



وزارة التعليم العالي والبحث العلمي
جامعة الشهيد حمه لخضر بالوادي
كلية العلوم الاقتصادية والتجارية وعلوم التسيير

مطبوعة موجهة لطلبة السنة الاولى جذع مشترك

احصاء - 1 -

من إعداد:
د. علي العبسي

الموسم الجامعي: 2019-2020

04.....مقدمة.

الفصل الأول : المفاهيم الإحصائية

05.....1-الإحصاء الوصفي والإحصاء الاستدلالي.

06.....1-1- الإحصاء الوصفي.

07.....1-2- الإحصاء الاستدلالي.

08.....2- البيانات.

10.....3- المتغيرات.

11.....3-1- المتغيرات الكيفية.

3-2- المتغيرات الكمية.

الفصل الثاني: توزيع وعرض البيانات

13.....1- تنظيم البيانات وعرضها جدولياً.

15.....2- بناء التوزيع التكراري.

19.....3- التوزيع التكراري النسبي.

20.....4- التوزيع التكراري المتجمع.

الفصل الثالث: مقاييس النزعة المركزية

25.....1- الوسط الحسابي.

27.....2- الوسط الهندسي.

28.....3- الوسط التوافقي.

29.....4- الوسط التربيعي.

- 5- العلاقة بين المتوسطات الأربعة (الحسابي، الهندسي، التوافقي، التريبيعي).....30
- 6- الوسيط.....31
- 7- المنوال.....35
- 8- العلاقة بين مقاييس النزعة المركزية.....36
- 9- التريبيعات والعشيرات والمئينات38

الفصل الرابع: مقاييس التشتت والشكل

- 1- المدى العام.....41
- 2- الانحراف المتوسط.....42
- 3- التباين والانحراف المعياري.....43
- 4- معامل الاختلاف.....47
- 5- العزوم.....48

الفصل الخامس: الانحدار والارتباط الخطي البسيط

- 1- الارتباط الخطي البسيط.....54
- 2- الغرض من تحليل الارتباط الخطي البسيط.....55
- 3- معامل الارتباط الخطي البسيط لبيرسون.....56
- 4- معامل الارتباط للرتب.....56
- 5- نموذج الانحدار الخطي البسيط.....58

الفصل السادس: الأرقام القياسية

- 1- مفهوم الأرقام القياسية.....54
- 2- الرقم القياسي البسيط.....56
- 3- خصائص الأرقام القياسية.....57
- 4- متوسط الأرقام القياسية.....58
- 5- الأرقام القياسية الترجيحية.....59

الفصل السابع: تحليل السلاسل الزمنية

- 1- مفهوم السلسلة الزمنية.....63
- 2- مكونات أو مؤشرات السلسلة الزمنية.....63
- 3- كشف وتحديد المركبة الفصلية أو الموسمية.....64
- 4- كشف وتحديد مركبة الإتجاه العام.....64

الفصل الثامن: توزيع بواسون

- مقدمة.....76
- 1- تعاريف.....76
- 2- نظرية التكافؤ.....77
- 3- قانون بواسون.....79
- 4- الأزمنة بين الوصول.....88
- 5- خاصية فدان الذاكرة.....90
- 6- الأزمنة الارتدادية.....92
- 7- الوضع العلوي لعماليات بواسون.....93

95.....	8- تفكيك عملية بواسون.....
97.....	9- عمليات بواسون المركبة.....
101.....	10- عمليات بواسون المستقرة
107.....	11- مجموعة تمارين.....

مقدمة:

علم الإحصاء هو فرع من فروع الرياضيات يشمل النظريات والطرق الموجهة نحو جمع البيانات ووصف البيانات و الاستقراء و صنع القرارات .

وعندما نتكلم عن علم الإحصاء لا نعنى بذلك البيانات الإحصائية وإنما نقصد حينئذ الطريقة الإحصائية. وهى الطريقة التي تمكننا من جميع الحقائق عن الظواهر المختلفة فى صورة قياسية رقمية وعرضها بيانيا ووضعها في جداول تلخيصيه بطريقة تسهل تحليلها بهدف معرفة اتجاهات هذه الظواهر وعلاقات بعضها ببعض.

ولقد كان الهدف الرئيسي من علم الإحصاء قديما هو عد أو حصر الأشياء المراد توفير بيانات إحصائية عنها ، وكانت الجهة التي تقوم بإعداد الإحصاءات على مستوى الدولة تعرف بمصلحة التعداد ولذلك كان التعريف القديم لعلم الإحصاء أنه علم العد، أي العلم الذي يشتمل على أساليب جمع البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة .

ولكن مع تطور المجتمعات وتشابه جوانب الحياة الاقتصادية والاجتماعية الحديثة بها ، لم يعد مجرد توفير البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة يفي بحاجات متخذي القرارات وصانعي السياسة العامة إلى تكوين صورة متكاملة الجوانب عن مجتمعهم والمجتمعات المحيطة به. فقام العلماء بتحديث نظريات علم الإحصاء وأساليبه وأدواته لكى يعين الباحثين وغيرهم على استخلاص استنتاجات معينة من البيانات الكمية التي أمكن لهم جمعها عن طريق العد.

من ذلك على سبيل المثال ، أن نظرية العينات ساعدت الباحثين على استخلاص استنتاجات عديدة من دراسة عدد صغير من الأفراد أو الأشياء - العينة - وتعميم تلك الاستنتاجات على المجتمع الذى سحبت منه العينة بأسره ولذلك يعرف علم الإحصاء حديثاً بأنه : (علم متكامل يتضمن الأسلوب العلمىالضروري لتقصى حقائق الظواهر واستخلاص النتائج عنها ، كما يتضمن أيضاً أيضاً النظرية اللازمة للقياس واتخاذ القرار فى كافة الميادين الاقتصادية والاجتماعية والسياسية).

الفصل الأول: المفاهيم الإحصائية.

يزخر كل علم من العلوم بالعديد من المصطلحات والمفردات اللغوية الخاصة به والتي يعد الإلمام بها خطوة هامة على طريق الدراسة والفهم المتعمق لموضوعات ذلك العلم وعلم الإحصاء لا يختلف في هذا الشأن عن غيره من العلوم فهو يتضمن عدد قليل من المصطلحات الأساسية التي نرى أن على الدارس أن يلم بتعريفاتها لكي يعي المقصود منها ويتسنى له معرفة كيفية التعامل معها عندما تعرض له في دراساته وبحوثه ومن ثم يتفادى الخطأ بين المصطلحات المختلفة عندما يحاول إختيار الأداة الإحصائية المناسبة لمعالجة البيانات التي قام بجمعها وتختلف الأساليب الإحصائية فيما بينها من حيث الهدف والتدرج من البساطة إلى التعقيد واختيار الأسلوب الملائم يتحدد وفقا لأهداف الباحث ونوعية البيانات المتاحة.

1- الإحصاء الوصفي والإحصاء الاستدلالي: قبل التعرض لمفهوم أنواع الإحصاء سنتطرق لتعريفات ومفاهيم إحصائية:

أ- **المجتمع الإحصائي:** هو مجموعة الوحدات الإحصائية التي تدور حولها الدراسة، ويجب أن يكون معرّفا تعريفًا دقيقًا في المكان و الزمان.

ب- **العينة الإحصائية:** هي جزء من المجتمع الإحصائي تختار من المجتمع بحيث تشمل جميع صفات المجتمع.

ج - **الوحدة الإحصائية:** هي موضوع البحث والدراسة، فد تكون شيء مادي أو شيء معنوي.

أما مصادر جمع البيانات الإحصائية فهناك:

(* **المصادر التاريخية:** وهي البيانات التي سبق تجميعها ونشرها في الكتب أو الجرائد أو المجالات أو المواقع الإلكترونية.

محاضرات في الإحصاء -1-

(المصادر الميدانية:** وهي البيانات التي يتم جمعها من المجتمع كله أو جزء منه إما عن طريق

المقابلة الشخصية أو عم طريق استخدام الاستبيانات الإحصائية لجمع المعلومات فهي بيانات مأخوذة

مباشرة من الميدان

1-1- الإحصاء الوصفي: يهدف إلى إدماج وتلخيص البيانات الرقمية بغية تحويلها من مجرد كم من

الأرقام إلى شكل أو صورة أخرى يمكن فهمها واستيعابها بمجرد النظر ومن أغلب الأساليب المستخدمة

مقاييس النزعة المركزية، مقاييس التشتت ومقاييس الارتباط والانحدار ويتوقف استخدام كل منهما على

نوعيه البيانات ومستوى القياس سواء أكان إسمياً أو وصفياً، أو ترتيبياً، أو فئوياً، أو نسبة.

يعتقد بعض الدارسين أن وظيفة الإحصاء تقتصر على معالجة مجموعة البيانات الوفيرة التي جمعها

الباحث بقصد استخلاص عدد من الجداول الإحصائية وعرضها في عدد من الأشكال والرسوم البيانية

وذلك على نحو ما نشاهده في إحصاءات السكان والاستهلاك والإنتاج وغيرها وقد يحسب المرء أن

العمليات الإحصائية تدور في جملتها حول إيجاد المتوسطات ودرجات التشتت في البيانات التي يجمعها

الباحثون ولكن في الحقيقة أن ما ذكرناه لا يمثل سوى جانب واحد من جوانب الإحصاء وهو الجانب

الوصفي ولهذا يطلق على العمليات الإحصائية التي تقوم بهذه الوظيفة مصطلح الإحصاء الوصفي وعلى

هذا يستخدم الإحصاء الوصفي تنظيم وتلخيص ووصف معلومات خاصة بعينة من العينات فمن عينة

محددة من العمال يمكن حساب متوسط الإنتاج الذي ينتجونه وحساب نسبة العمل بين أولئك العمال

ومعدل الزيادة في أجورهم وهذه المقاييس كلها وصفية بحتة لا تفيد في حد ذاتها، في الاستنتاج أو التنبؤ

وانما تصف الكيفية التي تتوزع بها البيانات التي تم الحصول عليها من العمال موضوع البحث.

تعتبر وظيفة الوصف من الوظائف الأولية لعلم الإحصاء التي تستخدم في تلمس حقائق الظواهر

المختلفة، وبإستخدام أسلوب التحليل الإحصائي للبيانات أصبح من السهولة إمكان تحديد خصائص

محاضرات في الإحصاء -1-

الظاهرة المدروسة حتى عن طريق الأشكال البيانية التي تمثل بيانات الظاهرة عملية تسهل وتبسط تحديد خصائص الظاهرة واتجاهاتها العامة .

إلى جانب ذلك يعتمد الوصف في الإحصاء على استخدام المقاييس والمؤشرات الإحصائية في تقصى الحقائق وتحديد الخصائص العامة لتوزيع بيانات الظاهرة دون الوصول إلى نتائج أو استدلاله خاصة بالمجموعات الأساسية التي تنتمي إليها الظاهرة.

إن عملية جمع البيانات ليست عملية منفصلة عن وظائف الإحصاء الأخرى فهناك صلة وثيقة - فالهدف واحد وهو الحصول على معلومات أو نتائج - وذلك يكون باستخدام مقاييس وأساليب وصف البيانات - وذلك بعد جمعها - وإذا كانت هذه البيانات خاصة بعينة أي جزء من المجتمع فإن وصف المجتمع يتطلب استخدام أساليب الاستقراء. وهذه المقاييس والأساليب لها شروط ومتطلبات يجب مراعاتها وتوفيرها عند جمع البيانات وذلك باستخدام التصميم التجريبي المناسب أو تصميم استمارة استبيان مناسبة واختيار طريقة المعاينة المناسبة وحجم العينة المناسب ومراعاة توفير مستوى القياس المناسب للمتغيرات، كما أن البيانات التي يتم جمعها يجب أن تكون محل ثقة حتى تكون النتائج المستخلصة منها محل ثقة، أي يجب أن يتوافر فيها الصدق والثبات أن تحديد ذلك واختياره يكون غالباً باستخدام الأساليب الإحصائية.

1-2- الإحصاء الاستدلالي: يستند هذا القسم من الأساليب الإحصائية إلى مجموعة من النظريات الإحصائية لعل أهمها نظرية الاحتمالات ونظرية العينات اللتان تمثلان حلقة الوصل بين الإحصاء الوصفي والاستدلالي.

ويسعى هذا النوع من الأساليب الإحصائية إلى الوصول إلى تقديرات لمعالم وخصائص مجتمعات الدراسة من خلال ما هو متوفر من معلومات عن العينات المختارة من تلك المجتمعات، فضلاً عن إختبار الفروض الإحصائية عن مجتمع البحث على أساس البيانات المتاحة عن عينات الدراسة.

محاضرات في الإحصاء -1-

ويطلق على هذا النوع من الأساليب أكثر من تسمية تؤدي جميعها إلى نفس المعنى فأحيانا يسمى بالإحصاء الاستدلالي، أو التعميمي، حيث يهدف إلى الوصول إلى تعميمات عن مجمع الدراسة من خلال العينة المسحوبة من هذا المجتمع. ويشمل هذا النوع من الأساليب الإحصائية، الاحتمالات، العينات، اختبار الفروض، الاستدلال من خلال عينة واحدة أو أكثر وما يتضمنه ذلك من اختبارات مختلفة.

يقصد بوظيفة الاستدلال اشتقاق النتائج من دراسة وفحص المقدمات والبيانات المتوفرة عن ظاهرة معينة. ولهذا يطلق على عملية الإحصائية التي تستخدم والاستدلالي على أساس المنطق الاستدلالي المبني على نظرية الاحتمالات الرياضية فمن عينة محددة من أعمال أحد المصانع وباستخدام الأسلوب الإحصاء الاستدلالي يكون من الممكن التنبؤ بمعدلات الزيادة في الإنتاج ومقدار التغير في نسبة الغياب وفي هذه الحالة نجد أن الدقة في التنبؤ تعتمد على عوامل كثيرة من أهمها ملائمة الأدوات الإحصائية المستخدمة وحجم العينة محل الدراسة والإجراءات الإحصائية اتخذت عند إختيارها.

تعتبر وظيفة الاستدلال أو الاستقراء من الأهمية بمكان في البحث العلمي فمثلاً:

إذا كانت الظاهرة موضوع الدراسة والتحليل ممثلة للمجتمع الذي تنتمي إليه فإنه يمكن الحصول على نتائج معنوية عن المجتمع بتحليل بيانات هذه الظاهرة وهو ما يعرف بالاستدلال ويعتمد هذا الأسلوب في البحث على الشروط التي يجب توافرها حتى يكون هذا الاستدلال سليماً - وبما أن الاستدلال لا يمكن أن يكون مؤكداً فإن لغة الاحتمال تستخدم عند عرض النتائج.

2- البيانات: من الشائع في مجال البحوث الاقتصادية توافر مجموعة من البيانات الإحصائية التي يحصل عليها الباحث باستخدام أدوات جمع بيانات مناسبة وعادة تتمثل تلك البيانات في شكل أرقام تعتبر

محاضرات في الإحصاء -1-

قياسا للمتغيرات تحت الدراسة ولما كانت تلك الأرقام تفنقر إلى الترتيب والتصنيف يطلق عليها البيانات الأولية أو البيانات الخام.

تعرف البيانات الإحصائية أنها كمية من المعلومات على هيئة أرقام وان تلك الأرقام إما أن تكون صحيحة مثل 10 ، 15 ، 30 وهكذا أو تكون أرقاما عشرية أو حقيقية مثل 8.5 ، 10.25 ، 1505 وهكذا:

ويتوقف حجم البيانات الخام على حجم المجتمع الأصلي فكلما إزداد حجم هذا المجتمع يتوقع مزيدا من الأرقام غير المرئية والتي يصعب مع كثرتها وعدم تصنيفها تفهم أو قياس متغير أو أكثر تحت الدراسة ومن ثم كان من الضروري أن يقوم الباحث بتصنيف وتبويب تلك البيانات بالشكل أو بالأسلوب الذي يخدم جيدا هدف الباحث من دراسة المتغيرات أو استنباط نوعية العلاقات أو المعلومات الهامة التي تتعلق بتلك المتغيرات .

يقصد بتعبير البيانات " أى كمية من المعلومات في صورة رقمية والصورة الرقمية للبيانات تبدو إما على شكل أرقام صحيحة مثل 10 ، 112 ، 464. أو على شكل أرقام حقيقية مثل 20.4 ، 61.8 ، 182.1 أي أنها الأرقام التي تحتوى على صفة عشرية. وتعتبر المعلومات الرقمية (البيانات) المادة الخام لأسلوب العمل الإحصائي كما أنها تلعب دورا كبيرا في تطبيق الأساليب الإحصائية .

وتسمى البيانات المتاحة – المنشورة أو التي تم جمعها – تسمى بيانات خام أو أولية – ذلك أنها تكون غير مجهزة فهي لاتفصح إلا عن القليل من المعلومات. كما أنه يستحيل استخلاص المعلومات منها. وفى سبيل ذلك نستعين بأساليب ومقاييس وصف البيانات، وهذه الأساليب كثيرة ومتنوعة فهي تختلف حسب عوامل أهمها عدد المتغيرات ومستوى قياسها.

محاضرات في الإحصاء -1-

ولعل أبسط الطرق الإحصائية لتنظيم وتلخيص البيانات طريقة التوزيع التكراري، أو بمعنى ضمني من

التوزيع التكراري يمكن استخدام وسيلة أو أكثر من الوسائل الثلاث التالية :

(أ) استخدام الجداول الإحصائية في عملية تصنيف وتبويب البيانات الخام.

(ب) استخدام التمثيل البياني والخرائط في عرض البيانات الإحصائية (تحويل التوزيع التكراري إلى

منحنى تكراري).

(ج) استخدام مقياس أو أكثر من المقاييس الإحصائية مثل المتوسط الخام الانحراف المعياري ومعامل

الارتباط في تلخيص البيانات الإحصائية في صورة رقم أو نسبة مئوية ونرى أهمية الوقوف على

نوعية البيانات الإحصائية من منظور مستويات القياس الإحصائي نظراً لأهمية تلك البيانات

الإحصائية وفقاً لمستويات القياس الإحصائي يرجع إلى أن المتغيرات التي تقاس كمياً تنقسم من

قيمتها العددية إلى المتغير المتصل والمتغير المنقطع .

3- المتغيرات: تشير كلمة المتغيرات إلى الخصائص التي تشترك فيها أفراد المجتمع الإحصائي ولكنها

تختلف من فرد إلى فرد آخر فالعمر ، درجة الذكاء، وطول القامة، واللياقة البدنية والقدرة على القراءة،

والدخول التي يحصل عليها الأفراد أمثلة للمتغيرات وتتميز هذه المتغيرات بأنها قابلة للقياس الكمي

وبإمكانية تحديد قيمة معينة لها.

والمتغيرات عبارة عن ظاهرات أو صفات تختلف قيمها باختلاف الحالات. ومن أمثلتها : درجة الحرارة في

مناطق مختلفة أو في فترات مختلفة لمكان واحد ، كميات الإنتاج الزراعي أو الصناعي.

ويمكن القول بأن المتغير هو أى ظاهرة أو حدث أو خاصية تأخذ فيها قيمة تتغير من ظرف لآخر.

والمتغير هو الوحدة الأساسية للتحليل الإحصائي ويمكن تعريفه بأنه مجموعة من العناصر أو التقسيمات

غير المتداخلة، وهذه المجموعة من التقسيمات تكون مقياس.

يمكن تصنيف المتغيرات إلى متغيرات كمية والمتغيرات كيفية:

3-1- المتغيرات الكيفية: وهي التي تصف الأشياء بصفات مثل متغير النوع الذي ينقسم إلى قسمين ذكور وإناث، والحالة العملية للفرد حيث تكون إما مزارع أو عامل غير ماهر، أو عامل ماهر أو موظف أو تاجر وما إلى ذلك من صفات، وهذه المتغيرات الكيفية تواجه مشاكل في معالجتها إحصائياً ما لم يميزها عن بعضها البعض، ويمكن أن نستخدم الأرقام فنرمز لمتغير الإناث برقم 1 و لمتغير الذكور برقم 2 أو العكس حسب الهدف المسطر، والرقم في هذه الحالة لا يعنى أكثر من أنه أداة للتمييز بين المتغيرات الكيفية لتسهيل تفريغ البيانات التي جمعت عنها من ميدان الدراسة تمهيداً لمعالجتها إحصائياً ولا تكون لها قيمة عددية في حد ذاته.

3-2- المتغيرات الكمية: وهي التي يمكن أن نصفها عددياً بأنها أكبر من أو أقل من قيمة معينة ويعتبر العمر وعدد سنوات التعليم أمثلة لهذه المتغيرات، وتنقسم المتغيرات الكمية إلى مستمرة وغير مستمرة (متقطعة).

3-2-1- المتغير المتصل: إن التعريف العام للمتغير هو ظاهرة أو صفات تختلف قيمها باختلاف الحالات فإن المتغير يكون متصلاً عندما يأخذ أي قيمة متدرجة على المقياس المستخدم. مثال ذلك قياس درجات الحرارة باستخدام الترمومتر فالمتغير يأخذ أي قيمة بين رقمين صحيحين، بمعنى أن المتغير يمكن أن يأخذ أي قيمة بين 36 درجة و 37 درجة مثل: 36.2.

3-2-2- المتغير المتقطع: عندما يأخذ المتغير قيماً محددة يطلق عليه متغيراً متقطعاً أو بمعنى آخر، المتغير المتقطع هو الذي يحتوى على عدد محدود من القيم أو يحتوى عدد لانهائي من القيم ولكن لكل منها قيمة محددة يمكن عدها أو ترتيبها في نهاية الأمر تعدد الأولاد أو الأفراد في الأسرة لابد أن يكون أعداداً

محاضرات في الإحصاء -1-

صحيحة غير حقيقة مثل 1، 2، 3، 4 وهكذا ومن أمثال المتغيرات المتقطعة، النوع، الحالة الزوجية، عدد أيام الإنتاج فيأحد المصانع، عدد حوادث السيارات.

3-3- المتغيرات التابعة والمستقلة والوسيطية: ويمكن تصنيف المتغيرات تصنيفاً آخر بحسب دورها في حدوث الظاهرة محل الدراسة وذلك إلى:

3-3-1- متغيرات تابعة: وهي تلك المتغيرات التي نحاول تفسيرها ومعرفة أسباب حدوثها وتحديد مدى إمكان التنبؤ بها.

3-3-2- متغيرات مستقلة: وهي التي لعبت دوراً مباشراً في حدوث المتغيرات التابعة ونستخدمها في تأييد تفسيرنا وفهمنا لما طرأ على هذه المتغيرات من تغيير، وفي التنبؤ بالحالة التي ستؤول إليها بعد ذلك.

3-3-3- متغيرات وسيطة: وهي تلك المتغيرات التي يمر من خلالها تأثير المتغيرات المستقلة إلى المتغيرات التابعة والمتغيرات الوسيطة بالغة الأهمية في تفسير حدوث الظواهر الاجتماعية إذ قد يغفل عنها الباحثون أو قد ينظرون إليها على أنها متغيرات مستقلة لإرتباطها المباشر بالمتغيرات التابعة، فإذا نظرنا إلى تفسير ظاهرة الفقر، على سبيل المثال سنجد أن بعض الأفراد ينظرون إلى حالة فقدان المعايير التي تؤدي إلى الفقر على أنها المتغير المستقل والفقر هو المتغير التابع ولكن فريقاً آخر من الباحثين الذين ينظرون إلى الظاهرة بطريقة أكثر تفصيلاً، ويرون أن المجتمع يمر بتغيرات اقتصادية واجتماعية قوية التفسير.

الفصل الثاني: عرض وتوزيع البيانات.

إن الهدف من عرض توزيع البيانات -البيانات الخام- في جداول مناسبة وذلك حتى يمكن تلخيصها وفهمها واستيعابها واستنتاج النتائج منها ومقارنتها بغيرها من البيانات، كما يسهل الرجوع إليها في صورة جداول دون الاطلاع على الاستمارات الأصلية التي قد تحمل معلومات مهمة لشركة مثلاً، مما يخل بمبدأ سرية البيانات الإحصائية .

كما يعتبر عرض وتبويب البيانات الإحصائية الخطوة الثانية (بعد تجميع هذه البيانات الخام) في مفهوم التحليل الإحصائي، ويلجأ الباحث إلى حصرو وتصنيف هذه البيانات وعرضها بطريقة مختصرة تساعد على فهمها وتحليلها إحصائياً للتعرف عليها ووصفها ومقارنتها بغيرها من الظواهر، والخروج ببعض المدلولات الإحصائية عن مجتمع الدراسة.

1- تنظيم البيانات وعرضها جدولياً: في حالة المتغير الكيفي، في الجدول نرتب كصفات الصفة حسب التكرارات المطلقة n_i .

في حالة المتغير الكمي، في الجدول نرتب الصفة حسب قيم المتغير x_i .

التكرار النسبي يساوي: $f_i = \frac{n_i}{\sum n_i}$ بحيث مجموع التكرارات النسبية يساوي الواحد 1 $\sum f_i = 1$

1-1- التكرار المتجمع: في حساب التكرار المتجمع الصاعد نبدأ من أعلى الجدول إلى أسفله ونقوم

بجمع التكرارات سواء المطلقة أو النسبية، أما في حساب التكرار المتجمع النازل نبدأ من أسفل

الجدول إلى أعلاه ونقوم بطرح التكرارات سواء المطلقة أو النسبية، ونرمز لتكرار المتجمع

الصاعد بـ $F_i(\uparrow)$ ونرمز لتكرار المتجمع النازل بـ $F_i(\downarrow)$

في حالة المتغير الكمي المتصل الجدول يعبر عنه بالفئات، وطول الفئة يساوي: $L = \frac{x_{\max} - x_{\min}}{k}$

محاضرات في الإحصاء -1-

بحيث: L : يمثل طول الفئة، و k : عدد الفئات، ويمكن حساب عدد الفئات باستخدام أحد الطرق الآتية:

❖ الطريقة الأولى: طريقة ستورج (Sturge): $k = 1 + 3.32 \log n$ أو $k = 1 + 1.32 \ln n$

❖ الطريقة الثانية: معادلة يول Yule التي تنص على أن: $k = 2.5 \times \sqrt[4]{n}$

❖ الطريقة الثالثة: عدد الفئات يساوي القسم الطبيعي للجذر التربيعي لعدد قيم n . $k = \sqrt{n}$

2-1- العرض البياني.

2-1-1- العرض البياني في حالة متغير كمي متقطع.

أ- الأعمدة البسيطة: هي عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة العينة للمتغير المدروس.

ب- العرض البياني للتكرار المتجمع الصاعد: هو عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات المتجمعة الصاعدة المقابلة لكل قيمة من قيم المتغير الإحصائي المدروس.

ج- العرض البياني للتكرار المتجمع النازل: وهو عبارة عن قطع مستقيمة متنازلة حسب تنازل التكرارات المتجمعة النازلة، حيث القطعة المستقيمة الأولى تقابل مجموع التكرارات ناقص التكرار الأول مع القيمة الثانية للمتغير وهكذا.

2-1-2- العرض البياني في حالة متغير كمي متصل (مستمر): إن العروض البيانية للمتغير الإحصائي

الكمي المستمر هي أكثر العروض البيانية استعمالاً ومن أهمها:

أ- المدرج التكراري: وهو عبارة عن مستطيلات متلاصقة طول كل منها يتناسب مع التكرار المقابل، وقاعدة كل منها تساوي طول الفئة المقابلة.

محاضرات في الإحصاء -1-

ب- **المضلع التكراري**: هو مجموع من قطع مستقيمة متصلة ومنكسرة تتحدد بنقاط أحداثياتها مركز الفئة والتكرارات المقابلة.

ج- **منحنى التكرارات المتجمعة الصاعدة والنازلة**: يرسم منحنى التكرار المتجمع الصاعد عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرار المتجمع الصاعد المقابل لها، ويرسم منحنى التكرار المتجمع النازل بإيصال مجموعة النقاط التي إحداثياتها: الحدود الدنيا للفئات والتكرار المتجمع النازل مقابل لها.

يبين كل من منحنى التكرار المتجمع الصاعد ومنحنى التكرار المتجمع النازل شدة أو ضعف تطور الظاهرة المدروسة عن مستوى معين من مجال الدراسة. إن فاصلة نقطة تقاطع المنحنيين تسمى بالوسيط. 1-2-3- العرض البياني في حالة متغير نوعي.

أ- **العرض الدائري**: ويتمثل في دائرة مقسمة إلى عدة أجزاء كل جزء يقابل زاوية مركزية تتناسب مع التكرارات المقابلة لكل خاصية من الخصائص المدروسة، ولتحقيق ذلك نضيف عموداً إلى جدول المعطيات يحتوي على الزوايا المركزية المقابلة لكل تكرار.

ب- **العمود المجزأ**: وهو عبارة عن مستطيل مقسم إلى عدة أجزاء كل جزء يقابل تكرار معين للخاصية المدروسة.

ج- **الأعمدة المستطيلة**: وهو عبارة عن مجموعة من الأعمدة المتجاوزة ذات القواعد المتساوية إلا أن ارتفاعها تتناسب مع تكرار كل خاصية، كما أن هذه الأعمدة تكون متباعدة بمسافات متساوية.

2- بناء التوزيع التكراري:

إذا كان مدى البيانات صغيراً ممكن أن يكون بناء التوزيع التكراري مباشرة ، أما إذا كان المدى كبيراً أو كان عدد البيانات كبيراً فإنه يجدر في هذه الحالة أن تقسم قيم البيانات إلى فئات يتراوح عددها ما بين 5،

محاضرات في الإحصاء -1-

15 فئة حسب كون عدد البيانات صغيراً أو كبيراً. وبعد تقسيم قيم البيانات إلى فئات تفرغ البيانات على الفئات وتجمع التكرارات المقابلة لكل فئة. نشرح الخطوات المتبعة لبناء التوزيع التكراري بدراسة المثال.

مثال(1): تبين العلامات التالية علامات 25 طالباً في امتحان الأعمال الموجهة لمقياس الإحصاء.

5	7	8	5	6
8	6	8	9	8
7	9	7	10	5
10	8	9	3	7
6	5	7	7	3

حسب هذا المثال يمكن بناء توزيع تكراري غير مبوب – مباشر -.

مثال (2): تمثل البيانات التالية علامات 60 طالباً في الامتحان النهائي لأحد المقاييس.

35	48	32	27	48	44
23	46	33	29	42	24
27	49	23	23	46	25
28	36	41	48	37	25
31	41	43	47	39	36
35	43	48	33	41	43
28	38	46	36	36	48
27	32	34	22	28	47
41	24	24	39	33	38
44	45	46	44	23	43

ولعرض هذه البيانات في توزيع تكراري ذي فئات متساوية نتبع الخطوات التالية:

1- نأخذ فئات متساوية ونفرض أن عددها 6.

2- المدى للبيانات = أكبر قيمة – أصغر قيمة.

من البيانات التي لدينا نجد:

$$\text{المدى} = 27 - 49 = 22$$

$$\text{ففي مثالنا يكون } \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{27}{6} = 4.5$$

وبما أن البيانات معطاة لأقرب عدد صحيح (لا يوجد كسور عشرية) نقرب العدد 4.5 إلى أعلى

فنأخذ طول الفئة $C = 5$.

محاضرات في الإحصاء -1-

3- يقصد بدرجة الدقة التي أعطيت بها البيانات، أنه إذا كانت البيانات أعداداً صحيحة فإن ذلك يعني أن طول الفئة يجب أن يكون عدداً صحيحاً، وإذا احتوت البيانات على خانة عشرية واحدة فيجب أن يكون طول الفئة عدداً صحيحاً أو ذا خانة عشرية واحدة، أي أن طول الفئة لا تزيد دقته عن الدقة المستعملة في البيانات.

4- نعين الحد الأدنى لأول فئة ويجب أن يكون هذا الحد مساوياً لأصغر قيمة في البيانات أو أقل منها بقليل، ويجب أن تكون درجة دقته نفس درجة دقة البيانات المستعملة، وإذن فالحد الأدنى للفئة الأولى هو 22، وبعد تعيين الحد الأدنى للفئة الأولى نعين الحد الأدنى الفعلي لتلك الفئة وهو عبارة عن الحد الأدنى ناقصاً نصف وحدة دقة أي نصف وحدة من الوحدات التي قربت إليها الأعداد في البيانات. وإذن فالحد الأدنى الفعلي للفئة الأولى هو: $22 - 0.5 = 21.5$ لأن التقريب في البيانات المستعملة كان لأقرب ساعة (لأقرب عدد صحيح). نلاحظ أنه لو كانت البيانات معطاة لأقرب عشر (خانة عشرية واحدة) لكان نصف وحدة الدقة $= 0.05$ أما لو احتوت البيانات على خانتين عشريتين فإن وحدة الدقة تكون 0.01 ونصف وحدة الدقة تكون 0.005 وهكذا.

5- نعين الحد الأعلى الفعلي للفئة الأولى وذلك بإضافة طول الفئة إلى الحد الأدنى الفعلي لتلك الفئة.

إذن الحد الأعلى الفعلي للفئة الأولى هو $21.5 + 5 = 26.5$

ثم نعين الحد الأعلى للفئة الأولى وهو يساوي الحد الأعلى الفعلي ناقصاً نصف وحدة الدقة، ولذا فالحد الأعلى للفئة الأولى هو $26 - 0.5 = 25.5$ ، وبهذا نكون قد حصلنا على الفئة الأولى وهي الفئة التي حدودها 22، 26 وحدودها الفعلية 21.5، 26.5.

6- نعين الحدود الدنيا والعليا لجميع الفئات الباقية وذلك بإضافة طول الفئة لكل حد. ثم نعين الحدود الدنيا الفعلية والعليا الفعلية وذلك بإضافة طول الفئة لكل حد فعلي.

7- نعين مراكز الفئات X_i ونجد مركز أي فئة بقسمة مجموع حديها على 2 ولذلك فمركز الفئة الأولى هو $24 = \frac{21.5 + 26.5}{2}$ ويمكن إيجاد مركز أي من الفئات الأخرى بإضافة طول الفئة إلى مركز الفئة التي قبلها فيكون مركز الفئة الثانية $29 = 24 + 5$ ومركز الفئة الثالثة $34 = 29 + 5$ وهكذا.

محاضرات في الإحصاء -1-

8- نفرغ البيانات المعطاة لدينا على الفئات التي أنشأناها وذلك باستعمال خط عمودي لكل قراءة وخط مائل للقراءة الخامسة في كل فئة وذلك لتسهيل جمع التكرارات.

9- نجمع التكرارات المقابلة لكل فئة ونسجله في عمود التكرارات f_i ثم نجمع التكرارات لجميع الفئات ونقارنه بعدد البيانات فإذا كان عدد البيانات n فإنه يجب أن يكون: حيث h عدد الفئات.

يظهر توضيح هذه الخطوات في الجدول التالي:

الجدول رقم (01-01) التوزيع التكراري لعلامات 60 طالباً.

حدود الفئة	الحدود الفعلية للفئة	مركز الفئة X_i	التكرار f_i
22 - 26	21.5 - 26.5	24	10
27 - 31	26.5 - 31.5	29	8
32 - 36	31.5 - 36.5	34	12
37 - 41	36.5 - 41.5	39	9
42 - 46	41.5 - 46.5	44	13
47 - 51	46.5 - 51.5	49	8
المجموع			60

وعند عرض التوزيع التكراري لا نكتب عمود إفراغ البيانات وفي بعض الأحيان لا نكتب عمود الحدود

الفعلية بل نكتفي بعرض العمودين (1) و (4) أو العمودين (2) و (4) أو (3) و (4).

ويمكننا تلخيص خطوات بناء التوزيع التكراري فيما يأتي:

- نعين عدد الفئات المتساوية.
- نعين المدى.
- نعين طول الفئة وذلك بقسمة المدى على عدد الفئات ثم التقريب إلى الأعلى.
- نعين الحد الأدنى للفئة الأولى ثم نطرح منه نصف وحدة دقة لنعين الحد الأدنى الفعلي للفئة الأولى.
- نعين الحد الأعلى الفعلي للفئة الأولى وذلك بإضافة طول الفئة إلى الحد الأدنى الفعلي لتلك الفئة، ثم نطرح منه نصف وحدة دقة لنعين الحد الأعلى للفئة الأولى.

محاضرات في الإحصاء -1-

- نعين الحدود الدنيا والعليا والدنيا الفعلية والعليا الفعلية للفئات الباقية وذلك بإضافة طول الفئة لكل حد على التوالي.
- نعين مراكز الفئات وذلك بتعيين مركز الفئة الأولى ثم إضافة طول الفئة له لتعيين مركز الفئة الثانية وهكذا.

$$\text{مركز الفئة الأولى} = \frac{\text{الأدنى منها 3 حدها الأعلى}}{4}$$

- نفرغ البيانات على الفئات.
- نسجل مجموع تكرارات كل فئة أمامها في عمود التكرارات.
- نجمع التكرارات لجميع الفئات ونقارنه بعدد البيانات.

3- التوزيع التكراري النسبي.

إن التكرار النسبي لكل فئة هو نسبة تكرار تلك الفئة إلى مجموع التكرارات، فإذا كان مجموع التكرارات n

وكان تكرار فئة معينة f فإن تكرارها النسبي هو $\frac{f}{n}$.

إن الجدول الذي يعطينا الفئات (أو مراكزها) مع تكراراتها النسبية يسمى التوزيع التكراري النسبي، وعليه

فإن الجدول رقم (2) يمثل التوزيع التكراري النسبي للمثال رقم (2).

محاضرات في الإحصاء -1-

الجدول رقم (02) : التوزيع التكراري النسبي لعلامات 60 طالباً.

حدود الفئة	التكرار النسبي
22 – 26	
27 – 31	
32 – 36	
37 – 41	
42 – 46	
47 – 51	

ونلاحظ أن مجموع التكرارات النسبية يجب أن يساوي 1.

$$\sum w_i = 1$$

4- التوزيع التكراري المتجمع.

نحتاج في كثير من الأحيان معرفة عدد المشاهدات التي تساوي قيمة معينة أو تكون أصغر منها، فعلى سبيل المثال، إذا حصل طالب على العلامة 80 في أحد الامتحانات فإنه يرغب في معرفة عدد الطلبة الحاصلين على العلامة 80 أو أقل في ذلك الامتحان.

وفي مثال (2)، ربما نحتاج الإجابة على أسئلة من نوع: كم طالباً حصل على العلامة 36 أو أقل؟.

بالنظر إلى التوزيع التكراري في جدول رقم (1) يكون عدد الطلاب $10 + 8 + 12 = 30$ وهذا هو التكرار المتجمع للفئة الثالثة.

محاضرات في الإحصاء -1-

وبهذا فإن مجموع تكرارات جميع القيم التي تساوي الحد الأعلى الفعلي لفئة ما أو تقل عنه هو التكرار المتجمع لتلك الفئة.

إن الجدول الذي يعطينا الحدود الفعلية للفئات مع التكرارات المتجمعة المقابلة لها يسمى التوزيع التكراري المتجمع، ونبدأ دائماً بالحد الأدنى الفعلي للفئة الأولى ونعتبر تكراره المتجمع صفراً، حيث لا يوجد بيانات تقل قيمتها عن ذلك الحد أو تساويه.

وعليه فإن الجدول رقم (03) يمثل التوزيع التكراري المتجمع للمثال رقم (2).

الجدول رقم (03): التوزيع التكراري المتجمع.

التكرار المتجمع	
0	أقل من 21.5
10	أقل من 26.5
18	أقل من 31.5
30	أقل من 36.5
39	أقل من 41.5
52	أقل من 46.5
60	أقل من 51.5

أما إذا استعملنا التكرارات النسبية بدلا من التكرارات فإننا نحصل على التوزيع التكراري المتجمع النسبي، ونحصل على هذه النتيجة نفسها لو قسمنا التكرار المتجمع لكل فئة على مجموع التكرارات.

مثال (3): بيانات التالية تمثل أوزان 60 طالبا بالكيلوغرام في أحد أقسام LMD بكلية العلوم الاقتصادية وعلوم التسيير بجامعة الوادي.

محاضرات في الإحصاء -1-

67	64	68	73	73	54	61	74	60	78
80	74	65	63	60	69	72	66	77	65
74	50	76	69	68	66	78	63	70	
67	67	64	76	61	72	72	57	65	77
59	71	79	78	58	63	74	66	73	67
61	71	69	68	73	81	64	61	84	55

المطلوب: أنشئ جدول التوزيع التكراري و أحسب التكرارات التجميعية الصاعدة والنازلة ثم مثلها بيانياً:

الحل: أول خطوة نقوم بها هي ترتيب البيانات السابقة ترتيباً تصاعدياً.

- حساب المدى: $E = \text{Max}(x_i) - \text{Min}(x_i) = 84 - 50 = 34$

- حساب عدد الفئات: $1 + 3.222 \log(n) = 6.9$

- حساب طول الفئة: $k = E / 1 + 3.22 \log(n) = 4.92$

$F_i^{\downarrow} \%$	$F_i^{\uparrow} \%$	$N_i^{\downarrow}$	$N_i^{\uparrow}$	$f_i \%$	f_i	C_i	عدد الطلبة n_i	أوزان الطلبة X_i
100	3,3	60	2	3,3	0,033	52,5	2	[55 – 50]
96,7	11,6	58	7	8,3	0,083	57,5	5	[60 – 55]
88,4	31,6	53	19	20	0,2	62,5	12	[65 – 60]
68,4	58,4	41	35	26,8	0,268	67,5	16	[70 – 65]
41,6	81,7	25	49	23,3	0,233	72,5	14	[75 – 70]
18,3	95	11	57	13,3	0,133	77,5	8	[80 – 75]
5	100	3	60	5	0,05	82,5	3	[85 – 80]
/	/	/	/	100	1	/	60	المجموع $\sum n_i$

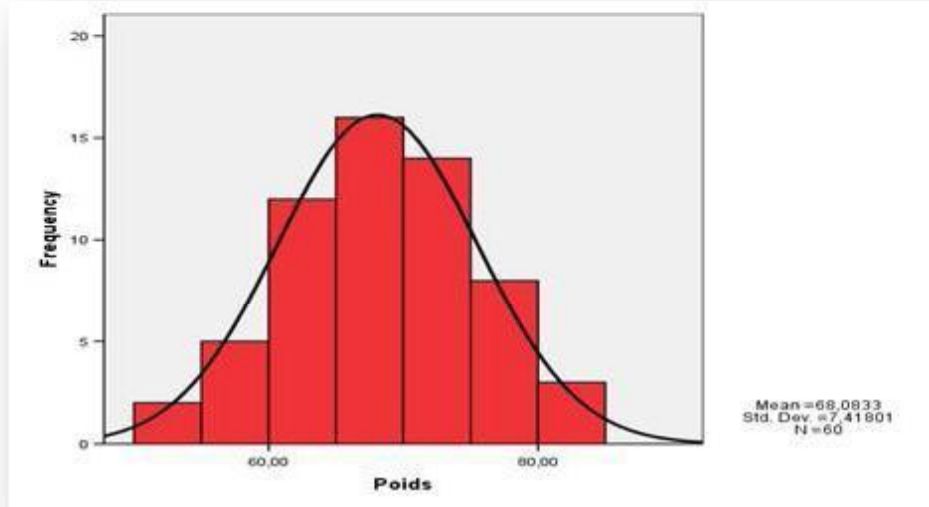
الشرح:

- $n_2 = 5$: هناك 5 طلبة من بين 60 طالبا أوزانهم تتراوح بين 55 و 60 كـلـغ.
- $N_2^+ = 7$: هناك 7 طلبة من بين 60 طالبا أوزانهم أقل تماما من 60 كـلـغ.
- $N_5^+ = 25$: هناك 25 طالبا من بين 60 طالبا أوزانهم أكبر أو يساوي 70 كـلـغ.
- $F_2^+ \% = 11,6\%$: هناك 11,6% من الطلبة أوزانهم أقل تماما من 60 كـلـغ.
- $F_5^+ \% = 41,6\%$: هناك 41,6% من الطلبة أوزانهم أكبر أو يساوي 70 كـلـغ.

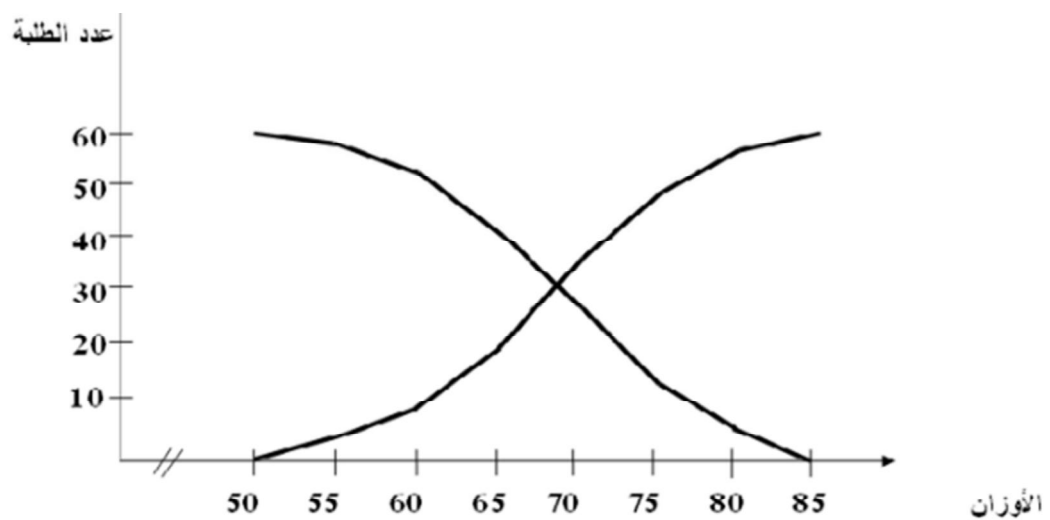
التمثيل البياني للمتغير الإحصائي المستمر:

- يمثل التكرار المطلق والنسبي للمتغير الإحصائي المستمر عن طريق المدرج التكراري حيث تتناسب مساحة المستطيل مع التكرار المطلق أو النسبي الموافق له، إذا ربطنا مراكز الفئات بواسطة خطوط مستقيمة مع بعضها البعض نحصل على المضلع التكراري، وإذا رسمنا منحنى بجوار المضلع التكراري نحصل على المنحنى التكراري .
 - يمثل التكرار المتجمع الصاعد والنازل للمتغير الإحصائي المستمر عن طريق المنحنى المتجمع الصاعد والنازل، حيث نرفق بكل قيمة للتكرار المتجمع الصاعد الحد الأعلى للفئة الموافقة لها ونرفق بكل قيمة للتكرار المتجمع النازل الحد الأدنى للفئة الموافقة لها.
- وبالرجوع إلى المثال الخاص بتوزيع الطلبة حسب الأوزان يكون التمثيل كالاتي:

الشكل (01-01): تمثيل بياني بواسطة المدرج التكراري والمنحنى التكراري لتوزيع الطلبة الجامعة حسب أوزانهم.



الشكل (01-02): تمثيل بياني للتكرار المتجمع الصاعد والمتجمع النازل.



الفصل الثالث: مقاييس النزعة المركزية.

إن الأسلوب البياني في تحليل ودراسة الظواهر لتحديد الخصائص والاتجاهات والعلاقات، يعتمد في دقته على دقة التمثيل البياني نفسه وبذلك ربما تختلف الخصائص من رسم إلى آخر لنفس الظاهرة، وعليه فإنه من الأفضل اللجوء إلى طرق القياس الكمي، حيث يستخدم الباحث الطريقة الرياضية في القياس.

فالهدف الأساسي من استخدام مقاييس النزعة المركزية ومقاييس التشتت هو تلخيص البيانات في محاولة أخرى لوصفها عن طريق التعرف على مركزها ومقدار تشتت البيانات حول هذا المركز (درجة تجانس البيانات) ومن خلال هذين المؤشرين يتمكن الباحث من فهم أبعاد الظاهرة قيد الدراسة.

ومن أهم مقاييس النزعة المركزية التي سنتعرض إليها بالدراسة الوسط الحسابي والوسيط والمنوال.

1. الوسط الحسابي: إذا كان لدينا n من الأعداد (قيم المشاهدات) ، $x_1, x_2, \dots, x_n$ ، فإن الوسط الحسابي لها هو حاصل قسمة مجموعها على عددها.
أي أن:

حيث: $\bar{x}$ يمثل الوسط الحسابي.

ويمكن استعمال الرمز Σ الذي يعني جمع الحدود التي في داخله. وبذلك يصبح تعريف الوسط الحسابي.

ويعرف الوسط الحسابي في حالة التوزيع التكراري كما يلي:

إذا كان لدينا توزيع تكراري عدد فئاته h وكانت مراكز الفئات (أوالقيم) $x_1, x_2, \dots, x_h$ والتكرارات المقابلة

$f_1, f_2, \dots, f_h$ فإن الوسط الحسابي يكون:

محاضرات في الإحصاء -1-

وباستعمال الرمز $\sum$ يكون:

مثال (1) : أحسب الوسط الحسابي للتوزيع التكراري التالي:

مركز الفئة X_i	التكرار f_i	
5	7	35
10	10	100
15	4	60
20	12	240
25	9	225
30	3	90
المجموع	45	750

1-1- الوسط الحسابي المرجح (الموزون) لأوساط حسابية.

إذا كان لدينا مجموعات ذات أعداد مختلفة من البيانات وعلم الوسط الحسابي لكل مجموعة، فكيف

نحصل على الوسط الحسابي للمجموعات كلها إذا دمجت معا؟

القاعدة التالية تعطينا الإجابة عند دمج مجموعتين معا:

محاضرات في الإحصاء -1-

قاعدة (1): إذا كان لدينا مجموعة ذات n_1 من القيم ووسطها الحسابي $\bar{x}_1$ ومجموعة ثانية ذات n_2 ووسطها الحسابي $\bar{x}_2$ ، فإن الوسط الحسابي للمجموعة ذات $n_1 + n_2$ من القيم الناتجة من دمج المجموعتين يكون:

وهذا هو الوسط الحسابي المرجح (الموزون) لأوساط حسابية. وهذه القاعدة صحيحة لو دمجنا أي عدد محدود من المجموعات مع بعضها البعض.

وكمثال على ذلك، لو دمجت مجموعتان أ، ب وكان عدد البيانات في أ يساوي 20 ووسطها الحسابي 45 وعدد البيانات في ب يساوي 30 ووسطها الحسابي 50 فإن الوسط الحسابي للمجموعة الناتجة من دمج المجموعتين يساوي:

لاحظ أن الوسط المرجح يقع بين وسطي المجموعتين.

ويمكن تعميم قاعدة (1) لإيجاد الوسط الحسابي المرجح لعدة مجموعات بعد دمجها.

2- الوسيط.

1.2 الوسيط للبيانات الأولية (الخام).

في كثير من الأحيان ينصب اهتمام الدارس على معرفة القيمة التي يكون عدد الأفراد الحاصلين على قيمة أقل منها مساويا لعدد الحاصلين على قيمة أعلى منها، فمثلا: عندما تحصل على علامة معينة في امتحان ما، هل تقع علامتك في النصف الأول من العلامات أم في النصف الثاني؟ في مثل هذه الحالة،

محاضرات في الإحصاء -1-

عليك أولاً أن تعرف تلك العلامة التي حصل نصف الطلبة على أقل منها ونصفهم الآخر على علامة أعلى منها. إذا عرفت هذه العلامة يكون بإمكانك الإجابة عن سؤالك السابق.

فمثلاً إذا كانت العلامات التي حصل عليها طلبة الصف هي:

20, 6, 14, 10, 9, 7, 20, 18, 11, 10, 19, 17, 12, 13, 8

فما هي العلامة التي حصل نصف الطلبة على أقل منها والنصف الآخر على أعلى منها؟ للإجابة عن هذا السؤال ترتب العلامات ترتيباً تصاعدياً أي من الأصغر إلى الأكبر، فيكون لدينا:

6, 7, 8, 9, 10, 10, 11, 12, 13, 14, 17, 18, 19, 20, 20

من الواضح أن العلامة المطلوبة هي 12 فهناك سبع علامات أصغر منها وسبع أعلى منها. وتسمى العلامة 12 هذه الوسيط.

أما الوسيط لمجموعة من الأعداد المرتبة ترتيباً تصاعدياً أو ترتيباً تنازلياً هو العدد الأوسط منها إذا كان عددها فردياً، والوسط الحسابي للعددين الأوسطين في المجموعة إذا كان عددها زوجياً.

مثال (2): ما هو الوسيط للأعداد 19, 18, 2, 9, 8, 17, 5, 12, 11, 10

الحل: إن عدد البيانات زوجي، وبعد ترتيبها ترتيباً تصاعدياً تصبح:

2, 5, 8, 9, 10, 11, 12, 17, 18, 19

والسادس وهما 10, 11 إذن يكون الوسيط $\frac{10+11}{2} = 10.5$ لاحظ أن هناك خمسة أعداد أقل من 10.5

وهي 2, 5, 8, 9, 10 وخمسة أعداد أكبر منها وهي 11, 12, 17, 18, 19.

وإذا كان $x_1, x_2, \dots, x_n$ مجموعة من الأعداد المرتبة تصاعدياً (أو تنازلياً) فإن الوسيط لهذه المجموعة

هو العدد $\frac{x_{\frac{n+1}{2}} + x_{\frac{n}{2}+1}}{2}$ إذا كان n زوجياً.

محاضرات في الإحصاء -1-

لاحظ أن $x_{\frac{n+1}{2}}$ هو العدد الذي رقمه هو $\frac{n+1}{2}$ وأن $x_{\frac{n}{2}}$ هو العدد الذي رقمه $\frac{n}{2}$ وذلك بعد ترتيب البيانات.

مثال (3) : ما هو الوسيط للبيانات 9, 15, 5, 40, 29, 25, 21, 17, 11, 18,

الحل: بعد ترتيب البيانات حسب قيمها تصاعدياً تصبح:

5, 9, 11, 15, 17, 18, 21, 25, 29, 35, 40

لاحظ أن عدد البيانات 11 أي فردي. إذن $\frac{11+1}{2} = 6$ وبذلك يكون العدد السادس هو الوسيط وهو 18.

مثال (4) : أوجد الوسيط للأعداد 7, 8, 10, 11, 12, 19, 23, 25

الحل: البيانات مرتبة ترتيباً تصاعدياً وعددها 8 (أي أنه زوجي) إذن فالوسيط هو: الوسط الحسابي

للعددين الرابع والخامس، وبالتالي فإنه يساوي $\frac{11+12}{2} = 11 \frac{1}{2}$

مثال (5) : أوجد الوسيط للأعداد 2, 21, 19, 11, 17, 5, 14, 3, 6

الحل: رتب الأعداد ترتيباً تصاعدياً فتصبح 2, 3, 5, 6, 11, 14, 17, 19, 21

والآن، لاحظ أن عدد البيانات 9 (أي أنه فردي). إذن الوسيط هو العدد الذي ترتيبه $\frac{9+1}{2}$ أي أنه العدد

الذي ترتيبه الخامس، وبالتالي فالوسيط هو العدد 11.

2.2 الوسيط للتوزيع التكراري ذي الفئات

بعد أن تعرّفنا على طريقة إيجاد الوسيط من البيانات الأولية نحتاج إلى معرفة طريقة إيجاد الوسيط

للتوزيع التكراري ذي الفئات، حيث يتم عرض كثير من البيانات بواسطة التوزيع التكراري.

ولما كان الوسيط هو القيمة التي يكون عدد التكرارات الأقل منها مساوياً لعدد التكرارات الأعلى منها ففي

هذه الحالة نحتاج إلى معرفة التكرار المتجمع المقابل لأي فئة. نحصل على هذا التكرار بإيجاد مجموع

محاضرات في الإحصاء -1-

التكرارات التي تساوي قيمها الحد الأعلى الفعلي لتلك الفئة أو تقل عنه. ونبدأ دائما بالحد الأدنى الفعلي للفئة الأولى، ونعتبر تكراره المتجمع صفرا، حيث لا توجد بيانات تقل قيمتها عن ذلك الحد أو تساويه.

مثال (6) : جد التكرار المتجمع في التوزيع التكراري:

حدود الفئات	التكرار
10 – 13	5
14 – 17	11
18 – 21	8
22 – 25	6

الحل: نجد الحدود الفعلية للفئات ومن الواضح أن الحدود الفعلية للفئة الأولى هي 9.5 – 13.5 ومن الواضح أن عدد التكرارات التي قيمها أقل من 9.5 هو صفر. أما عدد البيانات التي قيمها أقل من 13.5 هو 5.

استمر في جمع تكرار كل فئة لمجموع تكرارات الفئات التي تسبقها لتحصل على التكرار المتجمع كما في الجدول التالي:

التكرار المتجمع	الحدود الفعلية
0	أقل أو يساوي 9.5
5	أقل أو يساوي 13.5
16	أقل أو يساوي 17.5
24	أقل أو يساوي 21.5
30	أقل أو يساوي 25.5

محاضرات في الإحصاء -1-

إذا كان لديك توزيع تكراري مجموع تكراراته n فإن الفئة الوسيطة هي أول فئة يزيد تكرارها المتجمع عن $\frac{n}{2}$ أو يساويه.

مثال (7) : ما هي الفئة الوسيطة في المثال (6) .

$$\text{الحل: } n = 30, \quad = 15 \frac{n}{2}$$

أنظر التوزيع التكراري المتجمع. ما هي أول فئة يزيد تكرارها المتجمع عن 15 أو يساويه؟. لاحظ أن العدد 16 هو الذي يزيد عن 15، لذلك فالفئة الوسيطة (13.5 – 17.5) ولإيجاد الوسيط للتوزيع التكراري. أفرض ما يلي:

$$n = \text{مجموع التكرارات}$$

$$c = \text{طول الفئة الوسيطة}$$

$$a = \text{الحد الأدنى الفعلي للفئة الوسيطة}$$

$$f_m = \text{تكرار الفئة الوسيطة}$$

التكرار المتجمع للفئة التي تسبق الوسيطة مباشرة n_1 وبعبارة أخرى n_1 هو مجموع التكرارات التي تقل قيمها عن a .

تعريف: الوسيط للتوزيع التكراري هو:

حيث الرموز هي التي افترضناها سابقا.

محاضرات في الإحصاء -1-

مثال (8) : جد الوسيط للتوزيع التكراري للمصروف الشهري لستين طالبا المعطى في الجدول التالي:

حدود الفئات	التكرار
32 – 35	10
36 – 39	6
40 – 43	8
44 – 47	4
48 – 51	8
52 – 55	9
56 – 59	15

الحل: نكتب عمود التكرار المتجمع مقابل عمود الحدود الفعلية للفئات، فنحصل على الجدول:

الحدود الفعلية للفئات	التكرار	التكرار المتجمع
31.5 – 35.5	10	10
35.5 – 39.5	6	16
39.5 – 43.5	8	24
43.5 – 47.5	4	28
47.5 – 51.5	8	36
51.5 – 55.5	9	45
55.5 – 59.5	15	60

محاضرات في الإحصاء -1-

لاحظ أن عدد التكرارات $n = 60$ وطول الفئة $c = 4$. جد الفئة الوسيطة وهي أول فئة يكون تكرارها

المتجمع $\frac{60}{2}$ أو أكثر. ومن الجدول تجد أن الفئة الوسيطة هي $47.5 - 51.5$ إذن $28a = 47.5$

$$f_m = 8 \text{ وعليه } n_1 =$$

الوسيط يساوي

مثال (9) : إذا كانت الأرباح اليومية لأربعين مكتبة كما يلي :

الحدود الفعلية (بالدينار)	التكرار
10.5 – 16.5	5
16.5 – 22.5	6
22.5 – 28.5	7
28.5 – 34.5	9
34.5 – 40.5	6
40.5 – 46.5	3
46.5 – 52.5	4

المطلوب : أوجد الوسيط.

محاضرات في الإحصاء -1-

الحل: نبنى التوزيع التكراري المتجمع فيكون كما في الجدول.

الحدود الفعلية	التكرار	التكرار المتجمع
10.5 – 16.5	5	5
16.5 – 22.5	6	11
22.5 – 28.5	7	18
28.5 – 34.5	9	27
34.5 – 40.5	6	33
40.5 – 46.5	3	36
46.5 – 52.5	4	40

من الجدول: $n=40$ ، طول الفئة $c = 6$ الفئة الوسيطة هي أول فئة يزيد تكرارها المتجمع عن أو يساوي

$$\frac{n}{2} = \frac{40}{2} = 20 \text{ وهي الفئة } 28.5 - 34.5 \text{ وبالتالي فإن } a = 28.5 \text{ و } n_1 = 18 \text{ وعليه } f_m = 9.$$

الوسيط هو:

3. الربيعات، العشریات و المئينات:

1.3. الربيعات: إضافة إلى الوسيط الذي يجزء المعطيات إلى قسمين متساويين هناك مقاييس أخرى تقسم

المعطيات إلى أربعة أجزاء وتسمى الربيعات. وهناك ثلاث ربيعات Q_1 : هو الربيع الأول، Q_2 : هو الربيع الثاني

وهو نفسه الوسيط، Q_3 : هو الربيع الثالث، ويمكن حساب الربيعات كما يلي:

$$Q_1: \text{الربيع الأول} \quad Q_1 = L_{Q_1} + \left(\frac{\frac{N}{4} - \sum_{i=1}^n n_{Q_1-1}}{n_{Q_1}} \right) \cdot c, \quad Q_3: \text{الربيع الثالث} \quad Q_3 = L_{Q_3} + \left(\frac{\frac{3N}{4} - \sum_{i=1}^n n_{Q_3-1}}{n_{Q_3}} \right) \cdot c$$

محاضرات في الإحصاء -1-

2.3. العشرية الأولى D_1 : هو القيمة التي تقسم مجموع عدد البيانات إلى عشر أقسام بحيث

عشر عدد البيانات أقل منه وتسعة أعشار عدد البيانات أكبر منه وترتيبه هو $\frac{N}{10}$ ونحسب بالتالي:

$$D_1 = L_{D_1} + \left(\frac{\frac{N}{10} - \sum_{i=1}^n n_{D_1-1}}{n_{D_1}} \right) \cdot c$$

3.3. المئينات: إذا قسمت البيانات إلى مائة قسم متساوي فإن نقاط التقسيم هذه تسمى المئينات. فالمئين

الأول C_1 هو القيمة التي يسبقها 1% من البيانات ويليه 99% من البيانات ويحسب كالتالي:

$$C_1 = L_{C_1} + \left(\frac{\frac{N}{100} - \sum_{i=1}^n n_{C_1-1}}{n_{C_1}} \right) \cdot c$$

والمئين العشرين C_{20} يحسب كالتالي:

$$C_{20} = L_{C_{20}} + \left(\frac{\frac{20N}{100} - \sum_{i=1}^n n_{C_{20}-1}}{n_{C_{20}}} \right) \cdot c$$

مثال 11: باستعمال بيانات مثال 9 أحسب الربيع الأول والربيع الثالث والعشير الثالث والمئين العشرين؟.

4. المنوال: يمثل المنوال القيمة التي تقابل أكبر تكرار مطلق أو أكبر تكرار نسبي في السلسلة الإحصائية

وبالتالي فهو عبارة عن القيمة الأكثر شيوعاً أو القيمة المسيطرة ونرمز له بالرمز Mo .

مثال 12: لدينا البيانات التالية: 2-4-5-5-5-6-6-7-7-7-8-8-9-9 ماهي قيمة المنوال؟

أما في حالة البيانات المبوبة فإننا نستطيع حساب المنوال بعد تحديد الفئة المنوالية باستخدام القانون

$$Mo = L_1 + \left(\frac{d_1}{d_1 + d_2} \right) \cdot c$$

حيث: L_1 : الحد الأدنى لفئة المنوال، d_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي سبقتها،

d_2 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي تليها، c : طول الفئة المنوالية.

محاضرات في الإحصاء -1-

مثال 13: باستعمال بيانات مثال 03 أوجد قيمة المنوال؟

1.4. خواص المنوال:

1 - لا يأخذ بعين الاعتبار جميع البيانات المعطاة وبالتالي فهو لا يتأثر بالقيم المتطرفة.

2 - يمكن حسابه بيانياً.

3 - يمكن أن يوجد أكثر من منوال لتوزيع واحد.

4 - يمكن حساب من الجداول الإحصائية المفتوحة.

5 - يعتبر أفضل المتوسطات لوصف الظواهر النوعية.

والمنوال لمجموعة من القيم هو القيمة الأكثر تكراراً في توزيع تكراري للقيم أي أن المنوال يكون تلك القيمة التي يقابلها أكبر تكرار في جوارها.

مثال (10): إذا كانت الأجور اليومية التي يتقاضاها عدد من العمال بالدينار كما في الجدول التالي، جد المنوال:

الأجر اليومي	عدد العمال
4	4
5	5
6	10
7	18
8	9
9	8
10	3

محاضرات في الإحصاء -1-

الحل: لاحظ أن القيمة 7 يقابلها أكبر تكرار في جوارها، ولذلك فإن المنوال = 7 دينار.

أما إذا لم يكن هناك قيم يقابلها تكرار أكبر من غيرها فلا يكون هناك منوال.

وفي بعض الأحيان يكون هناك أكثر من منوال، وتكرر كل منها أكبر من تكرار القيم في جواره. ومن

الأمثلة التي يحدث فيها منوالان أن تحتوي البيانات على مجموعتين مختلفتين، لكن أفراد كل مجموعة لها

قيم متقاربة من بعضها البعض. فمثلاً، لو أخذت مجموعة من معلمي المدارس الإعدادية وأخرى من

الأساتذة الجامعيين واعتبرت الرواتب للمجموعتين هي البيانات التي لديك.

أما في التوزيعات التكرارية ذات الفئات فإننا نعطي التعاريف الآتية :

1- تسمى الفئة (أو الفئات) التي يقابلها أكبر تكرار الفئة المنوالية (الفئات المنوالية).

2- مركز الفئة المنوالية هو " المنوال التقريبي " وإذا كان هناك عدة فئات منوالية فإنه يكون لدينا

عدة منوالات تقريبية.

مثال (11): جد المنوال للتوزيع التالي:

حدود الفئة	التكرار
25 – 29	5
30 – 34	7
35 – 39	8
40 – 44	15
45 – 49	9
50 – 54	12
55 – 59	4

الحل: واضح أن الفئة 44 – 40 فئة منوالية، لأن تكرارها 15 وهو أكبر من تكرارات الفئات المجاورة لها،

ولذلك فإن مركزها $42 = \frac{40+44}{2}$ هو منوال تقريبي.

واضح أيضاً أن الفئة 54 – 50 فئة منوالية، لأن تكرارها 12 أكبر من تكرارات الفئات المجاورة لها،

ولذلك فمركزها $52 = \frac{50+54}{2}$ منوال تقريبي أيضاً.

إن المقاييس السابقة، الوسط الحسابي والوسيط والمنوال تعتبر كلها مقاييس موقع، حيث أنها تعين قيمة أو

قيما تتمركز حولها البيانات بمعنى أو بآخر، وهي قيم لها نفس وحدات البيانات وعند رسمها بيانياً فإننا

نرصدها على المحور الأفقي الذي نعين عليه حدود الفئات.

5. العلاقة بين مقاييس النزعة المركزية:

1- تنطبق هذه المقاييس على بعضها وتتساوى في حالة التوزيع التكراري المعتدل أو المتناظر

$\bar{X} = M_e = M_0$ وعندها يكون شكل منحنى التوزيع على شكل جرس.

2 - عندما يكون التوزيع التكراري المدروس غير متناظر من اليمين، أي عندما تكون البيانات الصغيرة

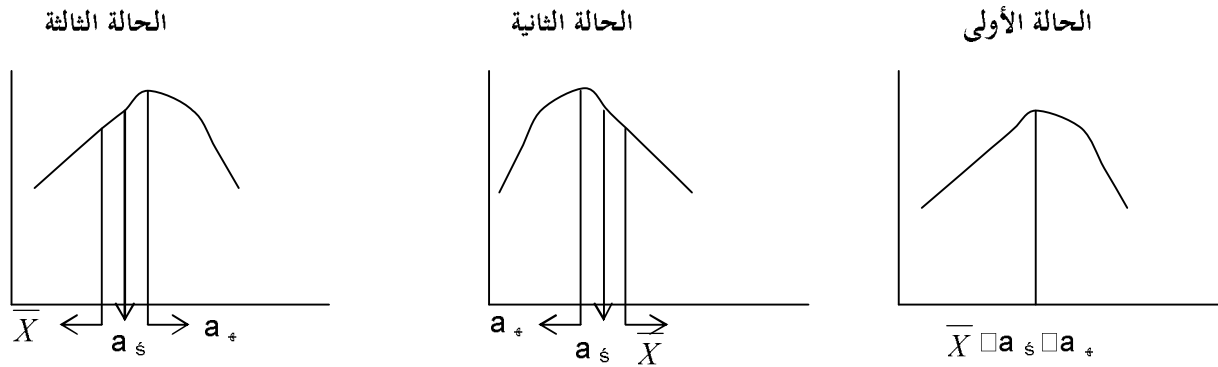
كثيرة (أكثر تكراراً) تكون المقاييس بالشكل التالي: $\bar{X} > M_e > M_0$ ويكون منحنى التوزيع ملتوي ناحية

اليمين.

3 - عندما يكون التوزيع التكراري المدروس غير متناظراً من اليسار، أي عندما تكون البيانات الكبيرة

كثيرة فإن المقاييس تكون بالشكل التالي: $\bar{X} < M_e < M_0$

محاضرات في الإحصاء -1-



وفي كل الحالات فقد حدد كارل بيرسون علاقة تجريبية بين المقاييس الثلاث إذا كان التوزيع قريباً جداً من التناظر وهي:

$$\text{المتوسط الحسابي} - \text{المنوال} = 3 (\text{المتوسط الحسابي} - \text{الوسيط}).$$

الفصل الرابع: مقاييس التشتت والشكل.

عرفنا في الفصل السابق أن مقاييس النزعة المركزية (من متوسط ووسيط ومنوال) تسمح لنا بالحصول على القيم المتوسطة للبيانات أو على تجمعها، غير أن هذه المقاييس لا تكفي لوحدها لمعرفة الصفات الإحصائية اللازمة لوصف الظواهر، لأن الفروق بين قيم الظواهر قد تزداد أو تنقص رغم تساوي المتوسطات لهذه الظواهر، ولتوضيح ما سبق نفترض أن طالبين تحصلا على النتائج التالية في خمس مواد دراسية: الطالب (X): 10، 11، 13، 14، 15. والطالب (Y): 8، 9، 13، 15، 18. فمتوسط درجات الطالب (X) يساوي 12،6 وكذلك متوسط درجات الطالب (Y) يساوي 12،6 ووسيط درجات الطالب (X) يساوي 13 وكذلك وسيط درجات الطالب (Y) يساوي 13. قد يفهم مما سبق أن الطالبين (X) و (Y) لهما نفس المستوى غير أن التمعن الجيد في الدرجات التي تحصل عليها الطالبين تبين أن الطالب (X) ناجح في كل المواد المدروسة في حين أن الطالب (Y) ناجح في ثلاث مواد فقط. إن هذه الحقيقة تبين أن مقاييس النزعة المركزية لا تعطي فكرة وافية عن اختلاف قيم الظواهر، ولا تحقق كل الأغراض التي نرغب الوصول إليها من دراستنا لذلك فإن مقاييس النزعة المركزية لا بد أن تكون مصحوبة بمقاييس أخرى لقياس مدى تباعد أو تقارب البيانات من بعضها البعض أو من متوسطها، تسمى هذه المقاييس بمقاييس التشتت. فما معنى التشتت؟.

محاضرات في الإحصاء -1-

تشتمل بيانات ظاهرة ما يقصد به درجة أو مقدار التفاوت أو الاختلاف بين مفردات هذه الظاهرة، وتعتبر بيانات الظاهرة متجانسة عندما تكون قيمتها قريبة من بعضها البعض ونقول في هذه الحالة أنها غير مشتتة. أما إذا كانت بيانات الظاهرة متباعدة وغير متجانسة فنقول أن مفردات الظاهرة مشتتة وغير مركزة. ويقاس تشتت البيانات بعدة مقاييس منها:

1- المدى العام: هو أسهل مقياس من مقاييس التشتت، فالمدى لمجموعة من البيانات هو الفرق

بين أكبر قيمة وأصغر قيمة لها ويرمز له بالرمز E .

$$\text{المدى} = \text{أكبر قيمة} - \text{أصغر قيمة} \quad E = X_{\max} - X_{\min}$$

أما المدى للتوزيعات التكرارية فيحسب بعدة طرق منها: المدى = مركز الفئة الأخيرة - مركز الفئة الأولى أو المدى = الحد الأعلى لفئة الأخيرة - الحد الأدنى للفئة الأولى

مثال 1: أوجد المدى للبيانات التالية: 30, 28, 22, 18, 12.

مثال 2: أوجد المدى للبيانات التالية 19, 18, 04, 17, 20, 65, 14.

نلاحظ المدى في هذا المثال قد تأثر بشكل كبير جدا بالقيم المتطرفة، إذ نلاحظ أن معظم البيانات متقاربة باستثناء القيمة 65 والقيمة 04، فإذا استبعدنا هذه القيم المتطرفة فإن المدى يصبح $E = 20 - 14 = 6$

وبسبب هذا العيب فإن المدى كمقياس للتشتت لا يستخدم إلا عندما نرغب في مقياس تقريبي وسريع لتشتت البيانات دون الاهتمام بالدقة في المقياس، أو عندما يكون للبيانات المتطرفة أهمية خاصة كتوزيعات درجات الحرارة على سبيل المثال، حيث تعلن درجات الحرارة اليومية بحددها الأقصى وحددها الأدنى خلال اليوم، كما يشيع استخدام هذا المقياس في حالات مراقبة جودة الإنتاج أو متابعة المبيعات

محاضرات في الإحصاء -1-

التي يحققها رجال البيع لمؤسسة ما. أما إذا أردنا أن نقلل من أثر القيم المتطرفة فإننا نقوم باستبعادها ويمكن أن يتم ذلك باستخدام الطرق التالية:

* المدى الربيعي = الربيع الثالث - الربيع الأول. * المدى العشري = العشير التاسع - العشير الأول.
* المدى المئيني = المئين 99 - المئين الأول.

1-1- خواص المدى:

1 - يتصف المدى بسهولة حسابه.

2 - يعتمد في حساب على قيمتين فقط هما القيمة الكبرى والقيمة الصغرى.

3- بسبب الخاصية الثانية فإن المدى شديد التأثر بالقيم المتطرفة.

2- الانحراف المتوسط: لأن مقاييس التشتت هي مقاييس لقوة تجمع البيانات حول بعضها، وحيث أن

التجمع يكون حول القيم المتوسطة، فإنه إذا كان مقدار الاختلاف بين القيم ومتوسطها كبيرا دل ذلك على

أن التشتت كبير والعكس صحيح. وحيث أن مجموع الاختلافات (الانحرافات) عن المتوسط يساوي صفرا

فإنه لو حسبنا القيم المطلقة لمقدار الاختلاف عن المتوسط يكون متوسط هذه الاختلافات مقياسا مناسباً

لمقدار التشتت، يسمى هذا المقياس بالانحراف المتوسط. ويعرف الانحراف المتوسط بأنه المتوسط

الحسابي للقيم المطلقة لانحرافات القيم عن متوسطها الحسابي، وسوف نرسم للانحراف المتوسط في

دراستنا بالرمز E_x وعليه:

إذا كانت لدينا القيم التالية: $X_1, X_2, X_3, X_4, \dots, X_n$ فإن الانحراف المتوسط لها هو:

$$E_x = \frac{|X_1 - \bar{X}| + |X_2 - \bar{X}| + |X_3 - \bar{X}| + \dots + |X_n - \bar{X}|}{n} = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

محاضرات في الإحصاء -1-

أما إذا كانت البيانات مكررة أو مبوبة في جداول توزيع تكرار فإن الانحراف المتوسط لها يعطي

$$E_X = \frac{\sum_{i=1}^n ni |X_i - \bar{X}|}{\sum_{i=1}^n ni} \text{ العلاقة:}$$

مثال 3: أوجد الانحراف المتوسط للبيانات التالية: 2، 4، 5، 6، 8؟

مثال 4: أوجد تشتت البيانات المبوبة في جدول التوزيع التكراري الآتي باستخدام الانحراف المتوسط؟

الفئة	2-0	4-2	6-4	8-6	المجموع
التكرار	2	3	4	3	12

ويعتبر الانحراف المتوسط أفضل من سابقه (المدى) لأنه أقل تأثر بالقيم المتطرفة غير أنه لا يستعمل

بشكل واسع بسبب اعتماده على القيمة المطلقة لانحرافات القيم عن متوسطها الحسابي.

1-2- خواص الانحراف المتوسط:

1- يعتمد في حسابه على جميع القيم وليس على القيمة الكبرى والصغرى فقط.

2- لا يمكن حسابه في حالة التوزيعات التكرارية المفتوحة.

3- يتأثر بالقيم المتطرفة، لأن انحرافها عن المتوسط الحسابي يكون كبيراً.

3- التباين والانحراف المعياري:

1-3- التباين: وهو عبارة عن المتوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي ومتوسطها

الحسابي، ونستخدم مربعات الفروق هنا تفادياً لاستخدام القيم المطلقة كما هو الشأن في الانحراف

المتوسط. فإذا كانت لدينا البيانات التالية: $X_1, X_2, X_3, X_4, \dots, X_n$ فإن التباين لهذه البيانات يعطي

بالعلاقة:

محاضرات في الإحصاء -1-

$$V_x = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{N}$$

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

مثال 5: أوجد التباين للبيانات التالية: 9,6,5,11,1,6,7,3.

في بعض الأحيان عندما يكون المتوسط الحسابي للبيانات عبارة عن كسر، فإن عملية حساب التباين تكون مضنية و عرضة للأخطاء الحسابية لذلك فإنه تم تطوير طريقة مختصرة لحساب التباين.

3-2- طريقة مختصرة لحساب التباين:

$$V_x = \frac{\sum (X_i - \bar{X})^2}{n}$$

انطلاقاً من العلاقة المتوصل إليها سابقاً:

$$V_x = \frac{\sum (X_i^2 - 2\bar{X}X_i + \bar{X}^2)}{n}$$

يمكن كتابة

$$V_x = \frac{\sum X_i^2}{n} - 2\bar{X} \frac{\sum X_i}{n} + \frac{n\bar{X}^2}{n}$$

$$V_x = \frac{\sum X_i^2}{n} - 2\bar{X}\bar{X} + \bar{X}^2$$

$$V_x = \frac{\sum X_i^2}{n} - 2\bar{X}^2 + \bar{X}^2$$

$$V_x = \frac{\sum X_i^2}{\sum ni} - \bar{X}^2$$

ومنه:

$$V_x = \frac{\sum n_i X_i^2}{\sum h_i} - \bar{X}^2$$

أما في حالة البيانات المبوبة فإن العلاقة تصبح:

3-3- الانحراف المعياري: ويعتبر الانحراف المعياري من أهم المقاييس الإحصائية للتشتت، وهو أكثر

استخداماً في النظريات والقوانين الإحصائية، لأنه يعطي فكرة سليمة ومنطقية عن ظاهرة التشتت، ويعرف

محاضرات في الإحصاء -1-

الانحراف المعياري بأنه الجذر التربيعي لمتوسط مجموع مربع انحراف القيم عن متوسطها، أي أنه الجذر التربيعي للتباين.

سوف نرمز للانحراف المعياري في دراستنا بالرمز (S_x) .

$$S_x = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} \quad \text{الانحراف المعياري لبيانات مفردة}$$

$$S_x = \sqrt{\frac{\sum h_i(X_i - \bar{X})^2}{\sum h_i}} \quad \text{الانحراف المعياري لبيانات متكررة أو مبوبة}$$

أما الصيغة المختصرة للانحراف المعياري فتعطى بالعلاقات التالية:

الانحراف المعياري لبيانات مفردة

$$S_x = \sqrt{\frac{\sum X_i^2}{n} - \bar{X}^2}$$

$$S_x = \sqrt{\frac{\sum h_i X_i^2}{\sum h_i} - \bar{X}^2} \quad \text{الانحراف المعياري لبيانات متكررة أو مبوبة}$$

مثال 6: أوجد التباين والانحراف المعياري بالصيغة الأصلية ثم بالصيغة المختصرة للبيانات المبوبة في

الجدول الإحصائي التالي:

الفئة	8-4	13-9	18-14	23-19	28-24	المجموع
التكرار	3	4	6	2	4	19

3-4 - خصائص الانحراف المعياري:

أ- إذا كان الانحراف المعياري للقيم $X_1, X_2, X_3, X_4, \dots, X_n$ هو S_x فإنه إذا أضيفت أو طرحت قيمة

$$S_x = \sqrt{\frac{\sum X_i^2}{n} - \bar{X}^2} \quad \text{ثابتة (a) إلى أو من جميع القيم فإن الانحراف المعياري لا يتغير}$$

بتربيع الطرفين

$$S_x^2 = V_x = \frac{\sum X_i^2}{n} - \bar{X}^2$$

$$nS^2 = \sum X_i^2 - n\bar{X}^2$$

بضرب الطرفين في (n)

نطرح قيمة ثابتة (a) من جميع القيم فنحصل على قيم جديدة $y_i = X_i - a$

$$\begin{aligned} nS_y^2 &= \sum (X_i - a)^2 - n(\bar{X} - a)^2 \\ &= \sum (X_i^2 - 2aX_i + a^2) - n(\bar{X} - a)^2 \\ &= \sum X_i^2 - 2a \sum X_i + na^2 - n(\bar{X}^2 - 2a\bar{X} + a^2) \\ &= \sum X_i^2 - 2a \sum X_i + na^2 - n\bar{X}^2 + 2an\bar{X} - na^2 \\ &= \sum X_i^2 - 2an \frac{\sum X_i}{n} - n\bar{X}^2 + 2an\bar{X} + na^2 - na^2 \\ &= \sum X_i^2 - 2a \sum X_i + na^2 - n\bar{X}^2 - 2an\bar{X} - na^2 \\ &= \sum X_i^2 - n\bar{X}^2 = nS_x^2 \end{aligned}$$

ومنه: $S_x = S_y$

ب- إذا كان الانحراف المعياري للقيم $X_1, X_2, X_3, X_4, \dots, X_n$ هو S_x فإنه إذا ضربت كل قيمة بالمقدار

a (قسمت كل قيمة على المقدار a) فإن الانحراف المعياري يتأثر بالمقدار نفسه أي أنه إذا كان $y_i = ax$

فإن:

$$nsy^2 = \sum y_i^2 - n\bar{y}^2 = a^2 [\sum X_i^2 - n\bar{X}^2]$$

$$nsy^2 = \sum (aX_i)^2 - na\bar{X}^2$$

$$nsy^2 = a^2 \sum X_i^2 - a^2 n\bar{X}^2$$

$$nsy^2 = a^2 (\sum X_i^2 - n\bar{X}^2)$$

$$Sy = asx$$

ومنه

ج- بالنسبة للتوزيع الطبيعي فإن:

* 68.27% من البيانات تقع في المجال $\bar{X} \pm Sx$

* 95.45% من البيانات تقع في المجال $\bar{X} \pm 2Sx$

* 99.73% من البيانات تقع في المجال $\bar{X} \pm 3Sx$

د- يأخذ الانحراف المعياري نفس وحدة القياس للمتغير الأصلي (كـلغ، متر، لتر) لذلك لا يمكن

استخدامها كأساس للمقارنة بين تشتت توزيعين لهما وحدات قياس مختلفة.

هـ- بما أن الانحراف المعياري يتأثر بالمتوسط الحسابي لبيانات الظاهرة فإنه لا يمكن استخدامه للمقارنة

بين تشتت بيانات توزيعين لهما متوسط حسابي مختلف ولو كان هذين التوزيعين من نفس النوعية

و- لا يمكن إيجاده بالنسبة للتوزيعات التكرارية المفتوحة من البداية أو النهاية.

ز- إذا كان لدينا مجموعة كلية مكونة من مجموعتين جزئيتين أو أكثر فإنه يمكن حساب الانحراف

المعياري لها من خلال العلاقة التالي:

$$S_x = \sqrt{\frac{1}{N}[n_1 S_{X_1}^2 + n_2 S_{X_2}^2] + \frac{1}{N}[h_1(\bar{X}_1 - \bar{X})^2 + h_2(\bar{X}_2 - \bar{X})^2]}$$

4- معامل الاختلاف: رأينا في الصفحات السابقة أن الانحراف المعياري هو مقياس واقعي ومؤشر

صحيح عن مقدار التشتت غير أن الخاصيتين 4 و 5 السابقتين تبينان أنه إذا استخدمنا هذا المقياس

للمقارنة بين تشتت ظاهرتين أو أكثر فإن المقارنة تكون واقعية وواقعية فقط إذا كانت الظواهر من

نوعية واحدة ولها متوسطات متساوية. أي يمكن مقارنة تشتت درجات مادة ما بدرجات مادة أخرى أو

مقارنة تشتت دخل مجموعة من العمال بدخل مجموعة أخرى، وتكون المقارنة أكثر واقعية إذا كانت

المتوسطات متساوية أو قريبة من بعضها.

أما إذا كانت الظواهر من صفات مختلفة أو إذا كانت متوسطاتها متباعدة، فإن المقارنة اعتمادا على

الانحراف المعياري ستكون غير منطقية وغير واقعية، ولهذا السبب وجدت مقاييس أخرى سميت بمقاييس

محاضرات في الإحصاء -1-

التشتت النسبي تعتمد على تمييز البيانات وتقيس التشتت كنسبة مئوية للمتوسط، أهم هذه المقاييس هو معامل الاختلاف.

$$CV = \frac{Sx}{X} \times 100 = \text{معامل الاختلاف}$$

مثال 7: إذا كان متوسط درجات مجموعة من الطلبة في مادة ما هو 15 بانحراف معياري 3 ومتوسط

درجاتهم في مادة أخرى هو 8 بانحراف معياري 2، فأى الدرجات في نظرك أكثر تشتتاً؟

مثال 8: ينتج مصنع نوعين من المصابيح الكهربائية، فإذا علمت أن المتوسط الحسابي والانحراف

المعياري لعمر المصباح في كل نوع هما:

$$\bar{X}_1 = 1500 \text{ ساعة} \quad Sx_1 = 300 \text{ ساعة}$$

$$\bar{X}_2 = 1800 \text{ ساعة} \quad Sx_2 = 325 \text{ ساعة}$$

أي المصابيح لها مدة حياة أكثر تشتت؟

إذا كانت مقاييس النزعة المركزية ومقاييس التشتت تسمح بتلخيص بيانات أي ظاهرة في صورة أرقام بسيطة تعطي فكرة عن خصائص توزيع هذه البيانات ودرجة تجانسها أو اختلافها فإن هذا الوصف الذي تعطيه يبقى تنقصه الدقة الكافية المطلوبة للتعرف على خواص التوزيع خاصة من حيث انتشار البيانات على المنحنى البياني الممثل لها من حيث التواءه أو تفلطحه عن الوضع الطبيعي* (المنحنى الطبيعي هو المنحنى المتماثل الذي يأخذ شكل الجرس والذي عنده تنطبق قيم المتوسط الحسابي والوسيط والمعدل)، كذلك دعت الحاجة لاستخدام مقاييس أخرى لتحقيق هذا الغرض سميت هذه المقاييس بمقاييس الالتواء والتفلطح التي سنتعرف عليها في هذا الفصل بعد التطرق لموضوع العزوم.

5- العزوم: العزوم قد تكون حول نقطة الأصل أو حول المتوسط الحسابي أو حول أي نقطة معينة،

فالعزم الأول حول نقطة الأصل مثلاً هو متوسط قيم الظاهرة والعزم حول المتوسط الحسابي هو متوسط

*

محاضرات في الإحصاء -1-

انحرافات قيم التوزيع عن المتوسط الحسابي لها. أما رتبة العزم فتحدد بدرجة القوة (الأس) التي ترفع إليها القيم أو انحرافاتهما عن المتوسط الحسابي. وتستخدم العزوم في إيجاد المعامل العزمي للإلتواء وكذلك معامل التفلطح.

5-1- العزوم حول نقطة الأصل: إذا كانت لدينا القيم $x_1, x_2, x_3, \dots, x_n$ فإن

$$\frac{\sum X_i}{n} = \text{العزم الأول} \quad \frac{\sum X_i^2}{n} = \text{العزم الثاني} \quad \frac{\sum X_i^3}{n} = \text{العزم الثالث} \quad \frac{\sum X_i^n}{n} = \text{العزم النوني}$$

مثال 1: إذا كانت لدينا القيم 1، 2، 5، 8 أوجد العزم الأول والثاني والثالث والرابع حول نقطة الأصل؟.

أما إذا كانت البيانات متكررة أو مبوبة في فئات فإن:

$$\frac{\sum n_i X_i}{\sum n_i} = \text{العزم الأول} \quad \frac{\sum n_i X_i^2}{\sum n_i} = \text{العزم الثاني} \quad \frac{\sum n_i X_i^3}{\sum n_i} = \text{العزم الثالث} \quad \frac{\sum n_i X_i^n}{\sum n_i} = \text{العزم النوني}$$

حيث n_i = التكرار.

$\sum n_i$ = مجموع التكرارات.

x_i = القيم أو مراكز الفئات.

مثال 2: أوجد العزم الأول والثاني والثالث للبيانات المبوبة في جدول التوزيع التالي، ثم أوجد التباين والانحراف المعياري؟.

الفئة	3 - 1	5 - 3	7 - 5	9 - 7	11 - 9	المجموع
التكرار	1	2	3	4	6	16

5-2- لعزوم حول المتوسط الحسابي:

ونرمز للعزم حول المتوسط الحسابي بالرمز μ .

محاضرات في الإحصاء -1-

العزم الأول حول المتوسط الحسابي: $\mu_1 = \frac{\sum (X_i - \bar{x})}{\sum n}$ ، العزم الثاني حول المتوسط الحسابي:

$$\mu_2 = \frac{\sum (X_i - \bar{x})^2}{\sum n}$$

العزم الثالث حول المتوسط الحسابي: $\mu_3 = \frac{\sum (X_i - \bar{x})^3}{\sum n}$ ، العزم النوني حول المتوسط الحسابي:

$$\mu_n = \frac{\sum (X_i - \bar{x})^n}{\sum n}$$

مثال 3: أوجد العزم الأول والثاني والثالث حول المتوسط الحسابي للقيم التالية 2، 4، 6؟

أما إذا كانت البيانات متكررة أو مبوبة في جداول توزيع تكراري فإن العزم حول المتوسط الحسابي يمكن حسابها على النحو التالي:

العزم الأول حول المتوسط الحسابي = $\mu_1 = \frac{\sum n_1 (X_i - \bar{x})}{\sum n_1}$ ، العزم الثاني حول المتوسط الحسابي:

$$\mu_2 = \frac{\sum n_1 (X_i - \bar{x})^2}{\sum n_1}$$

العزم الثالث حول المتوسط الحسابي: $\mu_3 = \frac{\sum n_1 (X_i - \bar{x})^3}{\sum n_1}$ ، العزم النوني حول المتوسط الحسابي:

$$\mu_n = \frac{\sum n_1 (X_i - \bar{x})^n}{\sum n_1}$$

مثال 4: أوجد العزم الأول والثاني حول المتوسط الحسابي للبيانات المبوبة في الجدول التالي:

الفئة	2 - 0	4 - 2	6 - 4	8 - 6	المجموع
التكرار	2	4	6	4	16

5-3- تحديد شكل التوزيع:

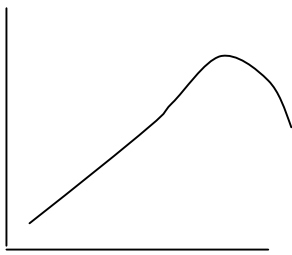
يمكن تحديد شكل التوزيع باستخدام مقاييس الالتواء والتفلطح.

5-3-1- الالتواء: (عدم التناظر من اليمين أو من اليسار مقارنة بتوزيع متناظر بالنسبة للقيمة

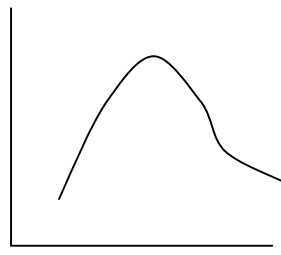
المركزية) يعتبر منحنى التوزيع التكراري المعتدل هاما جدا في الدراسات والتحليلات الإحصائية. إن هذا المنحنى الذي تتساوى عنده مقاييس النزعة المركزية الثلاث ($\bar{X} = Me = M0$) نظري ونادر الوقوع، فالمنحنيات التي نحصل عليها عادة تكون ملتوية ناحية اليمين أو ناحية اليسار أو قريبة من الاعتدال .

فقد عرفنا عند دراستنا لمقاييس النزعة المركزية أن التوزيعات الإحصائية يمكن أن تأخذ أحد الأشكال

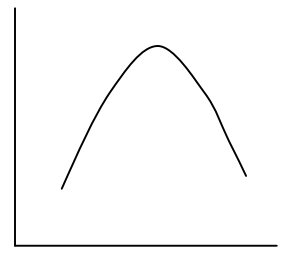
التالية حسب العلاقة بين المقاييس الثلاث:



التواء ناحية اليسار



التواء ناحية اليمين



توزيع متناظر

$$\bar{X} \quad Mo > Me > \quad \bar{X} \quad Mo < Me < \quad Me = M0 = \bar{X}$$

أما في هذا الفصل فسنحاول معرفة درجة تناظر (تماثل) توزيع البيانات حول المتوسط الحسابي باستخدام

العزوم والانحراف المعياري.

(أ) معامل فيشر للالتواء:

محاضرات في الإحصاء -1-

يقيس هذا المعامل درجة إلتواء شكل التوزيع الإحصائي ويعتمد في ذلك على قيمة العزم الثالث حول

المتوسط الحسابي، ولاستبعاد وحدة القياس نقسمه على الانحراف المعياري من نفس المرتبة. $F_1 = \frac{u^3}{S_x^3} *$

(تم إختيار $u3$ لقياس الالتواء في المعاملين لأن قيمته تساوي صفر في التوزيع المعتدل).

ويكون لدينا ثلاث حالات هي:

- توزيع إحصائي متناظر $\Rightarrow F_1=0$,
- منحنى التوزيع غير متناظر ملتوي ناحية اليمين $\Rightarrow F_1>0$
- منحنى التوزيع غير متناظر ملتوي ناحية اليسار $\Rightarrow F_1<0$.

(ب) معامل بيرسون للالتواء :

$$P_1 = \frac{(u3)^2}{(u_2)^3} \text{ وتكون لدينا ثلاث حالات كذلك}$$

توزيع إحصائي متناظر $\Rightarrow P_1=0$ ، منحنى التوزيع غير متناظر ناحية اليمين $\Rightarrow P_1>0$ ، منحنى

التوزيع غير متناظر ملتوي ناحية اليسار $\Rightarrow P_1<0$.

(ج) معامل يول وكندال للالتواء:

ويستعمل هذا المعامل بالنسبة للجداول الإحصائية المفتوحة.

$$Cyk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \text{ أما الحالات الممكنة فهي:}$$

- توزيع إحصائي متناظر $\Rightarrow Cyk = 0$
- منحنى التوزيع غير متناظر ملتوي ناحية اليمين $\Rightarrow Cyk>0$
- منحنى التوزيع غير متناظر ملتوي ناحية اليسار $\Rightarrow Cyk<0$.

محاضرات في الإحصاء -1-

5-3-2- التفلطح (تطاول أو تفلطح المنحنى مقارنة بالتوزيع الطبيعي):

يقصد بالتفلطح مدى اتساع وضعف قمة منحنى التوزيع ولقد أُصطلح على اعتبار منحنى التوزيع الطبيعي متوسط التفلطح. وتوجد كذلك عدة معاملات لقياس التفلطح أهمها:

(أ) معامل بيرسون للتفلطح:

$$P_2 = \frac{\mu_4}{(S_x)^4} = \frac{\mu_4}{(\mu_2)^2} \text{ والحالات الممكنة هي:}$$

توزيع معتدل التفلطح (توزيع طبيعي) $\Rightarrow P_2 = 3$. منحنى التوزيع متطاول (مدبب) $P_2 > 3$ منحنى

التوزيع متفلطح $P_2 < 3$

(ب) معامل للتفلطح : وهو عبارة عن معامل بيرسون مطروحا منه 3.

$$F_2 = P_2 - 3 = \frac{\mu_4}{(\mu_2)^2} - 3 \text{ والحالات الممكنة هي: منحنى التوزيع معتدل التفلطح } F_2 = 0 \text{ منحنى}$$

التوزيع معتدل التفلطح $F_2 > 0$ منحنى التوزيع متفلطح $F_2 < 0$

مثال 4: أدرس شكل منحنى التوزيع التكراري الآتي باستخدام معامل فيشر للالتواء ومعامل بيرسون

للتفلطح؟.

المجموع	4	3	2	1	X_i
20	1	4	9	6	n_i

3- معامل كيلي للتفلطح ويستخدم عندما يكون جدول التوزيع التكراري مفتوح من البداية أو من

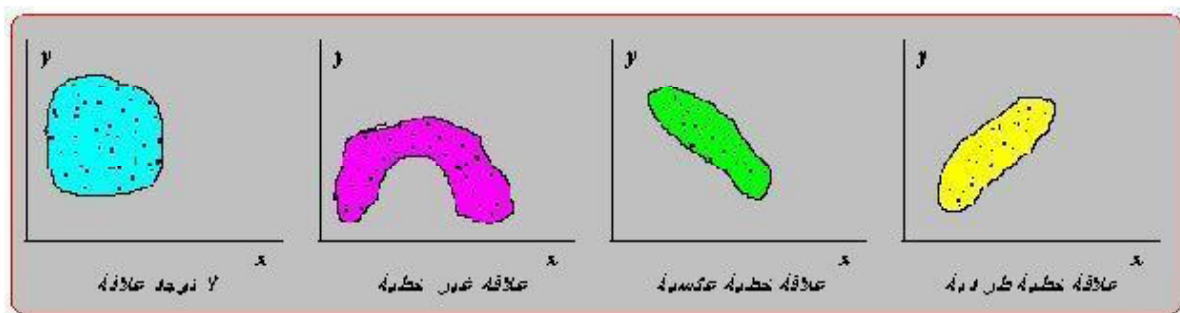
النهاية، ويعطي هذا المعامل بالعلاقة التالية:

$$C_k = \frac{1}{2} \cdot \frac{Q_3 - Q_1}{D_9 - D_1}$$

الفصل الخامس: الارتباط والانحدار.

يطلق الارتباط على العلاقة بين متغيرين $(X; Y)$ مثل العلاقة بين علامة الطالب في مادة الفيزياء وعلامته في مادة الرياضيات، أو العلاقة بين معدله في الدراسة وعدد ساعات الدراسة، أو العلاقة بين دخل الفرد وإنفاقه، أو العلاقة بين كميات السماد المستخدمة وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد، أو العلاقة بين وزن الجسم وضغط الدم. وهناك كثير من العلاقات... والأمثلة على ذلك في المجال التطبيقية كثيرة فإذا كان لدينا المتغيرين $(x; y)$ وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانياً في ما يسمى بشكلاً لانتشار (السحابة النقطية)، فإن العلاقة بينها تأخذ أشكالاً مختلفة علنا نلخصها في التالي:

الشكل رقم (05-01): شكلاً لانتشار لبيان نوع العلاقة بين y, x



1- الارتباط الخطي البسيط:

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين، يستخدم تحليل الارتباط، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر، يستخدم تحليل الانحدار، وفي

هذا الفصل يتم عرض أساليب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين

المتغيرين تأخذ الشكل الخطي، وسوف نعرض حسابات في حالة البيانات الكمية، والبيانات الوصفية المقاسة بمقياس ترتيبي.

محاضرات في الإحصاء -1-

2- الغرض من تحليل الارتباط الخطي البسيط:

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو) وفي حالة العينة بالرمز r ، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومنا لتحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

• نوع العلاقة: وتأخذ ثلاثاً أنوعاً حسب إشارة معامل الارتباط كما يلي:

✓ إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن

زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

✓ إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن

زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس.

✓ إذا كان معامل الارتباط قيمته صفراً ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

• قوة العلاقة: ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث

أن قيمة معامل الارتباط تقع في المدى ($-1 < r < +1$) وقد صنف بعض الإحصائيين درجات

لقوة العلاقة كما تمثيلها على الشكل التالي:

الشكل رقم (05-02): درجات قوة معامل الارتباط

ارتباط عكسي					ارتباط طردي					
قوي جدا	قوي	متوسط	ضعيف	شبه جدا	شبه جدا	ضعيف	متوسط	قوي	قوي جدا	
-1	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9	1
نام					نام					

محاضرات في الإحصاء -1-

3- معامل الارتباط الخطي البسيط لبيرسون (Pearson):

يستخدم إذا كان كلا المتغيرين مقاسا بمقياس كمي مثل إيجاد معامل الارتباط بين الدخل ومستوى الادخار، ويحسب باستخدام العلاقة التالية:

$$r_p = \frac{COV(X;Y)}{\sigma_X \cdot \sigma_Y} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}}$$

حيث أن: $\bar{X}$ و $\bar{Y}$: متوسط المتغيرين X_i و Y_i على الترتيب، σ_X و σ_Y : الانحراف المعياري

للمتغيرين X_i و Y_i على الترتيب، أما $COV(X;Y)$: فهو التباين المشترك للمتغيرين X_i و Y_i

ويمكن اختصار الصيغة السابقة علنا نحو التالي: $r_p = \frac{\frac{\sum_{i=1}^n X_i Y_i}{n} - \bar{X}\bar{Y}}{\sigma_X \cdot \sigma_Y}$ حيث: $\sum_{i=1}^n X_i Y_i$: مجموع حاصل ضرب X_i في Y_i .

مثال 01: الجدول التالي يبين دخل (X) ومستوى الادخار (Y) بآلاف الدينارات:

الدخل	14	12	08	16	11	10	09	07
الادخار	07	08	06	10	09	05	06	04

المطلوب حساب معامل الارتباط بين الادخار والدخل، وما هو مدلولة؟

4- معامل الارتباط للرتب:

يسمى أيضا بمعامل سبيرمان (Spearman) ويستخدم هذا المعامل إذا كان كلا من المتغيرين ترتيبيين أو

أحدهما كمي و الآخر ترتيبي مثل إيجاد العلاقة مستوى الدخل (مرتفع - متوسط - منخفض) وعدد

محاضرات في الإحصاء -1-

ساعات العمل اليومية (أكثر من 8 ساعات - من 5 ساعات إلى 8 - أقل من 5 ساعات)، كما يمكن استخدام مقياس سبيرمان في حالة المتغيرات الكمية أيضا، ويعبر عنها بالعلاقة التالية:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)}$$

حيث: d : الفرق بين رتب X و رتب Y (ترتيب) $\sum d = 0$.

مثال 02: فيما يلي نبين التقديرات التي حصل عليها ثمانية من الطلبة في مادتي الرياضيات و الإحصاء.

تقديرات الرياضيات	تقديرات الإحصاء
ضعيف	ضعيف
مقبول	مقبول
ممتاز	جيد جدا
جيد جدا	ممتاز
جيد	مقبول
مقبول	جيد
مقبول	ضعيف
ضعيف جدا	ضعيف جدا

المطلوب إيجاد معامل الارتباط باستعمال صيغة سبيرمان، وما هو مدلوله؟

إن

الهدف من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمّي على متغير كمّي آخر، ومنا أمثلة على

ذلك ما يلي: دراسة أثر الإنتاج على التكلفة، أثر الدخل على الإنفاق الاستهلاكي.

5- نموذج الانحدار الخطي البسيط:

فيتحلى الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير

المستقل، علما أن المتغير الثاني ويسمى بالمتغير التابع، ومن ثم يمكن عرض نموذج

الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي :

$$Y_i = a + bX_i + e_i \dots \dots \dots (1)$$

Y_i : هو المتغير التابع الذي يتأثر .

X_i : هو المتغير المستقل .

b : هو الميل ، a : الثابت .

e_i : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية Y_i ، والقيمة المقدرة $\hat{Y} = \hat{a} + \hat{b}X$: أي أن :

$$e = Y_i - \hat{Y} = Y_i - (\hat{a} + \hat{b}X)$$

5-1- تقدير نموذج الانحدار الخطي البسيط:

يمكن تقدير معاملات الانحدار $(a; b)$ في النموذج (1) باستخدام طريقة المربعات

$$\sum e_i^2 = \sum [Y_i - (\hat{a} + \hat{b}X)]^2$$

الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات أخطاء العشوائية

أقل ما يمكن، وبحسب هذا التقدير بالمعادلة التالية:

$$\hat{b} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$\hat{a} = \bar{Y} - \hat{b} \bar{X}$$

حيث أن: $\bar{X}$: هو الوسط الحسابي لقيم X_i ، و $\bar{Y}$: هو الوسط الحسابي لقيم Y_i . وتكون القيمة المقدرة للمتغير التابع هو :

$$\hat{Y} = \hat{a} + \hat{b}X$$

ويطلق على هذا التقدير تقدير معادلة انحدار Y_i على X_i .

محاضرات في الإحصاء -1-

5-2- الخطأ المعياري للتقدير:

يبين هذا المقياس انحراف القيم المحسوبة عن القيم الفعلية، الخطأ المعياري لتقدير Y على X هو:

$$\sigma_{YX} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n}}$$

$$\sigma_{XY} = \sqrt{\frac{\sum (X_i - \hat{X})^2}{n}} \text{ أما الخطأ المعياري لتقدير } X \text{ على } Y \text{ هو:}$$

5-3- معامل التحديد: (R^2):

هو النسبة بين التباين المفسر والتباين الإجمالي: $R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$ ويفسر على أنه نسبة التغير في

المتغير التابع التي يشرحها أو يفسرها التغير في المتغير المستقل أما النسبة المتبقية فهي راجعة لأسباب أو متغيرات أخرى لم تؤخذ بعين الاعتبار في النموذج.

5-4- معامل الارتباط: (r):

إن الجذر التربيعي لمعامل التحديد يعطي لنا قيمة معامل الارتباط الذي يدل على شدة الارتباط بين

$$r = \sqrt{\frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}} \text{ المتغيرين:}$$

مثال 03: باستعمال معطيات مثال 01:

المطلوب:

1- أوجد معادلة انحدار الادخار Y_i على الدخل X_i ؟

2- أوجد معادلة انحدار الادخار X_i على الدخل Y_i ؟

3- أحسب الخطأ المعياري لتقدير الادخار على الدخل؟

الفصل السادس: الأرقام القياسية.

يعود استخدام الأرقام القياسية الى أكثر من قرنين من الزمن، حيث استخدمها الإحصائي الإيطالي كارلي 1764 لمقارنة الأسعار في إيطاليا لسنة 1750 بالأسعار في سنة 1500، ويمكن تعريف الأرقام القياسية هي طريق مختصر لوصف المتغيرات الاقتصادية. وهي كثيرا ما تستعمل لوصف التغير في الأسعار أو الكميات أو القيمة في الزمن.

فالرقم القياسي للأسعار هو عدد واحد يرينا كيف تغيرت مجموعة كاملة من الأسعار. والرقم القياسي لتكاليف المعيشة يعطينا نسبة التغير في تكلفة الحياة بين سنة وأخرى مقاسة بتغير أسعار سلة من المواد من وقت لآخر.

1- مفهوم الأرقام القياسية وأنواعها واستخدامها: فمثلاً لو قلنا أن الرقم القياسي لتكاليف المعيشة بلغ 180 في عام 2000 مقارنة لعام 2005 فهذا يعني أن ما تحتاج إلى إنفاقه عائلة في عام 2000 لشراء عدد من المواد (عادة تحدد هذه المواد عند تعريف الرقم القياسي) يزيد 80% عما كانت تحتاجه هذه العائلة عام 2000. أي أن تكاليف الحياة قد ارتفعت 80 % في عام 2005 عما كانت عليه عام 2000.

ويمكن استعمال الأرقام القياسية لقياس التغير في الإنتاجية والبطالة ونمو الموارد ومعدلات الأجور وغيرها. ويستفاد من معرفة الأرقام القياسية لتكاليف الحياة في تقرير الزيادة في الرواتب والأجور وحساب راتب الضمان الاجتماعي والتقاعد.

وبمعرفة الأرقام القياسية التي تعطينا فكرة واضحة عن التغير في أسعار المواد الغذائية أو تكاليف العلاج تستطيع العائلة من تنظيم ميزانية مصروفاتها ونفقاتها لتناسب مع دخلها.

وعادة ما نحسب:

محاضرات في الإحصاء -1-

- 1- الرقم القياسي للأسعار: الذي يقارن التغير في السعر من فترة لأخرى.
 - 2- الرقم القياسي للكميات: الذي يقيس التغير في كمية أحد المتغيرات مع الزمن.
 - 3- الرقم القياسي للقيمة: الذي يقيس التغير في القيمة المالية للمتغير. أي أنه يقيس القيمة النقدية (بالدينار) لمتغير معين. والرقم القياسي للقيمة يركب التغير في الأسعار والكميات ليعطي رقما أكثر دلالة ومعنى.
- ومن أهم وجوه استخدام الأرقام القياسية تقدير الدخل الحقيقي للفرد، ويعرف الرقم القياسي للدخل الحقيقي للفرد في سنة معينة (سنة المقارنة) إلى سنة أخرى (سنة الأساس) بأنه:

R.

ويمكن الضرب بـ 100 لجعل الرقم نسبة مئوية.

فمثلا، إذا كان الرقم القياسي لدخل الفرد عام 1995 بالنسبة لعام 1990 يساوي 160 وكان الرقم القياسي لتكاليف المعيشة لعام 1995 بالنسبة لعام 1990 يساوي 195 فإن الرقم القياسي للدخل الحقيقي: $82.05 = 100 \times \frac{160}{195}$ أي أن الدخل الحقيقي للفرد عام 95 (القوة الشرائية لدخل الفرد) أصبحت 82 بالمائة مما كانت عليه 1990.

2- طرق تركيب الأرقام القياسية: هنا عدة طرق لتركيب الأرقام القياسية لإنها مهما اختلفت طريقة التركيب فإنها تشترك في الخطوات التالية:

1-2- اختبار السلع والقيم التي تدخل في تركيب الرقم القياسي وجمع البيانات عنها:

بعد معرفة الهدف الذي نسعى إليه من تركيب الرقم القياسي المطلوب تبدأ عملية

اختبار الأرقام والسلع التي فيها الغرض المطلوب ويوجب ملاحظة خصائص

السلاسل الزمنية المطلوبة من حيث طول السلسلة أو وحدة الفترة الزمنية وكذلك

محاضرات في الإحصاء -1-

ملاحظة الفرق فيما لأنواع مختلفة للسلع والوحدات الزمنية المستخدمة . ومن

هذا يجب اختبار القيمة التي تحقق الهدف المطلوب بالدقة الممكنة . ويمكن جمع

البيانات الخاصة بهذه القيم من مصادرها الصحيحة والدقيقة.

2-2- اختبار فترة الأساس: يعتمد اختبار فترة الأساس على العوامل التالية:

أ- الغرض من تركيب الرقم القياسي المطلوب.

ب- بعد فترة الأساس على التغييرات المفاجئة أو الدورية حتى لا تتأثر الأرقام المحسوبة منها.

ج- يجب أن تكون فترة الإحصاء قريبة كلما أمكن من فترة المقارنة حتى تتشابه الظروف والقائمة بقدر الإمكان وبذلك تسهل المقارنة.

2-3- اختبار الأوزان المناسبة ونظام الترجيح

من المعلوم هنا أن اختلاف أهمية السلع بعضها عن البعض الآخر ولذلك أخذ

نظام ترجيح بعض القيم والسلعنا الآخر بما يتناسب أهميتها ولذلك اختار

بعض الأوزان لتعطي لهذا القيم، فمنها ما يعطى حسب أهمية السلعة مثلاً أو حسب

ما يستهلك منها وهكذا أما بالنسبة للأسعار فقد ترجح سواء لكمية فترة الأساس أو كمية فترة المقارنة وحسب الظروف

3- الأرقام القياسية البسيطة: تعتمد هذه الأرقام في تركيبها على قيمة ظاهرة أو قيم ظواهر (متغيرات)

في أزمنة أو أماكن مختلفة، ويمكن حسابها بطريقتين ونقتصر في حسابها على الأسعار:

(1) الرقم القياسي التجميعي البسيط للأسعار: هي نسبة مجموع أسعار عدة سلع في سنة ما (تسمى نسبة

المقارنة) إلى مجموع أسعار هذه السلع في سنة أخرى (تسمى سنة الأساس).

فلو كان لدينا الجدول التالي: حيث P_{0i} سعر السلعة i في سنة الأساس، P_{ni} سعر السلعة i في سنة

المقارنة.

محاضرات في الإحصاء -1-

السنة	السعر عام 1992 (سنة الأساس)	السعر عام 1998 (سنة المقارنة)
(1)	P_{o1}	P_{n1}
(2)	P_{o2}	P_{n2}
.	.	.
.	.	.
(m)	P_{om}	P_{nm}

فيكون الرقم القياسي التجميعي البسيط للأسعار عام 1998 مقارنة مع أسعار 1990 هو:

ولتبسيط استعمال الرموز نتوقف عن استعمال (i) ونكتفي بالرمز لسنة الأساس بالحرف (o) وسنة المقارنة بالحرف (n) وبالتالي يكون:

تعريف (1): الرقم القياسي التجميعي للأسعار هو: $I_p(a) = \frac{\sum P_n}{\sum P_o} \times 100$

ونضع (a) لتدل على التجميعي ونضع P لتدل على الأسعار (Price).

نصطلح على إعطاء الأرقام القياسية كنسب مئوية ولذلك ضربنا بالعدد 100.

مثال (1): كانت الأسعار (بالدينار / كغم) لبعض المواد الاستهلاكية كما في الجدول التالي:

السنة	السعر سنة 2005	السعر سنة 2011
السكر	200	300
الأرز	240	400
الشاي	1500	1800
القهوة	2200	4500
المجموع	4140	7000

أحسب الرقم القياسي التجميعي البسيط للأسعار باعتبار عام 2005 سنة أساس.

الحل: الرقم القياسي التجميعي البسيط للأسعار = $\frac{\sum P_n}{\sum P_o} \times 100$

$$\frac{7000}{4140} \times 100 = 169.1 \text{ ويساوي:}$$

لاحظ أنه يجب أن تكون وحدة الأسعار هي نفسها لجميع السلع. أي إما جميعها بالدينار.

(2) الرقم القياسي النسبي البسيط للأسعار: هو الوسط الحسابي للأرقام القياسية للسلع. أي أننا نجد الرقم

القياسي لكل سلعة ثم نجد الوسط الحسابي لهذه الأرقام القياسية.

تعريف (2): الرقم القياسي النسبي البسيط للأسعار هو: $I_p(r) = \frac{1}{m} \sum \frac{P_n}{P_o}$

حيث m هو عدد السلع.

ونضع (r) لتدل على " النسبي "، ونضع P لتدل على الأسعار (Price).

مثال (2): أوجد الرقم القياسي النسبي البسيط للأسعار في مثال (1) باعتبار عام 2005 سنة أساس.

الحل:

الرقم القياسي النسبي البسيط لأسعار 2005 هو:

4- الأرقام القياسية المرجحة: إن من سلبيات الأرقام القياسية البسيطة أنها تعطي جميع السلع الأهمية

نفسها، أي أنها لا تعطي أهمية، أو وزناً، للتغير في سعر السلعة الأكثر استعمالاً أكبر مما تعطيه للسلعة

قليلة الاستعمال. فعلى سبيل المثال، الأرقام القياسية البسيطة للأسعار تعامل التغير في سعر التلفزيون أو

آلة التسجيل أو الآلة الحاسبة اليدوية مثلما تعامل التغير في سعر السكر والأرز والشاي مع أن شراء

السلع الثلاث الأول يكاد يكون مرة واحدة في الفترة الزمنية الطويلة (سنة أو أكثر) بينما يكون استعمال

السلع الثلاث الأخرى متكرراً في اليوم الواحد أو الأسبوع الواحد.

محاضرات في الإحصاء -1-

مما سبق بات ضرورياً أن نعطي لبعض السلع أهمية أكبر ووزناً أكبر مما نعطيه للسلع الأخرى. إن هذا الترجيح يعطينا الفرصة لاستعمال معلومات إضافية زيادة عن المعلومات المتعلقة بتغير الأسعار. وتبقى مسألة تحديد الوزن الذي نستعمله مع كل واحد من المتغيرات (السلع) تحت الدراسة.

وهناك العديد من الطرق التي تعين الأوزان ندرس منها:

4-1- طريقة لاسبير: تتلخص هذه الطريقة باستعمال الكميات المستهلكة والقيمة النقدية للكميات المستهلكة في سنة الأساس كأوزان لأسعار المواد الداخلة في حساب الرقمين القياسيين التجميعي والنسبي على التوالي.

تعريف (3): رقم لاسبير التجميعي للأسعار هو: نسبة ما تنفقه في سنة المقارنة إلى ما تنفقه في سنة الأساس على جميع السلع إذا أردنا استهلاك نفس كميات سنة الأساس في الحالين، أي:

حيث:

$$\bullet P_n = \text{أسعار سنة المقارنة.}$$

$$\bullet P_o = \text{أسعار سنة الأساس.}$$

$$\bullet Q_o = \text{الكميات المستهلكة في سنة الأساس.}$$

واستعملنا الرمز L ليدل على لاسبير.

تعريف (4): رقم لاسبير النسبي للأسعار هو:

$$\text{حيث: } W_o = \frac{P_o Q_o}{\sum P_o Q_o}$$

واستعملنا الرمز RL ليدل على لاسبير النسبي، والأوزان تعتمد على النفقات (القيم المالية) لسنة الأساس.

محاضرات في الإحصاء -1-

مثال (3): يبين الجدول التالي أسعار عدد من السلع (دينار / كغم) وكميات الاستهلاك منها بالكغم للعائلة الواحدة شهريا.

السلعة	السعر عام 1993	كمية الاستهلاك عام 1993	السعر عام 1999	كمية الاستهلاك عام 1999	
السكر	220	7	350	8	1540
الأرز	280	10	430	12	2800
الشاي	1700	1.5	3000	1.5	2550
لحم الضأن	2800	5.5	4000	6.5	15400
					22290

أ- أحسب رقم لاسبير القياسي التجميعي لأسعار 1999 باعتبار سنة 1993 سنة أساس.

ب- أحسب رقم لاسبير النسبي لأسعار 1999 باعتبار سنة 1993 سنة أساس.

الحل: (أ)

$$I_p(L) = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

$$= \frac{3.5}{2.5}$$

(ب) نحسب أولا W_0 ونضعه في عمود (6) في الجدول:

$$W_0 = \frac{P_0 Q_0}{\sum P_0 Q_0} = \frac{220 \times 7}{22290} = \frac{154}{2229} \text{ للسكر}$$

$$W_0 = \frac{P_0 Q_0}{\sum P_0 Q_0} = \frac{28}{2229} \text{ للأرز وهكذا لبقية السلع.}$$

نحسب الآن رقم لاسبير النسبي للأسعار وهو:

$$I_p(rL) = \left[\sum \right]$$

=

×

4-1- طريقة باش: نتلخص هذه الطريقة باستعمال الكميات المستهلكة (أو المنتجة) والقيمة النقدية

للكميات المستهلكة في سنة المقارنة أوزانا لأسعار المواد الداخلة في حساب الرقمين القياسيين

التجميعي والنسبي على التوالي.

تعريف (5): رقم باش التجميعي للأسعار هو: نسبة ما ننفقه في سنة المقارنة إلى ما ننفقه في سنة

الأساس على جميع السلع إذا أردنا استهلاك نفس كميات سنة المقارنة في الحالين، أي:

حيث:

• P_n = أسعار سنة المقارنة.

• P_o = أسعار سنة الأساس.

• Q_n = الكميات المستهلكة (أو المنتجة) في سنة المقارنة.

واستعملنا داخل القوس لتدل على باش.

تعريف (6): رقم باش النسبي للأسعار هو:

$$\text{حيث: } W_n = \frac{P_n Q_n}{\sum P_n Q_n}$$

واستعملنا الرمز rP ليدل على باش النسبي، والأوزان تعتمد على النفقات (القيم المالية) لسنة المقارنة.

مثال (4): في مثال (3)

محاضرات في الإحصاء -1-

أ- احسب رقم باش القياسي التجميعي لأسعار عام 1999 باعتبار عام 1993 سنة أساس.

ب- احسب رقم باش القياسي النسبي لأسعار عام 1999 باعتبار عام 1993 سنة أساس.

الحل: نعيد كتابة الجدول ونضيف عليه الأعمدة المطلوبة في الحل وسنتبع هذا النمط في حل المسائل بشكل عام.

السلعة	السعر عام 1993	كمية الاستهلاك 1993	السعر عام 1999	كمية الاستهلاك 1999	$P_o Q_n$	$P_n Q_n$	W_n
السكر	220	7	350	8	1760	2800	
الأرز	280	10	430	12	3360	5160	
الشاي	1700	1.5	3000	1.5	2250	4500	
لحم الضأن	2800	5.5	4000	6.5	18200	26000	
					25870	38460	

(أ) رقم باش القياسي التجميعي للأسعار هو: $I_p(P) = \frac{38460}{25870} \times 100 = 148.67$

(ب) رقم باش القياسي النسبي للأسعار هو:

$$100 \times \left[\frac{350}{220} \right] > \\ =$$

تعريف (7): الرقم الأمثل ليفشر .

مثال (5): أوجد الرقم الأمثل لفشر من المثالين (3) و (4).

الحل: الرقم الأمثل لفشر يساوي: $I_p(F) = \sqrt{149.2 \times 148.67} = 148.93$

الفصل السابع: السلاسل الزمنية

1- مفهوم السلسلة الزمنية: هي عبارة عن معطيات إحصائية كمية مرتبطة بالزمن، حيث نعتمد على عنصر الزمن كمتغير مستقل لتفسير ظاهرة معينة (متغير تابع)، ونستعمل نماذج تحليل السلاسل الزمنية في عدة حالات منها:

1/ في حالة غياب علاقة ارتباط أو سببية بين المتغيرات.

2/ في حالة عدم توفر المعطيات الكافية حول المتغيرات المستقلة.

3/ في حالة ضعف النماذج الانحدارية إحصائياً واقتصادياً.

ومن أجل تفسير قيم المتغير التابع هناك طريقتين على الأقل: إما باستعمال عنصر الزمن كمتغير مستقل لتحديد وتفسير المتغير التابع $y_t = f(t, \varepsilon_t)$ أو باستعمال قيم المتغير التابع لفترات سابقة أي سلوك هذا

المتغير في الماضي $y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-k}, \varepsilon_t)$

2- مكونات أو مؤشرات السلسلة الزمنية: تتكون السلسلة الزمنية بصفة عامة من أربعة مركبات.

2-1- مركبة الاتجاه العام: وتكون هذه المركبة بشكل عام على شكل خط مستقيم $Y_t = a + bt$ إما بميل موجب أو ميل سالب.

2-2- المركبة الفصلية أو الموسمية: تعبر هذه المركبة عن التغيرات الموسمية أو الفصلية الناتجة عن تأثير عوامل خارجية وهي تتم غالباً بطريقة منتظمة كما أن تغير هذه المركبات يكون في المدى القصير. وهذه المركبة لها ثلاث أشكال: الشكل التجميعي والشكل المضاعف والشكل المختلط (التجميعي والمضاعف معاً).

➤ الشكل التجميعي: يعبر عن علاقة خطية ويكون العرض البياني للسلسلة الزمنية متشابهاً عبر

مختلف الفترات الزمنية أي هي عبارة عن مجموع قيم مركبات السلسلة الزمنية و تكتب على

الشكل التالي: $Y_t = X_t + S_t + \varepsilon_t$.

محاضرات في الإحصاء -1-

➤ الشكل المضاعف: يعبر عن علاقة غير خطية و يكون شكل السلسلة الزمنية متزايداً على شكل

$$Y_t = X_t \times S_t (1 + \varepsilon_t)$$

تزايد متتالية هندسية، وتكتب على الشكل التالي:

2-3/ المركبة الدورية: تظهر هذه المركبة نتيجة التغيرات التي تطرأ على الدورات الاقتصادية مثلاً من

ارتفاع وهبوط تتناسب مع مراحل الدورة الاقتصادية بمدة تتجاوز السنة أي في المدى المتوسط و الطويل

وتتكرر باستمرار عبر الزمن.

2-4/ المركبة العشوائية: تعبر هذه المركبة عن التغيرات التي يصعب التحكم فيها و ضبطها

وعى ناتجة عن عوامل غير منتظمة و لا علاقة لها بعنصر الزمن.

3- كشف وتحديد المركبة الفصلية أو الموسمية:

3-1/ كشف المركبة الفصلية:

3-1-1/ الطريقة البيانية: إن استعمال الطريقة البيانية لتحديد وكشف مركبات السلسلة الزمنية

يتطلب دقة كبيرة في عرض بيانات السلسلة المدروسة ، وذلك نظراً للصعوبة الكبيرة التي يتلقاها

الباحث في كشف مركباتها في كثير من الحالات، غير أنه بصفة عامة ، يصعب تحديد وكشف

مركبات السلسلة الزمنية عن طريق العرض البياني ما عدا المركبة الفصلية أو الدورية التي تظهر

جلياً بالعين المجردة.

3-1-2/ طريقة إختبار كرسكال - والس Kruskall- wallis: يعتبر هذا الإختبار من أهم

الإختبارات للكشف عن المركبة الفصلية و يرمز له بالرمز KW ويكتب على الشكل

التالي: $KW = (12/n(n+1)) \sum (R_i^2/m_i) - 3(n+1)$ حيث أن هذا المقدار يتبع توزيع كاي تربيع

علماً أن R_i مجموع رتب الفصل m_i عدد القيم أو المشاهدات المقابلة للفصل وتكون غالباً عدد

السنوات. P عدد المركبات أو الفصول. وأهم خطوات هذا الاختبار:

محاضرات في الإحصاء -1-

(* تقدير بطريقة المربعات $x_t = a + bt$)

(* حساب البواقي $w_t = y_t - (a + bt)$)

(* تحديد الرتب R_t للبواقي w_t)

(* نحسب KW أي القيمة المحسوبة.

(* نستخرج القيمة المجدولة لتوزيع كاي تربيع.

(* مقارنة القيمة المحسوبة مع القيمة المجدولة فإذا كانت المحسوبة أكبر من المجدولة فإن

السلسلة تحتوي على المركبة الفصلية والعكس صحيح.

جزء من جدول توزيع كاي تربيع عند مستوى معنوية 5%

$\alpha = 5\%$					
p	1	2	3	4	5
القيمة المجدولة	3.841	5.991	7.815	9.488	11.070

مثال 1: تمثل السلسلة التالية إستهلاك مادة الدقيق في ولاية الوادي والمطلوب منك: تأكد من وجود أو عدم وجود المركبة الفصلية.

	1	2	3	4
2005	14	20	44	21
2006	10	19	64	32
2007	12	12	68	29
2008	07	18	60	36
2009	06	11	64	50

2-3/ تحديد شكل المركبة الموسمية أو الفصلية:

1-2-3/ الطريقة البيانية: أنظر إلى 1-1-3

2-2/ الطريقة التحليلية: توجد عدة طرق لتحديد شكل السلسلة الزمنية بالطريقة التحليلية وهي 1/

طريقة معامل الخشونة 2/ طريقة الوسط السنوي 3/ طريقة الإنحراف المعياري 4/ طريق المعادلة

الإنحدارية.

محاضرات في الإحصاء -1-

3-2-1/طريقة معامل الخشونة: هو معامل يقيس درجة إنكسار السلسلة ويحسب بمعادلة

التالية:

وإذا كان C كبيراً دل ذلك على شدة إنكسار منحنى السلسلة إما إذا كان العكس دل ذلك على أن السلسلة ملساء.

3-2-2/ طريقة الوسط السنوي: تستعمل هذه الطريقة عندما تكون المعطيات غير سنوية أي أن السنة

مقسمة إلى سداسيات أو ثلاثيات أو ثنائيات أو أشهر أو، حيث نحسب الوسط الحسابي السنوي لكل سنة. ثم نحسب الفرق بين قيم أجزاء السنة، فإذا كانت هذه الفروق متقاربة فهو شكل تجميعي أما إذا كانت هذه الفروق تتضاعف من سنة إلى أخرى فهو شكل مضاعف.

3-2-3/ طريقة الانحراف المعياري: تعتمد هذه الطريقة على مرحلتين، أولاً تحديد الانحراف المعياري

لكل سنة، وثانياً إذا كانت الانحرافات المعيارية متساوية أو متقاربة فنقول أن السلسلة لها شكل تجميعي أما إذا كانت الانحرافات متباينة أو متباعدة فتكون السلسلة في حالة شكل مضاعف.

3-2-4/ طريقة المعادلة الإنحدارية: وهي أهم الطرق التي سنعتمدها في تحديد شكل السلسلة الزمنية،

وتعتمد هذه الطريقة على معامل إنحدار المعادلة التالية: $SD_t = a + by$ فإذا كان:

✓ B أقل من 0.05 نكون في حالة نموذج تجميعي.

✓ B أكبر من 0.1 نكون في حالة نموذج مضاعف.

✓ $0.05 \leq B \leq 0.1$ نكون في حالة نموذج مختلط.

مثال 2: تبين السلسلة الزمنية التالية تطور مبيعات سلعة معينة خلال 5 سنوات حيث السنة مقسمة إلى ثلاثيات.

محاضرات في الإحصاء -1-

	1	2	3	4
2005	20	28	22	34
2006	19	39	25	44
2007	21	49	33	55
2008	23	60	37	66
2009	24	71	42	76

المطلوب: إستعمل طرق تحديد شكل السلسلة الزمنية .

4- كشف وتحديد مركبة الاتجاه العام:

هناك عدة إختبارات لكشف مركبة الاتجاه العام وتسمى بالإختبارات الحرة : ومن بين هذه إختبار معامل

الإرتباط لرتب سبيرمان حيث نحسب معامل الرتب ثم نحسب إحصائية ستودنت $t_{cal} = r\sqrt{n-1}$ ثم

نقارنها بإحصائية ستودنت المجدولة t_{tab} عند مستوى معنوية 5 % غالباً

فإذا كانت $t_{cal} > t_{tab}$ فإن السلسلة الزمنية تحتوي مركبة اتجاه عام.

مثال 3: تأكد من وجود مركبة اتجاه عام بإستعمال إختبار معامل الإرتباط الرتبي عند مستوى معنوية 5 %

علماً $t_{tab} = 2.086$

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y _t	20	28	22	34	19	39	25	44	21	49	33	55	23	60	37	66	24	71	42	76

5/ طرق تقدير السلسلة الزمنية:

5-1/ طريقة المربعات الصغرى: عندما تكون السلسلة الزمنية لا تحتوي على مركبة موسمية .

مثال 3: تمثل السلسلة الزمنية التالية تطور الإستثمار في مؤسسة ما خلال 9 سنوات:

t	1	2	3	4	5	6	7	8	9
y _t	710	688	672	645	631	610	601	572	540

المطلوب منك:

1- قدر معلمات هذه السلسلة (النموذج).

2- أكتب الشكل النهائي للنموذج.

3- حدد قيمة الإستثمار للسنة العاشرة.

5-2/ عندما تكون السلسلة الزمنية تحتوي على مركبة موسمية: هناك الطريقة التجريبية.

5-2-1/ الطريقة التجريبية: أهم خطواتها :

أ/ تحديد المتوسطات المتحركة: عندما يكون عدد قيم فترة الوسط المتحرك عبارة عن عدد زوجي

نضيف القيمة التالية لفترة الوسط المتحرك ليصبح عدداً فردياً، مثلاً إذا كان $n=4$ فإن الوسط المتحرك

$$\text{هو: } Z_1 = (1/2Y_1 + Y_2 + Y_3 + Y_4 + 1/2Y_5)/4$$

أما إذا كان عدد قيم الوسط المتحرك فردي فالوسط المتحرك هو عبارة عن الوسط الحسابي لهذه القيم.

ب/ تحديد المعاملات الموسمية: في حالة نموذج مضاعف فهي عبارة عن النسبة بين القيم الحقيقية و

المتوسطات المتحركة المقابلة لها: $C_t = Y_t / Z_t$ ، أما في حالة نموذج تجميعي فهي عبارة عن الفرق بين

القيم الحقيقية و المتوسطات المتحركة $C_t = Y_t - Z_t$.

ج/ نحسب متوسط كل شهر أو نحسب الوسيط .

د/ تصحيح السلسلة الزمنية الأصلية: هي عبارة عن قسمة قيم السلسلة الحقيقية على متوسط المعاملات

الموسمية.

محاضرات في الإحصاء -1-

مثال 4: صحح السلسلة بالطريقة التجريبية إذا كان هناك ضرورة لتصحيح.

	1	2	3	4	5	6	7	8	9	10	11	12
2006	330	364	310	312	350	357	332	200	388	355	351	331
2007	346	349	364	345	320	423	333	224	473	413	400	419
2008	327	442	447	401	351	449	350	228	381	423	479	434
2009	406	388	465	405	445	591	409	272	423	539	515	483
2010	510	570	573	491	602	555	553	367	588	696	627	475

الفصل الثامن: توزيع بواسون

مقدمة

بجانب البساطة الرياضية التي تتمتع بها عملية بواسون كمثال بسيط على عملية ماركوف، فإنها تتمتع بخاصية الوجدانية والتي تجعلها نظرية مركزية. عملية بواسون تكون عبارة عن عملية متصلة والتي تصف عدد الحوادث العشوائية النقطية التي تحدث في فترة زمنية معينة. غالباً نشير إلى هذه الحوادث بالوصول، فعلى سبيل المثال عدد المسائل أو المهام التي تصل إلى معالج بيانات، أو عدد المكالمات الهاتفية التي تصل إلى تحويل هاتفية، عدد الجزيئات التي تصل إلى عداد جيجر، عدد الأخطاء المطبعية في صفحة من صفحات كتاب معين، عدد الزبائن الذين يصلون إلى سوق معين، عدد الأشخاص الذين ينتظرون في صف لتلقي خدمة معينة، ... الخ.

(2، 7) تعاريف

عملية بواسون تكون عبارة عن عملية عشوائية بفضاء معلمة متصل وفضاء حالة متقطع $S = \{0, 1, 2, \dots\}$. يوجد تعريفين لعملية بواسون، سنبرهن في نظرية (1، 7) أنهما متكافئين. ليكن $N_{t,\tau}$ عدد الوصول خلال الفترة $(t, \tau]$. يمكن أحياناً كتابة N_t بدلاً من $N_{0,t}$ عدد الوصول في الفترة $(0, t]$. بيانياً الراسم $N_t \rightarrow t$ يكون غير تناقصي، متزايد بالقفزات فقط، ومتصل من اليمين ويحقق أن $N_0 = 0$ (انظر الشكل (1، 7)). أحياناً تسمى العملية $\{N_t; t \geq 0\}$ بعملية العد counting process.

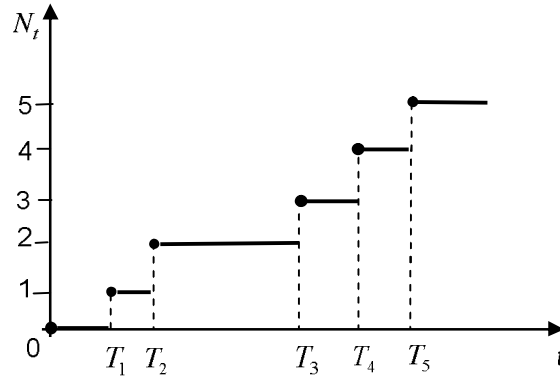
تعريف (1، 7):

تكون العملية $\{N_t; t \geq 0\}$ عملية بواسون إذا كان لأي $t, \tau \geq 0$ يتحقق ما يلي:

- (1) تكون العملية متجانسة مع الزمن time-homogeneous، أي أنه، لأي عدد صحيح غير سالب k ، فإن الاحتمال $P(N_{\tau, \tau+t} = k)$ يكون مستقل عن τ ، وهذا يكافئ أن $N_{0,t}$ ، $N_{\tau, \tau+t}$ لهما نفس التوزيع.
- (2) يكون للعملية زيادات مستقلة، بمعنى أن $N_{0,t}$ ، $N_{\tau, \tau+t}$ يكونان مستقلين.

محاضرات في الإحصاء -1-

(3) تكون عملية منظمة (مرتبة) orderly، بمعنى أن $\lim_{t \rightarrow 0} \frac{P(N_{\tau, \tau+t} \geq 2)}{t} = 0$



شكل (7، 1): الدالة N_t والتي لأجلها أزمدة الوصول هي $T_1 = 0.4$ ، $T_2 = 0.6$ ،

$$T_3 = 1.8, T_4 = 2.1, T_5 = 2.4, \dots$$

يمكن تعريف عملية بواسون بطريقة مكافئة كما في التعريف التالي.

تعريف (7، 2):

تكون العملية $\{N_t; t \geq 0\}$ عملية بواسون إذا وجد عدد صحيح موجب λ بحيث لأي

$t, h \geq 0$ يتحقق ما يلي:

$$(أ) P(N_{t,t+h} = 0) = 1 - \lambda h + o(h)$$

$$(ب) P(N_{t,t+h} = 1) = \lambda h + o(h)$$

$$(ج) P(N_{t,t+h} \geq 2) = o(h)$$

$$(د) N_{0,t}, N_{t,t+h} \text{ يكونان مستقلين لجميع قيم } t, h > 0.$$

يسمى العدد λ بمعلمة أو بمعدل عملية بواسون.

نظرية (7، 1):

التعريفان (7، 1)، (7، 2) متكافئان.

البرهان:

أولاً (2) هو نفس (د) و (3) هو نفس (ج)، علاوة على ذلك يمكن الحصول على (ب) من (أ) و (ج) وذلك لأن $P(N_{t,t+h} \geq 0) = 1$. الآن وحيث أن الأطراف اليمنى للعلاقات (أ)، (ب)، (ج) لا تحتوي على t ، فإن (1) تكون متحققة (سنقوم بالتحقيق من ذلك

محاضرات في الإحصاء -1-

بالتفصيل لاحقاً). يتبقى أن نبرهن على أنه يمكن استنتاج (أ) من النظرية (7، 1). بأخذ $p_k(t) = P(N_{0,t} = k) = P(N_{\tau, \tau+t} = k)$ لأي $\tau > 0$ وذلك باستخدام (1). وبالتالي يكون لدينا لأي $t, s > 0$

$$\begin{aligned} p_0(t+s) &= P(N_{0,s+t} = 0) \\ &= P(N_{0,s} = 0, N_{s,s+t} = 0) \quad \text{by (2)} \\ &= P(N_{0,s} = 0)P(N_{s,s+t} = 0) \quad \text{by (1)} \\ &= p_0(s)p_0(t) \end{aligned}$$

بالمثل يمكن التحقق باستخدام الاستنتاج الرياضي من أنه لأي $n \geq 1$ فإن

$$p_0(t_1 + t_2 + \dots + t_n) = p_0(t_1)p_0(t_2)\dots p_0(t_n)$$

وباستخدام $\sum_{i=1}^n (1/n) = 1$ ، ينتج أن

$$p_0(1) = [p_0(\frac{1}{n})]^n$$

بكتابة $p_0(1) = e^{-\lambda}$ ، حيث أن $0 < \lambda < \infty$ ، فإن

$$p_0(\frac{1}{n}) = e^{-\lambda/n}$$

وبنفس الطريقة نحصل على:

$$p_0(\frac{k}{n}) = e^{-k\lambda/n}$$

علاوة على ذلك فإن

$$p_0(s+t) \leq p_0(t)$$

وحيث أن $p_0(s+t) = p_0(t)p_0(s)$ إذن $p_0(t)$ تتناقص مع زيادة t . والآن ولأي $t \geq 0$ يوجد عدد صحيح $k > 0$ يحقق أن $(k-1)/n \leq t < k/n$. وبالتالي يكون لدينا

$$e^{-(k-1)\lambda/n} = p_0(\frac{k-1}{n}) \geq p_0(t) > p_0(\frac{k}{n}) = e^{-k\lambda/n}$$

ومن ثم فإنه عندما $t \rightarrow \infty$ يكون $p_0(t) \rightarrow e^{-\lambda t} = 1 - \lambda t + o(t)$ والتي تبرهن (أ).

ملاحظة (7، 1):

1. تعتبر عملية بواسون مثال لعملية عشوائية بفضاء معلمة متصل (عملية متصلة الزمن) وزيادات مستقلة ومستقرة. من التعريف (7، 1) الشرط (1) يعطي خاصية الاستقرار أما الشرط (2) فيعطي خاصية الزيادات المستقلة.

2. أحد تفسيرات خاصية الزيادات المستقلة هو أن الحوادث تحدث عشوائياً. يشار إلى

محاضرات في الإحصاء -1-

الحوادث التي تحدث تبعا لعملية بواسون بالحوادث التي تحدث عشوائيا. وبالتالي عندما نذكر أن الحوادث تحدث عشوائيا، بدون ذكر اسم العملية التي تحدث الحوادث تبعا لها، فهذا يعني أنها تحدث تبعا لعملية بواسون.

نظرية (2، 7):

المتغير العشوائي N_t يتبع توزيع بواسون.

البرهان:

من أفضل الطرق التي يمكن اتباعها لبرهان هذه النظرية هي استخدام الدوال المولدة للاحتمال كما سنوضح الآن. أولا نعتبر N_h لأي متغير h يأخذ قيما صغيرة. بالتالي وبناء على (أ)، (ب)، (ج) من تعريف (2، 7)، يمكن الحصول على ما يلي:

$$\begin{aligned} E[z^{N_h}] &= (1 - \lambda h)z^0 + \lambda h z^1 + o(h)z^2 \\ &= 1 - (1 - z)\lambda h + o(h)z^2 \\ &= e^{-\lambda(1-z)h + o(h)} \end{aligned}$$

و لحساب $E[z^{N_t}]$ يمكن تقسيم الفترة $[0, t]$ إلى n فترة جزئية طول كل واحدة منها h ، ومن ثم فإن

$$\begin{aligned} E[z^{N_t}] &= E[z^{N_{0,h} + N_{h,2h} + \dots + N_{t-h,h}}] \\ &= (E[z^{N_{0,h}}])^n \end{aligned}$$

وحيث أن $n = t/h$ ، إذن

$$\begin{aligned} E[z^{N_t}] &= (E[z^{N_h}])^{t/h} \\ &= [e^{-\lambda(1-z)h + o(h)}]^{t/h} \\ &= e^{-\lambda(1-z)t + t o(h)/h} \\ &\rightarrow e^{-\lambda(1-z)t} \text{ as } h \rightarrow 0 \end{aligned}$$

ملاحظة (2، 7):

1. يمكن برهان النظرية السابقة مباشرة باستخدام الاستنتاج الرياضي كما قدمنا في الجزء

(2، 2، 1).

2. أحد نتائج النظرية (2، 7) هي أن المتغير العشوائي الذي يتبع توزيع بواسون هو

محاضرات في الإحصاء -1-

المتغير الوحيد الذي يحقق الشروط المقدمة في التعريفين (1، 7)، (2، 7).

3. أيضا النظرية (2، 7) تعني أن عدد المرات التي تحدث عندها الحوادث في فترة

زمنية معينة بطول t يتبع توزيع بواسون بالمعلمة λt .

4. حتى الآن نلاحظ أن λ هو عبارة عن ثابت لم نقدم له أي معنى محدد. الآن نلاحظ

أن لأي $t \geq 0$

$$(1، 7) \quad E[N_t] = \lambda t$$

ومن ثم فإن λ هو عبارة عن عدد الوصول المتوقع في فترة طولها الوحدة، وبمعنى آخر هو

عبارة عن معدل الوصول. يمكن أيضا التحقق من أن

$$(2، 7) \quad \text{Var}[N_t] = \lambda t$$

مثال (1، 7):

ليكن $\{N_t; t \geq 0\}$ عملية بواسون بمعدل $\lambda = 8$ ، ونريد حساب

$$P(N_{2.5} = 17, N_{3.7} = 22, N_{4.3} = 36)$$

لدينا

$$\begin{aligned} \{N_{2.5} = 17, N_{3.7} = 22, N_{4.3} = 36\} = \\ \{N_{2.5} = 17, N_{3.7} - N_{2.5} = 5, N_{4.3} - N_{3.7} = 14\} \end{aligned}$$

من خاصية الزيادات المستقلة فإن المتغيرات $N_{2.5}$ ، $N_{3.7} - N_{2.5}$ ، $N_{4.3} - N_{3.7}$ تكون مستقلة، ومن

خاصية الاستقرار فإنها تتبع توزيع بواسون بالمعالم التالية $8 \times 2.5 = 20$ ، $8 \times (3.7 - 2.5) = 9.6$ ،

$8 \times (4.3 - 3.7) = 4.8$ على الترتيب. وبالتالي فإن الاحتمال المطلوب هو:

$$\frac{e^{-20} 20^{17}}{17!} \cdot \frac{e^{-9.6} 9.6^5}{5!} \cdot \frac{e^{-4.8} 4.8^{14}}{14!}$$

مثال (2، 7):

بفرض أن مكالمات هاتفية تصل عشوائيا إلى تحويلة هاتف بمعدل 15 مكالمات في

الساعة.

(أ) أوجد احتمال أن تصل على الأقل مكالمتين بين الساعة 8 والساعة 8:12 صباحا.

(ب) إذا تغيب العامل لمدة 10 دقائق، أوجد احتمال عدم فقد أي مكالمات.

الحل:

محاضرات في الإحصاء -1-

(أ) ليكن X هو عدد المكالمات التي تصل إلى التحويلة بين الساعة 8 والساعة 8:12

صباحاً. إذن X يتبع توزيع بواسون بالمتوسط $\lambda = \frac{12 \times 15}{60} = 3$ ، ومن ثم فإن:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - e^{-3} - 3e^{-3} \\ &= 1 - 4e^{-3} \end{aligned}$$

(ب) ليكن Y هو عدد المكالمات التي تصل إلى التحويلة في 10 دقائق. إذن Y يتبع توزيع

بواسون بالمتوسط $\lambda = \frac{10 \times 15}{60} = 2.5$ ، ومن ثم فإن احتمال عدم وصول مكالمات في عشرة

دقائق هو:

$$P(Y=0) = \frac{(2.5)^0 e^{-2.5}}{0!} = e^{-2.5}$$

مثال (7، 3):

تحدث أعطاب في توصيلة سلكية تحت الأرض بناء على عملية بواسون بالمعدل

$\lambda = 0.1$ لكل ميل.

(أ) ما هو احتمال عدم وجود أعطاب في الميلى الأولين؟

(ب) بشرط أنه لا يوجد أعطاب في أول ميلىين، فما هو احتمال أن لا يكون هناك أعطاب

بين الميل الثاني والثالث؟

الحل:

بفرض أن $N(t)$ يرمز إلى عدد الأعطاب التي تحدث حتى الميل t . إذن:

(أ) المتغير العشوائى $N(2)$ يتبع توزيع بواسون بالمعلمة $\lambda(2) = 0.2$. ومن ثم فإن:

$$P(N(2)=0) = e^{-0.2} = 0.8187$$

(ب) المتغير العشوائى $N(3) - N(2)$ مستقل عن $N(2) - N(0)$ ومن ثم فإن الاحتمال الشرطى

يكون نفس الاحتمال غير الشرطى وبالتالي:

$$P(N(3) - N(2) = 0) = e^{-0.1} = 0.9048$$

مثال (7، 4):

يصل زبائن إلى مركز خدمة ما بناء على عملية بواسون بمعدل $\lambda = 4$ لكل ساعة.

بفرض أن مركز الخدمة يفتح أبوابه من الساعة التاسعة صباحاً، ما هو احتمال وصول زبون

محاضرات في الإحصاء -1-

واحد فقط حتى اللحظة 9:30 وأن عدد الزبائن الكلي حتى الساعة 11:30 هو خمسة؟
الحل:

$$\begin{aligned} \text{حيث أن } N\left(\frac{1}{2}\right) \text{ و } N\left(\frac{5}{2}\right) - N\left(\frac{1}{2}\right) \text{ مستقلين، إذن} \\ P\left\{N\left(\frac{1}{2}\right)=1, N\left(\frac{5}{2}\right)=5\right\} &= P\left\{N\left(\frac{1}{2}\right)=1, N\left(\frac{5}{2}\right)-N\left(\frac{1}{2}\right)=4\right\} \\ &= \frac{e^{-4(1/2)} 4(1/2) e^{-4(2)} [4(2)]^4}{4!} \\ &= 2e^{-2} \left(\frac{512}{3} e^{-8}\right) \\ &= 0.0154965 \end{aligned}$$

مثال (7، 5):

بفرض أن زلزالاً يحدث في اليابان بناءً على عملية بواسون بمعدل $\lambda = 2$ في الأسبوع.
(أ) أوجد احتمال حدوث على الأقل ثلاث زلازل في أسبوعين متتاليين؟
(ب) أوجد التوزيع الاحتمالي لوقت انتظار حدوث زلزال ابتداءً من الآن؟
الحل:

بفرض أن $N(t)$ يرمز إلى عدد الزلازل التي تحدث حتى اللحظة t .
(أ) من المعلوم أن $N(2)$ يتبع توزيع بواسون بالمعلمة $\lambda \times t = (2)(2) = 4$. ومن ثم فإن:

$$\begin{aligned} P(N(2) \geq 3) &= 1 - P(N(2) < 3) \\ &= 1 - P(N(2) = 0) - P(N(2) = 1) - P(N(2) = 2) \\ &= 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2} e^{-4} \\ &= 1 - 13e^{-4} \\ &= 0.762 \end{aligned}$$

(ب) ليكن X وقت (وحدة القياس هي الأسبوع) انتظار وقوع الزلزال القادم. حيث أن الحادثة $\{X > t\}$ تحدث إذا وفقط إذا كان لم يحدث أي زلزال خلال فترة طولها t ، أي أن الحادثة $\{X > t\}$ تكافئ الحادثة $\{N(t) = 0\}$ ، وبالتالي فإن

$$P(X > t) = P(N(t) = 0) = e^{-\lambda t}, t \geq 0$$

ومن ثم فإن دالة التوزيع الاحتمالي للمتغير X هي

$$\begin{aligned} P(X \leq t) &= 1 - P(X > t) \\ &= 1 - e^{-2t}. \end{aligned}$$

(3، 7) الأزمنة بين الوصول

Times between arrivals

والآن سندرس الزمن بين وصولين متتالين لمتغير يتبع توزيع بواسون. يمكن التحقق من أن فترة الانتظار بين وصول ما والوصول رقم n بعده يكون عبارة عن متغير عشوائي يتبع توزيع إيرلانج n -(Erlang- n). ولأي $n=1$ نحصل على أن زمن الوصول الداخلي (زمن الانتظار بين وصولين متتالين) interarrival time يتبع التوزيع الأسّي. تبرهن النظرية التالية على صحة هذه النتيجة.

نظرية (3، 7):

لعملية بواسون بالمعلمة λ ، فإن كل من زمن أول وصول وزمن الانتظار بين وصولين متتالين (فترة الانتظار بين وصولين متتالين) يكون عبارة عن متغير عشوائي يتبع التوزيع الأسّي بالمعلمة λ .

البرهان:

ليكن X يرمز إلى زمن أول وصول (زمن الانتظار لحدوث أول وصول). إذن

$$P(X > t) = P(N_t = 0) = e^{-\lambda t},$$

وبالتالي فإن التوزيع الاحتمال لزمن أول وصول هو $P(X \leq t) = 1 - e^{-\lambda t}$ ومن ثم فإن زمن الوصول الأول يتبع توزيع أسّي بالمعلمة λ .

ليكن T يرمز إلى أزمنة الوصول الداخلية (زمن الانتظار بين وصولين متتالين). لحساب توزيع T نريد إثبات أنه بشرط معرفة حدوث وصول عند اللحظة τ ، فإن عدد الوصول في الفترة $(\tau, \tau + t]$ يساوي صفر. حيث أن $N_{s,t}$ يرمز إلى عدد الوصول في الفترة $(s, t]$ ، وحيث أننا فرضنا أنه يوجد وصول عند اللحظة τ ، إذن باستخدام الخاصية (2) من

تعريف (1، 7) والجزء الأول من هذه النظرية يمكن التحقق من الآتي:

$$\begin{aligned} P(T > t) &= \lim_{h \rightarrow 0} P(N_{\tau+h, \tau+h+t} = 0 \mid N_{\tau, \tau+h} = 1) \\ &= \lim_{h \rightarrow 0} P(N_{\tau+h, \tau+h+t} = 0) \\ &= P(N_t = 0) \\ &= e^{-\lambda t} \end{aligned}$$

ومن ثم فإن توزيع زمن الإنتظار بين وصولين متتالين يتبع التوزيع الأسّي بالمعلمة λ ،

وهذا يكمل البرهان.

مثال (7، 6):

بفرض أن زلزالاً يحدث عشوائياً في منطقة ما بناءً على عملية بواسون بمعدل $\lambda = 2$ في العام. إذا حدث زلزال في الوقت الراهن فما هو احتمال الانتظار على الأقل ثلاث أعوام قبل حدوث الزلزال المقبل (القادم).

الحل:

توجد طريقتان لحل هذا المثال وذلك إما باستخدام توزيع بواسون أو باستخدام التوزيع الأسّي. فيما يلي سنستخدم الطريقتين في الحل.

الطريقة الأولى: بفرض أن الزلزال القادم سيحدث بعد T عام، بالتالي فإن T يتبع توزيعاً أسياً بالمتوسط $\lambda = 2$ ومن ثم فإن احتمال الانتظار على الأقل ثلاث أعوام قبل حدوث الزلزال التالي هو

$$P(T \geq 3) = \int_3^{\infty} 2e^{-2t} dt = e^{-6}$$

الطريقة الثانية: ليكن X يرمز إلى عدد الزلازل التي تحدث في فترة زمنية طولها ثلاثة أعوام بعد آخر زلزال. وبالتالي فإن X يتبع توزيع بواسون بالمتوسط $2 \times 3 = 6$ ، ومن ثم فإن احتمال الانتظار على الأقل ثلاث أعوام قبل حدوث الزلزال التالي هو

$$P(X = 0) = \frac{6^0 e^{-6}}{0!} = e^{-6}.$$

تعطي النظرية التالية نتيجة أقوى من النتيجة المقدمة في النظرية (7، 3).

نظرية (7، 4):

ليكن $T_1, T_2, \dots$ عبارة عن الأزمنة المتتالية لوصول متتالي لعملية وصول $\{N_t\}$. إذن $\{N_t\}$ تكون عبارة عن عملية بواسون بالمعلمة λ إذا كان فقط إذا كان المتغيرات العشوائية $T_1, T_2 - T_1, T_3 - T_2, \dots$ مستقلة ومتطابقة ولها توزيع أسّي بالمعلمة λ .

البرهان:

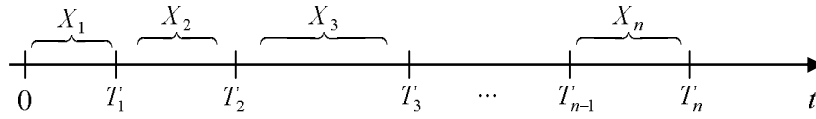
يترك كتمرين.

مثال (7، 7):

محاضرات في الإحصاء -1-

بفرض أن زمن حياة وحدة ما يتبع توزيع أسي بالمعلمة $\lambda=2$ ، عندما تتعطل هذه الوحدة فيتم استبدالها مباشرة بوحدة أخرى جديدة مطابقة للوحدة الأصلية، وعندما تتعطل الوحدة الجديدة يتم استبدالها بأخرى متطابقة، ... إلخ. وهذا يعني أن أزمنة الحياة للوحدة المتعاقبة في الاستخدام $X_1, X_2, \dots$ تكون عبارة عن متغيرات عشوائية مستقلة ومتطابقة ولها نفس التوزيع الأسي:

$$P(X_n \leq t) = 1 - e^{-\lambda t}, t \geq 0, n = 1, 2, \dots$$



شكل (7، 2): أزمنة الاخفاقات المتتالية.

إذا كان $T_1, T_2, \dots$ عبارة عن الأزمنة المتعاقبة للإخفاقات المتتالية، إذن $T_1 = X_1$ ، $T_2 = X_1 + X_2$ ، ، أنظر الشكل (7، 2). من السهل التحقق من أن $T_1, T_2, \dots$ تحقق شروط النظرية (7، 4). وبالتالي فإنه إذا كان N_t يرمز إلى عدد الاخفاقات في الفترة $[0, t]$ ، إذن يمكن اتباع نظرية (7، 4) في التحقق من أن $\{N_t; t \geq 0\}$ تكون عملية بواسون بالمعدل λ .

ملاحظة (7، 3):

باتباع خاصية أن $X_{n+1} = T_{n+1} - T_n$ يتبع توزيع أسي بالمعلمة λ ، يمكن الحصول على متوسط وتباين الزمن بين أي وصولين متتالين كآتي:

$$E[T_{n+1} - T_n] = E[X_{n+1}] = \frac{1}{\lambda},$$

$$\text{Var}[T_{n+1} - T_n] = \text{Var}[X_{n+1}] = \frac{1}{\lambda^2}$$

وبكتابة $T_n = T_1 + (T_2 - T_1) + \dots + (T_n - T_{n-1})$ يمكن الحصول على:

$$E[T_n] = E[X_1 + X_1 + \dots + X_n]$$

$$= nE[X_1]$$

$$= \frac{n}{\lambda},$$

$$\text{Var}[T_n] = \text{Var}[X_1 + X_1 + \dots + X_n]$$

محاضرات في الإحصاء -1-

وحيث أن $X_1, X_2, \dots$ مستقلة ومتطابقة، إذن

$$\begin{aligned}\text{Var}[T_n] &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] \\ &= n\text{Var}[X_1] = \frac{n}{\lambda^2} \\ &= \frac{n}{\lambda^2}\end{aligned}$$

مثال (7، 8):

بالعودة إلى مثال (7، 7) وبفرض أن $\lambda = 0.0002$ (مقياس وحدة الزمن هو ساعة). إذن متوسط زمن حياة الوحدة في الاستخدام هو: ساعة $E[X_1] = \frac{1}{\lambda} = 5000$ والتباين هو: ساعة $^2 \text{Var}[X_1] = \frac{1}{\lambda^2} = 25 \times 10^6$. ومن ثم فإنه إذا كانت معدة مكون من ثلاثة أجزاء متصلة بحيث أنه بمجرد أن يتعطل الجزء الأول يبدأ الثاني في العمل، ويبدأ الثالث في العمل بمجرد أن يتعطل الثاني، وبالتالي فإن زمن المعدة هو $T_3 = X_1 + X_2 + X_3$ ، وبالتالي فإن متوسط وتباين T_3 هما $E[T_3] = \frac{3}{\lambda} = 15000$ ساعة، $\text{Var}[T_3] = \frac{3}{\lambda^2} = 75 \times 10^6$ ساعة².

مثال (7، 9):

بالعودة إلى مثال (7، 7) وبفرض أنه سيتم استبدال الوحد بمجرد أن تتعطل، ونفرض أيضا أن تكلفة عملية الاستبدال هي β ريال، وأن معدل الخصم على المال (discount rate of money) هو $\alpha > 0$ ، وبالتالي فإن القيمة الشرائية لواحد ريال بعد مرور فترة زمنية طولها t تساوي القيمة الشرائية في الوقت الراهن للمبلغ $e^{-\alpha t}$ ريال (وبمعنى آخر فإن α تكون عبارة عن معدل الفائدة). وبالتالي فإن زمن الاخفاق رقم n هو T_n والقيمة الحالية لتكلفة الاستبدال هي $\beta e^{-\alpha T_n}$. ولحساب القيمة الحالية لتكلفة جميع عمليات الاستبدال في المستقبل نقوم بتجميع جميع القيم الحالية لتكلفة الاستبدال رقم n على جميع القيم الممكنة لـ n ، ويقودنا ذلك إلى النتيجة التالية:

$$C = \sum_{n=1}^{\infty} \beta e^{-\alpha T_n}$$

والمطلوب هو حساب القيمة المتوقعة الحالية لتكلفة الاستبدال. حيث أن القيمة المتوقعة للمجموع تكون عبارة عن مجموع التوقعات، إذن:

$$E[C] = \beta \sum_{n=1}^{\infty} E[e^{-\alpha T_n}]$$

محاضرات في الإحصاء -1-

لأي قيمة محددة لـ n ، يمكن كتابة $T_n = T_1 + (T_2 - T_1) + \dots + (T_n - T_{n-1})$ ، حيث أن $T_1, (T_2 - T_1), \dots, (T_n - T_{n-1})$ تكون مستقلة ومتطابقة، إذن

$$\begin{aligned} E[e^{-\alpha T_n}] &= E[e^{-\alpha T_1} e^{-\alpha(T_2 - T_1)} \dots e^{-\alpha(T_n - T_{n-1})}] \\ &= E[e^{-\alpha T_1}] E[e^{-\alpha(T_2 - T_1)}] \dots E[e^{-\alpha(T_n - T_{n-1})}] \\ &= \{E[e^{-\alpha T_1}]\}^n \end{aligned}$$

وحيث أن T_1 يتبع توزيع أسي بالمعلمة λ ، إذن

$$\begin{aligned} E[e^{-\alpha T_1}] &= \int_0^{\infty} e^{-\alpha x} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\alpha + \lambda)x} dx \\ &= \frac{\lambda}{\alpha + \lambda} \end{aligned}$$

وبالتالي وباستخدام العلاقة $1 + x + x^2 + \dots = \frac{1}{1-x}$ لأي $x \in [0, 1)$ ، يمكن الحصول على

$$E[C] = \beta \sum_{n=1}^{\infty} \left(\frac{\lambda}{\alpha + \lambda} \right)^n = \beta \frac{\lambda}{\alpha + \lambda} \left(1 - \frac{\lambda}{\alpha + \lambda} \right)^{-1} = \frac{\beta \lambda}{\alpha}$$

وكحالة خاصة نفرض أن زمن الحياة هو 5000 ساعة وأن تكلفة الإحلال هي 800

ريال ومعدل الفائدة يساوي 24% في العام، إذن $\beta = 800$ ، $\lambda = \frac{1}{5000}$ ، $\alpha = \frac{0.24}{365 \times 24} = \frac{0.01}{365}$ ، وبالتالي

$$E[C] = \frac{800 \times 36500}{5000} = 5840 \text{ ريال.}$$

وبالرغم من أنه يمكن عمل العديد من الحسابات بدون استنتاج توزيع المتغير العشوائي إلا أنه يوجد كثير من المواضيع التي نحتاج فيها إلى هذا التوزيع. إحدى الطرق التي يمكن اتباعها في الحصول على التوزيع هي طريقة معكوس تحويل لابلاس Laplace transform للمتغير T_n ، $E[e^{-sT_n}] = \left(\frac{\lambda}{s + \lambda} \right)^n$ لأي $s > 0$. وعموما فإنه من السهل أن نستخدم حقيقة أنه لأي $n \geq 0$ ، $t \geq 0$ فإن

$$(3, 7) \quad \{T_n \leq t\} = \{N_t \geq n\}$$

باستخدام علاقة المساواة (3، 7) مع نظرية (2، 7) يمكن برهان النظرية التالية.

نظرية (5، 7):

لأي $n \geq 0$ ، فإن توزيع المتغير العشوائي T_n يعطى بالعلاقة التالية:

محاضرات في الإحصاء -1-

$$P\{T_n \leq t\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}, t \geq 0.$$

البرهان:

حيث أن

$$\begin{aligned} P\{T_n \leq t\} &= P\{N_t \geq n\} \\ &= 1 - P\{N_t < n\} \\ &= 1 - \sum_{k=0}^{n-1} P\{N_t = k\} \end{aligned}$$

وحيث أن N_t يتبع توزيع بواسون بالمعلمة λt ، إذن بالتعويض عن توزيع بواسون في العلاقة السابقة نحصل على توزيع T_n ، وبه نكمل برهان النظرية.

يعرف توزيع المتغير العشوائي T_n ، المعطى في النظرية السابقة، بتوزيع إيرلانج- n . وينتمي هذا التوزيع إلى عائلة توزيعات جاما، وهو قابل للإشتقاق ومشتقته هي:

$$(4, 7) \quad \frac{\lambda (\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!}, t \geq 0$$

مثال (7، 10):

تصل مكالمة هاتفية إلى تحويلة هاتف بناءً على عملية بواسون بمعدل مكالمتين في الدقيقة. إذا تم استقبال مكالمات للتو، فما هو احتمال استقبال الثلاث مكالمات التالية بعد دقيقتين؟

الحل:

بفرض أن الثلاث مكالمات التالية ستصل بعد T_3 دقيقة. إذن المطلوب هو حساب

$$P(T_3 > 2) \text{ لدينا:}$$

$$P(T_3 > 2) = \{\text{عدد المكالمات في } (0, 2) \text{ أقل من أو يساوي } 2\}$$

$$n = \sum_{n=0}^2 P\{\text{عدد المكالمات في } (0, 2) \text{ يساوي } n\}$$

$$= \sum_{n=0}^2 \frac{(2 \times 2)^n e^{-(2 \times 2)}}{n!}$$

$$= 13e^{-4}$$

وبطريقة أخرى، يمكن الحل كالآتي:

محاضرات في الإحصاء -1-

$$\begin{aligned}
 P(T_3 > 2) &= 1 - P(T_3 \leq 2) \\
 &= 1 - P(N_2 \geq 3) \\
 &= P(N_2 < 3) \\
 n &= \sum_{n=0}^2 P\{ \text{عدد المكالمات في } (0,2) \text{ يساوي } n \} \\
 &= 13e^{-4}
 \end{aligned}$$

كما نلاحظ أنه يمكن الحصول على نفس الحل مباشرة باستخدام النظرية (7، 5).

مثال (7، 11):

بأخذ نقطة معينة على طريق سريع، وبفرض أن $U_1, U_2, \dots$ عبارة عن أوقات الانتظار الداخلية (أزمنة الوصول بين كل سيارتين متعاقبتين) للوصول المتعاقب للمركبات عند هذه النقطة. بفرض أن $U_1, U_2, \dots$ تكون مستقلة ومتطابقة بالتوزيع التالي:

$$P(U_k \leq t) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}, t \geq 0$$

والمطلوب هو تحديد التوزيع الاحتمالي للمتغير العشوائي الذي يمثل عدد المركبات التي تعبر هذه النقطة في الفترة $[0, t]$.

الحل:

بفرض أن M_t يرمز إلى عدد المركبات التي تعبر النقطة في الفترة $[0, t]$ ، إذن المطلوب هو استنتاج توزيع المتغير العشوائي M_t . نلاحظ أولاً أن المتغير العشوائي U_k يتبع توزيع إيرلانج-2. وبالتالي فإنه يمكن تفسير كل من U_k على أنه مجموع زمني وصولين متعاقبين لعملية بواسون بالمعلمة λ . وهذا يعني أنه يمكن اعتبار أزمنة عبور المركبات النقطة المحددة على أنه $U_1 = T_2$ ، $U_1 + U_2 = T_4$ ، $U_1 + U_2 + U_3 = T_6$ ، $\dots$ ، حيث أن $T_1, T_2, \dots$ تكون عبارة عن أزمنة الوصول في عملية بواسون $\{N_t\}$ بالمعلمة λ . وبالتالي فإن $M_t = 6$ إذا وفقط إذا كان عملية بواسون N_t لها 12 وصول أو 13 وصول في الفترة $[0, t]$. وبالتالي لأي $t \geq 0$ فإن:

$$\begin{aligned}
 P(M_t = k) &= P(N_t = 2k) + P(N_t = 2k + 1) \\
 &= \frac{(\lambda t)^{2k} e^{-\lambda t}}{(2k)!} + \frac{(\lambda t)^{2k+1} e^{-\lambda t}}{(2k+1)!}, \quad k = 0, 1, \dots
 \end{aligned}$$

ملاحظة (7، 4):

محاضرات في الإحصاء -1-

1. نلاحظ أن دالة الكثافة الاحتمالية تكون مطردة الزيادة. وبالتالي فإن احتمال وقوع أزمنة الوصول الداخلية في الفترة (تأخذ قيمة في الفترة) $[0, s]$ يكون أكبر من وقوعها في الفترة $[t, t+s]$ وذلك لأي $t > 0$. ومن ثم فإن عملية بواسون يكون لها عدد من الفترات القصيرة أكبر من عدد الفترات الطويلة. أي أن احتمال وقوع طول فترة زمن الوصول الداخلي في الفترة $[0, s]$ يكون أكبر من احتمال وقوعه في الفترة $[t, t+s]$ ، وذلك لأي $t > 0$.

2. التوزيع الأسّي السالب يكون توزيع مرافق لتوزيع بواسون في عملية بواسون حسب المفهوم التالي. أزمنة الوصول الداخلية في عملية نقطة point process تضمن أن عدد الوصول في فترة معينة يكون متغير عشوائي يتبع توزيع بواسون وأن عملية الوصول تكون عملية بواسون. وبالعكس، بفرض أن عدد الوصول في أي فترة زمنية يكون متغير عشوائي له توزيع بواسون، إذن أزمنة الوصول الداخلية تكون متغيرات عشوائية تتبع التوزيع الأسّي، وعملية الوصول تكون عملية بواسون. يستخدم التوزيع الأسّي بكثرة في النمذجة العشوائية وذلك بسبب بساطته التي تقود إلى تسهيل عملية الحسابات بالإضافة إلى أنه يتمتع بخاصية فقدان الذاكرة، وهذا هو الهدف الرئيس في البند التالي.

(7، 4) خاصية فقدان الذاكرة

Memoryless property

يقال للمتغير العشوائي X أنه يحقق خاصية فقدان الذاكرة memoryless property أو التوزيع الاحتمال للمتغير العشوائي X يكون فاقطاً للذاكرة إذا وفقط إذا كان

$$P(X > x+t | X > t) = P(X > x)$$

وهذا يعني أن زمن الوصول التالي يكون مستقل عن أزمنة الوصول التي تسبق زمن الوصول الراهن. التوزيع الأسّي هو التوزيع الوحيد الذي يحقق هذه الخاصية كما توضحه النظرية التالية.

نظرية (7، 6):

المتغير العشوائي المتصل يحقق خاصية فقدان الذاكرة إذا وفقط إذا كان يتبع توزيع

أسيًا .

البرهان:

أولا نفرض أن X متغير عشوائي يتبع توزيعاً أسياً بالمعلمة $\lambda > 0$ ، أي أن $P(X \leq x) = 1 - e^{-\lambda x}$ وبالتالي فإن

$$\begin{aligned} P(X > t+x | X > t) &= \frac{P(X > t+x, X > t)}{P(X > t)} \\ &= \frac{P(X > t+x)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} \\ &= e^{-\lambda x} = P(X > x) \end{aligned}$$

إذن المتغير X يحقق خاصية فقدان الذاكرة.

والآن نفرض أن X يحقق خاصية فقدان الذاكرة، والمطلوب برهان أنه يتبع توزيعاً أسياً .
نفرض أن $F(x)$ هو التوزيع الاحتمالي للمتغير X ، وأن $H(x) = 1 - F(x)$ ، إذن لدينا:

$$H(x) = P(X > x)$$

باستخدام خاصية فقدان الذاكرة نحصل على:

$$\begin{aligned} H(x) &= \frac{P(X > t+x | X > t)}{P(X > t)} \\ &= \frac{P(X > t+x, X > t)}{P(X > t)} \\ &= \frac{P(X > t+x)}{P(X > t)} \\ &= \frac{H(x+t)}{H(t)} \end{aligned}$$

ومن ثم فإن $H(x)H(t) = H(x+t)$. وحيث أن $0 \leq H(x) \leq 1$ ، إذن باتباع الطريقة المستخدمة في برهان النظرية (7، 1) يمكن التحقق من أن $H(x) = e^{-\lambda x}$ لقيمة $\lambda > 0$ ، وبالتالي فإن X يتبع توزيعاً أسياً بالمعلمة λ كما هو مطلوب، وهذا يكمل البرهان.

ملاحظة (7، 5):

توجد نتيجة مناظرة في حالة المتغيرات العشوائية المتقطعة وهي أن المتغير الهندسي هو التوزيع المنفصل الوحيد الذي يتمتع بخاصية فقدان الذاكرة. برهان هذه النتيجة يشبه

برهان النظرية (7 ، 6) أنظر تمرين (7 ، 8).

(7 ، 5) الأزمنة الارتدادية الأمامية

Forward recurrence times

ليكن $\{N_t\}$ عملية بواسون بالمعلمة λ ، وأن $T_1, T_2, \dots$ عبارة عن متتابعة من أزمنة الوصول المتتالية وبفرض أن $T_0 = 0$. لأي $t \geq 0$ فإن عدد الوصول في الفترة $[0, t]$ هو N_t ، وأن زمن آخر وصول قبل اللحظة t هو T_{N_t} ، وأن زمن الوصول التالي هو T_{N_t+1} . الزمن منذ آخر وصول يكون $t - T_{N_t}$ ، وأن طول الانتظار من اللحظة t حتى الوصول التالي هو

$$V_t = T_{N_t+1} - t$$

في المثال (7 ، 7) رمزنا بـ V_t إلى الزمن المتبقي لوحدة في الاستخدام عند اللحظة t .

نظرية (7 ، 7):

الاحتمال الشرطي التالي

$$P(V_t \leq u | N_s; s \leq t) = 1 - e^{-\lambda s}, u \geq 0$$

يكون مستقل عن t .

البرهان:

حيث أن

$$\begin{aligned} \{V_t \leq u\} &= \{T_{N_t+1} - t \leq u\} \\ &= \{T_{N_t+1} - t > u\}^c \\ &= \{T_{N_t+1} > u + t\}^c \\ &= \{N_{t+u} - N_t = 0\}^c \end{aligned}$$

إذن

$$\begin{aligned} P(V_t \leq u | N_s; s \leq t) &= P(\{N_{t+u} - N_t = 0\}^c | N_s; s \leq t) \\ &= 1 - P(N_{t+u} - N_t = 0 | N_s; s \leq t) \\ &= 1 - P(N_{t+u} - N_t = 0) \\ &= 1 - P(N_u = 0) \\ &= 1 - e^{-\lambda s}. \end{aligned}$$

وهذا يعني أن زمن انتظار الوصول التالي بعد اللحظة t يكون مستقل عن حالة العملية

محاضرات في الإحصاء -1-

حتى اللحظة t . وعلاوة على ذلك فإن له نفس التوزيع الذي يسلكه أي زمن وصول داخلي. وبشكل خاص لايعنينا طول الوقت الذي مضى بعد آخر وصول، فإن احتمال الانتظار لعدد من الوحدات مقداره n للوصول التالي يظل نفسه.

وكنتيجة، لدينا $E[V_t] = \frac{1}{\lambda}$. لاحظ أن $E[T_{n+1} - T_{N_t}] = \frac{1}{\lambda}$ وبالتالي فإن

$$\begin{aligned} E[T_{N_t+1} - T_{N_t}] &= E[T_{N_t+1} - t + t - T_{N_t}] \\ &= E[T_{N_t+1} - t] + E[t - T_{N_t}] \\ &= \frac{1}{\lambda} + E[t - T_{N_t}] \end{aligned}$$

وبالتالي فإن الفترة (T_{N_t}, T_{N_t+1}) والتي تغطي اللحظة t في المتوسط تكون أطول من أي فترة عادية أخرى (T_n, T_{n+1}) بين وصولين.

مثال (7، 12):

بفرض أن وصول الحافلات إلى محطة حافلات يكون بناءً على عملية بواسون بمعدل حافلة واحد كل خمس دقائق، وأنت وصلت إلى المحطة في المساء عند اللحظة $t = 617.23$ (الوقت مقاس من الساعة 7:00 صباحاً). تضمن النظرية (7، 7) أن زمن الانتظار حتى يصل أول أتوبيس يكون له نفس التوزيع الذي يتبعه زمن الانتظار إذا كان بمجرد وصولك إلى المحطة لم تتمكن من إدراك آخر أتوبيس. متوسط زمن انتظار الأتوبيس التالي هو 5 دقائق. وفي الحقيقة أنه يمكن توضيح أن متوسط زمن الإنتظار بعد وصول آخر أتوبيس هو 5 دقائق (وهذه الأخيرة تعتمد في العموم على t ولكنها تصبح 5 لقيم t الكبيرة). بمقارنة الطول المتوقع للفترة التي وصلت خلالها (بالتحديد 10 دقائق) بادعاء أصحاب شركة الحافلات أنه تصل حافلة كل خمس دقائق في المتوسط، يبدو أن هناك تناقض. إلا أنه في الحقيقة لا يوجد أي تناقض وذلك لأن متوسط طول الفترة التي وصلت خلالها يكون أكبر من متوسط الزمن بين وصول أي حافلتين متتاليتين.

(6، 7) الوضع العلوي لعمليات بواسون

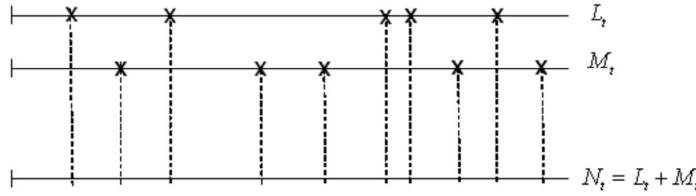
Superposition of Poisson processes

ليكن $\{L_t; t \geq 0\}$ ، $\{M_t; t \geq 0\}$ عمليتا بواسون مستقلتين بالمعلمتين λ و μ ، على الترتيب. ولأي $t \geq 0$ نفرض أن

محاضرات في الإحصاء -1-

$$N_t = L_t + M_t$$

فإن العملية العشوائية $\{N_t; t \geq 0\}$ تسمى بعملية بالوضع العلوي للعمليتين، أنظر الشكل (7) (3،).



شكل (7، 3): أزمنة وصول عملية الوضع العلوي N_t والتي تم الحصول عليها بوضع أزمنة وصول العمليتين M_t ، L_t معاً .

نظرية (7، 8):

العملية العشوائية $\{N_t; t \geq 0\}$ تكون عملية بواسون بالمعدل $\lambda + \mu$.

البرهان:

لدينا لأي $n = 0, 1, 2, \dots$

$$\begin{aligned} P(N_{s,t} = n) &= \sum_{k=0}^n P(L_{s,t} = k, M_{s,t} = n-k) \\ &= \sum_{k=0}^n P(L_{s,t} = k) P(M_{s,t} = n-k) \\ &= \sum_{k=0}^n \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{n-k}}{(n-k)!} \\ &= \frac{e^{-(\lambda+\mu)t} (\lambda + \mu)^n t^n}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\frac{\mu}{\lambda + \mu} \right)^{n-k} \\ &= \frac{e^{-(\lambda+\mu)t} (\lambda + \mu)^n t^n}{n!} \left[\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} \right]^n \\ &= \frac{e^{-(\lambda+\mu)t} ((\lambda + \mu)t)^n}{n!} \end{aligned}$$

ومن ثم فإن $N_{s,t}$ يتبع توزيع بواسون بالمعدل $\lambda + \mu$ ، وبالتالي فإن $\{N_t; t \geq 0\}$ تكون عملية بواسون بالمعدل $\lambda + \mu$ (المجموع في الخطوة قبل الأخيرة يساوي واحد وذلك لأنه عبارة مجموع التوزيع الاحتمالي لمتغير عشوائي يتبع توزيع ذي الحدين بالمعلمتين n ، $\frac{\lambda}{\lambda + \mu}$).

مثال (7، 13):

محاضرات في الإحصاء -1-

ليكن L_t يرمز إلى عدد المركبات التي تصل إلى أحد التقاطعات في الفترة $[0, t)$ من اتجاه ما، وأن M_t يرمز إلى عدد المركبات التي تصل إلى نفس التقاطع في الفترة $[0, t)$ من اتجاه آخر. وبالتالي فإن N_t يرمز إلى عدد المركبات التي تصل إلى ذلك تقاطع في الفترة $[0, t)$ من كلا الاتجاهين. إذا كان الوصول من كلا الاتجاهين يتبع عملية بواسون بالمعدل $\lambda + \mu$. فإن عملية الوصول الكلي تتبع عملية بواسون بالمعدل $\lambda + \mu$.

(7، 7) تفكك عملية بواسون

Decomposition of a Poisson process

ليكن $\{N_t; t \geq 0\}$ عملية بواسون بالمعلمة λ ، وليكن $\{X_n; n=1,2,\dots\}$ عبارة عن عملية برنولي باحتمال النجاح p ومستقلة عن $\{N_t; t \geq 0\}$ ، وبفرض أن S_n هو عدد مرات النجاح خلال عدد n من المحاولات الأولى. وبفرض أن المحاولة رقم n قد نفذت عند اللحظة T_n ، زمن الوصول رقم n . عدد المحاولات التي نفذت في الفترة $[0, t)$ هو N_t ومن ثم فإن عدد النجاحات الحاصلة في الفترة $[0, t)$ هي:

(5، 7)

$$M_t = S_{N_t}$$

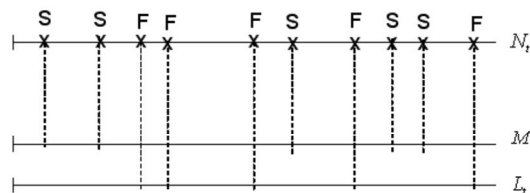
وأن عدد مرات الفشل تكون، أنظر الشكل (7، 4)،

(6، 7)

$$L_t = N_t - M_t$$

وبفرض أن $T_0 = 0$. لأي $t \geq 0$ فإن عدد الوصول في الفترة $[0, t]$ هو N_t وأن زمن آخر وصول قبل اللحظة t هو T_{N_t} وأن زمن الوصول التالي هو T_{N_t+1} . الزمن منذ آخر وصول يكون $t - T_{N_t}$ وأن طول الانتظار من اللحظة t حتى الوصول التالي هو

$$V_t = T_{N_t+1} - t$$



شكل (7، 4): أزمنة وصول V_t والتي عندها تكون النجاحات المتتالية العملية M_t والفشل المتتالي يكون العملية L_t .

نظرية (7، 9):

العمليتان العشوائيتان $\{M_t; t \geq 0\}$ ، $\{L_t; t \geq 0\}$ تكونان عمليتا بواسون بالمعدل λp ، λq ، $(q=1-p)$ على الترتيب وأنهما مستقلتين.

البرهان:

لبرهان هذه النظرية يكفي أن نثبت أنه لأي $s, t \geq 0$ ، $k, m = 0, 1, \dots$ فإن

$$(7، 7) P(M_{s+t} - M_t = m, L_{s+t} - L_t = k | M_u, L_u; u \leq t) = \frac{e^{-\lambda ps} (\lambda ps)^m}{m!} \frac{e^{-\lambda qs} (\lambda qs)^k}{k!}$$

الحادثة

$$\{M_{s+t} - M_t = m, L_{s+t} - L_t = k\}$$

تكافئ الحادثة

$$\{N_{s+t} - N_t = m + k, M_{s+t} - M_t = m\}$$

والتي بدورها تكافئ الحادثة

$$A = \{N_{s+t} - N_t = m + k, S_{N_{s+t}} - S_{N_t} = m\}$$

ومن الناحية الأخرى فإن عائلة المتغيرات العشوائية $\{M_u, L_u; u \leq t\}$ تكافئ العائلة $H = \{N_u; u \leq t, X_1, X_2, \dots, X_{N_t}\}$. وحيث أن $\{M_u, L_u; u \leq t\}$ عملية بواسون وأن العمليتين N ، S مستقلتين إذن المتغير العشوائي $N_{t+s} - N_t$ يكون مستقل عن H ، وبالمثل وحيث عدد النجاحات في المحاولات ذات الأرقام $N_{t+1}, N_{t+2}, \dots, N_{t+s}$ يكون مستقل عن الماضي $X_1, X_2, \dots, X_{N_t}$ وأن العمليتين N ، S مستقلتين إذن المتغير العشوائي $S_{N_{t+s}} - S_{N_t}$ يكون مستقل عن H . ومن ثم فإن الحادثة A تكون مستقلة عن H وبالتالي نحصل على الطرف الأيسر للعلاقة (7، 7) على النحو التالي:

$$\begin{aligned} P(A) &= P\{N_{s+t} - N_t = m + k, S_{N_{s+t}} - S_{N_t} = m\} \\ &= \sum_{n=0}^{\infty} P\{N_t = n, N_{s+t} - N_t = m + k, S_{N_{s+t}} - S_{N_t} = m\} \\ &= \sum_{n=0}^{\infty} P\{N_t = n, N_{s+t} = m + k + n, S_{m+k+n} - S_n = m\} \\ &= \sum_{n=0}^{\infty} P\{N_t = n, N_{s+t} = m + k + n\} P\{S_{m+k+n} - S_n = m\} \end{aligned}$$

وذلك باستخدام خاصية الاستقلال بين N ، S . وحيث أن $S_{n+m+k} - S_n$ له نفس توزيع المتغير

العشوائي S_{m+k} إذن

محاضرات في الإحصاء -1-

$$\begin{aligned}
 P(A) &= \sum_{n=0}^{\infty} P\{N_t = n, N_{s+t} - N_t = m + k\} P\{S_{m+k} = m\} \\
 &= P\{N_{s+t} - N_t = m + k\} P\{S_{m+k} = m\} \\
 &= \frac{e^{-\lambda s} (\lambda s)^{m+k}}{(m+k)!} \frac{(m+k)!}{m!k!} p^m q^k
 \end{aligned}$$

والذي يساوي الطرف الأيمن للعلاقة (7، 7) وهذا يكمل البرهان.

مثال (7، 14):

تصل مركبات إلى نقطة تحويل في أحد الطرق ثم تتحول إلى اليمين باحتمال 0.6 أو إلى اليسار باحتمال 0.4 بشكل مستقل. تمثل عملية الوصول إلى هذه التحويلة بعملية بواسون بالمعدل $\lambda = 30$ عربة في الدقيقة. إذن المركبات التي تتحول إلى اليسار تكون عملية بواسون بالمعدل $30 \times 0.6 = 18$ عربة في الدقيقة، وأن المركبات التي تتحول إلى اليمين تكون عملية بواسون بالمعدل $30 \times 0.4 = 12$ عربة في الدقيقة.

مثال (7، 15):

عملية وقوف المركبات عند أحد المطاعم على أحد جانبي طريق سريع تكون عملية بواسون بالمعدل $\lambda = 20$ في الساعة. تحتوي المركبة على شخص أو شخصين أو ثلاثة أو أربعة أو خمسة أشخاص باحتمالات 0.3، 0.3، 0.2، 0.1، 0.1 على الترتيب. نريد حساب العدد المتوقع للأشخاص الذين يصلون إلى المطعم خلال ساعة.

بفرض أن $N^1, N^2, \dots, N^5$ عدد المركبات التي تصل خلال ساعة وتحتوي على شخص، شخصين،...، خمس أشخاص على الترتيب. بناءً على نظرية (7، 8) فإن المتغيرات العشوائية $N^1, N^2, \dots, N^5$ تتبع توزيع بواسون بالمعدلات $20 \times 0.3 = 6$ ، $20 \times 0.2 = 4$ ، $20 \times 0.1 = 2$ على الترتيب. وبالتالي فإن $E[N^1] = E[N^2] = 6$ ، $E[N^3] = 4$ ، $E[N^4] = E[N^5] = 2$ ومن ثم فإن العدد المتوقع للأشخاص الذين يصلون إلى المطعم خلال ساعة هو

$$E[N^1 + 2N^2 + 3N^3 + 4N^4 + 5N^5] = 6 + 2 \times 6 + 3 \times 4 + 4 \times 2 + 5 \times 2 = 48$$

(7، 8) عمليات بواسون المركبة

Compound Poisson process

حتى الآن فرضنا أن القفزات في عملية بواسون كانت وحدة الحجم. والآن سنحذف هذا الفرض وسنسمح بقفزات بأي حجم.

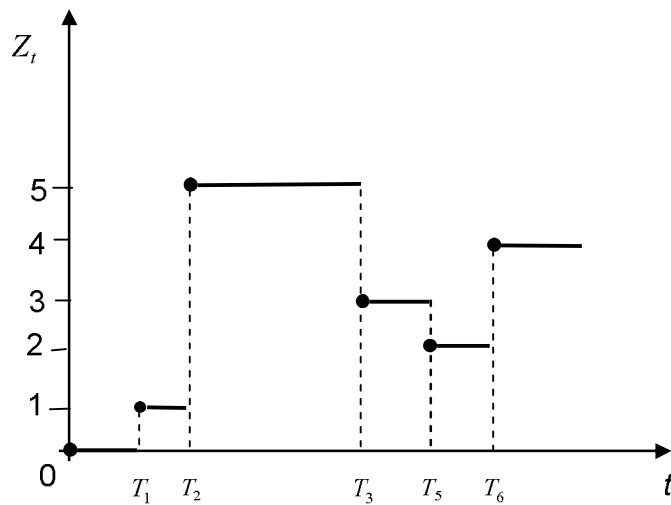
فيما يلي سنأخذ العملية العشوائية $Z = \{Z_t; t \geq 0\}$ حيث أن الدالة $t \rightarrow Z_t$ تكون متصلة من اليمين ولتأخذ قيماً حقيقية متزايدة أو متناقصة فقط مع القفزات وأن $Z_0 = 0$ ، أنظر الشكل (4، 7).

تعريف (3، 7):

يقال للعملية العشوائية $\{Z_t; t \geq 0\}$ أنها عملية مركبة لبواسون compound Poisson process إذا تحققت الشروط التالية:

- (أ) الدالة $t \rightarrow Z_t$ يكون لها عدد محدود من القفزات في الفترات المحدودة،
- (ب) لأي $t, s \geq 0$ فإن $Z_{t+s} - Z_t$ يكون مستقل عن الماضي $\{Z_u; u \leq t\}$ حتى اللحظة t .
- (ج) لأي $t, s \geq 0$ فإن توزيع $Z_{t+s} - Z_t$ يعتمد فقط على s (مستقل عن t).

ليكن $\{Z_t; t \geq 0\}$ عملية مركبة لبواسون. بناءً على الفرضية (أ) فإن لها عدد منتهى من القفزات خلال أي فترة محدودة ومن ثم فإنه يمكن ترتيب الأزمنة التي يحدث عندها القفزات. ليكن $T_1, T_2, \dots$ هي الأزمنة التي عندها أول قفزة، ثاني قفزة، ...، الخ، وأن مقادير القفزات عندها على الترتيب هي $X_1, X_2, \dots$.



شكل (5، 7): عملية مركبة لبواسون تتزايد أو تتناقص بقيم عشوائية عند أزمنة

محاضرات في الإحصاء -1-

الوصول لعملية بواسون. وكتحقيق وضعنا $T_1=1.5, T_2=3.7, T_3=4.2$ ،
 $T_4=4.5, \dots$ وأن حجم القفزات المقابلة هي $X_1=2.2, X_2=-0.8, \dots$
 $X_3=-1.2, X_4=0.5, \dots$

بناءً على الفرضية (ب) فإن عدد القفزات في الفترة $[t, t+s]$ يكون مستقل على تاريخ العملية في الماضي في الفترة $[0, t]$ ، وباتباع الفرضية (ج) فإن توزيع عدد القفزات في الفترة $[t, t+s]$ يكون مستقل عن t . ومن ثم فإن $T_1, T_2, \dots$ يجب أن تكون أزمنة الوصول في عملية بواسون $\{N_t; t \geq 0\}$ ، بمعنى أن N_t هو عدد قفزات Z في الفترة $[0, t]$ ، وبالتالي فإن $\{Z_t; t \geq 0\}$ تكون عملية بواسون. تختلف العملية العشوائية $\{Z_t; t \geq 0\}$ عن $\{N_t; t \geq 0\}$ في أن مقدار القفزات في $\{Z_t; t \geq 0\}$ تكون عبارة عن متغير عشوائي أما في $\{N_t; t \geq 0\}$ فتكون جميع القفزات محددة ومتساوية وتساوي الوحدة. وعلاوة على ذلك وبناءً على (ب)، (ج) فإن أحجام القفزات $X_1, X_2, \dots$ تكون عبارة عن متغيرات عشوائية مستقلة ولها نفس التوزيع والتي تكون أيضاً مستقلة عن $T_1, T_2, \dots$.

وبالعكس فإذا كانت $T_1, T_2, \dots$ أزمنة وصول لعملية بواسون وأن $X_1, X_2, \dots$ متغيرات عشوائية مستقلة ولها نفس التوزيع ومستقلة عن $T_1, T_2, \dots$ فإن العملية العشوائية $\{Z_t; t \geq 0\}$ والتي تنتج من تجميع جميع X_i والتي تحقق أن $T_i \leq t$ ، بمعنى $Z_t = \sum_{i: T_i \leq t} X_i$ ، تكون عملية مركبة لبواسون. وبالتالي يكون لدينا التخصيص الوصفي التالي.

نظرية (7، 10):

العملية العشوائية $\{Z_t; t \geq 0\}$ تكون عملية مركبة لبواسون إذا وفقط إذا كانت الأزمنة التي يحدث عندها القفزات تكون عملية بواسون وأن مقادير القفزات المتعاقبة تكون عبارة عن متغيرات عشوائية مستقلة ولها نفس التوزيع ومستقلة عن أزمنة القفزات.

مثال (7، 16):

تكون عملية وصول الزبائن إلى أحد المتاجر عملية بواسون $\{N_t; t \geq 0\}$ ، كمية النقود المنفقة بواسطة الزبون رقم n تكون متغير عشوائي X_n مستقل عن جميع الأزمنة التي يصل عندها الزبائن (بما فيهم الزبون رقم n) ومستقل أيضاً عن باقي النفقات التي ينفقها

محاضرات في الإحصاء -1-

باقي الزبائن. كمية النفقات التي تدفع بواسطة أول عدد n من الزبائن هو $Y_n = X_1 + \dots + X_n$ وذلك لأي $n \geq 1$ وأن $Y_0 = 0$. وحيث أن عدد الزبائن الذين يصلون في الفترة $[0, t)$ هو N_t وأن كمية المبيعات التي تباع للزبائن الذين يصلون في الفترة $[0, t)$ هو $Z_t = Y_{N_t}$. فرضيتنا على العملية $\{N_t; t \geq 0\}$ والمتغير العشوائي X_n هو أن العملية $\{Z_t; t \geq 0\}$ تكون عملية مركبة لبواسون.

مثال (7، 17):

بفرض أن أزمنة الأعطال المتعاقبة لجهاز حاسب آلي تكون متغيرات عشوائية مستقلة ومتماثلة وتتبع توزيع أسي. ومع كل عطل يوجد ثمن تكلفة للإصلاح. التكلفة المقابلة لكل إصلاح تكون عبارة عن متغيرات عشوائية مستقلة ومتماثلة (متطابقة)، وأن تكلفة الإصلاح مستقلة عن الزمن الذي يحدث عنده العطل. إذا كانت Z_t هو زمن التكلفة التراكمي للإصلاحات التي تمت في الفترة $[0, t)$ ، إذن العملية العشوائية $\{Z_t; t \geq 0\}$ تكون عملية مركبة لبواسون.

نظرية (7، 11):

الدالة المولدة لاحتمالات عملية بواسون المركبة $\{Z_t; t \geq 0\}$ هي:

$$E[u^{Z_t}] = e^{\lambda t(a(u)-1)},$$

حيث أن $a(u) = E[u^{X_i}]$.

البرهان:

يترك كتمرين.

تقدم النظرية التالية نتيجة أكثر شمولية.

نظرية (7، 12):

بفرض أن الأزمنة التي يتم عندها القفزات في العملية العشوائية $\{Z_t; t \geq 0\}$ تكون عملية بواسون بالمعدل λ وأن مقادير القفزات تكون متغيرات عشوائية مستقلة ومتطابقة ولها التوزيع ϕ وتأخذ قيمة غير سالبة ومستقلة عن الأزمنة التي يحدث عندها القفزات. وبالتالي فإنه لأي $\alpha \geq 0$ ، $t \geq 0$ يكون:

محاضرات في الإحصاء -1-

$$(8، 7) \quad E[e^{-\alpha Z_t}] = e^{-\lambda t(1-f(\alpha))},$$

حيث أن

$$f(\alpha) = E[e^{-\alpha X_t}] = \int_0^\infty e^{-\alpha u} d\phi(u)$$

البرهان:

باستخدام خاصية استقلال X_i عن $\{N_t; t \geq 0\}$ وحيث أن $Z_t = X_1 + \dots + X_n$ عندما $N_t = n$ ، إذن

$$E[e^{-\alpha Z_t} | N_t = n] = E[e^{-\alpha(X_1 + \dots + X_n)}]$$

وحيث أن $X_1, \dots, X_n$ مستقلين إذن

$$E[e^{-\alpha(X_1 + \dots + X_n)}] = E[e^{-\alpha X_1}] \dots E[e^{-\alpha X_n}] = [f(\alpha)]^n$$

وبالتالي يمكن كتابة أن

$$E[e^{-\alpha Z_t} | N_t] = [f(\alpha)]^{N_t}$$

ومن ثم فإن

$$E[e^{-\alpha Z_t}] = [f(\alpha)]^{N_t} = e^{-\lambda t(1-f(\alpha))}$$

وذلك لأنه لأي $\beta \geq 0$ يكون لدينا

$$E[\beta^{N_t}] = \sum_{k=0}^{\infty} \beta^k \frac{e^{-\lambda t} (\lambda t)^k}{k!} = e^{-\lambda t(1-\beta)}$$

(9، 7) عمليات بواسون غير المستقرة

Non-stationary Poisson process

في العديد من التطبيقات يتطلب الأمر إلى عمليات بواسون بمعدل متغير مع الزمن، أي أن $\lambda = \lambda(t)$. يسمى هذا النوع من عمليات بواسون بعمليات بواسون غير المستقرة أو غير المتجانسة. لعمليات بواسون غير المستقرة بالمعدل $\lambda = \lambda(t)$ فإن مقدار التغير $N_t - N_s$ يعطي عدد الوصول في الفترة $[s, t]$ ويتبع توزيع بواسون بالمعلمة $\int_s^t \lambda(u) du$ وأن مقادير التغير r على الفترات المتباعدة (غير المتقاطعة) تكون متغيرات عشوائية مستقلة.

مثال (7، 18):

تتبع عملية الطلب على خدمة متميزة في أحد الأحياء عملية بواسون غير المستقرة

محاضرات في الإحصاء -1-

بالمعدل $\lambda = \lambda(t)$ والمعرف كالآتي:

$$\lambda(t) = \begin{cases} 2t, & 0 \leq t < 1, \\ 2, & 1 \leq t < 2, \\ 4-t, & 2 \leq t \leq 4, \end{cases}$$

حيث يقاس الزمن t بالساعة من فتح محل الخدمة. ما احتمال حدوث طلبين خلال أول ساعتين من بدأ الخدمة وطلبين في الساعتين التاليتين؟
الحل:

حيث أن الطلب على الفترات المتباعدة عبارة عن متغيرات عشوائية مستقلة، فإنه يمكن الإجابة عن السؤال بشكل منفصل. متوسط الطلب في أول ساعتين، أي على الفترة $[0, 2]$ ، هو

$$\mu = \int_0^2 \lambda(t) dt = \int_0^1 2t dt + \int_1^2 2 dt = 3$$

وبالتالي فإن

$$P(N_2 = 2) = \frac{e^{-3} 3^2}{2!} = 0.2240$$

وبالمثل متوسط الطلب في الساعتين التاليتين، أي على الفترة $[2, 4]$ ، هو

$$\mu = \int_2^4 \lambda(t) dt = \int_2^4 (4-t) dt = 2$$

وبالتالي فإن

$$P(N_4 - N_2 = 2) = \frac{e^{-2} 2^2}{2!} = 0.2707$$

(7، 10) تمارين

(7، 1) يصل الزبائن إلى محل حلقة بناء على عملية بواسون بمعدل 3 أشخاص كل ساعة. باستخدام الطريقتين المقدمتين في مثال (7، 6) أوجد احتمال عدم وصول زبائن إلى المحل في الفترة من الساعة 8 إلى الساعة 10 صباحاً.

(7، 2) ليكن $\{N_t; t \geq 0\}$ عملية بواسون بالمعدل $\lambda = 15$ ، احسب ما يلي:

(أ) $P(N_6 = 9)$ ،

(ب) $P(N_6 = 9, N_{20} = 13, N_{56} = 27)$ ،

(ج) $P(N_{20} = 13 | N_6 = 9)$ ،

$$(د) \cdot P(N_6 = 9 | N_{20} = 13)$$

(3، 7) ليكن $\{N_t; t \geq 0\}$ عملية بواسون بالمعدل λ :

(أ) وضح أنه لأي $s < t$ فإن

$$P(N_s = k | N_t = n) = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}, k = 0, 1, \dots, n$$

(ب) بفرض أن $\lambda = 3$ في الساعة، أوجد احتمال وقوع حادثتين في أول ساعة بشرط وقوع خمس حوادث في الثلاث ساعات الأولى.

(4، 7) ليكن $\{N_t; t \geq 0\}$ عملية بواسون بالمعدل λ ، احسب $E[N_t \cdot N_{t+s}]$.

(5، 7) ليكن $\{N_t; t \geq 0\}$ عملية بواسون بالمعدل $\lambda = 2$ ، احسب:

$$(أ) E[N_t], \text{Var}[N_t]$$

$$(ب) E[N_{t+s} | N_t]$$

(6، 7) عدد الطائرات الخاصة التي تهبط إلى أحد المطارات يتبع عملية بواسون بقيمة متوقعة 24 خلال ثمان ساعات في يوم عمل. وبفرض أن الوقت مقاس بالساعة:

(أ) أوجد احتمال أن زمن الانتظار بين هبوط أي طائرتين متتاليتين يكون أقل من أو يساوي 30 دقيقة.

(ب) أوجد المتوسط والانحراف المعياري للزمن بين الهبوط.

(7، 7) تحدث العواصف الرعدية خلال فصل الصيف على إحدى المدن بناء على عملية بواسون بمعدل $\lambda = 0.8$ كل أسبوع، أوجد:

(أ) احتمال انتظار عاصفتين يكون أكثر من خمس أيام.

(ب) المتوسط الحسابي والانحراف المعياري لزمن مشاهدة عاصفتين.

(8، 7) برهن الخاصية (5، 7).

(9، 7) كون وصول المسافرين عند أحد محطات الحافلات عملية بواسون بالمعدل $\lambda = \frac{1}{3}$ في الدقيقة. بفرض أن الحافلة قد تركت المحطة للتو عند اللحظة $t = 0$ ولم تترك أي من المسافرين عندها. وليكن T زمن وصول الحافلة التالية ومن ثم فإن عدد

محاضرات في الإحصاء -1-

المسافرين عند المحطة عند لحظة وصوله هو N_T . وبفرض أن T ، N مستقلين، وأن T يتبع التوزيع ϕ ، احسب مايلي:

$$(أ) \quad E[N_T^2 | T], E[N_T | T]$$

$$(ب) \quad E[N_T], Var[N_T] \text{ لأي}$$

$$d\phi(t) = \begin{cases} \frac{1}{2}dt, & 9 \leq t \leq 11, \\ 0, & \text{otherwise.} \end{cases}$$

(7، 10) أعلن أحد المتاجر عن هدية لكل زبون يصل مع رقم 30، بفرض أن وصول الزبائن إلى هذا المتجر يكون عملية بواسون بالمعدل λ ، أوجد دالة الكثافة الاحتمالية للأزمنة بين الوصول المحظوظ (وصول الزبائن الذين سيحصلون على هدايا).

(7، 11) بفرض أن عدد الوصول في الفترة $B = (t, t+b]$ معلوم ويساوي k ولكن الأزمنة الفعلية لوصولهم غير معلومة. وبالتالي فإن كل وصول من هذا العدد k يتبع توزيع منتظم على الفترة B ومستقل عن باقي الـ $k-1$ وصول. بناءً على هذا التصور فإن معرفة عدد الوصول في فترة ما B لا يقدم أي معلومات عن الأزمنة التي يحدث عندها هذا الوصول (بخلاف المعلومة التافهة وهي أنهم وصول داخل هذه الفترة). ولصيغة هذه المسألة دعنا نفرض أن $A_1, A_2, \dots, A_n$ فترات متباعدة اتحادهم يعطي الفترة B (أي أنهم تجزئ للفترة B) وأن $a_1, a_2, \dots, a_n$ عبارة عن الأطوال المناظرة لهذه الفترات وبأخذ $b = a_1 + \dots + a_n$. إذن لأي $k = k_1 + \dots + k_n$ فإن $k_1, \dots, k_n \in \{0, 1, 2, \dots\}$

$$P\{N_{A_1} = k_1, \dots, N_{A_n} = k_n \mid N_B = k\} = \frac{k!}{k_1! \dots k_n!} \left(\frac{a_1}{b}\right)^{k_1} \dots \left(\frac{a_n}{b}\right)^{k_n}$$

استخدم ذلك في المسألة التالية. بفرض أن الزبائن يصلوا إلى أحد المتاجر بناءً على عملية بواسون. وبفرض أن أول زبونين قد وصلا خلال أول ساعتين. أوجد احتمال:

(أ) أنهما قد وصلا أثناء العشرين دقيقة الأولى.

(ب) على الأقل أحدهما قد وصل أثناء العشرين دقيقة الأولى.

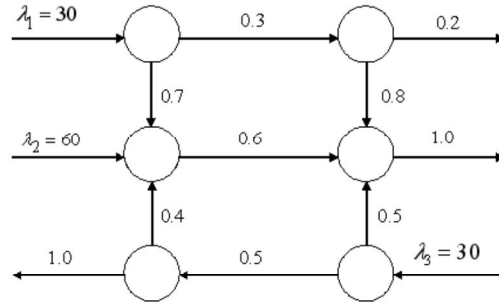
(7، 12) يصل الزبائن إلى أحد محلات المدينة بناءً على عملية بواسون $\{N_t; t \geq 0\}$ بمعدل

$\lambda = 20$ كل ساعة. أوجد العدد المتوقع للمبيعات خلال ثمان ساعات في يوم عمل إذا

كان احتمال أن الزبون يشتري شئ ما هو 0.3.

محاضرات في الإحصاء -1-

(7، 13) بفرض شبكة الخطوط الموضحة بالشكل (7، 6). المدخلات عبارة عن عملية بواسون بالمعدلات الموضحة بالرسم واحتمالات اختيار المركبات الاتجاهات الموضحة مكتوبة على الأسهم التي توضح اتجاه السير. أوصف تدفق السير على كل فرع من فروع هذه الشبكة.



شكل (7، 6): شبكة الخطوط لتمارين (7، 13).

(7، 14) يحتوي أحد أقسام مخزن ما ثلاثة أبواب. الوصول إلى كل من هذه الأبواب يكون عملية بواسون بالمعدلات $\lambda_1 = 110$ ، $\lambda_2 = 90$ ، $\lambda_3 = 110$ زبون في الساعة على الترتيب، وأن 30% من الزبائن ذكور. احتمال أن يشتري أحد الزبائن من الذكور شيئاً هو 0.8 أما سيطري أحد الإناث شيئاً باحتمال 0.1، وأن متوسط قيمة المشتريات هو 4.5 ريال.

- (أ) ما هو متوسط قيمة المبيعات الكلية خلال 10 ساعات في اليوم؟
 (ب) ما هو احتمال أن أحد الزبائن من الإناث والتي ستشتري شيئاً ما ستصل خلال خمسة عشر دقيقة الأولى؟ وما هو الزمن المتوقع لوصولها؟

(7، 15) استخدم نظرية (7، 11) لحساب المتوسط والتباين لعملية بواسون المركبة.

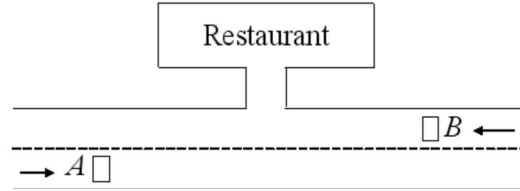
(7، 16) بفرض أن حركة المرور على إحدى الطرق الموضحة بالشكل (7، 7) هي كما يلي: عدد المركبات التي تعبر النقطة A يتبع توزيع بواسون بالمتوسط 60، وأن 20% منهم عربات نقل. عدد المركبات التي تعبر النقطة B يتبع توزيع بواسون بالمتوسط 80، وأن 30% منهم عربات نقل. وعلى العموم يقف 10% من جميع المركبات عند المطعم وأن كل عربة نقل تحتوي على شخص واحد وأما السيارة الواحدة تحتوي على شخص واحد باحتمال 0.3 وشخصين باحتمال 0.3، ثلاث

محاضرات في الإحصاء -1-

باحتمال 0.2، أربع باحتمال 0.1، أو خمس باحتمال 0.1.

(أ) أوجد العدد المتوقع $E[Z]$ للأشخاص الذين يصلون إلى المطعم خلال ساعة واحدة.

(ب) احسب $E[\alpha^Z]$ لأي $\alpha \in [0,1]$.



شكل (7، 7): صورة التمرين (7، 16).

(7، 17) يخضع جهاز ما لصدمات تحدث بناء على عملية بواسون $\{N_t; t \geq 0\}$ بالمعدل λ

. يمكن أن يخفق الجهاز نتيجة لصدمة واحدة فقط وأن احتمال أن صدمة ما ستسبب

فشل النظام هو p ومستقل عن الأزمنة التي حدثت عندها الصدمات السابقة. ليكن

K عدد الصدمات الكلي التي يتلقاها الجهاز قبل الاخفاق، وأن $T = T_K$ هو الزمن

الذي حدث عنده الاخفاق. احسب ما يلي:

(أ) $E[T]$ و $\text{Var}[T]$ ،

(ب) $E[T|K]$ ،

(ج) $E[T|K > 9]$.

مجموعة تمارين.

التمرين 01: إليك أسعار منتج البطاطا لـ 100 يوم في ولاية الوادي:

70-50-53-60-66-82-32-70-57-90-51-53-52-41-50-64-56-80-53-61-
54-53-60-61-45-45-58-42-60-38-80-73-60-69-63-43-
59-49-41-65-84-82-50-65-50-52-23-60-68-92-33-24-24-45-50-42-
56-48-55-50-45-53-43-80-100-55-33-57-66-56-54-61-30-33-70-58-
50-36-60-25-55-50-27-26-50-63-66-45-30-58-60-50-69-52-84-61.

المطلوب:

1/ حدد إحصائياً المجتمع، الوحدة، المتغير، نوعه.

2/ كون جدول توزيع تكراري بإستعمال قاعدة *Sturge*.

3/ أوجد عدد الطلبة الذين تفوق علاماتهم عن 73 نقطة؟.

4/ أوجد عدد الطلبة الذين تقل علاماتهم عن 53 نقطة؟.

التمرين 02: البيانات التالية تبين تقدير 30 طالب من الدفعة C في إمتحان الإحصاء.

جيد، مقبول، جيد، جيد جداً، جيد، جيد جداً، جيد، مقبول، ضعيف، ضعيف، ممتاز، جيد جداً، ممتاز،
ضعيف، مقبول، ضعيف، ممتاز، مقبول، جيد، جيد، مقبول، مقبول، جيد جداً، جيد، مقبول، جيد جداً
مقبول، جيد، جيد، مقبول، .

والمطلوب منك:

1- ضع هذه المعلومات في توزيع تكراري؟.

2- ضع العرض البياني المناسب؟.

محاضرات في الإحصاء -1-

3- هل يمكن حساب الوسيط والمنوال؟ علل إجابتك؟.

التمرين 03: الجدول التالي يمثل توزيع الأجور الأسبوعية لعمال في مؤسسة صناعية.

الفئات - الأجور	20-10	30-20	40-30	50-40	60-50	70-60	80-70
ت م الصاعد	04	11	21	36	43	47	50

المطلوب:

- 1- ما هو حجم العينة لهذا التوزيع؟.
- 2- كم عدد العمال الذين أجورهم تساوي أو يقل عن 45 ؟.
- 3- قارن بين الوسط الحسابي والوسيط والمنوال؟. وماذا تلاحظ ؟.
- 4- هل تتغير قيمة الوسط الحسابي والوسيط و المنوال إذا غيرنا طول الفئة؟.
- 5- أحسب التباين والانحراف المعياري ومعامل الاختلاف لهذا التوزيع؟.

التمرين 04: البيانات التالية تمثل معدل سعر منتج معين في 20 ولاية.

98.5 77.2 60 69.8 82.7 63.286.1 89.1 7365.3

84.1 87.261.3 66.6 907892 69.1 72.2 61.3

المطلوب:

- 1- ضع هذه البيانات في توزيع تكراري ذو 5 فئات متساوية.
- 2- أرسم المدرج التكراري والمضلع التكراري لهذا التوزيع.
- 3- حدد قيمة الوسيط وقيمة المنوال بيانياً وبدقة؟. وفسر إحصائياً ماذا تعني هتين القيمتين؟
- 4- عندما يكون طول الفئة غير متساوي هل نصحح التكرار عند حساب المنوال ؟ ولماذا؟.
- 5- أدرس شكل هذا التوزيع بإستعمال مقياس فيشر.

التمرين 05: أجب على الأسئلة التالية:

محاضرات في الإحصاء -1-

- 1- هل يمكن أن يكون الوسط الحسابي للتوزيع تكراري مبوب مساوٍ لقيمة سالبة؟ ولماذا؟.
- 2- هل إذا كان لدينا $X = \text{Med} = \text{Mod}$ نقول أن معامل الإلتواء لبرسون S وكذلك معامل يول وكاندال S' معدومين؟ ولماذا؟.
- 3- هل يمكن أن نستخدم معامل الارتباط الرتبي سبيرمان عند المتغيرات الكمية المستمرة؟ ولماذا؟.
- 4- هل من الضروري حساب الانحراف المعياري عند دراسة التشتت؟ ولماذا؟
- 5- أعط مثال من عندك تبين فيه كيفية تحديد الربع الأول والربع الثالث بيانياً؟
- 6- عندما يكون التوزيعاً قريباً لمتماثل التناظر فهناك علاقة بين الانحراف المعياري والمد بالربعين وكذلك علاقة بين الانحراف المعياري واريو الانحراف المتوسط، بينهما؟ .
- 7- هنا كطرف متعدد لحساب المنوال، بين طريقتين مختلفتين حسابياً؟.

التمرين

06: يمثل الجدول التالي توزيع عينتين من عمال وأعمال تحسباً لأجور اليومية بمئات الدينارات، والمطلوب منك أنتقارنتشتت وشكل أجور هذا العينةتان.

الفئات	15-5	20-15	25-20	35-25	45-35	55-45
عدد العمال	04	10	36	30	16	04
عدد العاملات	20	48	24	06	02	00

التمرين 07: فسر احصائياً علاقة المدخنين بمرضا السرطان لعشرة ولايات جزائرية كما يبينها الجدول التالي:

عدد المدخنين	450	490	400	600	1200	200	550	800	650	1000
عدد المصابين	10	11	08	12	16	06	10	11	09	13

محاضرات في الإحصاء -1-

التمرين 08: أعطت نتائج امتحان مقياس لإحصاء لدفعة C سنة أولنا المعلومات التالية :

$V(x)=10$ ، $CV=63.24\%$ ، وكان التوزيع معتماتلو متناظر، والمطلوب:

أوجد الوسط الحسابي والمدالي الربيعي والانحراف المتوسط.

التمرين 09: إليك التوزيع التكراري التالي:

الفئات	15-10	20-15	30-20	45-30	50-45	72-50	80-72
عدد الطلبة	12	18	28	28	15	10	07

1- صحح التكرار بطريقتين.

2- أحسب الوسط الحسابي والوسيط والمنوال البعد تصحيح التكرار بطريقة المضاعف، ثم قارن بينهما؟ وما تلاحظ؟

التمرين 10: إليك الجدول التالي:

x	09	08	07	06	04	03	01
y	42.4	32.6	24.6	18.6	10.8	9.2	4.6

1- تحديد شكل العلاقة الموجود بين المتغيرين.

2- قدر هذه العلاقة وفق النموذج الأسّي $Y=b(a^X)$

3- أحسب معامل الارتباط وعلق عليه.

التمرين 11: إليك الجدول التالي:

x	140	210	246	370	240	480	520	560
y	180	88	54	24	18	14	08	04

1- تحديد شكل العلاقة بين الكميات المطلوبة والسعر.

محاضرات في الإحصاء -1-

2- قم بتحديد النموذج وفقاً للصيغة التالية $y=b/x^a$

3- أحسب معامل التحديد وعلق عليه؟

4- قدر الكمية المتوقعة عندما تكون $x=60$

التمرين 12: نفس معطيات التمرين رقم 09، المطلوب:

1- أحسب الوسط الحسابي، الوسط التربيعي، الوسط التوافقي، الوسط الهندسي.

2- ما هي نسبة معامل التغير (الاختلاف).

التمرين 13: الجدول الآتي يبين معدل البطالة والجريمة في دولة ما.

السنة	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
معدل البطالة	5.0	5.2	4.9	7.3	6.5	10.5	10.8	8.5	8.0	6.5
معدل الجريمة	2.5	2.4	3.1	4.3	3.8	5.0	4.8	4.1	8.3	2.8

1- أوجد العلاقة ما بين الزمن ومعدل البطالة، ثم أوجد العلاقة ما بين الزمن ومعدل الجريمة؟

2- أوجد العلاقة ما بين معدل البطالة ومعدل الجريمة؟.

3- أوجد معامل الارتباط، ومعامل الارتباط الرتبي؟ أيهما أفضل؟

4- قدر معدل الجريمة إذا وصل معدل البطالة 12%.

التمرين 14: الجدول التالي يبين تطور أسعار وكميات 4 سلع إستهلاكية لفترتين:

q_1	P_1	q_0	P_0	
50	30	60	18	السلعة 1
80	18	72	22	السلعة 2
100	46	90	28	السلعة 3
80	48	120	32	السلعة 1

1- حساب الرقم القياسي التجميعي للأسعار والكميات؟

محاضرات في الإحصاء -1-

2- تحديد الوسط الحسابي للأرقام القياسية؟

3- تحديد الأرقام القياسية لكل من لاسبير، باش، فيشر، مارشال: للأسعار والكميات.

التمرين 15: إشتري شخص في بداية شهر، 2000 أورو من السوق الموازية بسعر 1 أورو = 15 دج، في

نصف الشهر إستبدل 5000 أورو من نفس السوق بسعر 1 أورو = 15.5 دج، وفي نهاية الشهر

إشتري 1000 أورو بسعر 1 أورو = 14.5 دج، المطلوب تحديد سعر الصرف المتوسط؟

التمرين 16:

18-16	16-14	14-12	12-10	10-08	08-06	06-04	04-02	x_i
05	06	14	22	29	45	34	15	$fi1$
02	03	13	20	32	48	28	21	$fi2$

تم تصحيح إمتحان الإحصاء لدفعة C والدفعة D فكانت النتائج وفق الجدول أعلاه، والمطلوب منك

تحديد:

1- ما هو العرض البياني المناسب؟.

2- أوجد الوسط الحسابي بطريقة الوسط الفرضي لدفعتين؟.

3- المعدل العام لدفعتين؟ .

4- العلامة المتحصل عليها من طرف أغلبية الطلبة لكل دفعة؟.

5- شكل التوزيعين بإستعمال الخصائص الثلاثة للنزعة المركزية؟.

6- علامة الطلبة ذو المرتبة 20%.

7- هل هناك فرق بين علامات الدفعتين؟.

8- وكيف يتم تقليص هذا الفرق؟.

9- قارن بين تشتت الدفعتين؟

محاضرات في الإحصاء -1-

التمرين 17: بين الجدول التالي درجات الحرارة لولاية باتنة على الساعة التاسعة صباحاً في أول جانفي

على مدى 10 سنوات.

السنة	درجات الحرارة
1	-3
2	2
3	6
4	0
5	-1
6	-2
7	0
8	-4
9	0
10	1

1- هل نسمي هذا الجدول توزيع تكراري غير مبوب؟ لماذا؟.

2- حدد متوسط درجات الحرارة؟.

3- حدد الانحراف المتوسط والانحراف المعياري؟.

4- ما هو توقعك لدرجة الحرارة لسنة 11.

التمرين 18: لدينا ثلاث مصانع A B C لإنتاج مادة معينة وكانت الكميات المنتجة وأسعار البيع لهذه

المادة في فترتين 2005-2010 ممثلة في الجدول التالي:

الكمية		السعر		
2010	2005	2010	2005	
90	70	14	12	A
120	100	11	10	B
70	80	15	16	C

1- أحسب الرقم القياسي لسعر المصنع A اعتباراً أن سنة الأساس هي 2005.

2- أذكر خصائص الأرقام القياسية الأولية.

3- أحسب الرقم القياسي للأسبير للأسعار مع التعليق على النتيجة؟.

4- أحسب الرقم القياسي لباش للكميات مع التعليق على النتيجة؟

محاضرات في الإحصاء -1-

التمرين 19: الجدول التالي يمثل تطور سعر لتر الحليب بالدينار .

السنة	2005	2006	2007	2008	2009	2010	2011
السعر	50	53	56	61	64	67	70

1- أحسب متوسط نمو السعر؟

2- أحسب الرقم القياسي البسيط للأسعار معتمداً سنة 2005 كسنة الأساس.

التمرين 20: ليكن لديك الجدول التالي:

20	15	10	5	3	xi
10	f4	9	f2	6	fi

مع العلم أن الوسط الحسابي يساوي 12.26، والوسط الهندسي يساوي 8.33

1- أوجد f4 و f2.

2- أحسب الوسط التربيعي والوسط التوافقي.

3- قارن بين الوسط الحسابي والوسط والمنوال؟

التمرين 21: أجرت شركة سونلغاز دراسة إحصائية لمعرفة مدة إشتغال مصابيح في مصنع معين:

مدة الإشتغال	أقل من 100 ساعة	100-200	200-300	300-400	400-500
عدد المصابيح	170	140	105	82	44

1/ أوجد قيمة الربع الثاني ومثله بيانياً

2/ إذا افترضنا أن كل الفئات متساوية، فما هو عدد القيم المحصورة في بين المجال $x - \sigma$ و $x + \sigma$ ؟

3/ أحسب معامل بيرسنل للتواء والتفرطح (التفلطح).

التمرين 22: قدم حكمان ملاحظتهما حول 8 لاعبين لفريق برشلونة لكرة القدم، علماً أن الملاحظات

مرتبة 1 إلى 8 فكانت النتائج كالتالي:

محاضرات في الإحصاء -1-

اللاعبون	الحكم 1	الحكم 2
A	6	7
B	5	6
C	4	3
D	7	5
E	1	2
F	3	1
G	8	7
H	4	4

لنفرض أنك رئيس الحكام ، هل يوجد إتفاق بين الحكمين في ملاحظتهما؟ قدم تقريراً إحصائياً محتضراً (وفق ما درست)..... لا تتجاوز 5 أسطر.

التمرين 23: أجب على الأسئلة التالية:

- 1/ أعطي مثال من عندك تستعمل فيه تعديل التكرار بطريقتين.
- 2/ ما الهدف من تقدير نموذج إنحدار خطي بسيط؟.
- 3/ عندما تكون التكرارات كلها متساوية فكيف يحدد المنوال؟.
- 4/ عندما يكون التوزيع غير متناسق وغير متزن، فكيف ندرس الشكل للتوزيع بمقاييس النزعة المركزية؟
- 5/ هل الوسيط يقسم المدرج التكراري إلى مساحتين متساويتين؟ لماذا؟.
- 6/ هل توجد طريقة لحساب طول الفئة وعدد الفئات متفق عليها؟ علل ذلك؟
- 8/ ما هي خصائص الانحراف المعياري؟.
- 9/ لماذا غالباً ما نتجه لتقدير لنماذج السلاسل الزمنية حتماً؟.

التمرين 24: يرتفع سعر التمر بولاية الوادي خلال أربع سنوات بالشكل التالي على التوالي 100، 150،

300، 400، اما سعر البطاطس فيتناقص خلال نفس الفترة بالشكل التالي على التوالي، 90، 75، 55،

35. والمطلوب منك:

- 1/ ما هي معدلات النمو من سنة لأخرى؟

محاضرات في الإحصاء -1-

2/ ما هو متوسط معدل النمو للفترة المدروسة؟ ماذا تلاحظ؟.

التمرين 25: إليك المعطيات التالية:

$$\sum_{i=1}^8 (y_i - \bar{y})^2 = 49748$$

$$\sum_{i=1}^8 (x_i - \bar{x})(y_i - \bar{y}) = 4127.5$$

$$\sum_{i=1}^8 y_i = 817 \quad , \quad \sum_{i=1}^8 x_i = 860 \quad , \quad r = 0.95$$

1/ أوجد دالة الإنحدار الخطي.

2/ أوجد الإنحراف المعياري للمعطيات.

3/ ماهي قيمة معامل التحديد.

التمرين 26: البيانات التالية تظهر تطور أسعار سلعة معينة بالدينار والكميات المطلوبة منها في إحدى

الولايات الجزائرية بمئات الأطنان.

الكمية	السعر
130	80
130	85
125	87
120	100
90	110
80	123
72	140
70	135

1/ حدد المتغير التابع والمتغير المتغير المستقل.

2/ على معلم متعامد إرسم شكل الانتشار؟ ماذا تلاحظ؟.

3/ أوجد معادلة إنحدار الكميات على الأسعار بشكل الرياضي المناسب؟ فسرهما إقتصادياً

4/ أوجد الإنحراف المعياري للأسعار والكميات؟

5/ أوجد الإنحراف المعياري للمعطيات؟

محاضرات في الإحصاء -1-

6/ أوجد معامل الارتباط الخطي وفسره؟

7/ أكتب الشكل النهائي للنموذج؟

8/ ما هي الكميات المتوقعة إستهلاكها عند الأسعار 180 دج، 50 دج

التمرين 27: قام أستاذ جامعي بدراسة العلاقة التي يمكن أن تكون بين عدد ساعات حضور الطلبة

والنتائج المتحصل عليها في الامتحان، ومن أجل ذلك أخذ عينة تقدر بـ 8 طلبة في مقياس الإحصاء

وتحصل على النتائج التالية:

نتائج الطلبة y	10	08	15	12	16	17	09	11
عدد ساعات حضور x	64	61	84	70	88	92	72	77

1/ ما هو شكل العلاقة بين عدد ساعات حضور الطلبة والنتائج المتحصل عليها في الامتحان؟.

2/ أوجد معامل الارتباط الخطي وفسره؟.

3/ أوجد معادلة خط إنحدار $f(x)$ ؟.

4/ ما هي نتائج الطلبة المتوقعة عندما تكون ساعات حضورهم تقدر بـ 90 ساعة.

التمرين 28: إليك الجدول التالي:

08	Y_3	04	02	y
02	04	06	X_1	x

1/ إذا كان $\bar{y} = \bar{x} = 5$ أو جد X_1 و Y_3

2/ أوجد معادلة الإنحدار y على x أي $f(x)$ ثم أوجد معامل الارتباط الخطي وعلق على النتيجة؟

3/ أحسب الإنحراف المعياري للمعلمة B ، ماذا تلاحظ؟.

محاضرات في الإحصاء -1-

التمرين 29: إليك التوزيع التكراري التالي:

الفئات	20-00	30-20	40-30	50-40	60-50	70-60	80-70
التكرار	10	12	16	20	12	06	04

1/ يمثل الجدول أعلاه توزيع تكراري يستعمل فئات مبوبة، فمتى نستعمل فئات غير مبوبة؟

2/ أحسب معامل الاختلاف لهذا التوزيع.

التمرين 30: البيانات التالية تظهر كل من الدخل الوطني X والواردات Y خلال الفترة الزمنية 2003-

2010.

X	10	15	20	25	30	35	40	45
Y	10	12	15	16	16	20	26	30

1/ ما هو شكل العلاقة بين الدخل الوطني والواردات؟

2/ أوجد معامل الارتباط الخطي وفسره؟.

3/ أوجد معادلة خط إنحدار الواردات على الدخل الوطني؟.

4/ أحسب الانحرافات المعيارية للمعاملات علماً أن: $\sum (x - \bar{x})^2 = 1050$ و $\sum e^2 = 25,75$

5/ أكتب الشكل النهائي للنموذج؟.

التمرين 31: فيما يلي جدولاً يبين القيمة الإيجار (الكراء) لعينة من الشقق في مدينة الجزائر العاصمة 10^3

دينار جزائري.

محاضرات في الإحصاء -1-

عدد الشقق	قيمة الإيجار
03	21-18
08	24-21
10	27-24
13	30-27
33	33-30
40	36-33
35	39-36
30	42-39
16	45-42
12	48-45
200	مج

1/ ما هو متوسط إيجار الشقة في هذه المدينة؟

2/ أحسب قيمة المدى الربيعي، ونصفه؟

3/ ما هو مبلغ الشقق الأكثر إيجاراً؟

4/ أحسب الانحراف المعياري وقارنه رياضياً مع المدى الربيعي؟

5/ أحسب معامل الاختلاف؟
