

جامعة الخرطوم
كلية التعليم عن بعد

الإحصاء في الجغرافيا

د. سمير محمد على حسن الرديسي
قسم الجغرافيا - كلية التربية
جامعة الخرطوم

المقدمة

لعلم الإحصاء أهمية قصوى في تلخيص البيانات وعرضها وبناء النماذج لتسهيل الدراسة وإجراء المسح الميداني لمجتمعات البحث وإيجاد لغة التفاهم بين العلماء من خلال اختيار أسلوب معين، وفي التنبؤ لفترات زمنية ماضية أو مستقبلية واتخاذ القرار عن طريق اختيار بديل من عدة بدائل متاحة والتحقق من صحة أو عدم صحة الفروض والرقابة على جودة السلع ومطابقتها للمواصفات والمقاييس. ونتيجة لذلك أصبح علم الإحصاء مستخدماً في شتى مناحي الحياة ، ويدرس في جميع المستويات التعليمية وخاصة الجامعات وفي مختلف التخصصات العلمية حيث يوظف لخدمتها وفق طبيعتها العلمية. ولا تنفك الجغرافيا أن تكون مثل نظيراتها من العلوم الأخرى تحتاج لعلم الإحصاء في مختلف مراحل البحث الجغرافي الطبيعي والبشري. لذلك جاء تأليف هذا الكتاب بغرض توفير أساسيات علم الإحصاء لطلاب الجغرافيا للاستفادة منها في التطبيقات الجغرافيا البحثية المختلفة. وتضمن الكتاب ثمانية أبواب ، اختص الباب الأول بتوضيح خصائص البيانات الإحصائية والجغرافية، والثاني بتبويب البيانات وطرق عرضها، والثالث بالمقاييس الوصفية في الإحصاء ، والباب الرابع بمقاييس التشتت ، والخامس بخصائص توزيع البيانات ، والسادس بالسلاسل الزمنية ، والباب السابع بالعلاقات والارتباطات.

المؤلف

د. سمير محمد على حسن الرديسي

جامعة الخرطوم - كلية التربية قسم الجغرافيا

مارس 2012

المحتويات

الصفحة	الموضوع
	الباب الأول: خصائص البيانات الإحصائية والجغرافية
	الباب الثاني: تبويب البيانات طرق عرضها
	الباب الثالث: المقاييس الوصفية في الإحصاء
	الباب الرابع: مقاييس التشتت
	الباب الخامس: خصائص توزيع البيانات
	الباب السادس: السلاسل الزمنية
	الباب السابع: العلاقات والارتباطات
	الباب الثامن: مبادئ تقدير معالم المجتمع واختبار الفروض

الباب الأول

طبيعة البيانات الإحصائية والجغرافية

مقدمة

ارتبط علم الإحصاء منذ نشأته بعمليات العد التي كانت تجريها الدول لحساب أعداد جيوشها والضرائب التي تجبى من المزارعين وجمع المعلومات عن الأراضي التي تسيطر عليها الدولة وغيرها. وقد يكون أصل كلمة الإحصاء statistics مشتق إما من اللاتينية status أو من الإيطالية statista أو من الألمانية statistik وكلها اشتقت من كلمة دولة state. وقد أشار المشهداني وهرمز (1989) لعدة تعريفات لعلم الإحصاء منها اختصاصه بإجراء التقديرات والاحتمالات، وجمع وتصنيف وتبويب الحقائق العددية كأساس لتفسير ووصف ومقارنة الظواهر، وجمع وتحليل وتفسير البيانات العددية، وأنه الطريقة العلمية التي تختص بجمع البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار. وعلى ضوء ذلك يمكن القول بأن علم الإحصاء علم يختص بجمع البيانات وتنظيمها وتلخيصها وعرضها بأسلوب علمي بهدف استخلاص النتائج لتعميمها على المجتمعات موضوع الدراسة بعد تحديد الأساليب المناسبة ومن ثم اتخاذ القرارات الملائمة.

أنواع علم الإحصاء

بحكم أن علم الإحصاء يتضمن عدة مراحل تضم أولاً المشاهدة والملاحظة، وثانياً صياغة الفرضيات بهدف تفسير الظاهرة، وثالثاً التحقق أو التأكد من صحة الفرضية والمعتمدة على تفسير الظاهرة فقد أدى ذلك لوجود نوعين من علم الإحصاء، أولهما

الإحصاء الوصفي، وثانيهما الإحصاء التحليلي. فالإحصاء الوصفي يعتمد على وصف ظاهرة ما في فترة زمنية أو مكانية معينة دون الحاجة لتعميمها على ظواهر أخرى من خلال الاهتمام بأساليب جمع البيانات وتبويبها وعرضها. ومن أهم وسائله استخدام الإشكال الهندسية للبيانات والجداول الإحصائية للبيانات والتوزيعات التكرارية والدراسة الرياضية للبيانات من خلال استخدام مقاييس النزعة المركزية و مقاييس التشتت ومقاييس الالتواء. أما الإحصاء التحليلي (الاستدلالي) فيحدد طرق الحصول على البيانات من المجتمع الإحصائي من خلال اخذ عينة من المجتمع بأساليب احصائية، وبالتالي قد لا تمثل هذه العينة المجتمع كله تماما مما يؤدي لحالة من عدم التأكد. ومن أهم وسائله اختبارات الفروض باستخدام البيانات للوصول لقرار يتم من خلاله قبول أو رفض الفروض، والتقدير من خلال إيجاد قيم تحل محل القيم الأصلية التي تمثل موضوع البحث. والتقدير إما أن يكون قيمة وحيدة ويسمى التقدير بنقطة أو فترة ويسمى تقدير بفترة.

الطريقة الإحصائية

تعني الطريقة الإحصائية في البحث العلمي بتوفير البيانات والمعلومات عن الظاهرة المطلوب دراستها، أي أمكانية التعبير عن الظاهرة تعبيراً كمياً، وتستخدم من قبل الباحثين في جميع فروع المعرفة الأخرى لأنها توفر أسلوباً علمياً موضوعياً محايداً دون تمييز أو تدخل من قبل الباحث. ويجب أن يراعي الباحث عند تصميم بحثه تحديد مشكلة البحث هدف البحث بشكل واضح ودقيق. لا بدّ أن تعبر المشكلة عن علاقة ما بين متغيرين أو أكثر وأن تصاغ المشكلة بطريقة لا تقبل الشك أو التأويل وأن تقبل

المشكلة إمكانية اختبارها التجريبي. كما على الباحث تحديد إمكانية التنفيذ الفعلي للبحث، والمتطلبات المادية والبشرية، وتحديد إطار البحث (المجتمع الإحصائي)، وتحديد أسلوب جمع البيانات والمعلومات (أسلوب التسجيل الشامل أو العينات)، وتصنيف وتبويب البيانات وعرضها، وحساب المؤثرات الإحصائية، ثم تحليل وتفسير النتائج. تعطي التحليل الإحصائية في الجغرافيا نتائج دقيقة جدا، ولكن من الضروري الانتباه إلى تحقق الآتي:-

أ - الدقة Precision

تشير الدقة إلى المستوى التفصيلي في القياس ، وترتبط بعملية موازنة جهاز القياس و تقويمه ، مثل جهاز قياس كمية المطر . فعند استعمال جهازين مختلفين في الدقة لقياس كمية المطر فان الأقل دقة قد يسجل الكمية (1.2)بوصة والآخر قد يسجل (1.26) بوصة . كذلك في العمليات الحسابية في بعض الحسابات اليدوية التي تختلف في درجة الدقة بعدد الأرقام العشرية . فكلما ازداد عدد هذه الأرقام ارتفعت نسبة الدقة . وتكون الدقة مطلوبة في الكثير من الطرائق الإحصائية عند حساب القيمة الحرجة لمقارنتها مع القيمة الجدولية .

ب - الضبط Accuracy

ترتبط عملية الضبط بانحراف عملية القياس عن النظام بصيغته الأوسع . فعندما يكون جهاز القياس دقيق جدا ، إلا انه غير مضبوط . فبالعودة إلى جهاز قياس المطر المشار إليه آنفا ، فعندما يكون تنظيم القياس غير صحيح ، فقد تكون الكمية المقاسة (1.19) بدلا من (1.26) فالنتيجة دقيقة و لكنها غير مضبوطة في درجتها . و لسوء الحظ فان

اكتشاف درجة الدقة و مستوى الضبط في الأجهزة غير سهل . وباعتماد عدد من المقاييس المختلفة في درجة دقتها و ضبطها فان النتائج تكون في النهاية غير دقيقة مالم يتم تعيير الأجهزة و توحيد درجة دقتها و ضبطها .

ج - الصلاحية Validity

في العديد من المشكلات الجغرافية يكون التوزيع المكاني أو نمط المواقع قيد التحليل ناتج عن عمليات معقدة . فعندما يكون المفهوم الجغرافي معقدا وغير واضح فان التعبير عنه يكون ضعيفا . وفي الجغرافيا ترد الكثير من المفاهيم الصعبة القياس ، متعددة الأوجه ومعقدة ، مثل : مستوى الفقر ، نوعية البيئة ، مستوى الرفاه الاقتصادي ، نوعية الحياة . و التعبير عن المعنى الحقيقي لمثل هذه المتغيرات أو المفاهيم غير ممكن . لذا يعتمد الجغرافيون تعاريف عملية تكون مقاييسها شبه مباشرة أو يتم تبنيها من بحوث و دراسات أخرى . ويبقى السؤال ، هل إن التعريف العملي (الإجرائي) صالح أم لا ؟ ومن الواضح فان درجة الصلاحية في الكثير من المشاكل التي يدرسها الجغرافيون صعب تقييمها .

د - درجة الثقة Reliability

عندما تحدث تبدلات في الأنماط المكانية عبر الزمن فان تحليلها يتطلب الإجابة عن مجموعة من التساؤلات المتعلقة باستقرارية البيانات و تبويبها من الناحية المكانية في الوحدات الإحصائية (الإدارية) . فعلى سبيل المثال ، فان اختبار النمط المكاني و التبدلات التي حصلت فيه خلال (20) سنة يتطلب معرفة التبدلات التي حصلت في الوحدات الإدارية خلال هذه المدة . تبرز هذه المشكلة بحدة أكثر عند مقارنة بيانات

إحصائية لدول مختلفة ، فالتباين ناجم عن الاختلاف في درجات الدقة و الضبط و في طبيعة البيانات و درجة الاعتماد عليها . وفي بعض الأحيان يكون الأمر كذلك في الدولة الواحدة ، بين الإقليم ، أو الأقسام الإدارية و الإحصائية المختلفة (زراعية ، صناعية ، تجارية ، عمرانية) .

المجتمع الإحصائي

يقصد بالمجتمع الإحصائي مجموعة من المفردات (أفراد , أعداد , أشياء , مقاييس) ذات خصائص مشتركة تدور الدراسة الإحصائية حولها. ويقسم المجتمع الإحصائي إلى: -
أ -مجتمع منتهي (محدد): وهو المجتمع الذي يمكن حصر مفرداته مثل (أعداد الطلبة / أعداد السلع) خلال فترة محدده.

ب -مجتمع غير منتهي (غير محدد): وهو المجتمع لا يمكن حصر مفرداته مثل (ذرات الهواء / قطرات الماء).

طبيعة البيانات الجغرافية

تهتم الجغرافيا بدراسة التنظيم المكاني لعناصر البيئة الطبيعية والبشرية ، إذ تدرس الإنسان وكل ما يؤثر عليه على سطح الأرض. ولذلك تعتبر الجغرافيا علم واسع تساهم فيه جميع العلوم الطبيعية والاجتماعية والانسانية ذات الصلة بالبيئة الطبيعية والبشرية. ولكن تمتاز الجغرافيا عن جميع هذه العلوم بقدرتها على الربط بين عناصر البيئة الطبيعية والبشرية وتفسير التنظيمات المكانية للظواهر الطبيعية والبشرية من خلال العلاقات بينها باعتبارها نظاما واحدا يكون بيئة الكرة الأرضية. ولدراسة هذه الظواهر تنتهج الجغرافيا نهجا علميا منظما ذاتيا وموضوعيا يتميز بالشمول والتعد والابتعاد عن

النظرة الأحادية الأبعاد والزوايا. ويتطلب هذا النهج العلمي تنوع وتعدد البيانات والمصادر والمقاييس المرتبطة بالمكان ، وتسمى بالبيانات الجغرافية عندما تجمع في شكل جداول بحيث تشكل قاعدة المعلومات المكانية.

يمكن تحديد طبيعة البيانات التي يستخدمها الجغرافيون في النقاط التالية:

(1) ارتباطها بالمكان ، حيث تمثل كل معلومة خاصية من خصائص المكان الذي تنتمي إليه.

(2) متنوعة المقاييس ، فقد تكون مفردة و مجدولة، متصلة ، متقطعة ، اسميه ، رتبيه.

(3) لا ينفصل المكان والزمان في المعلومة الجغرافية. فكل معلومة مكانية ترتبط بزمن تمثله وفي المقابل فان المعلومة الزمنية يجب بالضرورة أن تحدد مكانيا ليتسنى الاستفادة منها جغرافيا.

(4) متنوعة المصادر الأولية، والثانوية والثانوية متعددة المصادر، والأرضية و الفضائية. وقد تعاضمت تنوع هذه المصادر بتوافر بيانات من مصادر خارجية ، مثل المنظمات الدولية ، والاستشعار عن بعد ، وبنوك المعلومات ، وغيرها ويتطور تقنيات خزن المعلومات (ورقية ، آلية ، مساحية vector – raster خطية).

(5) أدى التراكم الكمي للمعلومات المكانية إلى زيادة الاهتمام بالبعدين الثالث والرابع في الدراسات الجغرافية (المساحة – ببعدين ، تكملها الكثافة ، الزمن).

مصادر البيانات الإحصائية

يوجد عدة مصادر تتبع في حالة جمع البيانات الإحصائية، منها:

1 - المصادر المباشرة (الميدانية أو الأولية) وتشمل كل البيانات التي يجمعها الباحث في حينها من مصادرها الأصلية بأي وسيلة كانت سواء عن طريق المراسلة أو المقابلة أو أي وسيلة أخرى خلال سنة معينة. ويعني هذا جمع البيانات عن ظاهرة ما أثناء حدوثها في ميدان العمل بحيث يمكن أن تشمل الملاحظة والملاحظة والتسجيل والاتصال الهاتفي والمقابلة الشخصية والاستبيان.

2 - مصادر غير مباشرة (تاريخية أو الثانوية) وتعني جمع البيانات من خلال سجلات سبق نشرها. وفي الغالب تشمل كل البيانات المتجمعة لدى أجهزة ومؤسسات ودوائر الدولة والمحفوظة لديها لسنوات سابقة مثل بيانات التعداد العام للسكان وإحصاءات التجارة الخارجية وغيرها.

مصادر البيانات الجغرافية

لا تتفك مصادر البيانات الجغرافية عن مصادر البيانات الإحصائية ، وكما تحتاج مصادر البيانات الإحصائية إلى التحليل ، كذلك تحتاج مصادر البيانات الجغرافية إلى تطبيق الأسلوب الكمي بغرض التحليل وتحديد الأنماط التي تشكلها، واكتشاف العلاقات التي تحتويها و العمليات التي توطرها. ويمكن تصنيفها في المصادر التالية:

أ - المصادر المكتنية

تعتبر المكتبة من أهم مصادر البيانات الجغرافية لأنها توفر المصادر التي تنظم فيها على أساس النوع الذي يضم المصدرية ، والمرجعية ، والدورية ، وغيرها. وتعتبر المكتبة نقطة البداية لأي مشروع بحث أو دراسة لأنها تمد الباحث بالمتوفر من المصادر و

المعلومات عن الموضوع المطلوب تفصيله. وقد دخلت الكثير من هذه المصادر في الشبكة الدولية للمعلومات فأصبحت قريبة جدا من الباحث.

ب - المصادر الثانوية للبيانات الجغرافية

ويمكن أن تضم البيانات المناخية والسكانية المتوفرة من الجهاز المركزي للإحصاء الذي يقوم بمسوحات عديدة عن السكان ، النشاطات الاقتصادية المختلفة ، والتخطيط الحضري والإقليمي، والتعدادات العامة للسكان، وجمع معلومات متنوعة بين حين وآخر . إضافة الى ذلك ، فان تصنيف منطقة الدراسة حسب الوحدات الإحصائية التي جمعت عنها المعلومات رسميا يساعد كثيرا في توفير قاعدة معلومات مكانية يمكن استخدامها في التحليل وفي رسم الخرائط ، وفي نظم المعلومات الجغرافية . ومن المصادر الثانوية للبيانات الجغرافية هناك المصادر التاريخية ، التي قد يحتاج لها في العديد من الدراسات الجغرافية لمعرفة تأريخ الظاهرة أو المنطقة قيد الدراسة. وقد تتوفر المعلومات التاريخية اما على شكل معلومات (بيانات) أولية او ثانوية. كما وهناك المصادر الخرائطية الرسمية التي تصدرها هيئة المساحة العامة في كل الدول.

ج - المصادر الميدانية

وتعرف أيضا باسم المصادر الأولية للبيانات ، والتي يقوم الباحث بجمعها مباشرة في الميدان تحت حيث يحدد نوعية المعلومات بما يتفق مع أهداف و حاجات بحثه. وتتنوع المصادر الميدانية في الدراسات الجغرافية بتنوع وتعدد مجالات البحث في الجغرافيا الطبيعي والبشرية. وللدراسة الميدانية أهمية خاصة في الجغرافيا ، لأنها:-

- (1) اختبار وتحليل ميداني لجزء من البلاد يسهل الوصول إليه لتوضيح واحد أو أكثر من معطيات التباين المكاني.
- (2) المختبر الحقيقي للجغرافيا خارج قاعات الدرس.
- (3) أفضل طريقة لتعلم الحقائق.
- (5) تعود الطالب على ملاحظة الأشياء وتطوير خبرة الملاحظة وتفسير ما يراه لأنها أسلوب رئيسي وأساسي لا يمكن للجغرافي الاستغناء عنه ، وتكسب الطلبة للمفردات الجغرافية اعتمادا على الملاحظة المباشرة.
- (6) لأن العمل المنجز في الحقل الميداني يشعل المخييلة ويحفزها لدراسة الجغرافيا في قاعات الدرس ويقود إلى تعظيم الأفكار الجغرافية الجوهرية ، ويطور بذلك ملكة النقد عندهم.
- (8) تطور النظرة للبيئة المحلية و البلد ، وتعود على التفكير بالمشاكل من اجل حلها.
- (9) تتطلب نوعية و قدرة عقلية مختلفة عن تلك التي تطورت من خلال التعلم من الكتب والمحاضرات.
- (10) توسع دائرة الخبرة المرئية و النجاح في استيعاب الجغرافيا اعتمادا على قدرة الطالب لتشكيل الصور الذهنية عن الأماكن.
- (13) تعمل على اتصال الطالب مباشرة مع الحقيقة والانغماس شخصا بالدراسة و امتلاك هذه المعرفة ، مما يجعله أكثر قدرة على الاتصال وأكثر تقديرا وإدراكا لعمله.
- (14) افضل طريقة لدراسة الجغرافيا.

د- نظم المحكاة

يقصد به نمذجة الواقع ، أو جزء منه بهدف الاستيعاب والدراسة والتحليل ، واستشفاف الحالات الممكنة و المتوقعة .وهي معروفة منذ القدم النمذجة في الجغرافيا ، إذ تعتبر الخريطة نمذجة للواقع باستخدام الصيغة الرمزية ، وكذلك استخدام بعض المعادلات الرياضية المعنية بتحديد طبيعة العلاقة بين المتغيرات و التنبؤ بما سيكون عليه الحال مستقبلا. وكانت تسمى النمذجة الساكنة. والتي تحولت إلى النمذجة الدينامية أو المتحركة بدخول التكنولوجيا الحديثة وبتراكم المعلومات مكانيا و زمنيا، حيث تجسد حالة النظام في فترات زمنية مختلفة . ومن أمثلة استخداماتها دراسة توقع تدفقات النقل بين المدن في العديد من دول العالم . وقد استكملت الصورة بتوافر تقنيات الحاسوب واعتماد تجسيد Simulation عناصر النظام بثلاث أبعاد أو أكثر Multi-Dimension ، وحالة التفاعل داخل النظام وكأنها حالة حقيقية Interactive mode ، مما يساعد في فهم النظام ، وفي قياس قوة تأثير العوامل الداخلية و الخارجية المؤثرة عليه، وهنا تلعب نظم المعلومات الجغرافية دورا مهما.

أنواع البيانات الإحصائية

تقسم البيانات الإحصائية إلى مجموعتين :

- 1 -البيانات النوعية: تقيس ظاهرة من الظواهر دون أن تأخذ قيما عددية. وقد نتج عن ذلك ما يعرف بالمتغيرات النوعية التي لا يمكن قياسها بوسائل قياس مألوفة وإنما تشكل صفات لذلك المتغير. ومن أمثلة ذلك لون العين كمتغير (سوداء ؛ خضراء؛ زرقاء)

والنوع كمتغير (ذكر ؛ أنثى) والحالة الاجتماعية كمتغير (أعزب؛ متزوج؛ مطلق؛ أرمل).
وتقسم البيانات النوعية إلى:

أ - بيانات نوعية أسمية: تعتمد على التصنيف النوعي بغض النظر عن أهمية الترتيب. مثلا : تصنيف موظفي إحدى الشركات حسب الجنسية او حسب التخصص.

ب - بيانات نوعية ترتيبية: يلعب الترتيب دورا أساسيا في تحديد معالم الظاهرة. مثلا : ترتيب موظفي إحدى الشركات حسب المؤهل ثانوي - دبلوم - جامعة - ماجستير - دكتوراه.

2 البيانات الكمية: تأخذ قيما عددية صحيحة أو كسرية حسب ظروف الحالة ، ويمكن قياسها بوسائل قياس مألوفة مثل عدد المرضى الراقدين في مستشفى الخرطوم العام؛ عدد رؤوس الماشية في قطيع معين ؛ درجات الطلبة في كلية ؛ أطوال الأشخاص بالسنتيمترات ؛ أوزان الأشخاص بالكيلوغرامات ؛ ودرجات الحرارة في مدينة معينة. وتقسم المتغيرات الكمية الى قسمين هما:-

أ - المتغيرات المستمرة (المتصلة) : اذا كانت مجموعة القيم التي يأخذها المتغير العشوائي (المتغير العشوائي هو دالة ذات قيمة حقيقية معرفه على فضاء العينة) مجموعة غير قابلة للعد سواء كانت محدودة او غير محدودة وانما تشكل قيم واقعة ضمن فترات. وهذا يعني وجود عدد غير منته من القيم مثل كمية الامطار المتساقطة على منطقة ما خلال سنة معينة ؛ أسعار سلعة معينة في فترة زمنية معينة وغيرها. وتعتمد البيانات الكمية المستمرة على وحدات القياس التي تأخذ قيم في مجال تغيراتها. مثلا وحدة قياس

الطول إما أن تكون بالمتراً أو السنتيمتر بفرض أن طول أحد الطالب يساوي 159 سنتيمتر وبمعنى أدق يساوي 159.4 سنتيمتراً أو أكثر دقة يساوي 159.45 سنتيمتراً. وقد مجموعة القيم التي يأخذها المتغير العشوائي مجموعة قابلة للعد أي يمكن عدّها سواء كانت مجموعة محدودة أو غير محدودة ، مثل عدد اشجار النخيل في قرية معينة ؛ أو عدد طلبة الصف الاول في مدارس بشير العبادي في أمدرمان ، وغيرها.

ب - بيانات كمية منقطعة : وهي البيانات التي تأخذ قيماً عددية صحيحة، مثل عدد موظفي الشركة خلال نصف القرن الماضي.

أنواع البيانات الجغرافية

يمكن الاستفادة من أنواع البيانات الإحصائية في تحديد أنواع البيانات الجغرافية. فالأساس الأول لتصنيف للبيانات الجغرافية ينبع من طبيعة تتابعها المكاني أو الزماني. فالبيانات الجغرافية الزمانية تكون ترتبط بمنطقة واحدة لسلسلة زمانية قد تكون متصلة أو منقطعة. وعندما ترتبط بوحدات مساحية محددة ولفترة زمنية واحدة تعتبر بيانات مكانية. ويعني هذا ضرورة ارتباط البيانات الزمانية بمكان معين، والعكس صحيح. أما التصنيف الثاني للبيانات فهو متولد عن طبيعتها ، نوعية أم كمية. تعتبر البيانات النوعية الجغرافية من أبسط المقاييس ، حيث يتم إعطاء قيمة أو عدد لواحد من مجموعتين فأكثر . ولكل فئة أو مجموعة اسم أو عنوان (المقياس الاسمي) وليس هناك علاقة افتراضية بين الفئات سوى إنها مختلفة عن بعضها البعض. ويصنف الجغرافيون المتغيرات الاسمية بطرائق عديدة ، فمثلاً الأفراد قد يتم التمييز بينهم على أساس الدين ، الجنس ، العرق ، و تصنف المدن على أساس الوظيفة التي تؤديها ، وهكذا . ويعتمد

التحليل الإحصائي لهذا المقياس على الاختبارات الإحصائية اللا معلمية non-parametric مثل : مربع كاي ، كولموكروف - سمرنوف ، معامل فاي ، معامل يول ، و معامل الجوار .

أما البيانات الكمية الجغرافية فهي البيانات فتخضع للعمليات الحسابية العادية (الجمع و الطرح و القسمة و الضرب) ، ويمكن تحويلها إلى بيانات نوعية أو ترتيبية ، وتتميز بأنها زمنية أو مكانية ، وقد تقاس بالقياسات المطلقة أو بالنسبة. تصنف البيانات الكمية الجغرافية إلى بيانات متقطعة مثل عدد المسافرين ، عدد السيارات ، وهكذا ، وإلى بيانات متصلة يمكن فيها تجزئة وحدة القياس ، مثل درجة الحرارة ، الارتفاعات و غيرها . وهناك مستوى آخر من القياس للبيانات الجغرافية ينتج من ترتيب القيم بناء على الحجم أو الوزن لتحديد أيها أكبر من ، وأيها أصغر من. والترتيب الناتجة هنا هي ليست فئات ، بل يبني الترتيب بناء على قوة أو ضعف المتغيرات. وهناك طرائق إحصائية تستخدم لترتيب القيم والمتغيرات تضم معامل سبيرمان ، و اختبار مان وتي . . ومن أمثلتها تصنيف الدول في شكل رتب بناء على معدلات الإنتاج القومي أو حصة التعليم العالي من الميزانية القومية.

إضافة لما سبق هناك تصنيف آخر للبيانات الجغرافية يرتبط بوجود / أو عدم وجود الصفر المطلق . فالبيانات المطلقة تضم مختلف أنواع البيانات الكمية التي لا يوجد لها صفر مطلق ، مثل درجة الحرارة . فمثلا عندما تكون درجة الحرارة في موقع جغرافي معين (30) درجة مئوية و في موقع آخر (15) درجة مئوية ، فإن هذا لا يعني أن الموقع

الأول درجة حرارته ضعف الموقع الثاني ، بل انه أكثر حرارة. ومن الضروري أن ينتبه الجغرافي إلى ذلك عند المقارنة والقياس وتفسير النتائج.

ويمكن القول بان الصفر في درجات الحرارة له معنى ، فهناك درجات حرارة دون الصفر (السالبة) ، كذلك الحال مع الارتفاع عن مستوى سطح البحر (الموجبة). ولكن عندما يكون الصفر هو الحد النهائي، في الإنتاج على سبيل المثال، حينها يكون القياس نسبيا . ويسمى هذا بمقياس الفاصلة حيث يحدد أصل القياس (الصفر) اعتباطيا ، مثل الصفر المئوي و الفهرنهايتي. كذلك الأمر مع المسافات و المساحات ، فالمزرعة التي مساحتها عشرة أفدنة هي ضعف مساحة مزرعة مساحتها خمسة أفدنة ، وهكذا. ويعرف هذا بمقياس النسبة ، حيث يكون الصفر حياديا (غير اعتباطي) مما يساعد في أخذ النسبة بين القيم. مع هذا النوع من البيانات تعتمد معظم إن لم يكن جميع الطرائق المعروفة بالمعلمية مثل معامل بيرسن، معامل الانحدار، اختبار (ت) ، وتحليل التباين . وفي هذين المقياسين الأخيرين يتم تحديد الفرق بين القيم على أساس الأصل أو الصفر الذي بدأ قياس القيم به .

طرق جمع البيانات

وتختلف طرق جمع البيانات حسب طبيعة الظاهرة والبيانات الإحصائية المطلوبة والهدف من الدراسة واسلوب التحليل المتبع. هناك أسلوبين لجمع البيانات. أول الأسلوبين هو أسلوب الحصر الشامل الذي يتناول دراسة كافة مفردات المجتمع الاحصائي مثل (التعداد السكاني) ، أي جمع بيانات عن جميع المفردات التي تؤلف المجتمع الإحصائي هنا يجب أن يكون المجتمع محدد. ومن مزايا هذا الأسلوب دقة

النتائج وعدم وجود أخطاء عشوائية، أما عيوبه فهي ارتفاع تكاليفها (جهد , المال)، وعدم إمكانية تطبيقها على المجتمعات ذات المفردات الكبيرة الحجم. وثاني الأساليب هو أسلوب المسح باستخدام العينة الذي يتناول جزء من المجتمع لتمثيله بطريقة العينة العشوائية بشرط أن تكون العينة ممثلة تمثيلاً صادقاً دون تحيز. ويعني هذا جمع البيانات والمعلومات عن جزء من المفردات التي تؤلف المجتمع الإحصائي. وهو أسلوب مفيد في المجتمعات الغير محدودة كما وتحتاج إلى وقت وجهد وموارد مادية وبشرية اقل مما يحتاجه أسلوب التسجيل الشامل ومن مزايا أيضا انخفاض تكاليفه وإمكانية تطبيقه مهما كان حجم المجتمع. أما عيوبه فتتمثل في نتائج التقريبية وإمكانية الخطأ نتيجة التحيز والاتساق.

تعريف العينة

هي مجموعة من مفردات المجتمع الإحصائي يتم جمعها بشكل عشوائي وتكون ممثلة للمجتمع الإحصائي ككل بهدف دراسة ظاهرة معينة للوصول إلى نتائج قابلة للتعميم وتعتبر 5 % نسبة مقبولة لحجم العينة.

أنواع العينات

تكون العينات إما عينات عشوائية (احتمالية) أو غير عشوائية (غير احتمالية). وتقسم العينات الاحتمالية إلى الآتي:

1 - العينة العشوائية البسيطة : هي عينة تختار مفرداتها بشكل عشوائي وتعطي نفس الفرصة لكل المفردات حتى نصل إل حجم العينة المطلوبة. وشرط سحبها أن يكون

المجتمع الإحصائي متجانس وكل مفردة في المجتمع لها نفس الفرصة في الظهور ضمن مفردات العينة.

2 - العينة المنتظمة : يتم اختيار المفردة الأولى عشوائيا وباقي المفردات مضاف لها (وهو ما يعرف بكسر المعاينة). ويتم حساب كسر المعاينة بقسمة حجم المجتمع / حجم العينة. ومثال لذلك إذا بلغ حجم المجتمع الإحصائي (560 مفردة) وحجم العينة المطلوبة (56 مفردة) علما بأن رقم المفردة المختارة عشوائيا (5 مفردات) فما هي قيمة الأرقام الأخرى؟ . يكون الحل كالآتي:-

كسر المعاينة = $56 / 560 = 10$. وعلى ذلك تكون بقية الأرقام كالآتي: (5 , 5 + 10) , (10 + 15)ألخ.

وشرط سحب العينة المنتظمة أن يكون المجتمع الإحصائي مرتبا ترتيبا تصاعديا أو تنازليا أو وفق أي ترتيب آخر. ويقسم المجتمع المرتب إلى L طبقة ثم يتم سحب عينة عشوائية بسيطة من الطبقة الأولى فقط ثم نضيف K لنسحب المفردة الثانية وهكذا بحيث نضمن سحب مفردة من كل طبقة.

3 - العينة الطبقيّة العشوائية : شرط سحبها أن يكون المجتمع غير متجانس بحيث نستطيع تحويله إلى مجتمع متجانس عن طريق تقسيمه إلى L طبقة بحيث نضمن مبدأ التجانس داخل كل طبقة. ويتم سحب عينة عشوائية بسيطة من كل طبقة بحيث يتناسب حجمها مع حجم الطبقة في المجتمع. ونختار العينة من خلال تقسيم المجتمع إلى مجموعات متجانسة ثم نختار اخذين بعين الاعتبار حجم الطبقة أو المجموعة. ويمكن توضيح ذلك في المثال الآتي : توجد مدرسة تضم 1000 طالب ، منهم 250 طالب ثانوي

, 500 طالب متوسط , 250 طالب ابتدائي ويراد أخذ عينة بحجم 10 % لتمثل هذه المدرسة. والمطلوب إيجاد عدد طلاب الثانوي وطلاب المتوسط وطلاب الابتدائي الواجب أخذها في العينة. ويكون الحل كالآتي:

نقوم بتحديد حجم العينة = حجم العينة نسبيا × عدد الطلبة = $100 = 1000 \times 10\%$
طالب

عدد مفردات العينة = حجم الطبقة × حجم العينة / حجم المجتمع

$$\text{طلبة الثانوي} = 250 \times 100 / 1000 = 25$$

$$\text{طلبة المتوسط} = 500 \times 100 / 1000 = 50$$

$$\text{طلبة الابتدائي} = 250 \times 100 / 1000 = 25$$

$$\text{حجم العينة المختارة} = 25 + 50 + 25 = 100$$

4 - العينة العنقودية أو المتعددة المراحل: نختار العينة من خلال تقسيم المجتمع الإحصائي إلى عدة أقسام يؤخذ منها كل قسم ويقسم إلى عدة أقسام حتى نصل لحجم العينة المطلوبة. وشرط سحبها هو أن يكون المجتمع الإحصائي كبير جدا وموزع على بقعة جغرافية واسعة كأن يكون المجتمع على مستوى البلد أو المحافظة أو المقاطعة أو إقليم. وهنا نبدأ بتقسيم المجتمع إلى وحدات أولية ثم نأخذ عينة عشوائية من هذه الوحدات كمرحلة أولى. ثم نقسم كل وحدة أولية مختارة إلى وحدات اصغر (الوحدات الثانوية) ويتم اختيار عينة من الوحدات الثانوية لكل وحدة أولية كمرحلة ثانية. ونستمر بالتقسيم والاختيار على هذا المنوال إلى أن نصل إلى عدد المفردات التي تؤلف العينة العشوائية.

(ب) - العينات الغير عشوائية: وهنا يتدخل الباحث شخصيا في اختيار مفردات العينة وبصورة متحيزة وليس على أساس عشوائي لاعتقاده بأن هذه المفردات هي خير ما تمثل المجتمع والتي ستغني الدراسة بآرائها وأطروحاتها و أهم أنواعها هي الحصصية والعمدية والعينة المقصودة وهي عينة فرضية وجدت لخدمة الباحث وتستخدم لغرض معين ولا يكون للباحث دخل أو رأي في اختيار هذه المفردة أو تلك.

أخطاء البيانات الإحصائية (والجغرافية)

- أ- خطأ عشوائي : هو خطأ غير مقصود يقل كل ماكبرت العينة.
- ب- خطأ تحيز هو خطأ في إعطاء إجابات متعمدة أو غير مقصودة. خطأ التحيز : هنا هذا الخطأ يرتكبه المصدر أو المفردة الإحصائية التي تزود الباحث بالمعلومات سواء بقصد أو بغير قصد أو يحدث هذا الخطأ أحيانا عندما يستقي الباحث معلومات بحثه ليست من مصادرها الأصلية بل من مصادرها الغير مباشرة.
- ج- خطأ الاتساق هو خطأ في وجود إجابتان توحيان إن أحدهما خطأ وهناك بعض الأخطاء التي تحدث عندما يقوم الباحث بجمع البيانات والمعلومات التي تخص بحثه :
- د- خطأ الصدفة: هذا الخطأ يرتكبه الباحث بنفسه سواء بتعمد أو بصورة غير متعمدة حيث يستقي معلومات بحثه بالاعتماد على ذاكرته بسبب بعد المفردة الإحصائية عنه أو لأي سبب شخصي آخر . هذا سيؤدي إلى الحصول على نتائج واستنتاجات غير دقيقة وبعيدة عن الواقع.

الفصل الثاني

تبويب البيانات الإحصائية وطرق عرضها

ان البيانات التي يتم جمعها حول مجتمع البحث غالبا ما تكون بيانات أولية ، غير مصنفة وغير منتظمة بحيث يتعذر على الباحث تكوين فكرة عن المشكلة قيد البحث ، وبالتالي لا يمكن إجراء التحليل الإحصائي للوصول الى نتائج البحث، مما يتطلب المراجعة والتدقيق للبيانات التي تم جمعها للتأكد من تكاملها ودقتها ووضوحها. فأولا ، يتم التأكد من دقة وتكامل البيانات التي تم جمعها الباحث ثم تأتي عملية تصنيف البيانات ثانيا، أي تشكيل مجموعات على أساس صفة معينة (ظاهرة الطول أو الوزن أو العمر أو الجنس مثلا). ثم يلي ذلك تبويب البيانات بمعنى تفريغ البيانات المصنفة وترتيبها في جداول خاصة ، يخصص كل جمع منها مستوى معين. والغرض من عملية التبويب هو إبراز وعرض البيانات في أضييق حيز يمكن الناظر إليها من تكوين فكرة عنها في وقت قصير دون عناء. وهناك أربعة أشكال لتبويب البيانات هي التبويب الزمني ؛ والجغرافي ؛ والكمي ؛ والتبويب على أساس صفة معينة (أن يكون نوع المتغير وصفا). ولكي تصبح البيانات قابلة للدراسة ، من الضرورة تجميعها وتصنيفها في مجموعات متجانسة تشترك فيها الوحدات في صفة معينة أو أكثر. ولتحقيق ذلك لابد من القيام بما يلي:

- المراجعة الأولية للبيانات: بمعنى التأكد من صحة البيانات المجمعة بمراجعتها وتدقيقها لاكتشاف أي أخطاء ومن ثم معالجتها. فمثلا إذا قام أحد الباحثين بدراسة فرق العمر ما بين الوالدين والأبناء في مجموعة من الأسر، ووجد أن عمر الأب لإحدى

العائلات 40 سنة وعمر الإبن الأكبر 33 سنة ، مما يدل على وجود خطأ بالنسبة للأعمار سواء للأب أو الابن. كما تهني المراجعة الأولية للبيانات إحلال الرموز أو الدلائل مكان الكلمات والصفات والأسماء. ومثال ذلك استبدال أسماء الكليات في الجامعة برموز معينة ، مثلاً كلية الطب بالرمز 01 وكلية الهندسة بالرمز 02.

- تفريغ البيانات الإحصائية : بمعنى تفريغ البيانات ضمن جدول يحتوى على أعمدة تأخذ الشكل التالي:

الكلية	الدليل	عدد الطلاب
الطب	01	25
الهندسة	02	33
التعليم عن بعد	03	45

عرض البيانات

يستخدم العرض الجدولي لعرض البيانات الإحصائية من خلال تفريغ البيانات في جداول تصمم حسب البيانات المجمعة. وتقسم الجداول إلى جداول مقفلة ومفتوحة وجداول منتظمة وغير منتظمة. الجداول المقفلة لها بداية ونهاية محدودة (مثلاً 3 - 6 23 - 26). أما الجداول المفتوحة إما أن يكون جدولاً مفتوحاً من أعلى 3 - 6 22 - أكثر، أو جدول مفتوح من أدنى (مثلاً أقل من 6 23 - 26)، أو جدولاً مفتوحاً الطرفين (مثلاً أقل من 6 22 فأكثر). وهناك الجداول المنتظمة (تكون أطوال فئاته متساوية 3 - 6 7 - 10 11 - 14).

وهناك الجداول غير المنتظمة (وتكون أطوال فئاته غير متساوية: 3 - 6 ، 7 - 12، 13 - 15. ويستخدم العرض الجدولي للبيانات النوعية والكمية. ومثال العرض الجدولي للبيانات نوعية البيانات التالية والمتعلقة بتوزيع موظفي إحدى الشركات حسب الجنسية (200 موظف بنغالي - 50 موظف مصري - 20 موظف سعودي - 5 موظفين أردنيين. والمطلوب : إعداد جدول لهذه البيانات واستخراج نسبة كل جنسية من مجموع موظفي الشركة. كما في الجدول أدناه:

البيان	الدول	العدد	النسبة 100 %
بنجلادش	200	72.7	
مصر	50	18.18	
السعودية	20	7.27	
الأردن	5	1.85	
المجموع	275	100 %	

ومثال للعرض الجدولي للبيانات الكمية التوزيع الطولي لعدد 50 تلميذ بالسنتيمتر لفصول إحدى المدارس كآتي : 125 - 130 - 138 - 142 - 151 - 134 - 143 - 154 - 167 - 170 - 139 - 140 - 150 - 126 - 152 - 162 - 172 - 141 - 131 - 179 - 120 - 121 - 128 - 153 - 135 - 147 -

164 – 145 – 137 – 145 – 144 – 180 – 132 – 136 – 141 – 175 –
 166 – 178 – 168 – 177 – 133 – 174 – 148 – 155 – 146 – 129 –
 . 127 – 149 – 160 – 176 –

175	166	153	147	141	135	129	120
176	167	154	148	142	136	130	121
177	168	155	149	143	137	131	125
178	170	160	150	144	138	132	126
179	172	162	151	145	139	133	127
180	174	164	152	146	140	134	128

ويتضح العرض الجدولي لهذا النوع من البيانات ترتيب هذه تصاعديا كالآتي:-
 كما يوجد العرض البياني حيث يتم تمثيل البيانات بأشكال هندسية ذات فائدة علمية
 تستخدم لغرض اعطاء فكرة واضحة وسريعة عن البيانات وتضم هذه الاشكال :-
 الاشرطة البيانية والدائرة البيانية والمستطيل البياني والخطوط البيانية والمربع التكراري
 والمضلع التكراري والمنحنى التكراري.

التوزيعات الإحصائية والتكرارات والعرض الجدولي لها:

التوزيع التكراري : هو عبارة عن بيانات المتغير العشوائي المجمعة والمصنفة الى عدد من المجاميع تسمى بالفئات (أي توزيع القيم حسب الفئات). وهذه الفئات تكون غالبا مرتبة ترتيبا تصاعديا او حسب طبيعة البيانات. ويقصد بالفئات تقسيم القيم الأصلية للظاهرة إلى مجموعات جزئية متقاربة لبعضها البعض في مدى بسيط ، وبالتالي فإن اتساع مدى الفئات قد يضيع بعض معالم التوزيع ما أن ضيق المدى قد يؤدي إلى فشل عملية التوبيخ للبيانات ، وبالتالي فإن تحديد الفئة يترك لتقدير الباحث عند القيام بإعداد الجداول التكرارية وذلك حسب طبيعة البيانات الأصلية. ويستحسن أن يتراوح عدد الفئات ما بين (5- 15) والا يزيد عن ذلك. ويفضل ان تكون اطوال الفئات متساوية في الطول لكي تسهل العمليات الحسابية. وفضل أنواع التوزيع التكراري ان يكون عدد الفئات متساوي الطول وان يكون التوزيع مغلقا ، لتسهيل العمليات الحسابية وحساب بعض مقاييس النزعة المركزية مثل الوسط الحسابي ، وبعض مقاييس التشتت مثل الانحراف المعياري والتباين وغيرها. ويرجع ذلك إلى أن حسابها يعتمد على تحديد مراكز الفئات ، والتي لايمكن تحديدها اذا كان الجدول مفتوحا. كما يجب أن تكون بداية ونهاية حدود الفئات اعداد صحيحة لتسهيل العمليات الحسابية. ولإعداد التوزيعات الإحصائية والتكرارات في الجداول لابد من إتباع الخطوات التالية:-

1 مدى الفئة = القيمة العليا - القيمة الدنيا أو يحسب مدى الفئة = القيمة العليا

- القيمة الد $3.322 + 1$ لوغ n

2 - عدد الفئات = تحدد حسب تقدير الباحث للبيانات الأصلي والتي تتراوح (5 - 15) عدد الفئات = القيمة العليا - القيمة الدنيا مدى الفئة.

3 +إشارات = تحدد حسب تكرار الأرقام

4 -الحدود الفعلية = (القيمة الدنيا - 0.5) , (القيمة العليا + 0.5)

5 -مركز الفئات = (القيمة الدنيا - 0.5) , (القيمة العليا + 0.5) 2

6 -التكرار النسبي = تكرار الفئة = مجموع التكرار

7 -التكرار المئوي = تكرار الفئة $\times 100\%$ = مجموع التكرار - مجموع قيم الفئة
= (مجموع الوحدات الإحصائية لكل فئة).

8 -إيجاد مركز الفئة = (التكرار \times مركز الفئة المجموع النظري

كيفية تكوين جدول توزيع تكراري:

لاجل تكوين جدول توزيع تكراري ننتبع الخطوات التالية:-

1- إيجاد المدى الكلي للتوزيع الذي يمثل الفرق بين اكبرقيمة واصغر قيمة مضافا لها العدد واحد

2- إيجاد عدد الفئات حسب صيغة ستور جيس $M = 1 + 3.322 \log$

3- إيجاد طول الفئة L حيث ان $L = R/M$

مثال : لدينا البيانات التالية والتي تمثل مجموع المبالغ التي حصل عليها 21 طالب أسبوعيا والتي تمثل مصروفاتهم بالريالات (12 , 8 , 10 , 15 , 11 , 12 , 14 , 17 , 16 , 22 , 32 , 25 , 18 , 30 , 18 , 10 , 9 , 17 , 14 , 19)
والمطلوب : وضع البيانات السابقة في جدول تكراري، كالاتي: -

الإشارات	التكرار	الحدود الفعلية	مراكز الفئات	التكرار النسبي	التكرار المئوي 100 %	الوحدات الإحصائية لكل فئة	مجموع قيم الفئة	إيجاد مركز الفئة
//	2	10.5 - 4.5	7.5	0.95	0.95	9+8	17	15
////////	7	15.5-9.5	12.5	0.33	0.33	11 + 10 + 12 + 14 + 12 + 14 + 10	83	87.5
////////	7	20.5-14.5	17.5	0.33	0.33	+ 17 + 15 + 18 + 16 + 17 + 18 19	120	122.5
//	2	25.5 - 19.5	22.5	0.95	0.95	21 + 22	43	45
///	3	30.5 - 24.5	27.5	0.14	0.14	+ 25 + 32 30	87	82.5
21	21	-	-	1	100		350	352.5

ويعرّف التوزيع التكراري المتجمع بأنه كمية التكرار المتجمع عند نقطة معينة من قيم المتغير العشوائي ويكون على نوعين هما:

1- التوزيع التكراري المتجمع الصاعد : وهو تراكم التكرارات ابتداء من الفئة الاولى وانتهاء بالفئة الاخيرة بالاعتماد على الحدود العليا للفئات وعبرة اقل من اقل من او يساوي.

2 التوزيع التكراري المتجمع الهابط: وهو تناقص التكرارات ابتداء من الفئة الاولى وانتهاء بالفئة الاخيرة وباعتماد على الحدود الدنيا للفئات وعبرة اكبر من او يساوي

البعد المكاني للتوزيع التكراري

بما أن البيانات الجغرافية ذات بعد مكاني، فإن تمثيلها على الخرائط يتم على أساس المساحة (وحدات احصائية ، ادارية) او بصيغة نقطية . وتعتمد مقاييس النزعة المركزية لوصف التوزيعات الجغرافية بغض النظر عن طبيعة التمثيل الخرائطي لها . إضافة الى ذلك ، فان البيانات الجغرافية قد تنظم بصيغة قيم مفردة (غير مبوبة) ، أو يتم تصنيفها الى فئات ويحسب تكرار فئاتها (قيم مبوبة) ، مما يحتم على الجغرافي تعلم استخدام مقاييس النزعة المركزية ، وغيرها في حالة البيانات المبوبة و غير المبوبة ، أو النقطية أو المساحية.

وتساعد مقاييس النزعة المركزية الجغرافي في وصف التوزيعات الجغرافية و التباين بين الأنماط المكانية ، وتعد بداية يتطلبها تطبيق الكثير من تقنيات التحليل المكاني. كما يؤدي استخدام مقاييس النزعة المركزية إلى توفير خلاصة دقيقة، سهلة الفهم ، عن خصائص مجموعة البيانات قيد التحليل في معظم المشكلات التي يعالجها الجغرافيون. ولكن عند تطبيق الاحصاء الوصفي على البيانات المكانية ، خاصة عند مقارنتها مكانيا

أو زمنيا ، لا بدّ من توخي الحذر و ذلك لوجود مؤثرات عديدة ، مثل اختلاف حدود منطقة الدراسة ، و حدود الوحدات الإحصائية للبيانات والتغيرات التي قد حصلت في حدود الوحدات الإحصائية و الادارية ، واختلاف مستويات Scales جمع البيانات . للموقع Location أهمية خاصة في الجغرافيا ، وقد تعززت هذه الأهميته لأن له معنيين ، إحصائي و مكاني. ويقصد بالمعنى الإحصائي للموقع موقع القيمة من نقطة معينة في توزيع قيم المتغير، أي أن لكل قيمة إحصائية في البيانات الجغرافية موقع مكاني مناظر له.

الدائرة المقسمة

هي عبارة عن رسم بياني دائري مقسم إلى قطاعات والتي تظهر بالتناسب حجم المجموعات المختلفة للمتغير تحت الدراسة. وتعتبر الدائرة المقسمة جيدة دا ومفيدة لعرض البيانات في شكل مجموعات.

ومثال للدوائر المقسم ، نفترض وجود 200 خريج من إحدى الكليات وقد أخذت الخلفية الاجتماعية - الاقتصادية لهم في الدراسة ووجد أن 100 منهم ذوي خلفية اقتصادية منخفضة ، 60 من المتوسطة و 40 من المرتفعة. ويمكن استخدام الدائرة القسمة لتمثيل هذه الخلفيات الثقافية المختلفة باستخدام المعادلة:

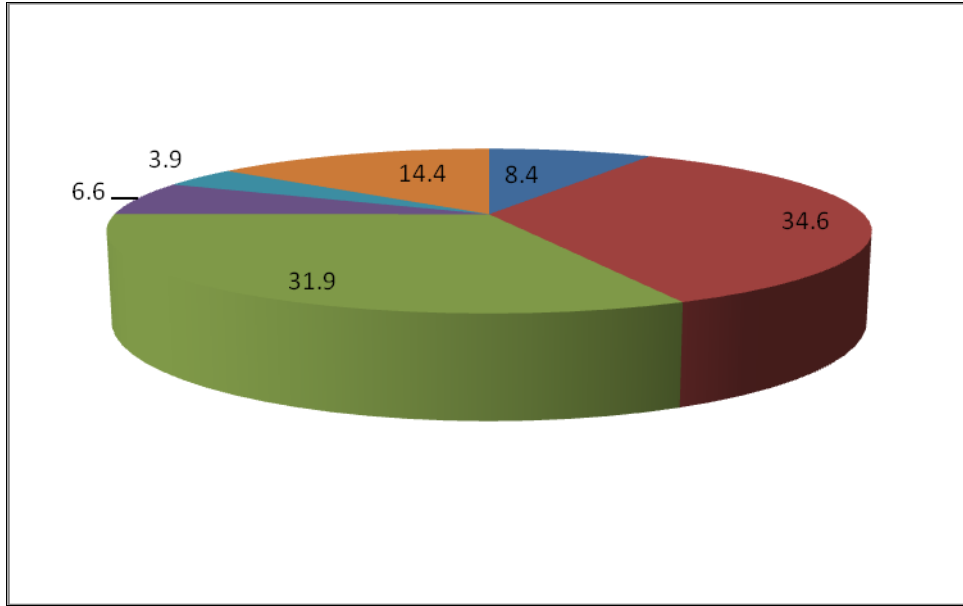
الدائرة المقسمة = تكرار كل مجموعة / مجموع عدد الحالات في كل المجموعات X

$$1/360$$

$$\text{عليه: المجموعة الأولى} = 1/360 \times 200/100 = 180$$

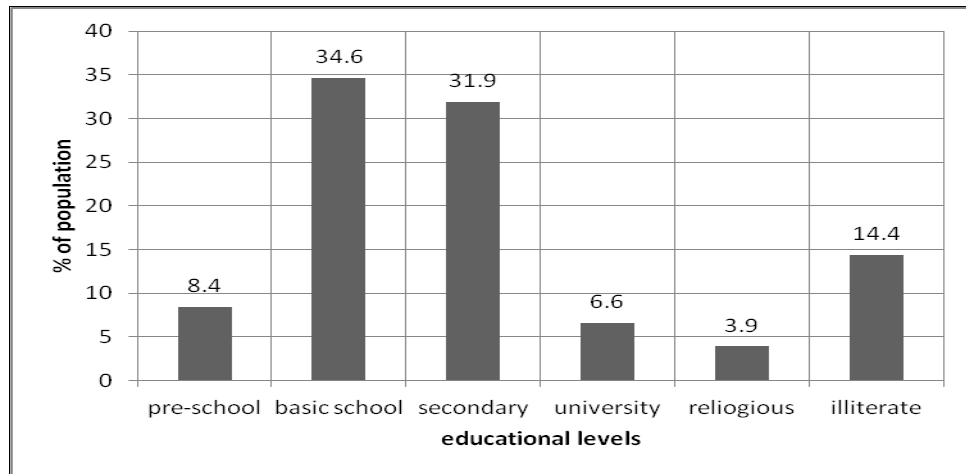
$$\text{المجموعة الثانية} = 1/360 \times 200 / 60 = 108$$

$$72 = 1/360 \times 200/40 = \text{المجموعة الثالثة}$$

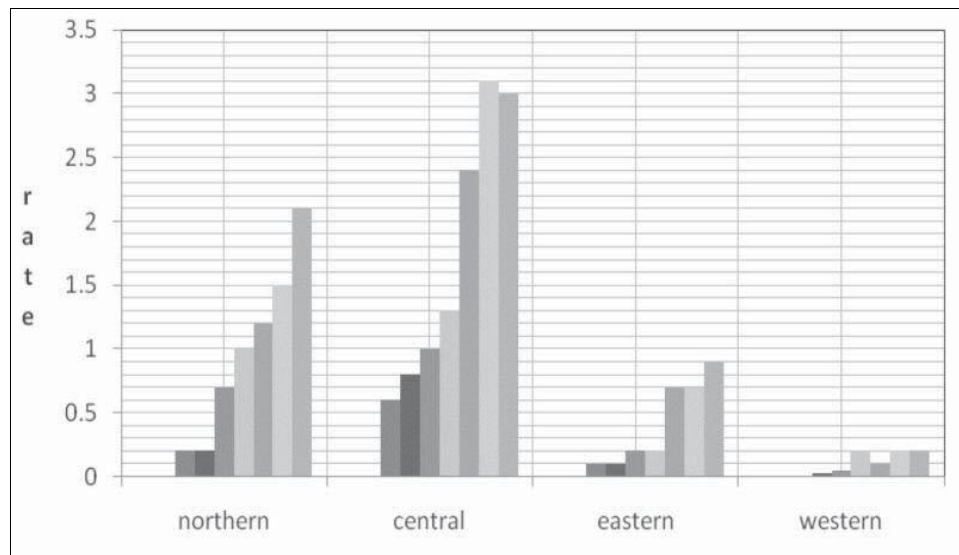


الأعمدة البيانية

هي طريقة تستخدم لعرض البيانات المصنفة على المقياسين الإسمي والترتيبي categorical والذين يظهران أحجام المجموعات للمتغير تحت الدراسة في شكل أعمدة. وفي الرسم البياني ذي الأعمدة البيانية تفصل الأعمدة في العادة عن بعضها البعض بفراغ بسيط بحكم أن المجموعات المختلفة لا تعتمد على فواصل متساوية معروفة (الشكل). وتختلف بذلك عن أعمدة الهستوگرام التي تلمس بعضها البعض (الشكل).



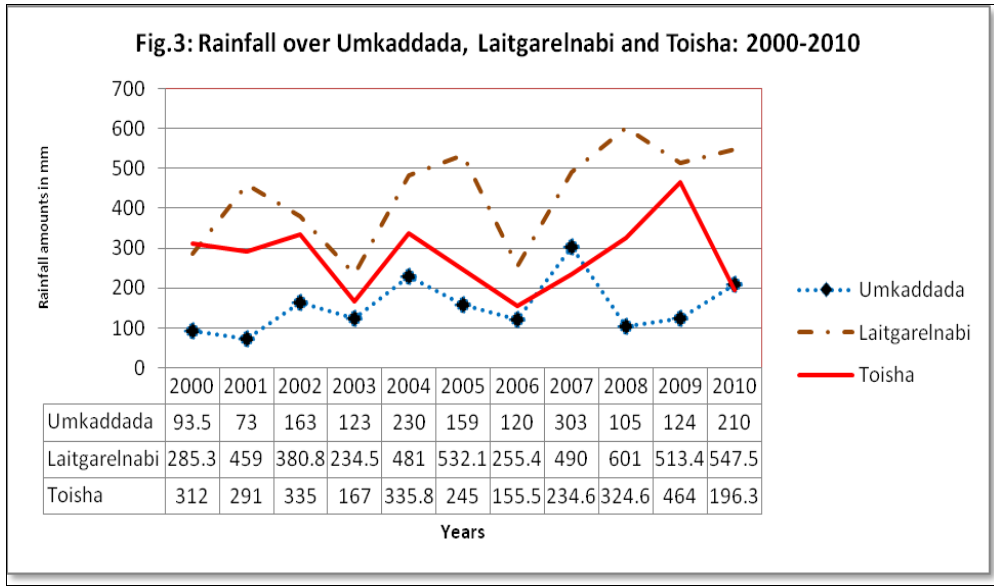
أعمدة بيانية متساوية الفواصل



أعمدة بيانية متلاصقة (الهستوگرام)

الخطوط البيانية

تستعمل لاطهار العلاقة بين الظاهرة المدروسة زمدي ارتباطها بعنصر الزمن بحيث تظهر أي نقطة على الرسم البياني حجم هذه الظاهرة خلال فترة زمنية معينة. مثال ذلك الرسم البياني الخطي الموضح في الشكل أدناه والذي يمثل كميات المطر الهائلة فوق أم كدادة ، ولعيت جار النبي وطويشة في ولاية شرق دارفور خلال الفترة من 2000 وإلى 2010.



الفصل الثالث

المقاييس الوصفية في الإحصاء

تطرق البابين السابقين إلى التعريف بالإحصاء وطبيعة البيانات الجغرافية وطرق تلخيصها في جداول تكرارية وغيرها من المواضيع ذات الصلة. ومن المعروف أن الرسوم تكون غير دقيقة ، لذلك يجب توفر بعض المقاييس العددية لتصف لنا البيانات بشكل أدق. وهذا ما سيتطرق له هذا الباب بذكر مقاييس النزعة المركزية أو مقاييس الموضع و المتوسطات، وهي مقاييس عددية تصف تعين موقع التوزيع ، وهي مهمة في حالة المقارنة بين التوزيعات المختلفة بشكل عام. وتكون فائدتها أكثر في حالة التوزيعات المتشابهة في طبيعتها وأشكالها ولكنها مختلفة في مواقعها.

يُعرّف الرقم بأنه القيمة المفردة التي تأخذ شكل المتوسط أو النسبة المئوية أو غيرها. ويعتبر الرقم في حد ذاته ذي معنى محدود، ولكن تزداد قيمته وضوحا عند مقارنته مع غيره من الأرقام . فعندما يعلن متجر ما عن تخفيض للأسعار بقيمة (100) أو (1000) جنيه ، فقد يكون هذا تلاعبا بالالفاظ وتكون هذه الأرقام غير ذات معنى إن لم تقارن بسعر السلعة ذاتها. فالرقم لا يفسر نفسه ، بل يتم ذلك من خلال أرقام أخرى ذات دلالة و معنى. ومن الأشياء الواجب معرفتها أيضا في المقاييس الوصفية أو غيرها من المقاييس الأخرى ما يعرف برمز التجميع أو سيقما (\sum). فإذا كان لدينا مجموعة من المشاهدات ($X_1, X_2, X_3, \dots, X_n$)، فإن حاصل جمع هذه المشاهدات يمكن التعبير عنه كما يلي:

$$\sum_{i=1}^n x_i = X_1 + X_2 + X_3 + \dots + X_n$$

أو باللغة العربية : $\text{مج}_1 = \text{س}_1 + \text{س}_2 + \text{س}_3 + \dots + \text{س}_n$
وفي بعض الأحيان يكتب : $\sum X$ ، أو مج س وهذا معناه جمع قيم س .

الوسط الحسابي (المتوسط)

يمكن تعريف المتوسطات بأنها القيمة النموذجية الممثلة لمجموعة من البيانات. وبحكم أن القيمة النموذجية تميل للوقوع في المركز لذلك فإنه يمكن أن تسمى المتوسطات بمقاييس النزعة المركزية. وتضم هذه المقاييس كل من الوسط الحسابي (المتوسط) والوسط المرجح، والوسيط والمنوال في حالتها البيانات الخام (غير المبوبة) والبيانات المبوبة.

يعتبر الوسط الحسابي أهم وأكثر مقاييس النزعة المركزية استخداماً في الإحصاء والحياة العملية. يعرف المتوسط بأنه حاصل جمع المشاهدات أو البيانات مقسوماً على عددها. ويرمز له بالمركز س^- ، ومعادلته هي:

$$\text{س}^- = \text{مج س}_1 + \text{س}_2 + \text{س}_3 + \dots + \text{س}_n / n = \text{الن مج}_1 = \text{س}_1$$

ومثال لذلك: إذا كانت درجات خمس طلاب في مقرر الجغرافيا الاقتصادية هي: 60،

72، 40، 80، 63

الحل:

$$\text{مج}_1 = 60 + 72 + 40 + 72 + 60 = 315$$

$$\text{إذن : س}^- = \text{الن مج}_1 = 315 / 5 = 63$$

استعمالات المتوسط

يستخدم المتوسط في وصف النزعة المركزية عند توفر الشروط الآتية :-

- 1) عندما تكون القيم موزعة بصورة متماثلة تقريبا.
- 2) عندما يتطلب البحث وصف النزعة المركزية للقيم.
- 3) عندما تكون قيم النزعة المركزية أساس لتحليل إحصائي لاحق.
- 4) عند البحث عن الصلة بين العينة و مجتمعها ، و تقدير خصائص المجتمع .

خصائص المتوسط الحسابي

- 1 -المجموع الجبري لانحرافات القيم عن وسطها الحسابي يساوي صفر .
- 2 سهل في حسابه ويخضع للعمليات الجبرية بسهولة.
- 3 يأخذ في الاعتبار جميع القيم موضع الدراسة.
- 4 يكون مجموع القيم المنحرفة عنه مساويا للصفر .
- 5 -النقطة التي تتوازن حولها مجموعة القيم، ويكون حساسا جدا للقيم المتطرفة.
- 6 يكون مجموع مربع انحراف القيم عن وسطها الحسابي هو الأقل في مجموع تربيع انحراف القيم عن أية قيمة أخرى عدا معدل المتغير نفسه.
- 7 -لا تتأثر قيمته كثيرا عند اعادة تنظيم التوزيع التكراري ، أي عند اعادة توزيع المشاهدات على فئات جديدة مغايرة في اطوالها للفئات الاصلية.
- 8 -لا يصلح الوسط الحسابي لتمثيل البيانات الاحصائية المبوبة التي تتوزع قيمها دون انتظام على الفئات المختلفة.

عيوب المتوسط الحسابي

من بعض عيوب المتوسط الحسابي:

- 1 أنه يتأثر بالقيم الشاذة (الصغيرة أو الكبيرة جدا) مقارنة بباقي القيم.

2 يصعب حسابه في حالة الجداول التكرارية المفتوحة لأنه يتطلب معرفة مركز كل فئة.

3 لا يمكن حسابه في حالة البيانات الوصفية.

الوسط الحسابي للبيانات المبوبة

يستعمل الجغرافيون في كثير من الأحيان بعض البيانات المبوبة في شكل فئات ، قد تكون عمرية ، أو مهنية ، أو تضاريسية ، وغيرها ، حيث يعتبر التعامل مع البيانات على هذا الشكل أفضل بكثير من الصيغة المفردة، وأكثر تناسبا مع هدف الدراسة و منهجها الدراسة . وتشترك الجغرافيا مع غيرها من العلوم في هذه الطريقة التي تعتمد التصنيف و التبويب أساسا في التحليل و الدراسة. لقد أوجد الإحصائيون أكثر من طريقة لمعالجة البيانات المبوبة، وميزوا بين : (1) البيانات المبوبة ذات الفئات المتساوية المدى ، و (2) والبيانات المبوبة ذات الفئات المفتوحة النهاية. ويتضمن حساب المتوسط الحسابي في كلتا الحالتين وجود بيانات في شكل فئات يقابلها تكرارات. ومثال لذلك إذا كان لدينا عدد ك من الفئات ذات المراكز (س₁ ، س₂ ، ، س_ك) على الترتيب، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\bar{س} = \frac{س_1 ت_1 + س_2 ت_2 + س_3 ت_3 + + س_ك ت_ك}{س_1 ت_1 + س_2 ت_2 + س_3 ت_3 + + س_ك ت_ك} = \frac{س_1 ت_1 + س_2 ت_2 + س_3 ت_3 + + س_ك ت_ك}{س_1 ت_1 + س_2 ت_2 + س_3 ت_3 + + س_ك ت_ك}$$

المثال التالي:

أحسب متوسط أعمار الطلاب للبيانات التالية:

فئات العمر	6- 5	8 - 7	10- 9	12-11	14-13
عدد الطلاب	2	5	8	4	1

الفئات	مراكز (س)	الفئات	التكرارات (ت)	س ت
6-5	5.5	2	11	
8-7	7.5	5	37.5	
10-9	9.5	8	76	
12-11	11.5	4	46	
14-13	13.5	1	13.5	
المجموع			184	

$$س^- = \frac{1}{n} \sum_{i=1}^n x_i = \frac{20}{184} = 9.2 \text{ سنة}$$

معدل قيم مجدولة متساوية المدى

لحساب قيمة المتوسط للبيانات المبوبة المتساوية المدى يمكن استخدام إحدى الطرق الثلاثة الآتية:

الطريقة الأولى

ويمكن تطبيقها باتباع الخطوات التالية (أنظر الجدول أدناه):

1) إيجاد مركز الفئة (العمود الثالث في الجدول).

(2) ضرب مركز الفئة بالتكرار المقابل لها (العمود 4).

(3) حساب مجموع حصل الضرب.

(4) تقسيم ناتج الخطوة (3) اعلاه على مجموع التكرارات

المتوسط الحسابي = مجموع (مركز الفئة X تكرارها) / مجموع التكرار

$$19.5 = 12 / 234 =$$

8	7	6	5	4	3	2	1
ت X	انحراف	ت X	انحراف	ت X	مركز الفئة م ف	تكرار	الفئة العمرية
أ	أ	ح	(ج)	م ف			
3-	3-	10-	10-	7	7	1	9 - 5
6-	2-	15-	5-	36	12	3	14 - 10
2-	1-	0	0	34	17	2	19 - 15
0	0	15	5	66	22	3	24 - 20
1	1	10	1	27	27	1	29 - 25
4	2	30	15	64	32	2	34 - 30
6-		30		234		12	المجموع

جدول يوضح معدل قيم مجدولة متساوية المدى

الطريقة الثانية

وتتبع فيها الخطوات الآتية (أنظر الجدول اعلاه):-

- 1 تحديد مراكز الفئات (العمود 3)
- 2 اختيار احد هذه المراكز ليكون وسطا فرضيا (وليكن مثلا 10- 14 والذي يقابله 17)
- 3 -ايجاد انحراف مراكز الفئات عن الوسط الفرضي (العمود 4)، أي طرح مراكز الفئات عن الوسط الفرضي (مثلا طرح 17 عن 12)
- 4 -ضرب التكرارات في الانحرافات المناظرة لها (العمود 5).
- 5 -حساب مجموع الفقرة (4) أعلاه وتقسيمه على مجموع التكرارات لينتج المتوسط الحسابي للانحرافات
- 6 -إضافة المتوسط الحسابي للانحرافات الى المتوسط الفرضي لينتج المتوسط الحقيقي .

$$\begin{aligned} \text{المتوسط الحسابي} &= \text{مركز الفئة الوسطى} + (\text{مجموع انحراف مركز الفئة عن مركز} \\ &\text{الفئة الوسطى} \times \text{تكرارها}) / \text{مجموع التكرارات} \\ 19.5 &= 2.5 + 17 = (12 \setminus 30) + 17 = \end{aligned}$$

الطريقة الثالثة

ويتبع فيها الخطوات التالية (أنظر الجدول السابق) :-

- 1) اختيار إحدى الفئات لتكون نقطة بداية، حيث يعد مركزها وسطا فرضيا ، ومن الأفضل أن تكون الفئة في وسط التوزيع لتسهيل العمليات الحسابية.

(2) تحديد الانحرافات الترتيبية للفئات عن الفئة الوسطية (وهي الفئة التي تحتوي على الوسيط)، (أنظر العمود 7)

(3) ضرب الانحرافات الترتيبية بالتكرارات المناظرة لها (العمود 8)

(4) ايجاد مجموع حاصل عمليات الضرب في الخطوة السابقة ، وحساب متوسطها الحسابي.

(5) ولما كانت الانحرافات الترتيبية تقل عن الانحرافات الاصلية بنسبة طول الفئة ، لذا يضرب ناتج الخطوة السابقة في طول الفئة

(6) يضاف ناتج الخطوة الاخيرة الى المتوسط الفرضي لينتج المتوسط الحقيقي

المتوسط الحسابي = مركز الفئة الوسيطة + طول الفئة X (مجموع انحراف رتبة

الفئة عن رتبة الفئة الفرضية X تكرار الفئة) \ مجموع التكرارات (

$$19.5 = (2.5-) + 22 = (0.5-) * 5 + 22 = (12 \setminus 6-) * 5 + 22 =$$

وكما توضحه المعادلة :

$$\bar{x} \approx x_0 + c * \frac{\sum fd}{n}$$

معدل بيانات مجدولة مفتوحة النهاية

يتعذر اتباع الطرق السابقة في حساب المتوسط اذا كان الجدول التكراري يضم فئات مفتوحة النهاية بسبب عدم امكانية تحديد مراكز مثل هذه الفئات . وقد يكمن الحل في تحديد الحدود الدنيا و العليا افئات فرضيا حسب موقعها في التوزيع وبما يتناسب مع طبيعة الموضوع قيد التحليل أوبناءا على خبرة سابقة . وفي جميع هذه تعتبر قيمة

المتوسط قيمة تقديرية وليست دقيقة تماما . ويمكن اعتماد الوسط الحسابي للفئات الاخرى كمقياس للفئات المفتوحة . وبحكم أن النتيجة تعتبر تقريبية ، من الأفضل اعتماد مقاييس النزعة المركزية الاخرى التي لا تتأثر بمثل هذه الحالة .

الوسط الهندسي

الوسط الهندسي لمجموعة من القيم : s_1 ، s_2 ، s_3 ، s_n يحسب بالمعادلة:

$$\text{الوسط الهندسي} = \sqrt[n]{s_1 \cdot s_2 \cdot s_3 \cdot \dots \cdot s_n}$$

ومثال لذلك أوجد الوسط الهندسي للأعداد: 2، 3، 5، 10،

الحل :

$$\text{الوسط الهندسي} = \sqrt[4]{2 \cdot 3 \cdot 5 \cdot 10} = 300$$

الوسط الهندسي للبيانات المبوية (المبوبة)

يكون بتطبيق المعادلة التالية:

$$\text{الوسط الهندسي} = \sqrt[n]{t_1 \cdot s_1 \cdot t_2 \cdot s_2 \cdot \dots \cdot t_n \cdot s_n}$$

$n = \text{مج } t$

الوسط التوافقي

$$\text{الوسط التوافقي} = n / \text{مج } \frac{1}{s_1}$$

مثال أوجد الوسط التوافقي للبيانات التالية: 2، 5، 7، 9،

الحل:

$$n = 4$$

$$\text{مج } \frac{1}{s} = \frac{1}{s_1} + \frac{1}{s_2} + \frac{1}{s_3} + \frac{1}{s_4}$$

$$9/1 + 7/1 + 5/1 + 1/2 = 4 = \text{مج } 1/س = 9/1 + 7/1 + 5/1 + 1/2 =$$

الوسط التوافقي للبيانات المبوبة

$$\text{الوسط التوافقي} = 1/ن (ت_1 / س_1 + ت_2 / س_2 + \dots + ت_n / س_n) = 1/ن \text{ مج } ت / س$$

مثال : أوجد الوسط التوافقي للبيانات المبوبة التالية:

15-13	12-10	9-7	6-4	3-1	الفئات
2	5	8	3	2	التكرارات

الحل

الفئات	مراكز الفئات (س)	التكرار (ت)	1/س	ت/س
3-1	2	2	0.5	1
6-4	5	3	0.2	0.6
9-7	8	8	0.125	1.0
12-10	11	5	0.1	0.5
15-13	12	2	.0	0.2
المجموع				3.1

$$\text{الوسط التوافقي} = 1/ن \text{ مج } ت / س = 5/3.1 = 1.6129$$

المتوسط الموزون :

عندما تختلف قيم المتغير عن بعضها ليس في قيمها العددية وفي أوزانها وأهميتها ، فإن معدل قيم المتغير (العددية) لا يعكس أهمية القيم الحقيقية ، لذلك يتطلب الأمر حساب معدل أهمية القيم ، والذي يعرف باسم المتوسط الموزون ، أو الوسط الحسابي المرجح، والذي يعتمد مع القيم المكررة اختصاراً لعملية الجمع ، و طالما لم توزع القيم الى فئات. وتحسب قيمة هذا المتوسط بضرب كل قيمة بما يقابلها من "وزن" ، أو تكرار ، ثم جمع حاصل الضرب الناتج ، ثم يقسم الناتج على مجموع التكرارات (الأوزان) ، طبقاً للمعادلة الآتية :-

$$\bar{x} = \frac{\sum wx}{\sum w}$$

وعندما تتكرر بعض القيم ، يفضل حساب تكرار كل قيمة ، وضربها بها ، وحساب مجموع حاصل الضرب ، ثم قسمته على مجموع عدد القيم (التكرارات) ، حسب الصيغة التالية: المعدل = مجموع حاصل ضرب القيمة في تكرارها / مجموع التكرارات

$$8,0 = 29 / 232 =$$

FX	التكرار f	القيمة X	FX	التكرار f	القيمة X
22	2	11	12	1	12
36	4	9	50	5	10
28	4	7	48	6	8
10	2	5	18	3	6
232	29	مجموع	8	2	4

كذلك تستخدم المتوسطات الموزونة لتحليل نتائج الاستبيانات. ويمكن أن نورد المثال الآتي والخاص باستطلاع أجري وسط طلبة إحدى الجامعات عن الوقت المستثمر لأغراض الدراسة الجامعية اسبوعيا وحسب الكليات. وقد خرج الاستطلاع بالنتيجة الآتية:

الكلية	عدد الطلبة	العينة	معدل الوقت
الآداب	245	36	42,1
العلوم	250	50	40,5
العلوم الاجتماعية	69	23	37.0
الكلية	564	109	40.8

$$\frac{(37.0 \times 69) + (40.5 \times 250) + (42.1 \times 245)}{69 + 250} = \text{المعدل}$$

$$564 \div (2553 + 10125 + 10314.5) =$$

$$564 \div 22992.5 =$$

$$40.7668 = \text{ساعة اسبوعيا}$$

وهذا هو المعدل لطلبة الجامعة وليس للعينة فقط ، بينما يكون معدل العينة هو :

$$39.866 = 3 \div (37.0 + 40.5 + 42.1) \text{ ساعة اسبوعيا للعينة .}$$

متوسط النسب Average Rating

يستعمل هذا النوع من المتوسط مع الاستبيانات التي تكون أجابة أسئلتها محددة بخيارات متباينة في قيمتها وأهميتها للبحث . فمثلاً ، في دراسة ما استطلع رأي خمسين مديراً لمؤسسات تعليمية ، و ترك الباب مفتوحاً لخمس خيارات للإجابة عن سؤال معين، والمهم هنا هو كيف سيتم حساب معدل نسب الاجابات ؟. ولتوضيح ذلك نعرض الجدول أدناه:

الهدف	مهم بالتأكيد 1	مهم 2	محايد 3	غير مهم 4	غير مهم بالتأكيد 5
3التحصيل العلمي	26	20	4	0	0
الابداع	13	2	5	20	10
المواطنة	11	24	10	5	0
تطور الشخصية	6	1	1	30	12

ولحساب معدل نسبة إجابة هؤلاء المدراء عن حالة الابداع مثلا ، تضرب الإجابة الخاصة بالابداع في الوزن الذي قدره الباحث للإجابة (الرقم في السطر الأول)، بحيث نتحصل على معدل النسبة كآلاتي:

$$\text{معدل النسبة} = (5 \times 10) + (4 \times 20) + (3 \times 5) + (2 \times 2) + (1 \times 13) = 50 / (50 + 80 + 15 + 4 + 13) = 3,24$$

وكلما كانت قيمة معدل النسبة قريبة من (1) كلما كانت أكثر أهمية ، وقد يكون العكس صحيحا إذا تم حساب طريقة تحديد أوزان كل إجابة ، كما نجد في المثال السابق أن معدل النسب كان كآلاتي : التحصيل العلمي (1,56) ، المواطنة (2,18) ، الابداع (3,24) ، وتطور الشخصية (3,82).

بعض التطبيقات الجغرافية للمتوسط

(1)المركز المتوسط ، The mean centre

يعرف المركز المتوسط للتوزيع النقطي بأنه نقطة ذات بعدين تحدد موقع معدل جميع النقاط على هذين المحورين (البعدين) ، أي انه النقطة التي يلتقي عندها معدلي المحورين الرأسي والأفقي لتمثل مركز معدل التوزيعات المكانية. وهو أبسط طرق قياس التوزيعات المكانية النقطية ، ويناظر معدل قيم مجموعة البيانات الرقمية المفردة حيث تستخدم نفس الطريقة لحسابها، والتي سبق ذكرها. ومعروف أن النقط في الخريطة ترمز إلى ظواهر جغرافية مختلفة قد تكون منطقة استقرار بشري أو أي من المرافق الخدمية أو الصناعية. وتتضمن خطوات حساب المركز المتوسط الآتي:

1 - رسم شبكة مربعات تغطي منطقة الدراسة حيث تقاس مواقع النقاط طبقا لمحوريها السيني و الصادي (x , y) . وقد تكون شبكة المربعات هذه اعتباطية ، او خطوط الطول والعرض نفسها . المهم انها تقيس مواقع النقاط بالاتجاهين الشمالي و الغربي (أو الجنوبي والشرقي) . وحتى بداية الشبكة هي الاخرى اعتباطية ، ولكن دون تجاوز شروط تعامد خطوط الشبكة على بعضها ، بزاوية قدرها (90) درجة ، وتوحيد قياسات المحورين السيني والصادي.

2 - تحديد موقع أية نقطة في التوزيع الجغرافي من خلال المحورين الأفقي والرأسي وذلك بقياس المسافة التي تفصل النقطة افقيا ورأسيا عن نقطة محددة حسب نظرية فيثاغورس، كالآتي:

$$C = \sqrt{(A_1 - A_2)^2 + (B_1 - B_2)^2}$$

ومثال لذلك إذا أراد أحد الجغرافيين وصف توزيع ثمان مستقرات بشرية في اقليم معين ، وبعد أن رسم شبكة المربعات ووجد أن مواقعها كما مبين في الجدول أدناه :

محور ص محور س	المستقرة س	محور ص	المستقرة	
3.2	1.8	ب	4.0	أ 1.7
3.2	3.7	د	3.0	ج 2.7
2.1	3.6	و	2.2	هـ 2.3
1.7	3.2	ح	1.0	ز 2.9
8	21.4	20.4	المجموع	

معدل س = $21.4 \div 8 = 2.67$

معدل المحور ص = $20.4 \div 8 = 2.55$

أي ان موقع مركز المتوسط يقع عند تلاقي هذين المحورين ، بإسقاط متوسط (س) على شبكة المربعات بدلالة متوسط (ص) يمكن تحديد المركز المتوسط .

يمكن تشبيه المركز المتوسط وكأنه مركز جذب للتوزيعات المكانية النقطية ، فعند تمثيل كل نقطة بقطعة نقود معدنية موزعة على ورق سميك (مقوى ، كارتون) و المطلوب وضع الورقة على حامل بحيث تكون متوازنة عند نقطة ارتكاز ، فمركز المتوسط هو هذه النقطة . يفيد هذا التشبيه عند تحديد موقع مركز يقدم خدماته لمجموعة من النقاط (قرى مثلاً) على صفحة اقليم معين.

(2) مركز معدل نقاط متباينة في الحجم والأهمية :

عند حساب المركز المتوسط هنا ، تعطى النقاط وزنا متساويا. ولكن هناك حالات لا تتساوي فيه ما تمثله النقاط من حيث الحجم أو الأهمية والتاثير كما هو الحال في أماكن الاستقرار البشري التي تختلف في الحجم وفي الانتاج.

يمكن حساب المركز المتوسط للإنتاج عندما يكون عدد الوحدات الانتاجية للمصانع معروفا، حيث يعطى كل مصنع وزنا مكافئا لكمية انتاجه ، وبالتالي إذا كان إنتاج أحد المصانع ضعف الآخر فإن تأثيره سيكون مضاعفا ايضا في تحديد موقع مركز الانتاج او مركز الجذب ، والذي قد يعرف بالمركز المتوسط الموزون.

ومثال لذلك إذا أراد مدير إحدى المؤسسات الانتاجية تحديد موقع المخازن المبردة لمجموعة مصانع تابعة لمؤسسه ، فعليه اختيار موقع يتناسب مع التباين في انتاجية المصانع، كما هو موضح في الجدول أدناه، والذي يوضح مواقع و أنتاجية المصانع قيد الدراسة:

المعمل	موقع س	موقع ص	الوزن	و س	و ص
أ	2	5	8	16	40
ب	1	4	5	5	20
ج	3	2	10	30	20
د	4	1	42	168	42
هـ	5	1	20	100	20
المجموع		85	319	142	

عليه:

$$1 \text{ يكون معدل س الموزون} = \text{مجموع (و X س)} / \text{مجموع الوزن} = 319 \setminus 85 = 3.75$$

$$2 \text{ معدل ص الموزون} = \text{مجموع (و X ص)} / \text{مجموع الوزن} = 142 \setminus 85 = 1.67$$

3 إذن الموقع المقترح للمخازن ، بما يتناسب مع مواقع و انتاجية المعامل ، هو عند التقاء المحور السيني ذي النقطة (3.75) مع المحور الصادي في النقطة (1.67) .

4 -ويحسب وزن الموقع بضرب الوزن في كل من قيمتيه السينية والصادية . فالمعمل (د) ذي طاقة انتاجية قدرها (42) ، لذا كان وزن موقعه على المحور السيني (4 * 42 = 168) و وزن موقعه على المحور الصادي (1 * 42 = 42) ، وهكذا .

5 وقد يقاس الوزن بنسبة الانتاج السنوي للمعمل من مجموع انتاجية المؤسسة او منطقة الدراسة ، ولكل نقطة وزن يتناسب مع هذه النسبة . وقد يكون عدد

العمال ، او الطاقة الانتاجية ، الانتاج الفعلي ، رأس المال هو المعيار الوزني ،
او حسب هدف البحث و موضوعه .

6 - وعند العودة الى تشبيه ابدن للمواقع واستبدالها بقطع النقود المعدنية ، فان وزن
الموقع يتحدد بعدد القطع المعدنية فيه ، وبهذا فان موقع مركز المتوسط
سيختلف عن موقع مركز المتوسط الموزون ، لذا يسمى بمركز الجذب ، والفرق
بين الاثنين يؤثر مناطق الجذب (السكاني ، الاقتصادي ، مثلاً) . والمقارنة بين
المركزين (المتوسط و الموزون) تؤثر الكثير من التباينات المكانية التي قد
تخفيها الخرائط التقليدية و التحليل غير المكاني .

(3)المركز المتوسط لنقاط مجدولة (مبوبة)

تلخص مقاييس النزعة المركزية التوزيعات وتنظم المقارنة بعيدا عن الذاتية (أي
بموضوعية) ، كما وتشير إلى الاختلافات غير المنظورة في الخرائط الخاصة
باستخدامات الأرض. وفي الأحوال التي يكون فيها عدد النقاط كبيرا ، يفضل معالجة
البيانات بصيغة المجاميع وذلك باستخدام المحورين السيني و الصادي و المربعات التي
شكلتها كحدود للفئات . ويعني هذا حساب عدد النقاط الواقعة في كل مربع على
المحورين الافقي والعمودي وليس كنقاط منفردة. وكما تم سابقا في اشتقاق معدل القيم
المبوبة ، تتبع الخطوات التالية لإيجاد المركز المتوسط للنقاط المبوبة:

- 1 (تقدير القيمة المحتملة للمعدل ، والتي تقع في الوسط في الغالب ، ثم عدّها لتمثل
المرتبة (0) ، وتكون الفئات التي تقع قبلها في السالب والتي تليها في الموجب.
- 2) حساب الفرق في المرتبة بين الفئات والفئة التي يقع فيها المتوسط الفرضي.

3) ضرب الفرق بين مرتبتي الفئة وفئة المتوسط بعدد النقاط في الفئة مع الانتباه الى الاشارة السالبة و الموجبة ، واستخراج المجموع.

4) إضافة او إنقاص النتيجة من مركز الفئة التي يعتقد بان المتوسط يقع فيها .
وبحكم أن المتوسط تقدر قيمته هنا ، لذلك يكون مختلفا عن معدل القيم غيرالمبوية.
ومثال لذلك إذا رغب أحد الجغرافيين في دراسة التنظيم المكاني لمحلات بيع الأواني المنزلية ومقارنته مع محلات بيع الخضر والفاكهة في سوق الخرطوم الكبير ، حيث كان عدد محلات بيع الأواني المنزلية (40) ، وتبين بعد إسقاط شبكة المربعات على خارطة سوق الخرطوم الكبير أنها تتوزع وكما مبين في الجدول أدناه :

الفئة	1	2	3	4	5	مجموع
1			2	5		7
2	4		5	5		14
3		5	2	4	1	12
4				1	1	12
5					5	5
مجموع	4	5	9	15	7	40

وبحكم أن :

1 مركز الفئة التي يقع بها معدل المحور السيني (3.5) ،

2 ومركز الفئة التي فيها معدل المحور الصادي (2.5) ،

لذلك :-

- 1 -تقدر قيمة معدل المحور السيني = مركز الفئة + (مجموع الفروقات مضروبة بالتكرار / مجموع النقاط) $2.9 = 0.6 - 3.5 = (40 \setminus 24-) + 3.5$
- 2 -وتقدر قيمة معدل المحور الصادي $2.1 = (40 \setminus 16-) + 2.5$
- 3 -يكون المركز المتوسط في النقطة (2.9) على المحور السيني حيث يلتقى فيها مع النقطة (2.1) على المحور الصادي، كما كوضح أدناه:

باتجاه الشمال			باتجاه الغرب			الفئة
fd	d	fy	fd	d	fx	
14-	2-	7	12-	3-	4	0.9 – 0
14-	1-	14	10-	2-	5	1.9 – 1
0000	0	12	9-	1-	9	2.9 – 2
2+	1+	2	0000	0	15	3.9 – 3
10+	2+	5	7+	1+	7	4.9 – 4
16-			40	24-		40
						مجموع

وتشير (f) تمثل للتكرار ، أي عدد النقاط في المربع ، (d) الفرق بين مرتبة المتوسط و الفئة الاخرى ، (fd) ضرب الفرق بالتكرار .

الوسيط The Median

هو القيمة التي تتوسط مجموعة قيم المتغير بعد ترتيبها ترتيبا تصاعديا أو تنازليا، بمعنى أنها القيمة التي تقسم عدد القيم الى نصفين متساويين بحيث يكون عدد القيم الأعلى مساويا للأدنى منها (50% من البيانات فوقها و50% من البيانات بعدها في الترتيب).

وإذا كان عدد البيانات زوجيا، فإن الوسيط هو متوسط المشاهدين اللتين تقعان في المنتصف. ونتيجة لاختلاف مجموعة القيم في تنظيمها وتكرارها ، فقد تعددت طرق تحديد القيمة الوسطى فيها.

وتتمثل أهم استعمالات الوسيط:

- 1) في البحث لتحديد النقطة الوسيطة في التوزيع ،
 - 2) لمعرفة كيفية تأثير القيم المتطرفة على قيمة المتوسط ،
 - 3) وعند السعي لمعالجة التوزيعات المتطرفة للقيم،
 - 4) أنسب الطرف مع البيانات المصنفة على المقياس الترتيبي.
- وللوسيط عدّة مميزات ، هي:

- 1 -لا يتأثر بالقيم المتطرفة ،
- 2 ويمكن حسابه في حالة الجداول التكرارية المفتوحة للبيانات الكمية ،
- 3 ويمكن إيجاده في حالة البيانات الوصفية التي يمكن ترتيبها ،
- 4 وأن مجموع الانحرافات المطلقة عن الوسيط أقل ما يمكن مقارنة بأي قيمة حقيقية.

أما عيوب الوسيط فتشمل:

- 1 -أنه لا يأخذ جميع القيم في الاعتبار عند حسابه ،
- 2 ولا يسهل التعامل معه في التحاليل الإحصائية والرياضية.

ويمكن وجه الاختلاف بين الوسيط والوسط الحسابي (المتوسط) في أن الوسيط لا يحسب كل قيمة مفردة في مجموعة البيانات ، لذا لا تؤثر عليه القيم المتطرفة ، كما يمكن استخراج الوسيط بطريقة الرسم البياني.

لإيجاد الوسيط لقيم مفردة غير مبنوية

لإيجاد الوسيط لقيم مفردة غير مبنوية، تتبع الخطوات التالية :-

- (1) ترتيب القيم ترتيبا تصاعديا أو تنازليا،
 - (2) عندما يكون عدد القيم فرديا فان القيمة الوسطى هي قيمة الوسيط ، (مثلا: 3 ، 4 ، 5 ، 6 ، 7)، تكون القيمة (5) هي الوسيط = $(1 + 2) / 2$
 - (3) أما عندما يكون عدد القيم زوجيا فان القيمة الوسطى تحسب من جمع قيمتي العددين في وسط القيم واستخراج معدلها ، (مثلا: 3 ، 4 ، 5 ، 6 ، 7 ، 8)، يكون الوسيط $((5 + 6) / 2 = 5,5)$
- مثال (1) : أوجد الوسيط لدرجات خمس طلاب في مقرر الجغرافيا الاقتصادية ، والتي كانت كما يلي: 60، 72، 40، 80، 63

الحل :

- 1 يتم ترتيب البيانات تصاعديا كالتالي: 40 ، 60 ، 63 ، 72 ، 80
- 2 وبما أن عدد المشاهدات فردي ، يكون الوسيط هو المشاهدة التي تقع في منتصف هذه البيانات.
- 3 وعليه فإن الوسيط هو المشاهدة رقم (3) وقيمتها 63
- 4 ويلاحظ وقوع بيانين يقعان في الترتيب قبل الوسيط وبيانين يقعان بعده.

مثال (2) : أوجد الوسيط لدرجات الطلاب الآتية: 63 ، 80 ، 40 ، 72 ، 60 ، 72
الحل:

- 1 يتم ترتيب البيانات تصاعديا كالتالي: 40 ، 60 ، 63 ، 72 ، 72 ، 80
- 2 وبما أن البيانات زوجي فنكون قيمة الوسيط هي متوسط قيمتي المشاهدين اللتين يقعان في المنتصف.

$$3 \text{ الوسيط} = (72+63) / 2 = 67.5$$

إيجاد وسيط لقيم مفردة مكررة

وهنا تبرز مشكلة بسيطة عندما تكون القيمة الوسيطة مكررة أكثر من مرة واحدة، حيث تعالج كالاتي:

- 1 يفترض أنها ، وعلى اختلاف عددها ، تحتل موقعا واحدا ،
 - 2 وأن موقع الوسيط يكون بين نصف وحدة قياس قبل و بعد القيمة الوسيطة المكررة ،
 - 3 وأن القيم المكررة تتوزع بانتظام على وحدة القياس هذه ، والتي تقسم على العدد الذي تكررت به هذه القيمة.
- ومثال لذلك إذا تكرر الرقم الوسيط ثلاث مرات كالاتي: 2 ، 3 ، 3 ، 4 ، 5 ، 5 ، 5 ، 7 ، 9 ، 10 ، نجد أن:

- 1 القيمة الوسطية هي (5) التي تكررت ثلاثة مرات .

2 وان فاصلة القراءة تقع بين (4,5 - 5,5) ، بعد أن قسمت الفاصلة (1 / 3 = 0,33) ، ويكون حينها موقع (5) بين : (4,5 - 4,83) ، (5,16 - 5,5) ،

3 وبترتيب القيم نجد الآتي : 2 ، 3 ، 3 ، 4 ، 4,5 ، 4,83 ، 5,16 ، 5,5 ، 7 ، 9 ، 10

4 عندئذ يتم اختيار الموقع الوسيط ، وهو (4.83) ليمثل قيمة الوسيط لأن عدد القيم أصبح فرديا .

مثال آخر إذا تكرر الرقم الوسيط أربع مرات كالآتي: 7 ، 6 ، 5 ، 5 ، 5 ، 5 ، 4 ، 4 ، 3 ، حيث نتبع الخطوات التالية لإيجاده:

1 تقع الفاصلة بين (4,5 و 5,5) ، حيث تقسم الى أربع أقسام ليكون موقع القيم (5) وبإتباع الصيغة الآتية : 7 ، 6 ، 5,5 ، 5,25 ، 5,0 ، الوسيط ، 4,75 ، 4,5 ، 4 ، 3

2 وبما أن عدد القيم أصبح زوجيا لذا تجمع قيمتي التسلسل الخامس و السادس لبعطي قيمة الوسيط : (5,0 + 4,75) / 2 = 4,88

إيجاد الوسيط في حالة البيانات المبوبة (الجدول التكرارية)

يعتبر استخراج قيمة الوسيط في التوزيعات التكرارية أكثر صعوبة و تعقيدا مقارنة مع البيانات غير المبوبة، حيث يستلزم إيجاد التكرار المتجمع الصاعد أو الهابط. ولإيجاد الوسيط للبيانات المبوبة تستعمل الطريقتين الحسابية والبيانية.

يمكن إيجاد الوسيط بيانيا من المنحنى المتجمع الصاعد أو الهابط أو تقاطع المنحنيين الصاعد والهابط معا في رسم واحد. وعموما، لإيجاد الوسيط للقيم المبوبة تتبع الخطوات التالية:

- (1) استخراج التكرارات التراكمية المقابلة للفئات ،
- (2) تعيين الفئة الوسيطة ، التي تقع فيها قيمة الوسيط ، التي يكون التكرار التراكمي فيها يساوي نصف المجموع العام للتكرارات ،
- (3) تعيين الحد الأدنى للفئة الوسيطة، والتكرارات التراكمية السابقة لها، و التكرارات المقابلة لها ،
- (4) تحديد طول الفئة ،
- (5) تطبيق المعادلة :

$$Med = L + \left\{ \left(\frac{n}{2} - \sum fi \right) / fmed \right\} * i$$

حيث أن : Med تمثل الوسيط ، (L) يمثل الحد الأدنى للفئة الوسيطة ، (n) يرمز لمجموع التكرارات، ($\sum fi$) تعني مجموع التكرارات التراكمية للفئات التي تسبق الفئة الوسيطة ، (fmed) تمثل التكرارات المقابلة للفئة الوسيطة ، (i) ترمز الى طول الفئة.

بالنسبة لحساب الوسيط حسابيا في حالة المنحنى المتجمع الصاعد، نتبع الخطوات التالية:

- 1 تكوين الجدول المتجمع الصاعد وذلك باستخدام الحدود الحقيقية.
 - 2 إيجاد رتبة الوسيط ($2/n$) سواء كانت فردية أو زوجية.
 - 3 تحديد نقطة $2/n$ على المحور الرأسي للتكرارات ونرسم منها خطاً أفقياً موازياً لمحور الفئات إلى أن يلتقي بالمنحنى في نقطة.
 - 4 نخفض من تلك النقطة عموداً رأسياً يلاقي محور الفئات في نقطة تكون قيمتها هي قيمة الوسيط بيانياً. تحديد مكان الوسيط بحيث يكون التكرار السابق له t_1 والتكرار اللاحق له t_2 أكبر من $2/n$.
 - 5 نأخذ الحد الحقيقي للتكرار السابق على أنه البداية الحقيقية للفئة الوسيطة ونرمز له بالرمز أ ،
 - 6 نعين طول الفئة الوسيطة ويساوي الحد الأدنى للفئة التالية مطروحاً منه الحد الأدنى للفئة الوسيطة ونرمز له بالحرف ل
 - 7 ويعطى الوسيط بالعلاقة: الوسيط = $A + (L/n - t_1) \times L$ حيث تشير t_1 للتكرار المتجمع الصاعد السابق للتكرار المتجمع الوسيطي ، t_2 للتكرار التجمع الصاعد اللاحق للتكرار المتجمع الوسيطي.
- أما في حالة المنحنى المتجمع الهابط ، نتبع نفس الخطوات السابقة للمتجمع الصاعد لتحديد قيمة الوسيط بيانياً.
- أما في حالة تقاطع المنحنيين الصاعد والهابط في نقطة، نقوم بإسقاط عمود رأسي على محور الفئات تكون نقطة تقاطعه مع محور الفئات هي قيمة الوسيط.

ومثال لذلك فقد درس أحد الجغرافيين اقليما صناعيا فوجد فيه (62) مصنعا للصابون، تتباين في كمية انتاجها في اليوم الواحد بين (22 - 30) كيلوغراما من الصابون ، وتحصل على التكرارات المبينة في الجدول أدناه.

جدول يوضح انتاجية مصانع الصابون في إقليم صناعي فرضي

الانتاج	التكرار	تصاعدي	تنازلي
22	1	1	62
23	1	2	61
24	2	4	60
25	8	12	58
26	9	21	50
27	15	36	41
28	12	48	26
29	10	58	14
30	4	62	4

لإيجاد الوسيط للتكرارات المتجمعة تصاعديا من المثال أعلاه ، نتبع الخطوات الآتية :-
(1) تقسيم مجموع التكرارات على العدد (2) لتحديد موقع التكرار الوسيط تراكميا ، وفي المثال أعلاه فان تكرار الوسيط هو : $(31 = 2 / 62)$.

(2) تحديد الفئة التي جاء بها التكرار الوسيط (31) ، وتقع ضمن التكرار

المتجمع التصاعدي (36) المناظر لكمية الانتاج (27) كيلوغرام .

(3) ومن الجدول يلاحظ أن (15) مصنعا قد بلغ انتاجهم (27) كيلوغرام

، وتراتبهم بين (22 - 36) ، والمطلوب معرفة الذي أنتج (31) كيلوغرام منهم .

(4) تطبق حينئذ المعادلة المشار إليها آنفا ، وبعد تحديد القيم المطلوبة

للمعادلة :-

$$Med = L + \left\{ \left(\frac{n}{2} - \sum fi \right) / fmed \right\} * i$$

الحد الأدنى للفئة الوسيطة = 26,5 ، مجموع التكرار = 62 ، التكرار المتجمع

التصاعدي للفئة السابقة للفئة الوسيطة = 21 ، تكرار الفئة الوسيطة = 15 ،

طول الفئة = 1 .

$$\text{الوسيط} = 26,5 + \{ 15 \setminus (21 - (2 \setminus 62)) \}$$

$$= 26,5 + (15 \setminus (21 - 31))$$

$$= 26,5 + 0,67 = 27,17 \text{ أي ان الانتاجية الوسيطة لمصانع}$$

الصابون في هذا الإقليم الصناعي هي (27,17) كيلوغرام من الصابون في اليوم الواحد .

ومن المثال السابق وإيجاد الوسيط للتكرارات بعد ترتيبها تنازليا، تحسب قيمة الوسيط

بالصيغة الآتية ، وذلك وفق الآتي:

- الحد الاعلى للفئة الوسيطة = 27,5، التكرار المتجمع التنازلي للفئة اللاحقة
للفئة الوسيطة = الوسيط = 27,5 - $\{ 15 (26 - (2 \setminus 62)) \} = 1 * 27,5 =$
 $27,17 = 0,33 - 27,5 = 1 * (15 (26 - 31)) - 27,5$
وهي النتيجة ذاتها .

مثال إضافي: أوجد الوسيط لأعمار الطلاب في المثال التالي:

الفئات	3-1	6-4	9-7	12-10	15-13
التكرارات	2	3	8	5	2

الحل

1 تكون جدول التكرار المتجمع الصاعد كالآتي:

التكرار المتجمع الصاعد	فئات التكرار المتجمع الصاعد
0	< 4.5
2	< 6.5
7	< 8.5
15	< 10.5
19	< 12.5
20	< 14.5

- 2 نحسب $n/2$ وهي تساوي $(20/2 = 10)$ ، ونلاحظ أن 10 تقع بين 7 و 15
فنضع خطاً أفقياً يمثل تكرار الوسيط المتجمع 10 ،
3 وعليه يكون :

$$أ = 8.5 ، ت_1 = 7 ، ت_2 = 15 ، ل = 10.5 - 8.5 = 2$$

4 ونطبق قانون الوسيط لنحصل على قيمته كالآتي:

$$\text{الوسيط} = 8.5 + 10 - 15 / 7 - 10 \times 2 = 9.25$$

أما في حالة تقاطع المنحنيين الصاعد والهابط في نقطة (أي تحديد الموقع الوسيط من

خلال اسقاط التوزيعات التكرارية تراكميا على ورق بياني) ، فنتبع الخطوات التالية:

(1) نرسم مستقيما أفقيا يمثل الفئات ، ومستقيما آخر عموديا يرمز للتكرار المتجمع التصاعدي.

(2) نرسم مستقيم أفقي يقطع المحور الرأسي عند النقطة التي تمثل التكرار المتجمع التصاعدي المناظر للفئة الاولى، ويلقي العمود المقام على المحور الافقي عند نقطة الحد الاعلى الحقيقي للفئة الاولى. وتمثل هذه النقطة التكرار المتجمع التصاعدي للفئة الاولى

(3) نكرر الخطوة السابقة مع الفئات الاخرى ،

(4) نرسم التكرار المتجمع الهابط بإتباع نفس الخطوات السابقة للمتجمع الصاعد.

(5) تمثل نقطة تقاطع المنحنيين الصاعد والهابط موقع القيمة.

المنوال

هو القيمة الأكثر شيوعا (تكرارا) في مجموعة البيانات. وقد يكون لمجموعة البيانات منوال واحد فتسمى وحيدة المنوال، أو يكون لها أكثر من منوال فتسمى متعددة المنوال ، أو لا يكون لها أي منوال فتسمى عديمة المنوال.

مثال : أحسب المنوال من البيانات التالية : 6 ، 10 ، 6 ، 4 ، 9 ، 6 ، 2

الحل: يوجد لهذه البيانات منوال واحد وهو القيمة 6 لأنها تكررت ثلاث مرات أكثر من غيرها.

مثال : أحسب المنوال من البيانات التالية: 4 ، 7 ، 4 ، 4 ، 9 ، 7 ، 3 ، 4

الحل: يوجد لهذه البيانات منوال واحد وهو القيمة 4 لأنها تكررت أربع مرات أكثر من غيرها.

مثال: أحسب المنوال من البيانات التالية: 10 ، 5 ، 7 ، 9 ، 8 ، 7 ، 5 ، 7 ، 5

الحل: نجد من خلال هذه البيانات أن القيمة 5 تكررت ثلاث مرات والقيمة 7 تكررت ثلاث مرات ، وعليه هذه البيانات لها منوالان هما 7 و 5.

مثال: أحسب المنوال من البيانات التالية: 15 ، 7 ، 11 ، 12 ، 8 ، 9 ، 4

الحل : لا يوجد لهذه البيانات أي قيمة تكررت أكثر من مرة ، وعليه لا يوجد منوال لهذه البيانات.

المنوال في حالة البيانات المبوبة (الجدول التكرارية)

في حالة البيانات المبوبة لا يمكن القول أن قيمة معينة يكون لها أكبر تكرار لأن القيم تذوب داخل الفئات المختلفة، ولذلك يمكن القول بأن هناك فئات منوالية وهي الفئات التي يقابلها أعلى تكرار. وفي حالة تساوي تكرار الفئة السابقة للفئة المنوالية مع تكرار الفئة اللاحقة للفئة المنوالية فإنه يمكن حساب قيمة المنوال بمركز الفئة المنوالية أي في منتصفها. وفي حالة عدم تساويهما في التكرار ، يمكن حساب المنوال كالتالي:

- نوجد أكبر تكرار ت ، وعليه يمكن إيجاد التكرار السابق له وهو t_1 والتكرار اللاحق له وهو t_2 .

- نأخذ قيمة الفئة المنوالية ويرمز لها بالرمز أ وهي الفئة التي يقابلها أعلى تكرار .
- نحدد طول الفئة المنوالية ل ، وهو يساوي الفرق بين بداية الفئة المنوالية وبداية الفئة التالية لها ، ويتم تطبيق القانون:

$$\text{المنوال} = أ + ت - ت_1 / 2 - ت_1 - ت_2 . ل$$

مثال : أوجد المنوال لأعمار الطلاب في الجدول التالي:

الفئات	3-1	6-4	9-7	12-10	15-13
التكرارات	2	3	8	5	2

من الجدول نجد أن : $ت = 8$ ، $ت_1 = 5$ ، $ت_2 = 4$ ، $أ = 8.5$ ، $ل = 10.5 - 2.0 = 8.5$

وبالتعويض في القانون ينتج أن:

$$\text{المنوال} = 8.5 + 8 - 5 - 16 / 5 - 2 = 9.36 \text{ سنة}$$

وتتمثل مميزات المنوال في أنه يسهل حسابه ولا يتأثر بالقيم الشاذة، كما يمكن إيجاد القيم الوصفية والتوزيعات التكرارية المفتوحة. أما عيوبه فإنه عند حسابه لا تؤخذ جميع قيم البيانات في الاعتبار ، وقد يكون لبعض البيانات أكثر من منوال وبذلك لا يمكن تحديد قيمة وحيدة للمنوال.

بعض التطبيقات الجغرافية للمنوال

1) المنوال المكاني :

يعتبر وجود منوال واحد ، أو أكثر في أي منطقة دراسة على تركيز مكاني (أو زمني) في بؤر معينة يتطلب تحليل أسبابها. فعند دراسة مرض معين ، أو ظاهرة مناخية أو

جيمورفولوجية معينة ، فان وجود المنوال دليل على وجود مسببات محلية تتطلب التحليل و النظر بعمق في الظروف البيئية التي أدت الى تكوينها و وجودها في هذا المكان دون غيره . أن وجود المنوال سبب كاف للجغرافي لدراسة معمقة لهذه الظاهرة أو الحالة .

(2) منوال الانماط النقطية :

تعتمد طريقة تحديد المنوال للانماط النقطية على حصر المنطقة على الخارطة و تقسيمها الى مربعات متساوية المساحة عن طريق انشاء مجموعة من الاحداثيات السينية و الصادية ، وحساب عدد النقاط في كل مربع ، والذي يضم أكبر عدد منها يكون هو الموقع.

مقارنة بين مقاييس النزعة المركزية

(1) المتوسط ، هو عضو نظام رياضي يسمح بعد استخدامه تطبيق تحليلات احصائية أكثر عمقا

(2) ان انحراف القيم عن الوسيط ذي تطبيقات محدودة في الطرائق الاحصائية المتقدمة،

(3) المتوسط أكثر استقرارية و فاعلية من غيره من مقاييس النزعة المركزية ، وعند أخذ عينات من المجتمع نفسه فان معدلاتها تكون أقل تذبذبا من قيم الوسيط و المنوال ، بعبارة اخرى ، يوفر المتوسط افضل تقدير لخصائص المجتمع (المصدر السابق) .

العلاقة بين الوسط الحسابي والوسيط والمنوال

توجد علاقة تجريبية بين الوسط الحسابي والوسيط والمنوال في حالة التوزيعات التكرارية أحادية المنوال وغير المتماثلة وذات الالتواء البسيط تأخذ شكل المعادلة التالية:

الوسط الحسابي - المنوال = 3 (الوسط الحسابي - الوسيط)
وقد وجد أن الوسيط تقع قيمته بين قيمتي الوسط الحسابي والمنوال.
أما في حالة التوزيعات التكرارية المتماثلة الوحيدة المنوال ، فإن قيمة الوسط الحسابي تكون مساوية لقيمة الوسيط وتكون مساوية لقيمة المنوال ، أي أن:

$$\text{الوسط الحسابي} = \text{الوسيط} = \text{المنوال}$$

أما في حالة التوزيعات التكرارية الملتوية (إلتواء موجب - إلتواء سالب) فإن العلاقة:
الوسط الحسابي - المنوال = 3 (الوسط الحسابي - الوسيط) تكون غير صحيحة.

النسب المئوية

تعتبر النسب المئوية من الطرق البسيطة المستخدمة لعرض وتلخيص البيانات، كما تعتبر النسبة المئوية أداة مفيدة عند مقارنة الحجم النسبي لكميتين مع بعضهما البعض. ولكن عند تفسير النسب المئوية من الضروري الانتباه الى أن النسب الصغيرة لكمية كبيرة تعني قيمة كبيرة ، والعكس صحيح النسب. فمثلا النسبة المئوية (10%) لكمية بالمئات تختلف عن غيرها لكمية بالالوف، كما أن نمو سكان مدينة ما بنسبة (2%) سنويا قبل عشرين سنة ليس نفسه الآن نظرا للتبدلات التي قد حصلت في حجمها و تركيبها السكاني. وبما أن النسب المئوية هي مقارنة نسبية فإنها تتطلب تقسيم البيانات الى مجاميع منفصلة عن بعضها طبقا لخصائص كل منها . فالارض الزراعية تصنف حسب جودتها ، أو ملكيتها ، أو طبيعة زراعتها ، أو نوع المحصول المزروع فيها. ويكون تلخيص هذه المعلومات كنسب مئوية مفيدة عند وصف منطقة الدراسة وتكوين صورة ذهنية موجزة عن طبيعة الزراعه واقتصاد منطقة معينة في زمن محدد عند جدولة

نسب استعمالات الارض فيها. وهذه الصورة ليست تحليلاً للبيانات ، بل ملخصا يصف إجمالي توزيع الاستعمالات في منطقة معينة في زمن محدد . ومن أوجه قصورها أنها تخفي الكثير من التفاصيل الجوهرية والعلاقات غير المنظورة بين المتغيرات قيد الدرس . وفي العديد من الحالات تعامل النسب المئوية كمعدلات. فالإحصاءات الرسمية (التعدادات العامة) تعامل هكذا عند دراسة المجاميع الثانوية او مقارنة نتائج الدراسات المحلية مع الحالة العامة او المعيارية . فالنسب المئوية للفئات العمرية على المستوى الوطني او الاقليمي تعامل كمعدلات تقارن معها نتائج المسوحات الميدانية المحلية و الاقليمية .

الرقم هو القيمة المفردة التي تأخذ شكل المتوسط أو النسبة المئوية أو غيرها. ويعتبر الرقم في حد ذاته ذي معنى محدود، ولكن تزداد قيمته وضوحا عند مقارنته مع غيره من الارقام . فعندما يعلن متجر ما عن تخفيض للأسعار بقيمة (100) أو (1000) جنيهه ، فقد يكون هذا تلاعبا بالالفاظ وتكون هذه الارقام غير ذات معنى إن لم تقارن بسعر السلعة ذاتها. . فالرقم لا يفسر نفسه ، بل يتم ذلك من خلال أرقام أخرى ذات دلالة و معنى. وما هو صحيح عن النسب المئوية هو كذلك عن النسب الاخرى ، مثل نسبة الولادات ، نسبة الوفيات ، نسبة الاعالة ، نسبة الخصوبة ، نسبة الجريمة ، نسبة الملكية ، وغيرها. فمقارنة هذه النسب لفترات مختلفة ، او لمناطق متباينة في أحجامها سكانها ، يعني إخفاء بعض الحقائق و التحيز غير العلمي ما لم تعرض مع الارقام التي تمثلها، أو مع المقاييس الأخرى التي توضح جوانب أخرى من الحقيقة.

الربيعات والعشيرات والمئينات

- إذا رتبت عينة من البيانات حسب قيمتها تصاعدياً أو تنازلياً ، فإن القراءة التي تكون في المنتصف والتي تقسم العينة إلى مجموعتين متساويتين في العدد هي الوسيط كما سبق الإشارة له. وتعميم الفكرة وتقسيم البيانات عد ترتيبها إلى أربعة أجزاء متساوية ويرمز لها ب 1 ، 2 ، 3 فإن 1 تسمى الربيع الأول ، 2 تسمى الربيع الثاني (الوسيط) ، 3 تسمى الربيع الثالث.
- وكذلك يمكن إيجاد القيم التي تقسم البيانات السابقة بعد ترتيبها إلى عشرة أقسام ونرمز لها ب ق1 ، ق2 ، ... ق9 ، حيث ق1 تسمى العشير الأول وهي تمثل القيمة التي يسبقها عُشر القراءات ، ق2 تسمى العشير الثاني وهي تمثل القيمة التي يسبقها 0.2 من القراءات وهكذا.
- كما يمكن إيجاد القيم التي تقسم البيانات السابقة بعد ترتيبها إلى مائة قسم ونرمز لها ب ج1 ، ج2 ج99 حيث ج1 تسمى المئين الأول وهي تمثل القيمة التي يسبقها 0.01 من القراءات، و ج2 تسمى المئين الثاني وهو يمثل القيمة التي يسبقها 0.02 من القراءات وهكذا لباقي المئينات.
- يعطى قانون حساب الربيعات في حالة البيانات المبوية مثل قانون الوسيط السابق مع استبدال 2/ن ب 4/ن للربيع الأول ، و 2/ن / 4 للربيع الثاني وهكذا.
- كذلك استبدال 2/ن ب 100/ن للمئين الأول ، 2/ن / 100 للمئين الثاني وهكذا.

ويمكن توضيح ذلك من المثال التالي:

نكون جدول التكرار المتجمع الصاعد كآتي:

حدود الفئات	التكرار المتجمع الصاعد
< 4.5	0
< 6.5	2
ق ₂	
< 8.5	7
< 10.5	15
ج ₉₀	
< 12.5	19
< 14.5	20

لإيجاد العشير الثاني ق₂ نستخدم القانون:

$$ق_2 = أ + (2ن / 10 - ت_1) / ت_2 - ت_1 \cdot ل$$

حيث أ بداية الفئة للعشير الثاني ، ن مجموع التكرارات ، ت₁ التكرار السابق ، ت₂

التكرار اللاحق ، ل طول فئة العشير الثاني:

$$ق_2 = 6.5 + (2 - 4) / 2 - 7 + 2 \cdot 2 = 7.3$$

لإيجاد المئين التسعين ج₉₀ نستخدم القانون:

$$ج_{90} = أ + (90 / ن - 100 / ت_1) / ت_2 - ت_1 \cdot ل$$

$$ج_{90} = 10.5 + (15 - 18) / 15 - 19 + 2 \cdot 15 = 12$$

الفصل الرابع

مقاييس التشتت

سيتناول هذا الفصل مقاييس التشتت المختلفة التي تعتبر ضرورية لوصف البيانات وتحليلها كميًا بغرض الوصول إلى فهم أكثر وضوحًا للظاهرة محل الدراسة، إذ تعتبر مقاييس النزعة المركزية ليست كافية للقيام بذلك. وهنا لا بدّ من ذكر مقاييس أخرى تبين مدى التفاوت في القيم حول المتوسط. فمثلاً تساوي المتوسطين وتساوي الحدين الأعلى والأدنى قد يخفيان وراءهما اختلافاً في شكل التوزيع مما يتطلب الأمر قياس التشتت وانتشار القيم حول متوسطاتها وآلا يقتصر الأمر فقط وعدم الاختصار على المتوسطات. وتوجد عدّة مقاييس تصلح لقياس درجة التشتت أهمها المدى والانحراف المتوسط والانحراف المعياري وغيرها.

المدى

هو الفرق بين أقل (أصغر) قيمة وأعلى (أكبر) قيمة.

مثال : أوجد المدى للدرجات التالية: 10، 3، 7، 5

الحل : المدى = $10 - 3 = 7$

مثال أوجد المدى لدرجات الطلاب المعطاة في الجدول التالي:

الفئات	49-40	59-50	69-60	79-70	89-80	99-90
التكرار	2	9	15	11	2	1

99.5

39.5

الحل : المدى = $99.5 - 39.5 = 60$

وكلما كبرت قيمة المدى دلت على تشتت أكثر وسط المجموعة.

ومن تعريف المدى يتضح أنه مقياس بسيط جدا وأن كل المطلوب معرفته هو أكبر قيمة وأصغر قيمة فقط. وعلى ذلك يعتمد المدى في حسابه على قيمتين فقط (أكبر قيمة وأصغر قيمة) وبالتالي يهمل باقي القيم. كما أنه لا يقيس تشتت البيانات عن متوسطها، فهو لا يشير أبدا إلى متوسط البيانات (أو مركزها). كما أنه حساس جدا لأي قيمة شاذة أو متطرفة. ومثال لذلك إذا كانت أعمار السلطة التشريعية في دارفور هي : 72 ، 78 ، 20 ، 74 ، 73 ، 75 ، 70 ، فإن المدى في هذه الحالة هم :

$78 - 20 = 58$. ويلاحظ وجود قيمة شاذة لباقي القيم وهي 20 وهي أصغر قيمة. فإذا أهملت هذه القيمة ، أو لم تكن موجودة أصلا، لأصبح المدى $= 8$ ($78 - 70$). وقد أدت مثل هذه الأسباب للحث عن طرق أخرى أكثر تعبيرا ودقة لتشتت البيانات، فأوجدوا التباين.

التباين

أحد أهم مقاييس التشتت. يأخذ جميع القيم في الاعتبار عند حسابه كما يقيس لتشتت عن الوسط الحسابي للقيم ، كما يسهل معالجته رياضيا ويدخل في تكوين عدد من المقاييس والاختبارات الاحصائية المهمة.

والفكرة الأساسية لتباين هي حساب انحرافات جميع القيم عن وسطها الحسابي (أي حساب الفرق بين كل قيمة والوسط الحسابي). وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الانحرافات) بالموجب، والبعض الآخر أصغر من الوسط فتكون الفروق بالسالب. زدائما يكون مجموع هذه الانحرافات مساويا للصفر. ويكون الحل

هنا إما إهمال العشرات السالبة ، أو تربيع هذه الانحرافات. وإهمال الإشارات السالبة ليس له مبرر رياضي ، فيكون الحل هو تربيع هذه الانحرافات ، ثم نحسب متوسط الانحرافات المربعة فنحصل على التباين. أي أن التباين يعرف كما يلي: هو متوسط مربعات انحرافات القيم عن وسطها الحسابي (التباين $\sigma^2 = 1/n \sum (x_i - \bar{x})^2$) مثال : أحسب تباين دخول الشريحة الاجتماعية التالية: 78 ، 76 ، 76 ، 74 ، 75 ، 70 ، 75 ، 71

الحل :

- 1 تحسب أولا الوسط الحساب = مج س / ن = $592 / 8 = 74$
- 2 تحسب انحرافات القيم عن الوسط الحسابي ، كما يلي : $4- ، 1+ ، 3- ، 1+ ، 0 ، 2+ ، 1- ، 4+$ (وبلاحظ أن مجموع هذه الانحرافات يساوي صفرا).
- 3 تربيع هذه الانحرافات كما يلي: $16 ، 1 ، 9 ، 1 ، 0 ، 4 ، 16$
- 4 التباين يساوي متوسط هذه المربعات ، أي : $16 + 1 + 9 + 1 + 0 + 4 + 16 = 56$

$$6 = 56 / 16$$

الانحراف المعياري

يلاحظ أن تمييز التباين سيكون وحدات مربعة لأنه يتم تربيع الانحرافات (أو الفروق). لذلك فإنه يؤخذ الجذر التربيعي للتباين فنحصل على الانحراف المعياري ، والذي سيكون تمييزه هو نفس الوحدات. ومن المثال السابق يكون الانحراف المعياري للدخول هو : $6\sqrt{}$ = 2.2449 دولارا.

والانحراف المعياري - عادة - يستخدم بدلا عن التباين كأهم مقياس للتشتت. فالتباين هو : مربع الانحراف المعياري ، والانحراف المعياري هو: الجذر التربيعي الموجب للتباين.

يعرف الانحراف المعياري على أنه : الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي. أي هو الجذر التربيعي الموجب للتباين ويرمز له عادة بالرمز S ، ولتباين العينة بالرمز S^2 ، وللانحراف المعياري للمجتمع بالرمز اللاتيني σ ، ولتباين المجتمع بالرمز σ^2 .

ويمكن التعبير عن كل من التباين والانحراف المعياري بالرموز كما يلي:
إذا فرضنا أن قيم العينة هي: s_1 ، s_2 ، ، s_n (أي عددها n) فإن الخطوات تكون كما يلي:

- 1 حساب الوسط الحسابي : $\bar{s} = \text{مجم } s / n$
- 2 حساب انحرافات القيم عن الوسط الحسابي: $s_1 - \bar{s}$ ، $s_2 - \bar{s}$ ، $s_n - \bar{s}$
- 3 تربيع هذه الانحرافات : $(s_1 - \bar{s})^2$ ، $(s_2 - \bar{s})^2$ ، $(s_n - \bar{s})^2$
- 4 التباين هو الوسط الحساب لهذه المربعات، أي:

$$\sigma^2 = (s_1 - \bar{s})^2 + (s_2 - \bar{s})^2 + \dots + (s_n - \bar{s})^2 / n = \text{مجم } (s_i - \bar{s})^2 / n$$
- 5 +لانحراف المعياري هو الجذر التربيعي الموجب للتباين ، أي: $\sigma = \sqrt{\text{مجم } (s_i - \bar{s})^2 / n}$

ويمكن حساب الانحراف المعياري بالطريقة المختصرة كآتي:

$$\text{الانحراف المعياري} = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

وفي الطريقة المختصرة فإن كل ما هو مطلوب معرفته لحساب التباين أو الانحراف المعياري هو مج س (أي مجموع القيم) ، مج س² (أي مجموع مربعات القيم) ثم التعويض في القانون.

في حالة المجتمع والبيانات الخام تستخدم المعادلة التالية:

$$\text{التباين} = \frac{\sum x^2}{n} - (\bar{x})^2$$

$$\text{التباين} = \frac{\sum x^2}{n} - (\bar{x})^2$$

في حالة العينات والبيانات الخام تستعمل المعادلة التالية:

$$\text{الانحراف المعياري} = \sqrt{\frac{\sum x^2}{n-1} - (\bar{x})^2}$$

$$\sigma = \sqrt{\frac{\sum x^2}{n-1} - (\bar{x})^2}$$

مثال : أوجد الانحراف المعياري للبيانات التالية: 5 ، 6 ، 7 ، 9 ، 8

س	س - س̄	(س - س̄) ²
8	1	1
9	2	4
7	0	0
6	-1	1
5	-2	4
35		10

$$\bar{s} = \text{مج س} / \text{ن} = 5 / 35 = 7$$

$$\sigma^2 = 1 - \text{ن} / 1 = \text{مج (س - س)}^2 / 1 = 10 \times 1 - 5 / 1 = 2.5$$

$$\sigma = 1.58 = 2.5 \sqrt{}$$

الانحراف المعياري في حالة البيانات المبوبة (العينات)

$$\sigma^2 = 1 - \text{ن} / 1 = \text{مج ت (س - س)}^2$$

$$\sigma = 1 - \text{ن} / 1 \sqrt{ \text{مج ت (س - س)}^2 }$$

ومثال لذلك البيانات التالية:

الفئات	49-40	59-50	69-60	79-70	89-80	99-90
التكرار	2	9	15	11	2	1

الحل

الفئات	مراكز الفئات (س)	التكرار (ت)	ت X س	س - س - س ²	(س - س ²)	ت (س - س ²)
49-40	44.5	2	89	21.25	457.56	903.13
59 - 50	54.5	9	490.5	11.25 -	126.56	1139.06
69 - 60	64.5	15	976.5	1.25 -	1.56	23.49
79 - 70	74.5	11	819.5	8.75	70.56	842.19
89 - 80	84.5	2	169	18.75	351.56	703.13
99 - 90	94.5	1	94.5	28.75	820.56	820.56
المجموع		40	2630			4437.5

$$\bar{s} = 40/2630 = 65.75$$

$$10.6 = \sigma \quad 113.7 = 1 - 40 / 4437.5 = \sigma^2$$

ومن خصائص الانحراف المعياري أن قيمته دائما موجبة أو أكبر من أو تساوي صفر .
 فاقل قيمة تساوي صفر (وذلك عندما تكون جميع القيم متساوية ، وفي هذه الحالة لا
 توجد فروق أو انحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم
 ، وبالتالي فإن قيمة الانحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).
 وكلما كان التشتت كبيرا حول الوسط الحسابي كلما كان الانحراف المعياري كبيرا،
 والعكس صحيح. كذلك إذا أضفنا مقدارا ثابتا من كل القيم ، فإن قيمة الانحراف المعياري
 (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع). كذلك
 إذا ضربنا كل قيمة في مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب
 القسمة على هذا المقدار الثابت. وإذا قسمنا كل قيمة على مقدار ثابت ثم حسبنا
 الانحراف المعياري للقيم الجديدة فإنه يجب الضرب في هذا المقدار الثابت.

معامل الاختلاف

لمقارنة تشتت مجموعتين (أو أكثر) من البيانات والتي تختلف في أوساطها الحسابية
 (مستواها العام) و / أو تختلف في وحدات القياس (مثلا مقارنة يانات الدخل التي تقاس
 بالريال ببيانات العمر وتقاس بالسنوات) فإن المقارنة لا تتم مباشرة بمقارنة الانحراف
 المعياري لكل منهما ، بل تتم من خلال مقياس آخر هو معامل الاختلاف أو ما يسمى
 أحيانا بمقياس التشتت النسبي حيث ينسب الانحراف المعياري لكل مجموعة إلى وسطها
 الحسابي والضرب في 100 فنحصل على مقياس نسبي أو مؤوي (وبدون تمييز)، أي تتم

المقارنة بحساب معامل الاختلاف لكل منهما، والمجموعة التي لها معامل اختلاف أكبر تكون أكبر تشتتاً والعكس صحيح أي أن

معامل الاختلاف = الانحراف المعياري / الوسط الحسابي $\sigma = 100 \times \text{س}^-$ أو = الربع الأعلى - الربع الأدنى مقسوماً على الربع الأعلى + الربع الأدنى ومثال لذلك :

الفئات	49-40	59-50	69-60	79-70	89-80	99-90
التكرار	2	9	15	11	2	1

$$\sigma = 10.6 ، \text{س}^- = 65.75$$

$$\text{معامل الاختلاف} = \sigma / \text{س}^- = 10.6 / 65.75 = 0.16$$

الطريقة الأخرى : الربع الأعلى - الربع الأدنى مقسوماً على الربع الأعلى + الربع الأدنى

$$r_1 = 58.39 ، r_3 = 73.14$$

$$0.111 = 58.39 + 73.14 / 58.39 - 73.14$$

$$\sigma^2 = 1 / n - 1 \text{ مج (س - س}^-) = 2$$

$$5.6 = 11 / 62 = 1 - 12 / 62 =$$

$$\text{معامل الاختلاف} = \sigma / \text{س}^- = 2.4 / 5 = 0.48$$

مثال آخر: إذا كان الوسط الحسابي والانحراف المعياري لدخول عينة من الناخبين بالريالات هو: $\text{س}^- = 1500$ ، $\sigma = 152$ ، وكان الوسط الحسابي والانحراف المعياري لأعمارهم (بالسنوات) هو: $\text{س}^- = 42$ ، $\sigma = 9.2$ ، فأيهما أكثر تشتتاً الدخل أم العمر؟.

الحل : لمقارنة التشتت نحسب معامل الاختلاف لكل من الدخل والعمر كما يلي:

$$1 \text{ معامل اختلاف الدخل} = \sigma / \text{س}^- = 100 \times 1500 / 152 = 10.13\%$$

$$2 \text{ معامل اختلاف العمر} = \sigma / \text{س}^- = 100 \times 42 / 9.2 = 21.0\%$$

3 بما أن معامل اختلاف العمر أكبر من معامل اختلاف الدخل ، فإن بيانات العمر تكون أكثر تشتتاً من بيانات الدخل.

نصف المدى الربيعي

نصف المدى الربيعي = الربيع الأعلى - الربيع الأدنى / 2

$$R = R_3 - R_1 / 2$$

مثال : أوجد نصف المدى الربيعي لأوزان مجموعة الطلاب الآتية : 70 ، 72 ، 71 ، 55 ، 58 ، 69 ، 65 ، 67

الحل : ترتيب البيانات : 55 ، 58 ، 65 ، 67 ، 69 ، 70 ، 71 ، 72

$$R_1 = 58 + 65 / 2 = 61.5$$

$$R_3 = 70 + 71 / 2 = 70.5$$

$$R = R_3 - R_1 / 2 = 70.5 - 61.5 = 4.5$$

نصف المدى الربيعي للبيانات المبوبة

$$R_1 = \frac{4}{n} (T_1 - T_2) + A_1$$

$$R_3 = \frac{4}{n} (T_3 - T_2) + A_3$$

مثال : أوجد نصف المدى الربيعي لدرجات الطلاب التالية:

99-90	89-80	79-70	69-60	59-50	49-40	الفئات
1	2	11	15	9	2	التكرار

الحل

- ترتيب الربع الأول = $4/40 = 10$
- ترتيب الربع الثالث: $4/120 = 30$
- نوجد التكرار المتجمع الصاعد والحدود الدنيا للفئات كالآتي:

حدود الفئات	التكرار الصاعد
أقل من 39.5	0
أقل من 49.5	2
أقل من 59.5	11
أقل من 69.5	20
أقل من 79.5	37
أقل من 89.5	39
أقل من 99.5	40

$$ل = 39.5 - 49.5 = 10$$

$$ر_1 = 4/ن_1 - 2/ن_2 + 10 = 10 \times 2 - 11 / 2 - 10 + 49.5 = 58.39$$

$$ر_3 = 4/ن_3 - 2/ن_2 + 69.5 = 10 \times 20 - 37 / 26 - 30 + 69.5 = 73.14$$

$$ر = ر_3 - ر_1 / 2 = 73.14 - 58.39 / 2 = 7.38$$

الانحراف عن المتوسط

في حالة البيانات الخام تستخدم المعادلة التالية:

$$\text{الانحراف عن المتوسط} = \frac{1}{n} \sum |س - س^-|$$

في حالة البيانات المبوبة تستخدم المعادلة التالية:

$$\text{الانحراف عن المتوسط} = \frac{1}{n} \sum ت |س - س^-|$$

مثال : أوجد الانحراف المتوسط لأعمار مجموعة من الطلاب : 5 ، 9 ، 9 ، 8 ، 7 ،

6 ، 5 ، 7

الحل :

س	س - س ⁻	س - س ⁻
6	1-	1
5	2-	2
7	0	0
7	0	0
8	1	1
9	2	2
9	2	2
5	2-	2
56		10

$$s^- = \text{مج س} / \text{ن} = 8 / 56 = 7$$

$$\text{الانحراف عن المتوسط} = 8 / 10 = 1.25$$

بالنسبة للبيانات المبوبة نذكر المثال التالي:

الفئات	49-40	59-50	69-60	79-70	89-80	99-90
التكرار	2	9	15	11	2	1

الحل

الفئات	مراكز الفئات (س)	التكرار (ت)	ت X س	س - س -	س س - س س -	ت س - س س -
49-40	44.5	2	89	21.25	21.25	42.5
50 - 59	54.5	9	490.5	- 11.25	11.25	101.25
60 - 69	64.5	15	976.5	1.25 -	1.25	18.75
70 - 79	74.5	11	819.5	8.75	8.75	96.25
80 - 89	84.5	2	169	18.75	18.75	37.5

28.75	28.75	28.75	94.5	1	94.5	90 - 99
325			2630	40		المجموع

$$\bar{s} = \text{مجموع } s / n = 65.75 = 40 / 2630$$

$$\text{الانحراف عن المتوسط} = \text{مجموع } |s - \bar{s}| / n = 8.125 = 40 / 325$$

س	س - \bar{s}	$ s - \bar{s} $	$ s - \bar{s} ^2$
1	4-	4	16
2	3-	3	9
3	2-	2	4
4	1-	1	1
4	1-	1	1
5	0	0	0
5	0	0	0
6	1	1	1
6	1	1	1
7	2	2	4
8	3	3	9
9	4	4	16
60		22	62

$$s^- = 12/60 = 5$$

$$1.8 = 12/22 = |s^- - s| \text{ مج } = \text{ الانحراف المتوسط}$$

المتغير المعياري (الدرجة المعيارية)

بالنسبة للعينات (البيانات المبوبة) تستخدم المعادلة:

$$z = \frac{s^- - s}{\text{الانحراف المعياري}}$$

بالنسبة للبيانات غير المبوبة، تستخدم المعادلة التالية:

$$z = \frac{s^- - \text{المتوسط}}{\text{التباين}}$$

مثال : حصل طالب على 82 درجة في مقرر الإحصاء حيث كان متوسط الدرجات هو 75 درجة وانحراف معياري 10 درجات. ثم حصل على 89 درجة في مقرر الرياضيات وكان متوسط الدرجات هو 81 درجة وانحراف معياري 16 درجة. في أي من المقررين كانت نسبة استيعاب الطالب أعلى؟.

$$z_1 = \frac{s_1^- - s_1}{\sigma_1} = \frac{10}{75 - 82} = 10 / 7 = 1.43$$

$$z_2 = \frac{s_2^- - s_2}{\sigma_2} = \frac{16}{81 - 89} = 16 / 8 = 2$$

الفصل الخامس

خصائص التوزيع

خصائص التوزيع

إن شكل المنحنى الذي تأخذه البيانات التيتم جمعها بغرض البحث ، يصف خصائص توزيع تلك البيانات. ويمكن أن يأخذ توزيع البيانات شكل المنحنى الطبيعي ، المنحنى المائل (الموجب أو السالب) ، المفلطح أو المنحنى ذو القمتين و القمة الواحدة.

التوزيع الطبيعي

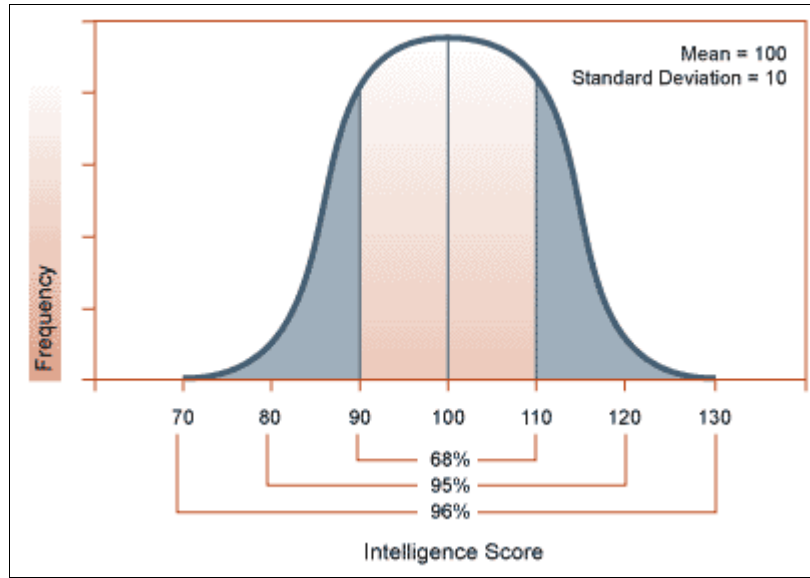
يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية وأكثرها استعمالا على الاطلاق. وقد اشتق اسمه من أن كثيرا من التوزيعات الطبيعية تأخذ شكلا قريبا منه ، كذلك فإن كثيرا من التوزيعات البيومترية كتوزيعات الطول والوزن ، وتوزيعات اخطاء المشاهدات مثل الفروق بين القيم الحقيقية والقيم المشاهدة) تأخذ شكلا قريبا منه.

يسمى منحنى هذا التوزيع بالمنحنى الطبيعي. وعند الحديث عن منحنى توزيع نظري يقصد به منحنى دالة كثافة احتمال المتغير العشوائي الذي له هذا التوزيع. زمنحنى التوزيع الطبيعي متمائل حول خط رأسي يمر بالوسط الحسابي الذي يساوي بسبب التماثل كلا من الوسيط والمنوال، وهو جرسى الشكل له قمة واحدة ويمتد طرفاه إلى ما لا نهاية يمينا ويسارا (فيقترب طرفاه من المحور الأفقي ولكنهما لا يلتقيان معه) ، ومع ذلك فإن المساحة تحت المنحنى تساوي الواحد الصحيح كما هو الحال في المساحة تحت منحنى دالة كثافة احتمال أي متغير عشوائي متصل آخر.

هناك عدد لا متناهي من المنحنيات الطبيعية ولكنها تختلف عن بعضها البعض حسب قيمة كل من الوسط الحسابي (التوقع) والانحراف المعياري. وقد تتفق منحنيات طبيعية في الانحراف المعياري ولكنها تختلف في الوسط الحسابي ، أو قد تتفق الوسط الحسابي وتختلف بالانحراف المعياري.

ويمتلك المنحنى الطبيعي عدّة خواص منها : أنه متماثل حول الوسط الحسابي ، أي أن المنحنى على يسار الوسط الحسابي يماثل المنحنى على يمينه. كما أن الوسط الحسابي في المنتصف يقسم المساحة تحت المنحنى إلى قسمين متساويين. كما أن مجموع المساحة تحت المنحنى تساوي واحد. ومهما كانت قيمة الوسط الحسابي والانحراف المعياري ، فإن للمنحنى الطبيعي الخواص التالية(أنظر الشكل المرفق أدناه):

- 1- حوالي 68.26% من المساحة يقع بين القيمتين (المتوسط + الانحراف المعياري ، المتوسط - الانحراف المعياري) ، 2
- 2- حوالي 95.46% من المساحة يقع بين القيمتين (المتوسط + 2 الانحراف المعياري ، المتوسط - 2 الانحراف المعياري) ، 3
- 3- حوالي 99.74% من المساحة يقع بين القيمتين (المتوسط + 3 الانحراف المعياري ، المتوسط - 3 الانحراف المعياري) ،
- 4- حوالي **9% من** المساحة يقع بين القيمتين (المتوسط + 1.96 الانحراف المعياري ، المتوسط - 1.96 الانحراف المعياري) ، 5
- 5- حوالي 99% من المساحة يقع بين القيمتين (المتوسط + 2.58 الانحراف المعياري ، المتوسط - 2.58 الانحراف المعياري).



شكل يوضح المنحنى الطبيعي

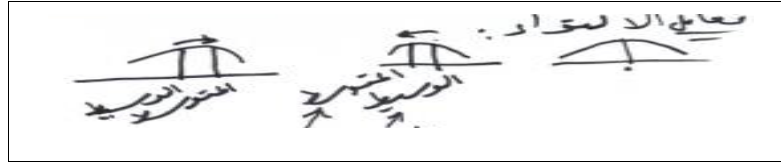
معامل الإلتواء

يشير الإلتواء إلى مدى التماثل أو عدم التماثل في شكل التوزيع. ويعرف التوزيع المتماثل أنه ذلك التوزيع الذي تقسمه القيمة الوسطى إلى قسمين متساويين بحيث يكون التزايد أو التناقص في التكرارات متشابهًا على جانبي المحور المقام عند منتصف التوزيع. أم التوزيع غير المتماثل فهو عكس التوزيع المتماثل حيث نجد أن التكرارات تتناقص أو تتزايد بشكل غير منتظم على جانبي المحور المقام عند منتصف التوزيع. **المنحنى المائل (أو الملتوي) الموجب** هو عبارة عن منحنى غير متجانس حيث تتكدس معظم الحالات (cases) جهة اليسار وعلى امتداد ذنب جهة اليمين ، حيث تشير إلى وجود حالات

قليلة عند النهاية اليمنى. وفي المنحنى المائل الموجب نجد أن المنوال أصغر من الوسيط ، والذي بدوره يكون أصغر من متوسط التوزيع. ما المنحنى المائل (أو الملتوي) السالب فهو منحنى غير متجانس حيث تكون معظم الحالات متكدسة ناحية اليمين مع امتداد لذنب طويل ناحية اليسار. وبسبب أن الكثير والكثير من الحالات توجد عند النهاية اليمنى والقليل والقليل من الحالات توجد عند النهاية اليسرى للمنحنى الاثل السالب ، يقال أنه مائل ناحية اليسار. ولمثل هذا النوع من التوزيع يكون المتوسط أصغر من الوسيط ، والذي بدوره يكون أصغر من المنوال.

مثال : إذا كانت $\bar{x} = 65.75$ ، الانحراف المعياري $(\sigma) = 10.67$ ، الوسيط = 65.5 أوجد الالتواء ن

$$ن = 3 (\bar{x} - \text{الوسيط}) / \sigma = 3 (65.75 - 65.5) / 10.67 = 0.7$$



مقياس التفلطح

$$ك = \sigma^4 / m_4$$

$$م_4 = \frac{1}{n} \sum (x - \bar{x})^4$$

$$م_4 = \frac{1}{n} \sum (x - \bar{x})^4$$

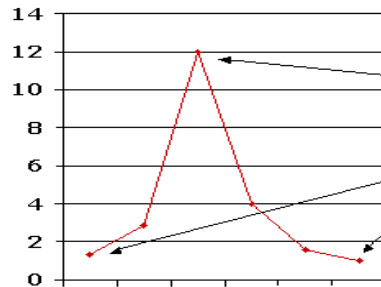
$$ك = \frac{\sigma^4}{m_4} \quad , \quad m_4 = \frac{\sum (x - \bar{x})^4}{n} \quad , \quad m_4 = \frac{\sum x^4 - 4\bar{x} \sum x^3 + 6\bar{x}^2 \sum x^2 - 4\bar{x}^3 \sum x + n\bar{x}^4}{n}$$

التوزيع المفطح

يقصد به مدى اختلاف التوزيع التكراري للظاهرة عن التوزيع الطبيعي. ومن أنواعه التوزيع المدبب Leptokurtic curve ، وهو شكل من اشكال التوزيع المفطح والذي يتميز أن معظم الحالات توجد في المركز حول المتوسط. وهو يميز نوع من التوزيع ذي تباين منخفض جدا (التباين، الانحراف المعياري، والمدى) بسبب أن درجة التجانس العالية للبيانات. وعند مقارنته مع المنحنى أو التوزيع الطبيعي، فإن هذا النوع من التوزيع يتميز أنه له قمة أعلى في الوسط مع أذنان سميكة. ومن أنواعه أيضا التوزيع المفرطح Platikurtic curve وهو نقيض للنوع السابق من التوزيع بحكم أنه يفقد تكس للحالات في المركز مع درجة عالية من التناقض (عدم التجانس) عند الأطراف. وله قمة أكثر وسعا عند القطة الوسطى للتوزيع وامتداد عظيم عند كلتا الطرفين.

1.16

Leptokurtic Graph

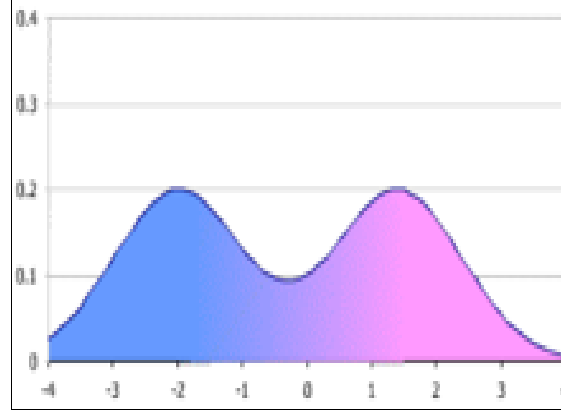


- This line graph is considered **leptokurtic** because it has:
- a. A high peak, showing that a lot of scores fall in the center of the distribution
 - b. A small range, representing only a small amount of variance or difference in the scores.

المنحنى ذي المنوالين bimodal curve

هو ذلك المنحنى الذي يميز توزيعا ذي نقطتين مرتفعتين متساويتين. ويستخدم لوصف أي توزيع يمتلك قمتين وا ضدين بصرف النظر عن أنهما متساويتين بالضبط أم لا. ونجد أن المتوسط والوسيط لهذا الشكل من المنحنيات متساويين في الغالب عند الأجزاء الدنيا أو المنخفضة من المنحنى بين القمتين.

وقد يأخذ التوزيع أيضا أكثر من قمتين منواليتين واضحتين. وعندما يأخذ التوزيع قمتين أو أكثر يسمى توزيعا متعدد المنوالات multimodal. في التوزيع ثنائي المنوال ، قد تتكدس بعض الدرجات scores أعلى المتوسط والبعض الآخر أسفله. والمنحنى ثنائي المنوال هو النقيض للمنحنى أحادي المنوال والذي يظهره المنحنى الطبيعي.



الفصل السادس

السلاسل الزمنية

بقدر اهتمام الجغرافيا بالمكان ، فإنها تهتم أيضا بالبعد الزمني للظواهر الجغرافية. وقد أشار بعض الجغرافيين إلى التسلسل الزمني وعرفه بأنه مشاهدة ظاهرة معينة بصورة متتالية خلال فترة زمنية محددة. ولتوضيح البعد الزمني استخدم الجغرافيون الرسوم البيانية الخطية والأشكال البيانية التكرارية مثل المضلع التكراري والمدرج التكراري وغيرها. وأهتم بشكل خاص بالمقاييس الخاصة بالنمو المرتبط ببعض الظواهر الجغرافية مثل أعداد السكان والإنتاج الزراعي وأعداد الثروة الحيوانية وغيرها. ومن ضمن الطرق التي استخدمها الجغرافيون دراسة التغيرات العددية التي تختص بما يطرأ على الظاهرة من تغيرات موجبة أو سالبة خلال فترة زمنية محددة وعرضها باستخدام الرسوم البيانية. وبحكم أن استخدام الأرقام الفعلية في التغيرات العددية لأنها تعتبر غير حقيقية مما تطلب استخدام التغيرات النسبية للظاهرة وذلك باستخدام الأرقام القياسية.

الأرقام القياسية

هي أداة تستخدم لقياس التغير النسبي (أو المئوي) في قيم الظواهر في زمن آخر أو من مكان إلى آخر ويكون هناك زمانان أو مكانان أحدهما يمثل سنة الأساس والثاني يمثل المقارن. ومثال لتوضيح ذلك ن سعر كغم واحد من الليمون لسنة 1975 هو 15 قرش واصبح سعره سنة 2007 يساوي 90 قرش. إذا اعتبرنا أن سنة 1975 هي سنة أساس وكانت سنة 2007 هي سنة المقارنة.

قياس التغير النسبي - الرقم القياسي

سنة المقارنة / سنة الأساس = $15/90 = 6$ يعني أن قيمة التغير الناتجة وهي 6 ،
تعني كمية الليمون التي كانت تشتري بقرش واحد سنة 1975 تشتري في سنة 2007 ب
6 قروش.

ماهية السلاسل الزمنية : عدد من المشاهدات الإحصائية نصف ظاهرة معينة مع مرور
الزمن ، أو مجموعة من المشاهدات التي أخذت على فترات زمنية متلاحقة ومتساوية.
أنواع السلاسل الزمنية:

1 **تتريه (تتعلق بالفترة) وطويلة الأجل (ممهدة):** مشاهدات ترصد لظاهرة معينة
على فترات محددة من الزمن (شهر ، ربع سنة ، فصل) مثلاً سعر سلعة على
مدار شهر كامل.

2 **لحظية (متذبذبة قصيرة الأجل) :** مشاهدات ترصد لظاهرة معينة في لحظات ()
تواريخ معينة) مثل سلسلة كمية المطر في اسبوع واحد من أحد الأشهر.

تمثيل السلسلة الزمنية بيانيا (المنحنى التاريخي للسلسلة)

مثال: أرسم المنحنى التاريخي الذي يمثل السلسلة الزمنية لعدد خريجي إحدى
الجامعات خلال السنوات 86 – 1995 في كلية من الكليات ولتخصص معين.

السنة	86	87	88	89	90	91	92	93	94	95
عدد الخريجين	6	4	8	3	7	2	6	1	5	9

إذا نظرنا إلى المنحنى التاريخي للسلسلة الزمنية السابقة نلاحظ أنها ترتفع في بعض
السنوات وتتنخفض في سنوات أخرى ، وهذا التذبذب يسمى خشونة السلسلة الزمنية.

ولحساب الخشونة يوجد مقياس يسمى مقياس الخشونة ، أو معامل الخشونة:

$$م.خ = مج_{ر=2} (س_ر - س_{ر-1})^2 / مج_{ر=2} (س_ر - س_{ر-2})^2$$

حيث أن $س_ر$: الملاحظة رقم (ر) في السلسلة الزمنية ، ن : عدد قيم السلسلة ، ر : رتبة كل قيمة في السلسلة

وكلما كان معامل الخشونة أقل كلما كانت السلسلة الزمنية ملساء أكثر ، ويحسب معامل الخشونة للظواهر وليس للزمن، ويلاحظ أن المجموع يبدأ من الملاحظة الثانية من $مج_{ر=2}$.

مثال : أحسب معامل الخشونة للسلسلة: 6 ، 4 ، 8 ، 3 ، 7 ، 5 ، 6 ، 7 ، 5 ،
9

الحل

اولا : نرتب مشاهدات السلسلة بحيث يعطي كل مشاهدة رقم صحيح موجب ابتداء من (1).

القيمة: 6 ، 4 ، 8 ، 3 ، 7 ، 5 ، 6 ، 7 ، 5 ، 9

$س_ر$: 1س ، 2س ، 3س ، 4س ، 5س ، 6س ، 7س ، 8س ، 9س ، 10س

ثانيا: نحسب الوسط الحسابي لمشاهدات السلسلة = $10/60 = 6$

ثالثا: نحسب كل من البسط والمقام في قانون معامل الخشونة سابق الذكر(أنظر

الوحة السابعة من كتابنا في سي دي)

وبما أن معامل الخشونة قيمته عالية فلا بد من تقليله وذلك عن طريق إيجاد سلسلة زمنية جديدة محل السلسلة الزمنية الأصلية بحيث يكون معامل الخشونة إليها أقل

من معامل الخشونة للسلسلة الأصلية. يتم ذلك من خلال ما يعرف بطريقة المتوسطات المتحركة والمتوسطات المتحركة أو الأوساط المتحركة.

إيجاد عناصر سلسلة جديدة بطريقة المتوسطات المتحركة

تستخدم المتوسطات المتحركة لتحليل السلاسل الزمنية وذلك بغرض دراسة التغيرات المختلفة التي تطرأ على الظاهرة المدروسة بسبب تأثير عدة مؤثرات مما يمكن من الاستفادة منها في إمكانية التنبؤ بحال أو وضعية الظاهرة في أزمنة لا نملك عنها أي بيانات. يمكن تمثيل السلسلة الزمنية بيانياً بتعيين أزواج مرتبة (الزمن ، قيمة الظاهرة) ثم نوصل تلك النقاط فينتج ما يعرف بالمنحنى التاريخي للسلسلة الزمنية. وعند تحليل البيانات الجغرافية ذات الاختلافات الزمانية يجب التركيز معرفة المؤثرات الاتجاهية (هي التي تؤثر في الاتجاه العام للظاهرة زيادة أو نقصاناً على مدار فترة زمنية لا تقل عن عشر سنوات) ، والمؤثرات الدورية (وهي التي تحدث تغيراً في الظاهرة خلال فترة زمنية محدودة مثل الأسبوع ، أو الشهر ، أو لفصل من السنة ، أو مرة كل سنة أو سنتين أو أكثر بحيث يتكرر هذا التأثير في نفس المواعيد على مدار نفس الفترة الزمنية) ، وهناك المؤثرات غير المنتظمة التي تحدث دون انتظام وتفتقد لنمط معين أو قاعدة ثابتة، وقد تتكرر أو لا تتكرر بغير انتظام زيادة أو نقصاناً.

- الطريقة تقوم على مبدأ متوسطات حسابية متتابعة لمجموعة متتابعة ومتداخلة والنتيجة هي إزالة بعض التعرجات الموجودة في السلسلة الزمنية الأصلية لتقليل خشونة السلسلة الزمنية.

- نفترض أن هناك السلسلة الزمنية س₁ ، س₂ ، س₃ ،.....، س_ر. إذا اردنا

إيجاد معدلات متحركة لها طول (2) نقوم بالآتي:

$$س_1 + س_2 / 2 ، س_2 س_3 / 2 ، س_3 + س_4 / 2 ،.....،$$

لو أردنا إيجاد متوسطات متحركة بطول (3) نقوم بالآتي:

$$س_1 + س_2 + س_3 / 3 ، س_2 + س_3 + س_4 / 3 ، س_3 + س_4 + س_5 / 3$$

لو أردنا معدلات متحركة بطول (4) نقوم بالآتي:

$$س_1 + س_2 + س_3 + س_4 / 4 ، س_2 + س_3 + س_4 + س_5 / 4 ، س_3 + س_4 + س_5 + س_6 / 4$$

$$.....، س_4 + س_5 + س_6 + س_7 / 4$$

والقاعدة هنا هي: عدد عناصر السلسلة الاصلية = عدد الأوساط المتحركة

$$الجديدة + طول الوسط المتحرك - 1 = ك + ت - 1$$

مثال: سلسلة عدد عناصرها (50) تم تعديلها باتجاه سلسلة جديدة بطريقة

المتوسطات المتحركة بطول (5) بناء على ما سبق حدد السلسلة الجديدة (عدد

الأوساط المتحركة الجديدة).

$$\text{الحل: } ن = 50 ، ل = 5 ، ك = ؟؟$$

$$ن = ك + ت - 1$$

$$50 = ك + 5 - 1 \Leftrightarrow 50 = ك + 4 \Leftrightarrow ك = 46$$

عناصر السلسلة الزمنية (مركبات السلسلة الزمنية)

تضم :

- 1 مركبة الاتجاه العام (ت) : وتمثل المشاهدات التي تأخذ شكل متزايد مستمر مع بعض التذبذبات . ومثال لها: إزدياد التحصيل بزيادة عدد ساعات الدراسة إلا أن هذا يتأثر بقلّة التركيز والتعب. وأفضل تقدير لها عن طريق معادلة خط انحدار قيمة الظاهرة (ص) على (س): ص - أس +
- 2 مركبة الدورة (د): هي المشاهدات التي تتكرر كل أربع أو خمس فترات زمنية (فترة تغير البيانات لمدة طويلة قد تزيد عن السنة). ومثال لها ارتفاع درجة الحرارة كل (5) سنوات ، فترة الرخاء ، فترة الكساد لدورة التغير للمشاهدات.
- 3 المركبة الفصلية (ف): التغيرات التي تظهر في الفصول ، والفصول قد تكون يومية (درجات الحرارة) ، أو أسبوعية ، أو شهرية.
- 4 مركبة الخطأ أو المركبة العشوائية (خ): المشاهدات التي تتذبذب بشكل عشوائي ويستحيل تفسيرها مثل الحروب والحرائق والزلازل والبراكين. المركبة الخاصة بما تبقى من العوامل الأخرى التي يمكن أن تؤثر في السلسلة غير المركبة سابقة الذكر.

ملاحظات عامة على مركبات السلاسل الزمنية:

- 1 أن السلسلة الزمنية الواحدة يمكن أن تتضمن أكثر من مركبة واحدة من مركبات السلسلة الزمنية (اتجاه عام ، دورة ، فصلية ، عشوائية).

2 في كل سلسلة يهمننا معرفى تأثير كل مركبة من مركبات السلاسل الزمنية.

أهداف دراسة السلاسل الزمنية (استعمالها)

1 وصف السلسلة: ويتضمن المنحنى التاريخي للسلسلة ، وخواص السلسلة (تزايد أو تناقص أو ثبات)

2 تفسير السلسلة والتنبؤ ، اي الاستفادة من بيانات السلسلة في التنبؤ بالمستقبل.

تحليل السلاسل الزمنية

هو إظهار تأثير إحدى المركبات السابقة بعد إلغاء تأثير المركبات الأخرى. وتتضمن طرق تحليل السلسلة الزمنية ، إما عن طرق الوصف أو عن طريق النماذج الاحتمالية. وتتضمن النماذج الاحتمالية كل من النموذج النسبي والنموذج التجميعي.

النموذج النسبي : $ص = ت$ (مركبة الاتجاه العام) X د (مركبة الدورة) X ف (المركبة الفصلية) X خ (مركبة الخطأ).

النموذج التجميعي : $ص = ت$ (مركبة الاتجاه العام) + د (مركبة الدورة) + ف (المركبة الفصلية) + خ (مركبة الخطأ).

حساب مركبات السلاسل الزمنية

أولا طرق حساب مركبة الاتجاه العام

تتضمن طرق حساب مركبة الاتجاه العام كل من طريقة المربعات الصغرى ، التمهيد باليد ، المتوسطات المركبة ونصف السلسلة.

ومثال لذلك : الجدول التالي يمثل درجات الحرارة فى أحد المدن على مدار 10 سنوات (1986 – 1995)

السنة	86	87	88	98	90	91	92	93	94	95
د. الحرارة	7	13	19	21	27	28	32	35	39	40

1 يمكن استخدام معادلة انحدار الظاهرة (ص) على الزمن (س) بطريقة المربعات الصغرى (أنظر الباب)

2 تتضمن طريقة نصف السلسلة عمل الآتي:

أ - تقسم السلسلة إلى نصفين متساويين ، وغذا كان عدد المشاهدات فردي تحذف المشاهدة المتوسطة.

ب نجد الوسط الحسابي س ، ص لكل نصف (النصف الأول / النصف الثاني)

ت نجد معادلة خط المسنقيم المار بالنقطتين الناتجتين من الخطوة (ب).

أما إيجاد معادلة الاتجاه العام بطريقة المتوسطات المتحركة فتشمل الآتي:

1 نجد الأوساط المتحركة بطول مناسب للسلسلة لينتج لدينا سلسلة زمنية جديدة من

المتوسطات المتحركة الناتجة ليكون أثر الاتجاه العام للسلسلة الجديدة ظاهرا

بشكل أفضل من السلسلة الاصلية.

2 نحدد معادلة الاتجاه العام للسلسلة الجديدة بإحدى الطرق السابقة (معادلة

الانحدار ، التمهيد باليد ...الخ).

تقدير المركبات الفصلية

هو إيجاد قيمة الظاهرة باعتبار المركبة الفصلية لا تتأثر إلا بالموسم.

وتتضمن الطرق كل من النسب للمعدل المتحرك ، والنموذج النسبي: ص = ت (مركبة الاتجاه العام) X د (مركبة الدورة) X ف (المركبة الفصلية) X خ (مركبة الخطأ).

وتشمل القوانين الخاصة بإيجاد المركبة الفصلية بطريقة النسب للمعدل المتحرك:

1 المتوسط الموسمي = المجموع الموسمي لكل ربع / عدد السنوات = المتوسطات الموسمية

2 المتوسط الكلي = مجموع المتوسطات الموسمية / عدد الأرباع = مجموع المجموع الكلي للمعدل الموسمي / عدد الأرباع

3 النسب الموسمية = المتوسط الموسمي / المتوسط الكلي X 100 % = الجزء / الكل X 100%

تاليا هو انتاج مصنع خلال 5 سنوات حيث أن كمية الانتاج مأخوذة كل 3 شهور

السنوات	76	77	78	79	80
ربع السنة الأول	7	12	8	20	25
ربع السنة الثاني	9	11	13	21	27
ربع السنة الثالث	10	14	15	23	28
ربع السنة الرابع	15	20	16	19	27

المطلوب : 1- إيجاد النسب الموسمية لهذا الانتاج باستخدام فكرة النسب للمعدل المتحرك ، 2- أحسب المتوسط الموسمي الخاص بكل ربع ، 3- أحسب المتوسط الموسمي العام (الكلي).

الحل

الربع	المجموع الموسمي لكل ربع	المتوسط الموسمي	النسب الموسمية
الأول	72=25+20+8+12+7	$\frac{5}{72}$ 14.4	$= \%100 \times 16.5/14.4$ %87.27
الثاني	81	$\frac{5}{81}$ 16.2	$\%100 \times 16.5 / 16.2$ %98.18 =
الثالث	90	$\frac{5}{90}$ 18	$= \%100 \times 16.5 / 18$ %109.09
الرابع	87	$\frac{5}{87}$ 17.4	$\%100 \times 16.5 / 17.4$ %105.45 =
المجموع		66	

المتوسط الكلي = $16.5 = 5 / 66$

الفصل السابع

العلاقات والارتباطات

يمكن دراسة العلاقة بين أي ظاهرتين أو متغيرين بإتباع إما طريقة الارتباط أو طريقة الاحدار. الارتباط هو وجود علاقة بين ظاهرتين بحيث إذا تغير أحدهما في اتجاه معين ، مال الآخر إلى التغير في نفس الاتجاه أو في الاتجاه المعاكس. ويسمى الارتباط في الحالة الأولى إرتباطا موجبا (طرديا) ، بينما يسمى في الحالة الثانية إرتباطا سالبا (عكسيا). ومن أمثلة النوع الأول الارتباط بين معدل الذكاء والتحصيل العلمي، والنوع الثاني الارتباط بين المستوى العام لنشاط الفرد ومدة تعضه للجوع أو العطش. ويعني انعدام الارتباط بين ظاهرتين أو متغيرين أن اتجاه التغير في قيم الظاهرة الأولى لا يدل على اتجاه التغير في قيم الظاهرة الثانية في كل الأحوال، والعكس صحيح. ويعني هذا أن قيم الظاهرة الثانية إذا مالت إلى الزيادة مثلا ، فقد تميل قيم الظاهرة الثانية إلى الزيادة أو النقصان ، أو قد تبقى على حالها دون تغيير. ويدل هذا على أن انعدام الارتباط بين ظاهرتين أو متغيرين لا يفيد بأن التغير في قيم أحدهما يفيد في معرفة التغير أو التنبؤ باتجاه التغير في قيم المتغير الآخر.

إن وجود ارتباط أو علاقة بين ظاهرتين ، سواء كان ارتباطا موجبا أو سالبا، لا يعني بالضرورة أن كل زوج من قيمهما المتناظرة يجب أن تتغير تبعا لتلك العلاقة. فقد شذ بعض القيم عن ذلك ،رغم أنه من الضروري نظريا في حالى الارتباط الموجب بين متغيرين او ظاهرتين أن تتغير قيمهما المتناظرو معا في نفس الاتجاه سلبا أو إيجابا. فالقيم التي قد تشذ عن هذه القاعدة في تغيرها تتناقص في وضع يجب أن تتزايد فيه ، أو

تتزايد في وضع يجب أن تتناقص فيه. وينطبق نفس الوضع على القيم ذات العلاقة السالبة. وهذا الوضع يكثر في حالة القيم الفردية ، ولكن بحكم أنالاستتباط في القواعد الإحصائية العامة يعني بالحالات والاتجاهات الجماعية لذلك لا تؤخذ هذه القم الشادة في الاعتبار عند تقرير الصوة العامة للارتباط.

وتنقسم العلاقات من حيث الطريقة التي يتم فيها قساس المشاهدات المأخوذة على الظواهر المختلفة الداخلة فيها إلى الارتباط correlation والإقتران association والتوافق contingency. يشير الارتباط إلى العلاقة بين الظواهر التي يمكن قياس المشاهدات المأخوذة عليها والتعبير عنها بصورة رقمية. ومثال لذلك العلاقة بين معدل الذكاء والتحصيل الدراسي ، أو بين الوزن والطول. أما الغقتران فيشير إلى العلاقة بين الظواهر بالتب لا يمكن قياس المشاهدات المأخوذة عليها والتعبير عنها صورة رقمية ، وغنما بصورة وصفية فقط. ومن أمثلة ذلك العلاقة بين مستوى الذكاء بدرجاته مرتفع جدا ، مرتفع ، متوسط ، مقبول والمستوى الاجتماعي بتقسيماته إلى عالي ومتوسط ومنخفض. ويشير التوافق إلى العلاقة بين الظواهر التي يمكن قياس المشاهدات المأخوذة على القسم الآخر بطريقة وصفية. ومن أمثلته العلاقة بين الجنس والفرد ومقدار تحصيله الدراسي ، أو العلاقة بين نوع العمل والمرتب الشهري للموظف أو العامل.

يقسم الارتباط من حيث عدد الظواهر الداخلة فيه ، وهو الارتباط الذي يتعلق بالظواهر التي يمكن قياسها والتعبير عنها بصورة رقمية ، إلى ثلاثة أقسام وودلك بالنسبة إلى عدد المتغيرات الداخلة فيه إلى الارتباط البسيط ، والارتباط المتعدد ، والارتباط الجزئي. يبحث الارتباط البسيط في العلاقة بين ظاهرتين أو متغيرين اثنين فقط ، بينما يبحث الارتباط

المتعدد في العلاقة بين متغير معين ومتغيرين آخرين أو أكثر مأخوذين معا. وفي الجانب الآخر يبحث الارتباط الجزئي في العلاقة بين متغيرين إثنين فقط من بين عدة متغيرات على فرض أن تأثير بقية المتغيرات الأخرى يبقى ثابتا.

وللمساعدة في تحديد المعادلة التي تربط بين المتغيرات تتبع الخطوات التالية:

1 جمع البيانات التي تظه القيم المقابلة للمتغيرات تحت الدراسة ولنفرض أنهما المتغيران س و ص.

2 وضع النقط في رسم بياني فيه الإحداثيات المتعامدة وتسمى النقط الناتجة بشكل الانتشار.

3 من شكل الانتشار يمكن بالنظر تمهيد منحنى لتقريب هذه البيانات ويسمى بالمنحنى التقريبي.

معامل الارتباط

الارتباط بين ظاهرتين س ، ص

معامل ارتباط بيرسون

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

مثال : إذا كان

س	13	9	19	15	11	8	16	11
ص	15	7	17	15	10	9	14	10

س	ص	س ⁻ = ص ⁻	ص ⁻ = ص ⁻	ص ⁻ = ص ⁻	ص ⁻ = ص ⁻	ص ⁻ = ص ⁻
		10-	10-			
13	15	3	5	9	25	15
9	7	1-	3-	1	9	9
19	17	9	7	81	49	63
15	15	5	5	25	25	25
11	10	1	0	1	0	0
8	9	2-	1-	4	1	2
16	14	6	4	36	16	24
11	10	1	0	1	0	0
		22	17	158	125	132

$$r = \frac{\sqrt{(8 \times 17 - 132)^2 + (8 \times 22 - 158)^2}}{(8 \times 17 - 132)^2 + (8 \times 22 - 158)^2} = 0.82 \text{ (طردي قوي)}$$

خصائص معامل الارتباط

لو أخذنا خطا مستقيما وقسمناه إلى نصفين متساويين ، فإن قيمة ر يمين الصفر (الذي يقسم الخط إلى قسمين متساويين) تتجه نحو ارتباط موجب تام (تام طردي) ، وإلى يسار الصفر نحو ارتباط سالب تام سالب. وبالتدرج من الصفر ناحية اليمين نجد ارتباط موجب ضعيف (عند صفر) ، يليه ارتباط موجب متوسط (عند 0.6) ، ثم ارتباط موجب

قوي (عند 0.8). ونجد عكس ذلك إذا اتجهنا ناحية اليسار من صفر التقسيم (ارتباط سالب ضعيف، ارتباط سالب متوسط، ارتباط سالب قوي).

معامل ارتباط الرتب لسبيرمان

القانون : $r = 1 - 6 \text{ مج ف}^2 / \text{ن} (1 - \text{ن}^2)$

مثال أوجد رتب س التي قيمتها معطاة في الجدول التالي:

رياضيات	A	C	C	C	C	B	D
إحصاء	B	B	B	D	C	A	E

نرتب القيم

س	ص	رتب س	رتب ص	الفرق = س - ص	ف ²
A	B	7	5.5	1.7	2.25
C	B	3.5	5.5	-2	4
C	B	3.5	2	1.5	2.25
C	D	3.5	3.5	0	0
C	C	3.5	3.5	0	0
B	A	6	7	-1	1
D	E	1	1	0	0

$r = 1 - 6 \text{ مج ف}^2 / \text{ن} (1 - \text{ن}^2) = 1 - 9.6 \times 6 / 7 (1 - 7^2) = -0.6$

(عكسي ضعيف)

مثال 2 : أحسب معامل ارتباط الرتب لسبيرمان للبيانات أدناه:

س	0	1	2	3	4	5	2
ص	1-	2	2	8	4	14	5

الحل

س	ص	رتب س	رتب ص	الفرق	الفرق ²
0	1-	1	1	0	0
1	2	2	2.5	0.5-	0.25
2	2	3.5	2.5	1	1
3	8	5	6	1-	1
4	4	6	4	2	4
5	14	7	7.5	0	0
2	5	3.5		1.5-	2.25
8.5					

$$r = -1 = 6 \text{ مج ف}^2 / (n - 2) = (1 - 2) 7 / 8.5 \times 6 - 1 = -0.85 \text{ (طردي قوي)}$$

خط الانحدار

يختص الانحدار بدراسة العلاقة بين متغيرين أحدهما مستقل والآخر متغير تابع. ويعني هذا أن المتغير المستقل مثبت عند مستويات معينة بينما يأخذ المتغير التابع قيما مختلفة عند كل مستوى من مستويات المتغير المستقل. ويعتبر شكل الانحدار من الأمور

الواجب معرفتها للحصول على فكرة عن نوع الارتباط بين هذين المتغيرين ، إضافة إلى درجة قوة هذا الارتباط ومعرفة طبيعة نقط الانتشار التي قد تقع على خط مستقيم تماما أم لا. ويستعمل شكل الانحدار عادة لتمثيل أزواج من القيم للظايرتين تحت الدراسة، حيث تتحدد مثلا قيمة س على أساس القيمة المناظرة لها في ص وهكذا، وينتج عن انتار النقط بين المحورية السيني والصادي المتاعدين ما يعرف بالشكل الانتشاري والذي يأخذ أشكالا مختلفة تبعا لاختلاف العلاقة بين قيم س وقيم ص حيث يتحدد شكل العلاقة تبعا لذلك سواء كانت طردية أو عكسية.

وخط الانحدار هو خط توفيق يمر بأكبر عدد من النقاط ويتوسط الباقي منها ويتم الحصول عليه عادة باستخدام الطرق الجبرية وأهما طريقة المربعات الصغرى والتي تستخدم لتحقيق هذا الهدف ، أي لتوفيق أفضل خط مستقيم (أو منحنى) لمجموعة من البيانات بحيث يكون مجموع انحراف النقط عن الخط المستقيم أصغر ما يمكن قدر الإمكان. ومن بعض خواص معامل الانحدار أن حاصل ضرب معامل انحدار ص على س (أي ص متغيرا مستقلا وس متغيرا تابعا) في معامل س على ص (أي س متغيرا مستقلا و ص متغيرا تابعا) يساوي مربع معامل الارتباط (r^2)، وأن خطي الانحدار يتقاطعان في نقطة احداثياتها متوسط قيمة س ومتوسط قيمة ص (أي س⁻ ، ص⁻).

خط انحدار س على ص ← ص =

$$م س + ج مج س ص = م مج س^2 + ج مج س$$

$$مج ص = م مج س + ن ج \quad (1)$$

$$ج مج س = مج س ص - ج مج س^2$$

$$\begin{aligned}
& \text{مج س مج ص} = \text{م (مج س)}^2 + \text{ن ج مج س} \\
& \text{مج س مج ص} = \text{م (مج س)}^2 + \text{ن (مج ص - م مج س)}^2 \\
& = \text{م (مج س)}^2 + \text{ن مج س ص} - \text{ن م مج س}^2 \\
& \text{مج س مج ص} - \text{ن مج س ص} = \text{م (مج س)}^2 - \text{ن مج س}^2 \\
& = \text{م [(مج س)}^2 - \text{ن مج س}^2] \\
& \text{م} = \text{مج س مج ص} - \text{ن مج س ص} / \text{(مج س)}^2 - \text{ن مج س}^2 \\
& = \text{ن مج س ص} - \text{مج س مج ص} / \text{ن مج س}^2 - \text{(مج س)}^2 \\
& \text{ومن (1) : ج} = \text{مج ص} - \text{م مج س} / \text{ن} \\
& \text{ج} = \text{مج ص} - \text{م مج س} / \text{ن}
\end{aligned}$$

$$\text{ص} = \text{م س} + \text{ج}$$

مثال : أوجد خط انحدار ص على س من المثال السابق

الحل

$$\text{مج س} = 17, \text{مج ص} = 24, \text{مج س ص} = 126, \text{مج س}^2 = 59, \text{ن} = 7$$

$$\text{ص} = \text{م س} + \text{ج}$$

$$\text{م} = \text{ن مج س ص} - \text{مج س مج ص} / \text{ن مج س}^2 - \text{(مج س)}^2$$

$$= 17^2 - 59 \times 7 / 34 \times 17 - 126 \times 7 =$$

$$\text{ج} = \text{مج ص} - \text{م مج س} / \text{ن} = 34 - 2.45 / 7 = 1.09$$

$$\text{خط انحدار ص على س} = \text{ص} = 2.45 \text{ س} + 1.09$$

الجدول التالي يمثل درجات الحرارة والمبيعات في إحدى محطات المحروقات ،
والمطلوب:

أ) أوجد خط انحدار ص على س

ب) أوجد خط انحدار س على ص

ت) أوجد قيمة المبيعات عندما تكون درجة الحرارة 45 درجة مئوية

الحل

ص	ص	ص = س -	ص = س -	ص = س -	ص = س -	ص = س -
30	30	30	30	30	30	30
25	44	5-	14	70-	25	196
30	45	0	15	0	0	225
32	33	2	3	6	4	9
33	38	3	8	24	9	16
35	30	5	0	0	25	0
40	27	10	3-	30-	100	9
37	41	7	11	77	49	121
		22	48	7	212	576

ص = م س + ج

$$م = \frac{ن \text{ م س} - \text{م س} \times \text{م س}}{ن \text{ م س} - (\text{م س})^2} = \frac{48 \times 22 - 212 \times 7}{48 \times 22 - 212^2} = 1 -$$

$$\begin{aligned} \text{ج} = \text{مج ص}^- + \text{م مج س}^- &= 48 = 7 / 22 \times 1 + 10 \\ \text{ص}^- = \text{س}^- &= 10 + 30 = 10 + (30 - \text{س}) = 10 + 30 \\ \text{ص}^- &= 10 + 10 + 30 \end{aligned}$$

خط انحدار ص على س	$\text{ص}^- = 70 - \text{س}$
-------------------	------------------------------

$$\text{ص} = 45 - 70 = 15$$

اذن المبيعات = 1500 جنيه

$$\text{س}^- = \text{أ ص}^- + \text{ب}$$

$$\begin{aligned} \text{أ} = \text{ن مج س}^- \text{ ص}^- - \text{م مج س}^- \text{ مج ص}^- / \text{ن مج س}^- - (\text{م مج ص}^-)^2 \\ = 7 \times 7 - 48 \times 22 / 48 - 567 \times 7 / 48 - 48^2 = 0.58 \end{aligned}$$

$$\text{ب} = \text{م مج س}^- - \text{أ مج ص}^- / \text{ن} = 22 - 7 / 48 \times 0.58 + 7.12$$

خط انحدار س على ص = $\text{س}^- = 0.58 - \text{ص}^- + 7.12$

$$\text{س}^- = 0.58 - (\text{ص}^- - 30) + 7.12$$

$$\text{س}^- - 30 = 0.584 + 0.58 \times 30 + 7.12$$

$$\text{خط انحدار س على ص} = \text{س}^- = 0.584 + 54.52$$

قائمة المراجع

محمود المشهداني، محمد , هرمز، 1989 : مبادئ الإحصاء، جامعة بغداد.

محاضرات في الإحصاء الجغرافي

الاستاذ الدكتور مضر خليل عمر الكيلاني

<https://sites.google.com/site/mutharalomar/>