أيضاً معلومة. أما بالنسبة للمقدار

$$\frac{X-\mu}{\sigma}$$

فإن الناتج سيمثل متغير عشوائي وليس قيمة ثابتة، وذلك لأن X متغير عشوائي، وتحويل متغير عشوائي ينتج عنه متغير عشوائي أيضا في حال كون التحويل خطي.

شـــكل 2-4-1-7

العلاقة الاحتماليـــة بين المتغيرين Z و X



مثال 2-4-1-7

إذا كانت الأجور الشهرية للعمال فـــي مدينـــة الريـــاض تتبـــع التوزيـــع الطبيعـــي بمتوســـط 800 ريـــال وانحراف معياري 150 أي،

$$X \sim N\left(800\ ,\ (150)^2\right)$$

حيث X تمثل الأجر الشهري لأي عامل في مدينـــة الريـــاض. وإذا تـــم اختيـــار عامـــل عشـــوائياً مـــن مدينة الرياض، فأوجد الاحتمالات التالية:

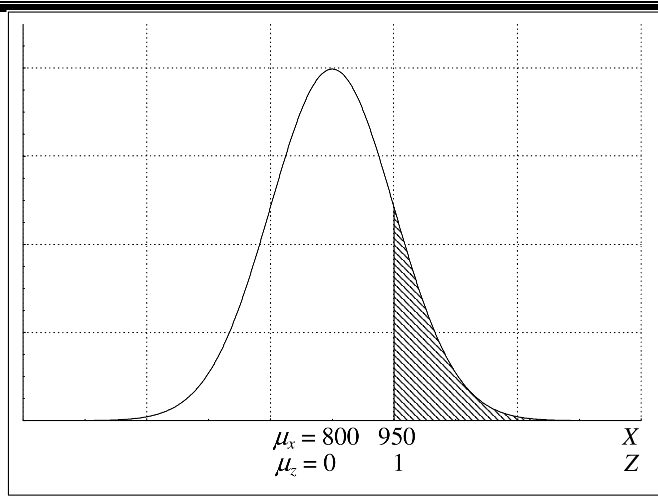
1) أن يكون اجر العامل أكثر من 950 ريال.
2) أن يكون اجر العامل اقل من 1200 ريال.
3) أن يكون اجر العامل بين 850 و 770 ريال.

الحـــل

1) لإيجاد احتمال أن يكون اجر العامل الـــذي تـــم اختيـــاره عشـــوائيا أكثـــر مـــن 950 ريـــال، يتطلـــب الأمر، أول، تحويل القيمة المعطاة إلى قيمة معيارية، ومن ثم يتم إيجاد الاحتمال كما يلي:

$$\begin{aligned}\Pr(X > 950) &= \Pr\left(\frac{X-\mu}{\sigma} > \frac{950-\mu}{\sigma}\right)\\ &= \Pr\left(Z > \frac{950-800}{150}\right)\\ &= \Pr(Z > 1)\\ &= 1-\phi(1)\\ &= 1-0.8413 = 0.1587\end{aligned}$$

الشكل (2-4-1-8) يوضح عملية التحويل والاحتمال المطلوب في الفقرة السابقة.



شـــكل 8-1-4-2

احتمال أن يكـــون الأجـــر الشـــهري لعامـــل اختيـــر عشوائيا أكثر مـــن 950 ريال

2)

بنفس الطريقة المستخدمة لحل الفقرة (1)، يمكن إيجاد احتمال أن يكـــون اجـــر العامـــل اقـــل مـــن 1200 ريال، كما يلي:

$$\begin{aligned}\Pr(X < 1200) &= \Pr\left(\frac{X-\mu}{\sigma} < \frac{1200-\mu}{\sigma}\right)\\ &= \Pr\left(Z < \frac{1200-800}{150}\right)\\ &= \Pr(Z < 2.6)\\ &= \phi(2.6)\\ &= 0.9962\end{aligned}$$

3)

لإيجاد احتمال أن يكون اجر العامل بين 850 و 770 ريال نقوم بالتالي:

$$\begin{aligned}\Pr(770 < X < 850) &= \Pr\left(\frac{770-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{850-\mu}{\sigma}\right)\\ &= \Pr\left(\frac{770-800}{150} < Z < \frac{850-800}{150}\right)\\ &= \Pr\left(\frac{-30}{10} < Z < \frac{50}{150}\right)\\ &= \Pr(-0.20 < Z < 0.33)\\ &= \phi(0.20)+\phi(0.33)-1\\ &= 0.5793+0.6293-1\\ &= 0.2086\end{aligned}$$

معادلات مهمة

القيمة المتوقعة للمتغير العشوائي المتقطع

$$E(X) = \sum_{x} \left[x \Pr(X = x) \right]$$

الانحراف المعياري وتباين المتغير العشوائي المتقطع

$$\begin{aligned}\sigma &= \sqrt{\sigma^2} \\ &= \sqrt{\sum_{x} \left[x^2 \Pr(X - x) \right] - \mu^2} \\ \sigma^2 &= \sum_{x} \left[x^2 \Pr(X = x) \right] - \mu^2\end{aligned}$$

توزيع ذي الحدين

$$\Pr(X = x) = \binom{n}{x} P^x (1-P)^{(n-x)}$$

حيث

n عدد المحاولات.

x عدد مرات النجاح المطلوب إيجاد الاحتمال لها.

P احتمال النجاح و $(1-P)$ احتمال الفشل.

توقع توزيع ذي الحدين

$$\mu = E(X) = np$$

تباين توزيع ذي الحدين

$$\sigma^2 = np(1-p)$$

الانحراف المعياري لتوزيع ذي الحدين

$$\sigma = \sqrt{nP(1-P)}$$

توزيع بواسون

$$\Pr(X=x)=\frac{\lambda^{x}e^{-\lambda}}{x!}$$

حيث:

$x=0,1,2,....$

λ =متوسط أو توقع المتغير العشوائي X

e = العدد الطبيعي ويساوي تقريبا 2.71828183

$x!$ = مضروب x $\left[x!=x(x-1)(x-2)\cdots1\right]$

توقع توزيع بواسون

$$\mu=E(X)=\lambda$$

تباين توزيع بواسون

$$\sigma^{2}=\lambda$$

الانحراف المعياري لتوزيع بواسون

$$\sigma=\sqrt{\lambda}$$

دالة كثافة الاحتمال للتوزيع الطبيعي

$$f(x:\mu,\sigma^{2})=\frac{1}{\sqrt{2\pi\sigma^{2}}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}$$

حيث

$-\infty<X<\infty$ ، $-\infty<\mu<\infty$ ، $\sigma^{2}>0$

$e\approx2.7182818285$ ، $\pi\approx3.1415926$

المساحة تحت المنحنى الطبيعي المعياري

$$\phi(a)=\int_{-\infty}^{a}f(Z)dz$$

تمارين

1)

إذا تم رمي حجري نرد متزنين مرة واحدة، وكان المتغير العشوائي هو مجموع النقاط على السطحين العلويين للحجريين فأوجد:

- التوزيع الاحتمالي للمتغير العشوائي X
- توقع المتغير العشوائي X
- تباين المتغير العشوائي X

2)

اعد حل التمرين السابق (رقم 1) على افتراض أن المتغير العشوائي X يمثل حاصل ضرب الرقمين الناتجيين من رمي الحجريين.

3)

إذا كانت الوحدات المعيبة في إنتاج إحدى المصانع تتبع توزيع ذي الحدين باحتمال يساوي 0.03 فإذا تم سحب عينة عشوائية من إنتاج المصنع حجمها 10 وحدات، احسب ما يلي:

- احتمال عدم وجود أي وحدة معيبة في العينة
- احتمال وجود وحدتين معيبتين في العينة
- احتمال وجود وحدتين على الأكثر معيبتين في العينة
- الوسط الحسابي (التوقع) والانحراف المعياري لعدد الوحدات المعيبة

4)

اذكر أي الجداول التالية تمثل توزيع احتمالياً مع ذكر السبب:

أ

X	$\Pr(X)$
1	$1/8$
0	$1/4$
1	$1/8$
2	$3/8$

ب

X	$\Pr(X)$
-2	0.1
-3	0.4
-4	0.3
-5	0.2

جـ

X	$\Pr(X)$
6	0.25
7	0.25
8	0.125
9	0.25
10	0.125

5)

إذا كان احتمال أن يكون فرد الأسرة في مدينة الرياض ذكراً مساوي لـــ 0.52 وإذا تم اختيار إحدى الأسر بطريقة عشوائية فوجد بها 5 مواليد فاحسب الاحتمالات التالية:

- أن يكون في الأسرة ذكر واحد فقط

- أن يكون في الأسرة أنثى واحدة فقط
- أن يكون في الأسرة ذكر واحد على الأقل

6)

إذا كانت الأخطاء المطبعية في صفحات إحدى الجرائد تتبع توزيع بواسون بمتوسط 4 أخطاء فاحسب لإحدى صفحات هذه الجريدة والتي تم اختيارها بطريقة عشوائية ما يلي:

- احتمال وجود خطاءين مطبعيين
- احتمال وجود خطاءين مطبعين على الأكثر
- احتمال وجود خطاءين مطبعين على الأقل
- الانحراف المعياري لعدد الأخطاء المطبعية

7)

إذا تم تطعيم 500 شخصاً ضد مرض معين وكان احتمال الإصابة بتأثير عكسي رغم التطعيم هو 0.02 فما احتمال:

- إصابة شخص واحد بتأثير عكسي
- إصابة شخص واحد على الأكثر بتأثير عكسي
- إصابة شخص واحد على الأقل بتأثير عكسي

8)

إذا كانت عدد الطائرات التي تطلب الإذن بالهبوط الاضطراري في إحدى المطارات خلال أسبوع واحد متغير عشوائي يتبع توزيع بواسون بمتوسط 3 طائرات. إذا تم اختيار أسبوع بطريقة عشوائية فأوجد الاحتمالات التالية:

- احتمال أن تطلب الإذن بالهبوط الاضطراري 4 طائرات
- احتمال أن تطلب الإذن بالهبوط الاضطراري 4 طائرات على الأكثر
- احتمال أن يكون عدد الطائرات التي تطلب الإذن بالهبوط الاضطراري اقل من 4 طائرات
- الوسط الحسابي والانحراف المعياري لعدد الطائرات التي تطلب الإذن بالهبوط الاضطراري

9)

إذا كان X متغيراً عشوائياً يتبع توزيع ذي الحدين بوسط حسابي $\mu = 2$ وانحراف معياري $\sigma = 1.6$ فأوجد قيمتي معلمتي التوزيع n و P

10)

إذا كانت المكالمة الهاتفية التي تصل إلى سنترال الجامعة خلال مدة زمنية محددة تتبع توزيع بواسون بمعدل يساوي 5 مكالمات فاحسب الاحتمالات التالية لفترة تم اختيارها عشوائيا لنفس السنترال:

- أن يكون عدد المكالمات 3
- أن يكون عدد المكالمات 3 على الأكثر
- أن يكون عدد المكالمات 2 على الأقل

11)

إذا كان الوسط الحسابي لعمر نوع محدد من البطاريات هو 600 ساعة بتباين مساوي لـ 49 وبافتراض أن عمر البطاريات متغير عشوائي يتبع التوزيع الطبيعي احسب الاحتمالات التالية لبطارية تم اختيارها عشوائيا:

- أن تعمل أكثر من 618 ساعة
- أن تعمل مابين 587 و 610 ساعة
- اقل عدد من الساعات التي تعملها أفضل 5% من البطاريات
- أن تعمل 604 ساعة بالضبط

---

12)

إذا كانت درجات الطلاب في إحدى المواد تتبع التوزيع الطبيعي بمتوسط 75 درجة وانحراف معياري 5.4 درجة فأوجد التالي:

- اقل درجة حصل عليها أفضل 10% من الطلاب
- اكبر درجة حصل عليها اضعف 5% من الطلاب
- الدرجتان اللتان تحصران بينهما 50%من الطلاب المتوسطين

---

13)

إذا كانت درجات الطلاب في إحدى المواد تتبع التوزيع الطبيعي بمتوسط 72 درجة وانحراف معياري 5 درجة، إذا أدى الامتحان 120 طالب فأوجد عدد الطلاب الذين:

- تقل درجاتهم عن 70 درجة
- تنحصر درجتهم بين 70 و 80 درجة
- تنحصر درجتهم بين 75 و 82 درجة
- تزيد درجاتهم عن 68 درجة

---

14)

افترض أن نسبة الربح الصافي إلى راس المال في الصفقات التي ينجزها أحد البنوك المحلية، يعتبر متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 0.21 وتباين 0.16 آي:

$$X \sim N(0.21, 0.16)$$

إذا تم اختيار صفقة عشوائيا، فأوجد احتمال أن تتجاوز نسبة الربح في الصفقة 0.17

---

15)

تعتمد إحدى الشركات الكبرى على شبكة الإنترنت في عملية الحصول على المعلومات المتعلقة ببعض العمليات التجارية التي تقوم بها. افترض أن الوقت المستغرق بالدقائق في عملية الحصول على المعلومة المعينة من خلال شبكة الإنترنت متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 45 دقيقة وانحراف معياري 18 دقيقة. إذا رغب أحد العاملين في الشركة باستخدام الشبكة للحصول على معلومة معينة فأوجد احتمال أن يستغرق أكثر من 40 دقيقة؟

16)

تتميز أسواق مدينة الرياض بعدم وجود تسعيرة لأسعار السلع بشكل عام، مما يعطي فرصة للمستهلك للتأثير على الأسعار عن طريق علاقة العرض والطلب، وعليه، يفترض أن يكون سعر سلعة ما متغير من مكان إلى آخر. افترض أن سعر سلعة معينة X متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 57 ريال وتباين 81، أي

$$X \sim N(57,81)$$

استخدم المعلومات السابقة لإيجاد الاحتمالات التالية:

- إذا تم اختيار بطريقة عشوائية معرض يبيع السلعة فأوجد احتمال:
  - أن يكون سعرها اقل من 48
  - أن يكون سعرها بين 50 و 68
- إذا تم بطريقة عشوائية اختيار 73 معرض يبيع السلعة المحددة، فأوجد العدد المتوقع للمعارض التي يكون سعر السلعة فيها أكثر من 48

17)

افترض أن أطوال طلاب المرحلة الابتدائية متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 112 سم وانحراف معياري 25 سم

- إذا تم اختيار طالبا عشوائيا من المستوى الأول فأوجد احتمال أن يكون طول الطالب أكثر من 83 سم.
- إذا كان عدد طلاب المرحلة الابتدائية في مدرسة ما مساوي لـ 525 طالبا فأوجد عدد الطلاب المتوقع الذين تكون أطوالهم اقل من 90 سم.

18)

افترض أن الزمن المستغرق لإنهاء معاملة في أحد الدوائر الحكومية متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 27 ساعة وتباين 100 ساعة.

- إذا تم اختيار معاملة عشوائيا، أوجد احتمال أن يكون الوقت المستغرق لإنجازها اقل من 20 ساعة
- إذا تم اختيار 81 معاملة عشوائيا، فأوجد العدد المتوقع للمعاملات التي يكون الوقت المستغرق لإنجازها بين 25 و 30 ساعة

19)

افترض أن المبيعات الأسبوعية لكل عامل بيع في إحدى الشركات متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 528 ريال وانحراف معياري 81 ريال

- إذا اختير إحدى عمال البيع عشوائيا أوجد احتمال أن تكون مبيعاته الأسبوعية اقل من 400 ريال
- إذا كان عدد عمال البيع في الشركة مساوي لـ 80 عامل فأوجد عدد العمال المتوقع الذين تكون مبيعاتهم بين 485 ريال و 412 ريال

20)

افترض أن أعمار طلاب المستوى الأول في جامعة الملك سعود متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 19 سنه وانحراف معياري 1.2 سنه.

- إذا تم اختيار طالبا عشوائيا من المستوى الأول فأوجد احتمال أن يكون عمره أكثر من 18.5 سنه
- إذا كان عدد الطلاب في المستوى الأول في الجامعة مساوي لــ 8000 طالــب فأوجــد عــدد الطــلاب المتوقــع التــي تكون أعمارهم اقل من 18 سنه.

(21

إذا كانت X تمثل متغير عشوائي يتبع التوزيع الطبيعــي بمتوســط μ و انحــراف معيــاري σ فــإن التحويــل $Z=\dfrac{X-\mu}{\sigma}$ يمثل أيضا متغير عشوائي يتبع التوزيع الطبيعي. ما هي قيمة كل من μ_z و σ_z؟

(22

افترض أن س تمثل متغير عشوائياً يرمز لعدد الحفر فــي الكيلــومتر الواحــد علــى طريــق ســريع. افتــرض كــذلك أن المتغير س يتبع توزيع بواسون بمتوسط 1.2 ، إذا تم اختيار كيلومتر واحــد مــن الطريــق عشــوائياً فأوجــد احتمــال أن يكون هناك أكثر من مطب واحد في الطريــق؟ كــذلك أوجــد احتمــال أن يكــون هنــاك 4 مطبــات بالضــبط علــى أي كيلومتر في الطريق؟

(23

افترض أن متوسط المعدل التراكمي لطلاب جامعة الملك ســعود يتبــع التوزيــع الطبيعــي بمتوســط 2.8 و تبــاين 0.04 . إذا تم اختيار طالب عشوائياً فأوجد احتمال أن يكون معدلهُ اكبر من 3.2

(24

إذا كان سعر السيارات المستعملة يتبع التوزيع الطبيعي بمتوسط 14000 و انحراف معياري 2000 فأوجد

- احتمال أن تباع سيارة مستعملة ما بقيمه بين 12000 و 20000 ريال
- احتمال أن تباع سيارة مستعملة ما بسعر اقل من 10000 ريال

(25

إذا كان عدد الوحدات المعيبة المنتجة بواسطة آلة معينه خــلال يــوم واحــد يتبــع توزيــع بواســون بمتوســط 4 وحــدات معيبة فأوجد

- التوقع و التباين لعدد الوحدات المعيبة المنتجة من الآلة خلال يوم واحد
- إذا اختير يوم عشوائيا فأوجد احتمال أن يكون عدد الوحدات المعيبة اكبر من 5 ولكن اقل من أو يساوي 8

(26

أوجد التوقع و التباين للمتغير العشوائي X إذا كان يتبع التوزيع التالي

X	1	2	3	4	5
$\Pr(X)$	0.2	0.1	0.2	0.1	0.5

(27

إذا كان المتغير العشــوائي Y يتبــع التوزيــع الطبيعــي بتوقــع $E(Y)=23$ و انحــراف معيــاري $\sigma(Y)=3$، فأوجــد احتمال أن تكون قيمة المتغير العشوائي Y اكبر من القيمة 20 وفي نفس الوقت اقل من القيمة 17 .

(28

بافتراض أن المتغير العشوائي X يتبع توزيع ذي الحدين حيث $n=10$ و $P=0.4$. أوجد احتمال أن تكون قيمة المتغير العشوائي X اقل من 4 و اكبر من أو يساوي 2.

(29

افترض أن

$$X \sim N\,(12,9)$$

أوجد

- $\Pr(X > 10)$
- $\Pr(10 < X < 15)$

(30

افترض أن

$$X \sim Poi\,(3.2)$$

أوجد

- $\Pr(X > 2)$
- $E\,(4X - 7)$

(31

إذا كانت درجات الطلاب النهائية في أحد الشعب تتبع التوزيع الطبيعي بمتوسط 74 درجة و انحراف معياري 8 درجات. إذا كان هناك 60 طالب في الشعبة، فكم تتوقع أن يكون عدد الطلاب الناجحين بافتراض أن درجة النجاح هي 60؟

(32

إذا كانت الطلبيات التي يقوم مورد سلعة معينة بطلبها تأتي متأخرة باحتمال 20% . إذا قام هذا المورد بطلب خمس طلبيات فأوجد

- احتمال أن تأتى جميع الطلبيات في وقتها؟
- احتمال أن يكون عدد الطلبيات المتأخرة واحدة فقط؟
- احتمال أن يكون عدد الطلبيات المتأخرة 2 أو أكثر؟
- توقع و تباين المتغير العشوائي الذي يمثل عدد الطلبيات المتأخرة؟

(33

إذا كان الوقت المستغرق لإنجاز معاملة في أحد الدوائر الحكومية يعتبر متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 23 ساعة وتباين 72 ساعة. إذا تم اختيار معاملة عشوائيا فأوجد الاحتمالات التالية

- احتمال أن يكون الوقت المستغرق لإنجازها اقل من 10 ساعات
- احتمال أن يكون الوقت المستغرق لإنجازها اكبر من 20 ساعة

- احتمال أن يكون الوقت المستغرق لإنجازها بين 30 و 20 ساعة

34)

افترض أن احتمال إنتاج وحدة معيبة من آلة معين مساوي لـ 0.07 وافترض انه تم الحصول على 7 وحدات منتجة من الآلة فالوجد الاحتمالات التالية

- أن يكون عدد الوحدات المعيبة ٥ وحدات
- أن يكون عدد الوحدات المعيبة اقل من وحدتين
- توقع عدد الوحدات السليمة (ت(ن – س))

35)

الجدول التالي يمثل متغير عشوائي X والاحتمالات المرافقة للقيم التي يأخذها

X	8	6	5	3
$\Pr(X)$	0.20	0.40	0.10	0.30

- بين لماذا يمثل الجدول السابق توزيع احتمالي؟
- أوجد التوقع الرياضي $E(X)$
- أوجد $E(4X+5)$
- أوجد $E(X^2)$

36)

افترض أن أعمار البيوت المعروضة للبيع في مدينة الرياض متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 12.5 سنة وتباين 17.8 . إذا تم عشوائيا اختيار بيت معروض للبيع في مدينة الرياض فأوجد التالي:

- احتمال أن يزيد عمره عن 16 سنة.
- احتمال أن يكون عمره اقل من 5 سنوات.
- احتمال أن يكون عمره بين 23 و 16.5 سنة.

37)

تبين من دراسة سابقة أن احتمال أن تصل حافلة النقل الجماعي لمحطة محددة قيمة ثابتة ومساوية لـ 0.16 . إذا كان يتوقع أن يصل اليوم إلى المحطة أربعة حافلات، فان عدد الحافلات التي ستصل في الموعد يمثل متغير عشوائي يتبع توزيع ذي الحدين بـ 5 محاولات واحتمال نجاح 0.16

- اكتب الصيغة الرياضية للتوزيع السابق.
- أوجد احتمال أن يكون عدد الحافلات التي تصل في الموعد مساوي لـ 5
- أوجد احتمال أن يكون عدد الحافلات المتأخرة أكثر من 4

38)

افترض أن المتغير العشوائي X يتبع توزيع بواسون بمتوسط 4، أوجد توقع هذا المتغير العشوائي و احتمال أن تكون قيمة X اكبر من 5 وفي نفس الوقت اقل من أو تساوي 8

(39

إذا كان إنتاج آلة معينه يتبع توزيع ذي الحدين باحتمال 10% بأن تكـون الوحـدة المنتجـة معيبـة. إذا تـم اختيـار عينـه من 18 قطعه للفحص، فأوجد

- التوقع و التباين للمتغير العشوائي
- احتمال أن لا تقبل العينة، إذا كانت العينة تقبل إذا كان عدد الوحدات المعيبة اقل من أو تساوي 2
- احتمال أن يكون عدد الوحدات المعيبة في العينة أكثر من أو يساوي 4 و في نفس الوقت اقل من ٦ وحدات

(40

افترض أن المتغير العشوائي X يتبع التوزيع الطبيعي بتوقـع 8 و انحـراف معيـاري 4، أوجـد احتمـال أن تكـون قيمـة المتغير العشوائي بين 8 و 4

(41

إذا كان احتمال أن يكون المنتج في مصنع ما معيب يساوي 0.02 و أخـذت عينـة مكونـة مـن 50 وحـدة منتجـة فأوجـد احتمال أن يكون هناك علي الأقل وحدتين معيبتين بافتراض أن المتغير العشوائي (عدد الوحدات المعيبة) يتبع:

- توزيع ذي الحدين
- توزيع بواسون

جدول التوزيع الطبيعي المعياري Z :

(المساحات تحت المنحنى على يسار القيمة المعيارية)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	1.0000
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000