كلية الآداب
جامعة جنوب الوادي
قسم الجغرافيا ونظم المعلومات الجغرافية

الإحصاء الجغرافي

إعـــــــــــداد

أ.د/ محمد أحمد إبراهيم نعينع

أستاذ ورئيس قسم الجغرافيا ونظم المعلومات الجغرافية كلية الآداب – جامعة جنوب الوادي

د/ شيماء أحمد عبدالله رضوان

مدرس بقسم الجغرافيا ونظم المعلومات الجغرافية كلية الآداب – جامعة جنوب الوادي

قنا
العام الدراسي 2022 / 2023م
1444هـ

محتويات الكتاب

مقدمة

طبيعة البيانات الإحصائية والجغرافية

ارتبط علم الإحصاء منذ نشأته بعمليات العد التي كانت تجريها الدول لحساب أعداد جيوشها، والضرائب التي تجبى من المزارعين، وجمع المعلومات من 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 و تفسير 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، وثانيًا: صياغة الفرضيات بهدف تفسير الظاهرة، وثالثا: 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 زمانية أو مكانية معينة دون الحاجة لتعميمها على ظواهر أخرى، من خلال الاهتمام بأساليب جمع البيانات وتبويبها وعرضها. ومن أهم وسائله استخدام الأشكال الهندسية للبيانات والجداول الإحصائية للبيانات والتوزيعات التكرارية، والدراسة الرياضية للبيانات من خلال استخدام مقاييس النزعة المركزية، ومقاييس التشتت، ومقاييس الالتواء، أما الإحصاء التحليلي (الاستدلالي)، فيحدد طرق الحصول على البيانات من المجتمع الإحصائي من خلال اخذ عينة من المجتمع بأساليب إحصائية، وبالتالي قد لا تمثل هذه العينة المجتمع كله تمامًا، 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 تعبيرًا كميًّا، وتستخدم من قبل الباحثين في جميع فروع المعرفة الأخرى؛ لأنها توفر أسلوبًا علميًّا موضوعيًّا محايدًا دون تمييز أو تدخل من قبل الباحث. ويجب أن يراعي الباحث عند تصميم بحثه تحديد مشكلة البحث هدف البحث بشكل واضح ودقيق. لا بد 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 التحاليل الإحصائية في الجغرافيا نتائج دقيقة جدًّا، ولكن من الضروري الانتباه إلى تحقق الآتي:-

أ- الدقة Precision

تشير الدقة إلى المستوى التفصيلي في القياس، وترتبط بعملية موازنة جهاز القياس وتقويمه، مثل جهاز قياس كمية المطر، فعند استعمال جهازين مختلفين في الدقة لقياس كمية المطر؛ فان الأقل دقة قد يسجل الكمية (١.٢) بوصة، والأخر قد يسجل (١.٢٦) 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 Curacy

ترتبط عملية الضبط بانحراف عملية القياس عن النظام بصيغته الأوسع. فعندما يكون جهاز القياس دقيق جدًّا، إلا انه غير مضبوط، فبالعودة إلى جهاز قياس المطر المشار إليه آنفا، فعندما يكون تنظيم القياس غير صحيح فقد تكون الكمية المقاسة (١.١٩) بدلا من (١,٢٦) فالنتيجة دقيقة، ولكنها غير مضبوطة في درجتها. و لسوء الحظ فإن اكتشاف درجة الدقة ومستوى الضبط في الأجهزة غير سهل. وباعتماد عدد من المقاييس المختلفة في درجة دقتها و ضبطها؛ فان النتائج تكون في النهاية غير دقيقة ما لم يتم تعيير الأجهزة وتوحيد درجة دقتها و ضبطها.

جـ - الصلاحية Validity

في العديد من المشكلات الجغرافية يكون التوزيع المائي أو نمط المواقع قيد التحليل ناتج عن عمليات معقدة. فعندما يكون المفهوم الجغرافي معقدا وغير واضح فإن التعبير عنه يكون ضعيفًا. وفي الجغرافيا ترد الكثير من المفاهيم الصعبة القياس، متعددة الأوجه ومعقدة، مثل: مستوى الفقر، توعية 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 ودراسات أخرى، ويبقى السؤال، هل التعريف العملي الإجرائي صالح أم لا؟

ومن الواضح فإن درجة الصلاحية في الكثير من المشاكل التي يدرسها الجغرافيون يصعب تقييمها.

د- درجة الثقة Reliability

عندما تحدث تبدلات في الأنماط المكانية عبر الزمن فان تحليلها يتطلب الإجابة عن مجموعة من التساؤلات المتعلقة باستقرارية البيانات وتبويبها من الناحية المكانية في الوحدات الإحصائية الإدارية. فعلى سبيل المثال، فإن اختبار النمط المكاني والتبدلات التي حصلت فيه خلال (٢٠) عينة يتطلب معرفة التبدلات التي حصلت في الوحدات الإدارية خلال هذه المدة، تبرز هذه المشكلة بحدة أكثر عند مقارنة إحصائية لدول مختلفة، فالتباين ناجم عن الاختلاف في درجات الدقة والضبط وفي طبيعة البيانات ودرجة الاعتماد عليها. وفي بعض الأحيان يكون الأمر كذلك في الدولة الواحدة، بين الإقليم، أو الأقسام الإدارية والإحصائية 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 محددة.

ب- مجتمع غير منتهي (غير محدد): وهو المجتمع الذي 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 والإنسانية 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 نظامًا واحدًا يكون بيئة الكرة الأرضية.

ولدراسة هذه الظواهر تنتهج الجغرافيا نهجًا علميًّا منظمًا ذاتيًّا وموضوعيًّا، يتميز بالشمول والتعدد 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:

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

٢) متنوعة المقاييس، فقد تكون مفردة ومجدولة، متصلة، متقطعة، اسميه، رتبيه.

٣) لا ينفصل المكان والزمان في المعلومة الجغرافية. فكل معلومة مكانية ترتبط بزمن تمثله، وفي المقابل فإن المعلومة الزمنية يجب بالضرورة أن تحدد مكانيًّا ليتسنى الاستفادة منها جغرافيًّا.

٤) متنوعة المصادر الأولية، والثانوية، والثانوية متعددة المصادر، والأرضية، والفضائية. وقد تعاظم تنوع هذه المصادر بتوافر بيانات من مصادر خارجية، مثل المنظمات الدولية، والاستشعار عن بعد، وبنوك المعلومات، وغيرها، وبتطور تقنيات خزن المعلومات (ورقية، آلية، مساحية raster - vector خطية).

٥) أدي التراكم الكمي للمعلومات المكانية إلى زيادة الاهتمام بالبعدين الثالث والرابع في الدراسات الجغرافية المساحة - بعدين، تكملها الكثافة، الزمن).

مصادر البيانات الإحصائية

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

١- المصادر المباشرة (الميدانية أو الأولية) وتشمل كل البيانات التي يجمعها الباحث في حينها من مصادرها الأصلية بأي وسيلة كانت سواءً 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 المشاهدة والملاحظة والتسجيل والاتصال الهاتفي والمقابلة الشخصية والاستبيان.

٢- 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 تنفك 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 والعمليات التي تؤطرها. ويمكن تصنيفها في المصادر التالية:

أ- المصادر المكتبية

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 والمعلومات عن الموضوع المطلوب تقصيه. وقد دخلت الكثير من هذه المصادر في الشبكة الدولية للمعلومات فأصبحت قريبة جدًّا 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 إلى ذلك، فإن تصنيف منطقة الدراسة حسب الوحدات الإحصائية التي جمعت عنها المعلومات رسميًا يساعد كثيرًا 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 تاريخ الظاهرة أو المنطقة قيد الدراسة. وقد تتوفر المعلومات التاريخية إما على شكل معلومات (بيانات) أولية أو ثانوية. كما وهناك المصادر الخرائطية الرسمية التي تصدرها هيئة المساحة العامة في كل الدول.

جـ - المصادر الميدانية

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

(١) اختبار وتحليل ميداني لجزء من البلاد، يسهل الوصول إليه لتوضيح واحد أو أكثر من معطيات التباين المكاني.

(٢) المختبر الحقيقي للجغرافيا خارج قاعات الدرس.

(٣) أفضل طريقة لتعلم الحقائق.

(٤) تعود الطالب على ملاحظة الأشياء، وتطوير خبرة الملاحظة، وتفسير ما يراه؛ لأنها أسلوب رئيسي وأساسي لا يمكن للجغرافي الاستغناء عنه، وتكسب الطلبة للمفردات الجغرافية اعتمادًا على الملاحظة المباشرة.

(٥) لأن العمل المنجز في الحقل الميداني يشعل المخيلة ويحفزها لدراسة الجغرافيا في قاعات الدرس، ويقود إلى تعظيم الأفكار الجغرافية الجوهرية، ويطور بذلك ملكة النقد عندهم.

(٦) تطور النظرة للبيئة المحلية و البلد، وتعود على التفكير بالمشاكل من أجل حلها.

(٧) تتطلب نوعية وقدرة عقلية مختلفة عن تلك التي تطورت من خلال التعلم من الكتب والمحاضرات.

(٨) توسع دائرة الخبرة المرئية والنجاح في استيعاب الجغرافيا اعتمادًا على قدرة الطالب لتشكيل الصور الذهنية عن الأماكن.

(٩) تعمل على اتصال الطالب مباشرة مع الحقيقة، والانغماس شخصيًّا بالدراسة، وامتلاك هذه المعرفة، مما يجعله أكثر قدرة على الاتصال، وأكثر تقديرًا وإدراكًا لعمله.

(١٤) أفضل طريقة لدراسة الجغرافيا.

د- نظم المحاكاة

يقصد به نمذجة الواقع، أو جزء منه بهدف الاستيعاب والدراسة والتحليل، واستشفاف الحالات الممكنة والمتوقعة، 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، والتنبؤ بما سيكون عليه الحال مستقبلًا. وكانت تسمى النمذجة الساكنة والتي تحولت إلى النمذجة الدينامية أو المتحركة بدخول التكنولوجيا الحديثة، وبتراكم المعلومات مكانيًّا وزمانيا، 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 بثلاثة 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 والخارجية المؤثرة عليه، وهنا تلعب نظم المعلومات الجغرافية دورًا مهمًا.

أنواع البيانات الإحصائية

تقسم البيانات الإحصائية إلى مجموعتين:

١- **البيانات النوعية:** تقيس ظاهرة من الظواهر دون أن تأخذ قيمًا 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 اسمية: تعتمد على التصنيف النوعي بغض النظر عن أهمية الترتيب. مثلا: تصنيف موظفي إحدى الشركات حسب الجنسية أو حسب التخصص.

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

٢- **البيانات الكمية:** تأخذ قيمًا عددية صحيحة أو كسرية حسب ظروف الحالة، ويمكن قياسها بوسائل قياس مألوفة مثل عدد المرضى الراقدين في إحدى المستشفيات ، عدد رؤوس الماشية في قطيع معين، درجات الطلبة في كلية، أطوال الأشخاص بالسنتمترات، أوزان الأشخاص بالكيلوغرامات، ودرجات الحرارة في مدينة معينة. وتقسم المتغيرات الكمية إلى قسمين هما:-

أ- المتغيرات المستمرة (المتصلة): إذا كانت مجموعة القيم التي يأخذها المتغير العشوائي (المتغير العشوائي هو دالة ذات قيمة حقيقية معرفة على فضاء العينة) مجموعة غير قابلة للعد سواء كانت محدودة أو غير محدودة، وإنما تشكل قيمًا واقعة ضمن فترات. وهذا يعني وجود عدد غير منتهٍ من القيم مثل: كمية الأمطار المتساقطة على منطقة ما خلال سنة معينة، أسعار سلعة معينة في فترة زمنية معينة، وغيرها. وتعتمد البيانات الكمية المستمرة على وحدات القياس التي تأخذ قيمًا في مجال تغيراتها. مثلًا وحدة قياس الطول إما أن تكون بالمتر أو السنتيمتر بفرض أن طول أحد الطلاب يساوي ١٥٩ سنتيمتر وبمعنى أدق يساوي ١٥٩,٤ سنتيمترًا أو أكثر دقة يساوي ١٥٩,٤٥ سنتيمترًا.

ب- بيانات كمية منقطعة: وهي البيانات التي تأخذ قيمًا عددية صحيحة، وقد تكون مجموعة القيم التي يأخذها المتغير العشوائي مجموعة قابلة للعد أي يمكن عدها سواء أكانت مجموعة محدودة أو غير محدودة، مثل عدد أشجار النخيل في قرية معينة، أو عدد طلبة الصف الأول في مدرسة معينة ، و عدد موظفي شركة ما خلال نصف القرن الماضي ، وغيرها.

أنواع البيانات الجغرافية

يمكن الاستفادة من أنواع البيانات الإحصائية في تحديد أنواع البيانات الجغرافية. فالأساس الأول لتصنيف البيانات الجغرافية ينبع من طبيعة تتابعها المكاني أو الزماني. فالبيانات الجغرافية الزمانية تكون مرتبطة بمنطقة واحدة لسلسلة زمانية قد تكون متصلة أو متقطعة. 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، ولكل فئة أو مجموعة اسم أو عنوان (المقياس الإسمي). وليس هناك علاقة افتراضية بين الفئات سوى أنها مختلفة عن بعضها البعض. ويصنف الجغرافيون المتغيرات الاسمية بطرائق عديدة، فمثلًا: الأفراد قد يتم التمييز بينهم على أساس الدين، الجنس، العرق، وتصنف المدن على أساس الوظيفة التي تؤديها، وهكذا. ويعتمد التحليل الإحصائي لهذا المقياس على الاختبارات الإحصائية اللامعلمية non parametric، مثل: مربع كاي، كولموكروف - سمرنوف، معامل فاي، معامل يول، ومعامل الجوار.

أما البيانات الكمية الجغرافية فتخضع للعمليات الحسابية العادية (الجمع والطرح والقسمة والضرب)، 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، وغيرها.

وهناك مستوى آخر من القياس للبيانات الجغرافية ينتج من ترتيب القيم بناء على الحجم أو الوزن لتحديد أيها أكبر من، وأيها أصغر من. والرتب الناتجة هنا هي ليست فئات، بل يبنى الترتيب بناء على قوة أو ضعف المتغيرات. وهناك طرائق إحصائية تستخدم لترتيب القيم والمتغيرات تضم معامل سبيرمان، و اختبار مان وتني... ومن أمثلتها تصنيف الدول في شكل رتب بناءً 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 (٣٠) درجة مئوية وفي موقع آخر (١٥) درجة مئوية، فإن هذا لا يعني أن الموقع الأول درجة حرارته ضعف الموقع الثاني، بل إنه أكثر حرارة. ومن الضروري أن ينتبه الجغرافي إلى ذلك عند المقارنة والقياس وتفسير النتائج.

ويمكن القول بأن 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 نسبيًّا. ويسمى هذا بمقياس الفاصلة، حيث يحدد أصل القياس (الصفر) اعتباطيًّا، مثل: الصفر المئوي والفهرنهايتي. 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 حياديًّا (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 وأسلوب التحليل المتبع. هناك أسلوبان لجمع البيانات، أول الأسلوبين هو أسلوب الحصر الشامل الذي يتناول دراسة كافة مفردات المجتمع الإحصائي، مثل (التعداد السكاني)، أي جمع بيانات عن جميع المفردات التي تؤلف المجتمع الإحصائي، هنا يجب أن يكون المجتمع محددًا. ومن مزايا هذا الأسلوب دقة النتائج وعدم وجود أخطاء عشوائية، أما عيوبه فهي ارتفاع تكاليفها (الجهد، المال)، وعدم إمكانية تطبيقها على المجتمعات ذات المفردات كبيرة الحجم، وثاني الأساليب هو أسلوب المسح باستخدام العينة الذي يتناول جزءًا من المجتمع لتمثيله بطريقة العينة العشوائية، بشرط أن تكون العينة ممثلة تمثيلًا صادقًا دون تحيز، ويعني هذا جمع البيانات والمعلومات عن جزء من المفردات التي تؤلف المجتمع الإحصائي. وهو أسلوب مفيد في المجتمعات غير المحدودة، كما وتحتاج إلى وقت وجهد وموارد مادية وبشرية أقل مما يحتاجه أسلوب التسجيل الشامل، ومن مزاياه أيضًا انخفاض تكاليفه وإمكانية تطبيقه مهما كان حجم المجتمع. أما عيوبه فتتمثل في نتائجه 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 ٥% نسبة مقبولة لحجم العينة.

أنواع العينات

تكون العينات إما عينات عشوائية (احتمالية)، أو غير عشوائية (غير احتمالية).

(أ) العينات العشوائية (الاحتمالية) وتقسم إلى الاتي:

١- العينة العشوائية البسيطة: هي عينة تختار مفرداتها بشكل عشوائي، وتعطي نفس الفرصة لكل المفردات، حتى نصل إلى حجم العينة المطلوبة. وشرط سحبها أن يكون المجتمع الإحصائي متجانسًا، وكل مفردة في المجتمع لها نفس الفرصة في الظهور ضمن مفردات العينة.

٢- العينة المنتظمة: يتم اختيار المفردة الأولى عشوائيًّا، 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 (٥٦٠ مفردة)، وحجم العينة المطلوبة (٥٦ مفردة)، علمًا بأن رقم المفردة المختارة عشوائيًّا (٥ مفردات)، فما هي قيمة الأرقام الأخرى؟ يكون الحل كالآتي:-

كسر المعاينة = ٥٦٠ / ٥٦ = ١٠ ، وعلى ذلك تكون بقية الأرقام كالآتي: (٥)، (٥ +١٠)، (١٥ + ١٠) ... إلخ.

وشرط سحب العينة المنتظمة أن يكون المجتمع الإحصائي مرتبًا ترتيبًا تصاعديًّا او تنازليًّا 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.

٣- العينة الطبقية 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 آخذين بعين الاعتبار حجم الطبقة أو المجموعة. ويمكن توضيح ذلك

في المثال الآتي: توجد مدرسة تضم ١٠٠٠ طالب، منهم ٢٥٠ طالب ثانوي، ٥٠٠ طالب متوسط، ٢٥٠ طالب ابتدائي، و يراد أخذ عينة بحجم ١٠% 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 = ١٠% × ١٠٠٠ = ١٠٠ طالب

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

طلبة الثانوي = ٢٥٠ × ١٠٠ / ١٠٠٠ = ٢٥

طلبة المتوسط = ٥٠٠ × ١٠٠ / ١٠٠٠ = ٥٠

طلبة الابتدائي = ٢٥٠ × ١٠٠ / ١٠٠٠= ٢٥

حجم العينة المختارة = ٢٥ + ٥٠ + ٢٥ = ١٠٠

٤- العينة العنقودية أو متعددة 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 جدًّا، 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 أصغر (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 غير العشوائية:

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، وأهم 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- خطا عشوائي: هو خطأ غير مقصود يقل كلما كبرت 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 سواءً بقصد أو بغير قصد، أو يحدث هذا الخطأ أحيانا عندما يستقي الباحث معلومات بحثه ليست من مصادرها الأصلية، بل من مصادرها غير المباشرة.

جـ - خطأ الاتساق هو خطأ في وجود إجابتين توحيان بأن إحداهما خطأ، وهناك بعض الأخطاء التي تحدث عندما يقوم الباحث بجمع البيانات والمعلومات التي تخص بحثه.

د- خطأ الصدفة: هذا الخطأ يرتكبه الباحث بنفسه سواءً 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 الأول

جمع البيانات الجغرافية وتصنيفها

الفصل الأول

جمع البيانات الجغرافية وتصنيفها

Collection& Classification

أولاً: مصادر البيانات الجغرافية

- المصادر الخاصة
- المصادر العامة
- المصادر الميدانية
- العينات

ثانيًا: تصنيف البيانات

- التصنيف الكمي
- التصنيف النوعي
- التصنيف الزمني
- التصنيف الجغرافي

أولًا : جمع البيانات

تتوقف طبيعة البيانات التي يقوم الباحث بجمعها من مصادرها المختلفة على نوع البحث، والهدف الذي يرمي إليه، ونظرًا لتباين الموضوعات الجغرافية، فإن جمع البيانات لأي موضوع منها يتحدد مصدره تبعًا للأساس الذي يقوم عليه، والواقع أن البيانات أيا كان نوعها، فهي عبارة عن حقائق ومعلومات تكون مجتمعة المادة الخام للموضوع قيد البحث، وهي بمثابة الهيكل الكمي الذي يقوم الباحث بتغليفه، حسب قدراته الخاصة ومهارته، بثوب وصفي تحليلي، يرقى هذا الثوب بكل من الأسلوبين إلى المنهج التجريبي الحديث، وبقدر ما للباحث من مهارة ودراية يأتي هذا الثوب متقنًا مضبوطًا، على ذلك الهيكل الكمي، وهناك خاصيتان أساسيتان للبيانات الجغرافية وغير الجغرافية:

أ- مدى دقة البيانات، ومدى خضوعها للمعالجة الإحصائية، ومدى سهولة تخزينها، ليس فقط في صورة ملفات من الورق أو القوائم، ولكن أيضًا في نظام الكروت المثقوبة، أو شرائط التخزين، أو الحاسب الآلي.

ب- الهدف الذي جمعت من أجله البيانات، لأن تحديد الهدف يقلل من الوقت والجهد المبذولين في معالجة هذه البيانات، وما يصاحبها من قياسات ودراسات ميدانية للتأكد من مطابقة هذه البيانات للواقع أو بُعدها عنه.

وليس للبيانات الجغرافية من خصائص تميزها تمامًا عن سواها من البيانات. وإجمالاً للقول، فإن أي بيانات يمكن أن تصبح جغرافية إذا أمدتنا، ما يسهل علينا تصوير العلائق أو حل المشاكل ذات الطبيعة الجغرافية. وحين يعتمد الجغرافي على البيانات الطبيعية أو البشرية عليه أن يكون أهلاً لهذه البيانات متفهمًا الأسس والمبادئ التي تقوم عليها، ومتعلمًا الأساليب التي تمكنه من تحليلها وتفسيرها، وعمومًا الخاصية التي يمكن أن تميز البيانات الجغرافية، هي توزيعها المكاني، ومدى تأثيرها، أو تأثرها بهذا المكان.

مصادر البيانات الجغرافية:

يمكن تقسيم مصادر البيانات الجغرافية إلى ثلاثة مصادر أساسية: -

١- **المصادر الخاصة**: وهي المصادر وثيقة الصلة بموضوع البحث. فمن المعروف أن كل مصلحة من المصالح، أو هيئة من الهيئات، أو مصنع من المصانع، أو مجمع من المجمعات الصناعية أو التجارية، أو وزارة من الوزارات، أو أمانة من الأمانات، توجد بها بيانات تخص القطاع الذي يدخل في نطاق خدمتها. ولهذا إذا طرق الباحث بابًا منها سوف يحصل على بعض البيانات الخاصة بموضوعه.

٢- المصادر العامة: وتشمل الموسوعات العلمية، ودوائر المعارف، مثل دائرة المعارف البريطانية، أو الأمريكية، أو الكندية، أو الفرنسية، أو موسوعة العلوم الاجتماعية، أو الكتب السنوية، والتقاويم (Almanacs) التي تمد الباحث بالإحصاءات الخاصة عن العلوم الجغرافية، والتاريخية، والسياسية، والاقتصادية، والاجتماعية التي تتناسب مع بحثه، وكذلك الحوليات (Annals) العالمية، والمحلية، والأطالس، وهي الأساس الذي يعتمد عليه الجغرافي في بحثه. والرسائل العلمية العامة والمتخصصة، من المصادر الهامة، وخاصة إذا كانت في مجال التخصص الذي اختاره الباحث، فتصبح في هذه الحالة مصدر من المصادر وثيقة الصلة بموضوع البحث. وكذلك المطبوعات والنشرات والتقارير الحكومية التي تصدر عن السكان، الزراعة، التجارة، الصناعة، والخدمات. ويقوم الجهاز المركزي للتعبئة العامة والإحصاء في مصر بإصدار هذه المطبوعات، وتصدر مصلحة الاقتصاد السياسي والإحصاء والتشريع مجلة مصر المعاصرة. كما تصدر الجمعية الجغرافية المصرية مجلتين واحدة باللغات الأجنبية، والأخرى باللغة العربية. وتصدر بعض الدول العربية الأخرى مجلات جغرافية، ولا سيما سوريا، وتونس، والجزائر، والعراق، والكويت. وتقوم وزارة المالية والاقتصاد الوطني بالمملكة العربية السعودية بعامة، ومصلحة الإحصاءات العامة بخاصة بإصدار كافة الإحصاءات التي تخدم جميع الغايات، هذا إلى جانب الحوليات الجامعية المنتظمة التي تصدرها الجامعات داخل المملكة. وأخيرًا هناك مطبوعات ونشرات هيئة الأمم المتحدة، والمعاجم الجغرافية، فقد صدر في مصر، على سبيل المثال، معجم المصطلحات الجغرافية ليوسف تونى من جامعة عين شمس عام ١٩٦٤م، وكذلك الدوريات العالمية التي تهتم بالدراسات والأبحاث الجغرافية بمختلف اللغات منها:-

أ. الدوريات البريطانية :

1- Geographical Journal.
2- Geographical Magazine.
3- Journal of the Manchester Geographical Society.
4- Journal of the Royal Geographical Society.
5- Middle East Economic Digest.
6- Proceedings of the Royal Geographical Society
7- Quarterly Economic Review.
8- Quarterly Journal of the Geological Society of London.
9- Scottish Geographical Magazine

10- World Affairs.

ب. الدوريات الأمريكية:

1- Economic Geography.
2- Geographical Review.
3- National Petroleum News.
4- Political Science Quarterly.
5- United Nations Review.

جـ ـ الدوريات الفرنسية:

1- Annales de Geographie.

دـ الدوريات الكندية:

1- Canadian Geographical Journal.

ويعتبر الكتاب السنوي للسكان "Demographic Yearbook"، والكتاب السنوي للإحصاء "Statistical Yearbook"، من أهم المصادر الخاصة بالبيانات السكانية والاقتصادية على المستوى العالمي، إلى جانب كتب منظمة الفاوFAO المختلفة.

٣- المصادر الميدانية: ويقوم الباحث بجمعها بعدة وسائل منها: استمارة الاستبيان (Questionaire)، وهي تتضمن مجموعة من الأسئلة تتعلق بموضوع البحث لم يتمكن الباحث من العثور عليها بين ثنايا المصادر السابقة. وتجدر الإشارة هنا إلى أنه يجب أن تكون هذه الأسئلة واضحة، محددة، ودقيقة، وألا يكون لأي سؤال فيها احتمالين في الفهم وألا تتكرر الإجابات، وتعجز الاستمارة عن تحقيق الهدف منها. وهناك أيضًا كشف البحث (Schedule)، الذي يستخدم عندما يكون الباحث وجهًا لوجه مع المبحوثين. وإن كان عيب اللقاء أنه يُعَرِّض المبحوثين لبعضٍ من الحرج من بعض الأسئلة، ولكنه رغم ذلك وسيلة نافعة، وأسلوبًا ناجحًا في الحصول على المعلومات دون الشك فيها، أو ضياعها إذا استخدم البريد، أو كانت الاجابات كاذبة، خاصة إذا استخدم التليفون كوسيلة من وسائل الاتصال التي تدخل ضمن وسائل جمع البيانات.

٤- **العينات**: وفي حالة ما إذا كان موضوع البحث يشمل مجموعة كبيرة من المفردات، أو الأشخاص الذين لا يتمكن الباحث من حصرهم حصرًا شاملًا، ففي هذه الحالة، يعتمد الباحث على أسلوب العينات (Samples). واختيار العينات هو أحد الطرق الإحصائية المفيدة في جمع البيانات في المجتمعات كبيرة الحجم. والمشكلة التي تصادف أي باحث يود تطبيق نظام العينات، هي كيفية تحديد حجم العينة التي تمثل المجتمع الذي يود دراسته تمثيلًا صحيحًا.

ومن المتفق عليه عند تحديد حجم العينة (نَ) أن تكون في حدود ١٠ – ١٥٪ من حجم المجتمع الأصلي (ن) الذي سيتم سحب العينة منه. وبالرغم من أن هذا الأسلوب غير علمي إلا أنه يسهل العمل، ولا يُدْخل الباحث في التعقيدات الرياضية التي حاولنا تبسيطها في هذا الكتاب. وهناك بعض الرموز التي يجب التعرف عليها قبل تحديد حجم العينة وهي: (نَ) وهي حجم العينة، (ن) حجم المجتمع الأصلي الذي سوف تسحب منه العينة، (ق) القيمة المعيارية، (م) معامل التشتت، (خ) الاختلاف النسبي، وبالتالي تكون معادلة تحديد حجم العينة بالصورة الآتية:

$$\text{حجم العينة} = \frac{\text{حجم المجتمع} \times \text{مربع القيمة المعيارية} \times \text{مربع معامل التشتت}}{\text{حجم المجتمع} \times \text{مربع الاختلاف النسبي} + \text{مربع القيمة المعيارية} \times \text{مربع معامل التشتت}}$$

$$\text{أو نَ} = \frac{\text{نَ} \times (\text{ق})^{٢} \times (\text{م})^{٢}}{\text{ن} \times (\text{خ})^{٢} + (\text{ق})^{٢} \times (\text{م})^{٢}}$$

مثال:

قام أحد الباحثين باختيار منطقة لدراسته بلغ عدد سكانها ١٠٠٠ نسمة، وأراد أن يختار من هذه المنطقة عينة ممثلة لها، وذلك بناء على عدة افتراضات هي:

١- نسبة الخطأ المسموح به لا تزيد على ٥ ٪.

٢- معامل التشتت بين المفردات في حدود ٢٠ ٪ .

٣- متوسط معامل التشتت في أقل الحدود ٢ ٪ .

٤- عند نسبة خطأ مسموح به = ٥ ٪ على الأكثر، تكون القيمة المعيارية المناظرة في جدول

التوزيع الطبيعي=١.٩٦
معامل الاختلاف النسبي = ٠.٠٤

$$\therefore \text{حجم العينة} = \frac{١٠٠٠ \times (١.٩٦)^{٢} \times (٠.٢)^{٢}}{١٠٠٠ \times (٠.٠٤)^{٢} + (١.٩٦)^{٢} \times (٠.٢)^{٢}}$$

$$= \frac{١٠٠٠ \times ٣.٨ \times (٠.٠٤)}{١٠٠٠ \times ٠.٠٠١٦ + ٣.٨ \times (٠.٠٤)}$$

$$= \frac{١٥٢}{١.٦ + ٠.١٥}$$

$$= \frac{١٥٢}{١.٧٥}$$

= ٨٦.٨ مفردة (٨٧ مفردة تقريبًا)

ويمثل حجم هذه العينة ٨.٧ ٪ فقط من حجم المجتمع الأصلي، ومن الأفضل في مثل هذه الأحوال اختيار افتراضات واعية، لأنها ترشدنا في تطبيق طريقة من الطرق الإحصائية بأقل خطأ ممكن.

١ - العينة العشوائية البسيطة: Random Sample

ويناسب هذا النوع من العينات الدراسات التي ترمي إلى تحديد الخصائص العامة للمجتمعات التي تمتاز بتجانس مفرداتها. كأن يكون هذا المجتمع نوع من الإنتاج مفرداته عبوات متساوية في الحجم، متجانسة في النوع. أو أن يكون مجتمعًا من الشباب في سن معين. كما أن هذا النوع من العينات يعطي كل مفردة من مفردات العينة فرصتها في الظهور.

فإذا كان لدينا مجتمع مكون من ٦٠ مفردة، وأردنا اختيار عينة عشوائية مكونة من مفردتين من هذا المجتمع، ففي هذه الحالة، نقوم بإحضار ٦٠ بطاقة، وندون على كل منها رقما من (١ – ٦٠) يمثل كل رقم مفردة من مفردات هذا المجتمع، ونخلط البطاقات خلطًا جيدا، ونسحب منها بطاقتين. فإذا وجدنا الرقمين المدونين على هاتين البطاقتين هما ٢، ١٢ تكون العينة العشوائية البسيطة التي تم اختيارها من هذا المجتمع هي العينة المكونة من المفردتين الثانية والثانية عشرة.

أما إذا كان لدينا مجتمع مكون من ٢٥٠٠ عامل، وأردنا اختيار عينة من ٥٠٠ عامل، فإنه يمكن استخدام الجداول العشوائية في ذلك. فإذا أعطينا العمال أرقامًا مسلسلة من ١ – ٢٥٠٠ (مع مراعاة أن كل رقم يتكون من أربع خانات فنكتب ٠٠٠١ ، ٠٠٠٢ ، ٠٠٠٣، ٠٠٠٤ ، ٠٠٠٥ ، ٠٠٠٦ ، ٠٠٠٧، حتى نصل إلى ٢٥٠٠).

بعدها ننظر إلى جداول الأعداد العشوائية، ونختار أربعة أعمدة، ونقرؤها من أعلى إلى أسفل، ونكتب الأرقام التي تساوي، أو تقل عن ٢٥٠٠، ونستبعد الأرقام التي تزيد على ٢٥٠٠. وكلما وصلنا إلى نهاية الصفحة نختار أربعة أعمدة أخرى، ونقرؤها أيضا من أعلى إلى أسفل، حتى يصل عدد الأرقام التي تكتبها إلى (٥٠٠) رقمًا فنكون قد حصلنا على العينة العشوائية المطلوبة. ويجب مراعاة عدم تكرار أي رقم يتم اختياره في العينة حتى لا نسمح باختيار أي مفردة من مفردات المجتمع في العينة اكثر من مرة واحدة.

٢- العينة العشوائية المنتظمة: Systematic Sample

يتناسب هذا النوع من العينات العشوائية مع المجتمعات المتجانسة أيضا، والتي تنتمي مفرداتها إلى نوعية واحدة، ولكن الاختيار والسحب هنا يتم بطريقة منتظمة. فإذا فرضنا مثلاً أن حجم المجتمع الأصلي ١٠٠٠٠ نسمة، وأن حجم العينة ٢٠٠ مفردة، فإن كل

مفردة من مفردات العينة سوف تمثل $\frac{١٠٠٠٠}{٢٠٠}$ = ٥٠ مفردة من مفردات المجتمع الأصلي.

فإذا تم اختيار الرقم ٢٠ عشوائيًا من مجموعة الأرقام التي تقع بين ١ – ٥٠ ، فإنه بإضافة مقدار التمثيل بطريقة منتظمة إلى الرقم ٢٠ نحصل على المفردات الممثلة للعينة كما يأتي:

٢٠	٧٠	١٢٠	١٧٠	٢٢٠	٢٧٠	٣٢٠	٣٧٠	٤٢٠
٤٧٠	٥٢٠	٥٧٠	٦٢٠	٦٧٠	٧٢٠	٧٧٠	٤٢٠	٨٧٠
٩٢٠	٩٧٠	١٠٢٠	١٠٧٠	١١٢٠	١١٧٠	١٢٢٠	١٢٧٠	١٣٢٠
١٣٧٠	١٤٢٠	١٤٧٠	١٥٢٠	١٥٧٠	١٦٢٠	١٦٧٠	١٧٢٠	١٧٧٠
١٨٢٠	١٨٧٠	١٩٢٠	١٩٧٠	٢٠٢٠	٢٠٧٠	٢١٢٠	٢١٧٠	٢٢٢٠
٢٢٧٠	٢٣٢٠	٢٣٧٠	٢٤٢٠	٢٤٧٠	وهكذا حتى آخر رقم وهو			٩٩٧٠

ورغم أن هذا الأسلوب يمتاز ببساطته، وقلة تكاليفه، إلا أنه يتطلب منا الحذر الشديد، والتحفظ عند استخدامه.

٢- العينة العشوائية الطبقية: Stratified Sample

يقسم المجتمع الأصلي في هذا النوع من العينات إلى طبقات Strata، ونأخذ عينة عشوائية من كل طبقة، ومن ثم نضمن تمثيل العينة لكل طبقاته. ولذلك يتناسب هذا النوع من العينات مع المجتمعات المتباينة، التي يمكن تقسيمها من الداخل إلى مجموعات متجانسة. فإذا كنا بصدد دراسة أحجام المدن كموضوع في جغرافية المدن، نأخذ فئات الحجم المتساوية، ونختار من داخل كل فئة، على اعتبار أنها طبقة، العينة التي نريدها من المدن، فنكون بذلك قد حصلنا على عينة طبقية. وعادة ما تفضل العينة الطبقية العينة العشوائية في الاختيار.

وإذا كنا في دراسة لجغرافية السكان، نستطيع تقسيمهم إلى فئات عمر رئيسية، ومن ثم يكون اختيارنا لعينة طبقية. وكذلك يمكننا اختيار العينة الطبقية بعد توزيع السكان في منطقة من المناطق حسب أماكن إقامتهم، أو أماكن ميلادهم، أو حسب أطوالهم، أو أوزانهم طالما نعتقد في وجود علاقة ما بين هذه المتغيرات، وبين الظاهرة التي نقوم بدراستها. وعند تقسيم السكان حسب النوع ذكورًا وإناثًا، أو تقسيمهم حسب الريف والحضر. كذلك دراسة خطوط الارتفاعات المتساوية، وخطوط المطر المتساوية، واختيار محطات الرصد الممثلة للإقليم، أو عدد من الأقاليم المتجانسة موضوع الدراسة.

ويراعى هنا أن اختيار عدد المفردات يكون وفقا لقانون الاختيار الذي ينص على أنه إذا كان عدد أفراد العينة ع، وعدد أفراد المجتمع ن، فإن عدد أفراد الطبقة الأولى $ع_١$، ومن ثم يكون عدد الحالات المطلوب أخذها من العينة هو:

$$ع \times \frac{ع_١}{ن}$$

علمًا بأنه إذا حصلنا على الناتج كسرًا يجب علينا أن نقربه إلى عدد صحيح. كما يجب أن يراعى عند اختيارنا هذا أحجام الطبقات التي نختار منها، وفي هذه الحالة، إذا كان أمامنا مجتمع من المجتمعات مثلاً مدينة أو إمارة الرياض مثلاً، وقد بلغ حجمها ١.٢٥٩.١٢٥ سنة ١٩٧٤، ونسبة سكان الحضر فيها ٨٣٪، وسكان الريف ١٧٪، فعند اختيار عينة من كلتا الفئتين أو الطبقتين، يجب علينا إذا اخترنا من الحضر ١٧ شخصًا، أن نختار من الريف أربعة أشخاص فقط، وذلك مراعاة لاختلاف حجم كل فئة من جهة، وإمكانية تمثيل العينة للمجتمع الأصلي باقل نسبة خطأ ممكن من جهة أخرى.

وليست هذه الطريقة مقصورة على المجتمعات غير المتجانسة، ففي إمكاننا تطبيقها على المجتمعات المتجانسة أيضًا بُغية الوصول إلى نتائج أحسن، وتحقيق أقل نسبة من الخطأ. كما أنه لدينا الخيار في طريقة اختيار أفراد العينة بالطريقة العشوائية أو الطريقة المنتظمة، ومن ثم تعتبر العينة الطبقية من أفضل أنواع العينات، لأنها تقوم بدراسة أي

مجتمع بشرائحه المختلفة، وهذا يضفي مزيدًا من تقصي الحقائق، والوقوف على كافة خصائص هذا المجتمع.

٤- العينة العشوائية متعددة المراحل : Multi-Stage Sample

وفي حالة المجتمعات الكبيرة قد يصعب على الباحث اتباع أي من طرق العينات السابقة. ولهذا لا يكون أمامه إلا أن يقسم هذا المجتمع الكبير إلى عدد من المجموعات، يختار منها مجموعة واحدة بالطريقة العشوائية. ويقسم المجموعة التي وقع عليها اختياره إلى عدد آخر من المجموعات الفرعية، وهو بهذا الأسلوب يتجنب تبعات كبر حجم المجتمع، وتصغيره إلى الحد الذي يستطيع عنده دراسة فئة حجمها مناسب، وممثلة للمجتمع الأصلي الذي اختيرت منه بقدر الإمكان.

وتصلح هذه الطريقة في المجتمعات المترامية الأطراف، متسعة المساحة، حيث لا يستطيع الباحث بإمكانياته المحدودة أن يغطي هذه المساحة الواسعة، وقد تكون في نفس الوقت سُبل الاتصال على مدى المسافات البعيدة غير جيدة، مما يقف حائلاً دون تغطية المساحة تغطية كاملة. هذا بالإضافة إلى الإمكانيات المادية التي يتطلبها مثل هذا العمل.

وتأتي فكرة المراحل من طريقة اختيار الباحث، حيث يقوم أولاً باختيار منطقة معينة من بين عدد المناطق الموجودة داخل هذا المجتمع. ولتكن هذه المنطقة محافظة من المحافظات، أو بلدية من البلديات، أو ولاية من الولايات، أو إمارة من الإمارات مثلا، ثم يقوم بعد ذلك باختيار قرية من قرى هذه المحافظة، أو فرع بلدي من فروع البلدية، ثم يقوم باختيار مجموعة معينة من السكان، أو المساكن في هذه القرية، أو الفرع البلدي ليخصهم بدراسته اعتبارًا منه بأن هذه المجموعة سوف تمثل المجتمع كله.

وتختلف هذه الطريقة عن الطرق السابقة، في استخدامها لأكثر من وحدة واحدة للمعاينة في العينة النهائية، حيث تتكون الأخيرة من عدة عينات متتالية، مع ملاحظة أنه كلما زاد عدد المراحل تبع ذلك كبر حجم العينة، حتى تكون أكثر تمثيلًا لمجتمع الدراسة. وهذا في الواقع، تجنبًا للعيب الذي تعاني منه هذه الطريقة، وهو استعانتها في نهاية المعاينة بعدد قليل من المفردات. ورغم هذا العيب إلا أن طريقة المراحل المتعددة في اختيار العينة غير مكلفة، ولا مجهدة من النواحي الإدارية أو الفنية، إذا ما قيست بالطرق السابقة، كما أنها لا تحتاج إلى عمل إطار كامل لكافة مفردات المجتمع، وإنما يقتصر الإطار على الوحدات أو المناطق الرئيسية، والأفرع البلدية أو القرى.

٥. العينة العمدية :Purposive Sample

وهذا النوع من العينات غير العشوائية، يعتمد على الاختيار الشخصي، ولا يراعى فيها تكافؤ الفرص لظهور المفردات في العينة المختارة. والباحث عندما يميل إلى هذا النوع فهو يتعمد ويصر على عدم اتباع أي من الطرق السابقة، وينحاز في اختياره. وتصلح هذه الطريقة في المجتمعات شديدة التباين والاختلاف، فعند البحث عن مستوى كفاءة الخدمات في مدينة ما، متعددة الأقسام أو المحلات، قد يلجأ في بعض الأحيان إلى اختيار قسم من هذه الأقسام أو محلة بعينها، خاصة إذا كانت المدينة كبيرة الحجم متباينة الخصائص والسمات. والباحث في التجمعات البشرية المتناثرة هنا وهناك على مسطح جغرافي واسع، قد يلجأ لأخذ هذا النوع من العينات أسلوبًا في بحثه متوخيًا فيه التحيز إلى منطقة دون غيرها. وكذلك فإن الباحث في قرى الصيادين، ومراكز استقرارهم المتباعدة المتباينة، وأمام صعوبة حصرهم، أو عمل إطار يشمل جميع أفرادهم، فإنه يلجأ أيضًا إلى هذا الأسلوب. ودارس توزيع الرعاة والبدو ليس أمامه إلا أن يتخذ مثل هذه الطريقة.

٦- العينة بالحصص : Quota Sample

يتم إجراء هذا النوع من العينات غير العشوائية، بعد تقسيم مجتمع الدراسة إلى عدة أقسام متجانسة، حسب خصائص معينة يفترضها الباحث، ويتحكم في حصة كل قسم من هذه الأقسام على قدر ما لها من أهمية، وما بها من تجانس. وعادة ما يستخدم هذا النوع من العينات في الدراسات العامة غير الدقيقة، التي لا تتطلب من الباحث سوى معرفة الاتجاه العام للظاهرة موضوع بحثه.

ولا تقف أنواع العينات عند هذا الحد، فهناك العينات المتشابكة، أو الشبكية Network of Samples، أو المتشاركة والمتداخلة Interpenetrating، والمتناظرة Matched Samples، والعينات الموصولة Linked Samples، والعينات المزدوجة Duplicated Samples، والعينات الرئيسية Master Samples، والعينات المتنافرة Discordant Samples، وعلى الرغم من ذلك فإن العينات العشوائية تفضلها جميعًا.

وعلى العموم أيًّا كان نوع العينة التي يختارها الباحث في دراسته للمجتمع الذي وقع اختياره عليه، فإن هذا الاختيار تتحكم فيه خبرة الباحث، وإمكاناته المادية والمعنوية وموضوع بحثه سواءً كان عامًّا أو متخصصًا، وكذلك الوقت المحدد لدراسته من قِبَله أو من الجهة المشرفة عليه، والعلاقات العامة التي تربطه بأفراد مجتمع الدراسة. والعامل الأخير أصبح في الآونة الأخيرة وعلى كافة المستويات، يلعب دورا كبيرا في نجاح أو فشل أي عمل من الأعمال العلمية في جميع التخصصات.

وإذا كان الباحث جادًّ في بحثه، فإنه يراعي عدم الاعتماد على الأفراد الذين يقعون تحت أي نوع من التأثير كالقرابة أو الصداقة والمعارف. وكذلك البعد عن الأماكن الخاصة المحدودة جدًّا عند أخذ العينة. وعليه عدم القيام بتوزيع استماراته في الميادين، أو على المارة في الطرقات، ويجب ألا يستخدم دليل التليفونات في اختيار العينة، لأن الدليل لا يشمل سوى طبقة معينة من المجتمع، وربما تكون محدودة، وألا يستخدم قوائم الانتخابات، أو التصعيد أو الضرائب في اختيار عينة. ففي هذا تفاديًا لكل الأخطاء التي يمكن أن تقلل من نجاح وفاعلية العينة، ومدى تمثيلها للمجتمع الذي اختيرت منه.

وتعتبر الخرائط الحقلية التي يقوم الباحث بملئها في ميدان دراسته، مصدرًا من المصادر الميدانية الهامة جدًّا، حيث يستطيع الباحث أن يوقع على عدد من الخرائط الصماء بعض المعلومات من أماكنها المحددة، ومواضعها الخاصة بها، فهي في ذلك مثلها مثل استمارة الاستبيان، أو كشف البحث سابقي الذكر.

في الجغرافيا الطبيعية مثلًا يلزم الباحث مجموعة من خرائط المناسيب المتساوية "الخرائط الكنتورية"، والخرائط الجيولوجية، وخرائط المناخ، والخرائط الطبوغرافية وثيقة الصلة بموضوع بحثه. وحينما يقوم بتطبيق الخريطة الطبوغرافية على الخريطة الجيولوجية فإن ذلك يساعده على تفسير العديد من ظاهرات السطح، وتفسير تكوينها، وطبيعتها الصخرية، وتتابعها الطبقي. كما تظهر له الخريطة الطبوغرافية الانعكاس العام للتفاعل الذي تم بين التكوين الجيولوجي وظاهرات السطح المختلفة، وأثر ذلك في توزيع مراكز الاستقرار البشري من حيث تركزها في مكان، وتشتتها في آخر.

وفي دراسة استغلال الأرض، يحتاج الباحث إلى الخرائط التفصيلية التي توضح الأحواض الزراعية، خرائط فك الزمام، مقياس ١ : ٢٥٠٠، حيث يستطيع من خلالها الوقوف على أنماط الاستغلال المختلفة، وصور استخدام الأرض المتباينة، وكذلك العوامل المختلفة الطبيعية، والبشرية التي تؤثر في صورة هذا الاستخدام. ودراسة جميع مشكلات التربة، درجة انحدارها، انجراف التربة، المياه المتوفرة للزراعة، الترع والمصارف، أو الوديان، الدورة الزراعية للمحاصيل في كل فصل من فصول السنة، وأخيرًا فعليه أن يتابع حقل دراسته عدة مرات حتى يقف على أبعاد صور استغلال الأرض وفق نظام الدورة الزراعية السائد.

وفي جغرافية المدن، يجد الباحث نفسه بدون الخرائط الخاصة بالمدينة موضوع دراسته، وكأنه لا يرى أمامه شيئًا، ولهذا تأتي أهمية الخريطة التي تقوم إلى جانب وظيفتها في تخزين المعلومات، وترتيبها، بتقديم الاتساع المكاني للمدينة. الاتساع المكاني القديم في خريطة، والاتساع المكاني الحديث في أخرى. فإذا أمعن النظر في الخريطتين، أو في أكثر

منهما، وقف على مدى التغير الذي طرأ على هذا الكائن، وإذا وضع الخريطتين إحداهما على الأخرى، فإنه يستطيع أن يتبين أماكن الامتداد العمراني الجديدة، وأن يلاحظ مدى اتباعها لنمط معين من أنماط التخطيط، وكذلك مدى ارتباطها بالطرق الرئيسية أو الخدمات، أو المراكز الصناعية، خاصة في مناطق الضواحي.

وفي أثناء تطبيق الباحث لنظريات الحجم والتباعد، إذا شملت دراسته أكثر من مدينة، فسوف تكون الخريطة مساعدة له في قياس المسافات بين مراكز العمران المختلفة، وإيجاد العلاقة بين توزيع هذه المراكز وأحجامها، وعند التعرف على نواة المدينة، وبما لها من صفات وخصائص مميزه، فإن الخريطة هي مصدره الأساس لمثل هذه المعلومات، وكذلك الصور الجوية. وعند دراسة كثافة السكان داخل كل قسم من أقسام المدينة يحتاج إلى بعض الخرائط التفصيلية، أو الكروكروماتية Chrochromatic، أو خرائط التظليل والتلوين المساحي في توضيح التوزيع العام لكثافة السكان في كل قسم. وخرائط الكوروبلث Choropleth، وهي التي تختص بالتوزيع النسبي، وتعتمد على المعادلات الرياضية والأساليب الإحصائية في عملها. ووظيفة خرائط التوزيع النسبي الأساسية في ضبط ظاهرة الانتشار، أو التجمع في المحلات وفي المباني التي تقام على سطح الأرض، آخذة في الاعتبار عدد هذه المحلات أو عدد المساكن أو حجم السكان في كل محلة أو مساحة الإقليم الذي تخدمه كل محلة، والمسافات التي تفصل بين هذه المحلات المختلفة. وقد ذكر "ديمانجون Demangon"، أن معامل التباعد والانتشار في الريف أيضا، يمكن الحصول عليه من المعادلة التالية:

$$\frac{\text{عدد سكان المحلات} \times \text{عدد المحلات}}{\text{مجموع السكان}}$$

وفي بريطانيا يحصل على معامل التباعد بقسمة عدد المحلات في أي قسم إداري على عدد المنازل المنعزلة، على أساس أن المنزل المنعزل هو الذي يقع على بعد ميل عن المحلة. وفي المؤتمر الجغرافي الدولي عام ١٩٣٤م قدمت "جين برنارد"، معادلة لحساب معامل التجمع للعمران في الريف مؤداها:

$$\frac{\text{عدد المنازل في كل محلة} \times \text{المسافة}}{(\text{عدد المحلات})^{2}}$$

ومن الخرائط الهامة أيضًا في دراسة المدن، خرائط الأيسوبليث Isopleth، وهي خرائط خطوط التساوي، ومنها خرائط الأزمان المتساوية، والنفوذ المتساوي، ويقصد

بخطوط الأزمان المتساوية، الخطوط التي تربط بين الأماكن التي تستطيع بلوغها من نقطة معينة بوسيلة معينة في وقت واحد.

هذا بالإضافة إلى خرائط الحركة Flow-Line-Maps، التي تبين العلاقة بين المدن والقرى المجاورة التي تتبادل الخدمات معها. وتصور لنا هذه الخرائط الحركة بخطوط تتبع طرق المواصلات الموجودة، وترسم بسمك يتناسب وحجم الحركة على كل منها. وعادة ما يقدر نشاط الحركة إما بعدد وحدات النقل، أو بمقدار الحمولة، أو بعدد المسافرين. وإذا رسم الباحث عددًا من خطوط الحركة من وإلى مدينة ما في خريطة واحدة، أصبح بإمكانه تحديد مناطق نفوذ المدينة، ويستطيع أن يقارن مناطق النفوذ في سائر المدن الأخرى، بغية وضع المدن في مراتب معينة، وفق حجم إقليم كل مدينة.

وهناك أيضًا الخرائط الإشعاعية Ray-Maps، وتفيد في تحديد مجالات وظائف المدينة والطريقة، أن يقوم الباحث بتوصيل عدد من الخطوط بين المركز الحضري، والمحلات المستفيدة من كافة الخدمات التي يقدمها هذا المركز.

وختامًا فالخريطة تعتبر عين الباحث التي يكتشف بها نفسه في الاتساع المكاني الذي يقوم بدراسته أيًّا كان نوعه، وفي عمليات التحليل والتفسير هي التي تجرد الرقم من جفافه، وتُلِين بالتحليل من صلابته وتربطه بمواضعه السليمة. ومن ثم يستقيم أسلوب الباحث في التعليل والتفسير اعتمادًا على مدى درايته بمنطقة الدراسة، ومعلوماته التي جمعها عنها، ومدى خبرته في هذا المجال، وإن كانت الخريطة نوعًا من التخزين للمعلومات التي يدونها الباحث، فهي في نهاية المطاف معرضًا يفصح فيه عن مهارته في الدراسة ودقته في الأداء.

وبعد انتهاء الباحث من جمع المادة العلمية الخاصة ببحثه من المصادر الخاصة والمصادر العامة والمصادر الميدانية السابقة، فإنه يجد نفسه أمام خضم هائل من المعلومات، ويظهر أمامه التساؤل الذي يزيده عناءً فوق معاناته، بعد جمع هذه المعلومات، وهو نفس التساؤل الذي يسلمنا إلى الجزء التالي، وهو الخاص بتصنيف البيانات.

ثانيًا: تصنيف البيانات: Classification

بعد جمع البيانات تصبح مهمة الباحث هي ترتيب هذه البيانات وتنسيقها ومراجعتها وتدقيقها، والوقوف على ما بها من أخطاء لإمكان تفاديها قبل تصنيفها وتبويبها. فقد يكون أحد الأخطاء التي وقع فيها الباحث هو عدم تحديده بدقة لموضوع بحثه، أو المشكلة التي يريد التوصل إلى حلها. ويظهر ذلك في تباين واختلاف البيانات التي جمعها، وتعددها وابتعادها عن الهدف الذي كان يرجوه منها. حينئذ يكون عليه استبعاد ما يشك فيه من بيانات، والتخلص من البيانات التي لا تمت إلى مشكلته بصلة.

وقد يكون خطأ الباحث هو عدم تحديده المجال الجغرافي الذي يحتوي على مشكلته موضوع الدراسة، وفي هذه الحالة، وبسبب عدم شمول البيانات لمنطقة البحث بكل أقسامها ووحداتها، فسوف يجد الباحث أن البيانات أصبحت عاجزة عن الوفاء بكل ما تتطلبه الدراسة، لأنها لم تشمل المكان كله، واقتصرت على أجزاء معينه في داخل المكان، ولهذا يكون على الباحث أن يقوم باستكمال كافة البيانات الناقصة، من المصادر التي يرى أن العجز محصور فيها.

وربما تكون الإجابات الخاطئة من قبل المبحوثين، والتي يفهمها الباحث من خلال تكرار بعض الأسئلة التي توضع بقصد الوصول إلى إجابة صحيحة، سواء كان ذلك في استمارة الاستبيان أو كشف البحث، والتي تأتي نتيجة عدم فهم الأسئلة أو الألفاظ التي يجدها الباحث أثناء مراجعة الاستمارات، أو تكون ناتجة من مجرد السهو أو النسيان، أي تكون أخطاء بسبب التحيز، أو عدم الاستجابة، كأن يرفض أحد المبحوثين الإجابة إذا أحس بالإحراج، أو تغيب بعض المبحوثين من محال إقامتهم المعتادة وقت جمع المعلومات، حينئذ يجب على الباحث معاودة الكَرَّة مرةً ومرتين، وإن اقتضى الأمر أن يعاود الكرة مرات ومرات حتى يصل إلى مرماه.

ونظرًا لاحتمال ضرورة الرجوع إلى المبحوثين في بعض هذه الحالات، فمن الأفضل أن تتم مراجعة الاستمارات على الطبيعة في نفس المكان. وربما يتمكن الباحث من تصحيح ما جاء في الاستمارة من إخطاء بالإجابات الواردة عن بعض الأسئلة المتشابهة، أو الأسئلة التي تفيد إجابتها في تصحيح ذلك الخطأ. ويتم هذا من قبل الباحث بوضع بعض الأسئلة التي تسمى بأسئلة المراجعة Check Questions، فمثلًا، يمكن أن يسال عن عمر الشخص، وبعد قليل يسأل عن تاريخ ميلاده، أو يسأل الشخص عن محل إقامته السابق، فإذا أجاب بالنفي، يوجه إليه سؤالًا عن تاريخ تواجده في محل الإقامة الحالي، وهكذا تكون الأسئلة في حد ذاتها حلاً لمشكلة خطأ من هذا النوع.

ويجب على الباحث أثناء المراجعة والتدقيق، أن يتأكد من توحيد وحدة القياس التي اختارها، فوحدة القياس إذا اختلفت لن تمكن الباحث من الحصول على نتائج مقبولة. وكذلك يجب عليه مراعاة التجانس التام في جميع الوحدات المستخدمة في البحث.

وعلى العموم فإن الهدف الرئيسي من تصنيف البيانات هو التخلص من كمياتها الضخمة، وتقسيمها حسب أوجه التصنيف التالية:

١- التصنيف الكيفي، أو النوعي: Qualitative

يهتم الباحث في التصنيف الكيفي أو النوعي، بأنواع مفردات بحثه. إذا كان عن السكان ركز على نوع هؤلاء السكان، ذكورًا أو إناثًا، صغارًا أو كبارًا، زراع أو تجار أو صناع، مهاجرين أو مستقرين، وكذلك حالتهم الزواجية، وإذا كان عن السكن فيصنف السكن إلى انماط تتفق وطبيعة استخدام السكن، منزل، فيلا، مباني حكومية، وكالات تجارية، أو مصانع. وإذا كان السؤال عن التجهيزات الهندسية المعمارية لتحديد كافة الخدمات، مثل المياه والكهرباء والغاز، والصرف الصحي، وشكل النوافذ، ومواقعها، ومواقع دورات المياه بالنسبة لاتجاه الرياح ، وكذلك أشكال الطرق (متسعة –ضيقة –حارات – أزقة – ملتوية – مستقيمة – مسدودة – لها أرضية للمشاه ، وهكذا نجد أن هذا النوع من التصنيف يدخل في حسابه كل الظواهر التي يصعب إعطاؤها قيمة رقمية .

٢- التصنيف الكمي : Quantitative

يهتم الباحث في هذا النوع بالكم وليس بالكيف أو النوع (كم عدد السكان – كم حجم الأسرة – كم عدد الأسر – كم حجم المدينة – كم دخل الأفراد – كم طالب في الكلية – كم مدرسة – كم مدرس – كم مستشفى – كم سرير – عدد المصانع – عدد العاملين في الصناعة – مساحة الأرض الزراعية – مساحة الأراضي الفضاء - كثافة المرور – حجم تيارات الهجرة – كميات المطر الساقطة – درجات الحرارة – درجة الانحدار – زاوية الانحدار .....وهكذا يدخل ضمن هذا النوع كل متغير يسهل اعطاؤه قيمة رقمية .

٣- التصنيف الزمني : Chronological

يعتمد التصنيف الزمني على تتبع أي ظاهرة خلال فترة زمنية معينة فمثلا تتبع نمو السكان خلال الفترات التعدادية ، أو نمو المدن في فترة زمنية محددة ، وكذلك اختلاف الأسعار من وقت إلى آخر ، ودراسة الاتجاه العام للأسعار العالمية وتوقعات السكان لفترات زمنية مقبلة لتجنب المشكلات التي تترتب على زيادتهم ، ودراسة المراحل التطورية التي تمر بها أشكال سطح الأرض ، ودراسة التطور التاريخي الذي تمر به الوحدات السياسية منذ نشأتها حتى الوقت الحاضر .

٤- التصنيف الجغرافي : Geographical

يعتمد التصنيف الجغرافي على الكم والنوع والزمن معًا ، وذلك لأنه يدرس أي ظاهرة داخل الإطار المكاني لها متأثرة به ومؤثرة فيه ؛ففي الدراسات السكانية مثلا يقوم الباحث بدراسة تطور النمو السكاني في مكان معين ، والمراحل المختلفة التي مر بها هؤلاء السكان وتركيبهم العمري والنوعي والاقتصادي في كل منطقة من المناطق ، وحين يذكر الباحث رقمًا مرتفعًا للكثافة السكانية مثلا فإنه يوحي له بالمكان المزدحم وحين يذكر أن

معظم السكان يتركزون في الوادي والدلتا في مصر يكون هذا المكان هو أقرب المناطق التي يوحي إليها رقم الكثافة المرتفعة ، وعندما يطالع الباحث بيانات عن إنتاج الأسماك يوحي له ذلك بقرب المكان من المسطحات المائية أو جوانب الأنهار .

<u>أسئلة وتطبيقات</u>

١- إذا كان حجم المجتمع الأصلي هو ١٥٠٠٠٠ مفردة والمطلوب حساب الحجم الأمثل للعينة العشوائية التي ستجرى عليها الدراسة بدقة ٩٥% ، فإذا أخذنا عينة استرشاديه وأجريت عليها بعض القياسات وكانت النتائج كالتالي :

- معامل التشتت بين مفردات العينة الاسترشادية هو ١٥%
- متوسط معامل التشتت بين مفردات العينة الاسترشادية هو ١.٥%
- درجة الثقة المطلوبة في اختيار حجم العينة هي ٩٥% (أي أن القيمة المعيارية هي ١.٩٦ من جداول التوزيع الطبيعي).
- **بناء على ما سبق أوجد حجم العينة**

الفصل الثاني

تجهيز وتبويب البيانات

الفصل الثاني
تجهيز وتبويب البيانات

١- تبويب البيانات الوصفية

٢- تبويب البيانات الكمية:

تبويب البيانات الكمية المتقطعة (الوثابة).

تبويب البيانات الكمية (المستمرة).

- تبويب البيانات الكمية المستمرة في شكل جدول تكراري بسيط.

- تبويب البيانات الكمية المستمرة في شكل جدول تكراري مزدوج.

- تبويب البيانات الوصفية أو الكمية في شكل جدول تكراري بسيط أو مزدوج في صورة تكرارات نسبية.

- تبويب البيانات في شكل جدول تكراري متجمع.

إن الخطوة التالية لجمع البيانات هي تبويب هذه البيانات، فالبيانات المجمعة تكون غير منظمة، فيجب ترتيبها وتصنيفها في شكل جدول لتسهيل تفسيرها وتحليلها. وكثيرا ما تتم هذه العملية بطريقة آلية، باستخدام آلات التثقيب Punching، والفرز Coding، والتبويب Tabulating. وفي الآونة الأخيرة، كنتيجة لانتشار الحاسب الآلي، فلقد استخدم لهذا الغرض، إلا أنه لا يجب استخدام آلات التبويب الإحصائي ولا الحاسب الآلي إلا في حالة وجود عدد كبير جدًّا من البيانات، أما إذا كان عدد البيانات عددًا بسيطًا نسبيًّا فالتبويب في هذه الحالة يتم يدويا.

وسوف نتناول بالشرح طرق التبويب اليدوي.

تجهيز البيانات أو تبويبها، هو تنظيم وحفظ وتخزين واسترجاع وتصنيف البيانات بشكل يتلاءم مع الاحتياجات الحالية والمستقبلية لمستخدمي تلك البيانات.

ويمكن تقسيم البيانات الإحصائية إلى نوعين من البيانات تسمى بالمتغيرات الأول يمثل المتغيرات الكمية الخاصة بالظواهر القابلة للقياس، وقد تكون هذه المتغيرات وثابة (متقطعة) مثل عدد المستخدمين لأنواع معينة من التليفون المحمول، أو تكون مستمرة (متصلة) مثل الأطوال والأوزان والأحجام والسرعات، والثاني يمثل المتغيرات الوصفية وهي المتغيرات التي يعبر عنها بالوصف، وليس بالأرقام، والتي يلزم تحديد الصفات التي تتميز بها البيانات، ويتم حصر عدد المفردات التي تنتمي إلى كل صفة من تلك الصفات، مثال ذلك البيانات الخاصة بإنتاجية الفدان من محاصيل معينة، وتنقسم المتغيرات الوصفية إلى متغيرات وصفية ترتيبية يمكن وضعها في ترتيب معين مثل الحالة التعليمية (لا يقرأ ولا يكتب، يقرأ بالكاد، يقرأ ويكتب، ابتدائي، إعدادي، ثانوي، جامعة، فوق الجامعي) وإلى متغيرات وصفية اسمية، التي لا يمكن وصفها في ترتيب معين مثل النوع (ذكر، أنثى) ولون الشعر والعينين، وبعد الانتهاء من جمع البيانات سواء بأسلوب الحصر الشامل أو المعاينة وسواء كانت البيانات كمية أو وصفية فإنه يستلزم مراجعة تلك البيانات قبل البدء في تصنيفها ووضعها في الجداول المناسبة بهدف إلغاء تلك البيانات المتناقضة أو الناقصة الإجابة أو استيفائها في حالة إمكانية ذلك.

والهدف من تصنيف البيانات أن البيانات المجمعة تكون غير منظمة، ولتسهيل تفسيرها وتحليلها يجب ترتيبها وتصنيفها في شكل جدول إحصائي.

وهناك العديد من القواعد التي يجب اتباعها عند تصميم الجداول الإحصائية منها:

- أن يشتمل الجدول الواحد على بيانات متشابهة (متجانسة) أو توجد علاقة بينها.

- أن يعطي لكل جدول رقم لتسهيل الرجوع إليه إلى جانب عنوان مختصر وواضح ومحددًا لما يحتويه من معلومات.

- أن تكون عناوين الأعمدة والصفوف مختصرة ودقيقة ومرتبة وفقا لتسلسل زمني أو حسب أهميتها، وذلك لتسهيل الإشارة إلى بيانات الجدول تبعا لترقيم تلك الأعمدة والصفوف.

- يجب إيضاح طبيعة المتغيرات الواردة في الجدول، فيبين الجدول ما إذا كانت الأرقام الموجودة مطلقة أو في شكل نسبة مئوية، وفيما إذا كانت مقربة، وتوضح وحدات المقياس للأرقام المطلقة بدقة.

- يحسن عدم التوسع في البيانات المعروضة في الجدول الواحد بقدر الإمكان حتى يكون أكثر وضوحًا، كما يجب مراعاة تنسيق حجم الجدول من حيث طوله وأتساعه.

- وللأمانة العلمية يلزم كتابة مصدر البيانات الواردة بالجدول.

وهناك تقسيمات عديدة تقسم بها البيانات، وأكثر هذه التقسيمات انتشارًا هي تقسيمها إلى بيانات وصفية أو نوعية، وبيانات كمية، وفيما يلي شرح لطريقة تبويب هذين النوعين من البيانات.

١- تبويب البيانات الوصفية:

وتبعا لنوع المتغيرات يمكن وضع البيانات في جداول مناسبة يطلق عليها اسم الجداول التكرارية، يتم عن طريقها عرض البيانات بطريقة مختصرة دون المساس الشديد بشكل البيانات، وتنقسم الجداول التكرارية إلى نوعين، الأول: يمثل الجداول التكرارية البسيطة التي تعرض بيانات ظاهرة واحدة، والثاني: يمثل الجداول التكرارية المزدوجة التي تعرض بيانات خاصة بظاهرتين فيما بينهما علاقة سببية ما.

أ- تبويب البيانات الوصفية في شكل جدول تكراري بسيط:

حالة تبويب البيانات الوصفية، تقسم البيانات إلى صفات معينة، مثال ذلك تقديرات امتحان الطلبة في مادة الجغرافيا الكمية إلى مقبول وجيد وجيد جدًّا وممتاز، أو تقسيم المجتمع إلى أمي ومتعلم، أو تقسيم الأشخاص إلى (أعزب، ومتزوج، ومطلق، وأرمل)إلخ....

مثال ذلك توزيع الطلبة في بعض كليات جامعة معينة في أحد السنوات الدراسية كما يلي:

الكلية	الآداب	التجارة	الزراعة	الهندسة	العلوم	الطب
عدد الطلاب	١٠٠٠	٩٠٠	٨٠٠	٦٠٠	٣٠٠	٢٠٠

وإنشاء جداول تكرارية بسيطة في حالة البيانات الوصفية يتطلب إنشاء جدول تفريغ من ثلاثة أعمدة يطلق على العمود الأول عمود الصفات (توضع به الصفات حسب تنسيق أو ترتيب معين)، والعمود الثاني يوضح العلامات وذلك بوضع علامة (/) أمام كل صفة، ونستمر عملية التبويب (التفريغ) حتى ننتهي من تفريغ جميع الصفات.

ويلاحظ أن كل خمس علامات تكون ما يسمى بالحزمة (////) لتسهيل عد العلامات ، والعمود الثالث نطلق عليه اسم التكرار ويمثل عدد العلامات أمام كل صفة من الصفات.

مثال: فيما يلى بيان بدرجة توطن محصول معين في محافظات جمهورية مصر العربية، والمطلوب تفريغ هذه البيانات في شكل جدول تكراري مناسب: توطن شديد، ضعيف، ضعيف جدًّا، متوسط، شديد، متوسط، شديد، شديد، شديد، متوسط، ضعيف، ضعيف، ضعيف جدًّا، ضعيف، شديد، شديد جدًّا، متوسط، شديد جدًّا، ضعيف، ضعيف جدًّا، متوسط، متوسط ، شديد ، شديد جدًّا، ضعيف، ضعيف.

درجة التوطن	العلامات	التكرارات
شديد جدًّا	///	٣
شديد	// ~~////~~	٧
متوسط	/ ~~////~~	٦
ضعيف	// ~~////~~	٧
ضعيف جدًّا	///	٣

وبأخذ العمود الأول (الصفة)، والعمود الأخير (التكرارات) نحصل على ما يسمى بجدول توزيع تكراري بسيط للبيانات الوصفية مثل:

التوطن	شديد جدًّا	شديد	متوسط	ضعيف	ضعيف جدًّا
التكرار	٣	٧	٦	٧	٣

ب - تبويب البيانات الوصفية في شكل جدول تكراري مزدوج:

إذا كان هناك علاقة بين ظاهرتين وصفيتين فإن تلك البيانات تبوب في شكل جدول تكراري مزدوج للبيانات الوصفية، وتنقسم الجداول التكرارية المزدوجة في حالة البيانات الوصفية إلى نوعين: الأول يسمى بجدول الاقتران، والذي نستطيع عن طريقه عرض بيانات وصفية عن ظاهرتين كل ظاهرة تنقسم إلى قسمين فقط، ومن الأمثلة على ذلك تقسيم الأشخاص إلى مدخنين أو غير مدخنين للظاهرة الأولى، وتقسيم الأشخاص إلى مصابين بمرض معين أو غير مصابين بذلك المرض للظاهرة الثانية، كما يتضح من بيانات المثال التالي:

جدول اقتران لامتلاك عينة من الرجال والنساء

لبعض الأراضي الصالحة للاستصلاح

	يملك	لا يملك	المجموع
سيدات	٥	١٨	٢٣

رجال	١٧	١٠	٢٧
المجموع	٢٢	٢٨	٥٠

والنوع الثاني يسمى بجدول التوافق والذي نستطيع عن طريقه عرض بيانات وصفية عن ظاهرتين، كل ظاهرة منهما تنقسم إلى أكثر من قسمين، أو تكون إحدى الظاهرتين على الأقل تتكون من أكثر من قسمين، ومن الأمثلة على ذلك تقسيم الأسر تبعًا للمستوى الاجتماعي (غنية، متوسطة، فقيرة) للظاهرة الأولى، وإلى حجم الأسرة (صغيرة، متوسطة، كبيرة) للظاهرة الثانية، كما يتضح من بيانات المثال التالي:

		المستوى الاجتماعي			المجموع
		غنية	متوسطة	فقيرة	
حجم الأسرة	صغير	١٠٠	١٥٠	٥٠	٣٠٠
	متوسط	٧٥	١٠٠	٢٠٠	٣٧٥
	كبير	٢٥	٢٠٠	٥٠٠	٧٢٥
المجموع		٢٠٠	٤٥٠	٧٥٠	١٤٠٠

وكما فعلنا في حالة تبويب البيانات الوصفية شكل جدول تكراري بسيط، يمكن إنشاء جدول تكراري مزدوج في حالة البيانات الوصفية، بإنشاء جدول تفريغ من تقسيمات الصفة الأولى أفقيا، وتقسيمات الصفة الثانية رأسيا، مكونًا بذلك عدد من الخلايا المشتركة في جزء من الصفة الأولي، وجزء من الصفة الثانية ، ثم يضاف صف أخير لمجموع كل منهما، وتفرغ البيانات الزوجية تلك تبعًا للخلية التي تقع فيها، بوضع علامة (/) حتى يتم الانتهاء من تفريغ جميع الصفات المزدوجة، وبجمع عدد العلامات داخل كل خلية نحصل على تكرار الخلايا، ويمثل عدد العلامات التي تشترك في جزء من الصفة الأولى، وجزء من الصفة الثانية، وبأخذ تقسيمات الصفة الأولى والثانية والتكرارات في كل خلية، إلى جانب المجاميع في الصف والعمود الأخير نحصل على جدول توزيع تكراري مزدوج للبيانات الوصفية.

مثال: فيما يلي بيان بعدد ٣٠ رجل وامرأة تبعا للعمل الزراعي في إحدى القرى، والمطلوب تفريغ جدول توزيع تكراري مناسب.

يعمل	يعمل	تعمل	تعمل	لا يعمل	لا تعمل
لا تعمل	تعمل	يعمل	يعمل	لا تعمل	لا تعمل
يعمل	تعمل	يعمل	يعمل	تعمل	يعمل
يعمل	تعمل	لا تعمل	يعمل	يعمل	تعمل
لا يعمل	يعمل	لا تعمل	لا تعمل	تعمل	لا تعمل

جدول التفريغ

	رجل	امرأة	المجموع
يعمل	~~////~~ ~~////~~ /	~~////~~ ///	١٩
لا يعمل	//	~~////~~ ////	١١
المجموع	١٣	١٧	٣٠

وبالحصول على جدول توزيع تكراري لتلك الظاهرة نجد أن

	رجل	امرأة	المجموع
يعمل	١١	٨	١٩
لا يعمل	٢	٩	١١
المجموع	١٣	١٧	٣٠

٢- تبويب البيانات الكمية:

يمكن تقسيم البيانات الكمية إلى نوعين: بيانات مستمرة وبيانات وثابة (متقطعة). وتختص البيانات المستمرة بقياس متغيرات مستمرة، بينما تقوم البيانات الوثابة بقياس المتغيرات الوثابة، ويقصد بالمتغير المستمر أي متغير يمكن أن يأخذ أي قيمة بين قيمتين ومثال ذلك الأطوال، والأعمار، ودرجات الطلبة في الامتحان، بينما يقصد بالمتغير الوثاب المتغير الذي يأخذ قيم معينة فقط ولا يأخذ أي قيمة بين هذه القيم، وعادة تكون المتغيرات الوثابة أعدادًا صحيحة لا يمكن تجزئتها إلى كسور، ومثال المتغيرات الوثابة عدد الأشخاص، عدد الحجرات، عدد الكتب ..... إلخ.

أ- تبويب البيانات الكمية المتقطعة (الوثابة):

- تبويب البيانات الكمية (المتقطعة) في شكل جدول تكراري بسيط:

يقصد بالمتغير الكمي (المتقطع) ذلك المتغير الذي يأخذ قيم معينة ومحددة، ولا يأخذ قيمة بين كل قيمتين، مثال ذلك عدد الحجرات بالمنزل، وعدد الأفراد، وفي الغالب تكون المتغيرات الوثابة أعدادًا صحيحة لا يمكن تجزئتها إلى كسور، ولكن ذلك لا يمنع من وجودها في شكل كسري محدد ومعين، وكما في حالة البيانات الوصفية في شكل جدول تكراري بسيط، يمكن إنشاء جدول توزيع تكراري بسيط للبيانات الكمية الوثابة (المتقطعة)، وذلك بإنشاء جدول تفريغ من ثلاث خانات (أعمدة)، العمود الأول يوضح قيم المتغير العشوائي الوثاب، والثاني للعلامات، أما الثالث فيمثل التكرارات، وبأخذ العمود الأول والأخير نحصل على ما يسمى بجدول التوزيع التكراري البسيط للبيانات الكمية المتقطعة.

مثال: لنفرض أن لدينا ٣٠ أسرة، أحجامها كالآتي:

٢، ٣، ٥، ٦، ٧، ٦، ٥، ٣، ٦، ٧، ٣، ٥، ٧، ٩، ٢، ١٠، ٧، ٢، ٦، ٥، ٥، ٦، ٥، ٣، ١٠، ٦، ٢، ٥، ٩.

أول خطوة هي تحديد أكبر وأصغر قيمة، والفرق بينهما، أي المدى، فأكبر قيمة هنا هي ١٠ وأصغر قيمة ٢، والمدى هنا (١٠-٢=٨)، ثم نكون جدول التفريغ بحيث تكون لدينا القيم من ٢ إلى ١٠ في العمود الأول، بينما يخصص العمود الثاني للعلامات، والعمود الثالث يختص بعدد الأسر. ثم نضع العلامات مثل ما فعلنا في حالة القيم الوصفية، ويكون جدول التفريغ كما هو مبين في الجدول التالي، ونحصل على جدول التوزيع التكراري بأخذ العمود الأول والأخير من هذا الجدول، كما سبق وأن فعلنا في حالة البيانات الوصفية.

تفريغ لعدد الأفراد في ٣٠ أسرة

عدد الأفراد	العلامات	عدد الأسر
٢	////	٤
٣	////	٤
٥	~~////~~ ///	٨
٦	~~////~~ /	٦
٧	////	٤
٩	//	٢
١٠	//	٢
المجموع		٣٠

وواضح أن المدى في المثال السابق صغير، كما هو الحال في أغلب البيانات الوثابة، إلا أنه كثيرًا ما يكون المدى كبير، وفي هذه الحالة، يجب تقسيم القيم إلى فئات تضم كافة القيم المتقاربة.

ويكون جدول التوزيع التكراري للمتغير العشوائي الوثاب (عدد الأفراد في الأسر) كما يلي:

عدد الأفراد	٢	٣	٥	٦	٧	٩	١٠	المجموع
التكرار	٤	٤	٨	٦	٤	٢	٢	٣٠

- تبويب البيانات الكمية (الوثابة) في شكل جدول تكراري مزدوج:

كما فعلنا في حالة تبويب البيانات الوصفية في شكل جدول تكراري مزدوج، يمكن تبويب البيانات الكمية الوثابة في شكل جدول تكراري مزدوج، بإنشاء جدول تفريغ من تقسيمات البيانات الكمية الوثابة للظاهرة الأولى أفقيا، وتقسيمات البيانات الكمية الوثابة للظاهرة الثانية، مكونًا بذلك عدد من الخلايا تشترك في جزء من الظاهرة الأولى والثانية معًا، وتفريغ البيانات الزوجية عن كل من تلك الظاهرتين معا بعلامة (/)، حتى يتم الانتهاء من تفريغ جميع البيانات المزدوجة للظاهرتين، وبجمع عدد العلامات داخل كل خلية نحصل على تكرار الخلايا، وبأخذ تقسيمات الظاهرة الأولى، وتقسيمات الظاهرة الثانية مع تكرارات

الخلايا، إلى جانب المجاميع الأفقية والرأسية نحصل على جدول توزيع تكراري مزدوج لتلك البيانات.

مثال: لنفرض أن لدينا ٢٩ قطعة أرض موزعة توزيعًا عشوائيًّا، تختلف فيما بينها في المساحة، لكل منها عدد معين من الملاك على النحو التالي:

مساحة المنطقة (فدان)	عدد الملاك	مساحة المنطقة (فدان)	عدد الملاك
٧	٥	٢	٨
٦	٤	٣	٦
٥	٣	١	٢
٤	٤	٤	٥
٣	٥	٥	٤
٢	٦	٦	٤
١	٣	٧	٥
٦	٤	٦	٤
٥	٦	٧	٣
٣	٧	٥	٨
٦	٧	٤	٥
٧	٢	٣	٢
٥	٦	٢	٤

من الواضح أن عدد ملاك الأراضي يتراوح بين ٢، ٨ أفراد على حين أن مساحة قطع الأرض يتراوح بين ١، ٧ فدان، وعلى ذلك، يمكن تكوين جدول التفريغ كما يلي:

عدد الملاك / عدد الأفدنه	٢	٣	٤	٥	٦	٧	٨	المجموع
١	/	/						٢
٢			/		/		/	٣
٣	/			/	/	/		٤
٤			/	//				٣
٥		/	/		//		/	٥
٦			////			/		٥
٧	/	/		//				٤
المجموع	٣	٣	٧	٥	٤	٢	٢	٢٦

وبالتالي يكون جدول التوزيع التكراري المزدوج لعدد ملاك تلك الأراضي، وعدد الأفدية التي تتكون منها كل قطعة أرض على الصورة التالية:

عدد الملاك / عدد الأفدنه	٢	٣	٤	٥	٦	٧	٨	المجموع
١	١	١						٢

٢			١		١		١	٣
٣	١			١	١	١		٤
٤			١	٢				٣
٥		١	١		٢		١	٥
٦			٤			١		٥
٧	١	١		٢				٤
المجموع	٣	٣	٧	٥	٤	٢	٢	٢٦

ب - تبويب البيانات الكمية (المستمرة):

- تبويب البيانات الكمية (المستمرة) في شكل جدول تكراري بسيط:

تبويب البيانات الكمية (المستمرة) المتصلة في شكل جدول تكراري بسيط يتطلب تكوين جدول مكون من ثلاثة أعمدة، الثاني يمثل العلامات، والثالث يمثل التكرارات، أما العمود الأول فيمثل ما يطلق عليه الفئات، حيث يتم تقسيم المدي بين أصغر قيمة تأخذها الظاهرة وأكبر قيمة تأخذها الظاهرة إلى عدة تقسيمات (فئات)، يكون لكل فئة حد أدنى وحد أعلى والفرق بينهما يسمى بطول الفئة حيث:

طول الفئة = الحد الأعلى للفئة - الحد الأدنى للفئة

وإذا كانت أطوال جميع الفئات المقسمة إليها الظاهرة متساوية، يسمى الجدول التكراري في هذه الحالة بالجدول التكراري المنتظم، أما إذا كان هناك فئة واحدة على الأقل طولها يختلف عن أطوال باقي الفئات، فيقال أن الجدول التكراري غير منتظم، أي إذا كانت البيانات مفصلة في جزء، ومجملة في جزء آخر، إلى جانب أن هناك رغبة في دراسة فئات ذات أطوال معينة لأهمية هذه الفئات بغض النظر عن الفئات الأخرى، وإذا كان الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة معروفين (موجودين)، يقال أن الجدول التكراري مغلق، أما إذا كان الحد الأدنى للفئة الأولى، وكذا الأعلى للفئة الأخيرة غير معروفين (غير موجودين) يقال أن الجدول التكراري مفتوح، وإذا كان الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة فقط غير موجود يقال الجدول التكراري مفتوح من طرفه الأعلى، ويطلق على القيمة التي تقع في نصف المسافة بين الحد الأدنى والحد الأعلى للفئة بمركز هذه الفئة، حيث:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة + الحد الأعلى للفئة}}{\text{٢}}$$

وفي الغالب يرمز لمركز الفئة بالرمز س ر.

ويلزم أن يكون عدد الفئات، وأطوال كل فئة من الفئات مناسب، بحيث لا يكون عدد الفئات كبيرًا أو طول الفئة صغيرًا، وبالتالي يفقد التلخيص أهميته، وألا يكون عدد الفئات صغيرًا، أو طول الفئة كبيرًا، فيفقد التوزيع كثيرًا من تفاصيله، وإذا كانت ن تمثل عدد مفردات الظاهرة، فإن عدد الفئات يمكن الحصول عليها من العلاقة:

عدد الفئات = ١ + ٣,٣ لوغاريتم ن ٠٠٠٠٠٠٠ أو

عدد الفئات = ٢,٥ $\sqrt[4]{ن}$ ، أي = ٢,٥ للجذر الرابع لـ ن

وتقريبا يمكن تحديد العلاقة بين عدد الفئات وعدد مفردات الظاهرة على الصورة:

	١٠	٢٠	٣٠	٤٠	٥٠	١٠٠	١٠٠٠
عدد الفئات	٤,٣	٥,٣	٥,٩	٦,٣	٦,٦	٧,٦	١٠,٩

ومن الواضح أن عدد الفئات يلزم ألا يقل عن ٥ فئات، وقد وجد أنه من غير المستحب أن يزيد عدد الفئات عن ٢٠ فئة، وبعد ذلك يتم تحديد الحدود الدنيا والعليا للفئات، سواءً كان التوزيع منتظمًا أو غير منتظم، ويمكن في حالة تكوين جدول تكراري منتظم مغلق أن يتم تحديد عدد الفئات من العلاقة السابقة، ثم يتم تحديد طول الفئة من العلاقة:

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

ويعرف طول الفئة بأنه مدى الفئة، أي الفرق بين حدها الأعلى وحدها الأدنى، والفئة في الجدول حدها الأدنى هو الحد الأدنى للقيم، وطولها هو طول الفئة، ونهايتها عبارة عن الحد الأدنى + طول الفئة.

وحدود الفئات التي تصلح في حالة المتغيرات المستمرة، هي التي تجعل كل فئة تبدأ مباشرة حيث تنتهي الفئة السابقة لها مباشرة، دون أن يحدث تداخل بين الفئات أو يترك ثغرات بينها، فإذا كان الحد الأدنى للفئة الأولي ١٠، والحد الأدنى للفئة الثانية ٢٠، فإن الفئة الأولى تكون ١٠ وأقل من ٢٠ بمقدار طفيف، وتكتب عادة ١٠-، وإذا كان الحد الأدنى للفئة الثالثة ٣٠، فإن الفئة الثانية تكون ٢٠ وأقل من ٣٠ وتكتب ٢٠-، وأما الفئة الأخيرة فهي فقط التي يكتب حديها، فإذا كان الحد الأدنى للفئة الأخيرة ٧٠ والأعلى ٨٠، فيكتب ٧٠ إلى ٨٠ أو ٧٠- ٨٠.

وبعد ذلك يتم تفريغ البيانات كما سبق بوضع علامة (/)، تعبر عن البيان في الفئة المناسبة له دون تداخل بين الفئات، وبأخذ العمود الأول (الفئات)، والعمود الأخير (التكرارات)، نحصل على جدول توزيع تكراري بسيط للبيانات الكمية المستمرة (المتصلة)، ويراعى على عكس الحالات الأربعة السابقة، أن معالم القيم الأصلية تضيع في الجداول التكرارية للبيانات الكمية المتصلة، ولا يعرف شيئًا عن أي مفردة من المفردات الأصلية،

سوى أنها تنتمي إلى فئة معينة محدودة بحدين معلومين، ولا يمكن من واقع الجدول التكراري معرفة القيمة الأصلية، وما إذا كانت تقع بالقرب من بداية الفئة أو بالقرب من نهايتها، ولذلك يفترض أن جميع القيم تحصل على قيم متساوية، كل منها يساوي قيمة مركز هذه الفئة، وهذا الافتراض مناسب وعملي، والخطأ الناشئ عنه يكون من الصغر بحيث يمكن التجاوز عنه، خاصة إذا كان طول الفئة صغير.

مثال: البيانات التالية توضح أطوال بعض الطرق الفرعية (خمسون طريقًا)، والمطلوب عرض هذه البيانات في صورة جدول تكراري.

١٠	٩	١٥	١٣	١١
٩	١٠	١٥	٨	٢٠
١٠	١٢	١٣	٨	١٠
٩	١٣	١٥	١٢	١٠
١٦	١١	١٠	١٧	١١
١٦	١٣	٩	١١	١٤
١٠	١٨	١٧	١٢	١٠
١٥	١٣	١٣	١٥	٨
١٢	١٤	١٢	٩	١١
١١	١٣	١٦	١٤	١٢

خطوات الحساب:

١- حساب المدى:

المدى = ٢٠ – ٨ = ١٢، أي أن المدى هو الفرق بين أكبر قيمة وأصغر قيمة في التوزيع.

٢- تحديد عدد الفئات، ويجب ألا يكون عدد الفئات كبيرًا، فيقل المعنى من اختصار البيانات، ولا يكون صغيرًا فتفقد البيانات جدواها، وفي أغلب الأحوال لا يزيد عدد الفئات عن عشرين فئة، ولا يقل عن خمس فئات.

ويمكن تحديد عدد الفئات بإحدى المعادلتين التاليتين:

(أ) معادلة يول YULE

عدد الفئات = ٢,٥ $\sqrt[4]{\text{عدد المفردات}}$

وتصلح هذه المعادلة عندما يكون عدد القيم أقل من ١٠٠٠

(ب) معادلة STURGES

عدد الفئات = ١+ ٣,٣ × لوغاريتم عدد القيم. وتصلح تلك المعادلة عندما يكون عدد القيم أكثر من ١٠٠٠.

وفي مثالنا السالف الإشارة له، وبتطبيق القاعدة الأولى فإن:

عدد الفئات = ٢,٥ $\sqrt[4]{\text{عدد المفردات}}$

= ٢,٥ × ٢.٦٥٩١ = ٦.٦٤ = ٧ تقريبًا

ومن ثم فإن طول الفئة = المدى / عدد الفئات = ١٢ / ٧ = ١.٧ = ٢ تقريبًا

وتكون الفئات كما يلي:

الحد الأعلى للفئة الأولى = الحد الأدنى للقيم + طول الفئة

= ٨ + ١٠ = ١٠، وتكتب هكذا (٨ - ١٠)

وبنفس الطريقة تكون الثانية (١٠–١٢)، (١٢–١٤)، (١٤–١٦)، (١٦–١٨)، (١٨–٢٠)، (٢٠–٢٢).

الفئات	العلامات (الحزم)	التكرار
٨-	𝍸 ///	٨
١٠-	𝍸 𝍸 //	١٢
١٢-	𝍸 𝍸 ////	١٤
١٤-	𝍸 ///	٨
١٦-	𝍸 /	٦
١٨-٢٠	/	١
٢٠-٢٢	/	١
الإجمالي		٥٠

وفي الغالب يستبعد العمود الثاني؛ لأنه ما هو إلا وسيلة للتعرف على التكرارات داخل كل فئة، ويكون الجدول عندئذ على الصورة التالية:

التوزيع التكراري البسيط

الفئات	التكرار
٨-	٨
١٠-	١٢
١٢-	١٤
١٤-	٨
١٦-	٦

١٨-٢٠	١
٢٠-٢٢	١
الإجمالي	٥٠

مثال: فيما يلي أطوال الروافد النهرية من الرتبة الثالثة لأحد النظم النهرية الناضجة (كم)، والمطلوب وضعها في جدول تكراري بسيط.

٨٣	٦٩	٩٠	٦٣	٧٧	٦٩	٩٤	٧٥	٨٩	٨٣
٧٧	٦٠	٩٤	٧٢	٨٥	٧١	٧٥	٨٩	٥٨	٧٤
٦٥	٧٣	٦٣	٩٥	٦٤	٦٣	٧١	٧٥	٧٧	٨٨
٧٦	٩٣	٧٦	٦٨	٦٩	٦٧	٦٠	٨٣	٧١	٧٥
٧٦	٧٥	٩٥	٦٠	٨٤	٩٥	٧٣	٨٧	٧٧	٦٤
٦٤	٧٩	٨٥	٧٧	٧٣	٧٨	٧٥	٦٦	٦٥	٩٥
٨٠	٨٧	٧٩	٦٣	٥٤	٦٥	٧١	٧٤	٧٧	٥٩
٦٧	٧١	٦٤	٧٧	٨٥	٨٧	٧٨	٧٤	٧٥	٨٣

نبدأ أولاً بتحديد أصغر قيمة وأكبر قيمه بينهم (أي المدى)، فأصغر قيمة هي ٥٤، وأكبر قيمة هي ٩٥، والمدى = ٤١، وإذا ما رتبنا القيم من ٥٤ إلى ٩٥، ووجدنا تكراراتها، كما فعلنا في المثال السابق، فستكون لدينا ٤١ قيمة وتكراراتها، وواضح أنه بوجود هذا العدد الكبير من القيم، فإن التلخيص على النحو المبين في المثال السابق لا يفي بالغرض في هذه الحالة، لذلك تكون الخطوة التالية هي وضع القيم المتقاربة في مجموعات أو "فئات"، ويجب ملاحظة ألا يكون عدد الفئات كبيرًا، وبالتالي يفقد التلخيص أهميته، وألا يكون عدد الفئات صغيرًا جدًّا فيفقد التوزيع كثيرًا من تفاصيله، ويختلف العدد المناسب للفئات من توزيع إلي آخر وفقا للهدف من إجراء التوزيع التكراري، وعمومًا يمكن القول أن عدد الفئات يجب ألا يقل عن ٥، وألا يزيد عن ٢٥. وعلى قدر الإمكان من الأفضل أن تكون الفئات متساوية.

المدى = ٩٥ ــ ٥٤ = ٤١

عدد الفئات = ١ + ٣,٣ لو ٨٠ = ٧,٢٨ = ٨ فئات تقريبًا

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ = $\frac{٤١}{٨}$ = ٥,١٢٥ = ٥ تقريبًا

ونظرًا لعملية التقريب لعدد الفئات قد يزيد عددها قليلاً أو يقل عددها قليلاً عن عدد الفئات التي تم الحصول عليه من المعادلة السابقة، بعد أن تمدنا أن يكون طول الفئة ٥ وهو عدد دائري حتي يسهل قراءة حدود الفئات مباشرة.

٥٠ إلى أقل من ٥٥	٥٠ ــ	٥٠ ــ ٥٥
٥٥ إلى أقل من ٦٠	٥٥ ــ	٥٥ ــ ٦٠

٦٠ إلى أقل من ٦٥	٦٠ ـ	٦٠ ـ ٦٥
٦٥ إلى أقل من ٧٠	٦٥ ـ	٦٥ ـ ٧٠
٧٠ إلى أقل من ٧٥	٧٠ ـ	٧٠ ـ ٧٥
٧٥ إلى أقل من ٨٠	٧٥ ـ	٧٥ ـ ٨٠
٨٠ إلى أقل من ٨٥	٨٠ ـ	٨٠ ـ ٨٥
٨٥ إلى أقل من ٩٠	٨٥ ـ	٨٥ ـ ٩٠
٩٠ إلى أقل من ٩٥	٩٠ ـ	٩٠ ـ ٩٥
٩٥ إلى أقل من ١٠٠	٩٥ ـ	٩٥ ـ ١٠٠

ويجب الحذر من كتابة الفئات على الشكل الآتي:

٥٠ ـ ٥٤
٥٥ ـ ٥٩
٦٠ ـ ٦٤
٦٥ ـ ٦٩
٧٠ ـ ٧٤
٧٥ ـ ٧٩
٨٠ ـ ٨٤
٨٥ ـ ٨٩
٩٠ ـ ٩٤
٩٥ ـ ١٠٠

إذ أن هذه الطريقة في الكتابة تكون صحيحة في حالة القيم الوثابة فقط، فالقيمة ٥٤ تكون في الفئة الأولى، والقيمة ٥٠ تكون في الفئة الثانية، أما في حالة القيم المستمرة لا يمكن كتابة حدود الفئات بهذه الطريقة؛ لأن القيم ما بين ٥٤ و ٩٥ مثال ٥٤,٣ ، ٥٤,٧، لا يمكن تمثيلها في الفئة الأولى، ولا في الفئة الثانية.

وفي مثالنا هذا سنكتب الفئات كالآتي: ٥٠ ـ ، ٥٥ ـ ، ....، وآخر فئة سنكتبها على الصورة ٩٥ وأقل من ١٠٠، لتحديد الحد الأعلى للفئة الأخيرة، وبعد كتابة الفئات بهذه الصورة في العمود الأول من جدول التفريغ، نضع العلامات على النحو المبين سلفًا، ثم تضع في العمود الأخير عدد الطلبة (أو التكرار)، ويمثل الجدول التالي جدول التفريغ التكراري الممثل لهذه البيانات.

جدول تكراري لأطوال الروافد النهرية من الرتبة الثالثة لأحد النظم النهرية

أطوال الروافد النهرية	العلامات	عدد الروافد
٥٠ ـ	/	١
٥٥ ـ	//	٢
٦٠ ـ	///// ///// /	١١

٦٥ –	~~////~~ ~~////~~	١٠
٧٠ –	// ~~////~~ ~~////~~	١٢
٧٥ –	/ ~~////~~ ~~////~~ ~~////~~ ~~////~~	٢١
٨٠ –	/ ~~////~~	٦
٨٥ –	//// ~~////~~	٩
٩٠ –	////	٤
٩٥ وأقل من ١٠٠	////	٤
المجموع		٨٠

وأحيانًا يكون من المفيد وضع التكرارات في صورة نسب، وهي التكرارات النسبية، ونحصل عليها بقسمة التكرار في كل فئة على مجموع التكرارات الكلية، ويوضح الجدول التالي التكرارات النسبية للتوزيع السابق.

جدول تفريغ لأطوال ٨٠ رافد نهري من الرتبة الثالثة لأحد النظم النهرية

عدد الروافد النهرية	التكرارات	التكرار النسبي
٥٠ –	١	١,٢٥
٥٥ –	٢	٢,٥
٦٠ –	١١	١٣,٧٥
٦٥ –	١٠	١٢,٥
٧٠ –	١٢	١٥
٧٥ –	٢١	٢٦,٢٥
٨٠ –	٦	٧,٥
٨٥ –	٩	١١,٢٥
٩٠ –	٤	٥
٩٥ وأقل من ١٠٠	٤	٥
المجموع	٨٠	١٠٠٪

ب - تبويب البيانات الكمية المستمرة في شكل جدول تكراري مزدوج:

رأينا كيفية عمل جدول توزيع تكراري مزدوج في حالة بيانات وصفية، وسنرى هنا كيفية عمل هذا الجدول في حالة بيانات كمية، فإذا كانت هناك علاقة بين ظاهرتين كميتين كل منهما متصل (مستمر)، فإن تلك البيانات تبوب في شكل جدول تكراري، حيث يتم إنشاء جدول تفريغ يتكون أفقيًا من فئات الظاهرة الأولى، ورأسيًا من فئات الظاهرة الثانية، مكونًا عددًا من الخلايا المشتركة في الفئة من الظاهرة الأولى، وفئة من الظاهرة الثانية، وتفرغ البيانات الزوجية تبعًا للخلية التي تقع فيها بوضع علامة (/)، حتى يتم الانتهاء من تفريغ جميع البيانات المزدوجة، وبجمع عدد العلامات داخل كل خلية نحصل على تكرار الخلايا، وبأخذ فئات الظاهرة الأولى والثانية والتكرارات في الخلايا إلى جانب المجاميع في

الصف والعمود الأخير نحصل على جدول توزيع تكراري مزدوج للبيانات الكمية المستمرة، وعلى سبيل المثال يمكن تكوين جدول توزيع تكراري للعلاقة بين أطوال الروافد النهرية في عدد ٥٠ حوض نهري، ومساحة هذه الأحواض، حتى نعرف العلاقة بين طول النظم التصريفية، ومساحة أحواضها النهرية.

المساحة (كم^{2}) / الطول (كم)	١٦٠٠ –	١٧٠٠ –	١٨٠٠ –	١٩٠٠ – ٢٠٠٠	المجموع
٦٠٠ –	٤	٢	١		٧
٧٠٠ –	٣	٥	٤		١٢
٨٠٠ –		٦	٧	٣	١٦
٩٠٠ – ١٠٠٠		٤	٢	٩	١٥
المجموع	٧	١٧	١٤	١٢	٥٠

ويلاحظ أنه يمكن دمج الجداول التكرارية المزدوجة للبيانات الكمية الوثابة والمستمرة مع البيانات الوصفية، فيمكن أن يكون المتغير أفقيًا في صورة وصفية، والمتغير رأسيًا في صورة كمية، فعلى سبيل المثال يمكن أن يكون جدول التوزيع التكراري المزدوج على الصورة التالية فيما بين السن ومرض السكر في منطقة معينة:

العمر / المرض	أقل من ١٥ سنة	١٥ – ٣٥	٣٥ – ٦٥	أكثر من ٦٥	المجموع
مريض	٥	١٠	١٥	٢٠	٥٠
غير مريض	٢٥٠	٣٠٠	٢٥٠	٢٠٠	١٠٠٠
المجموع	٢٥٥	٣١٠	٢٦٥	٢٢٠	١٠٥٠

أو على الصورة التالية فيما بين عدد أفراد الأسرة في قرية معينة وعدد الحجرات في المسكن.

عدد الأفراد / عدد الغرف	٥	٦	٧	٨	المجموع
٣ –	٥	٤			٩
٤ –	٤	٧	١		١٢
٥ –	٦	٧	٢		١٥
٦ – ٧	٧	١٠	٧	١٢	٣٦
المجموع	٢٢	٢٨	١٠	١٢	٧٢

جـ - تبويب البيانات الوصفية أو الكمية في شكل جدول تكراري بسيط أو مزدوج في صورة تكرارات نسبية:

في بعض الأحيان قد يكون من المناسب أن تعرض البيانات في شكل جدول توزيع تكراري نسبي، بإظهار تكرار كل فئة أو خلية، كنسبة من المجموع الكلي للتكرارات، حيث:

$$\text{التكرار النسبي} = \frac{\text{التكرار الأصلي}}{\text{مجموع التكرارات}}$$

ويمكن إعادة تبويب البيانات الخاصة بالتوزيعات التكرارية البسيطة والمزدوجة، سواءً كانت البيانات وصفية أو كمية، في شكل جداول توزيع تكراري نسبي، فعلى سبيل المثال نجد أن البيانات الوصفية تبوب في شكل جدول تكراري بسيط نسبي، على الصورة التالية:

جدول التوزيع التكراري النسبي للمتغير العشوائي الكمي الوثاب (عدد الأفراد في الأسر):

عدد الأفراد	٢	٣	٥	٦	٧	٩	١٠	المجموع
التكرار	٤	٤	٨	٦	٤	٢	٢	٣٠
التكرار النسبي	٠,١٣	٠,١٣	٠,٢٧	٠,٢٠	٠,١٣	٠,٠٧	٠,٠٧	١

ومن الملاحظ أن مجموع التكرارات النسبية لجميع قيم الظاهرة (أو صفاتها) يساوي الواحد الصحيح.

	رجل	امرأة	المجموع
تعمل	٠.٣٧	٠.٣٠	٠.٦٧
لا يعمل	٠.٠٧	٠.٢٦	٠.٣٣
المجموع	٠.٤٤	٠.٥٦	١,٠٠

د- تبويب البيانات في شكل جدول تكراري متجمع:

في كثير من الأحيان نحتاج لمعرفة عدد المفردات (التكرارات) التي تقل قيمتها عن حد معين، أو تلك التي تساوي قيمتها أو تزيد عن حد معين، ويتم ذلك عن طريق تبويب البيانات في شكل جدول تكراري متجمع، وتنقسم الجداول المتجمعة إلى نوعين من الجداول، الأول يسمى جدول التوزيع التكراري المتجمع الصاعد، والذي يتكون من خانتين: الأولى تمثل الفئات العليا (أو بصفة عامة حدود المتغير الكمي الوثاب أو المستمر، وحدود المتغير الوصفي الذي يمكن ترتيبه)، وتمثل عد التكرارات التي تقل عن الحدود العليا للفئات، والخانة الثانية تمثل التكرارات المتجمعة الصاعدة، حيث يتم تجميع البيانات (التكرارات) من جهة الفئات الصغيرة إلى الكبيرة.

والنوع الثاني من الجداول يسمى جدول التوزيع التكراري المتجمع الهابط والذي يتكون من خانتين أيضًا، الأولى تمثل حدود الفئات الدنيا وتمثل عند التكرارات التي تساوي أو تزيد عن الحدود السفلي للفئات، والخانة الثانية تمثل التكرارات المتجمعة الهابطة، حيث يتم تجميع البيانات (التكرارات) من جهة الفئات الكبيرة إلى الصغيرة.

جدول تكراري متجمع صاعد:

إذا كان لدينا توزيع تكراري وأردنا معرفة عدد المفردات التي تكون قيمتها أقل من قيمة معينة، نوجد ما يسمى "بالتكرار المتجمع الصاعد"، ولتوضيح ذلك إذا نظرنا إلى الجدول التالي، الذي يوضح الدخل الشهري لعدد ١٠٠ أسرة في حي معين من أحياء القاهرة الكبرى، وأردنا معرفة كم أسرة دخلها الشهري أقل من ٤٠٠ جنيه، فسيكون الجواب: ليس هناك أسر يقل دخلها الشهري عن ٤٠٠ جنيه، وإذا أردنا معرفة كم أسرة دخلها الشهري أقل من ٤٥٠ جنيه سيكون الجواب: أسرة واحدة فقط، وإذا أردنا معرفة كم أسرة دخلها الشهري أقل من ٥٠٠ جنيه سيكون الجواب (١ + ٤ = ٥)، وبالمثل فإن عدد الأسر التي يقل دخلهم الشهري عن ٥٥٠ جنيه = (١ + ٤ + ١٣ = ١٨)، ولتنفيذ جدول التوزيع التكراري المتجمع الصاعد نتبع الخطوات الآتية:

- نضيف إلى الفئات فئة قبل الأولى، وسيكون التكرار المناظر لها صفر، فمثلا يأخذ الجدول التالي فئة أخرى قبل الفئة الأولى وهي: أقل من ٤٠٠ جنيه، سيكون التكرار المناظر لها صفر.

- نضيف عمودين: الأول يبين "أقل من الحد الأعلى للفئة"، الثاني خاص بالتكرار المتجمع الصاعد.

- نحسب التكرار المتجمع الصاعد، فبالنسبة لأقل من ٤٠٠ جنيه يكون التكرار المتجمع الصاعد = صفر، ثم بعد ذلك نضيف التكرار المتجمع الصاعد إلى تكرار الفئة التالية، فيكون التكرار المتجمع الصاعد ١، ثم ٥، ثم ١٨ ....، وهكذا إلى أن نصل إلى أقل من الحد الأعلى للفئة الأخيرة، وهنا يجب أن يكون التكرار المتجمع الصاعد مساوي لمجموع التكرارات، أي أمام أقل من ٩٠٠ جنيه يكون التكرار المتجمع الصاعد ١٠٠.

التكرار المجتمع الصاعد للدخل الشهري لعدد ١٠٠ أسرة

في حي من أحياء مدينة القاهرة الكبرى

الفئة	التكرار	أقل من الحد الأعلى للفئة	التكرار المتجمع الصاعد	التكرار المتجمع الصاعد النسبي
أقل من ٤٠٠	٠	أقل من ٤٠	٠	٠
٤٠٠ –	١	أقل من ٤٥	١	٠,٠١
٤٥٠ –	٤	أقل من ٥٠	٥	٠,٠٥
٥٠٠ –	١٣	أقل من ٥٥	١٨	٠,١٨
٥٥٠ –	١٧	أقل من ٦٠	٣٥	٠,٣٥
٦٠٠ –	٢١	أقل من ٦٥	٥٦	٠,٥٦
٦٥٠ –	١٨	أقل من ٧٠	٧٤	٠,٧٤
٧٠٠ –	١٥	أقل من ٧٥	٨٩	٠,٨٩
٧٥٠ –	٨	أقل من ٨٠	٩٦	٠,٩٦
٨٠٠ –	٣	أقل من ٨٥	٩٩	٠,٩٩
٨٥٠ وأقل من ٩٠٠	١	أقل من ٩٠	١٠٠	١,٠٠
المجموع	١٠٠			

ويمكن إيجاد التكرار المتجمع الصاعد النسبي بقسمة كل تكرار متجمع صاعد على مجموعة التكرارات.

التكرار المتجمع الهابط:

إذا أردنا معرفة أحد المفردات التي تكون قيمتها متساوية، وأكبر من قيمة معينة نوجد التكرار المتجمع الهابط، فمثلا لو أردنا معرفة كم أسرة دخلها الشهري ٤٠٠ جنيه فأكثر، فالجواب يكون: كل الأسر أي ١٠٠ أسرة، وإذا أردنا أن نعرف كم أسرة دخلها الشهري ٤٥٠ جنيه فأكثر فالجواب يكون: (١٠٠ – ١ = ٩٩ أسرة)، وبالمثل فعدد الأسر الذين يحصلون على دخل شهري ٥٠٠ جنيه فأكثر يكون: (٩٩ – ٤ = ٩٠ أسرة)، وهكذا إلى أن نصل إلى ٩٠٠ فأكثر، فيكون عدد الأسر صفر.

ولإيجاد التكرار المتجمع الهابط نضيف عمودين إلى جدول التوزيع التكراري، العمود الأول يعطي الحد الأدنى للفئة فأكثر، والعمود الثاني خاص بالتكرار المتجمع الهابط، ثم بعد ذلك نبدأ بالفئة الأولى، فأمام ٤٠٠ جنيه فأكثر يكون التكرار المتجمع الهابط ١٠٠، ثم بعد ذلك نحصل على التكرار المتجمع الهابط للفئة الثانية عن طريق طرح تكرار الفئة من ١٠٠، وهكذا بالطرح المتتالي نحصل على نفس هذا التكرار عن طريق الجمع المتتالي للتكرارات من أسفل الجدول، فأمام ٩٠٠ جنيه فأكثر يكون التكرار المتجمع يساوي صفر، وأمام ٨٥٠ جنيه فأكثر يكون التكرار المتجمع (٠ + ١ = ١)، وأمام ٨٠٠ جنيه فأكثر يكون

التكرار المتجمع (١ + ٣ = ٤)، وهكذا نحصل على التكرار المتجمع عن طريق جمع التكرارات من أسفل.

التكرار المجتمع الهابط للدخل الشهري لعدد ١٠٠ أسرة

في حي من أحياء مدينة القاهرة

الفئة	التكرار	أقل من الحد الأعلى للفئة	التكرار المتجمع الصاعد	التكرار المتجمع الصاعد النسبي
٤٠٠ –	١	٤٠ فأكثر	١٠٠	١,٠٠
٤٥٠ –	٤	٤٥ فأكثر	٩٩	٠,٩٩
٥٠٠ –	١٣	٥٠ فأكثر	٩٥	٠,٩٥
٥٥٠ –	١٧	٥٥ فأكثر	٨٠	٠,٨٢
٦٠٠ –	٢١	٦٠ فأكثر	٦٥	٠,٥٦
٦٥٠ –	١٨	٦٥ فأكثر	٤٤	٠,٤٤
٧٠٠ –	١٥	٧٠ فأكثر	٢٦	٠,٢٦
٧٥٠ –	٧	٧٥ فأكثر	١١	٠,١١
٨٠٠ –	٣	٨٠ فأكثر	٤	٠,٠٤
٨٥٠ وأقل من ٩٠٠	١	٨٥ فأكثر	١	٠,٠١
المجموع	١٠٠	٩٠ فأكثر	٠	٠

ولإيجاد التكرار المتجمع الهابط بالنسب، نقسم التكرار المتجمع الهابط على مجموعة التكرارات.

مثال:

البيانات التالية - الإيرادات السنوية لخمسين مجرى نهري في العالم (أرقام افتراضية) بالمليار متر مكعب، والمطلوب عمل جدول تكراري بسيط.

٤٣	٢٦	٢٥	٨	٢٦
٢٠	٢٦	٤٠	٣٥	٢٠
١٦	٣٨	٤٠	٣٦	٣٨
٣٦	١٧	٣٢	٤٢	٢٧
٣٢	٢٤	٣٣	٣٢	٣٠
١٥	٢٧	٣	٣٠	٣٢
٣٥	٣٧	٢٦	٣٠	٢٧
٤٥	٢٧	٢٥	٢٠	١٢
٣٧	١٢	٣٥	٣٥	٣٥
٢٢	٣٤	٣٥	٣٥	٣٦

الحل:

(أ) حساب المدى: المدى أكبر قيمة في التوزيع - أصغر قيمة في التوزيع.

= ٤٥ – ٣ = ٤٢

(ب) حساب عدد الفئات: نلاحظ أن عدد المفردات أقل من ١٠٠٠، ولحساب عدد الفئات نطبق القانون التالي:

عدد الفئات = ٢,٥ $\sqrt[4]{\text{عدد المفردات}}$

عدد الفئات = ٢,٥ $\sqrt[4]{٥٠}$ = ٢,٥ × ٢.٦٥٩١ = ٦.٦٤ فئة أي تقريبًا ٧ فئات.

(جـ) طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ = $\frac{٤٢}{٧}$ = ٦

ويعرف طول الفئة بأنه مدى الفئة.

ولتركيب الجدول نبدأ بالفئة الأولى، فنلاحظ أن حدها الأدنى هو أقل القيم في التوزيع وهو (٣)، ولحساب حدها الأعلى نقول:

الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة = ٣ + ٦ = ٩

وتكتب كما يلي (٣ – ٩)، أي ثلاثة فأكثر حتى بداية ٩ والفئة الثانية (عندما تكون الفئات متصلة، بداية الفئة الثانية هو نهاية الفئة الأولى)، حدها الأدنى ٩، ولحساب الحد الأعلى نقول:

الحد الأعلى للفئة الثانية = ٩ + ٦ = ١٥

وتكتب كما يلي (٩ – ١٥)، والثالثة تحسب بنفس طريقة الحساب السالفة، وتكتب (١٥ - ٢١)، والرابعة (٢١ - ٢٧)، والخامسة (٧٧ – ٣٢) والسادسة (٣٣ - ٣٩)، والسابعة (٣٩ - ٤٥).

أما إذا كانت الفئات منفصلة فتكتب كما يلي:

(٣ – ٩)، (١٠ – ١٦)، (١٧ – ٢٣)، (٢٤ – ٣٠)، (٣١ – ٣٧)، (٣٨ – ٤٤)، ومن ثم يكون عدد الفئات في تلك الحالة ستة بدلاً من سبعة، هذا إذا أهملنا الكسر في قاعدة YULE فيكون عدد الفئات ستة.

ولتنفيذ جدول التكرار البسيط، ننفذ قبله جدولاً بعلامات التكرار، والذي يتم بعمل علامات للأرقام التي تقع داخل كل فئة، وكل أربع علامات تحزم بعلامة عكسية كما يلي:

الإيرادات السنوية لخمسين مجرى نهري في العالم

فئات الإيرادات السنوية	العلامات	التكرار
٣ – ٩	//	٢
٩ – ١٥	//	٢
١٥ – ٢١	𝍸 /	٦
٢١ – ٢٧	𝍸 ///	٨
٢٧ – ٣٣	𝍸 𝍸 /	١١
٣٣ – ٣٩	𝍸 𝍸 𝍸 /	١٦
٣٩ – ٤٥	𝍸	٥
المجموع		٥٠

ومن الجدول يمكن إعداد جدول خالٍ من العلامات، ويصبح الجدول من عمودين الأول به الفئات (متصلة في هذا المثال)، والثاني به التكرارات كما يلي:

الفئات	التكرارات
٣ – ٩	٢
٩ – ١٥	٢
١٥ – ٢١	٦
٢١ – ٢٧	٨
٢٧ – ٣٣	١١
٣٣ – ٣٩	١٦
٣٩ – ٤٥	٥
المجموع	٥٠

ويلاحظ من الجدول السابق أن أكثر الفئات تكرارًا، هي الفئة (٣٣ – ٣٩) متر مكعب سنويًا، إذ يبلغ عدد المجاري النهرية بها ١٩ مجرىً، ويطلق على هذه الفئة (الفئة المنوالية).

كما يلاحظ أنه في حالة وجود رقم بين التوزيعات كبير جدًّا، فإن الباحث سيضطر إلى عمل جدول مفتوح من أسفل، هذا في حالة ما إذا كانت البيانات منتظمة قبل ذلك، ففي المثال السابق إذا كان هناك مجرىً نهريًّا حقق إيرادًا أعلى من٤٥ مليار متر مكعب سنويًّا أي ٨٠ مليار متر مكعب سنويًّا مثلاً، فتكتب الفئة الأخيرة ٤٥- أي ٤٥ فأكثر، ذلك لأن هناك رقما شاذًّا لن نصل إليه إلا بعمل فئات عددية بعد ذلك.

كما أنه يلاحظ إذا وجد رقمًا صغيرًا جدًّا أقل بكثير من الحد الأدنى للفئة الأولى، كأن يكون في مثالنا ٠,٥ مليار متر مكعب، فعندئذ يكون الجدول مفتوحًا لأسفل، وتكتب الفئة الأولى أقل من٣، وإذا تعرض الباحث للحالتين الآنف الإشارة لهما، فإن الجدول عندئذ يكون مفتوح الطرفين، بمعنى أن الفئة الأولى تقرأ ٣ فأقل والفئة الأخيرة تقرأ ٤٥ فاكثر.

وفي بعض الأحيان قد يضطر الباحث إلى إعداد جدول التكرار، ويلتزم بتساوي أطوال الفئات، وذلك عند عمل فئات توزيع الدخل في دولة ما، أو توزيع الأرض الزراعية على مالكيها، أو فئات الدخول للسائحين القادمين لزيارة جمهورية مصر العربية، ويمكن عمل جدول التكرار المتجمع الصاعد والهابط من المثال السابق كما يلي:

أولا : جدول التكرار المتجمع الصاعد:

يلاحظ أن هذا الجدول مكون من أربعة أعمدة، الأول يوضح الفئات، والثاني يوضح التكرارات، والثالث يوضح الحدود العليا للفئات، والأخير يوضح التكرار المتجمع الصاعد كما يلي:

جدول التكرار المتجمع الصاعد

الفئات	التكرارات	الحدود العليا للفئات	التكرار المتجمع الصاعد
٣ – ٩	٢	أقل من ٩	٢
٩ – ١٥	٢	أقل من ١٥	٤
١٥ – ٢١	٦	أقل من ٢١	١٠
٢١ – ٢٧	٨	أقل من ٢٧	١٨
٢٧ – ٣٣	١١	أقل من ٣٣	٢٦
٣٣ – ٣٩	١٦	أقل من ٣٩	٤٥
٣٩ – ٤٥	٥	أقل من ٤٥	٥٠

ويلاحظ من العمود الأخير، الجمع الجبري للتكرارات تباعًا.

ويمكن بنفس الطريقة إعداد جدول التكرار المتجمع الهابط كما في الجدول التالي:

التكرار المتجمع الهابط

الفئات	التكرارات	الحدود العليا للفئات	التكرار المتجمع الصاعد
٣	٢	٢ فأكثر	٥٠
٩ ـ ١٥	٢	٩ ـ	٤٨
١٥ ـ ٢١	٦	١٥ ـ	٤٦
٢١ ـ ٢٧	٨	٢١ ـ	٤٠
٢٧ ـ ٣٣	١١	٢٧ ـ	٣٢
٣٣ ـ ٣٩	١٦	٣٣ ـ	٢١
٣٩ ـ ٤٥	٥	٣٩ ـ	٥

الفصل الثالث

مقاييس النزعة المركزية

الفصل الثالث : مقاييس النزعة المركزية

Measures of Central Tendency

لقد استخدمنا الجداول الإحصائية في الفصل السابق بهدف تلخيص وعرض البيانات الإحصائية عن الظواهر المختلفة ، وفي هذا الفصل سنلجأ إلى تلخيص هذه البيانات بصورة رقمية، وذلك باستخدام بعض المقاييس المعروفة باسم المتوسطات.

ففي كثير من التوزيعات التكرارية نجد أن عددًا كبيرًا من المفردات يميل إلى التجمع حول قيمة متوسطة معينة، ويقل عدد المفردات تدريجيًا كلما بعدنا عن هذه القيمة المتوسطة التي تمثل مركز التوزيع وتسمى هذه الظاهرة: بالنزعة المركزية، أي نزعة المفردات المختلفة إلى التجمع حول مركز التوزيع . ويتضح من ذلك أن لكل مجموعة من البيانات قيمة متوسطة معينة خاصة بها تميزها عن مجموعات البيانات الأخرى، والتي يمكن استخدامها لوصف المجموعة، حيث إنها تحدد مركز أو متوسط المجموعة.

إن مقاييس النزعة المركزية أو المتوسطات شائعة الاستعمال بين الناس وينظر إليها دائمًا على أنها قيم تعبر عن سلوك الصفات المختلفة، ولذلك يهتم بدراستها المشتغلون بالكثير من الشؤون الاقتصادية والاجتماعية ؛ فعلى سبيل المثال يهتم الجغرافيون وواضعوا البرامج الاقتصادية والإنتاجية بمعرفة معدل الزيادة في عدد السكان، أو متوسط إنتاج الفدان من المحاصيل المختلفة، أو متوسط دخل الفرد أو متوسط الإنتاج اليومي لمصنع معين، أو متوسط تكاليف سلعة معينة إلى آخر ذلك من الأمور التي يراد تبيان صفة توزيعها بطريقة مختصرة .

ونود أن نوضح هنا وقبل أن نستطرد في دراسة هذا الموضوع أننا حين نستخدم لفظ متوسط فإننا نعني بذلك عدة مقاييس لتقدير القيمة الوسطى لمجموعة البيانات. وأهم هذه المقاييس هي الوسط الحسابي، والوسيط ، والمنوال. ولكل من هذه المقاييس مميزاته وعيوبه ؛ ولذلك لا نستطيع أن نفضل أحداهما على الآخر تفضيلًا مطلقا ، وفيما يلي ندرس كيفية حساب كل من هذه المقاييس:

أولًا : المتوسط الحسابي Arithmetic Mean:

يعتبر المتوسط الحسابي أكثر أنواع المتوسطات استعمالًا، وأسهلها فهمًا بالنسبة لعامة الناس. وهو عبارة عن مجموع قيم الظاهرة المدروسة مقسومة على عددها أي أن:

المتوسط الحسابي = $\frac{\text{مجموع القيم}}{\text{عددها}}$ أو $\bar{\text{س}} = \frac{\text{مجـ س}}{\text{ن}}$

حيث (سَ) هي الوسط الحسابي، (س) قيم الظاهرة المدروسة (مجـ)تعني المجموع العام لقيم الظاهرة، (ن) عدد قيم الظاهرة .

فإذا كان لدينا القيم التالية: وهي المتوسط الشهري لدرجة الحرارة خلال عام (١٢ شهرًا) في مدينة جدة مثلًا مرتبة على النحو التالي:

٢٥ ، ٣٦ ، ٢٦ ، ٣٠، ٢٨ ، ٢٦ ، ٢٨ ،٣٤ ، ٣٠ ، ٢٨ ، ٣٢، ٣٧

فإذا أردنا إيجاد المتوسط السنوي لدرجة الحرارة خلال ال١٢ شهرًا نقوم بما يلي:

١- نوجد مجموع القيم وذلك بجمع الأرقام السابقة للحصول على مجـ س الذي = ٣٦٠.

٢- إن عدد القيم (ن) = ١٢.

٣- المتوسط الحسابي (سَ) = $\frac{\text{مجـ س}}{\text{ن}}$ = $\frac{٣٦٠}{١٢}$ = ٣٠م

أي أن متوسط درجة الحرارة السنوية من واقع بيانات درجات الحرارة لمدة ١٢ شهرًا يساوي ٣٠م.

ويمكننا الحصول على هذه النتيجة نفسها باستخدام طريقة مختصرة تعتمد على ما يلي:

١- اختيار أي وسط فرضي (و)، والوسط الفرضي: هو أي رقم يقع عليه اختيارنا، ويستحسن أن يكون قريبًا من الوسط الحسابي.

٢- نحسب الانحرافات (ح) عن هذا الوسط الفرضي. وكل انحراف هو عبارة عن قيمة القراءة – قيمة الوسط الفرضي أي أن : ح= س- و

٣- نجمع الانحرافات ونقسم المجموع على عدد القراءات ؛ أي نحسب متوسط الانحرافات، ثم نضيف الناتج إلى الوسط الفرضي فينتج الوسط الحسابي (أي نطبق القانون):سَ = و+ $\frac{\text{مجـ ح}}{\text{ن}}$

أي أن الوسط الحسابي يساوي الوسط الفرضي + متوسط الانحرافات. فإذا كان المطلوب حساب المتوسط الحسابي لمثال درجات الحرارة السابق بهذه الطريقة فإننا نقوم بهذه العملية على النحو المذكور في الجدول ؛ حيث نستخرج مجموع الانحرافات عن الوسط الفرضي (و) الذي اختير وهو رقم (٢٨)، وبقسمة مجموع الانحرافات على عددها مجـ ح / ن

وإضافة الناتج إلى الوسط الفرضي نحصل على وسط حسابي . ومن خلال الجدول نجد أن :
سَ = ٣٠ وهو الجواب السابق نفسه .

حساب الوسط الحسابي عن طريق الوسط الفرضي لقيم غير مبوبة

القيم س	**ح (س- و)**
٢٥	٣-
٣٦	٨+
٢٦	٢-
٣٠	٢+
٢٨	صفر
٢٦	٢-
٢٨	صفر
٣٤	٦+
٣٠	٢+
٢٨	صفر
٣٢	٤+
٣٧	٩+
مجـ ح = ٢٤	

و=٢٨

مجـ ح =٢٤

ن=١٢

$$\text{سَ} = \text{و} + \frac{\text{مجـ ح}}{\text{ن}}$$

$$= ٢٨ + \frac{٢٤}{١٢}$$

=٢٨ + ٢ = ٣٠م " وهي نفسها النتيجة السابقة"

وإذا دققنا النظر في المثال السابق عن درجات الحرارة خلال ١٢ شهرًا نجد أن بعض القيم تتكرر أكثر من مرة؛ لذلك يمكننا حساب المتوسط الحسابي من واقع جدول التوزيع التكراري لهذه القيم الذي يكون على الشكل التالي:

استخراج الوسط الحسابي لقيم مبوبة

فئات درجات الحرارة	عدد الأشهر (التكرارات – ك)	مراكز الفئات (س)	س× ك
٢٤-٢٥	١	٢٤.٥	٢٤.٥
٢٦-٢٧	٢	٢٦.٥	٥٣.٠
٢٨-٢٩	٣	٢٨.٥	٨٥.٥
٣٠-٣١	٢	٣٠.٥	٦١.٠
٣٢-٣٣	١	٣٢.٥	٣٢.٥
٣٤-٣٥	١	٣٤.٥	٣٤.٥
٣٦-٣٧	٢	٣٦.٥	٣٧.٠
المجموع	مجـ ك =١٢		مجـ س ك = ٣٦٤

إن المتوسط الحسابي يمكن استخراجه في حالة البيانات المبوبة باستعمال القانون التالي:

$$\bar{س} = \frac{مجـ س ك}{مجـ ك} = \frac{مجـ س ك}{ن}$$ وذلك باتباع الخطوات التالية :

١- نضيف عمودًا لمراكز الفئات (س).

أ- نضرب تكرار كل فئة في مركز هذه الفئة، ونضع حاصل الضرب (س×ك) في العمود الأخير من الجدول.

٣- نوجد قيمة المتوسط الحسابي باستخدام القانون السابق .

$$\bar{س} = \frac{مجـ س ك}{مجـ ك} = \frac{٣٦٤}{١٢} = ٣٠,٢$$

والواقع أن هذه القيمة تختلف بعض الشيء عن القيم السابقة ؛ وذلك لأنه عند توزيع القيم الأصلية للظاهرة على الفئات في الجدول التكراري اختفت القيم الأصلية وضاعت معالمها، وكل ما بقي هو أن كل قيمة من القيم الأصلية أصبحت مفردة في فئة معينة، والقاعدة في هذه الحالة أن كل المفردات التي في فئة تكرارية واحدة متساوية، وقيمتها تساوي مركز الفئة التي تناظرها.

وهناك طريقة مختصرة لحساب المتوسط الحسابي باستخدام الوسط الفرضي على النحو التالي:

١- نحسب مراكز الفئات (س).

٢- نطرح من كل مركز من مراكز الفئات وسطًا فرضيًا مناسبًا (و) وكما سبق فإن أي رقم يصلح أن يكون وسطًا فرضيًا.

٣- بواقي الطرح هي عمود الانحرافات (ح = س - و).

٤- نضرب كل انحراف في التكرار المناظر له، فنحصل على العمود (ح ك).

٥- نحسب الوسط الحسابي باستخدام المعادلة :

$$\bar{س} = و + \frac{مجـ ح ك}{مجـ ك}$$

ويمثل الجدول الذي يبين التوزيع التكراري لدرجات الحرارة خلال ١٢ شهرًا طريقة حساب المتوسط الحسابي بواسطة الوسط الفرضي.

جدول استخراج الوسط الحسابي بطريقة الوسط الفرضي لقيم مبوبة

فئات درجات الحرارة	عدد الأشهر (التكرار- ك)	مراكز الفئات (س)	الانحراف عن وسط فرضي ح=(س-و) و=٢٨.٥	ح ك
٢٤-٢٥	١	٢٤.٥	-٤	-٤
٢٦-٢٧	٢	٢٦.٥	-٢	-٤
٢٨-٢٩	٣	٢٨.٥	صفر	صفر
٣٠-٣١	٢	٣٠.٥	+٢	٤
٣٢-٣٣	١	٣٢.٥	+٤	٤
٣٤-٣٥	١	٣٤.٥	+٦	٦
٣٦-٣٧	٢	٣٦.٥	+٨	١٦
المجموع	مجـ ك =١٢			مجـ ح ك=٢٢

وبتطبيق المعادلة السابقة وهي :

$$\bar{س}=و+\frac{مجـ ح ك}{مجـ ك}$$ نجد أن :

$$٢٨.٥+\frac{٢٢}{١٢} = ٢٨.٥+١.٨ = ٣٠.٣٣$$ وهو الجواب السابق نفسه.

مزايا وعيوب المتوسط الحسابي:

أ) المزايا:

١- لا يمكن إهمال أي قيمة من قيم الظاهرة المدروسة عند حسابه ، إذ تدخل في حسابه جميع القيم.

٢- أكثر المتوسطات استخدامًا وأيسرها فهمًا بالإضافة إلى سهولة إيجاده، إذ لا نحتاج عند حسابه إلا لمعرفة مجموع القيم وعددها.

ب) العيوب:

١- يتأثر بالقيم المتطرفة؛ ولذلك فهو يهضم حق القيم المعتدلة . مثال ذلك :

$$\frac{٦+١٥+٢٥+١٢٦}{٤} = ٤٣$$

وهو متوسط يزيد على ثلاث قيم، ولا يقل إلا عن قيمة واحدة في مثالنا هذا.

٢- لا يمكن إيجاده بالرسم .

ثانيًا : الوسيط Median

الوسيط هو القيمة الوسطى بحيث أن أعداد القيم قبلها يساوي عدد القيم بعدها ، بعد ترتيب القيم تصاعديا أو تنازليا ،فإذا كان لدينا الأرقام :

٤ ،٦، ٨، ٩، ، ١ ،٣، ٥، وطلب إلينا استخراج الوسيط نقوم بما يلي :

١- نرتب القيم تصاعديا أو تنازليا على الشكل التالي :
١ ، ٣ ، ٤ ، ٥ ، ٦ ، ٨ ، ٩

٢- نحسب ترتيب الوسيط من واقع القانون التالي :

$$\frac{ن+١}{٢}$$ حيث (ن) عدد القيم

فيكون ترتيب الوسيط = $\frac{٧+١}{٢}$ = ٤

٣- نحدد قيمة الوسيط وهو القيمة التي ترتيبها الرابع في المجموعة والقيمة التي ترتيبها (الرابع) هي رقم (٥).

أما إذا كان عدد القيم زوجيا لا فرديا كما لو كانت الأرقام كالآتي :
١ ، ٣ ، ٤ ، ٤ ، ٥ ، ٦ ، ٨ ، ٩

صادفنا مشكلة تعيين الرقم الأوسط في الترتيب ، وهو في مثالنا السابق ليس رقما واحدا بل رقميين هما : ٤ ، ٥ ، ولإيجاد الوسيط نستخرج متوسطهما الحسابي وهو ٤+٥ ÷٢=٤,٥ فيكون هو قيمة الوسيط ، ونستنتج من ذلك أنه إذا كان عدد القراءات زوجيا ففي هذه الحالة نجد قيمتين وسيطتين ترتيبهما هو ن÷٢ ، ن+١÷٢ ويلاحظ أن أي قيمة منهما أو تقع بينهما تصلح لأن تكون الوسيط ، ولكن العرف جرى على استخدام الوسط الحسابي لهاتين القيمتين ، أي نجمعهما ونقسمهما على ٢ .

أما في حالة البيانات المبوبة والموزعة في جداول تكرارية فيتم حساب الوسيط بالحساب والرسم ، وفيما يلي شرح لكل من الطريقتين :

- **طريقة الحساب**

تتبع الخطوات التالية :

١- نكون من الجدول التكراري البسيط جدولا تكراريا متجمعا صاعدا .

٢- نعين ترتيب الوسيط وهو مجموع التكرارات ÷٢ = مجـ ك ÷ ٢ ؛إذا لا نستخدم في حالة القيم المبوبة القاعدة نفسها التي استخدمناها للقيم غير المبوبة فيكتفى هنا بقسمة ن÷٢ دون إضافة (١) في الوسيط (ن تعادل ك في التوزيع التكراري).

٣- نحدد فئة الوسيط ، وهي الفئة التي تقع قيمة الوسيط بين حديها الأدنى والأعلى ؛ أي نوجد الفئة التي تقع بها القراءات ذات الترتيب ن÷٢ ، ويتم ذلك بأن نبحث في عمود التكرار المتجمع الصاعد (أو النازل) عن قيمتين متتاليتين يقع بينهما ترتيب الوسيط ، هاتان القيماتان تناظران رقمان في عمود حدود الفئات ، وهذان الرقمان هما الحد الأدنى والأعلى لفئة الوسيط.

٤- يمكن الوصول إلى الحل بإتباع الصيغة الآتية :

الوسيط = الحد الأدنى للفئة الوسيطية

$$+\left[\frac{\text{ترتيب الوسيط} - \text{التكرار الصاعد السابق للفئة الوسيطية}}{\text{التكرار الصاعد الوسيطي} - \text{التكرار الصاعد السابق للفئة الوسيطية}} \times \text{طول الفئة}\right]$$

ويلاحظ أنه كان بالإمكان استخدام الصيغة الآتية ، وذلك في حالة التكرار المتجمع الهابط وهي :

الوسيط = الحد الأعلى للفئة الوسيطية

$$+\left[\frac{\text{التكرار الصاعد الوسيطي} - \text{موقع الوسيط}}{\text{التكرار الصاعد الوسيطي} - \text{التكرار الصاعد السابق للفئة الوسيطية}} \times \text{طول الفئة}\right]$$

ولتطبيق الخطوات السابقة على مثال درجات الحرارة لجدة نقوم بما يلي:

١- نكون الجدول التكراري المتجمع الصاعد .

٢- ترتيب الوسيط = $\frac{\text{مجموع التكرارات}}{٢}$ = $\frac{١٢}{٢}$ = ٦ ومنه نجد أن الفئة الوسيطية هي الفئة (٢٨-)

٣- التكرار الصاعد السابق للفئة الوسيطية = ٣

٤- التكرار الصاعد الوسيطى = ٦

٥- طول الفئة = ٢

٦- قيمة الوسيط بناء على المعادلة السابقة =

$$٢٨ + \frac{٦-٣}{٦-٣} \times ٢$$

=٢٨ +٢

=٣٠م

التكرار المتجمع الصاعد لدرجات الحرارة في مدينة جدة خلال ١٢ شهر

فئات درجات الحرارة	عدد الأشهر (التكرارات – ك)	مراكز الفئات (س)	س×ك
٢٤-	١	أقل من ٢٦	١
٢٦-	٢	أقل من ٢٨	٣
٢٨-	٣	أقل من ٣٠	٦
٣٠-	٢	أقل من ٣٢	٨
٣٢-	١	أقل من ٣٤	٩
٣٤	١	أقل من ٣٦	١٠
٣٦-٣٧	٢	أقل من ٣٧	١٢
المجموع	١٢		

- **طريقة الرسم :**

يتم إيجاد قيمة الوسيط من المنحنى المتجمع الصاعد باتباع الخطوات

١- نكون جدول التكرار المتجمع الصاعد.

٢- نرسم المنحنى المتجمع الصاعد في شكل بياني.

٣- نعين ترتيب الوسيط ويساوي مجـ ك ÷ ٢. ثم نحدد هذه النقطة على المحور الرأسي ونرسم منها خطًا أفقيًا حتى يقابل المنحنى المتجمع الصاعد في نقطة ، نسقط منها عمودًا على المحور الأفقي يقابله في نقطة الوسيط .

٤- النتيجة نفسها يمكن الوصول إليها من التوزيع المتجمع النازل بالطريقة نفسها.

٥- وللمزيد من الدقة نرسم المنحنيين المتجمع الصاعد والنازل في شكل واحد. وهذان المنحنيان سيتقابلان في نقطة تقابل ترتيب الوسيط على المحور الرأسي، وقيمة الوسيط على المحور الأفقي

مزايا وعيوب الوسيط :

أ) المزايا:

١- تتوقف قيمته على موقعه أو موضعه .

٢- لا يتأثر بالقيم المتطرفة الشاذة وإنما يتأثر بعدد القيم .

أ- يمكن حسابه إذا كان التوزيع مفتوحًا من أحد طرفيه أو من كليهما .

٤- يمكن الحصول عليه بالرسم

ب) العيوب:

١- لا يدخل في حسابه سوى قيمة واحدة أو قيمتين من المجموعة كلها .

٢- ليس له شيوع المتوسط الحسابي نفسه .

ثالثًا : المنوال Mode:

المنوال هو القيمة الأكثر شيوعًا بين القيم ؛ أي القيمة التي تتكرر أكثر من غيرها ففي المثال الآتي :

٢ ، ٤، ٤، ٥، ٦ ، ٨ ، ٩

تكون قيمة المنوال ٤ باعتبارها قد تكررت أكثر من غيرها ولم تتكرر أي قيمة أخرى مثلها، أما إذا تكررت كل القيم بالعدد نفسه فإن هذا يعني أن لكل قيمة درجة الشيوع نفسها، فلا يصبح للتوزيع منوال. مثال ذلك :

٢، ٢، ٤، ٤، ٥، ٥.. إلخ.

وبالمثل لا يكون للتوزيع منوال إذا لم تتكرر أي قيمة في التوزيع أكثر من مرة، مثال ذلك :

٢ ، ٤ ، ٦، ٨، .... إلخ

أما إذا تكررت قيمتان أو أكثر بالعدد نفسه بين مجموعة كبيرة من القيم، بمعنى أن هذه القيمة المعينة لها درجة الشيوع نفسها بين كل القيم فإن هذا يعني تعدد المناويل مثال ذلك:

٥ و ٦ و ٦ و ٢ و ٨ و ٤ و ٢ و ٩

حيث للتوزيع هنا منوالان هما ٢، ٦ ويسمى هذا التوزيع Bimodal

وفي حالة البيانات المبوبة في جداول تكرارية فيمكن حساب المنوال بعدة طرق رياضية، وسوف نقتصر على واحدة فقط بالإضافة إلى إيجاده بالرسم.

إن المنوال يفترض أن يحتل مركز الفئة المنوالية لو كان التوزيع التكراري متماثلًا، أي إن المنحني الذي نرسمه له متماثلًا كذلك، فإذا ما اختل التماثل يميل المنوال نحو أقوى التكرارات اجتذابًا له من بين التكرارين المجاورين لتكرار الفئة المنوالية ذاتها .وهناك قانون خاص لحساب المنوال بهذه الطريقة مشتق من قوانين العزوم ومعادلة هذا القانون هي على النحو التالي :

$$\text{المنوال} = \text{بداية الفئة المنوالية} + \frac{\text{التكرار اللاحق}}{\text{التكرار اللاحق} - \text{التكرار السابق}} \times \text{طول الفئة}$$

وبتطبيق هذا القانون على مثال درجات الحرارة السابق نجد أن:

١- بداية الفئة المنوالية = ٢٨.

٢- التكرار اللاحق للفئة المنوالية = ٢.

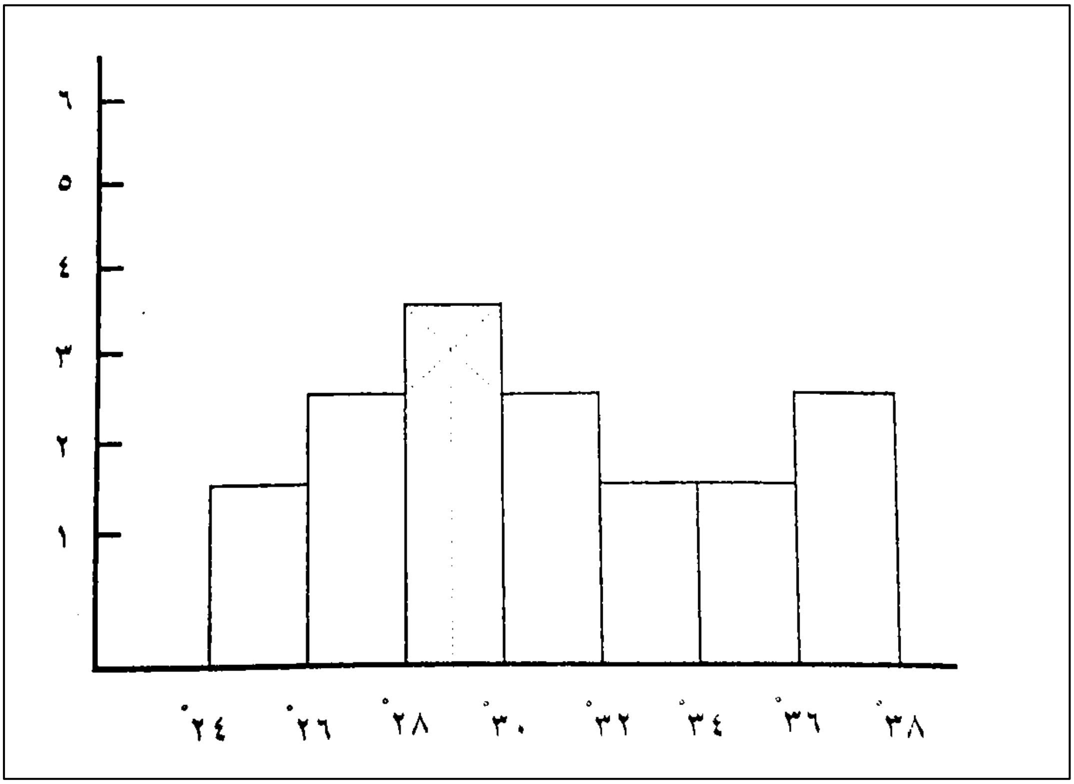
٣- التكرار السابق للفئة المنوالية = ٢.

٤ - طول الفئة = ٢.

٥-المنوال= $٢٨ + \frac{٢}{٤} \times ٢$

=٢٨+١ =٢٩م.

حساب المنوال بالرسم : يتم حساب قيمة المنوال بالرسم من المدرج التكراري، وإن كان يكتفي برسم المستطيلات التي تمثل الفئة المنوالية والفئة السابقة واللاحقة لها، ثم نصل الرأس الأيمن العلوي لمستطيل الفئة المنوالية بالرأس الأيمن العلوي للمستطيل السابق له ، وكذلك نصل الرأس الأيسر العلوي لمستطيل الفئة المنوالية بالرأس الأيسر العلوي للمستطيل اللاحق له، فيتقاطعان في نقطة ، نسقط منها عمودًا على المحور الأفقي يقابله في نقطة تكون هي قيمة المنوال، كما يتضح من الشكل التالي الذي يوضح كيفية إيجاد قيمة المنوال بالرسم من المدرج التكراري لدرجات الحرارة في مدينة جدة السابق . ويجب ألا نتوقع أن تتفق هذه النتيجة مع نتيجة الحساب السابقة؛ وذلك لأن جميع عمليات المنوال تعطي نتائج تقريبية ، وكلما كان الرسم دقيقًا كلما كانت النتائج أقرب إلى الصحة.



استخراج فئة المنوال بالرسم

مزايا وعيوب المنوال

أ) المزايا:

١- يمكن إيجاده بسهولة سواء بالرسم أو بالطريقة الرياضية .

٢- لا يتأثر بالقيم المتطرفة كالمتوسط الحسابي .

ب) العيوب:

١- يصعب تقديره إذا زاد عدد القيم زيادة كبيرة وتساوت التكرارات الكبيرة في فئات متجاورة.

٢- لا غبار عليه كمقياس من مقاييس الموضع إذا كان التوزيع الذي يمثله هذا المنوال متماثلًا، أما إذا لم يكن التوزيع متماثلًا فإن قيمة المنوال تبدو بعيدة عن مركز التوزيع، ويفقد المنوال بذلك جودته كأحد مقاييس الموقع .

أسئلة و تطبيقات

(١) تم جمع معلومات عن المكافآت الشهرية لـ ٦٢ عاملًا يعملون في أحد المصانع فوجد أن مقدار المكافآت موزعة بحسب الجدول التكراري (بالجنيه) على النحو التالي:

فئة المكافأة	عدد العمال
٢٠٠ – ٢٣٩,٩	٧
٢٤٠ – ٢٧٩,٩	١٠
٢٨٠ – ٣١٩,٩	١٦
٣٢٠ – ٣٥٩,٩	١٤
٣٦٠ – ٣٩٩,٩	٩
٤٠٠ – ٤٣٩,٩	٤
٤٤٠ – ٤٧٩,٩	٢
	٦٢

١- احسب المتوسط الحسابي :

أ) بالطريقة المباشرة.

ب) باستخدام الوسط الفرضي.

٢- احسب قيمة الوسيط .

٣- ارسم المدرج التكراري للمكافآت الشهرية، ومن خلاله استخرج قيمة المنوال وقيمة الوسيط .

(٢) البيانات التالية تمثل إنتاج مجموعة من المزارع (بالألف طن)
١٥ ، ١٠، ٧ ، ٧ ، ٩، ١٢ ، ٨ ، ٧ ، ١١

أجب عما يلي:

١- احسب إجمالي إنتاج المزارع المذكورة .

٢- أوجد قيمة الوسيط .

٣- أوجد قيمة المنوال .

(٣) فيما يلي أعمار عينة من السكان في إحدى المناطق السكنية :

الفئة	عدد السكان
١ ـ ٩	٤٥
١٠ ـ ١٩	٤٠
٢٠ ـ ٢٩	١٠٠
٣٠ ـ ٣٩	٦٠
٤٠ ـ ٤٩	٣٠
٥٠ ـ ٥٩	١٠
٦٠ ـ ٦٩	١٥
	٣٠٠

١- استخرج قيمة المنوال والوسيط بالرسم .

الفصل الرابع

مقاييس التركز والتخصص

الفصل الرابع
التركز والتخصص

أولًا : مقاييس التركز :

١- دليل التركز

٢- معامل التوطن

٣- منحنى لورنز

٤ - دليل التركز من منحنى لورنز

ثانيًا : مقاييس التنوع والتخصص

١- قياس التنوع الصناعي من منحنى لورنز

٢- مقياس جيبس مارتن للتنوع

٣- دليل عدم التماثل

مقاييس التركز والتخصص

أولًا : مقاييس المركز:

١- دليل المركز :

ويقيس مدى تركز توزيع أى ظاهرة في اطار مساحة جغرافية معينة ويمكن تطبيقه في حالات توزيع السكان أو الإنتاج الزراعي لمحصول معين أو العاملين بالصناعة في إطار وحدات إدارية.

وعلى سبيل المثال إذا كان لديك جدولا يبين توزيع السكان في الدول العربية الأسيوية ومساحاتها في عام ١٩٨٦ على النحو التالي:

الدولة	السكان (ألف نسمة)	المساحة (ألف كم^2)	الدولة	السكان (ألف نسمة)	المساحة (ألف كم^2)
العراق	١٦٤٥٠	٤٣٨	الكويت	١٧٩١	١٨
الأردن	٣٦٥٦	٩٨	البحرين	٤١٢	٠,٧
لبنان	٣٤٩٢	١٠	قطر	٣٣٥	١١
السعودية	١٢٠٠٦	٢١٥٠	الإمارات	١٣٨٤	٨٤
سوريا	١٠٦١٢	١٨٥	عمان	٩٩٠	٢١٢
اليمن	٩٤١١	٤٢٨	فلسطين	٤٢٩٦	٢١

فإنه لحساب دليل التركز تتبع الخطوات التالية:

١- تحسب النسبة المئوية لسكان كل دولة من الدول العربية المشار إليها لإجمالي سكان هذه الدول و هر ٦٤٨٣٥ الف نسمة.

٢- نحسب النسب المئوية لمساحة كل دولة لإجمالي مساحات هذه الدول وهي٣٦٥٥.٧ الف كم^2.

٣- نحصل على الفرق بين النسبة المئوية لمساحة الدولة والنسبة المئوية لسكانها بغض النظر عن الاشارة سالبة أو موجبة.

٤ - تجميع الفروق السابقة بغض النظر عن إشاراتها.

ويعني دليل التركز هذا أنه إذا كانت نسبة مساحة كل دولة تتفق تمامًا مع نسبة سكانها فإن التوزيع السكاني سيكون توزيعًا عادلًا أي أن ما يخص الدولة من السكان يمثل نصيبها في المساحة ونتيجة الفروق تساوي صفرًا، أما إذا كان الناتج بعيدًا عن الصفر فكلما كبر أشار إلى بعد التوزيع عن المثالية، وهذا يعني أن زيادة الفروق (التباينات) في التوزيع بين نسب الظاهرة الأولى المراد قياس تركزها والظاهرة الثانية المراد قياس التركز فيها (المساحة في هذه الحالة) تعطي قيمة رقمية أكبر. وبتطبيق الخطوات السابقة على الجدول ينتج الجدول التالي:

الدولة	نسبة السكان	نسبة المساحة	الفرق	الدولة	نسبة السكان	نسبة المساحة	الفرق
العراق	٢٥,٥	١١,٩	١٣,٤	الكويت	٢,٨	٠,٥	٢,٣
الأردن	٥,٦	٢,٧	٢,٩	البحرين	٠.٦	٠,٠١	٠,٦
لبنان	٥,٤	٠,٣	٥,١	قطر	٠,٥	٠,٣	٠,٢
السعودية	١٨,٥	٥٨,٨	٤٠,٣	الإمارات	٢,١	٢,٣	٠,٢
سوريا	١٦,٤	٥,١	١١,٣	عمان	١,٥	٥,٨	٤,٣
اليمن	١٤,٥	١١,٧	٢,٨	فلسطين	٦,٦	٠,٦	٦,٠

ويلاحظ أن مجموع نسب السكان والمساحة تنتهي إلى١٠٠ ٪، كما أن التناقضات بين نسب سكان الدول ومساحتها في التوزيع تظهر مدى التباين ؛ففي حالة السعودية مثلًا تمثل مساحتها ٨٥,٨% من إجمالي مساحة الجناح العربي الأسيوي على حين لا يجاوز سكانها ١٨,٥٪ ومن ثم يصل الفرق إلى ٤٠,٣٪ ، وإذا جمعت الفروق تصل إلى ٨٩,٤٪

وتطبق بعد ذلك المعادلة الآتية :

$$\text{دليل التركز} = \frac{١}{٢} \text{ مجـ } |\text{س} - \text{ص}|$$

حيث تشير س إلى نسب المساحة، ص إلى نسب السكان بينما يبين الخطان الرأسيان أن مجموع الفروق يكون بغض النظر عن الاشارة ، وهذا القانون معناه أن دليل التركز يساوي نصف مجموع الفروق الموجبة بين نسب توزيع الظاهرتين في الوحدات المكانية ، وعلى ذلك ففي حالة المثال السابق تكون قيمة دليل التركز مساوية لنصف القيمة ٨٩,٤ أي ٤٤,٧ الأمر الذي يشير لعدم العدالة في توزيع السكان قياسًا للمساحة في الدول العربية الاسيوية لبعد القيمة الناتجة عن الصفر. ويعتمد تطبيق هذا الأسلوب على متغيرين أحدهما الظاهرة المراد قياس تركزها في إطار المكان بجانب الوحدات المكانية ذاتها، ويمكن استخدامه أيضا في قياس تركز ظاهرة تمثل جزء من كل في إطار الوحدات المحددة مثل استهلاك الكهرباء بالنسبة للسكان أو الخدمات مثل عدد الأطباء لإجمالي السكان ... وهكذا.

٢- معامل التوطن :

ويسمى نسبة النسب أو نسبة التركز الموقعي ويستخدم كثيرًا في الدراسات الجغرافية، وتقوم فكرته على اعتبار متوسط نسب وجود ظاهرة ما في منطقة معينة أساسًا يقاس عليه مدى انحراف توزيع نسب الظاهرة ذاتها في الوحدات المكانية الأصغر التي تتكون منها المنطقة، ولإيضاح ذلك بالنسبة لتوطن محصول الأرز اعتمادًا على مساحاته المزروعة في المحافظات المصرية لعام ١٩٧٩ تتبع الخطوات التالية:

١- نحصل على المساحة المزروعة أرزًا في كل محافظة ولتكن في عام ١٩٧٩ مثلًا.

٢- نحصل على المساحة المزروعة أرزًا في انحاء الجمهورية في نفس السنة.

٣- تقسم المساحة المزروعة بمحصول الأرز في عام ١٩٧٩ في كل محافظة على اجمالي المساحة المزروعة بالمحاصيل المختلفة (المساحة المحصولية) في نفس المحافظة وتستخرج نسبتها المنوية .

٤ - تحسب النسبة المئوية لما يشغله محصول الأرز في الجمهورية من المساحة المحصولية.

٥ - تقسم النسبة الناتجة من رقم ٣ على النسبة المستخرجة من رقم ٤ ، وينتج عنها معامل التوطن.

وعلى ذلك يكون معامل التوطن: $\frac{\text{المساحة المزروعة بالأرز في المحافظة}}{\text{إجمالى المساحة المحصولية فى نفس المحافظة}}$

مقسومة على $\frac{\text{المساحة المزروعة بالأرز في الجمهورية}}{\text{إجمالى المساحة المحصولية فى الجمهورية}}$

وفيما يلي تطبيق لهذه الطريقة:

المحافظة	المساحة المزروعة بالأرز (فدان)	النسبة المئوية من المساحة المحصولية	المحافظة	المساحة (فدان)	النسبة % من المساحة المحصولية
كفر الشيخ	٢٣١١٨٣	٢٦,٧	الإسكندرية	٤٦٠١	٢,٩
الدقهلية	٢٦٧٣٨٩	٢٠,٠	الغربية	٩٠٣٠٤	١١,٧
دمياط	٥١٣٣٨	٢٥,٧	الشرقية	١٤٧٠٢٥	١١,٧
البحيرة	١٨٣٥٥٥	١٣,٢	القليوبية	٢٣١٤	٠,٢

ولما كان إجمالي المساحة المزروعة أرزًا في الجمهورية يبلغ ٩٧٧٧٥٠ فدانًا، والمساحة المحصولية تبلغ حوالي ١١ مليون فدان، فإن النسبة تكون:

$$\frac{٩٧٧٧٥٠ \times ١٠٠}{١١٠٠٠٠٠٠} = ٨,٩ \%$$

وهذه النسبة تمثل نسبة ما يشغله الأرز في مصر كلها لجملة مساحات المحاصيل، وللحصول على درجات التوطن تقسم نسب المحافظات في الجدول السابق على ٨,٩ % وعلى ذلك يكون معامل التوطن لهذه المحافظات:

كفر الشيخ= $\frac{٢٦,٧}{٨,٩} = ٣$ الدقلهية = $\frac{٢٠}{٨,٩} = ٢,٢$

دمياط = $\frac{٢٥,٧}{٨,٩} = ٢,٩$ البحيرة = $\frac{١٣,٢}{٨,٩} = ١,٥$

وهكذا تكون نتيجة المحافظات الأربع التالية هي : ٠,٣ للإسكندرية، ١,٣ للغربية، ١,٣ للشرقية، ٠,٠٢ للقليوبية، ويمكن بعد ذلك الخروج بنتيجة مؤداها أن المحافظات التي يزيد فيها العامل التوطن عن واحد صحيح ترتفع فيها نسبة المساحة المزروعة أرزًا عن مثيلها في الجمهورية كلها أي يتوطن فيها المحصول ، و كلما زاد الرقم دل ذلك على شدة التوطن وعلى العكس إذا قل الرقم عن واحد فإن نصيب المحافظة من المساحة المزروعة يكون أقل من تلك النسبة المزروعة في الجمهورية كلها. وتوقع هذه الأرقام على خرائط تبين توطن الأرز حسب الدرجات أو القيم التي حسبت بحيث توضع في فئات وبظلال متدرجة، ويمكن تطبيق معامل التوطن على أي ظاهرة خلاف الزراعة والمساحات المزروعة، وعند تفسير الخرائط الناتجة لابد من الالمام بالأوضاع السائدة في كل منطقة حيث تعطي أحيانًا ظلالًا ذات درجات عالية ولكنها لا تعني سوى تمركز الظاهرة قياسًا بما هو موجود في نفس الإقليم بسبب الاعتماد على النسبة في المساحة الأكبر كوحدة معايرة.

٣- منحنى لورنز :

وهو أحد أساليب قياس العلاقة بين توزيع ظاهرة ما في اطار مساحة جغرافية أي أنه يحاول التعرف على درجة بعد توزيع معين عن المثالية، وإذا أخذت محافظات الوجه القبلي كمثال لتطبيق منحنى لورنز على توزيع سكان وعلاقاتهم بالمساحة فيمكن رسم المنحنى باتباع الخطوات التالية :

١ - نحصل على توزيع السكان والمساحات لمحافظات الوجه القبلي وليكن في تعداد ١٩٧٦.

٢- تحسب النسب المئوية للمساحة والسكان في كل محافظة لجملة المحافظات في كل حالة.

٣- ترتب المحافظات ترتيبًا تصاعديًا حسب نسب مساحتها وتوضع نسبة السكان المقابلة لكل محافظة.

٤- تجمع نسب المساحة والسكان بعد الخطوة السابقة جمعًا تراكميًا أي في صورة تكرار متجمع مساعد في كل حالة.

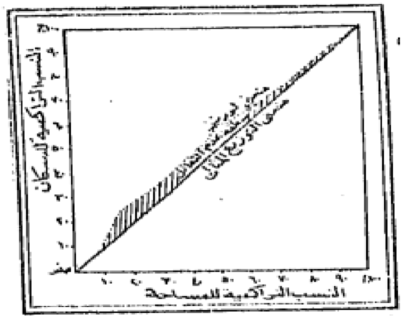
٥- يرسم محوران أحدهما أفقي تبين عليه النسب التراكمية للمساحة، والآخر رأسي تبين عليه النسب التراكمية للسكان، وتوقع النسب المجموعة تراكميًا عليهما، ويوصل بين النقاط لينتج منحنى لورنز وفيما يلى تطبيق هذه الطريقة.

المحافظة	السكان ١٩٧٦	المساحة ك م^{٢}	نسبة السكان%	نسبة المساحة
الجيزة	٢٤١٦٦٥٩	١٠٥٨	١٩,١	٨,٧
بني سويف	١١١٠١٣٢	١٣٢٢	٨,٨	١٠,٩
الفيوم	١١٤١٨٧٩	١٨٢٧	٩,٠	١٥,١
المنيا	٢٠٥٤١٠٥	٢٢٦٢	١٦,١	١٨,٧
أسيوط	١٦٩٧٤٢٢	١٥٥٣	١٣,٤	١٢,٩
سوهاج	١٩٢٤٨١٤	١٥٤٧	١٥,٢	١٢,٨
قنا	١٧٠٩٢٩٩	١٨٥١	١٣,٥	١٥,٣
أسوان	٦١٨٥١٨	٦٧٩	٤,٩	٥,٦
جملة وجه قبلي	١٢٦٧٢٨٢٨	١٢٠٩٩	١٠٠	١٠٠

ترتيب المحافظات تصاعديًّا حسب نسب مساحتها:

١ المحافظة	٢ المساحة %	٣ السكان %	٤ المتجمع الصاعد للمساحة	٥ المتجمع الصاعد للسكان
أسوان	٥,٦	٤,٩	٥,٦	٤,٩
الجيزة	٨,٧	١٩,١	١٤,٣	٢٤,٠
بني سويف	١٠,٩	٨,٨	٢٥,٢	٣٢,٨
سوهاج	١٢,٨	١٥,٢	٣٦,٠	٤٨,٠
أسيوط	١٢,٩	١٣,٤	٤٨,٩	٦١,٤
الفيوم	١٥,١	٩,٠	٦٤,٠	٧٠,٤
قنا	١٥,٣	١٣,٥	٧٩,٣	٨٣,٩
المنيا	١٨,٧	١٦,١	١٠٠,٠٠	١٠٠

يرسم المنحنى بعد ذلك من واقع العمودين ٤، ٥ كما يلي:



ويشير هذا التوزيع إلى الاقتراب بصورة كبيرة من المثالية حيث يتوزع ٦١,٤% من السكان في ٥٠,٩ ٪ من المساحة، ويبدو في الشكل اقتراب منحنى لورنز من خط التوزيع المثالي، ويمكن تطبيق هذا الأسلوب لقياس العلاقة بين السكان والمساحة في نفس الوحدات الادارية لأكثر من تعداد شريطة تثبيت الحدود الادارية. ويمكنك ملاحظة أن لدينا

متغيران أحدهما مستقل والآخر تابع، والمتغير التابع هر السكان لأنه يراد معرفة علاقته بالمساحة و بسبب أن احتمالات التغير في المساحة أقل حدوثًا خلال الزمن من السكان، كما أن الوحدات المكانية رتبت تصاعديًا حسب قيم المتغير الأول المراد قياس مدى التركز المكاني لقيم المتغير الثاني (السكان) فيه.

وعلى ذلك فإن منحنى لورنز إما يميل للاقتراب من المحور الرأسي ويتعدى خط التوزيع المثالي إلى أعلى مشيرًا للتركز السكاني الشديد في اطار مساحة محدودة، فعلى سبيل المثال إذا كان لدينا ٩٥% من السكان يتركزون في ٣٠ ٪ من المساحة وبقية الوحدات المكانية لا تضم سوى ٥٪ فقط فان بداية المنحنى ستكون مرتفعة القيمة عند المتغير التابع ومنخفضة على محور المتغير المستقل، ويحدث العكس إذا كان ٥ ٪ من السكان ينتشرون في ٩٠ ٪من المساحة مثلًا حيث يقترب المنحنى من عند نهاية الركن الأيمن للمحور الافقي مشيرًا إلى شدة الانتشار، وبين هذين الحدين الأدنى والأقصى يتباين التوزيع في اقترابه أو بعده من الصورة المثالية والتي تتحقق إذا كان التوزيع النسبي للظاهرتين في الوحدات المكانية متماثلًا وذلك على النحو الذي يوضحه الجدول التالي :

المنطقة	نسبة الظاهرة الأولى %	نسبة الظاهرة الثانية %	المنطقة	ترتيب الظاهرة الأولى	ما يقابلها من الظاهرة الثانية	المتجمع الصاعد في الحالتين
أ	٥	٥	أ	٥	٥	٥
ب	١٥	١٥	ب	١٥	١٥	٢٠
جـ	٣٠	٣٠	هـ	١٦	١٦	٣٦
د	٣٤	٣٤	جـ	٣٠	٣٠	٦٦
هـ	١٦	١٦	د	٣٤	٣٤	١٠٠

٤ - دليل التركز من منحنى لورنز:

عند الحصول على دليل التركز من منحنى لورنز يشترط الاعتماد على الوحدات المكانية كأساس وقياس تركز نوع من الظاهرات أو الأنشطة بالنسبة لباقي الأنشطة، وتشبه الطريقة المتبعة في رسم المنحنى تلك المستخدمة في المثال السابق عدا بعض الاختلافات تتمثل في ترتيب الوحدات المكانية حسب تركز الظاهرة موضع البحث ترتيبًا تصاعديًا قبل جمعها تراكميًا.

وإذا أخذت مجموعة دول مجلس التعاون الخليجي الست كنموذج لتطبيق هذه الطريقة في قياس مدى تركز العاملين في الزراعة وصيد الأسماك بهذه الدول بالنسبة لإجمالي العاملين بكل أوجه النشاط الاقتصادي لعام ١٩٨٦ وكانت النسب على النحو التالي :

الدولة	العاملين بالزراعة والصيد		إجمالي العاملين		الدولة	العاملين بالزراعة والصيد		إجمالي العاملين	
	العدد بالألف	%	العدد بالألف	%		العدد بالألف	%	العدد بالألف	%
السعودية	٤٣٢	٧١,٥	٣٠٢٢	٥٥,٣	عمان	١٠٩	١٨,٠	٤٦٧	٨,٥
الكويت	١٤,٢	٢,٣	٧١٢	١٣,٠	البحرين	٤	٠,٧	١٨٣	٣,٣
الإمارات	٤٥,٠	٧,٤	٨٩١	١٦,٣	قطر	٠,٤	٠,٩	١٩٥	٣,٦

في مثل هذه الحالة تحسب نسبة التركز للعاملين بالزراعة والصيد لإجمالي العاملين في كل الأنشطة لكل دولة بقسمة كل نسبة على قرينتها

ففي السعودية ٧١,٥ ÷ ٥٥,٣ = ١,٣ والكويت = ٢,٣ ÷ ١٣ = ٠,٢

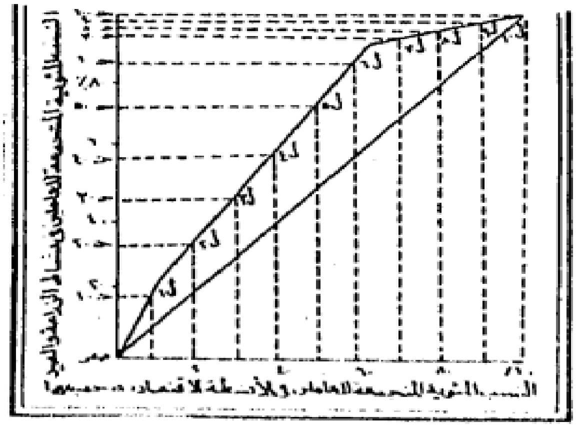
وهكذا تكون النتائج في الإمارات ٠,٥، وعمان ٢,١

والبحرين ٠,٢ ، وقطر ٠,٠٢.

على ذلك يتكون جدول ترتب فيه الدول الست حسب نسب التركز فنبدأ بعمان ثم السعودية والإمارات والكويت، والبحرين و قطر ويوضع ما يقابلها من نسب في الحالتين وتجمع جمعًا تراكميًا متصاعدًا على النحو التالي :

الدولة	نسبة التركز	العاملين بالزراعة	متجمع صاعد	العاملين بالأنشطة كلها	متجمع صاعد
عمان	٢,١	١٨,٠	١٨,٠	٨,٥	٨,٥
السعودية	١,٣	٧١,٥	٨٩,٥	٥٥,٣	٦٣,٨
الإمارات	٠,٥	٧,٤	٩٦,٩	١٦,٣	٨٠,١
الكويت	٠,٢	٢،٣	٩٩,٢	١٣,٠	٩٣,١
البحرين	٠,٢	٠,٧	٩٩,٩	٣,٣	٩٦,٤
قطر	٠,٠١	٠,١	١٠٠	٣,٦	١٠٠

ويمكن من خلال النسب التراكمية السابقة رسم منحنى لورنز على النحو المبين في الشكل الآتي:



منحنى لورنز للعاملين في قطاع الزراعة وصيد الأسماك

في دول الخليج العربية سنة ١٩٨٦م

وبعد تطبيق منحنى لورنز للتوزيعات المكانية في المثال السابق يمكن استخدام الشكل النهائي للمنحنى في حساب دليل التركز وذلك باتباع الخطوات التالية :

١- تحدد عشر نقاط على مسافات متساوية بطول المحور الأفقي.

٢- تقام أعمدة رأسية من النقاط العشر حتى تلتقي منحنى لورنز عند النقاط ل١،ل ٢، ل ٣ ..... حتى ل١٠.

٣- تسقط اعمدة افقية من نقاط التلاقي السابقة إلى المحور الراسي لتلتقي به عند جـ ا ،جـ ٢،جـ ٣ .... حتى جـ ١٠.

٤- تجمع قيم جـ ا ... حتى جـ ١٠ من المحور الرأسي لنحصل على مجموع جـ ، في المثال السابق الذي تم تطبيقه، كانت قيم جـ كالاتي :

جـ١= ١٩، جـ٢ = ٢٧، جـ٣ = ٤٥، جـ٤ = ٥٩، جـ٥ = ٧٠، جـ٦ = ٨٥، جـ٧ = ٩٢، جـ٨ = ٩٥، جـ٩ = ٩٨، جـ١٠ = ١٠٠.

مجـ جـ = ١٩ + ٢٧ + ٤٥ + ٥٩ + ٧٠ + ٨٥ + ٩٢ + ٩٥ + ٩٨ + ١٠٠ = ٦٩٠

٥- تطبق المعادلة:

$$\text{دليل التركز} = \frac{\text{مجـ جـ} - ٥٥٠}{١٠٠٠ - ٥٥٠} = \frac{\text{مجـ جـ} - ٥٥٠}{٤٥٠}$$

وتشير القيمة ٥٥٠ لمجموع قيم جـ تراكميًا عندما تقترب درجة التركز من المثالية وفيها تكون قيم جـ عند التقاء الأعمدة بالمحور الراسي كالآتي :

جـ١ = ١٠، جـ٢ = ٢٠، جـ٣ = ٣٠، جـ٤ = ٤٠، جـ٥ = ٥٠، جـ٦ = ٦٠، جـ٧ = ٧٠، جـ٨ = ٨٠، جـ٩ = ٩٠، جـ١٠ = ١٠٠ = ٥٥٠

وعندها يقترب منحنى لورنز أن لم ينطبق تمامًا على منحنى التوزيع المثالي :

وتشير القيمة ١٠٠٠ لأقصى تركز للظاهرة وتصبح قيم جـ العشرة متساوية و كل واحدة يخصها ١٠٠ ولذلك يكون مجموع جـ ١٠× ١٠٠=١٠٠٠ وعندها ينحرف منحنى لورنز عن نمط التوزيع المثالي.

وبعد التعويض في المعادلة يكون دليل التركز لقطاع الزراعة وصيد الأسماك في دول الخليج العربية

$$= \frac{٦٩٠ - ٥٥٠}{١٠٠٠ - ٥٥٠} = \frac{١٤٠}{٤٥٠} = ٠,٣١$$

وحيث أن الناتج من المعادلة إذا كان واحد صحيح يدل على أن الظاهرة موضوع الدراسة بلغت أقصى تركيز لها ويتناقص مدى التركز ببعد القيمة عن الواحد الصحيح ،و طالما أن دليل التركيز لقطاع الزراعة وصيد الأسماك في دول الخليج العربية يساوي (٠,٣) فإن القيمة تبتعد عن الواحد الصحيح مما يدل على قلة تركيز الظاهرة بل أن الظاهرة تكاد تكون قد بلغت ادني تركز للبعد الكبير عن الواحد الصحيح ، ويعود ذلك إلى ظهور النفط في المنطقة وتغير أنماط الحياة الاجتماعية والاقتصادية وظهور وظائف جديدة ومتنوعة جذبت السكان اليها، مما أدى إلى تراجع العاملين بقطاع الزراعة وصيد الأسماك، وإذا أخذ في الاعتبار أيضًا أن الأراضي الزراعية في تلك الدول محدودة ... لهذه الأسباب كانت الانشطة الأخرى مثل قطاع الصناعة والخدمات تنافس هذا القطاع وإن كان الوضع يحتاج إلى اعادة النظر في الاهتمام بهذا النوع من القطاعات لأهميته في توفير الغذاء محليًا بدلًا من الاعتماد على المستورد.

ثانيًا : مقاييس التنوع والتخصص :

وتقوم فكرتها على محاولة التعرف على تخصص أقاليم بالذات في الاستئثار بوجود ظاهرات معينة او قياس درجة توزع هذه الظاهرات بالتساوي بين عدد من الأقاليم، وعادة يطبق فيها أكثر من طريقة منها استخدام منحنى لورنز للتعرف على التنوع في توزيع الصناعات بين عدد من الأقاليم الجغرافية أو مقياس جيبس مارتن للتنوع أو دليل عدم التماثل لقياس مدى الاختلاف في توزيع مجموعتين من النسب في تاريخين مختلفين أو إقليميين جغرافيين أو لمقارنة نسب نظرية بأخرى واقعية.

١-قياس التنوع الصناعي من منحنى لورنز:

إذا كانت لديك أعداد العاملين في خمسة أنواع من الصناعات تتوزع جغرافيًا في ثلاث مناطق جغرافية في القاهرة الكبرى والاسكندرية والغربية وتريد قياس التنوع الصناعي بها تتبع الخطوات التالية : -

١- تحسب النسبة المئوية للعاملين بكل صناعة في المناطق المختلفة لإجمالي العاملين بالصناعات كلها.

٢- ترتب النسب المئوية ترتيبًا تنازليًا ويكون جدول تكراري متجمع صاعد لكل منطقة حسب ترتيب صناعاتها.

٣- يرسم منحنى تكراري متجمع صاعد لكل منطقة يوضع على محوره الأفقي أنواع الصناعات، وعلى محوره الرأسي النسب المئوية التراكمية المتجمعة لكل منطقة، وإذا كان الهدف مقارنة المناطق المختلفة فيجب توحيد ترتيب الصناعات حسب أنواعها.

٤ - تقارن المنحنيات المرسومة بمنحنى التوزيع المثالي الذي يفترض توزيع النسب المئوية بالتساوي بين أنواع الصناعات ، وعلى ذلك يصبح الجدول على النحو التالي :

القاهرة الكبرى			الإسكندرية			الغربية		
نوع الصناعات	نسبة العاملين%	متجمع صاعد	نوع الصناعات	نسبة العاملين%	متجمع صاعد	نوع الصناعات	نسبة العاملين%	متجمع صاعد
غزل ونسيج	٤٠,٠	٤٠,٠	غزل ونسيج	٧١,٤	٧١,٤	غزل ونسيج	٨٦,٥	٨٦,٥
غذائية	٣٥,٠	٧٥,٠	غذائية	١٤,٣	٨٥,٧	غذائية	٥,٤	٩١,٩
معدنية	١١,٣	٨٦,٣	معدنية	٧,٢	٩٢,٩	معدنية	٣,٨	٩٥,٧
هندسية	٧,٤	٩٣,٧	هندسية	٣,٨	٩٦,٧	هندسية	٢,٧	٩٨,٤
كيماوية	٦,٣	١٠٠,٠	كيماوية	٣,٣	١٠٠,٠	كيماوية	١,٦	٥٠,٠
المجموع	١٠٠	٣٩٥,٠		١٠٠	٤٤٦,٧		١٠٠	٤٧٢,٥

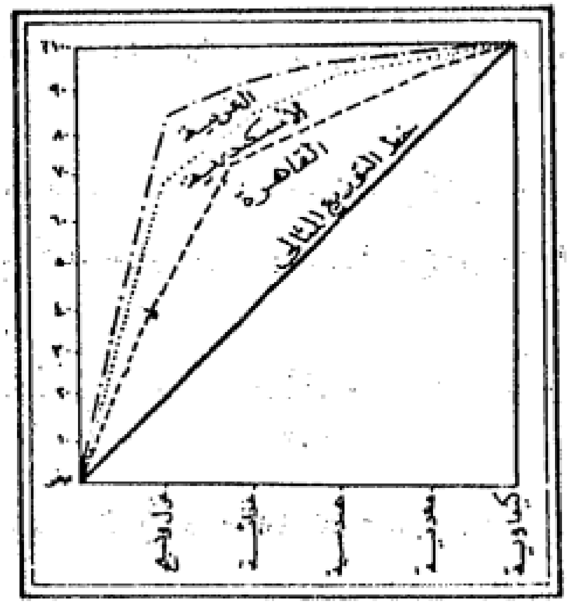
ومن هذا الجدول يلاحظ اختلاف ترتيب الصناعات في محافظة الغربية عن الإسكندرية والقاهرة لاختلاف الأهمية النسبية لأعداد العاملين بها، كما أن مجموع النسب التراكمية يتزايد مع ارتفاع درجة تركز العاملين في صناعة واحدة، ولما كانت الصناعات المبحوثة خمس فإن مجموع النسب التراكمية يمكن مقارنته بحدين الأقصى منهما يمثل حالة التركز الكامل أو التخصص في صناعة واحدة فقط هي الغزل والنسيج ومجموع النسب التراكمية فيها = ١٠٠ ٪ يتكرر خمس مرات أي ٥٠٠ أما الحد الأدنى فيوضح التوزيع المثالي ويخص كل صناعة فيه ٢٠ ٪من عدد العاملين إذا جمعت تراكميًا ستصبح ٢٠+٤٠+٦٠+٨٠ +١٠٠ إي ٣٠٠، ويمكن مقارنة النسب التراكمية الناتجة، بأي من هذين الحدين على النحو التالي :

القاهرة الكبرى = ٥٠٠ - ٣٩٥ = ١٠٥ : تنوع إلى حد ما

الإسكندرية = ٥٠٠ - ٧, ٤٤٦ = ٥٣.٣ : تخصص

الغربية = ٥٠٠ - ٤٧٢,٥ = ٢٧,٥ : تخصص شديد

والواضح أنه كلما كبرت القيمة الناتجة اشارت لبعد الصناعات عن التركز في نمط واحد أو إلى التنوع وبالعكس إذا صغرت أشارت إلى التخصص، ورغم أهمية هذه الطريقة إلا أن عيوبها تتمثل في صعوبة إجراء مقارنات بين عدد كبير من الأقاليم لاختلاف ترتيب صناعاتها استنادًا لأي معيار يتخذ كمقياس، بجانب ذلك فإن قيم أقصى تركز أو التوزيع المثالي تزيد وتنقص حسب عدد الأنشطة أو الصناعات المبحوثة .



منحنى لورنز لقياس تنوع الصناعات في ثلاث مناطق جغرافية

٢- مقياس جيبس - مارتن للتنوع:

استخدمه كلًا من جيبس ومارتن لأول مرة عام ١٩٦٢ في دراسة مدى التنوع في توزيع العاملين بالأنشطة الاقتصادية، فإذا كانت قوة العمل في منطقة ما تتمثل في نشاط واحد كانت نتيجة تطبيق المقياس تساوي صفرًا، وإذا كانت موزعة بالتساوي على كل الأنشطة فإن المقياس يساوي واحدًا صحيحًا أما المعادلة المستخدمة فهي كما يلي:

$$\text{مقياس التنوع} = \text{١} - \frac{\text{مجـ س}^{\text{٢}}}{(\text{مجـ س})^{\text{٢}}}$$

وتشير س إلى عدد العاملين في كل نشاط اقتصادي.

وإذا افترضنا منطقة معينة تضم تنوعًا كاملًا في انشطتها ورمز لها بالرمز أ ، ومنطقة أخرى تضم تركزًا كاملًا ورمز لها بالرمز ب فيمكن أن يكون لدينا الجدول التالي :

رقم النشاط الاقتصادي

رقم النشاط	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	مجموع
المنطقة أ	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٢٠
المنطقة ب	١٢٠	–	–	–	–	–	–	–	–	–	–	–	١٢٠

وفي هذه الحالة يكون مجموع حالة التنوع الكامل = ١٢٠ أي أن مجموع س = ١٢٠ ومجموع مربعات س = ١٢٠٠ لـ حيث تربع كل قيمة للصناعات المختلفة وتجمع.
وعلى ذلك يكون مقياس جيبس ومارتن للتنوع في حالة التنوع الكامل :

$$= ١ - \frac{١٢٠٠}{(١٢٠٠)^{٢}} = ١ - \frac{١٢٠٠}{١٤٤٠٠} = ١ - ٠,٠٨٣ = ٠,٩١٧$$

أما في حالة المركز الكامل فإن مجموع س^{٢} =١٤٤٠٠،ولما كان مجموع س ١٢٠

$$\text{فإن المقياس يكون} = ١ - \frac{١٤٤٠٠}{(١٢٠)^{٢}} = ١ - ١ = \text{صفر}$$

ويتميز مقياس جيبس – مارتن بمزايا عدة منها استخدامه في المقارنة بين توزيع الأنشطة الاقتصادية في الأقاليم المختلفة كما أنه لا يحتاج إلى تحويل الأرقام إلى نسب مئوية ، ولكن يعيبه أن قيمته تتأثر بعدد الأنشطة موضع الدراسة فقد وجد من دراسات مختلفة أن قيمته تبلغ عند التنوع الكامل إذا كانت الأنشطة عددها أربعة = ٠,٧٥ بينما إذا كانت الأنشطة عشرة فانه يساوي ٠,٩٠ ولذلك لا يمكن استخدامه في المقارنة إلا إذا كان عدد الأنشطة الاقتصادية متساويًا في كل الحالات. وفيما يلي توزيع السكان حسب الأنشطة الاقتصادية في بعض محافظات مصر عام ١٩٧٦ والمطلوب حساب مقياس جيبس - مارتن لقياس مدى التنوع في النشاط الاقتصادي بين هذه المحافظات والتعليق على النتائج.

نوع النشاط	القاهرة (ألف نسمة)	الدقهلية (ألف نسمة)	الإسكندرية (ألف نسمة)	سوهاج (ألف نسمة)
الزراعة	١٧	٤٦١	٣٦	٣٧٣
التعدين	٦	١	٣	١
الصناعة	٣٨٤	٦٥	٢٠٨	٢٣
الكهرباء	١٦	٤	٨	١
التشييد	١١٩	٢١	٤١	١٣
التجارة	٢١١	٥٦	٩٠	٣٨
النقل	١٢٦	٢٧	٦٠	١٢
التمويل	٣٠	٥	٩	٣
الخدمات	٤٨٩	١١٩	١٦٣	٥٥
المجموع	١٣٩٨	٧٥٩	٦١٨	٥١٩

و لحساب مقياس جيبس ومارتن للتنوع في هذه المحافظات الأربع نحصل على مجموع مربع أعداد العاملين في كل الأنشطة الاقتصادية على النحو التالي :

المحافظة	مجموع مربع عدد العاملين	المحافظة	مجموع مربع عدد العاملين
القاهرة	٤٦٢٦١٦	الإسكندرية	٨٤٦٣٧
الدقهلية	٢٣٥٢٥٥	سوهاج	١٤٤٤٥١

ويجب التأكيد هنا على أنك تقوم بتربيع عدد العاملين في كل نشاط اقتصادي أولًا ثم جمع هذه القيم بعد ذلك وتكون النتيجة:

$$\text{مقياس التنوع في القاهرة} = ١ - \frac{٤٦٢٦١٦}{(١٣٩٨)^٢} = ١ - ٠,٢٤ = ٠,٧٦$$

$$\text{وفي الدقهلية} = ١ - \frac{٢٣٥٢٥٥}{(٧٥٩)^٢} = ١ - ٠,٤١ = ٠,٥٩$$

$$\text{وفي الإسكندرية} = ١ - \frac{٨٤٦٣٧}{(٦١٨)^٢} = ١ - ٠,٢٢ = ٠,٧٨$$

$$\text{وفي سوهاج} = ١ - \frac{١٤٤٤٥١}{(٥١٩)^٢} = ١ - ٠,٥٤ = ٠,٤٦$$

وعلى ذلك ترتب المحافظات حسب درجة التنوع في انشطتها الاقتصادية كالتالي : الاسكندرية - القاهرة - الدقهلية - سوهاج، ويفسر ذلك بأن المحافظتين الأخيرتين تستأثر الزراعة فيهما بنسب عالية من عدد العاملين بين الانشطة الاقتصادية.

٣- دليل عدم التماثل:

وهو يشبه دليل التركز وإن كان اسهل في حسابه من ناحية ولا يتطلب رسومًا بيانية للحصول عليه ، ويمكن من خلاله معرفة درجة الاختلاف بين توزيع مجموعتين من النسب سواء كانت واقعية أو نظرية.

فإذا كان لدينا الجدول التالي الذي يمثل توزيع نسب سكان الولايات في استراليا خلال فترات زمنية مختلفة والمطلوب حساب درجة عدم التماثل في التغيرات التي حدثت في التوزيع لأعوام ١٨٨١- ١٩٢١ من ناحية وعدم التماثل الذي حدث بين ١٩٢١ - ١٩٦١ من ناحية ثانية أي خلال فترتين كل منهما ٤٠ عامًا.

توزيع نسب السكان في الولايات الأسترالية ١٨٨١، ١٩٢١، ١٩٦١

الولاية	الشمالية	نيوسوث ويلز	فيكتوريا	الجنوبية	الغربية	كوينزلاند	تسمانيا
١٨٨١	٠,١٥	٣٣,٣٢	٣٨,٢٩	٢٨,٢٨	١,٣٣	٩,٤٩	٥,١٤
١٩٢١	٠,٠٧	٣٨,٦٩	٢٨,١٧	٩,١١	٦,١٢	١٣,٩١	٣,٩٣
١٩٦١	٠,٢٦	٣٧,٨٤	٢٧,٨٨	٩,٢٢	٧,٠١	١٤,٤٥	٣,٣٣
نسبة المساحة	١٧,٦٢	١٠,٤٢	٢,٩٦	١٢,٧٩	٣٢,٨٥	٢٢,٤٥	٠,٨٨

والمعادلة المستخدمة لحساب دليل عدم التماثل هي : مجـ (س - ص) اذا كانت س أكبر من ص او مجـ (ص- س) إذا كانت ص اكبر من س

وتشير س ، ص إلى مجموعات النسب التي يراد مقارنتها، وإذا كان مجموع النسب المئوية في المجموعتين = ١٠٠ فإن كلا من شطري المعادلة سيعطي نفس النتيجة، وعادة ما تتراوح النتائج في هذا المقياس بين صفر في حالة التماثل الكامل و ١٠٠ عند أقصى حد لعدم التماثل.

وإذا طبقت المعادلة السابقة على توزيع سكان استراليا حسب الولايات في عامي ١٨٨١، ١٩٢١ والتي يمكن أن يشار إليها بـ س، ص فإن الولايات التي تظهر فيها قيمة س (السكان عام ١٩٨٨) أكبر من قيمة ص (السكان عام ١٩٢١) هي الولاية الشمالية - فيكتوريا - جنوب أستراليا - تسمانيا على حين تقل النسب في باقي الولايات ، ولذلك فإذا ما حصلنا على الفروق تكون كالتالي:

بين ١٨٨١ – ١٩٢١	بين ١٩٢١ – ١٩٦١
الولاية الشمالية ٠,١٥ – ٠,٠٧ = ٠,٠٨	الشمالية ٠,٢٦ – ٠,٠٧ = ٠,١٩
فيكتوريا ٣٨,٢٩ – ٢٨,١٧ = ١٠,١٢	الجنوبية ٩,٢٢ – ٩,١١ = ٠,١١
جنوب أستراليا ٢٨,٢٨ – ٩,١١ = ١٩,١٧	الغربية ٧,٠١ – ٦,١٢ = ٠,٨٩
تسمانيا ٥,١٤ – ٣,٩٣ = ١,٢١	كوينزلاند ١٤,٤٥ – ١٣,٩١ = ٠,٣٤
المجموع ٣٠,٥٨	١,٥٣

ويبدو من ذلك أن مجموع الفروق في التوزيع خلال التاريخين الأول والثاني تساوي ٥٨ ، ٣٠

وبتطبيق نفس المعادلة على الأرقام الخاصة بعامي ١٩٢١، ١٩٦١ فان النتيجة ستكون ١,٥٣ وذلك يعني أن التغيرات التي حدثت في توزيع سكان ولايات استراليا خلال الفترة الأولى أكثر منها خلال الفترة الثانية.

ويمكن بعد ذلك استخدام مساحات الولايات المبينة في الجدول في حساب درجة التركز الجغرافي للسكان فإذا ما توزع السكان بصورة عادلة تمامًا فإن نصيب الولاية منهم لابد وأن يساوي ما تشغله من مساحة ولكن ليس هذا هو واقع الحال ويمكننا بمقارنة نسب كل من السكان والمساحة الحصول على دليل التركز فالولايات التي يجاوز حجمها السكاني عام ١٩٦١ ما تشغله من مساحة هي ثلاث ولايات فقط توضع نسب سكانها مقابل مساحتها كالتالي:

نيو سوث ويلز	٣٧,٨٤	–	١٠,٤٢	= ٢٧,٤٢
فيكتوريا	٢٧,٨٨	–	٢,٩٦	= ٢٤,٩٢
تسمانيا	٣,٣٣	–	٠,٨٨	= ٢,٤٥
المجموع	٦٩,٠٥		١٤,٢٦	٥٤,٧٩

وقد تستخدم أرقام السكان والمساحة أحيانا لقياس درجة التركز السكاني فيقال أن ٦٩٪ من سكان استراليا يستوطنون ولايات مساحتها ١٤,٢٦٪ من مساحة الدولة.

الفصل الخامس

مقاييس التباين والانتشار

الفصل الخامس

التباين والانتشار

أولًا : مقاييس التباين :

١- المدى

٢- الانحراف عن المتوسط

٣- التباين

٤- الانحراف المعياري

٥- معامل الاختلاف

ثانيًا : مقاييس الانتشار :

١- الربيع الجغرافي

٢- معامل الانتشار

٣- الانتشار حول موقع معين

٤- المسافة المعيارية

٥- مقياس أقرب جار أو صلة الجوار

التباين والانتشار

أولًا : مقاييس التباين:

تهتم مقاييس التشتت بالتعرف على مقدار انتشار البيانات أو القيم ؛ فالمتوسط وحده لا يكفي لتقديم فكرة دقيقة عن مجموعة بيانات من حيث طبيعة توزيعها ؛ فعلى سبيل المثال قد توجد مجموعتان من القيم لهما نفس المتوسط ولكن يختلف تشتتهما، وفي بعض الأحيان يكون حساب المتوسط لا معنى له ، فإذا كان لدينا في مصر مثلًا ٣٠ مدينة تتراوح أحجامها السكانية بين ٥٠ ألف نسمة و ٦ ملايين نسمة فإن حساب المتوسط أو الوسيط في هذه الحالة لن تكون له دلالة كبيرة ، ولذا تستخدم نوعية أخرى من المقاييس للتعرف على درجة انتشار البيانات أو تشتتها هي:-

(١)المدى: Range

وهو أبسط المقاييس لمعرفة درجة انتشار البيانات ويقصد به الفرق بين أكبر القيم وأقلها في توزيع مكاني أو غير مكاني فإذا كان لدينا مجموعة من القيم على النحو التالي :

٥ ، ٨ ، ٧ ، ٩ ، ١١

فإن المدى يصل إلى ١١- ٥ = ٦ ويلاحظ أن المدي كمقياس للتشتت له عيوب هي أنه لا يستخدم من القيم سوى قيمتين فقط، كما تتأثر قيمته بالحد الأقصى والأعلى لتوزيع القيم ؛أي أنه إذا كان لدينا عدد من القيم يبلغ ١٠٠، والمدى فيها يتراوح بين ٥، ٨٠ أي يساوي ٧٥ فإنه من الممكن أن تكون ٩٩ قيمة منها تقع بين ٢٠، ٨٠ وقيمة واحدة هي التي تبلغ ٥.

(٢) الانحراف عن المتوسط : Mean deviation

إذا كانت هناك مجموعة من القيم تمثل توزيع عدد الأطفال في عشر أسر على النحو التالي:

٥ ، ٣ ، ٢ ، ٥ ، ٦ ، ٨ ، ٥ ، ١ ، ٤ ، ٣

فإن متوسط عدد أفراد الأسرة يكون ٤,٢ فرد، والمدى يتراوح بين ١، ٨ أو يساوي ٧ فكيف يمكننا التعرف على مدى التشتت في هذه القيم بصورة افضل؟ . . يأتي ذلك بحساب الانحراف عن المتوسط كما يلي:

(القيم)	الانحرافات عن الوسط الحسابي	
س	(س – سَ)	حَ (الانحراف)
٥	(٥ – ٤,٢)	٠,٨
٣	(٣ – ٤,٢)	– ١,٢
٢	(٢ – ٤,٢)	– ٢,٢
٥	(٥ – ٤,٢)	٠,٨
٦	(٦ – ٤,٢)	١,٨
٨	(٨ – ٤,٢)	٣,٨
٥	(٥ – ٤,٢)	٠,٨
١	(١ – ٤,٢)	– ٣,٢
٤	(٤ – ٤,٢)	– ٠,٢
٣	(٣ – ٤,٢)	– ١,٢
المجموع ٤٢	مجموع الانحرافات الموجبة + ٨ مجموع الانحرافات السابة – ٨	

يلاحظ أن مجموع انحرافات القيم عن متوسطها الحسابي لابد أن يساوي صفرًا، والانحراف عن المتوسط ما هو إلا مجموع انحرافات القيم عن وسطها الحسابي بغض النظر عن الاشارة موجبة أو سالبة مقسومًا على عددها، وفي حالة المثال السابق يكون:

$\frac{١٦}{١٠}$ = ١,٦ ويمكن أن يحسب بتطبيق إحدى المعادلتين:

$\frac{\text{مجـ ح}}{\text{ن}}$ أو $\frac{\text{مجـ } |\text{س} - \text{سَ}|}{\text{ن}}$

وكلتا الصيغتين تحقق نفس الغرض طالما أهملت الإشارات.

ويتميز الانحراف عن المتوسط بكونه مقياسًا بسيطًا في حسابه وفهمه بجانب بلورته لمدى تشتت مجموعة من القيم أخذًا في اعتباره قيمة كل رقم منها ورغم هذا فقلما يستخدم في الجغرافيا ربما لأنه يقدم من خلال مقياسين آخرين أكثر شيوعًا هما التباين والانحراف المعياري.

(٣) التباين : Variance

وهو من المقاييس الهامة المطبقة في الدراسات الجغرافية على نطاق واسع لأنه يظهر درجة التفاوت في توزيع ظاهرة ما مكانيًا، ويمكن أن يستخدم التباين في قياس التفاوت في توزيع ظاهرة واحدة بين الاقاليم الجغرافية في وقت معين، بمعنى إذا كان الجغرافي يريد معرفة درجة التباين في توزيع احجام سكان المدن في الدلتا والوجه القبلي مثلًا أو التباين في توزيع العاملين بالصناعة في اقسام محافظة الإسكندرية والقاهرة يمكنه استخدام هذا المقياس،

أو قد يستخدم في تتبع مدى اختلاف الظاهرة الواحدة في مجموعة من المناطق الجغرافية خلال فترات زمنية مختلفة كأن تحسب درجة التباين في توزيع الأميه عام ١٩٦٠ مثلًا بين مراكز إحدى المحافظات ثم نقارنها بالتباين في عام١٩٧٦ لتوضيح مقدار التكافؤ بين المراكز المختلفة في حصولها على نصيب من الخدمات التعليمية خلال الفترة بين ١٩٦٠،١٩٧٦.

ويمكن تعريف التباين بأنه بمجموع مربعات انحرافات القيم عن وسطها الحسابي مقسومًا على عددها، ونلجأ عادة لتربيع الانحرافات للتخلص من الإشارات السالبة فيها، وهو اسلوب شائع جدًا لدرجة أنك إذا سألت جغرافيًا عن النموذج الذي يرمي لاستخدامه سيرد عليك انني لم اختر نموذجًا محددًا وإنما طبقت تحليلًا للتباين، وتختلف الطرق المستعملة لقياس التباين فمنها على سبيل المثال التصنيفات وتحليل التمايز Discriminant Analysis المستخدم للفصل بين المجتمعات الإحصائية المتداخلة. ويمكن حساب التباين للأرقام المطلقة والنسبية وللبيانات المبوبة وغير المبوبة أي التي تأخذ صورة تكرارات أو أرقام مطلقة وفيما يلي أمثلة على ذلك.

مثال:

إذا كانت النسب التالية تمثل درجة الاستغلال الزراعي للأرض

في مراكز محافظة أسوان في عامي ١٩٦١ و١٩٩١

فيمكن حساب التباين لها على النحو التالي:

المركز	النسبة % عام ١٩٦١	الانحرافات	النسبة % عام ١٩٩١	الانحرافات	مربع ١٩٦١	مربع ١٩٩١
ادفو	٨٢	+ ٢٧	٨٣	– ٢	٧٢٩	٤
كوم أمبو	٤٩	– ٦	٨٨	+ ٣	٣٦	٩
نصر	٤٣	– ١٢	٨٩	+ ٤	١٤٤	١٦
أسوان	٤١	– ١٤	٦٧	– ١٨	١٩٦	٣٢٤
المتوسط العام	٥٥ %		٨٥	=	١١٠٥	٣٥٣
المتوسط الإجمالي	٥٣,٧٥ %		٨١,٧٥ %			

وعلى ذلك يكون التباين في عام ١٩٦١ = $\frac{\text{مجـ ح}^{٢}}{\text{ن}}$ = $\frac{١١٠٥}{٤}$ = ٢٧٦,٢٥

أما التباين في عام ١٩٩١ = $\frac{\text{مجموع مربعات الانحرافات}}{\text{عدد القيم}}$ = $\frac{٣٥٣}{٤}$ = ٨٨,٢٥

ومن الواضح أن درجة التباين في استغلال الارض اقتصاديًا كانت أكبر في التاريخ الأول منها في التاريخ الثاني، ويمكن من خلال الجدول السابق ملاحظة أن انحرافات القيم عن المتوسط العام في المحافظة لا تساوي صفرًا، وذلك لوجود فرق بين المتوسط الإحصائي إذا حصلنا عليه بجمع نسب المراكز المختلفة وقسمتها على عددها وبين الحصول على نسبة استغلال الأرض في المحافظة ككل لأن النسبة العامة في المحافظة يمكن أن تتأثر باختلاف توزيع المساحات بين المراكز، بمعنى إذا كان لديك واحدًا من هذه المراكز يضم نصف مساحة المحافظة فإن نسبة استغلال أراضيه ستؤثر بلا شك على النسبة العامة السائدة في المحافظة كلها إذا ارتفعت ترتفع معها وعند انخفاضها تتأثر بها.

أما إذا حسب المتوسط الإحصائي فإن النتيجة لابد وأن تساوي صفرًا؛ فالمتوسط الإحصائي يمكن الحصول عليه بجمع النسب وقسمتها على عددها فيكون في عام ١٩٦١ يساوي ٥٣,٧٥ ٪ بينما نسبة الاستغلال في المحافظة ٥٥ ٪، وكذلك فان متوسط عام ١٩٩١ يساوي ٨١,٧٥ ٪على حين أن نسبة الاستغلال في المحافظة ٨٥٪، والتباين في هذه الحالة يمكن أن يطلق عليه التباين الجغرافي لأنه يقيس درجة الاختلاف في توزيع ظاهرة (معينة) في فترتين مختلفتين بعدًا عن المتوسط العام لها في إطار المساحة الكلية والتي تتألف من الوحدات الأصغر، وهو أفضل من قياس التباين استنادًا إلى المتوسط الحسابي للنسب لأن هذا الأخير لا يعكس درجة إسهام الوحدات المساحية في النسبة العامة.

وقد يحسب التباين في توزيع الأمية بين مراكز محافظات الجمهورية المختلفة في سنة معينة مثلًا ليبين مدى التجانس والاختلاف في توزيع هذه الظاهرة بين المراكز في كل محافظة أو قد تحسب درجة التباين في توزيع الأمية بين محافظات الحضر والوجه البحري والقبلي على النحو التالي :

المحافظة	نسبة الأمية	المحافظة	نسبة الأمية	المحافظة	نسبة الأمية
القاهرة	٣٤,٠	الدقهلية	٥٧,٥	الجيزة	٥١,٦
الإسكندرية	٣٦,٩	الشرقية	٦١,٩	بني سويف	٦٨,١
بورسعيد	٣٥,٨	القليوبية	٥٣,٩	الفيوم	٧٢,٠
السويس	٤٢,٨	كفر الشيخ	٧٠,٤	المنيا	٦٩,٠
دمياط	٤٩,٩	الغربية	٥٥,٤	أسيوط	٦٨,٣
جملة الحضر	٤٢,٧	المنوفية	٥٨,٥	سوهاج	٧٠,٣
		البحيرة	٦٦,٦	قنا	٧١,٣
		الإسماعيلية	٥٠,٩	أسوان	٥٥,٨
		جملة وجه بحري	٥٦,١	جملة وجه قبلي	٦٥,٦

أولا: المتوسط الحسابي = $\frac{\text{مجموع نسب الأمية}}{\text{عدد المحافظات}}$ = $\frac{١٩٩,٤}{٥}$ = ٣٩,٨٨ = ٣٩,٩ تقريبًا

تباين المجموعة الأولى في محافظات الحضر:

				مربع الانحرافات
الانحرافات:	القاهرة	٣٤,٠ – ٣٩,٩	= – ٥,٩	٤٣,٨١
	الإسكندرية	٣٦,٩ – ٣٩,٩	= ٣,٠	٩,٠٠
	بورسعيد	٥٣,٨ – ٣٩,٩	= ٤,١	١٦,٨١
	السويس	٤٢,٨ – ٣٩,٩	= ٢,٩	٨,٤١
	دمياط	٤٩,٩ – ٣٩,٩	= ١٠,٠	١٠٠,٠٠
				١٦٩,٠٣

التباين = $\frac{١٦٩,٠٣}{٥}$ = ٣٣,٨

ويمكن تطبيق نفس الأسلوب لحساب تباين محافظات الوجهين البحري والقبلي لتكون النتائج ٣٦,٢، ٥٣,٧٥ على الترتيب.

وهكذا يبدو أن التجانس في ترتيب نسب الأمية يظهر واضحًا في محافظات الحضر يليها الوجه البحري ثم الوجه القبلي الذي تظهر قيمة التباين فيه مرتفعة.

وقد يحسب التباين بطريقة أخرى في هذه الحالة الأخيرة بالذات فبدلًا من الحصول على انحراف كل قيمة عن المتوسط الحسابي ثم تربيع الانحرافات يمكن الحصول على مربع كل قيمة ثم جمعه واستخدام المعادلة التالية :

$$\text{التباين } (\text{ع}^2) = \frac{\text{مجـ س}^2}{\text{ن}} - \left(\frac{\text{مجـ س}}{\text{ن}}\right)^2$$

ويعني هذا الحصول على مجموع مربعات القيم ثم يربع مجموع القيم ويقسم على عددها مع قسمة كل ذلك في النهاية على عدد القيم ويوضح المثال التالي ذلك :

القيم (س)	٥	٣	٢	٥	٦	٨	٥	١	٤	٣	= ٤٢
مربعها (س^2)	٢٥	٩	٤	٢٥	٣٦	٦٤	٢٥	١	١٦	٩	= ٢١٤

$$\text{التباين} = \frac{٢١٤}{١٠} - \left(\frac{٤٢}{١٠}\right)^2 = ٢١,٤ - (٤,٢)^2 = ٢١,٤ - ١٧,٦٤ = ٣,٧٦$$

ويلاحظ أن الشق الثاني من المعادلة الواقع بعد الإشارة ما هو إلا مربع المتوسط الحسابي للقيم ويمكن أيضًا حساب التباين بهذه الصورة:

$$ع^{٢} = \frac{\text{مجـ} (س - \bar{س})}{ن}$$

حيث تشير سَ إلى قيمة المتوسط الحسابي للأرقام، وفي هذه الحالة عليك الحصول على مجموع مربعات انحرافات كل قيمة عن الوسط الحسابي ثم تقسمه على عدد القيم الواردة في مجموعة البيانات كما يلى :

القيم	٥	٣	٢	٥	٦	٨	٥	١	٤	٣	
المتوسط	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	٤,٢	
الانحراف	٠,٨	١,٢	٢,٢	٠,٨	١,٨	٣,٨	٠,٨	٣,٢	٠,٢	١,٢	
المربع	٠,٦٤	١,٤٤	٤,٨٤	٠,٦٤	٣,٢٤	١٤,٤٤	٠,٦٤	١٠,٢٤	٠,٤	١,٤٤	=٣٧,٦

وبتطبيق القانون السابق تكون النتيجة :

$$ع^{٢} = \frac{٣٧,٦}{١٠} = ٣,٧٦$$ وهي ذات القيمة السابقة

- حساب التباين من الجدول التكرارية:

يمكن الحصول على التباين من جدول التوزيع التكراري بنفس الطريقة التي يتم حساب المتوسط بها حيث نحصل على مراكز الفئات والانحرافات عن الوسط الفرضي وتربع الانحرافات و تضرب في التكرارات ويطبق القانون بعدها على النحو التالي :

الفئات	التكرار	م	ح	ح × ك	ح² ك
٢ ـ	٧	٣	ـ ٦	ـ ٤٢	٢٥٢
٤ ـ	٨	٥	ـ ٤	ـ ٣٢	١٢٨
٦ ـ	٩	٧	ـ ٢	ـ ١٨	٣٦
٨ ـ	٦	٩	صفر	صفر	صفر
١٠ ـ	٤	١١	+ ٢	٨	١٦
١٢ ـ	٢	١٣	+ ٤	٨	٣٢
١٤ ـ	٤	١٥	+ ٦	٢٤	١٤٤
	٤٠			ـ ٩٢	٦٠٨
				+ ٤٠	
				ـ ٥٢	

$$(ع^{٢}) \text{ التباين} = \frac{\text{مجـ ح}^{٢} \text{ ك}}{\text{مجـ ك}} - \left(\frac{\text{مجـ ح ك}}{\text{مجـ ك}}\right)^{٢}$$

وبالتطبيق على المثال السابق:

$$\text{التباين } (\text{ع})^{٢} = \frac{٦٠٨}{٤٠} - \left(\frac{-٥٢}{٤٠}\right)^{٢} = ١٥,٢ - (-١,٣)^{٢} = ١٥,٢ - ١,٦٩ = ١٣,٥١$$

وفي هذا الجدول يمكن أيضًا تطبيق الانحرافات المختصرة التي سبقت الاشارة اليها عند حساب المتوسط الحسابي مع تعديل طفيف في القانون تضرب من خلاله القيمة الناتجة في الجزء الأخير من المعادلة في مربع القيمة التي تمت القسمة عليها وهي عادة تساوي طول الفئة (٢ في هذه الحالة) وتكون صيغة القانون:

$$\text{ع}^{٢} = \frac{\text{مجـ ح}^{٢}\text{ ك}}{\text{مجـ ك}} - \left(\frac{\text{مجـ حَ ك}}{\text{مجـ ك}}\right)^{٢} \times \text{ل}^{٢}$$

ويتطلب ذلك الحصول على مجـ حَ ك بإضافة عمود جديد للجدول السابق نقسم فيه كل قيمة من قيم الانحرافات على ٢ لتصبح قيم حَ كالتالي : - ٣ ، -٢،-١ ثم +١، + ٢، + ٣ وبضرب هذه الانحرافات المختصرة في التكرارات نحصل على -٢١، -١٦ ، - ٩ ثم +٤، +٤، +١٢ وعليه تكون نتيجة مجموع القيم السالبة

-٤٦ والموجبة ٢٠، والفرق = - ٢٦ وبعدها يطبق القانون :

$$\text{ع}^{٢} = \frac{٦٠٨}{٤٠} - \left(\frac{-٢٦}{٤٠}\right) \times (٢)^{٢} = ١٥,٢ - ٠,٤٢٢٥ \times ٤ = ١٥,٢ - ١,٦٩ = ١٣,٥١$$

ومعنى ذلك ببساطة أن التباين هو: "مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسومًا على مجموع التكرارات في القسم الأول من القانون ويطرح منه مجموع انحرافات القيم عن الوسط الحسابي مقسومًا على مجموع التكرارات مضروبًا في نفسه".

ويبين الجدول التالي توزيع المدن المصرية التي يزيد حجمها عن ٥٠ ألف نسمة في تعداد ١٩٨٦ حسب أعداد سكانها في فئات وذلك في كل من الوجهين البحري والقبلي والمطلوب حساب التباين في هذا التوزيع بين الوجهين وفي الجمهورية كلها:

فئات الحجم	٥٠ –	١٠٠ –	١٥٠ –	٢٠٠ –	٢٥٠ –	المجموع
وجه بحري	١٧	٣	٢	٢	٦	٣٠
وجه قبلي	١١	١	٣	١	١	١٧
مجموع المدن	٢٨	٤	٥	٣	٧	٤٧

تباين المجموعة الأولى:

الفئات	التكرارات (ك)	مركز الفئة م	الانحرافات ح	حَ(+٥٠)	حَ × ك	حَ^{٢} ك
٥٠ ــ	١٧	٧٥	ــ ١٠٠	ــ ٢	ــ ٣٤	+ ٦٨
١٠٠ ــ	٣	١٢٥	ــ ٥٠	ــ ١	ــ ٣	+ ٣
١٥٠ ــ	٢	١٧٥	صفر	صفر	صفر	صفر
٢٠٠ ــ	٢	٢٢٥	+ ٥٠	+ ١	+ ٢	+ ٢
٢٥٠ ــ	٦	٢٧٥	+ ١٠٠	+ ٢	+ ١٢	+ ٢٤
المجموع	٣٠				ــ ٣٧ + ١٤	٩٧
					ــ ٢٣	

$$ع^{٢} = \frac{\text{مجـ حَ}^{٢}\text{ ك}}{\text{مجـ ك}} - \left(\frac{\text{مجـ حَ ك}}{\text{مجـ ك}}\right)^{٢} \times ل^{٢}$$

$$= \frac{٩٧}{٣٠} - \left(\frac{-٢٣}{٣٠}\right)^{٢} \times (٥٠)^{٢}$$

$$= ٣,٢ - ٠,٥٨٧ \times (٥٠)^{٢} = ٢,٦ \times ٢٥٠٠ = ٦٥٠$$

وهكذا يمكن حساب تباين المجموعة الثانية :

ويعتبر التباين مقياسًا إحصائيا له قيمة كبيرة خاصة عندما تريد معرفة مقدار الاختلاف في بيانات عينة أو أكثر ويعرف هذا إحصائيًا بتحليل التباين.

٤ - الانحراف المعياري : Standard Deviation

"هو عبارة عن الجذر التربيعي للتباين سواء كان محسوبًا للقيم المطلقة أو للأرقام الموضوعة في جداول تكرارية ويفضل التباين في استخدامه لأن قيمته عادة ما تكون صغيرة وخصوصًا في حالة كبير الأرقام التي تهدف إلى التعرف على مدى تشتتها حيث يودي الاكتفاء بحساب التباين في هذه الحالات للحصول على أرقام كبيرة ولذا نحصل على الجذر التربيعي له فيقدم قيمة أصغر.

ويكون حساب الانحراف المعياري من القيم غير المبوبة في جداول تكرارية بنفس الطريقة التي حسب بها التباين مع اضافة الوصول إلى الجذر التربيعي في النهاية على النحو التالي:

إذا كان الجدول التالي يبين متوسطات إنتاج بعض المحاصيل في مراكز محافظة اسوان بين عامي ١٩٧٨ - ١٩٨١ والمطلوب حساب درجات الانحراف المعياري في انتاج المحاصيل المختلفة.

المنطقة	القمح (أرادب)	الذرة الرفيعة (أرادب)	الذرة الشامية (أرادب)	قصب السكر (بالطن)
إدفو	٦,٢	٩,٣	٨,١	٣٤,٢
كوم أمبو	٦,٦	٩,٨	٨,٧	٣٣,٩
نصر	٥,٥	٣,٦	٤,٧	٢٦,٤
أسوان	١١,١	١١,٥	١١,٦	–
وادي عبادي	٦,٠	٣,٦	٣,٥	٣٥,٢
المتوسط العام للمحافظة	٦,٤	٧,٥	٦,٦	٣٢,٨

ولحساب الانحراف المعياري نحصل على الانحرافات في كل محصول عن المتوسط العام للمحافظة مع ملاحظة أنه لا يمثل المتوسط الإحصائي الناتج عن قسمة مجموع متوسطات إنتاجية الفدان في كل المراكز على عدد المراكز وإنما يمثل المتوسط العام للمحافظة كلها الناتج عن قسمة مجموع الإنتاج لكل محصول على المساحة المزروعة به ، ثم تنتقل بعد ذلك للخطوة التالية وتربع فيها الانحرافات و تجمع ،و نحصل الجذر التربيعي لكل محصول على حدة وسنشير الى الانحرافات عن المتوسط في المحصول الأول بالرمز ح١، والثاني ح٢، والثالث ح٣ والرابع ح٤.

المنطقة	ح١	$ح_{١}^{٢}$	ح٢	$ح_{٢}^{٢}$	ح٣	$ح_{٣}^{٢}$	ح٤	$ح_{٤}^{٢}$
إدفو	- ٠,٢	٠,٠٤	+ ١,٨	٣,٢٤	+ ١,٥	٢,٢٥	+ ١,٤	١,٩٦
كوم أمبو	+ ٠,٢	٠,٠٤	+ ٢,٣	٥,٢٩	+ ٢,١	٤,٤١	+ ١,١	١,٢١
نصر	+ ٠,٩	٠,٨١	- ٣,٩	١٥,٢١	- ١,٩	٣,٦١	+ ٦,٤	٤٠,٩٦
أسوان	+ ٤,٧	٢٢,٠٩	+ ٤,٠	١٦,٠	+ ٥,٠	٢٥,٠٠	-	-
وادي عبادي	- ٠,٤	٠,١٦	٣,٩	١٥,٢١	- ٣,١	٩,٦١	- ٧,٦	٥٧,٧٦
المجموع		٢٣,١٤		٥٤,٩٥		٤٤,٨٨		١٠١,٨٩

وبتطبيق القانون: $\sqrt{\frac{\text{مجـ ح}^{٢}}{\text{ن}}}$ يمكن الحصول على الانحراف المعياري لكل محصول من المحاصيل ، وتكون النتائج كالتالي:

نوع المحصول	الانحراف المعياري	نوع المحصول	الانحراف المعياري
القمح	٢,١٤	الذرة الشامية	٣,٠
الذرة الرفيعة	٣,٣٢	قصب السكر	٥,٠٤

ويلاحظ أنه لا يمكن مقارنة الانحراف المعياري للقصب ببقية المحاصيل لأن وحدة القياس المستخدمة فيه تختلف عن المحاصيل الثلاثة الأخرى، ونخلص لنتيجة مؤداها أن الانحراف المعياري في توزيع الإنتاجية بين هذه المناطق أعلى ما يكون في حالة محصول الذرة الرفيعة، ويليها الذرة الشامية ثم في النهاية القمح الذي تميل متوسطات إنتاجيته للتجانس.

كذلك يمكن حساب الانحراف المعياري لتوزيع ظاهرة معينة مكانيًا في تاريخين مختلفين ويقارن بين النتائج في الحالتين، وبناءً عليه يظهر مدى التكافؤ في توزيع العوامل المسئولة عن الظاهرة في الوحدات الإقليمية، و لإيضاح ذلك يمكن حساب الانحراف المعياري في توزيع نسبة الأمية في مراكز محافظة أسوان من الجدول التالي:

المنطقة	النسبة عام ١٩٦٠	ح	ح^2	النسبة عام ١٩٧٦	ح	ح^2
إدفو	٦٧,٧	+ ٧,٥	٥٦,٣	٦٢,٠	+ ٦,٢	٣٨,٤٤
كوم أمبو	٧١,٢	+ ١٠,٠	١٠٠	٦٦,٤	+ ١٠,٦	١١٢,٣٦
نصر	٤٢,٨	- ١٨,٤	٣٣٨,٦	٥٢,٦	- ٣,٢	١٠,٢٤
أسوان	٥٦,٢	- ٥,٠	٢٥,٠	٤٩,٤	- ٦,٤	٤٠,٩٦
مدينة أسوان	٣٢,٦	- ٢٥,٦	٦٥٥,٤	٣٦,٢	- ١٩,٦	٣٨٤,١٦
متوسط المحافظة	٦١,٢		١١٧٥,٣	٥٥,٨		٥٨٦,١٦

وتكون قيم الانحراف المعياري لتوزيع الأمية عام ١٩٦٠ = √‾١١٧٥,٢‾ = ٣٤,٢

وفي عام ١٩٧٦ = √‾٥٨٦,١٦‾ = ١٠,٨٣

ويبدو من هذه القيم أن درجة الانحراف في توزيع الامية بين مراكز المحافظة في عام ١٩٧٦ أصبحت أقل بما يشير إلى أن توزيع الخدمات التعليمية أصبح أكثر ميلًا للتكافؤ بين المراكز عام ١٩٧٦ عنه في ١٩٦٠ والذي ظهرت فيه درجة الانحراف بصورة أكبر، ويمكن التوسع بحساب الانحراف المعياري في توزيع الأمية بين الذكور والاناث في التاريخين ومعرفة مدى التجانس أو التباين في التوزيع الجغرافي لكل منها في فترتين أو أكثر من ذلك.

ويحسب الانحراف المعياري أحيانًا بطريقة أخرى إذا كانت الارقام بسيطة وذلك على النحو التالي :

الأرقام (القيم) ٣، ٢، ١، ٢، ٣، ٤، ٣، ٧، ٦، ٥ = ٣٦

وفي هذه الحالة نحصل على مربعات القيم مباشرة كما يلي :

س^2: ٩، ٤، ١، ٤، ٩، ١٦، ٩، ٤٩، ٣٦، ٢٥ = ١٦٢

ويطبق القانون التالي:

$$\text{ع (الانحراف المعياري)} = \sqrt{\frac{\text{مجـ س}^{٢}}{\text{ن}} - \bar{\text{س}}^{٢}}$$

$$\text{وعلى ذلك تكون قيمة ع} = \sqrt{\frac{١٦٢}{١٠} - (٣,٦)^{٢}} = \sqrt{٣,٢٤} = ١,٨$$

ولا يقتصر استخدام الانحراف المعياري على الجوانب البشرية في الجغرافيا وإنما يمكن حسابه في الجغرافيا الطبيعية فإذا كانت لدينا كميات المطر السنوي بالسنتيمتر في مدينين خلال الفترة من ١٩٧٢ الى عام ١٩٨١ يمكن حساب الانحراف المعياري لها على النحو التالي :

السنة	المدينة الأولى		المدينة الثانية	
	الكمية	مربع الكمية	الكمية	مربع الكمية
١٩٧٢	١٢٥	١٥٦٢٥	١٦٠	٢٥٦٠٠
١٩٧٣	١٢٨	١٦٣٨٤	٨٤	٧٠٥٦
١٩٧٤	١٣٢	١٧٤٢٤	١٣٩	١٩٣٢١
١٩٧٥	١٢٧	١٦١٢٩	١٨٧	٣٤٩٦٩
١٩٧٦	١٢٠	١٤٤٠٠	١٢٨	١٢٦٣٨٤
١٩٧٧	١٢٣	١٥١٢٩	٦٧	٤٤٨٩
١٩٧٨	١٣٥	١٨٢٢٥	٩٧	٩٤٠٩
١٩٧٩	١٢٣	١٥١٢٩	١٠٠	١٠٠٠٠
١٩٨٠	١١٨	١٣٩٣٤	٨٣	٦٨٨٩
١٩٨١	١٢٢	١٤٨٨٤	٦٩	٤٧٦١
الجملة	١٢٥٣	١٥٧٢٥٣	١١١٤	١٣٨٨٧٨

وتظهر المتوسطات أن المدينة الأولى أعلى قليلاً من الثانية من حيث متوسطها الحسابي (١٢٥,٣سم^{٣}، مقابل ١١١,٤ سم^{٣})

أما إذا حسب الانحراف المعياري، فسيعطي نتيجة مختلفة إلى حد ما، وذلك على النحو التالي:

$$\text{ع} = \sqrt{\frac{\text{مجـ س}^{٢}}{\text{ن}} - \left(\frac{\text{مجـ س}}{\text{ن}}\right)^{٢}}$$

$$\text{ففي المدينة الأولى} = \sqrt{\frac{١٥٧٢٥٣}{١٠} - \left(\frac{١٢٥٣}{١٠}\right)^{٢}}$$

$$= \sqrt{١٥٧٢٥,٣ - ١٥٧٠٠,٠٩} = \sqrt{٢٥,٢١} = ٥,٠١$$

وفي المدينة الثانية = $\sqrt{\frac{١٣٨٨٧٨}{١٠} - \left(\frac{١١١٤}{١٠}\right)^{٢}}$

$$= \sqrt{١٣٨٨٧,٨ - ١٢٤٠٩,٩٦} = \sqrt{١٤٧٧,٨٤} = ٣٨,٤٤$$

وعلى ذلك فعند المقارنة بين المدينتين ستكون النتائج كالآتي:

	المتوسط	الانحراف المعياري
المدينة الأولى	١٢٥,٣	٥,٠١
المدينة الثانية	١١١,٤	٣٨,٤٤

ويبين ذلك أهمية حساب الانحراف المعياري حيث أن المدينة الأولى متوسطها أكبر قليلًا من الثانية على حين ظهر الانحراف المعياري في الحالة الثانية كبيرًا (أكثر من سبعة أضعاف المدينة الأولى) ومعنى ذلك أن معظم قيم المطر في المدينة الأولى تتركز حول المتوسط بينما تتذبذب أرقام المطر في المدينة الثانية بصورة كبيرة.

والخلاصة أنه لحساب الانحراف المعياري لأي مجموعة من القيم غير المبوبة تتبع الخطوات التالية :

١- توضع القيم في صورة جدول وتربع و يجمع مربعها ويقسم على عددها.

٢- يحسب المتوسط الحسابي للقيم ويربع.

٣ - يطرح مربع المتوسط من ناتج مجموع مربعات القيم و ونحصل علي جذر القيمة.

- حساب الانحراف المعياري من الجداول التكرارية :

يختلف حساب الانحراف المعياري من البيانات المبوبة في جداول عن الأرقام المطلقة في ناحيتين هما :

أ- أن الانحراف في هذه الحالة سينصب على الارقام الموضوعة في فئات ولذلك سيكون انحرافًا عن الوسط الفرضي.

ب - أنه لابد من ضرب الناتج في طول الفئة المستخدمة في الجدول التكراري .

وفيما يلي مثال لحساب الانحراف المعياري :

الفئة (١)	مركز الفئة (٢)	التكرار (٣)	ح (٤)	حَ (٥)	حَ×ك (٦)	حَ^{٢} ك (٧)
صفر –	٥	١٠	– ٢٠	– ٢	– ٢٠	٤٠
١٠ –	١٥	١٣	– ١٠	– ١	– ١٣	١٣
٢٠ –	٢٥	٢٥	صفر	صفر	صفر	صفر
٣٠ –	٣٥	٢٠	+ ١٠	+ ١	+ ٢٠	٣٠
٤٠ –	٤٥	٨	+ ٢٠	+ ٢	+ ١٦	٣٢
٥٠ –	٥	٤	+ ٣٠	+ ٣	+ ١٢	٣٦
		٨٠			– ٣٣ + ٤٨	١٤١
					+ ١٥	

ويمكنك ملاحظة أن مجموع التكرارات (مجـ ك) = ٨٠، ومجموع حاصل ضرب الانحرافات المختصرة بعد قسمتها على ١٠ يساوي الفرق بين ٣٣، ٤٨ أي ١٥، وقد رمز له بالرمز مجـ ح/ك ، أما مجموع حاصل ضرب مربعات الانحرافات في التكرارات فيساوي ١٤١ ويرمز له مجـ ح/٢ك ، وبناءً على ما سبق يلزم لحساب الانحراف المعياري من الجداول التكرارية مجـ ك، مجـ حَ ك، مجـ ح٢ك ثم ل (طول الفئة) وهو في هذه الحالة يساوي عشرة ، وبعد ذلك تطبق المعادلة :

$$\text{ع} = \text{ل}\sqrt{\frac{\text{مجـ حَ}^{٢}\text{ ك}}{\text{مجـ ك}} - \left(\frac{\text{مجـ حَ ك}}{\text{مجـ ك}}\right)^{٢}}$$

$$= ١٠\sqrt{\frac{١٤١}{٨٠} - \left(\frac{١٥}{٨٠}\right)^{٢}}$$

$$= ١٠\sqrt{١{,}٧٦ - ٠{,}٠٤} = ١٠\sqrt{١{,}٧٢} = ١٠ \times ١{,}٣١ = ١٣{,}١$$

والخلاصة أنه لحساب الانحراف المعياري من القيم التكرارية تتبع الخطوات التالية:

(١) تحسب مراكز الفئات (عمود رقم ٢ في الجدول).

(٢) يختار وسط فرضي مناسب من بين مراكز الفئات السابقة وهو في الحالة السابقة (٢٥)

(٣) تحسب الانحرافات عن الوسط الفرضي لكل مراكز الفئات (ح) عمود ٤

(٤) تختصر الانحرافات بقسمتها على طول الفئة في التوزيع التكراري وهي في الجدول السابق =١٠ لنحصل على الانحرافات المختصرة (حَ) عمود ٥.

(٥) تضرب الانحرافات المختصرة في التكرارات على النحو المستخدم في حساب الوسط الحسابي لنحصل على (حَ ك) عمود ٦

(٦) تربع الانحرافات في العمود رقم ٥ ثم تضرب في التكرارات (عمود رقم ٣) لنحصل على ح/٢ك في العمود رقم ٧.

(٧) نجمع العمود رقم ٣ والعمود رقم ٦ والعمود رقم ٧ لنحصل على مجـ ك، مجـ حَ ك ، مجـ ح/٢ك ، ويطبق القانون.

وللانحراف المعياري عدة خصائص إحصائية هي :

١- أن ٥٦٪ من القيم الواقعة في أي توزيع تنحصر على الأقل بين المتوسط الحسابي وما يعادل مرة ونصف المرة من قيمة الانحراف المعياري زيادة ونقص.

٢ - أن ٧٥ ٪ من هذه القيم يقع بين المتوسط الحسابي وضعف قيمة الانحراف المعياري سالبًا وموجبًا.

٣- أن ٨٩ ٪من القيم يقع بين المتوسط وثلاثة أمثال الانحراف المعياري، زيادة ونقصًا.

٤- أن ٩٤٪ من القيم تقع على الأقل بين المتوسط وأربعة أمثال الانحراف المعياري زيادةً ونقصًا.

(٥) معامل الاختلاف :

وهو أحد مقاييس التشتت أيضًا وتقوم فكرته على قسمة الانحراف المعياري على الوسط الحسابي للقيم وتحويله إلى نسبة مئوية بضربه في ١٠٠. وفي حالة المثال السابق الخاص بكميات الأمطار في المدينتين فان معامل الاختلاف للمدينة الأولى يكون :

$$ف_{ن} = \frac{ع}{\bar{س}} \times ١٠٠$$

حيث ترمز ف للمعامل، ع الانحراف المعياري، سَ المتوسط

وبالتعويض فإن ف = $\frac{٥,٠١}{١٢٥,٣} \times ١٠٠$ = ٤ ٪ تقريبًا.

أما المدينة الثانية فإن معامل الاختلاف = $\frac{٣٨,٤٤}{١١١,٤} \times ١٠٠$ = ٣٤,٥ ٪

ويظهر ذلك أن نسبة الاختلاف في توزيع قيم المطر في المدينة الثانية أكبر بكثير من مثيله في المدنية الأولى، ويمكن ملاحظة ذلك بوضوح بمجرد النظر الى تشتت كميات المطر السنوي في كل حالة.

ويمكن حساب معامل الاختلاف في متوسط درجات الحرارة الشهرية في المدن المصرية المختلفة خلال مجموعة من السنوات وتوقيعه على خرائط لمعرفة الاختلافات

المحلية، كذلك من السهل حساب الاختلافات في توزيع أي ظاهرة بشرية باستخدام نفس الأسلوب وذلك على النحو التالي :

إذا كان لديك التوزيع التالي للأنشطة الاقتصادية بين السكان العاملين في جمهورية مصر العربية و مدينتي القاهرة والإسكندرية في عام ١٩٧٦ فاحسب الانحراف المعياري ومعامل الاختلاف في توزيع الأنشطة في الجمهورية والقاهرة والإسكندرية.

الحرفة	الجمهورية ٪	القاهرة	ح	$ح^٢$	الإسكندرية٪	ح	$ح^٢$
الزراعة والصيد	٤٨,٤	١,٢	ـ ٤٧,٢	٢٢٢٧,٨٤	٥,٩	ـ ٤٢,٥	١٨٠٦,٢٥
التعدين	٠,٣	٠,٤	+ ٠,١	٠,٠١	٠,٤	+ ٠,١	٠,٠١
الصناعة	١٣,٦	٢٧,٥	+ ١٣,٩	١٩٣,٢١	٣٣,٦	+ ٢٠,٠	٤٠٠,٠
الكهرباء والغاز	٠,٦	١,١	+ ٠,٥	٠,٢٥	١,٣	+ ٠,٧	٠,٤٩
التشييد والبناء	٤,٢	٨,٥	+ ٤,٣	١٨,٤٩	٦,٧	+ ٢,٥	٦,٢٥
التجارة والمطاعم	٨,٦	١٥,١	+ ٦,٥	٤٢,٢٥	١٤,٥	+ ٥,٩	٣٤,٨١
النقل والتخزين	٤,٨	٩,٠	+ ٤,٢	١٧,٦٤	٩,٧	+ ٤,٩	٢٤,٠١
التمويل والتأمين	٠,٩	٢,١	+ ١,٢	١,٤٤	١,٥	+ ٠,٦	٠,٣٦
الخدمات	١٨,٦	٣٥,١	+ ١٦,٥	٢٧٢,٢٥	٢٦,٤	+ ٧,٨	٦٠,٨٤
الإجمالي	١٠٠	١٠٠		٢٧٧٣,٣٨	١٠٠		٢٣٣٣,٠٢

وفي هذه الحالة اعتبرت نسبة العاملين في الجمهورية من كل نوع من أنوع النشاط الاقتصادي ممثلة للمتوسط ليحسب بعدها الانحراف المعياري ومعامل الاختلاف لمدينتي القاهرة والإسكندرية كما يلى:

القاهرة = $\sqrt{\frac{٢٧٧٣,٣٨}{٩}}$ = $\sqrt{٣٠٨,١٥}$ = ١٧,٥٥

الإسكندرية = $\sqrt{\frac{٢٣٣٣,٠٢}{٩}}$ = $\sqrt{٢٥,٢٢}$ = ١٦,١

أما معامل الاختلاف فتقسم فيه قيم الانحراف المعياري على المتوسط العام للعاملين في الحرف المختلفة وهو:

العاملين الحرف المختلفة وهو = $\frac{١٠٠٪}{٩}$ = ١١,١ ٪ تقريبًا

ومن ثم فالناتج: في القاهرة = $\frac{١٧,٥٥ \times ١٠٠}{١١.١}$ = ١٥٨,١

وتعنى هذه القيم أن مدى التشتت في توزيع السكان حسب أوجه النشاط الاقتصادي في مدينة القاهرة أكبر منه في مدينة الاسكندرية وذلك إذا قورنت المدينتان بالجمهورية ككل من حيث توزيع العاملين فيهما على الأنشطة الاقتصادية .

وفي مثل هذه الحالة حسب معامل الاختلاف استنادًا إلى المتوسط الحسابي الذي قيست الانحرافات بعدًا عنه وهو متوسط نسبة العاملين في كل حرفة في الجمهورية ، وهو متفاوت بين الزراعة والتعدين والصناعة وهكذا ، لذا يكتفي عندئذ بحساب الانحراف المعياري او الانحراف المتوسط أو التباين لأن متوسط نسبة العاملين في كل من الجمهورية والاسكندرية والقاهرة في الحرفة الواحدة سيكون مساويًا لمجموع النسب المئوية والتي تساوي ١٠٠٪ في كل حالة مقسومًا على عدد الحرف وهي ٩ أي أنه يساوي ١١,١ في كل الحالات.

ثانيًا: مقاييس الانتشار

تقاس درجة الانتشار عادة حول نقاط معينة قد تكون الوسط أو الوسط الجغرافي أو الهندسي أو أي نقطة أخرى يراد قياس انتشار صورة توزيعية محددة حولها ، وهذا النوع من المقاييس له قيمته في إظهار مدى التباعد أو التقارب المكاني للظاهرات ، ويشترط فيه غالبًا معرفة المساحة الأصلية لمنطقة التوزيع وعدد النقاط أو المساحات المحددة وتوقيعها في أماكن على الخريطة بدقة وقياس المسافات الفاصلة بينها، ووضوح مقياس رسم الخريطة المستخدمة ، والإلمام الجيد بالقواعد الإحصائية الأساسية .

وأهم مقاييس الانتشار هي :

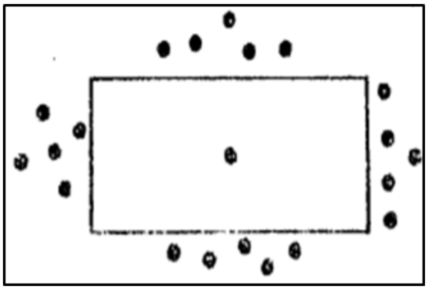
١- الربيع الجغرافي :

يمكن بصفة عامة قياس الانتشار حول الوسيط الجغرافي ما يعرف باسم الربيع الجغرافي وفيه تقسم المنطقة حول الوسيط حسب الجهات الجغرافية الأصلية الأربع (شمالي - جنوبي - شرقي - غربي) ويحدد الخط الذي يقع خارجه ربع عدد النقاط بعدًا عن الوسط فإذا كانت هذه النقاط شمال الخط أطلق عليه اسم ربيع شمالي وإذا كانت جنوبه فهو ربيع جنوبي وهكذا الشرقي والغربي ، ومعنى هذا أن الربيع يعني الخط الذي يقطع أو يخترق توزيعًا معينًا لمجموعة من النقاط بحيث تتوزع بنسبة ١:٣على جانبيه ؛أي أنه يعزل ربع عدد النقاط في جانب واحد منه و الثلاثة أرباع الأخرى على الجانب الآخر ويرسم الربيع أو يحدد من قبل الباحث حسب صورة التوزيع الواقع على الخريطة وليس من الضروري أن يكون متفقًا مع الجهات الأصلية الأربع إنما قد يساير اتجاهات فرعية ، والمهم في النهاية خروج المساحة المحصورة بين هذه الربيعات في صورة شكل هندسي منتظم (مستطيل) .

٢- معامل الانتشار :

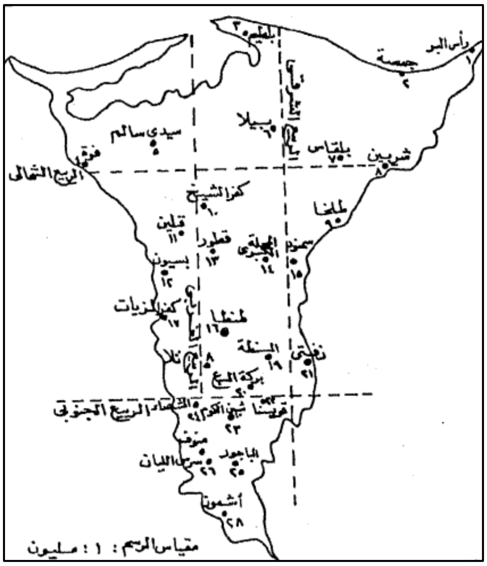
تحسب المساحة الواقعة داخل الربيعات الأربع، وكلما كانت كبيرة دل " ذلك على عظم الانتشار وعندما تقل يميل التوزيع للتقارب مكانيًا .

$$\text{فمعامل الانتشار} = \frac{\text{مساحة المستطيل المحدد بالربيعات}}{\text{إجمالي المساحة الواقع فيها التوزيع}}$$



شكل يوضح كيفية تعيين الربيع الجغرافي

وتتروح قيمة المعامل بين صفر في حالة التركز الكامل حول نقطة المركز ، ومعنى هذا أن الربيعات تتطابق ولا تترك بينها أي مساحة ،وعندما تقع كل المساحة في إطار الربيعات يكون المعامل مساويًا لواحد صحيح ويظهر التوزيع بعيدًا كلية عن النقطة المتوسطة محققًا أقصى انتشار ،أما التوزيع المثالي فيتحقق عندما يعادل المستطيل ربع المساحة الكلية، وإذا طبق هذا المعامل على مدن الدلتا الثماني والعشرين (المبينة على الخريطة) وباعتبار طنطا نقطة مركزية تتوسطها فيمكن تعيين الربيعات الأربع على النحو المبين بحيث تقع كل سبع منها شمال و شرق وجنوب وغرب كل ربيع ثم يحسب معامل الانتشار على النحو التالي :



طول المستطيل = ٦سم

عرض المستطيل = ٢,٥ سم

ولما كان مقياس رسم الخريطة ١: مليون

فإن: الطول الحقيقي للمستطيل يكون٦٠٠٠٠٠٠سم أي٦٠٠٠٠متر =٦٠ كيلو متر

والعرض = ٢,٥سم = ٢٥٠٠٠٠٠سم = ٢٥٠٠٠متر = ٢٥ كيلو متر

ومساحة الدلتا بين الفرعين ١٠ آلاف كيلو متر مربع

$$\text{فمعامل الانتشار} = \frac{٦٠ \times ٢٥}{١٠٠٠٠} = \frac{١٥٠٠}{١٠٠٠٠} = ٠,١٥$$

والمقام هنا يساوي مساحة الدلتا (عشرة آلاف كيلو متر مربع).

ومن الواضح أن الرقم يقترب من الصفر وبالتالي تميل هذه المدن إلى التركز حول مدينة طنطا بصورة أكبر من ميلها إلى الانتشار بعيدًا عنها أو بمعنى آخر هي أقرب إلى التوزيع المثالي حيث تكون مساحة المستطيل تساوي ربع مساحة المنطقة : وهنا مساحته تصل إلى ٠,١٥ منها .

والمشكلة التي تواجه هذه الطريقة هي صعوبة تقسيم عدد نقاط التوزيع أحيانًا بين أربع جهات أصلية أو فرعية كأن يكون العدد ٣٠ مدينة في الحالة السابقة ، ولحل ذلك توزع المدن الإضافية السابقة على أي جهتين بحيث يكون ٧ ،٧ ، ٨ ، ٨.

٣- الانتشار حول موقع معين:

وقد يقاس مدى الانتشار حول موقع معين بطريقة أخرى تقوم على رسم مجموعة من الدوائر المتعاقبة يكون مركزها الموقع الذي يراد معرفة أبعاد التوزيع حوله وغالبًا ما تستخدم هذه الطريقة في دراسات السكان والعمران حيث يمكن في الحالة الأولى تعيين نقطة الوسط السكاني و ترسم حولها بأنصاف أقطار تتزايد بقيم معينة وتحسب نسبة السكان التي تضمها هذه الدوائر من جملة سكان المنطقة أو يعين قلب المدينة العمراني وترسم نفس هذه الدوائر ويعرف منها مدى تركز أو انتشار الوظائف التي تؤديها المدينة حول هذا القلب ، وتمثل الوظائف عادة من خلال أعداد المحلات التي تقدم كل نوع منها.

ولتوضيح ذلك إذا فرض أنك تقوم بدراسة عن مدينة الإسكندرية وحددت قلبها التجاري في ميدان المنشية مثلًا ، وبدأت ترسم مجموعة من الدوائر مركزها هذا القلب ، وكانت أنصاف أقطار هذه الدوائر تبدأ بربع كيلو متر ثم نصف ثم كم ثم ١,٠ كم ثم ٢ كم ، وبدأت بعد ذلك في توقيع محلات الأحذية كنوع من محلات تجارة التجزئة ، وعيادات الأطباء باعتبارها تمثل نوعًا من الخدمات الصحية التي تقدمها المدينة لسكانها وروادها، ووجدت أن عدد محلات الأحذية وعيادات الأطباء تتوزع بعدًا عن قلب المدينة كما يلى :

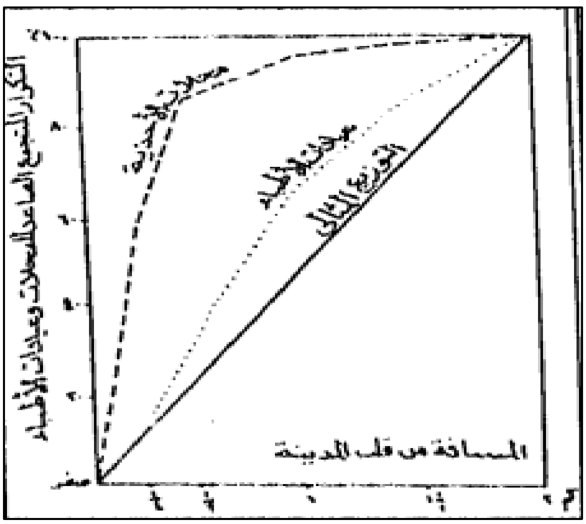
المسافة من القلب	محلات الأحذية			عيادات الأطباء		
	العدد	٪	النسب التراكمية	العدد	٪	النسب التراكمية
١/٤ كيلو متر	٨٠٠	٥٦,٨	٥٦,٨	٩٨	١٣,٢	١٣,٢
١/٢ كيلو متر	٤٠٠	٢٨,٤	٨٥,٢	١٦٢	٢١,٩	٣٥,١
١ كيلو متر	١٥٠	١٠,٦	٩٥,٨	٢٥٠	٣٣,٨	٦٨,٩
١,٥ كيلو متر	٥٠	٣,٥	٩٩,٣	١٥٠	٢٠,٣	٨٩,٢
٢ كيلو متر	١٠	٠,٧	١٠٠	٨٠	١٠,٨	١٠٠
المجموع	١٤١٠	١٠٠		٧٤٠	١٠٠	

وهكذا يمكن الخروج بنتيجة مؤداها أن حوالى ٥٧٪ من محلات الأحذية تقع في الدائرة الأولى والتي لا تبعد سوى ١/٤ كم من المركز ، وإذا اتسعت الدائرة لتصبح نصف كم تضم حوالي ٨٥ ٪من عدد هذه المحلات.

أما إذا انتقلنا إلى الدائرة الأكبر والتي يبلغ نصف قطرها كيلومتر تصبح النسبة حوالى ٩٦ ٪ وهكذا فإنه في دائرة نصف قطرها كيلو متر واحد تقع أغلبية محلات الأحذية (٩٦٪) على حين يقع نصف عدد هذه المحلات في حدود ربع كيلو متر من قلب المدينة.

اما بالنسبة لعيادات الأطباء فالأمر مختلف حيث لا تجاوز نسبة الموجود منها في حدود ربع كيلومتر ١٣٪ وحتى إذا أضيفت الدائرة الثانية فإن عدد العيادات لا يتعدى ثلثها الكلى بكثير ومن هنا فهي أقرب إلى التوزع بصورة أكثر انتشارًا من النوع السابق .

وإذا جمعت هذه النسب جمعًا تراكميًا (تكرار متجمع صاعد) وعرفت علاقتها بذلك الجزء من مساحة المدينة الواقع داخل الدائرة فإنه في الإمكان معرفة صورة التوزيع خصوصًا إذا ما رسم ذلك في صورة منحنى تكراري متجمع صاعد على النحو التالي :



ومن هذا الشكل يظهر أن توزيع عيادات الأطباء اقرب إلى التوزيع المثالي من توزيع محلات بيع الأحذية . ويمكن بعد ذلك حساب وسيط المسافة والذي يعين نصف قطر الدائرة التي يقع داخلها ٥٠% من عدد المحلات أو العيادات وهي في حالة محلات الأحذية تقل قليلًا عن١/٤ كيلومتر وفي حالة عيادات الأطباء تقل عن كيلومتر .

٤ - المسافة المعيارية :

وتقوم فكرتها على حساب الجذر التربيعي لمجمع مربعات انحرافات القيم في س ، ص عن الوسط الحسابي مع قسمته على عدد من قيم س ، ص بحيث يكون الناتج في النهاية رقما يبين تركز ٦٨ ٪من القيم حول نفطة الوسط ، وبالتالي فهي المسافة التي تظهر مدى انتشار مجموعة من النقاط حول نقطة الوسط الجغرافي وتحسب بالقانون التالي :

$$\text{المسافة المعيارية} = \sqrt{\frac{\text{مجـ س}^{٢}}{\text{ن}} - (\bar{\text{س}})^{٢} + \frac{\text{مجـ ص}^{٢}}{\text{ن}} - (\bar{\text{ص}})^{٢}}$$

حيث تشير (مجـ س^{٢} – سَ) لمجموع مربعات انحرافات القيم في حالة س عن الوسط الحسابي (سَ) وفي حالة ص تكرر نفس الانحرافات، ن تشير إلى عدد القيم.

<u>وفيما يلي مثال لحساب المسافة المعيارية:</u>

إذا كانت لديك أبعاد عشر مدن عن الإحداثي الشرقي (السينى) والإحداثي الشمالي (الصادي) فاحسب المسافة المعيارية لها.

المدن	الإحداثي الشرقي (س)	مربع س	الإحداثي الشمالي (ص)	مربع ص
أ	٥	٢٥	٨	٦٤
ب	٣	٩	٧	٤٩
ج	٦	٣٦	٧	٤٩
د	٨	٦٤	٧	٤٩
هـ	٤	١٦	٦	٣٦
و	٥	٢٥	٦	٣٦
ز	٢	٤	٥	٢٥
ح	٥	٢٥	٥	٢٥
ط	٦	٣٦	٩	٨١
ي	٦	٣٦	١٠	١٠٠
المجموع	٥٠	٢٧٦	٧٠	٥١٤

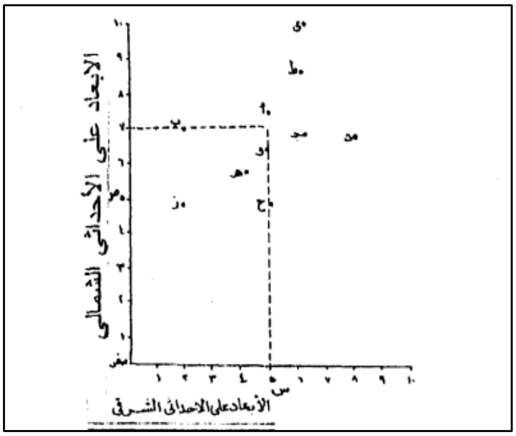
والخطوة الأولى هي حساب المتوسط الحسابي لقيم س، ص وهو ٥ ، ٧ على الترتيب.

والخطوة الثانية هي إيجاد حاصل جمع مربع كلا من س ، ص على النحو المبين في الجدول السابق، ثم يطبق القانون السابق على النحو التالي:

$$\text{المسافة المعيارية} = \sqrt{\frac{\text{٢٧٦}}{\text{١٠}} - (\text{٥})^{\text{٢}} + \frac{\text{٥١٤}}{\text{١٠}} - (\text{٧})^{\text{٢}}}$$

$$= \sqrt{(\text{٢٧,٦} - \text{٢٥}) + (\text{٥١,٤} - \text{٤٩})} = \sqrt{\text{٢,٦} + \text{٢,٤}} = \sqrt{\text{٥}} = \text{٢,٢٤}$$

ومعنى هذه القيمة أنه في دائرة قطرها يساوي هذه المسافة يقع ٦٨ ٪من النقاط حول نقطة الوسط الجغرافي المعينة في الشكل التالي، وقد جاءت نسبة ٦٨% هذه من الحقيقة المتصلة بطبيعة العلاقة بين المتوسط الحسابي و الانحراف المعياري حيث يقع ٦٨٪ من القيم أو المساحة الجغرافية بين المتوسط الحسابي والانحراف المعياري.



ه - مقياس أقرب جار أو صلة الجوار :

تستخدم المعادلة التالية في الجغرافيا لقياس مدى بعد توزيع معين عن العشوائية ، وذلك من خلال توقيع مجموعة من النقاط تمثل توزيعًا معينًا على محورين :

$$\text{ر} = \text{٢سَ} \sqrt{\frac{\text{ن}}{\text{م}}}$$

وتشير لقيمة أقرب جار، سَ = متوسط التباعد بين عدد من المراكز العمرانية المتجاورة، ن = عدد مراكز العمران، م = مساحة المنطقة موضع البحث ،و تنحصر القيمة الناتجة عادة بين صفر، ٢,١٤٩١ ، وكلما اقتربت القيمة من الرقم الأخير أخذ التوزيع شكلًا قريبًا من المثالية (الشكل السداسي في نموذج كريستلر).

أي أن كل نقطة تكون على بعد مساو من النقاط الست الأخرى وعند الرقم صفر تتجمع كل النقاط في شكل عنقود، ويشير الرقم (١) إلى توزيع عشوائي كامل. وعلى سبيل

المثال إذا كان الجدول التالي يبين توزيع مراكز العمران في نطاق ترعة الاسماعيلية تبعًا لمراكزها الإدارية في عام ١٩٩٠ فيمكن من خلاله حساب مقياس الجار الأقرب على النحو التالي إذا كان ذلك للمنطقة كلها:

المركز	المساحة كيلو متر مربع	عدد مراكز العمران	متوسط التباعد
شبين القناطر	١٣	٥	١,٧
مشتول السوق	١٦	٢	٣,٠٠
أبو حماد	٢٦٢	٢٩	٣,٢
قليوب	٤٩	٥	٣,٤
الخانكة	١٢٦	١٢	٣,٥
بلبيس	٢١٧	١٦	٤,٠
التل الكبير	١٩٧	٥	٦,٧
الإسماعيلية	٩٧٣	٨	١١,٩
الإجمالي	١٨٥٣	٨٢	٥,١

$$\text{ر} = ٢ \times ٥,١ \sqrt{\frac{٨٢}{١٨٥٣}}$$

$$= ١٠,٢ \sqrt{٠,٠٤٤٢}$$

$$= ١٠,٢ \times ٠,٢١٠ = ٢,١٤٥٧$$

ولذا فالقيمة يقترب بها نمط التوزيع من الشكل المثالي إلى حد كبير ويمكنك حساب هذا المقياس لكل مركز من المراكز المبينة في الجدول ومقارنة النتائج.

الفصل السادس

المقَارنَات

الفصل السادس

المقَارَنات

يهتم الجغرافي في دراسته بإجراء المقارنات بين الظاهرات المختلفة، حيث تمثل مجالًا هامًّا لإبراز أوجه التشابه والاختلاف. وتلك خطوة أساسية لتحليل الظاهرات الجغرافية وتعليلها. وهناك ثلاثة أنواع من المقارنات للوصول إلى هذه الغاية، وهذه الأنواع هي:

١- المقارنات الوصفية البحتة:

وهذه أبسط أنواع المقارنات، ويتميز الجغرافيون بصفة عامة بإجراء المقارنات الوصفية البسيطة في دراستهم، فكثير من الكتابات الجغرافية تتضمن مقارنة بين إقليمين مختلفين أو أكثر سواءً من نواحيها الطبيعية أو الحضارية، وقد ظهرت بعضًا من أعظم الكتابات الجغرافية تحت اسم «الجغرافيا الإقليمية المقارنة»، وسنتناول هذا النوع من المقارنة بعد قليل.

٢- المقارنات الاستنتاجية التفسيرية:

وهذا النوع من المقارنات هو الذي نجريه لكي نفسر بعض الخصائص التي سبق شرحها ثم استنتاج الحقائق الكامنة بعد ذلك، فعلى سبيل المثال إذا أجرينا دراسة عن مستوى أداء الطلاب في امتحان ما، فيتعين علينا أن نفسر أولًا لماذا حصل كل طالب على درجة معينة، ولن يتأتى ذلك إلا إذا أخذنا في الاعتبار كل العوامل التي قد نراها هامة وراء توزيع الدرجات التي حصل عليها الطلاب، وذلك هو الجزء الصعب من الدراسة ذلك لأن علينا آنذاك أن نفكر بعناية شديدة في كل العوامل الممكنة التي قد يكون لها تأثير في ذلك، وفي هذا المثال بالذات قد نرى أنه من بين العوامل: أ- مستوى الذكاء، ب – الخلفية الاجتماعية، جـ - ميل الطالب للعمل والدراسة. وكل من هذه العوامل الثلاثة ذو أثر فعال فيما يعرف بالمتغيرات التابعة، ذلك لأن أداء الطالب في الامتحان يعتمد عليها. وبعد ذلك علينا أن نختبر إلى أي مدى كانت هذه المتغيرات التابعة مرتبطة بالأداء في الامتحان، وهنا نقارن أداء كل طالب في الامتحان برتبته بالنسبة لكل من هذه المتغيرات التابعة، وعلى ذلك تقارن درجات الامتحان بما يلي:

أ- ترتيب الطالب حسب مستوى الذكاء.

(باستخدام اختبارات الذكاء).

ب - ترتيب الطالب حسب حالته الاجتماعية.

(باستخدام مقياس ما للدلالة على ذلك).

جـ- ترتيب الطالب حسب ميله للدراسة

(باستخدام مقياس ما للدلالة على ذلك).

وبعد أن نجري هذه المقارنات يصبح في مقدورنا أن نوضح كيف أثرت هذه المتغيرات التابعة، وبالتالي يمكن أن نستنتج ما هو العامل الأكثر أهمية في تفسير مستوى الأداء الذي تحقق في الامتحان.

وما نفعله في دراستنا الجغرافية يشبه ذلك إلى حد كبير، فنحن في حاجة مستمرة إلى تفسير الأنماط المختلفة التي ندرسها، فعلى سبيل المثال في دراستنا عن استخدام الأرض في منطقة الدراسة (إحدى المقاطعات البريطانية) رأينا أن الأرض الرعوية تتباين بدرجة ملحوظة بين قرى المنطقة (م = ١٤,٨%).

ومن الواضح أننا إذا أردنا إجراء دراسة ذات قيمة لتوجب علينا أن نسأل عن أسباب هذا الاختلاف في نسبة الأراضي الرعوية. وكما هي الحال في المثال السابق عن مستوى الأداء في الامتحان، فإننا نحتاج إلى فرز وتصنيف كل العوامل التي نراها هامة في أحداث هذه الاختلافات في نمط استغلال الأرض، كأن نقرر مثلًا أن العوامل التالية كانت من أكثر العوامل أهمية في تحديد مساحة الأرض الرعوية:

١- نوع التربة.

٢ – المناخ.

٣- ميول المزارعين واتجاهاتهم.

٤- ثراء المزارعين.

٥- المساعدة الحكومية.

٦- ارتفاع سطح الأرض.

وتمثل هذه المتغيرات الستة - متغيرات تابعة، وحتي يمكن أن نقف على ما إذا كانت مرتبطة بمساحة الأرض الرعوية ومدى هذا الارتباط، فعلينا أن نقارن مراعي كل قرية برتبتها (ترتيبها) حسب كل متغير من هذه المتغيرات الستة، فعلى سبيل المثال إذا كانت القرى التي بها مساحة كبيرة من الأرض الرعوية بها أيضا نسبة كبيرة من الأراضي المرتفعة، والقرى التي بها مساحة قليلة من الأرض الرعوية بها مساحة قليلة من الأراضي المرتفعة، ويكون ذلك مدعاة إلى الاستنتاج أن مساحة الأراضي العالية كانت عاملًا في تحديد نمط استخدام الأرض.

وتسمح لنا المقارنة بين المتغيرين أن نستنتج تفسيرًا للظاهرة، ومن ثم يعرف هذا النوع بالمقارنات الاستنتاجية التفسيرية.

٣ - أما النوع الثالث من المقارنات فهو في طبيعته تفسيري كذلك، ولكنه يتضمن بعض البيانات النظرية الافتراضية مقارنة بالبيانات الفعلية المتاحة، ويعرف هذا النوع بالمقارنات النظرية التفسيرية وسنشير إلى ذلك لاحقا.

١- المقارنات الوصفية البحتة:

لا ريب أن وصف مظاهر الاختلاف أو التشابه بين مجموعتي بيانات ليس أمرًا يسيرًا، ويبدو ذلك على سبيل المثال عندما نقارن مساحة المراعي في عينة من عشر قرى في إقليمين من الأقاليم، فنبدأ بوضع بيانات كل إقليم جنبًا إلى جنب مع الإقليم الآخر في جدول واحد كما يبدو من الجدول التالي وحتى يمكن أن نميز بين هاتين المجموعتين من البيانات، فسنشير إلى إحداهما بالمجموعة س والأخرى بالمجموعة ص.

جدول

حساب الوسط الحسابي والانحراف المعياري

(النسبة المئوية للأراضي الرعوية في ١٠ قرى في الإقليم س و١٠ قرى في الإقليم ص)

مجموعة البيانات س (إقليم س)			مجموعة البيانات ص (إقليم ص)		
رقم القرية	س	س٢	رقم القرية	ص	ص٢
١	١٥	٢٢٥	١	٦٠	٣٦٠٠
٢	٢٠	٤٠٠	٢	٥٣	٢٨٠٩
٣	٢٥	٦٢٥	٣	٤٧	٢٢٠٩
٤	٣٠	٩٠٠	٤	٤٢	١٧٦٤
٥	٣٥	١٢٢٥	٥	٥٩	٣٤٨١
٦	٤٠	١٦٠٠	٦	٦٢	٣٨٤٤
٧	٣٨	١٤٤٤	٧	٤٨	٢٣٠٤
٨	٣٣	١٠٨٩	٨	٥٦	٣١٣٦
٩	٢٤	٥٧٦	٩	٥٧	٣٢٤٩
١٠	٢٢	٤٨٤	١٠	٥٤	٢٩١٦

سَ = ٢٨,٢   مجـ س^{٢} = ٨٥٦٨   صَ = ٥٣,٨   مجـ ص^{٢} = ٢٩٣١٢

$$\text{ع س} = \sqrt{\frac{\text{مجـ س}^{٢}}{\text{ن}} - \text{سَ}^{٢}} \qquad \text{ع ص} = \sqrt{\frac{\text{مجـ ص}^{٢}}{\text{ن}} - \text{صَ}^{٢}}$$

$$= \sqrt{(\frac{٨٥٦٨}{١٠} - ٧٩٥,٢٤)} = ٧,٨٥ \qquad = \sqrt{\frac{٢٩٣١٢}{١٠} - ٢٨٩٤,٤٤} = ٦,٠٨$$

وبطبيعة الحال فإننا سنستهل حديثنا بمقارنة المساحة النموذجية للأرض الرعوية في الإقليم س، بمثيلتها في الإقليم ص، وكما سبق أن أشرنا من قبل أن هناك عدة طرق لقياس النموذج، لعل أكثرها استخدامًا هي (قيمة المتوسط الحسابي)، ومن ثم سنحسب أولًا متوسط عينة الأرض الرعوية في كل من الإقليمين: فالإقليم س، يبلغ متوسط مساحة الأرض الرعوية به ٢٨,٢٪، بينما يبلغ هذا المتوسط في الإقليم ص ٥٣,٨٪، كذلك يمكن أن نحسب الانحراف المعياري في كلا المجموعتين: فعلى سبيل المثال يبلغ الانحراف المعياري للإقليم س (أي ع س) = ٧,٨٥، وللإقليم ص (أي ع ص) = ٦,٠٨، ولا جدال في أننا سنبادر إلى القول بأن الإقليم س مختلف عن الإقليم ص، ذلك لأن المتوسطات والانحرافات المعمارية لكل من الإقليمين مختلفة عن بعضها البعض.

وينبغي أن نكون على حذر عندما نصف مجموعتين للبيانات معًا، ذلك لأنه قد يظهر على سبيل المثال أن متوسط العينة في كلا المجموعتين مختلف تمامًا، ومن ثم يكون هناك فرق بين الإقليم س والاقليم ص، ولكن هذا الفرق قد يكون ظاهريًّا وليس حقيقيًّا، وعلى ذلك نبدأ في التساؤل الهام وهو كم حجم الفرق الذي يجب أن يكون بين مجموعتي البيانات حتى يكون هناك فرق معنوي بينهما؟ .

وعلى ذلك فإن المقارنة بالمتوسط الحسابي والانحراف المعياري كما سبق تعد أبسط أشكال المقارنة، فهو مجرد وصف لأوجه التشابه والاختلاف بين مجموعتين أو أكثر من البيانات التي ترتبط بنفس الموضوع، وفي مثالنا السابق كان الموضوع هو (نسبة الأرض الرعوية في الإقليمين). ومهما يكن من أمر فإن أهم نقطة هي أن هذا النوع من المقارنة لمثالين من موضوع واحد ليس تفسيرًا لأي شيء؛ ولذا يسمى هذا النوع من المقارنة بالوصف المجرد.

٢ - المقارنات الانتاجية التفسيرية:

سبق القول أن المقارنة بين المتغيرات التابعة والمستقلة قد تؤدي على الأرجح إلى استنتاج السبب والنتيجة Cause and effect ففي دراستنا عن استخدام الأرض - على سبيل المثال - رأينا أن مساحة الأرض الرعوية في قرية ما قد تكون مرتبطة بمساحة الأرض المرتفعة (أي الأرض التي يزيد منسوبها على ٥٠٠ متر) في هذه القرية. فكيف نحدد مدى هذا الارتباط ومقداره؟

أولًا علينا أن نرى - بعد وضع مجموعتي البيانات جنبًا إلى جنب - كيف تختلف كل منهما عن الأخرى، ويمكن أن نستنتج من الوهلة الأولى وبصفة عامة - أن مساحة الأرض الرعوية تتزايد بتزايد نسبة الأرض المرتفعة، وبطريقة نستطيع معها أن نقول بأن هناك درجة ما من الارتباط بين مجموعتي البيانات، إلا أن الارتباط ليس كاملًا، ذلك لأنه إذا كان كذلك فإن النسب ستتمشى مع بعضها البعض زيادةً أو نقصانًا، وبمعنى آخر أن كل قرية

تكون فيها نسبة الأرض المرتفعة أعلى من المتوسط ستكون فيها نسبة الأرض الرعوية أعلى من المتوسط أيضًا، والعكس. ولكن الواقع كما توضحه البيانات ليس على هذه الصورة إطلاقًا، فكيف - والحال كذلك - يمكننا أن نقيس بدقة درجة وشكل الارتباط بين مجموعتي البيانات؟

هناك ثلاث طرق رئيسية لكي تظهر مثل هذه الارتباطات إحصائيًّا:

أولًا: أن تقوم بحساب التغاير Co - Variance في مجموعة البيانات.

ثانيًا: أن نقوم بحساب ما يعرف بمعامل الارتباط Correlation.

ثالثًا: أنه يمكننا أن نوضح شكل الارتباط بيانيًّا برسم ما يعرف بخطوط الانحدار Regression Lines.

أ- التغاير: Co-Variance

سبق أن ذكرنا منذ قليل في حديثنا عن استخدام الأرض أن الارتباط بين الأرض الرعوية والأرض المرتفعة ليس ارتباطًا كاملًا، ذلك لأن كل القرى التي كانت نسبة المراعي بها أكثر من المتوسط لم تكن هي القرى التي كانت نسبة الأرض المرتفعة بها أكثر من المتوسط كذلك، وطريقة مقارنة هاتين المجموعتين من البيانات هي في الواقع الطريقة التي تشكل الأساس لحساب التغاير إحصائيًّا، فنستطيع أن نقيس المدى الذي تتغاير (تختلف) فيه مجموعة البيانات بالنسبة لقيم الوسط بها وانحرافاتها المعيارية.

والخطوة الأولى ستكون حساب قيمة المتوسط الحسابي لكل مجموعة من مجموعات البيانات، ثم نقيس الانحراف لكل قرية عن متوسط كل مجموعة، ثم نحاول بعد ذلك تحديد أي العوامل هو التابع وايها هو المستقل. ومن الواضح في مثالنا هذا أن نسبة الأرض الرعوية هي المتغير التابع. ومع فرض صحة ما نقول فهو تابع للأرض المرتفعة، وحتي يمكننا التمييز بين مجموعتي البيانات فسنشير إلى مجموعة البيانات التي نعتبرها تابعة بالرمز ص، أما المجموعة التي نعتبرها مستقلة فسنشير إليها بالرمز س [من القواعد المتعارف عليها إحصائيا كما أشرنا أنه في التمثيل البياني نوقع المتغير التابع على المحور الصادي، بينما نوقع المتغير المستقل على المحور السيني، ومن ثم يعرف المتغير التابع ومجموعة البيانات الدالة عليه بمجموعة البيانات ص].

ويمكن حساب المتوسط الحسابي لمجموعة البيانات ص (نسبة الأرض الرعوية) ثم نحسب انحراف كل قرية عن هذا المتوسط (ص – صَ)، ثم نضعه في جدول آخر ، وبطريقة مماثلة نحسب س، وقيم (س – سَ)، ثم نضعها في نفس الجدول.

وسنجد بطبيعة الحال أن بعض القرى أكثر من المتوسط في مجموعة البيانات (أي أن كلًّا من س – سَ و ص – صَ موجبتان)، بينما البعض الآخر أقل من المتوسط (أي أن

كلًّا من س – سَ و ص – صَ سالبتان). كذلك سنجد أن هناك بعض القرى (مثل القريتين رقم ٥ ، ٧ في الجدول) تزيد عن المتوسط في ناحية، وتقل عن المتوسط في ناحية أخرى، ولكن إذا استمررنا في حصر كل القائمة لمعرفة وضع كل قرية على حدة فسيطول بنا الوقت، ولن نستطيع أن تستنتج على الفور إلى أي مدي ترتبط المتغيرات التابعة والمستقلة.

جدول

نسبة الأرض المرتفعة الرعوية

(١٢ قرية في إحدى المقاطعات)

رقم القرية	مجموعة البيانات (س) الأراضي التي تزيد على ٥٠٠ مترًا %	مجموعة البيانات (ص) الأراضي الرعوية %
١	٢٦	٣٧
٢	٣٠	٤٦
٣	٣٣	٤٨
٤	٣٦	٤٩
٥	٥٤	٥٠
٦	٤٦	٥٤
٧	٥٢	٥٦
٨	٥٨	٦٢
٩	٦٥	٦٥
١٠	٦٨	٦٧
١١	٧٠	٧٤
١٢	٧٤	٧٩
	س = ٥١	ص = ٥٧

جدول

حساب التغاير

(اعتمادًا على البيانات الواردة في الجدول السابق)

رقم القرية	الانحراف مجموعة البيانات س (س – سَ)	الانحراف مجموعة البيانات ص (ص – صَ)	حاصل ضرب (س – سَ) × (ص – صَ)
١	-٢٥	-٢٠	٥٠٠
٢	-٢١	-١٤	٢٩٤
٣	-١٨	-٩	١٦٢
٤	-١٥	-٨	١٢٠
٥	٣	-٧	-٢١
٦	-٥	-٣	١٥
٧	١	-١	-١
٨	٧	٥	٣٥
٩	١٤	٨	١١٢
١٠	١٧	١٠	١٧٠
١١	١٩	١٧	٣٢٣
١٢	٢٣	٢٢	٥٠٦

مجـ (س – سَ) (ص – صَ) = ٢٢١٥

$$\text{مجـ} = \frac{\text{(س – سَ) (ص – صَ)}}{\text{ن}} = \frac{\text{٢٢١٥}}{\text{١٢}} = \text{١٨٤,٥٨}$$

وحينئذ نجد أننا في حاجة إلى ضغط بياناتنا مرة أخرى حتى يمكننا أن نحسب رقمًا مفردًا واحدًا يمكن أن يصف كلًّا من مدي واتجاه انحرافات كل القرى مجتمعة.

وتحديد اتجاه الانحراف أمر من السهل توضيحه، فقد سبق أن قلنا من قبل أن ضرب العلامات الجبرية يتضمن تغييرًا في هذه العلامات، وعلى ذلك فإن +×+=+، -×-=+ و+x-=- و-x+ =-. وإذا استخدمنا هذا المفهوم في محاولتنا لمعرفة إلى أي حد تتباين هاتان المجموعتان مع بعضهما البعض (أي لمعرفة التغاير)، فيمكننا أن نعتبر أن كل القرى التي تكون علاماتها بالزائد تكون أرقامها دالة على وجود علاقة، وأن كل القرى التي تكون أرقامها عكس ذلك (واحدة أعلى من المتوسط وواحدة أقل المتوسط) فإن علاماتها ستكون بالناقص. ومعنى ذلك فإن كل ما نحتاج إليه لكي نعبر عن علاقة انحراف كل زوجين من

المشاهدات (المراعي والمرتفعات) هو أن نضرب الانحراف في المجموعة س بالانحراف في المجموعة ص - آخذين العلامات الجبرية في الاعتبار. فعلى سبيل المثال في الجدول يمكن ملاحظة أن انحراف القرية الخامسة هو – ٢١، مما يدل على أنها أعلى من المتوسط في ناحية وأقل من المتوسط في ناحية أخرى، بينما يصل انحراف القرية التاسعة إلى +١١٢، مما يدل على أن كلًّا من المشاهدتين (المراعي والمرتفعات)، إما أن يكونا أعلى من المتوسط أو أقل من المتوسط.

وإذا واصلنا العمل بهذه الطريقة وجمعنا كل هذه الانحرافات وطرحنا عندما تكون العلامة – ، فإننا سنحصل على رقم يبين مجموع حاصل ضرب الانحرافات (مجـ (س – سَ)(ص – صَ))، وإذا قسمنا الناتج على عدد العناصر (ن) فإننا نحصل على متوسط مجموع حاصل ضرب الانحرافات، [أي مجـ (س – سَ)(ص – صَ)/ ن]، حيث ن = عدد أزواج المشاهدات (أي نسبة المراعي والأرض المرتفعة في كل قرية)، وتعرف هذه القيمة بالتغاير، أي تغاير مجموعتي البيانات (أي التباين المترابط في الواقع)، وفي المثال المذكور فإن قيمة التغاير هي +١٨٤,٥٨. فماذا يدل عليه هذا الرقم؟

الواقع أن هذا المقياس يوضح لنا بعضًا من العلاقة بين مجموعتي البيانات معًا، وقد تكون قيمة التغاير هذا - قيمة موجبة أو قيمة سالبة، ويعتمد ذلك على ما إذا كانت مجموعتا البيانات اللتين نقارنهما ترتبطان في نفس الاتجاه أم لا.

فإذا كانا يرتبطان في نفس الاتجاه كما في الجدول [أي أن أية زيادة في الأرض الرعوية ترتبط بأية زيادة في الارتفاع] فستكون النتيجة التي سنحصل عليها موجبة كما في هذا المثال. ولكن إذا كانت مجموعتا البيانات ترتبطان معًا في اتجاه عكسي فإن القيم ستكون سالبة حينذاك، ومثال ذلك أننا كنا نتعامل مع النسبة المئوية للأرض الزراعية مقارنة بالارتفاع ، فإنه يمكن أن نرى أن القيم في المجموعة ص تتناقص كلما تزايدت القيم في المجموعة س. وتكون النتيجة أن التغاير الناتج يكون سالبا (-١٨٤,٥٨) كما يبدو من (الجدول) وهذه في الواقع نقطة هامة سنعود إليها بعد قليل في حديثنا عن الارتباط، فإذا كانت العلاقة بين المجموعة س والمجموعة ص تأخذ نفس الاتجاه، فإن التغاير سيكون بالزائد (+)، وإذا كانت العلاقة بين المجموعة س والمجموعة ص تأخذ اتجاهًا عكسيًّا فإن التغاير سيكون بالناقص (-).

ب - الارتباط correlation

١- معامل بيرسون للارتباط Pearson Correlation:

سبق أن تناولنا في حديثنا عن قياس النماذج (المتوسطات) أن أية مجموعة من البيانات يمكن حساب متوسطها الحسابي وانحرافها المعياري (ع)، ويفيدنا ذلك عندما نقارن هذه المجموعة بمجموعة أخرى، فإذا أشرنا إلى المجموعة الأولى بالرمز س، والمجموعة

الثانية بالرمز ص، فإن الانحراف المعياري للمجموعة س سيكون (ع $_{س}$)، والانحراف المعياري للمجموعة ص سيكون (ع $_{ص}$)، وإذا ضربنا كلًّا منهما في الآخر (ع $_{ص}$ ع $_{س}$)، نحصل على مقياس آخر لانحرافات مجموعتي البيانات معًا يعرف بمقياس التغاير.

وعلى ذلك فعندما نريد معرفة ما إذا كانت مجموعتا البيانات س، ص مرتبطتين معًا فهناك نوعان من الانحرافات يمكن حسابهما للوصول لهذا الغرض.

جدول

حساب التغاير

(نسبة الأرض الزراعية والارتفاع في ١٢ قرية بإحدى المقاطعات)

رقم	مجموعة البيانات س		مجموعة البيانات ص		(س – سَ) × (ص – صَ)
	الأرض التي تزيد على ٥٠٠ متر%	(س – سَ)	الأرض الزراعية %	(ص – صَ)	
١	٢٦	– ٢٥	٦٣	٢٠	– ٥٠٠
٢	٣٠	– ٢١	٥٧	١٤	– ٢٩٤
٣	٣٣	– ١٨	٥٢	٩	– ١٦٢
٤	٣٦	– ١٥	٥١	٨	– ١٢٠
٥	٥٤	٣	٥٠	٧	٢١
٦	٤٦	– ٥	٤٦	٣	– ١٥
٧	٥٢	١	٤٤	١	١
٨	٥٨	٧	٣٨	– ٥	– ٥٣
٩	٦٥	١٤	٣٥	– ٨	– ١١٢
١٠	٦٨	١٧	٣٣	– ١٠	– ١٧٠
١١	٧٠	١٩	٢٦	– ١٧	– ٣٢٣
١٢	٧٤	٢٣	٢١	– ٢٢	– ٥٠٦

سَ = ٥١، صَ = ٤٣، مجـ (س – سَ)(ص – صَ) = – ٢٢١٥

مجـ $\frac{\text{(س – سَ) (ص – صَ)}}{\text{ن}}$ = – ١٨٤,٥٨

أولهما: التغاير وهو الذي يقيس انحراف مجموعتي البيانات معًا، وفي نفس الوقت كما سبق أن رأينا، وثانيهما: الانحراف المعياري وهو الذي يقيس الانحرافات بصرف النظر عن بعضهما البعض، وإذا قارنا هذين المقياسين معًا (أي نسبتا أحدهما إلى الآخر)، فسنعرف إلى أي حد ترتبط المجموعتان مع بعضهما البعض.

فإذا كان للتغاير والانحراف المعياري قيمة مشابهة فمعنى ذلك أن هناك درجة عالية من التشابه بين المجموعتين، وإذا لم يكن هذان المقياسان متشابهين فإن الارتباط سيكون أقل. وطالما أننا سنعتبر أحد المقياسين نسبة من الأخر، فسنصل بذلك إلى قيمة بسيطة تدل على مدى الترابط بينهما تعرف بمعامل الارتباط، ويرمز لها بالحرف (ر) [يعرف أحيانا بمعامل ارتباط بيرسون نسبة إلى الإحصائي الذي توصل إليه]، ويأخذ الصيغة التالية:

$$\text{ر} = \frac{\text{تغاير مجموعة البيانات س، ص مأخوذًا كأزواج من القيم}}{\text{الانحراف المعياري لمجموعة البيانات س مضروبًا في الانحراف المعياري لمجموعة البيانات ص}}$$

أو بالصيغة الإحصائية المختزلة: $\text{ر} = \frac{\text{مجـ}(\text{س} - \bar{\text{س}})(\text{ص} - \bar{\text{ص}})/\text{ن}}{\text{ع}_{\text{س}}\ \text{ع}_{\text{ص}}}$

ويقع معامل الارتباط (ر) دائمًا في المدى +١ إلى -١، وعلامة الموجب (+) والسالب (-)، توضح ما إذا كانت العلاقة طردية أي في نفس الاتجاه، وبالتالي تكون موجبة أو علاقة عكسية تسير في الاتجاه العكسي، وبالتالي تكون سالبة، ومن هنا فإن الارتباط الكامل يدل عليه +١ أو -١، وكلما قلت قيمة (ر) يصبح الارتباط أقل كمالًا، حتى إذا أصبحت القيمة صفرًا، كان معنى ذلك أنه لا يوجد ارتباط على الإطلاق بين هاتين الظاهرتين المدروستين، ومعني ذلك أنه كلما كانت قيمة (ر) اقرب إلى (+١) أو (-١) كلما كان الارتباط أكثر اكتمالاً سواءً بالإيجاب أو بالسلب.

وبصفة عامة فإن وصف الارتباط أو مستوياته يمكن أن يتحدد في ضوء القيم الحرجة التالية:

قيمة معامل الارتباط	مدى الحكم عليه
من ± ٠,٧ إلى +١,٠	درجة ارتباط عالية وقوية.
من ± ٠,٤ إلى ± ٠,٧	درجة ارتباط جوهرية (حقيقية).
من ± ٠,٢ إلى ± ٠,٤	درجة ارتباط منخفضة وضعيفة.
أقل من ± ٠,٢	درجة ارتباط ضعيفة للغاية أو منعدمة.

وباختصار فإن الباحث الجغرافي يلزمه في كثير من دراساته عن تحليل العلاقات بين الظاهرات المختلفة، أن يقف على درجة العلاقة بينها - أي مدى التغير الذي تحدثه ظاهرة ما - سلبًا أو إيجابًا في ظاهرة أخرى - وتعرف هذه العلاقة حينئذ بالارتباط، ووجودها يعني أنه إذا تغيرت إحدى الظاهرتين فإن الظاهرة الأخرى (أي المتغير الآخر) يميل إلى التغير في نفس الاتجاه أو الاتجاه العكسي، فإذا حدث التغير في الظاهرتين في نفس الاتجاه فإن الارتباط يكون موجبًا، أي أنه إذا زادت قيم أحد المتغيرين فإن قيمة المتغير الثاني تميل إلى الزيادة أيضًا بصفة عامة، وإذا تناقصت هذه القيم فإن قيمة المتغير الآخر تميل إلى التناقص هي الأخرى بوجه عام، أما إذا كان التغير في الظاهرتين في اتجاه عكسي،

فإن الارتباط يكون سالبًا بمعنى أنه إذا زادت إحدى قيم أحد المتغيرين، فإن قيمة المتغير الثاني تتناقص بصفة عامة، والعكس. ومقياس الحكم على هذه العلاقة هو معامل الارتباط.

ونأتي الآن إلى المثالين اللذين نتعامل معهما لنرى كيف نطبق ذلك. ونبدأ بمثال «الأرض الرعوية»، وقد سبق أن حسبنا بالفعل مقياس التغاير وقيمته + ١٨٤,٥٨ ، ويمكن حساب الانحراف المعياري بالطريقة التي ذكرت في سياق الحديث عن الانحراف، وحساب هذه القيم لمجموعتي البيانات س، ص كالتالي:

$$ع_{س} = ١٦,٠١$$

$$ع_{ص} = ١٢,١٢$$

وبالتعويض في المعادلة:

$$ر = \frac{مجـ (س - \bar{س})(ص - \bar{ص})/ن}{ع_{س}\ ع_{ص}} = \frac{١٨٤,٥٨}{١٦,٠١ \times ١٢,١٢} = + ٠,٩٥$$

ويعني ذلك أن هناك درجة عالية من الارتباط بين نسبة الأرض الرعوية والأرض التي يزيد منسوبها على ٥٠٠ متر، وتتمشى نسبة الأرض الرعوية مع الارتفاع زيادةً أو نقصانًا.

ومن ناحية أخرى وجدنا في مثال الأرض الزراعية أن التغاير هو – ١٨٤,٥٨، وباستخدام الانحراف المعياري الذي سبق حسابه وقدره:

$$ع_{س} = ١٦,٠١$$

$$ع_{ص} = ١٢,١٢$$

نجد أن :

$$ر = \frac{- ١٨٤,٥٨}{١٦,٠١ \times ١٢,١٢} = - ٠,٩٥$$

ويعني ذلك أن هناك درجة عالية من الارتباط بين مقدار الأرض الزراعية والارتفاع الذي يزيد على ٥٠٠ متر، ولكن كلما تزايد الارتفاع تناقصت نسبة الأرض الزراعية.

مثال :

نفرض أننا نريد قياس درجة الارتباط بين إنتاجية الفدان من القمح والفول في إحدى المحافظات على امتداد عشر سنوات (١٩٦١ - ١٩٧٠) كما تبين البيانات التالية:

(متوسط إنتاج الفدان بالأردب سنويا)

السنة	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	المتوسط
القمح	٥,٠	٥,٢	٦,٨	٥,٠	٤,٨	٥,٦	٥,٤	٥,٨	٥,٦	٥,٨	٥,٥
الفول	٤,٢	٣,٤	٧,٠	٤,٢	٣,٨	٤,٤	٥,٢	٤,٢	٥,٢	٥,٢	٤,٧

فمن الملاحظ أن الإنتاج يتباين من سنة لأخرى بالنسبة للمحصولين، ولكن التغيرات ليست متشابهة دائمًا لكليهما، كذلك فإن انحرافات القيم السنوية عن المتوسط الحسابي تختلف لكلا المحصولين.

ويبين ذلك الجدول الذي وضع لحساب معامل الارتباط بين المحصولين، وتبدو انحرافات القيم عن المتوسط في العمودين الرابع (س - سَ)، والخامس (ص - صَ)، وبتطبيق صيغة معامل الارتباط لقياس العلاقة بين التغير في قيم س (القمح)، والتغير في قيم ص (الفول)، وأفضل طريقة لقياس هذا التغير هي إيجاد الفرق بين قيم كل متغير ومتوسطه الحسابي - ثم نحصل على تربيع الانحراف وحاصل الجمع لكل منهما، ونوجد الانحراف المعياري لقيم س وكذلك لقيم ص - ونضرب كل انحراف لقيم س في الانحراف المناظر لقيم ص ونوجد حاصل الجمع، وبقسمة حاصل الجمع هذا على حاصل ضرب الانحراف المعياري لكل من المتغيرين نحصل على معامل الارتباط، ويبين الجدول خطوات الحساب.

ويبدو أن معامل الارتباط بين المحصولين قد وصل إلى + ٠,٨٧، ويدل ذلك على أن هناك درجة عالية من الارتباط الموجب بينهما، فحينما تزداد قيم الإنتاج في محصول منهما، فإن هناك نزعة قوية لتزايد هذه القيم في المحصول الآخر، رغم أن التزايد ليس مطلقًا في كميته أو في توزيعه السنوي.

وينبغي الإشارة هنا إلى أن معاملات الارتباط التي تتراوح بين + ٠,٥ و +١,٠ من ناحية و – ٠,٥ و – ١,٠ من ناحية أخرى تعد معاملات جوهرية (أي تكون ذات دلالة للعلاقة الموجبة أو السالبة - وتزداد الدلالة قوة بتزايد المعامل إيجابًا وبتناقصه سلبًا، أما إذا كانت المعاملات واقعة بين - ٠,٥، + ٠,٥ فإن الارتباط لا يكون جوهريًّا بدرجة مقبولة، ومن ثم تقل دلالته سلبًا وإيجابًا، وبديهي أن معامل الارتباط إذا كان يساوي صفرًا، فإن ذلك يدل على أن كلًّا من القيمتين يتغير دون أي ارتباط بالآخر ومن ثم فلا يوجد بينهما أي ارتباط في هذه الحالة.

جدول

حساب معامل الارتباط بين التغير في إنتاج القمح والفول

في إحدى المحافظات في الفترة (١٩٦١ - ١٩٧٠)

(متوسط إنتاجية الفدان بالإردب)

السنوات	القمح س	الفول ص	(س – سَ)	(ص – صَ)	(س – سَ)^{٢}	(ص – صَ)^{٢}	(س – سَ) × (ص – صَ)
١	٥,٠	٤,٢	– ٠,٥	– ٠,٥	٠,٢٥	٠,٢٥	+ ٠,٢٥
٢	٥,٢	٣,٤	– ٠,٣	– ١,٣	٠,٠٩	١,٦٩	+ ٠,٣٩
٣	٦,٨	٧,٠	+ ١,٣	+ ٢,٣	١,٦٩	٥,٢٩	+ ٢,٩٩
٤	٥,٠	٤,٢	– ٠,٥	– ٠,٥	٠,٢٥	٠,٢٥	+ ٠,٢٥
٥	٤,٨	٣,٨	– ٠,٧	– ٠,٩	٠,٤٩	٠,٨١	+ ٠,٦٣
٦	٥,٦	٤,٤	+ ٠,١	– ٠,٣	٠,٠١	٠,٠٩	– ٠,٠٣
٧	٥,٤	٥,٢	– ٠,١	+ ٠,٥	٠,٠١	٠,٢٥	– ٠,٠٥
٨	٥,٨	٤,٤	+ ٠,٣	– ٠,٣	٠,٠٩	٠,٠٩	– ٠,٠٩
٩	٥,٦	٥,٢	+ ٠,٣	+ ٠,٥	٠,٠٩	٠,٢٥	+ ٠,١٥
١٠	٥,٨	٥,٢	+ ٠,٣	+ ٠,٥	٠,٠٩	٠,٢٥	+ ٠,١٥
	٥٥,٠	٤٧,٠	–	–	٣,٠٦	٩,٢٢	+ ٤,٦٤

المتوسط س = ٥,٥، ص = ٤,٧.

وينبغي أن يكون الباحث على حذر في استخدامه لمعاملات الارتباط، ذلك لأن قيمة الارتباط الموجب (+ ٠,٨٧ كما في الجدول السابق مثلًا) لا تتضمن إجابة لسؤال هو لماذا توجد هذه العلاقة؟ كما لا تدل في الواقع على أن نفس الأسباب تؤدي إلى نفس النتائج، ذلك لأن هناك عوامل عديدة أخرى أدت إلى التغيرات الملاحظة في إنتاجية الفدان من المحصولين، أما كل ما يعنيه معامل الارتباط فهو وجود درجة من العلاقة الإحصائية بين القيم المبينة، أما العوامل الأخرى فتتطلب تفسيرات وتحليلات أخرى.

الانحراف المعياري لقيم س = $\text{ع}_{\text{س}} = \sqrt{\frac{٣,٠٦}{١٠}} = ٠,٥٥$

الانحراف المعياري لقيم ص = $\text{ع}_{\text{ص}} = \sqrt{\frac{٩,٢٢}{١٠}} = ٠,٩٦$

وبتطبيق قانون الارتباط $\text{ر} = \frac{\frac{١}{\text{ن}}\ \text{مجـ}(\text{س} - \bar{\text{س}})(\text{ص} - \bar{\text{ص}})}{\text{ع}_{\text{س}} \times \text{ع}_{\text{ص}}}$

$$= \frac{\frac{١}{١٠}\,(+\,٤,٦٤)}{٠,٥٥ \times ٠,٩٦} = \frac{٠,٤٦٤}{٠,٥٢٨} = +\,٠,٨٧$$

مثال :

يبين الجدول بيانات عن نسبة الأمية بين سكان المحافظات المصرية ونسبة الأطفال دون سن الثانية عشرة. وإذا أردنا معرفة درجة الارتباط بين هاتين الظاهرتين، فإننا ننشئ جدولًا مشابهًا، ونبدأ في تحديد انحرافات الأرقام المعطاة في كل عمود عن متوسطها كما هو موضح بالجدول، ونستمر للحصول على مربع هذه الفوارق (لإزالة علامات السالب)، ونحسب الانحراف المعياري ثم نطبق قانون الارتباط بعد ذلك، وأظهرت العمليات الإحصائية في هذه الحالة أن معامل الارتباط بين نسبة الأمية ونسبة الأطفال دون سن الثانية عشرة (مؤشر للخصوبة) يصل إلى + ٠,٨٣ وهي تدل على ارتباط موجب قوي بين هاتين الظاهرتين.

جدول

حساب معامل الارتباط بين الأمية ونسبة الأطفال الصغار في محافظات مصر
(باستثناء المحافظات الصحراوية) سنة ١٩٧٦
(النسبة المئوية للأمية بين السكان ١٠ سنوات فأكثر، ونسبة الأطفال أقل من ١٢ سنة إلى جملة السكان)

المحافظة	الأمية % (س)	الأطفال % (ص)	(س – سَ)	(ص – صَ)	(س – سَ)²	(ص – صَ)²	(س – سَ) × (ص – صَ)
القاهرة	٣٤,٦	٢٧,٣	– ٢٢,٩	– ٤,٥	٥٢٤,٤١	٢٠,٢٥	+ ١٠٣,٠٥
الإسكندرية	٣٧,٤	٢٧,٧	– ٢٠,١	– ٤,١	٤٠٤,٠١	١٦,٨١	+ ٨٢,٤١
بورسعيد	٣٥,٩	٢٤,٥	– ٢١,٦	– ٧,٣	٤٦٦,٥٦	٥٣,٢٩	+ ١٥٧,٦٨
السويس	٤٤,٤	٣٠,٠	– ١٣,١	– ١,٨	١٧١,٦١	٣,٢٤	+ ٢٣,٥٨
دمياط	٤٩,٤	٣٢,٧	– ٨,١	+ ٠,٩	٦٥,٦١	٠,٨١	– ٧,٢٩
الدقهلية	٥٦,٣	٣١,٩	– ١,٢	+ ٠,١	١,٤٤	٠,٠١	– ٠,١٢
الشرقية	٦٢,٦	٣٣,٤	+ ٥,١	+ ١,٦	٢٦,٠١	٢,٥٦	+ ٨,١٦
القليوبية	٥٣,٧	٣٣,٥	– ٣,٨	+ ١,٧	١٤,٤٤	٢,٨٩	– ٦,٤٦
كفر الشيخ	٧٠,١	٣٣,٣	+ ١٢,٦	+ ١,٥	١٥٨,٧٦	٢,٢٥	+ ١٨,٩
الغربية	٥٤,٩	٣٠,٤	– ٢,٦	– ١,٤	٦,٧٦	١,٩٦	+ ٣,٦٤
المنوفية	٥٦,٩	٣١,٦	– ٠,٦	– ٠,٢	٠,٣٦	٠,٠٤	+ ٠,١٢
البحيرة	٦٦,٢	٣٣,٧	+ ٨,٧	+ ١,٩	٧٦,٦٩	٣,٦١	+ ١٦,٥٣
الإسماعيلية	٥٠,٨	٣٢,١	– ٦,٧	+ ٠,٣	٤٤,٨٩	٠,٠٩	– ٢,٠١
الجيزة	٥٣,٥	٣٢,٦	– ٤,٠	+ ٠,٨	١٦,٠٠	٠,٦٤	– ١,٢٠
بني سويف	٦٨,٤	٣٣.٢	+ ١٠,٩	+ ١,٤	١١٨,٨١	١,٩٦	+ ١٥,٢٦

الفيوم	٨٣,٦	٣٥,٢	+ ١٦,١	+ ٣,٤	٢٥٩,٢١	١١,٥٦	+ ٥٤,٧٤
المنيا	٧٠,٩	٣٣,٠	+ ١٣,٤	+ ١,٢	١٧٣,٥٦	١,٤٤	+ ١٦,٠٨
أسيوط	٦٨,٥	٣٣,٦	+ ١١,٠	+ ١,٨	١٢١,٠٠	٣,٢٤	+ ١٩,٨
سوهاج	٧٢,٨	٣٣,١	+ ١٥,٣	+ ١,٣	٢٣٤,٠٩	١,٦٩	+ ١٩,٨٩
قنا	٧١,٢	٣٢,٥	+ ١٣,٧	+ ٠,٧	١٨٧,٦٩	٠,٤٩	+ ٩,٥٩
أسوان	٥٦,٠	٣٢,٩	– ١,٥	+ ١,١	٢,٢٥	١,٢١	– ١,٦٥
الجملة	١٢٠٨,١	٦٦٨,٢	–	–	٣٠٧٩,١٦	١٣٠,٠٤	+ ٥٢٨,٧

المتوسط: سَ = ٥٧,٥ ، صَ = ٣١,٨

الانحراف المعياري لقيم س = ع $_{\text{س}}$ = $\sqrt{\frac{\text{٣٠٧٩,١٦}}{\text{٢١}}}$ = $\sqrt{\text{١٤٦,٦٣}}$ = ١٢,١١

الانحراف المعيار لقيم ص = ع $_{\text{ص}}$ = $\sqrt{\frac{\text{١٣٠,٠٤}}{\text{٢١}}}$ = $\sqrt{\text{٦,١٩}}$ = ٢,٤٩

وبتطبيق قانون الارتباط:

$$\text{ر} = \frac{\text{مجـ (س – سَ)(ص – صَ)/ن}}{\text{ع}_{\text{س}}\ \text{ع}_{\text{ص}}} = \frac{\text{٢١ / ٥٢٨,٧}}{\text{٢,٤٩} \times \text{١٢,١١}} = \frac{\text{٢٥,١٨}}{\text{٣٠,١٥}} = + \text{٠,٨٣}$$

٢- معامل سبيرمان لارتباط الرتب:

The Spearman rank correlation coefficient

رغم سهولة حساب معامل بيرسون للارتباط إلا أنه طويل نوعًا ما، ولا يمكن حسابه إلا لبيانات تتكون من قيم مطلقة فقط (مثل مقدار الأرض المرتفعة)، ، وغالبًا ما لا يكون لدينا وقت كافٍ لحساب قيم (ر) بالطريقة التي شرحناها، كذلك لا تتاح عناصر لنا دائمًا ببيانات على شكل قيم مطلقة، ولكنها تكون بشكل آخر كأن تكون على هيئة عناصر في قائمة ما مرتبة ترتيبًا نسبيًّا، وتعرف مثل هذه البيانات بلغة الإحصاء بمجموعة بيانات الرتب، وتتطلب مثل هذه البيانات طرقًا إحصائيةً مختلفةً لحساب الارتباط بينها، لعل أكثرها استخدامًا وشهرةً ما يعرف بمعامل سبيرمان لارتباط الرتب. والذي يعتمد غالبًا على استخدام الرتب وليس القيم المطلقة الفعلية، وحتى يمكن أن نميز بينه وبين معامل ارتباط بيرسون (ر) فإننا نضيف حرف س (من سبيرمان) بجوار (ر)، وعلى ذلك فإن الرمز ر $_{\text{س}}$ يشير إلى معامل سبيرمان لارتباط الرتب. والأساس في حساب ر $_{\text{س}}$ هو مقارنة فروق الرتب في كل عنصر من عناصر مجموعتي البيانات، وكنقطة بداية فإننا سنكون في حاجة إلى وضع مجموعتي البيانات جنبًا إلى جنب حتى يمكن أن نحسب الفوارق بين الرتب، ونشير إلى الفرق في الرتب بقيمة (ف)، ثم نطبق بعد ذلك قانون سبيرمان لارتباط الرتب على النحو التالي:

$$\text{ر}_{\text{س}} = \text{١} - \frac{\text{٦ مجـ ف}^{\text{٢}}}{\text{ن}^{\text{٣}} - \text{ن}}$$

حيث:

$ف^2$ = الفرق بين رتبتي كل قيمتين متناظرتين.

ن = عدد أزواج الرتب.

ويعطي هذا القانون قيمة تقريبية لمعامل الارتباط، ولكنها تتميز بالسهولة والسرعة، ويسمى المعامل الناتج بمعامل ارتباط الرتب.

فإذا كان لدينا مثلًا خمسة أقاليم صناعية يختلف مركزها في الإنتاج الصناعي بعناصره المختلفة خاصة في صناعتين هما:

الصناعات الهندسية عمومًا، وصناعة السيارات على وجه الخصوص، على النحو الذي يبين رتبة كل إقليم في هاتين الصناعتين.

الإقليم	(أ)	(ب)	(جـ)	(د)	(هـ)
رتب الصناعات الهندسية	١	٢	٣	٤	٥
رتب صناعة السيارات	٣	٢	١	٥	٤

فنلاحظ أن هذه الأقاليم الخمسة تختلف في ترتيب الصناعتين بها. فالأقاليم الثلاثة الأولى مثلًا في الصناعات الهندسية ليست كذلك في صناعة السيارات، فالإقليم الأول في الصناعات الهندسية يحظى بالمركز الثالث في صناعة السيارات، والإقليم الثالث يحظى بالمركز الأول على الترتيب، أما الإقليمان الأخيران فهما أقل أهمية بين الأقاليم من حيث رتبة الإنتاج الصناعي، ويكون من المفيد الوقوف على درجة الارتباط بين هذين النمطين من الصناعة في هذه الأقاليم الخمس.

وتكون الخطوة الأولى لحساب معامل ارتباط الرتب هي إعادة تبويب البيانات وترتيبها حسب الرتبة، وحساب الفرق بين كل من المجموعتين في كل حالة (ف)، ثم تربيع هذه الفروق للتخلص من إشاراتها السالبة ($ف^2$)، ونحصل على مجموع هذه المربعات (مجـ $ف^2$)، وذلك على النحو التالي:

الإقليم	الصناعات الهندسية (الرتبة)	صناعة السيارات (الرتبة)	ف	$ف^2$
أ	١	٣	٢	٤
ب	٢	٢	صفر	صفر
جـ	٣	١	٢	٤
د	٤	٥	١	١
هـ	٥	٤	١	١
			مجـ $ف^2$ = ١٠	

وبتطبيق قانون سبيرمان للرتب:

$$\text{ر} = ١ - \frac{٦ \text{ مجـ ف}^{٢}}{\text{ن}^{٣} - \text{ن}}$$

$$\therefore \text{ر} = ١ - \frac{٦ \times ١٠}{٥^{٣} - ٥} = ١ - \frac{٦٠}{١٢٥ - ٥} = ١ - \frac{٦٠}{١٢٠} = ١ - ٠,٥ = + ٠,٥$$

وتدل هذه القيمة على أن هناك علاقة موجبة ليست قوية بين الصناعتين في هذه الأقاليم الخمس، وبدرجة يمكن التغاضي عنها.

وهكذا يبدو واضحًا أن أحسن الطرق الإحصائية لمعرفة الارتباط بين مجموعتين إحصائيتين مرتبتين هي استخدام معامل سبيرمان لارتباط الرتب، الذي يستخدم الرتب وليس الأرقام الأصلية الواردة في البيانات، والأساس في حساب هذا المعامل هو مقارنة الفروق في الرتبة لكل عنصر من عناصر المجموعتين الإحصائيتين، وتكون الخطوة الأولى كما لاحظنا هي أن نضع المجموعتين في جدول جنبًا إلى جنب حتى يمكن استخراج الفوارق بسهولة، ويبين ذلك الجدول التالي:

جدول

حساب معامل الارتباط بين نسبة الأطفال (أقل من ١٢ سنة)

ونسبة الأمية لدى السكان في محافظات مصر حسب بيانات تعداد ١٩٧٦

المحافظة	نسبة الأطفال (%)	نسبة الأمية (%)	ترتيب المحافظة	الأطفال	الأمية	ف	ف٢
القاهرة	٢٧,٣	٣٤,٦	بورسعيد	٢١	٢٠	١	١
الإسكندرية	٢٧,٧	٣٧,٤	القاهرة	٢٠	٢١	−١	١
بورسعيد	٢٤,٥	٣٥,٩	الإسكندرية	١٩	١٩	٠	٠
السويس	٣٠,٠	٤٤,٤	السويس	١٨	١٨	٠	٠
دمياط	٣٢,٧	٤٩,٤	الغربية	١٧	١٣	٤	١٦
الدقهلية	٣١,٩	٥٦,٣	المنوفية	١٦	١٠	٦	٣٦
الشرقية	٣٣,٤	٦٢,٦	الدقهلية	١٥	١١	٤	١٦
القليوبية	٣٢,٥	٥٣,٧	الإسماعيلية	١٤	١٦	−٢	٤
كفر الشيخ	٣٣,٣	٧٠,١	قنا	١٣	٣	١٠	١٠٠
الغربية	٣٠,٤	٥٤,٩	الجيزة	١٢	١٥	−٣	٩
المنوفية	٣١,٦	٥٦,٩	دمياط	١١	١٧	−٦	٣٦
البحيرة	٣٣,٧	٦٦,٢	أسوان	١٠	١٢	−٢	٤
الإسماعيلية	٣٢,١	٥٠,٨	المنيا	٩	٤	٥	٢٥
الجيزة	٣٢,٦	٥٣,٥	سوهاج	٨	٢	٦	٣٦
بني سويف	٣٣,٢	٦٨,٤	بني سويف	٧	٧	٠	٠
الفيوم	٥٣,٢	٧٣,٦	كفر الشيخ	٦	٥	١	١
المنيا	٣٣,٠	٧٠,٩	الشرقية	٥	٩	−٤	١٦
أسيوط	٣٣,٦	٦٨,٥	القليوبية	٤	١٤	−١٠	١٠٠

سوهاج	٣٣,١	٧٢,٨	أسيوط	٣	٦	– ٣	٩
قنا	٣٢,٥	٧١,٢	البحيرة	٢	٨	– ٦	٣٦
أسوان	٣٢,٩	٥٦,٠	الفيوم	١	١	٠	٠
المجموع	–	–	–	–	–	–	٤٤٦

وبتطبيق قانون سبيرمان للرتب وهو $\text{ر} = ١ - \frac{\text{٦ مجـ ف}^{٢}}{\text{ن}^{٣} - \text{ن}}$

نحصل على:

$$\text{ر} = ١ - \frac{٦ \times ٤٤٦}{٢١^{٣} - ٢١} = ١ - \frac{٢٦٧٦}{٩٢٦١ - ٢١} = ١ - \frac{٢٦٧٦}{٩٢٤٠} = ١ - ٠,٢٨٩٦ = + ٠,٧١$$

وتدل هذه القيمة على أن هناك علاقة موجبة قوية بين ارتفاع نسبة الأطفال وارتفاع نسبة الأمية بين السكان بدرجة كبيرة لا يمكن التغاضي عنها.

وكما ذكرنا فإن استخراج معامل ارتباط الرتب عملية تتميز بالسرعة في الحساب، ولكنها أقل دقة من طريقة بيرسون المطولة، وينعكس ذلك على قيمة (ر)، وعلى سبيل المثال فإن ارتباط بيرسون للعلاقة بين الأمية ونسبة الأطفال في محافظات مصر وصلت إلى +٠,٨٣ مقابل + ٠,٧١ في طريقة سبيرمان.

معامل ارتباط الرتب للبيانات الوصفية :

بالإضافة إلى ما سبق من حساب الارتباط من قيم رقمية قد يحتاج الباحث إلى حساب معامل ارتباط الرتب لبيانات وصفية، وفي تلك الحالة لا بد من تحويل الأوصاف إلى قيم رقمية ويضعها في جدول رتب، فعلى سبيل المثال إذا أردنا دراسة العلاقة بين الحالة التعليمية لرب الأسرة والمستوى الاقتصادي للأسرة في عينة من سبع أسر مختلفة في حي معين، وكان الجدول الوصفي على النحو التالي:

رقم الأسرة	الحالة التعليمية لرب الأسرة	المستوى الاقتصادي للأسرة
١	يحمل شهادة متوسطة	فقيرة
٢	أمي	معدمة
٣	يقرأ ويكتب	فقيرة
٤	يحمل شهادة عالية	غنية
٥	أمي	معدمة
٦	أمي	متوسطة الحال
٧	يقرأ ويكتب	فقيرة

وللوقوف على مدى العلاقة بين هاتين الظاهرتين يجب أن نستخدم معامل سبيرمان لارتباط الرتب، والفكرة الأساسية في قياس هذا المعامل هي مقارنة رتبتي الأسرة الواحدة في الظاهرتين، فإن اختلفتا كثيرًا دل ذلك على قلة الارتباط، وإن اتفقا دل ذلك على شدة الارتباط - أي أن أساس المعامل - هو الفروق بين الرتب المتقابلة، فكلما كبرت هذه

الفروق في المتوسط كلما ضعف الارتباط بين الظاهرتين، والعكس كلما صغرت هذه الفروق.

ولحساب هذا العامل في هذا المثال فإننا نعطي لكل أسرة رتبة حسب الحالة التعليمية لرب الأسرة والحالة الاقتصادية للأسرة نفسها، ثم نحسب الفرق بين الرتبتين ثم نربع هذا الفرق ثم نطبق قانون سبيرمان السابق ذكره ويبين الجدول التالي خطوات العمل.

رتب الحالة التعليمية		رتبة الحالة الاقتصادية		الفرق بين الرتب المعدلة	ف^{٢}
أصلية	معدلة	أصلية	معدلة		
٦	٦	(١)	٤	٢	٤
(١)	٢	(١)	١,٥	٠,٥	٠,٢٥
(٤)	٤,٥	(٤)	٤	٠,٥	٠,٢٥
٧	٧	٧	٧	صفر	صفر
(٢)	٢	(٢)	١,٥	٠,٥	٠,٢٥
(٣)	٢	٦	٦	٤	١٦,٠٠
(٥)	٤,٥	(٥)	٤	٠,٥	٠,٢٥
مجموع ف^{٢}	–	–	–	–	٢١,٠٠

∴ $ر_{س}$ (معامل ارتباط الرتب) $= ١ - \frac{٦ \times ٢١}{٧\,(٤٩ - ١)} = ١ - \frac{١٢٦}{٣٣٦} = ١ - ٠,٣٧٥ = +٠,٦٢٥$

أي أن الارتباط طردي وقوي نوعًا ما.

ويلاحظ أننا عندما نرتب القيم المفردة في مجموعة البيانات ترتيبًا تنازليًّا وفق رتبتها، فقد نجد أن قيمتين أو ثلاثة تشترك في نفس الرتبة (تعرف بالمراتب المشتركة)، وقبل أن نحسب $ر_{س}$ فمن الضروري أن نعطي قيمة فعلية للمراتب المشتركة، ويكون ذلك بأن نعطي متوسط القيمة للمراتب المشتركة، ففي المثال السابق وضعنا الرتب المشتركة بين أقواس حسب ترتيبها الأصلي في المجموعة، ثم عدلنا هذه الرتب بأخذ المتوسط لها، فالحالة التعليمية يقرأ ويكتب تكررت مرتين وترتيبها الأصلي ٤,٥ فأخذنا المتوسط وهو ٤,٥ كترتيب معدل، والحالة التعليمية (أمي) تكررت ثلاث مرات وترتيبها ١، ٢، ٣ فأخذنا المتوسط وهو ٢ كترتيب معدل، وهكذا بالنسبة للظاهرة الثانية وهي الخاصة بالحالة الاقتصادية للأسرة ذاتها.

ومن الواضح كما ذكرنا أن معامل سبيرمان يتميز ببساطة الحساب، فالعمليات الحسابية التي يحتاجها سهلة للغاية، ولكن لا يجب أن نستخدمه في قياس الارتباط إلا في حالة أن يكون ترتيب الظواهر أمرًا منطقيا معقولا.

ج- الانحدار regression

(التوضيح البياني للارتباط - الانحدار)

سبق القول أن معاملات الارتباط تبين مدى ارتباط كل من المتغيرين أو الظاهرتين، فبدلالة ر، $ر_س$ على سبيل المثال يمكن أن نقول بدقة أكثر ما إذا كان المتغيران يظهران درجة قوية من الارتباط أو ارتباطًا متوسطًا أو ضعيفًا، ومع ذلك كله فغالبًا ما تكون في حاجة إلى معرفة المزيد عن العلاقة بين الظاهرتين، وبمعنى آخر نحتاج إلى أن نعرف شكل العلاقة. فإذا كان لدينا عدة أزواج من القيم لظاهرتين فإن خط الانحدار لظاهرة منهما على الأخرى هو الخط البياني الذي يمثل الظاهرتين، إحداهما الظاهرة المستقلة independent variable ممثلة على المحور الأفقي، والأخرى الظاهرة الفرعية أو التابعة dependent variable ممثلة على المحور الرأسي، فإذا كانت هناك علاقة بين الظاهرتين كان لهذا الخط وجود، أما إذا انعدمت العلاقة لما أمكن رسم مثل هذا الخط، بمعنى أن وجود العلاقة بينهما يؤدي ذلك كله إلى وجود اتجاه عام للنقط التي نحددها في الرسم أمام قيم الظاهرة المستقلة والقيم التابعة لها للظاهرة الأخرى، هذا الاتجاه العام إما أن يكون مستقيمًا فتكون العلاقة من الدرجة الأولى، ويسمى الارتباط مستقيمًا (أو خطيًّا)، وإما أن يكون غير مستقيم فتكون العلاقة من درجة أعلى من الأولى، ويسمى الارتباط في هذه الحالة غير مستقيم أو غير خطي.

على أنه يجب أن نلاحظ ما نقصده هنا بالاتجاه العام، حيث إن ما نعنيه بذلك لا يستلزم وجود جميع النقط على خط مستقيم أو غير مستقيم (خط الانحدار)، وإنما يكون هناك اتجاه عام لها يمكن تحديده بالرسم. وإذا وقعت جميع النقط على خط الانحدار كان ذلك دليلًا على الارتباط الكامل بين الظاهرتين، وكلما قربت النقط من خط الاتجاه العام أو وقع معظمها عليه كان ذلك دليلًا على شدة الارتباط بين الظاهرتين، بينها إذا بعدت معظم النقط عن خط الانحدار كان ذلك دليلًا على ضعف الارتباط بينهما. وبمعنى آخر أنه كلما كان تشتت النقط حول الانحدار كبيرًا كلما ضعف الارتباط، والعكس كلما كان تشتتها ضعيفًا كان ذلك دليلًا على شدة الارتباط بين الظاهرتين اللتين ندرس العلاقة بينهما.

وخط الانحدار بذلك يوضح لنا بيانيًّا العلاقة بين الظاهرتين، فهو يبين لنا كيف تميل الظاهرة التابعة إلى التغير نتيجة تغير معين في الظاهرة المستقلة، على أن هذا التغير في الظاهرة التابعة الذي يظهره الرسم ليس هو حتمًا نفس التغير الذي يحدث في الواقع العملي - تمامًا مثل المتوسط لعدة قيم - فلا يعني هذا المتوسط أن جميع القيم متساوية حتمًا، لذلك يسمى خط الانحدار أحيانًا بخط العلاقة المتوسطة بين الظاهرتين حيث يعطينا القيمة النظرية للمتغير التابع التي تقابل قيمة معينة للمتغير المستقل. وهذه القيمة الثالثة قد تكون هي نفسها القيمة الواقعية، وقد تختلف عنها بعض الشيء، وكلما قربت القيم النظرية للمتغير التابع من قيمته الواقعية كلما كان ذلك دليلا على أن خط الانحدار يمثل العلاقة بين الظاهرتين تمثيلًا

صادقًا، وكلما بعدت القيم النظرية عن القيم الواقعية كلما دل ذلك على ضعف تمثيل خط الانحدار للعلاقة بين الظاهرتين.

شكل الانتشار: Scatter Diagram

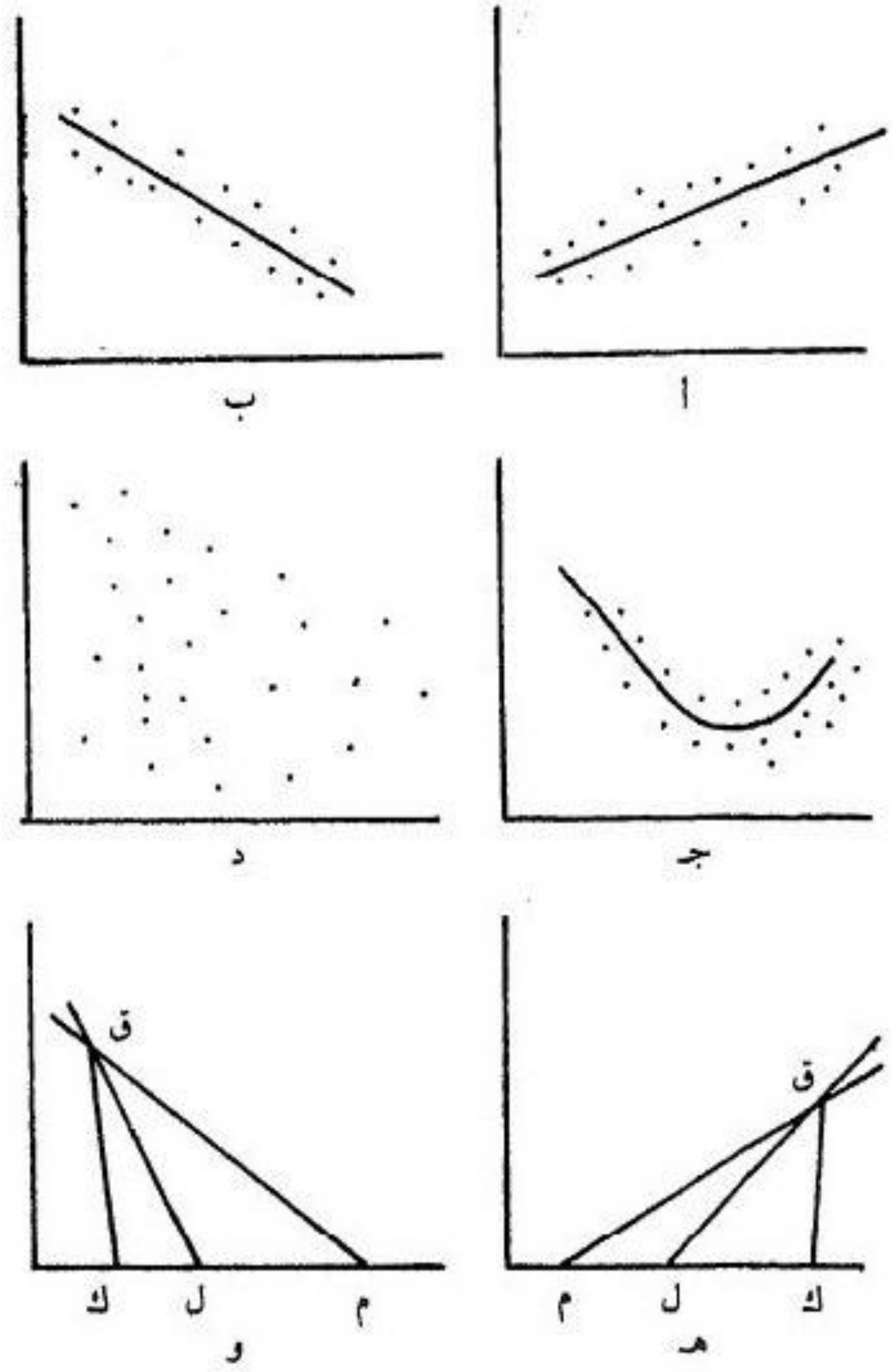
مما سبق يبدو أن الارتباط والانحدار بينهما علاقة متبادلة، فإذا عُلِم أحدهما أمكن إيجاد الآخر منه، وبمعنى آخر أنه في الإمكان حساب معامل الارتباط بين متغيرين بمعلومية مجموعة من أزواج القيم المتناظرة لهما، ونستطيع أن ننظر إلى هذه الأزواج من القيم من ناحية أخرى نظرة هندسية على أن كل زوج منها يعين نقطة في مستوى الورقة بالنسبة إلى محورين متعامدين، يمثل أحدهما قيم أحد المتغيرين، ويمثل الآخر قيم المتغير الأخر، وعند تمثيل هذه القيم بالنقط كما ذكرنا نحصل على ما يسمى «شكل الانتشار».

والنقط في شكل الانتشار قد تتجمع حول اتجاه معين، فيقال أن المتغيرين بينهما ارتباط، أو تكون مبعثرة بلا نظام أو اتجاه فلا يكون بين المتغيرين أي ارتباط كما في شكل (د)، وعند تجمع النقط حول اتجاه معين قد يكون هذا الاتجاه مستقيمًا فيقال إن شكل الانتشار مستقيم مثل الشكلين (أ)، (ب)، أو منحنٍ، فيقال إن شكل الانتشار منحن مثل (ج).

ونحاول عندئذ أن نوفق خطًا مستقيمًا أو خطًا منحنيًا يطابق هذه النقط أحسن مطابقة. ومن أهم الشروط اللازمة لذلك أن يمر الخط بأكبر عدد ممكن من هذه النقط، ويمر خلال باقي النقط بالتوازن والخط الناتج يسمي خط الانتشار أو خط الانحدار، ويمكن الحصول على هذا الخط بمحاولة رسمه ليطابق النقط، أو الأفضل باستخدام طريقة جبرية هي طريقة المربعات الصغرى Method Of Least Squares .

وتعني هذه الطريقة أن الخط الذي نريد توفيقه للنقط سواء كان مستقيمًا أو منحنيًا يشترط فيه أن يمر بجميع النقط، فلا بد أن بعض النقط على العموم سوف لا تقع عليه بل تنحرف عنه، أي أنه إذا أخذنا قيمة س لأي نقطة وقدرنا بين معادلة الخط قيمة ص المناظرة لها، قد نجدها مساوية لقيمة ص الفعلية للنقطة، أي قيمة ص المشاهدة بين بيانات المجموعة، وقد نجدها غير مساوية لها. والفرق بين قيمة ص المشاهدة وقيمتها المقدرة يسمى انحراف النقطة عن خط الانحدار. وأساس هذه الطريقة اعتبار الخط الذي يطابق النقط أحسن مطابقة هو الخط الذي يكون مجموع مربعات انحرافات النقط عنه أصغر ما يمكن.

أشكال الانتشار وخطوط الانحدار



وللدلالة على ذلك نعود إلى مثالنا عن الأرض الرعوية ، فقد أوضحنا أن مقدار الأرض الرعوية مرتبط بمقدار الأرض المرتفعة في قرى الإقليم (ر =+٠,٩٥). وقد يكون من المفيد أن نعرف أي قدر من الأرض الرعوية هو الذي يرتبط بأي ارتفاع. [فإذا كانت نسبة الأراضي التي تزيد على ٥٠٠ مترٍ مثلًا في قرية ما هي ٥٠٪، فكم من هذه المساحة سيكون رعويًا؟]

وكما سبق القول إن شكل العلاقة يمكن أن يتضح بسهولة تامة بواسطة رسم بياني وهو الرسم الذي يبين العلاقة بين متغيرين أحدهما تابع Dependent نوقعه على المحور الصادي، والأخر مستقل Independent نوقعه على المحور السيني، وعلى ذلك نبدأ في توقيع بيانات الأرض الرعوية والارتفاعات الخاصة بكل قرية من القرى على مثل هذا الرسم. و في بادئ الأمر علينا أن نحدد ما هو المتغير التابع، وما هو المتغير المستقل، من الواضح أن الأرض الرعوية تعتمد على الارتفاع، ومن ثم فنسبة الأرض الرعوية هو المتغير التابع ونوقعه على المحور الصادي (شكل أ).

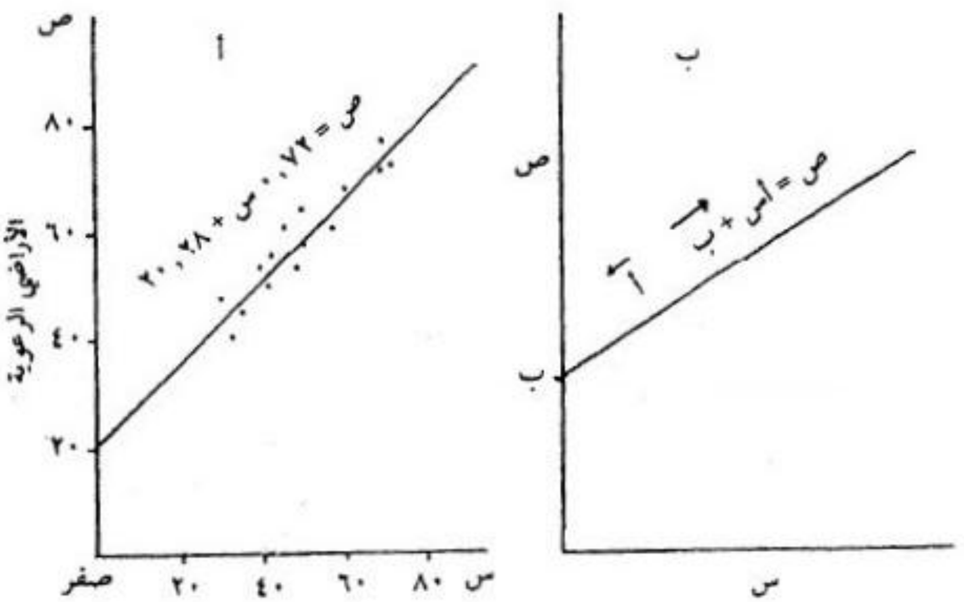
ومن هذا الرسم البياني يمكن ملاحظة أن الشكل العام للعلاقة بين المتغيرين يأخذ - بصفة عامة - خطًا مستقيمًا (تسمى بالعلاقة الخطية)، وبين الخط المستقيم الاتجاه المتوسط للعلاقة، فإذا عرفنا الشكل الدقيق لهذا الاتجاه المتوسط فيمكننا أن نبدأ في إجابة السؤال الذي سألناه من قبل وهو ما هو المقدار المعين في متغير ما الذي يرتبط مقدار معين في المتغير الآخر؟».

وفي المثال السابق يمكن أن نلاحظ اتجاه الخط بمجرد النظر، ولكن حتى نكون أكثر دقة علينا أن نحسب رياضيًّا الموقع الدقيق لهذا الخط المتوسط في الرسم البياني والذي يعرف بخط الانحدار، ويمثل «أحسن توفيق» لخط يمر بمجموعة من النقط على رسم بياني.

خطوط الانحدار

أ- الأراضي الرعوية والارتفاع

ب - الشكل العام لخط الانحدار



ولما كان الخط المستقيم على أي رسم بياني يظهر بوضوح علاقة القيم على المحور الصادي بمثيلتها على المحور السيني، فقيمة ص ترتبط بقيمة س، فإن الشكل النهائي (الكلي) لهذه العلاقة يعكسه موقع وانحدار هذا الخط على الرسم البياني. وعلى ذلك إذا كنا نريد معرفة كيف ترتبط قيم ص بقيم س، فعلينا أن نأخذ في الاعتبار موقع وانحدار أحسن خط متوسط تم توفيقه والذي يمر بالنقط الفعلية.

وبصفة عامة فقد وجد أن العلاقة بين ص إلى س تبينها المعادلة التالية:

ص = أ س + ب

حيث أ هي مقياس انحدار الخط،

ب تحدد موقع الخط على الرسم البياني وذلك بتحديد النقطة التي يقطع فيها الخط المحور الصادي (شكل ب).

وبعد أن نحدد قيمة أ، ب لعلاقة ما يكون من السهل حساب قيمة ص (المراعي في مثالنا) المرتبطة بقيمة س (الارتفاع)، وإذا كان لدينا ارتفاع معلوم (س) فسيكون من السهل حساب قيمة (ص) بتعويض الأرقام في المعادلة

ص = أ س + ب.

ولكن قبل إجراء ذلك علينا أن نحسب أ، ب. وهناك معادلة تساعدنا في حساب قيمتهما، فنستطيع حساب قيمة أ بالتعويض في المعادلة:

$$\text{أ} = \frac{\text{مجـ س ص} - (\text{مجـ ص})\ (\text{مجـ ص}) / \text{ن}}{\text{مجـ س}^{٢} - (\text{مجـ س})^{٢} / \text{ن}}$$

حيث ن = عدد د أزواج المشاهدات.

وبنفس الطريقة نستطيع أن نحسب قيمة ب بالتعويض في المعادلة:

بَ = صَ – أ سَ

وقبل كل شيء لكي نحسب قيمة أ، ب علينا أن نحسب مجـ س، مجـ ص ، مجـ س ص، مجـ س^{٢} ، س ص. وكل هذه القيم سبق حسابها عند استخراج قيمة ر وذلك باستثناء مجـ س ص. ولذلك فلن يأخذ ذلك قدرًا كبيرًا من العمل**!**.

وفي هذا المثال بالذات تم حساب القيم التالية:

مجـ س = ٦١٢	مجـ س^{٢} = ٤٣٢٨٦
مجـ ص = ٦٨٤	س = ٥١,٠
مجـ س ص = ٣٧٠٩٩	صَ = ٥٧,٠

وبالتعويض للحصول على قيمة «أ»:

$$\text{أ} = \frac{٣٧٠٩٩ - ٣٤٨٨٤}{٣٤٢٨٦ - (٦١٢)^{٢} / ١٢} = ٠,٧٢$$

وللحصول على قيمة (ب):

ب = ٥٧ – ٠,٧٢ (٥١,٠) = ٢٠,٢٨.

وعلى ذلك يصبح في إمكاننا استخدام المعادلة ص = أ س + ب لكي نرسم خط الانحدار، فقد سبق أن حسبنا قيمة أ، ب وعلى ذلك يكون خط الانحدار في هذا المثال هو:

ص = ٠,٧٢ س + ٢٠,٢٨

وحتى يمكن أن نحدد موقع الخط على الرسم البياني فإننا نحسب قيمتي ص التاليتين:

(١) قيمة ص عندما تكون س هي الوسط الحسابي (سَ) لمجموعة البيانات.

(٢) قيمة ص عندما تكون س أية قيمة أخرى.

وعلى ذلك نجد في مثالنا الذي نتعامل معه - حيث قيمة سَ تساوي ٥١,٠ - أن المعادلة هي:

ص = (٠.٧٢) (٠,٧٢) + ٢٠,٢٨ = ٧٢,١٢.

وهاتان هما القيمتان اللتان نوقعهما على رسمنا البياني الأصلي (ص = ٥٧,٠ عندما تكون س = ٥١,٠ ، ص = ٧٢,١٢ عندما تكون س = ٧٢,٠) (ويبدو ذلك في الشكل أ)، وكما نلاحظ فإن هذا الخط يمر كمتوسط في منتصف مفردات منظومة النقط على الرسم البياني. لاحظ أيضًا كيف أن هذا الخط يقطع المحور الصادي عند قيمة ٢٠,٢٨ - أي قيمة (ب). ومن المعتاد أن نبين على خط الانحدار المعادلة التي تمثله، ص = ٠,٧٢ س + ٢٠,٢٨ (شكل أ).

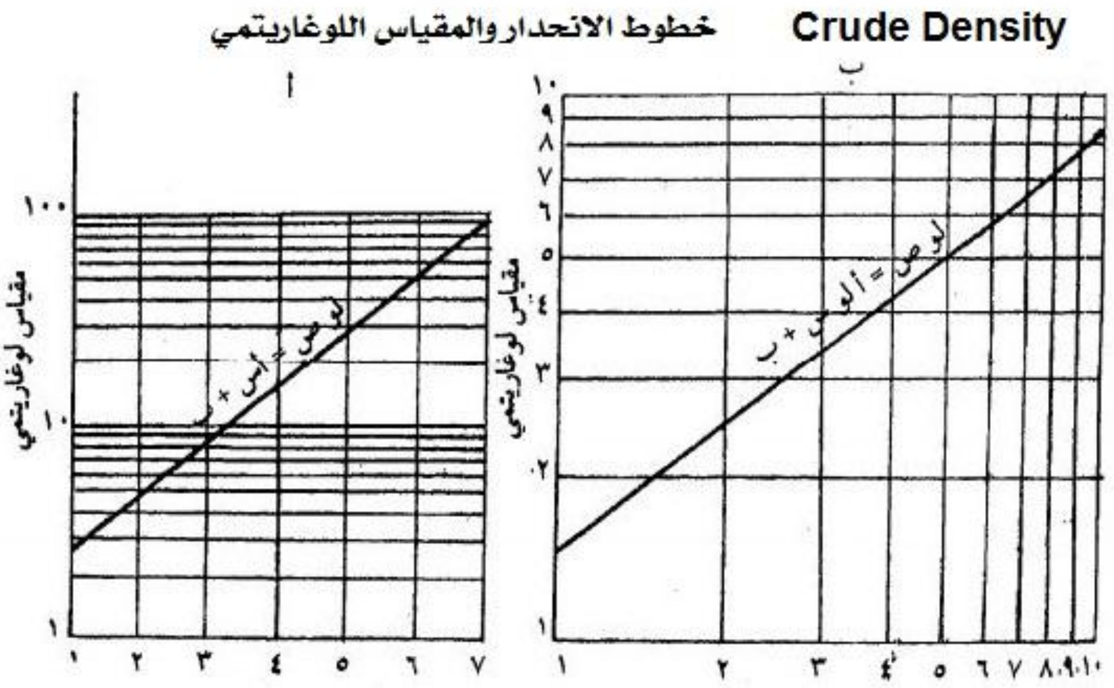
وعلى أية حال يكون من الضروري في بعض الأحوال ألا نستخدم ورق الرسم البياني العادي (والذي وقعنا عليه حتى الآن خط الانحدار) ولكن نستخدم إما الورق البياني شبه اللوغاريتمي أو اللوغاريتمي.

وفي مثل هذه الأحوال فإن أحد المحورين المستخدمين في الرسم البياني أو كليهما سيكون بمقياس لوغاريتمي. ومن الضروري حينئذ عندما نحسب شكل خط الانحدار أن نستخدم اللوغاريتمات المناسبة وليس الأرقام الأصلية.

وعلى العموم يمكننا أن نقول:

١- عندما نستخدم ورق بياني شبه لوغاريتمي فإن المحور السيني سيكون بمقياس حسابي عادي بينما المحور الصادي سيكون بمقياس لوغاريتمي. وعلى ذلك فإن كل الحسابات التي تتضمن قيم ص ستتم باستخدام لوغاريتمات قيم ص، ففي تلك الحالة سيكتب الشكل النهائي لخط الانحدار على النحو التالي:

لو ص = أ س + ب (شكل - أ).



٢ - عندما نستخدم ورق بياني لوغاريتمي، فإن كلًّا من قيم س ، ص ستتحول إلى قيم لوغاريتمية، ومن ثم سيكون شكل خط الانحدار على النحو التالي:

لو ص = أ لو س + ب (شكل ب).

وعلى ذلك فمن الممكن تقدير قيمة ص إذا كانت لدينا قيمة س على اعتبار أن هناك علاقة ارتباطية بين القيمتين، وهذا إذن انحدار ص بالنسبة إلى س (أو انحدار ص على س كما يسمى ذلك بالتعبير الإحصائي) والآن ما الذي فعلناه بالضبط؟

لعلك تذكر أننا بدأنا الحديث بالقول أننا نريد الحصول على خط يبين لنا نوع العلاقة المتوسطة لمجموعة النقاط على رسم بياني وذلك هو ما فعلناه تماما فقد وضعنا الخط عبر منتصف النقاط بطريقة كانت المسافات بين النقاط وهذا الخط هي أقل ما يمكن. (أي أننا جعلنا المسافة بين الخط والنقاط هي أدنى مسافة ممكنة) وحتى الآن فقد قمنا بقياس قيم «ص» بالنسبة إلى قيم «س»، وبمعنى آخر قللنا المسافات إلى أدنى حد ممكن بشكل رأسي على الرسم البياني. وعلى ذلك فإن المسافات الرأسية للنقط عن الخط هي الحد الأدنى، وقد حسب قياس هذه المسافة في معادلتنا (بتربيع) المسافات حتى يمكن إزالة العلامات الجبرية - لاحظ عدد القيم المربعة في معادلتنا لحساب (أ) - ولهذا السبب تعرف طريقة توفيق الخط بين كل النقاط بطريقة (المربعات الصغرى).

وينبغي القول بأنه من السهل أن نحسب خط انحدار لقيمة س على ص، والتي تقيس (المربعات الصغرى) في اتجاه أفقي، ومن الواضح أن هذا الخط سيكون ذا انحدار مختلف

عن انحدار خط ص على س، ولكن سيقطع خط ص على س عند نقطة القيمة المتوسطة. وعلى ذلك فهناك باستمرار خطان للانحدار في أية مجموعة بيانات، أحدهما يتنبأ بقيمة ص على س، والآخر يتنبأ بقيمة س على ص، وبطبيعة الحال فإن شكل خط الانحدار يختلف من مسألة إلى أخرى من المسائل التي نتعامل معها، ففي بعضها يكون الانحدار أكبر بكثير من الخط الذي حسبناه الآن، وفي البعض الآخر سيكون أقل من ذلك. والمهم دائمًا أن نتذكر أن خط الانحدار يصف الشكل الإجمالي للعلاقة بين متغيرين، بينما يدل معامل الارتباط ببساطة على مدى هذه العلاقة.

وبالإضافة إلى ذلك فإن معادلة خط الانحدار (الخط المستقيم) يمكن تطبيقها في تحديد الاتجاه العام للظاهرات السكانية، ومنها على سبيل المثال تحديد اتجاه معدلات الوفيات في فترة زمنية محددة، ومثال ذلك أننا إذا أردنا أن نحسب الاتجاه العام لمعدلات وفيات الأطفال الرضع الناتجة عن أمراض الجهاز التنفسي بالإسكندرية في الفترة من ١٩٥٠ – ١٩٦١، والتي يبينها الجدول التالي فإننا نطبق معادلة الخط المستقيم.

جدول

معدل وفيات الأطفال الرضع بالإسكندرية

الناجمة عن أمراض الجهاز التنفسي في الفترة من ١٩٥٠ - ١٩٦١

السنة	معدل الوفيات في الألف
١٩٥٠	٢٠,٦
١٩٥١	٢٦,٧
١٩٥٢	٢٣,٩
١٩٥٣	٢٧,٥
١٩٥٤	٢٥,٩
١٩٥٥	٢٤,١
١٩٥٦	١٨,٨
١٩٥٧	١٧,٤
١٩٥٨	١٩,٣
١٩٥٩	١٩,٨
١٩٦٠	٢٠,٠
١٩٦١	١٩,٣

ص = أ س + ب

حيث ص = معدل الوفيات   س = الزمن بالسنوات

ثم نحصل على قيمة أ، ب و بطريقة أخرى وذلك باستخدام المعادلتين:

(١) مجـ ص = أ مجـ س + ن ب

(٢) مجـ س ص = أ مجـ س + ب مجـ س

ثم يعمل الجدول التالي:

السنة	س	ص	$س^٢$	س ص
١٩٥٠	١	٢٠,٦	١	٢٠,٦
١٩٥١	٢	٢٦,٧	٤	٥٣,٤
١٩٥٢	٣	٢٣,٩	٩	٧١.٧
١٩٥٣	٤	٢٧,٥	١٦	١١٠,٠
١٩٥٤	٥	٢٥,٩	٢٥	١٢٩,٥
١٩٥٥	٦	٢٤,١	٣٦	١٤٤,٦
١٩٥٦	٧	١٨,٨	٤٩	١٣١,٦
١٩٥٧	٨	١٧,٤	٦٤	١٣٩,٢
١٩٥٨	٩	١٩,٣	٨١	١٧٣,٧
١٩٥٩	١٠	١٩,٨	١٠٠	١٩٨,٠
١٩٦٠	١١	٢٠,٠	١٢١	٢٢٠,٠
١٩٦١	١٢	١٩,٣	١٤٤	٢٣١,٦
الجملة	٧٨	٢٦٣,٣	٦٥٠	١٦٢٣,٩

وبالتعويض في المعادلة الأولى ينتج أن:

(٣) ٢٦٣,٣ = ٧٨ أ + ١٢ ب

(٤) ١٦٢٣,٩ = ٦٥٠ أ + ٧٨ ب

و بضرب (٣) في ٦,٥ ينتج أن:

١٧١١,٤٥ = ٥٠٧ أ + ٧٨ ب

، (٤) = ١٦٢٣,٩ = ٦٥٠ + ٧٨ ب

وبالطرح ٨٧,٥٥ = - ١٤٣ أ

∴ أ $= \frac{٨٧,٥٥}{-١٤٣} =$ - ٠,٦١٢

، ب = ٢٦٣,٣ = ٧٨ (- ٠,٦١٢) + ١٢ ب

= ٢٦٣,٣ = - ٤٧,٧٣٦ + ١٢ ب

∴ ٢٦٣,٣ + ٤٧,٧٣٦ = ١٢ ب

∴ ب $= \frac{٣١١,٠٣٦}{١٢} =$ ٢٥,٩٢

∴ المعادلة هي:

ص = - ٠,٦١٢ س + ٢٥,٩٢.

ومن الواضح أن الميل في الاتجاه العام ميل سلبي، أي يتجه نحو الهبوط. ومعامل س وهو -٠,٦١٢ يسمى الاتجاه، ومعنى هذا أن الظاهرة تميل قيمها إلى الهبوط بمقدار ٠,٦١٢ سنويًّا.

٣- المقارنات النظرية النسبية:

Theoretic Explanatory Comparisons

هناك طريقة مختلفة نوعًا ما عن الطرق السابقة لشرح التوزيعات التي نتعامل معها تعرف بالطريقة الاستنتاجية النظرية، ونستطيع بهذه الطريقة أن نقارن التوزيعات الفعلية لظاهرة ما، والتوزيعات الافتراضية لنفس الظاهرة (أي مقارنة البيانات الواقعية بالبيانات المستنبطة على أساس افتراض معين، وهو افتراض نظري بطبيعة الحال). كذلك نستطيع بها مقارنة نتائج عينة أجريناها في منطقة ما بالمجتمع الإحصائي الذي سحبنا منه هذه العينة، حيث نحصل على بيانات من العينة تختلف بعض الشيء عن المتوقع، وتصبح المشكلة أمام الباحث هي اختبار التطابق بين البيانات الواقعية والبيانات الافتراضية - أو بين البيانات التي حصل عليها من عينة ما والبيانات المتوقعة - وذلك حتى يمكن أن يحكم على افتراضه فإما أن يطمئن إليه ويعتبره مناسبًا غير بعيد عن الحقيقة، وإما ألا يعتبره كذلك فيستبعده، فإذا كان كل من التوزيعين (الفعلي والمتوقع) متطابقين، فسنعرف أن العوامل التي افترضناها كعوامل هامة في توزيعنا النظري ستكون بالتالي عوامل هامة في التوزيع الحقيقي.

والآن لنأت إلى مثال يوضح لنا تطبيق هذه الطريقة الاستنتاجية النظرية[١]، ولنفترض أننا جمعنا بيانات عن مساحة الأراضي الرعوية في مزارع ١٣٠ قرية بأحد الأقاليم، ووجدنا أن التوزيع التكراري الفعلي (المشاهد) للمزارع التي تزيد نسبة المراعي بها على ٥٠٪ موقعة مقابل نسبة الأراضي التي تزيد على ٥٠٠ مترٍ كما يبين الجدول التالي، فيبدو لنا من أول نظرة على الجدول أن معظم المزارع التي تزيد نسبة المراعي بها على ٥٠٪ من جملة مساحتها توجد في قرى ترتفع بها نسبة الأراضي التي تزيد على ٥٠٠ مترٍ، وسيكون الاستنتاج أن الارتفاع هو العامل الهام في تحديد مساحة المراعي، ولكن هذا الاستنتاج تقريبي جدًّا في التحليل الإحصائي، ذلك لأنه في النهاية مجرد تخمين بديهي قد يكون صائبًا وقد لا يكون، ويحدد بأن الوضع في هذه الحالة هو هكذا.

---

[١] تستخدم هذه الطريقة فقط إذا كانت البيانات مقسمة إلى مجموعات، وبمعنى آخر فهي طريقة للبيانات المتقطعة فقط.

جدول

التوزيع التكراري

(١٣٥ قرية تزيد بها نسبة المراعي على ٥٠ ٪)

عدد القرى	نسبة الأراضي التي تعلو على ٥٠٠ متر (٪)				
	صفر ــ ٢٠	٢١ ــ ٤٠	٤١ ــ ٦٠	٦١ ــ ٨٠	٨١ ــ ١٠٠
	٥	١٠	١٠	٣٠	٨٠

أ. فرض العدم: The Null Hypothesis

نستطيع في الواقع أن نقيس مدى الارتباط الفرضي السابق الذكر بأن نقوم بعمل توزيع نظري للمزارع التي يفترض أن تكون موجودة إذا لم يكن عامل الارتفاع ذا أهمية في تحديد المراعي بها، (أي ننشئ توزيعًا افتراضيًّا) فإذا كان التوزيع الفعلي مشابهًا في شكله مع التوزيع النظري فنستطيع القول بأن الارتفاع ليست له أهمية كعامل مؤثر في التوزيع، وإذا لم يكن التوزيع النظري مشابهًا للتوزيع الفعلي فنستطيع القول بأنه من المحتمل أن الارتفاع له أهمية كعامل مؤثر في التوزيع.

وعلى ذلك يصبح من الضروري أن نقوم بعمل توزيع نظري معقول، وهذا التوزيع النظري نسميه (فرض عدم) ذلك لأنه توزيع غير موجود في الواقع، بل افتراضي، والذي يمكن أن يحدث فقط إذا لم يكن للعوامل التي نراها في الواقع أي تأثير، وفي مثالنا هذا فإن (فرض العدم) هو أن الارتفاع ليس عاملًا هامًا في توزيع المراعي، فإذا كان الأمر كذلك فإن عدد المزارع الموجودة في كل فئة من فئات الارتفاع سيكون متطابقًا، (أي نفس العدد)، وحيث إن هناك خمس فئات للارتفاع وهي (صفر – ٢٠ ... حتى ٨١-١٠٠)، وأن عدد المزارع ١٣٥ فيترتب على ذلك أن نتوقع (في ضوء فرض العدم) أن هناك ٢٧ مزرعة في كل مجموعة من مجموعات الارتفاع، ويكون التوزيع التكراري المتوقع حينئذ كما يبينه الجدول التالي:

جدول

التوزيع التكراري المتوقع في ١٣٥ قرية نسبة المراعي بكل منها أكثر من ٥٠ %

اعتمادًا على فرض العدم أن الارتفاع ليس له تأثير في تحديد مساحة المراعي

عدد القرى	نسبة الأراضي التي تزيد على ٥٠٠ متر (٪)				
	٠ ــ ٢٠	٢١ ــ ٤٠	٤١ ــ ٦٠	٦١ ــ ٨٠	٨١ ــ ١٠٠
	٢٧	٢٧	٢٧	٢٧	٢٧

ب - مقارنة القيم المشاهدة بالقيم المتوقعة:

بعد ذلك يمكننا الآن أن نقارن القيم المشاهدة (ش) بالقيم المتوقعة (م)، ويمكن أن نفعل ذلك بتطبيق ما يعرف باختبار كاي تربيع Chi-Squared (يكتب $كا^{٢}$)، وهو أداة إحصائية تساعدنا في مقارنة البيانات الواقعية بالبيانات الافتراضية، حتى يمكن أن نحكم على افتراضنا فإما أن نطمئن إليه ونعتبره مناسبًا غير بعيد عن الواقع، وإما ألا نعتبره كذلك فنرفضه، ويتم حساب كاي تربيع بإضافة الفرق بين كل تكرار مشاهد وتكرار متوقع، ثم بتربيع هذا الفرق وقسمته على التكرارات المتوقعة.

أي أن:

$$كا٢ = مجـ \frac{(ش - م)^{٢}}{ش}$$

وسنجد أنه من السهل حساب $كا^{٢}$ إذا وضعنا بياناتنا في جدول مماثل للجدول التالي:

جدول

حساب كاي تربيع

(بيانات الجدولين السابقين)

	نسبة الأراضي التي تزيد على ٥٠٠ متر (٪)				
	٠ – ٢٠	٢١ – ٤٠	٤١ – ٦٠	٦١ – ٨٠	٨١ – ١٠٠
التكرارات المشاهدة (ش) (جدول رقم ٢٧)	٥	١٠	١٠	٣٠	٨٠
التكرارات المتوقعة (م) (جدول رقم ٢٨)	٢٧	٢٧	٢٧	٢٧	٢٧
ش – م	– ٢٢	– ١٧	– ١٧	٣	٥٣
$(ش - م)^{٢}$	٤٨٤	٢٨٩	٢٨٩	٩	٢٨٠٩
$\frac{(ش - م)^{٢}}{ش}$	١٧,٩٣	١٠,٧	١٠,٧	٠,٣٣	١٠٤,٠٤
$مجـ \frac{(ش - م)^{٢}}{ش} = ١٤٣,٧$					

وقيمة $كا^{٢}$ هذه تشبه إلى حد ما قيمة معامل الارتباط [ر أ و ر س] في أنها توضح كيف تتباين مجموعة بيانات معينة عن مجموعة بيانات أخرى، ولكنها لا تعطي معنى فوريًا كرقم في حد ذاته.

وهذا الاختبار ($كا^{٢}$) باختصار يعرف باختبار حسن المطابقة الذي يمكننا من تأكيد الرأي أو استبعاده، وينبغي أن نختبره لكي نرى ما إذا كان الفرق بين مجموعتي البيانات التي يمثلها فرقًا معنويًا (حقيقيًا) أم لا.

الفصل السابع

التمثيل البياني للمعلومات

الفصل السابع

التمثيل البياني للمعلومات

إن الغرض من دراسة التوزيعات العددية أو التكرارية هو تلخيصها وتمثيلها بيانيًا، وهذا يجعلها في صورة يسهل بها تعليلها، ويساعد على مقارنتها بتوزيعات أخرى، ولا شك بأن هذا التلخيص ينقص من تعقيد وصعوبة البيانات الأولية ويزيد من فائدتها، وكمثال على ذلك نجد أن البيانات الموجودة في جدول توضح كميات الأمطار الساقطة خلال ١٠٠ يوم على مدينة معينة على الرغم من قلتها نسبية ، إلا أنها لا بد وأن تلخص بطريقة يمكن الاستفادة منها عند مقارنتها بتوزيعات أخرى. وقد ذكرنا سابقا أن تبويب وجدولة هذه البيانات يساعد على فهمها واستيعابها.

وهناك وسائل أخرى تساعد على عرض البيانات عرض مبسطة يسهل فهمه واستيعابه ألا وهي الأشكال والرسوم البيانية . إن الرسوم البيانية هي أحد الوسائل البصرية التي تساعد على تصوير البيانات تصوير تقريبية. وسنعالج من هذه الأشكال والرسوم البيانية: المدرج التكراري، والمضلع التكراري، والمنحنى التكراري والرسوم البيانية الخطية.

يعد التمثيل البياني Graphic Presentation ، روح الإحصاءات وسبيل الوصول إلى ما تحتويه من معلومات، كما إنه سبيل الربط بين العوامل المختلفة، ويعتبره البعض لغة ثانية للباحثين. ويتوقف نجاح الباحث في الحصول على نتائج من التمثيل البياني لمادته العلمية على مدى نجاحه في اختيار الطريقة المناسبة للتمثيل.

وتنقسم طرق تمثيل البيانات - غير التكرارية - من وجهة النظر الكارتوجرافية Cartogaphy تعنى: فن رسم الخرائط والأشكال إلى عدة طرق نذكر منها:

١- طرق وصفية: وتهتم بوصف الاتجاه العام للظاهرة قيد الدراسة.

٢- طرق تحليلية : وتهتم بتحليل العلاقة بين المتغيرات أو تحديد معدلات

التغير النسبي للمتغير قيد البحث. والطرق الوصفية تضم عدد كبير من طرق التمثل أهمها :

أ- الطرق البيانية ذات البعد الواحد:

وتشمل الخطوط أو المنحنيات والأعمدة البيانية .

ب- الطرق البيانية ذات البعدين (أو المساحية)

وتضم الأشكال البيانية ذات المساحة كالدوائر والمربعات والمستطيلات.

ج- الطرق البيانية ذات الثلاثة أبعاد (الحجمية) وتتمثل في المكعبات والكرات والأسطوانات.

أما الطرق التحليلية للبيانات فمن أهمها المنحني اللوغاريتمى، ومنحنى لورنز، والأهرامات السكانية ، ومثلث التعادل.

وفيما يتعلق بالبيانات التكرارية - المبوبة - فيمكن تمثيلها بعدة طرق منها:

أ- المدرج التكراري

ب- المضلع التكراري

ج- المنحنى التكراري

وفيما يلي إشارة سريعة للعديد من هذه الطرق دون الإسهاب في كيفية إنشائها.

أولًا: طرق التمثيل الوصفية:

أ- الطرق البيانية ذات البعد الواحد:

١- الخطوط أو المنحنيات البيانية : Line Graphs

تستخدم هذه الطريقة لبيان تغير أو تطور ظاهرة خلال فترة زمنية، أي البيان علاقة بين متغيرين وعادة ما يكون أحدهما الزمن. والخط البياني الناتج عن هذه الحالة يسمى بالخط البياني البسيط.

مثال:

يبين الجدول التالي تطور كل من قيمة ونسبة رأس المال المستثمر في قطاعات الصناعة تبعا لنوع الصناعة بمدينة العاشر من رمضان خلال الفترة الممتدة بين عامي ١٩٧٩، ١٩٨٨ .

القيمة مليون جنيه

نوع الصناعة / السنة	الصناعات الغذائية		منتجات الأثاث والأخشاب المعدني		الصناعات النسيجية		الصناعات الهندسية		الصناعات المعدنية		صناعات مواد البناء والإنشاء		الصناعات الكيميائية		صناعات وخدمات صناعية متنوعة		الإجمالي	
	القيمة	%	القيمة	%	القيمة	%	القيمة	%	القيمة	%	القيمة	%	القيمة	%	القيمة	%	القيمة	%
١٩٧٩	-	-	٠.١٨٥	١.٩٣	-	-	-	-	-	-	٩.٢٥	٩٦.٤	-	-	٠.١٦	١.٦٧	٩.٥٩٥	١٠٠
١٩٨٠	٠.٢٥	١.٣٧	٣.٨٧٥	٢١.٢٥	١.٣٣	٧.٢٩	-	-	١.٧٢	٩.٤٣	١٠.٩٠	٥٩.٨٧	-	-	٠.١٦	٠.٨٨	١٨.٢٣٥	١٠٠
١٩٨١	٠.٢٥	١.١٣	٣.٨٧٥	١٧.٥٩	٤.٣٣	١٩.٦٥	٠.٣	١.٣٦	١.٧٢	٧.٨١	١١.٤	٥١.٧٣	-	-	٠.١٦	٠.٧٣	٢٢.٠٣٥	١٠٠
١٩٨٢	٠.٤٢٥	١.١٣	٨.١	٢١.٤٤	٧.٤٠٥	١٩.٦١	٢.٣	٦.٠٩	٤.٢٢	١١.١٧	١١.٧٢	٣١.٠٣	٣.٤٤٢	٩.١١	٠.١٦	٠.٤٢	٣٧.٧٧٢	١٠٠
١٩٨٣	٢.٦٩٥	٤.٢٠	٨.٦١	١٣.٤٢	١١.٤٠٥	١٧.٧٨	٢.٣	٣.٥٨	٧.٩٢	١٢.٣٥	١٩.٩٨٨	٣١.١٦	٦.٦٨٤	١٠.٤٢	٤.٥٥	٧.٠٩	٦٤.١٥٢	١٠٠
١٩٨٤	٤.٨٤٥	٤.٤٩	٨.٨٤٦	٨.٢٠	١٥.٢٢٥	٣٢.٦٧	٤.٥٥	٤.٢٢	٧.٩٢	٧.٣٤	٢٠.٤٣٨	١٨.٩٤	١٥.٤١٤	١٤.٢٨	١٠.٦٤٥	٩.٨٦	١٠٧.٩١٣	١٠٠
١٩٨٥	١٢.٢٠٣	٦.٤٣	٩.٣٤٦	٤.٩٢	٤٦.٠٤٦	٢٤.٢٥	٢٣.٤٦	١٢.٣٦	١٣.٧٨	٧.٢٦	٣٣.٩١٣	١٧.٨٦	٣٨.٢٨٤	٢٠.١٧	١٢.٨٢	٦.٧٥	١٨٩.٨٥٣	١٠٠
١٩٨٦	١٤.١٠٣	٦.٠٢	١٣.١٤٦	٥.٦٩	٥٢.٣٤٦	٢٢.٣٣	٣٤.٧٩٥	١٤.٨٤	١٦.٧٨	٧.١٦	٣٧.٩٠٣	١٦.١٧	٤٨.١٦٤	٢٠.٥٤	١٦.٩٩٥	٧.٢٥	٢٣٤.٤٣٢	١٠٠
١٩٨٧	٢١.٠٤٣	٧.٢٦	١٤.٨٣١	٥.١٢	٦١.٨٢٨	٢١.٣٤	٣٨.١٨٩	١٣.١٨	١٩.٦٦	٦.٧٩	٤٢.٣٣٩	١٤.٦١	٦١.٨٠٢	٢١.٣٢	٣٠.٠٣	١٠.٣٧	٢٨٩.٧٢٢	١٠٠
١٩٨٨	٢١.٠٤٣	٦.٩٦	١٥.٣٣١	٥.٠٧١	٦١.٨٢٨	٢٠.٤٦	٤٦.١٨٩	١٥.٢٨	١٩.٦٦	٦.٥١	٤٢.٣٣٦	١٤.٠١	٦٣.٣٠٢	٢٠.٩٥	٣٢.٥٣	١٠.٧٦	٣٠٢.٢٢٢	١٠٠

رسم الخط البياني البسيط :

يفضل دائما استخدام الأرقام النسبية عن الأرقام المطلقة في التمثيل البياني للمعلومات. من الجدول السابق : إذا كان الهدف هو التمثيل البياني لتطور رأس المال لصناعة واحدة ولتكن صناعة مواد البناء والإنشاء (أي متغير واحد خلال فترة زمنية) نتبع الآتي:

- نختار مقياس رسم مناسب للمحور الرأسي الذي سوف نمثل عليه قيمة رأس المال وليكن كل اسم = ٥ مليون جنيه.

- ومقياس رسم مناسب للمحور الأفقي الذي سوف نمثل عليه السنوات وليكن اسم = عام.

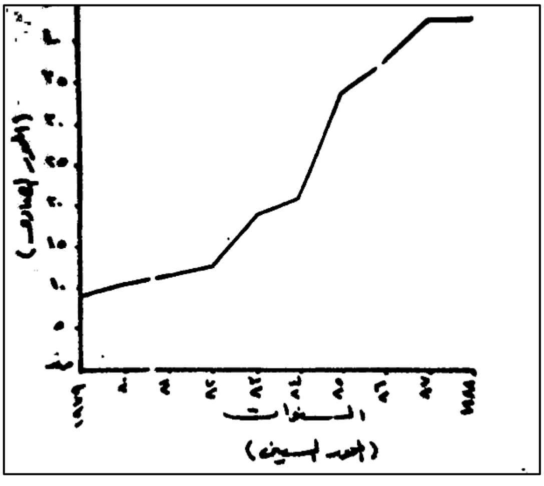
- ويراعى دائما أن يكون هناك توافق بين المحورين (الرأسي والأفقي) حتى لا يظهر الشكل الناتج بصورة مشوهه ، أي أن النسبة بين المحورين ١:١ تقريبًا.

- من نقطة الأصل (نقطة التقاء المحورين السيني والصادي) يبدأ التقسيم إلى الوحدات المختارة تبعًا لمقياس الرسم.

- نقوم بتوقيع كل قيمة باستخدام (س، ص) نظام الإحداثيات على نفس ترتيب سنوات الظاهرة، وفي المثال بداية من عام ١٩٧٩ حتى عام ١٩٨٨ للمحور الأفقي.

- بعد الانتهاء من توقيع القيم المختلفة نصل بين النقاط - التي سبق وأن وقعناها - بخط مستقيم لينتج في النهاية الخط البياني البسيط لتطور رأس المال المستثمر في الصناعة قيد البحث كما هو موضح بالشكل .

الشكل يوضح تطور قيمة رأس المال المستثمر في صناعة مواد البناء والأنشاء بمدينة العاشر من رمضان خلال الفترة الممتدة بين عامي ١٩٧٩ ، ١٩٨٨

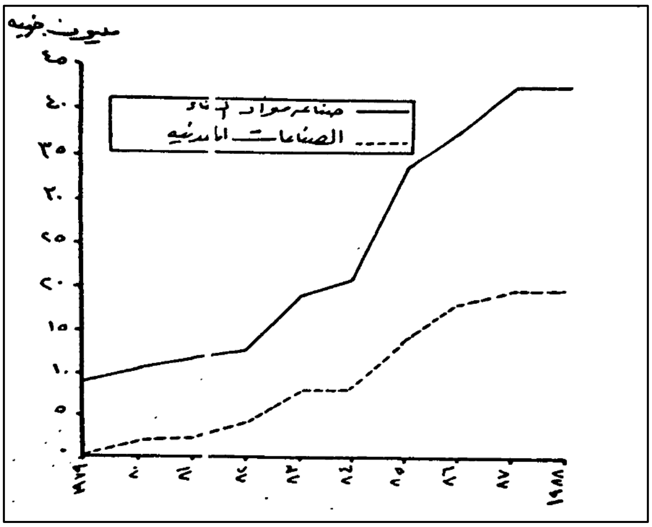


شكل يوضح الخط البياني البسيط

وبعد الانتهاء من التمثيل البياني للظاهرة، يقوم الباحث بوصف الاتجاه العام لها، ومن الشكل نتبين أن قيمة رأس المال المستثمر في هذه الصناعة في زيادة مستمرة وسريعة وخاصة خلال الفترة بين عامي ١٩٨٢ ، ١٩٨٧ والسبب في ذلك يرجع إلى زيادة أعداد المنشآت العاملة في هذا المجال من الصناعات بالمدينة لسد احتياجات النشاط العمراني لها في تلك الفترة .

أما اذا كان الهدف من الدراسة هو عقد مقارنة بين تطور رأس المال لأكثر من نوع صناعي أو بين كل الصناعات المذكورة بالجدول؛ ففي هذه الحالة يختار مقياس الرسم على أساس أكبر رقم في الإحصائية (٦٣,٣٠٢ مليون جنيه)للصناعات النسيجية، ويمكن استخدام المقياس السابق أي كل ١سم = ٥ مليون جنيه. ثم نقوم بتوقيع قيمة رأس المال المستثمر لكل صناعة خلال سنوات الدراسة، باستخدام نفس الطريقة السابقة (س، ص) ثم يرسم الخط البياني الخاص بها. وفي النهاية سوف ينتج شكل تتداخل فيه الخطوط البيانية أو تتعدد. ويطلق على هذا الشكل اسم الخطوط البيانية المتداخلة أو المتعددة.

والشكل يوضح تطور قيمة رأس المال المستثمر لكل من صناعة مواد البناء والإنشاء والصناعات المعدنية بمدينة العاشر من رمضان خلال الفترة الممتدة بين عامي ١٩٧٩، ١٩٨٨ اعتمادا على بيانات الجدول السابق. باستخدام الخط البياني المتعدد.



شكل يوضح الخط البياني المتعدد

أما النوع الثالث من الخطوط أو المنحنيات البيانية فيعرف بالمنحنيات البيانية المجمعة. ويختلف عن النوعين السابقين في أنه يرسم على أساس التجميع التراكمي لقيم أو نسبة كل مجموعات الدراسة ولتوضيح هذه الطريقة سوف نقوم برسم المنحني البياني النسبي المتجمع لقيمة رأس المال المستثمر في كافة الصناعات بمدينة العاشر من رمضان اعتمادًا على أرقام الجدول السابق.

١- نقوم بحساب النسبة المئوية لرأس مال كل صناعة من إجمالي رأس المال المستثمر خلال السنة الواحدة. مثلا: نسبة رأس المال المستثمر في صناعة الأخشاب عام ١٩٧٩ =

$$\frac{\text{رأس المال للصناعة عام 1979}}{\text{إجمالي رأس المال للصناعات عام 1979}} \times ١٠٠$$

$$= \frac{0.185}{9.595} = ١.٩٣\%$$

والجدير بالذكر أن النسب المئوية المذكورة في الجدول قد تم حسابها بنفس الطرقة السابقة

وفي هذه الحالة سيكون طول المحور الرأسي = ١٠٠ ٪

(أما أذا كان الرسم على أساس مطلق فيرسم المحور الرأسي على أساس المجموع الكلي لقيمة رأس المال).

أما المحور الأفقي فسوف يخصص للسنوات كما سبق.

٢- نقوم بتوقيع الخط البياني النسبي للصناعة الأولى ولتكن الصناعات الغذائية.

٣- عند توقيع الصناعة الثانية (الصناعات الخشبية) يتم جمع نسبة الأولى (الصناعات الغذائية) سنة ٧٩ + نسبة الثانية (الصناعات الخشبية) سنة ٧٩ ثم جمع نسبة الأولى سنة ٨٠+ نسبة الثانية سنة ٨٠ وهكذا لباقي السنوات .

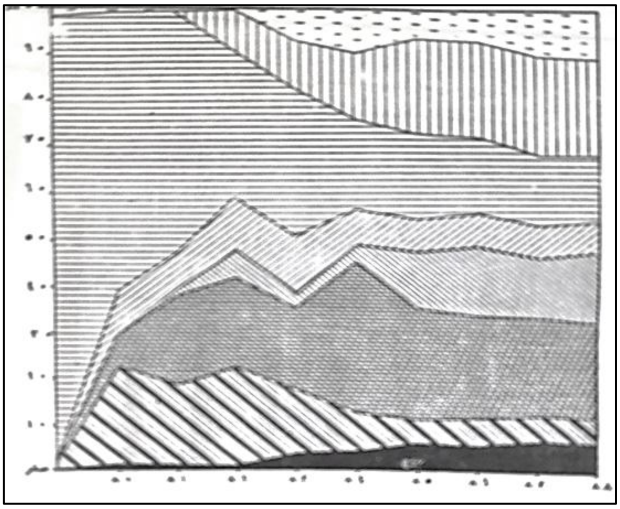
ثم يتم توصيل النقط المحددة للنسبة المجمعة بالخط المستقيم.

٣- عند توقيع الصناعة الثالثة يتم الحساب على النحو التالي:

تحديد نقطة عام ٧٩ للصناعة الثالثة = نسبة الأولى عام ٧٩ + نسبة الثانية

عام ٧٩ + نسبة الثالثة عام ٧٩ وهكذا لباقي السنوات ثم يصل بين النقط بالخط المستقيم وتستكمل الطريقة حتى آخر صناعة في الجدول .

وفي النهاية ينتج الشكل التالي : الذي يوضح الخط البياني النسبي المجمع لقيمة نسبة رأس المال المستثمر في كافة أنواع الصناعة بمدينة العاشر من رمضان خلال الفترة الممتدة بين عام ١٩٧٩ ، ١٩٨٨ .



٢- الأعمدة البيانية : (Bar Graphis)

هي أبسط طرق التمثيل البياني التي تستخدم للمقارنة بين الأرقام المطلقة أو النسبية لتغير ظاهرة خلال فترة زمنية ، أو لتبين تغير ظاهرة مقسمة إلى أكثر من قسم أو مجموعة، ولا تختلف عن الطريقة السابقة - الخط البياني - في توقيع قيم الظاهرة، إلا أنه في هذه الطريقة يستخدم أعمدة أو مستطيلات ذات عرض متساوي أو قاعدة متساوية وطول متناسب مع الكميات أو النسب حسب مقياس الرسم، ويمكن رسم الأعمدة البيانية أفقية أو رأسية حسب الهدف من الدراسة وتقسم الأعمدة البيانية تبعًا لما تظهره من معلومات إلى:

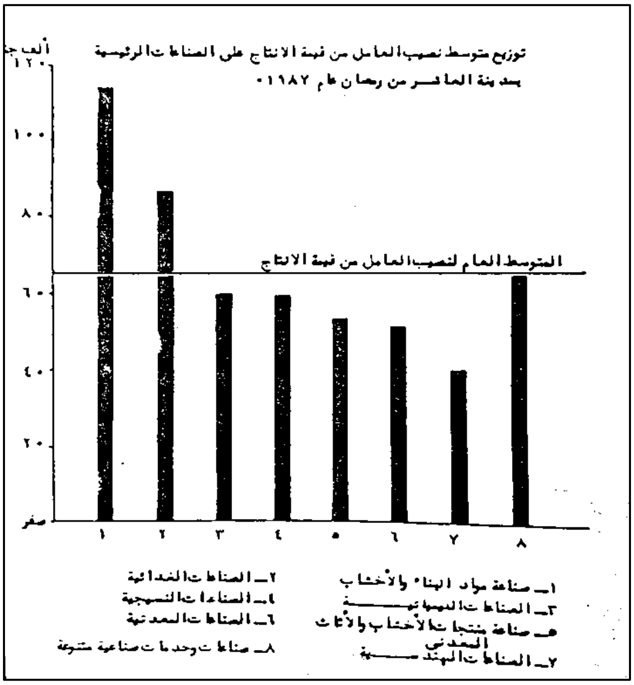
أعمدة بيانية بسيطة: ويستخدم هذا النوع لبيان تطور ظاهرة واحدة خلال فترة زمنية، أو للمقارنة بين مجموعة من القيم لظاهرة معينة .

مثال: يبين الجدول التالي متوسط نصيب العامل من قيمة الإنتاج الصناعي بمدينة العاشر من رمضان تبعًا لنوع الصناعة عام ١٩٨٧ .

نوع الصناعة	متوسط نصيب العامل من قيمة الإنتاج بالألف جنيه
الصناعات الكيميائية	٦.١١
صناعات مواد البناء والإنشاء	١١٤.٠٤
الصناعات النسيجية	٥٦.٠٧
الصناعات الغذائية	٨٦.٣٥

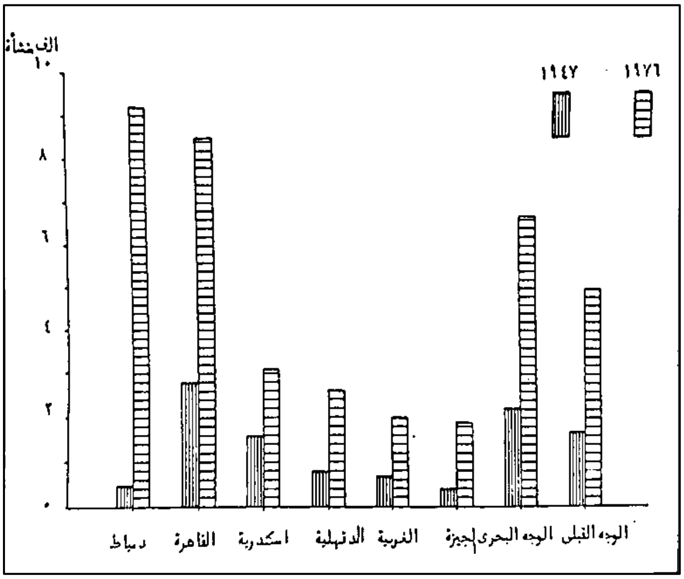
الصناعات الهندسية	٤٠.٣١
صناعة منتجات الأخشاب	٥٦.٠٨
الصناعات المعدنية	٥٠.٦٤
صناعات متنوعة	٦٥.٢٠
المتوسط العام للمدينة	٦٥.٥١

وباستخدام طريقة الأعمدة البيانية البسيطة لتوضيح الاختلافات بين مجموعات الدراسة (أنواع الصناعات) تم رسم الشكل التالي.

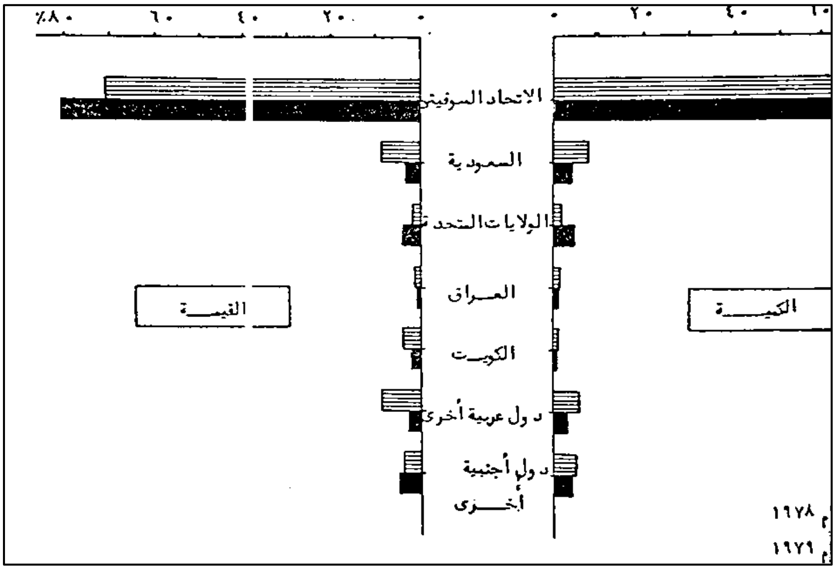


الأعمدة البيانية المتعددة: وهذا النوع من الأعمدة يستخدم اذا كانت الظاهرة المدروسة مكونة من أكثر من مجموعة أو أن الظاهرة لمنطقتين مختلفتين خلال فترة زمنية محددة، أو عند دراسة التغير في الظاهرة خلال تعدادين مختلفين - أو أكثر -

والشكل التالي يبين استخدام الأعمدة المتعددة لمقارنة أعداد منشآت الصناعات الخشبية خلال عام ١٩٧٦، ١٩٤٧ في بعض المناطق الرئيسية بجمهورية مصر العربية.

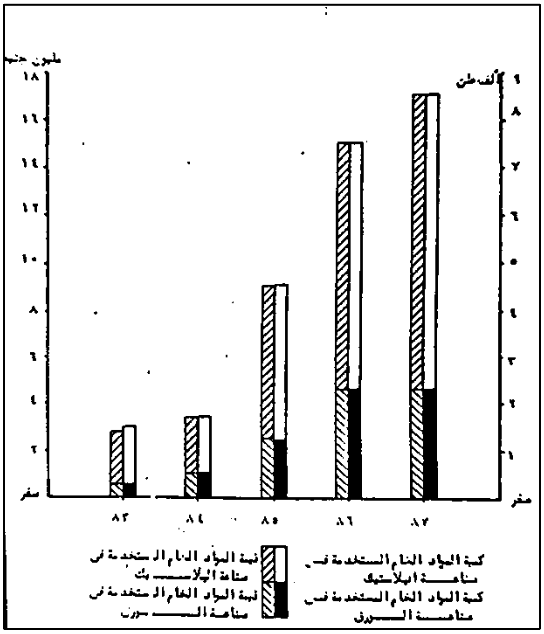


والشكل التالي يبين صورة أخرى من صور استخدام الأعمدة البيانية المتعددة . إذ يبين نسبة حركة الصادر من الأثاث الخشبي المصري موزعة على الأسواق الخارجية الرئيسية خلال عامي ١٩٧٨ ، ١٩٧٩ .



الأعمدة البيانية المجمعة: وتقوم فكرة تمثيل هذه الطريقة على أساس تجميع الأرقام المطلقة أو النسبية للظاهرة - مثل الخطوط البيانية المجمعة .

والشكل التالي يبين استخدام طريقة الأعمدة المجمعة لتوضيح التغيرات في كمية وقيمة المواد الخام المستخدمة في كل من صناعة الورق وصناعة البلاستيك خلال الفترة الممتدة بين عامي ١٩٨٣ ، ١٩٨٧ بمدينة العاشر من رمضان.

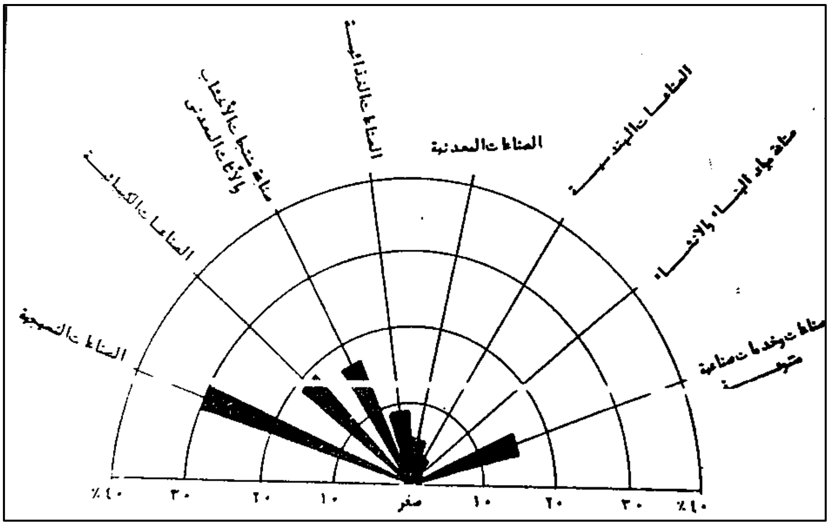


شكل يوضح تطور كمية وقيمة المواد الخام المستخدمة في صناعات كلا من الورق والبلاستيك خلال الفترة الممتد ة بين عامي ١٩٨٣ – ١٩٨٧.

ويلاحظ من الشكل أنه تم تنفيذه على أساس متغيرين (القيمة والكمية) لسهولة المقارنة بين سنوات الدراسة ، أي أنه يمثل الأعمدة المجمعة والمتعددة في نفس الوقت.

الأعمدة البيانية الدائرية : وهي عبارة عن شكل للأعمدة البيانية البسيطة إلا أن الفارق هو تحويل المحور الرأسي من خط مستقيم إلى نصف دائرة أو دائرة كاملة. تتباعد عن بعضها بمسافات متساوية تبعًا لمقياس الرسم المستخدم ، وتقسم الدائرة (٣٦٠°) أو نصف الدائرة (١٨٠°) إلى أقسام متساوية من الدرجات حسب عدد أقسام الظاهرة.

والشكل التالي يوضح التوزيع النسبي للكمية المستهلكة من الطاقة الكهربائية على منشآت الصناعات الرئيسية بمدينة العاشر من رمضان عام ١٩٨٧ باستخدام طريقة الأعمدة البيانية الدائرية.



ب- الطرق البيانية ذات البعدين (المساحية):

سبقت الإشارة إلى أن هناك طرق عديدة تستخدم في متطلبات البحث لإظهار الاختلافات في مكونات الظاهرة قيد الدراسة، ومن هذه الطرق: الدوائر النسبية والمربعات ....الخ. وسوف تقتصر الدراسة هنا على طريقة رسم الدوائر المقسمة .

التمثيل البياني باستخدام الدوائر المقسمة: Divided Circles

تفيد طريقة الدوائر المقسمة في إظهار الاختلافات النسبية بين العناصر المكونة للظاهرة المدروسة، وتتميز بسهولة تفسيرها إلا نها تحتاج لبعض العمليات الحسابية عند تنفيذها.

مثال: إذا أستخدمنا الأرقام المذكورة في الجدول السابق الذي يبين تطور قيمة رأس المال المستثمر في الصناعة بمدينة العاشر من رمضان خلال الفترة بين ٧٩، ١٩٨٨. وأردنا أن نستخدم طريقة الدور المقسمة لإظهار الاختلافات في التوزيع على مستوى نوع الصناعة لعام ١٩٨٨ فقط.

فنقوم بتحويل الأرقام المطلقة إلى درجات بنفس الأسلوب عند حسابنا النسبة المئوية ولكن هنا يستخدم درجات الدائرة فمن المعروف أن إجمالي درجات الدائرة ٣٦٠°

إذا الدرجة الخاصة بالصناعات الكيميائية مثلًا

$$= \frac{\text{قيمة رأس المال للصناعة عام } 198}{\text{إجمالي رأس المال المستثمر في الصناعات عام } 1988} \times ٣٦٠^\circ$$

أي $= \frac{63.302}{302.222} \times ٣٦٠^\circ = ٧٥.٤^\circ$

الدرجة الخاصة بالصناعات النسيجية $= \frac{61.828}{302.222} \times ٣٦٠^\circ = ٧٣.٦^\circ$

وهكذا حتى آخر نوع، ومن الضروري أن يكون المجموع الكلي للدرجات المحسوبة = ٣٦٠° (أي دائرة كاملة).

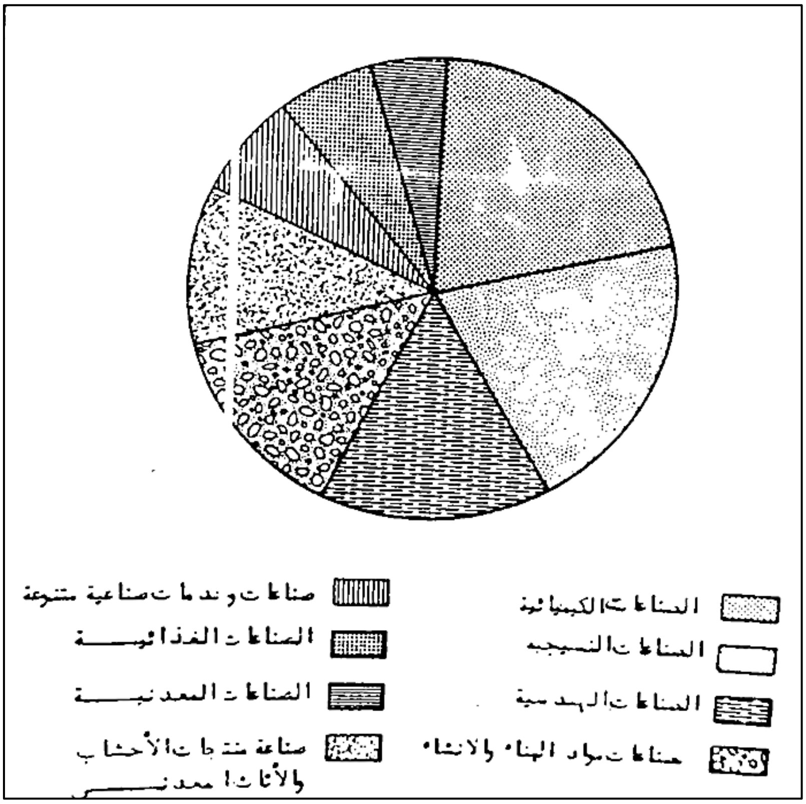
ثم نقوم برسم دائرة بنصف قطر مناسب، ومن أي اتجاه نرسم نصف قطر الدائرة (ويفضل الاتجاه الرأسي)، ونبدأ في قياس درجة أول صناعة ثم بتجميع درجة النوع الأول + درجة النوع الثاني لرسم الصناعة الثانية كما فعلنا في الخطوط البيانية المجمعة.

أي أن الجزء الخاص بالصناعة الأولى = ٧٥,٤ °

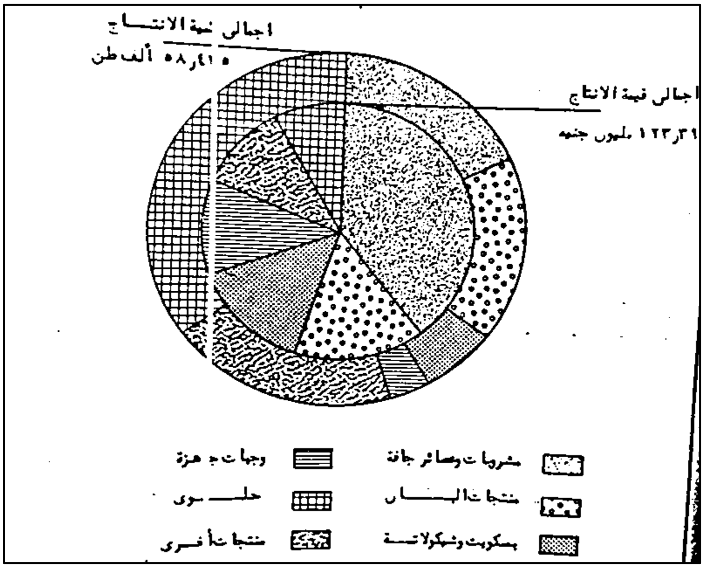
والجزء الخاص بالصناعة الثانية يكون عند =٧٥,٤ +٧٣,٦ = ١٩٤° وهكذا حتى الأخيرة =٣٦٠°.

ويجب على الباحث تحرى الدقة في توقيع الدرجات وخاصة اذا تعددت مكونات الظاهرة. ثم نقوم بعد ذلك بتظليل أو تلوين كل قسم من الأقسام حتى يظهر الاختلاف فيما بينها. ويوضع التظليل في مفتاح يفسر كل قسم من أقسام الظاهرة .

والشكل التالي يوضح التوزيع النسبي لقيمة رأس المال المستثمر في كافة الصناعات بمدينة العاشر من رمضان عام ١٩٨٨ باستخدام طريقة الدوائر المقسمة.



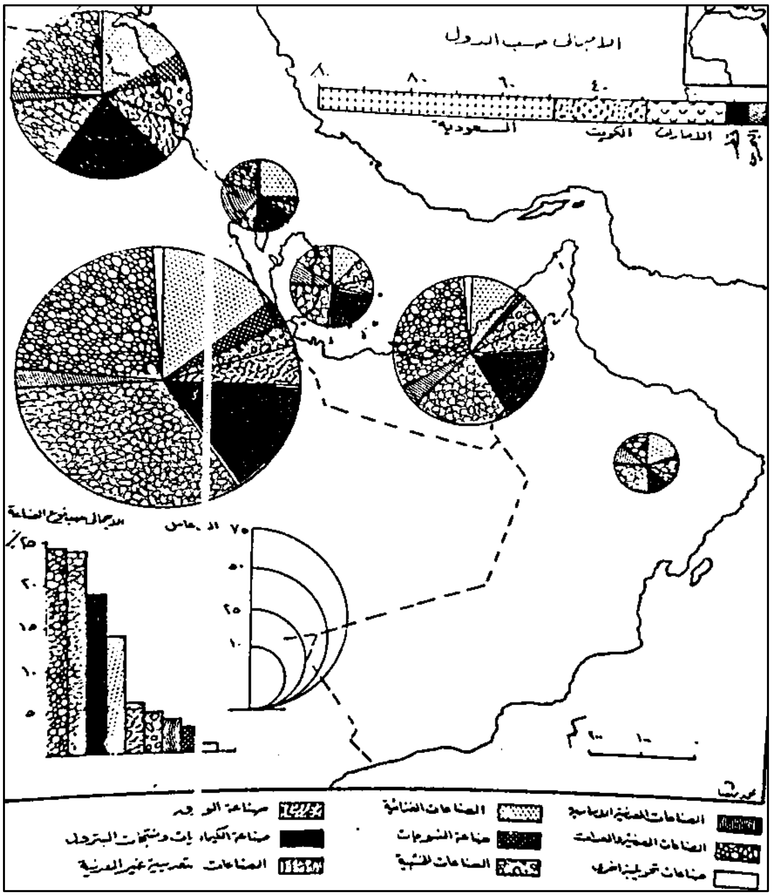
ويمكن تنفيذ أشكال عديدة باستخدام طريقة الدوائر المقسمة اذا دعت الحاجة إلى ذلك، وعلى سبيل المثال لا الحصر عند عقد مقارنة بين متغيرين لظاهرة واحدة مقسمة إلى أكثر من قسم، كما هو واضح من الشكل التالي والذي يوضح التوزيع النسبي لكمية وقيمة أهم منتجات الصناعات الغذائية بمدينة العاشر من رمضان عام ١٩٨٧ .



ومن الشكل نتبين أنه على الرغم من أن منتجات المشروبات والعصائر تستأثر بأكثر من ١/٣ القيمة الإجمالي لإنتاج الصناعات الغذائية إلا أن كمية إنتاجها أقل من ٢٥ ٪ من إجمالي الكمية المنتجة. وهذا يؤخذ كمؤشر لارتفاع تكاليف منتجات المشروبات والعصائر، وعلى العكس من ذلك في منتجات الحلوى التي تستأثر بنحو ٣٣ ٪ من الكمية المنتجة لكافة الصناعات الغذائية إلا أن قيمتها تنخفض عن هذه النسبة بكثير وهذا يدل على انخفاض قيمة تكاليف المنتج منها.

والجدير بالذكر أنه يمكن استخدام جميع طرق التمثيل السابقة لرسم بعض خرائط التوزيعات الكمية، والتي تساعد في عقد مقارنات على مستوى التوزيع الجغرافي لمتغير واحد متعدد الأقسام، أو عند المقارنة بين عدة متغيرات.

والشكل التالي يوضح التوزيع الجغرافي للعاملين بالصناعات التحويلية حسب نوع الصناعة بدول مجلس التعاون لدول الخليج العربية عام ١٩٨٧



يلاحظ من الشكل أنه تم استخدام طريقة التمثيل بالأعمدة النسبية المتعددة والمجمعة وطريقة التمثيل بالدوائر المقسمة النسبية والتي تتباين - الدوائر - في مساحتها حسب القيمة الإجمالية لكل واحدة منها.

ثانيًا: الطرق البيانية التحليلية:

تستخدم هذه الطرق لبيان المعدل والاتجاه الصحيحين لتغير الظاهرة، وهناك العديد من هذه الطرق اشهرها المنحني اللوغاريتمي ومنحنى لورنز.

أ- المنحني اللوغاريتمى: Logarithmic Graphs

يفضل دائمًا استخدام هذا النوع من التمثيل عند دراسة التغيرات في الظاهرات البشرية أو الاقتصادية خلال فترة زمنية معينة حيث باستخدامه يتضح لنا الاتجاه الصحيح لتغير الظاهرة، ومن مميزاته أيضًا إمكانية استخدامه اذا كانت التغيرات في الظاهرة متباينة بشكل واضح من عام لآخر ومن فترة إلى أخرى (أو بمعنى آخر إذ كان المدي بين أرقام أو قيم الظاهرة كبير جدا).

والمنحني اللوغاريتمي على نوعين:

- منحنى لوغاريتمى كامل: ويتم تنفيذه على أساس تقسيم المحورين (س، ص) إلى أقسام أو دورات لوغاريتمية ، ويفضل استخدامه عند دراسة الاتجاه الصحيح لظاهرة ذات بيانات زوجية (مزدوجة القيم).

- منحنى نصف لوغاريتمي: وينفذ هذا المنحني على أساس أن يكون أحد المحورين مقسم إلى أقسام أو دورات لوغاريتمية والمحور الآخر مقسم إلى أقسام متساوية تبعًا للغرض الذي سوف يظهره.

ويمكن استخدام المنحني اللوغاريتمي لدراسة تطور متغير واحد، أو عند دراسة الظاهرات المكونة من عدة مجموعات أو أقسام ، وفيما يلي عرض لكيفية إنشاء **المنحنى النصف لوغاريتمى.**

مثال: إذا استخدمنا بيانات الجدول السابق الخاص بتطور قيمة رأس المال المستثمر في الصناعات المختلفة بمدينة العاشر من رمضان خلال الفترة بين عامي ١٩٧٩، ١٩٨٨. لتمثيلها بطريقة المنحني النصف لوغاريتمى نتبع الآتي:

- نقوم أولًا بإلقاء نظرة سريعة على الجدول لتحديد المدى الذي تقع فيه أرقام الإحصائية، فأصغر رقم (٠.١٦ مليون جنيه) خاص بالصناعات المتنوعة، وأكبر رقم ١٦,٣٠٢ مليون جنيه للصناعات الكيميائية ، فإذا أردنا توقيع هذه الأرقام باستخدام المنحنى البياني المتعدد سوف تكن هناك صعوبة في إظهار جميع أرقام الإحصائية نظرًا لتباينها الشديد، ولهذا يفضل استخدام المنحنى النصف لوغاريتمي في هذه الحالة.

- ثانيًا: نقوم بتحديد عدد الدورات أو الأقسام اللوغاريتمية على النحو التالي عدد الدورات اللوغاريتمية = عدد أرقام أكبر رقم - عدد أرقام أصغر رقم

عدد أرقام أكبر رقم (٦٣,٣٠٢) مكون من خمس أرقام ، عدد أرقام أصغر رقم (٠.١٦) مكون من رقمين

∴عدد الدورات = ٥ - ٢ = ٣ (ثلاث دورات لوغاريتمية)

ثالثًا: يختار اتساع مناسب للدورة أو مقياس رسم الدورة، و ليكن ٤ سم

∴طول المحور الرأسي = عدد الدورات X مقياس الرسم للدورة

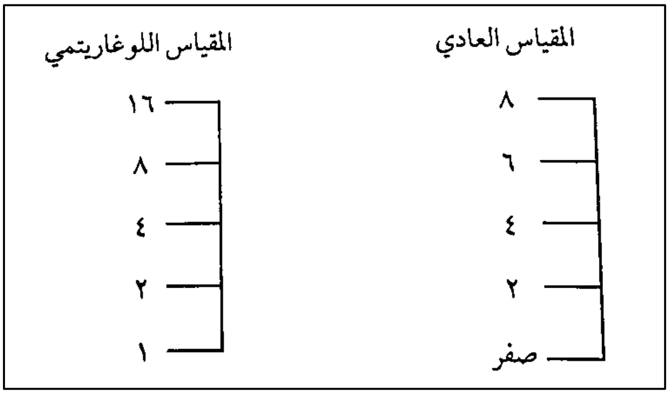
= ٣×٤ = ١٢ سم.

والدورة اللوغاريتمية الأولى تبدأ من (١) وتنتهى عند (١٠)

والدورة الثانية تبدأ من (١٠) وتنتهي عند (١٠٠) والثالثة تبدأ من (١٠٠) وتنتهى عند (١٠٠٠) وهكذا..

وهذا يعني أن الدورة اللوغاريتمية لا تبدأ بالصفر لأن لوغاريتم صفر =ε ولوغاريتم ١ = صفر.

والتقسيم اللوغاريتمي يبدأ القياس فيه من أي رقم خلاف الصفر؛ لأن وجود الصفر في مقام أي نسبة معناه رياضية أن هذه النسبة تساوي ما لا نهاية . أي أنه لا يمكن قياس التغير النسبي من أساس مقداره صفر بل لا بد من أن يكون الأساس عدد حقيقي صحيح؛ لأن القياس اللوغاريتمي قياس نسبي دائمًا.



شكل يوضح المقياس العادي والمقياس اللوغاريتمي

- ثم نقوم بتقسيم الدورة اللوغاريتمية أو المحور الرأس إلى أقسام تختلف باختلاف لوغاريتمات الأعداد .

فمثلا عند توقع القيمة ١ على المحور الرأسي

= لو ٢ × اتساع الدورة = ٠,٣٠ × ٤ = ١,٢ سم

أي أن القيمة ٢ تبعد عن الرقم واحد صحيح بمقدار ٢، ١سم

ثم القيمة ٣ = لو ٣× ٤ = ٠,٤٨ × ٤= ١,٩١ سم

وهكذا حتى نهاية الدورة الأولى. ثم نقوم برسم خطوط موازية للمحور الأفقي تبعًا للمسافات التي سبق وأن حصلنا عليها وبطول يساوى طول المحور الأفقي المقسم إلى أقسام متساوية والذي يوضح الزمن أو السنوات.

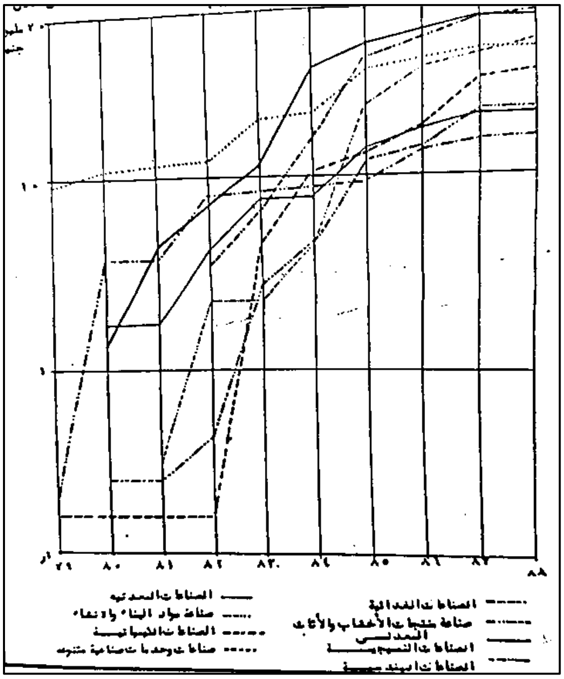
ونكرر العمل السابق لرسم الدورة الثانية والثالثة ، والملاحظ أن الدورة الأولى في الشكل والتي تقل عن الواحد صحيح تم تقسيمها أيضًا بنفس الأسلوب بدون استخدام اللوغاريتمات السالبة للأعداد أقل من ١ صحيح أي أن المسافة بين ٠,١ ، و ٠,٢. على المحور الرأسي = تساوي المسافة بين واحد صحيح و (٢) على نفس المحور.

كما أن المسافة بين ١٠، و ٢٠ = المسافة بين ١٠٠، ٢٠٠

- بعد تصميم الشكل اللوغاريتمي السابق نبدأ في توقيع كل قيمة للسنة الواحدة باستخدام (لوغاريتم القيمة × اتساع الدورة) أي أن بعد الرقم أو القيمة عن المحور الأفقي =

لوغاريتم القيمة X اتساع الدورة .

ثم نصل بين نقط كل صناعة – أو متغير - بخط مستقيم ينتج الخط أو المنحني الخاص بالصناعة وهكذا في باقي الصناعات مع رسم مفتاح خاص يبين أشكال الخطوط المختلفة لأنواع الصناعات المدروسة.



شكل يوضح تطور قيمة رأس المال المستثمر في الصناعات المختلفة بمدينة العاشر من رمضان خلال الفترة (١٩٧٩- ١٩٨٨)

ب- منحنى لورنز: Lorenz Curve

يستخدم منحنى لورنز عند تحديد الشكل العام لاتجاه توزيع الظاهرة قيد الدراسة والتي تتصف بازدواجية الأرقام، أي لتحديد الشكل العام لمتغيرين من حيث مدى التعادل أو التماثل في توزيع الظاهرة وتقوم فكرة المنحنى علي أساس :

١- رسم شكل مربع تمامًا بطول ضلع مناسب حسب مقياس الرسم المستخدم ويقسم كل محور (س، ص) إلى أقسام من مئوية متساوية (كل١٠ ٪ مثلا).

٢- تحويل الأرقام المزدوجة للمتغيرين إلى نسبة مئوية.

٣- رسم خط يصل بين نقطة الأصل والنقطة المقابلة لها ويسمى هذا الحظ باسم خط التعادل أو التماثل الذي من أي نقطة تقع عليه تكون النسبة المناظرة لها على المحور الأفقي = النسبة المناظرة لها على المحور الرأسي.

٤- يتم توقيع نقط المتغير قيد الدراسة على أساس التكرر المتجمع النسبي الصاعد حيث أن كل زوجين من التكرارات المتجمعة تقع في فئة معينة (س، ص).

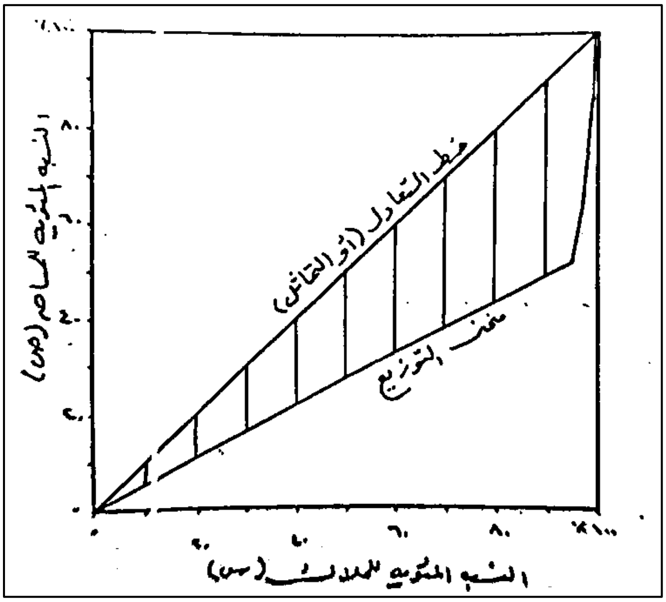
فإذا وقعت قيم الظاهرة على خط التعادل دل ذلك، على أن الظاهرة تتوزع بصورة معتدلة وكلما بعدت القيم (المنحنى المرسوم) عن خط التعادل دل ذلك على ابتعاد الظاهرة عن التماثل أو التوزيع المتساوي.

واستخدام منحنى لورنز كان في بادئ الأمر قاصرًا على الدراسات الاقتصادية الخاصة بتوزيع الدخل القومي أو المحلي على السكان . ويستخدمه الآن كثير من الباحثين في أغراض مختلفة والتي منها دراسة : توزيع حجم الملكية الزراعية في أي منطقة أو إقليم.

مثال : يوضح الجدول التالي توزيع الملكية في جمهورية مصر العربية عام ١٩٨٤

فئات حجم الملكية بالفدان	الملاك		المساحة		التكرار المتجمع الصاعد	
	العدد بالألف	%	بالألف فدان	%	للملاك	للمساحة
أقل من ٥	٣٢٨٨	٩٥.٣	٢٩٠٤	٥٣.٥	٩٥.٣	٥٣.٥
٥-	٨٦	٢.٥	٥٧٦	١٠.٦	٩٧.٨	٦٤.١
١٠-	٤٦	١.٣	٥٨٩	١٠.٨	٩٩.١	٧٤.٩
٢٠-	٢٣	٠.٦	٦٢١	١١.٤	٩٩.٧	٨٦.٣
٥٠-	٦	٠.٢	٤٠٧	٧.٥	٩٩.٩	٩٣.٨
١٠٠ فدان +	٢	٠.١	٣٣٥	٦.٢	١٠٠	١٠٠
الإجمالي	٣٤٥١	١٠٠	٥٤٣٢	١٠٠	-	-

وباستخدام النسبة المئوية المتجمعة (التكرار النسبي المتجمع الصاعد) نقوم بتوقيع أزواج القيم النسبية المتجمعة لكل فئة باستخدام س ، ص (كما سبق) وينتج لدينا الشكل التالي:



ومن الشكل يتضح لنا منحنى توزيع الظاهرة بعيدًا عن خط التعادل وهذا يعني أن مساحة الأرض الزراعية في مصر تتوزع بصورة غير متعادلة أو غير متساوية على الملاك ، والجدير بالذكر أنه يمكن تمثيل أو رسم اكثر من منحنى على شكل واحد - أو في داخل شكل واحد - لعقد مقارنات بين المناطق المختلفة من حيث مدى التعادل في توزيع الملكية الزراعية ، كما يمكن استخدام منحنى لورنز عند عقد المقارنات للمتغيرات التي تتصف بنفس الخصائص السابقة.

ثالثًا: التمثل البياني للتوزيعات التكرارية:

ذكرنا فيما سبق أن المنهج الكمي يستخدم الجداول الإحصائية والأرقام بغرض تلخيص ووصف الظاهرة التي تهم الباحث، والجداول الإحصائية بالرغم من دقتها في تلخيص الظواهر وعرض ما بينها من علاقات إلا أنها لا تستهوي القارئ العادي، وفي كثير من الأحيان لا يستطيع قراءتها أو استقراءها واستيعابها إلا القارئ المتخصص؛ لذلك كان العرض البياني أو التصويري أكثر جاذبية وسهولة ؛ فعند دراسة درجة الحرارة في مختلف أشهر السنة في منطقة ما نجد أنه من الأيسر فهمها إذا وضعت بشكل بياني كالمدرج أو المضلع أو المنحنى التكراري، وكذلك الحال بالنسبة لكمية المطر، أو حركة المرور في مختلف ساعات النهار، أو توزيع السكان على فئات السن المختلفة .

إن التمثيل البياني لمثل هذه الظواهر وغيرها من الظواهر الجغرافية يسهل فهمها واستيعابها وقراءتها .

مقارنة الظواهر الجغرافية بيانيًا: إذا أريد مقارنة ظاهرتين مختلفتين، أو ظاهرة واحدة في تاريخين مختلفين يمثل توزيع كل منهما بمدرج تكراري أو مضلع تكراري، ومن ثم تسهل المقارنة بين التوزيعين من حيث شكل التوزيع وتماثله أو بعده عن التماثل ومن حيث درجة انتشار مفردات كل من التوزيعين .

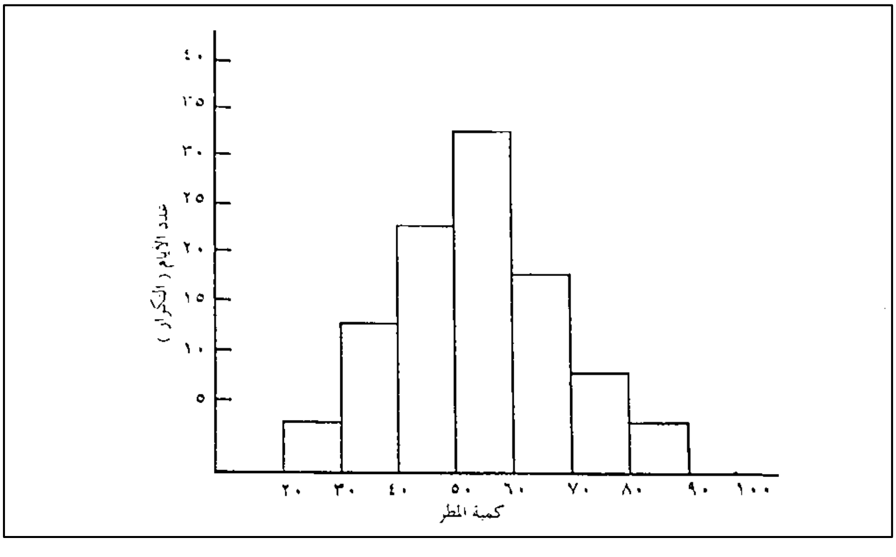
فمثلًا عند مقارنة المواليد بالوفيات في بلد ما خلال فترة معينة أو مقارنة الوفيات في فئات السن المختلفة في بلدين أو أكثر، أو مقارنة الواردات مع الصادرات خلال فترة معينة ، أو أجور العمال في مصنعين مختلفين ... إلخ، فإننا نقوم برسم المدرج أو المضلع التكراري لهاتين الظاهرتين، وطبيعي أنه في حالة استخدام شكل المدرج التكراري للمقارنة، فإنه يفضل رسم مدرجين تكراريين في اتجاهين : يخصص أحدهما لأحد التوزيعين، ويخصص الثاني للتوزيع الآخر، إذ أن رسم المدرجين في اتجاه واحد يؤدي إلى تداخل الشكلين ومن ثم يصعب مقارنة التوزيعين .

أما إذا استخدم شكل المضلع التكراري للمقارنة فإنه لا حاجة لرسم مضلعين في اتجاهين مختلفين ، وإنما يكتفي بتمييز كل من المضلعين بإحدى طرق الرسم المناسبة وبذلك تتيسر عملية المقارنة، ومن ثم نستطيع القول بأننا إذا أردنا مقارنة ثلاثة توزيعات أو أكثر فإنه من الأفضل في هذه الحالة استخدام شكل المضلع التكراري أو المنحنى التكراري، إذ تتعذر المقارنة البصرية بين مدرجات تكرارية في أكثر من اتجاهين .

هذا وإذا اختلف المجموع الكلي للتكرارات في التوزيعين بشكل واضح فإنه يفضل استخدام النسب المئوية بدلا من التكرارات، أي نستخدم التوزيعات التكرارية النسبية مع تخصيص المحور الرأسي في الرسم للتكرارات النسبية .

المدرج التكراري Histogram:

هو عبارة عن مجموعة من الأعمدة المتلاصقة يمثل ارتفاع كل منها تكرارات معينة لفئة معينة ، وذلك في حالة تساوي فئات التوزيع . فعند تمثيل توزيع كمية الأمطار الساقطة خلال ١٠٠ يوم حسب فئات كمية المطر من واقع البيانات الموجودة في الجدول بواسطة مدرج تكراري ينتج لدينا الشكل التالي الذي يمكن رسمه عن طريق رسم محورين متعامدين، و نأخذ المحور الأفقي - عادة - لتمثيل الفئات، والمحور الرأسي لتمثيل التكرارات، ونقوم بتدريج المحور الرأسي حسب مقياس رسم مناسب بحيث يسمح بظهور قيمة أكبر تكرار في الجدول، ويلزم في هذه الحالة أن يبدأ التدريج من الصفر ، أما المحور الأفقي فلا يلزم بالضرورة أن يبدأ من نقطة الصفر .



شكل يوضح المدرج التكراري لتوزيع كمية الأمطار

إن أطوال الأعمدة في المدرج التكراري المبين تتناسب مباشرة مع قيمة التكرارات، وذلك فقط لتساوي أطوال الفئات في هذا التوزيع، إذ أن الأساس في رسم المدرج أن تتناسب مساحات الأعمدة فيه مع قيمة التكرارات، ولما كانت مساحة المستطيل هي عبارة عن حاصل ضرب القاعدة في الارتفاع ، فإن تساوي القاعدة بين هذه المستطيلات المتعاقبة يجعل في الإمكان أن تتناسب التكرارات مع ارتفاعات هذه المستطيلات.

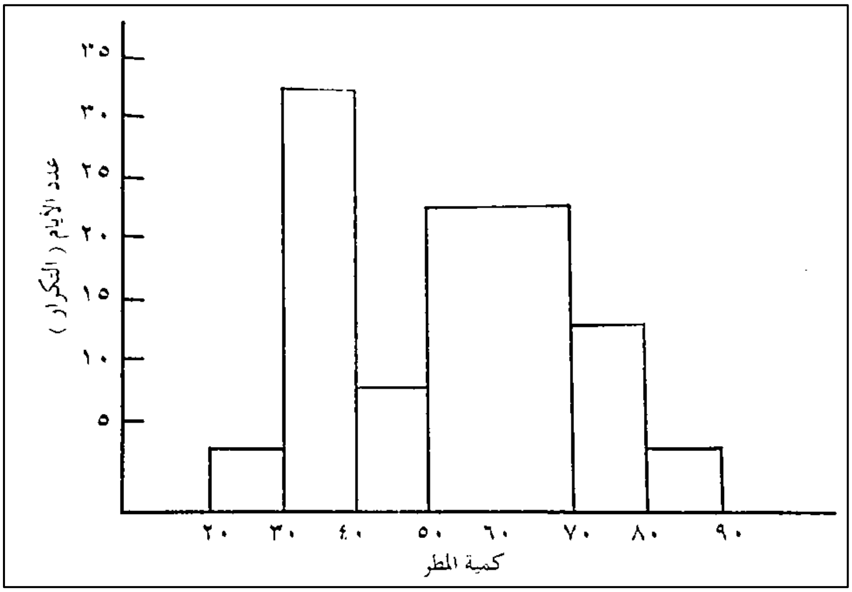
أما إذا لم تكن الفئات متساوية الطول فلا يصح لنا في هذه الحالة أن نرسم على الفئات مستطيلات تتناسب ارتفاعاتها مع التكرارات، كما هو الحال في الفئات المتساوية ، ولا بد لنا من تعديل التكرارات بحيث تتناسب ارتفاعات المستطيلات مع التكرارات المعدلة وتتناسب مساحة هذه المستطيلات مع التكرارات الأصلية ، ويتم هذا بقسمة التكرار الأصلي على طول الفئة ، وتسمى هذه العملية : تعديل التكرارات.

فإذا نحن عدنا إلى مثال كمية الأمطار الساقطة على مدينة ما نفسه ووجدنا التوزيع التكراري على الصورة التالية الواردة في الجدول، ومن الواضح أن أرقام العمود الخامس في الجدول المذكور هي التكرارات المعدلة الواجب تثبيتها على المحور الرأسي كما في الشكل التالي:

جدول حساب التوزيع التكراري المعدل

الفئات (أ)	**التكرارات الأصلية (٢)**	**أطوال الفئات(٣)**	**التكرارات المعدلة** (٤)= (٢)÷(٣)	**ملاءمة التكرارات لمقتضيات الرسم**
٢٠-	٤	١٠	٠.٤	٤
٣٠-	١٥	٥	٣.٠	٣٠
٣٥-	٢١	١٥	١.٤	١٤
٥٠-	٤٠	٢٠	٢.٠	٢٠
٧٠-	١٦	١٠	١.٦	١٦
٨٠-٩٠	٤	١٠	٠.٤	٤
المجموع	١٠٠			

نضرب الأرقام في (١٠) لإزالة العلمة العشرية وذلك تبسيطا لعملية الرسم ، وهذه العملية لا تؤثر كما هو واضح فيما هو كائن بين التكرارات من علاقات تناسبية بعضها ببعض.



شكل يوضح الجدول التكراري المعدل

إن التكرارات المعدلة لا يقتصر حسابها على المدرج التكراري بل يجب حسابها في حالة تمثيل التوزيعات ذات الفئات غير المتساوية بمضلع أو منحني أيضًا على النحو الذي سيأتي. وواضح أنه إذا كان التوزيع التكراري مفتوحًا من أحد الطرفين أو من كليهما فإنه لا يمكن تمثيله بيانيًا برسم المدرج التكراري، إذ إن طول إحدى الفئتين أو كليهما ليس معروفة ؛ ولهذا لو اضطررنا إلى الرسم نهمل عادة مثل هذه الفئات المفتوحة على أن نشير إليها في أسفل الرسم.

المضلع التكراري Frequency Polygon :

إذا أردنا تمثيل توزيعين تكراريين بيانيًا على المحور نفسه عن طريق مدرجيهما التكراريين بقصد إجراء المقارنة بينهما، فإننا نجد أن المستطيلات المتناظرة تتداخل مع بعضها البعض مما يصعب معه إجراء المقارنة والتمييز بين التوزيعين ؛ لذلك فإننا نلجأ إلى تمثيل البيانات بما يسمى بالمضلع التكراري.

ونحصل عليه بتقسيم المحورين كما في حالة المدرج التكراري تمامًا ، ثم ننصف كل فئة في نقطة تسمى مركز الفئة ، ومركز الفئة : هو النقطة الوسطى التي تقع على بعدين متساويين من بداية الفئة ونهايتها، ولتعيين مركز الفئة نتبع أي من القواعد التالية :

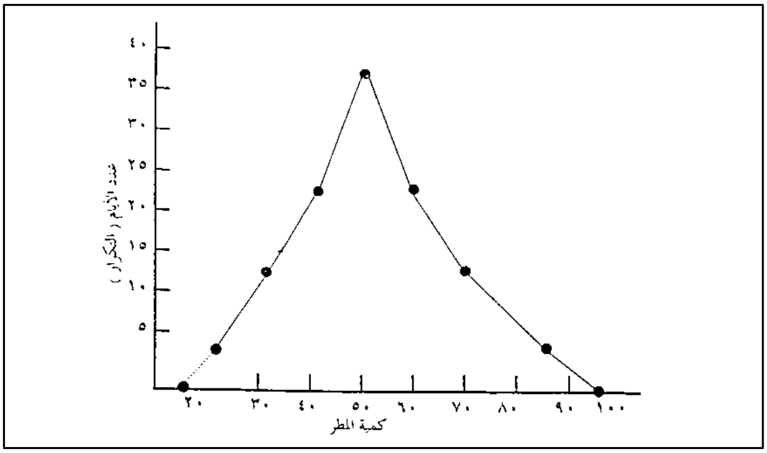
مركز الفئة = الحد الأدنى للفئة + الحد الأعلى للفئة.

٢

= الحد الأدنى للفئة + نصف طول الفئة

= الحد الأعلى للفئة - نصف طول الفئة

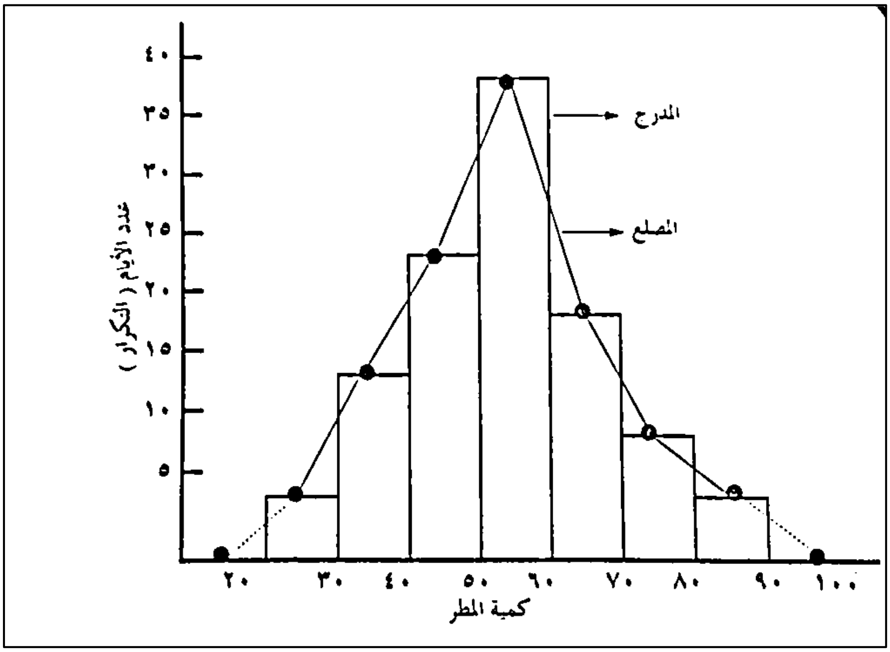
ونحن إذا افترضنا أن مركز كل فئة يصلح لتمثيل الفئة، أي إننا نستطيع بأن نفترض أن تكرار كل فئة مركز عند مركزها، وعليه فإذا رسمنا محورين متعامدين خصصنا الأفقي منهما لمراكز الفئات، والرأسي للتكرارات، ثم قمنا برصد التكرارات المناظرة لمراكز الفئات، ثم وصلنا النقط الناتجة بخطوط مستقيمة فكان الشكل الناتج هو المعروف باسم : المضلع التكراري كما في الشكل التالي:



شكل يوضح المضلع التكراري لتوزيع كمية الأمطار

ويحسن إقفال المضلع التكراري مع المحور الأفقي، وذلك بأخذ فئة سابقة للفئة الأولى ومساوية لها في الطول، وفئة لاحقة للفئة الأخيرة ومساوية لها في الطول، وتكرار كل منهما صفر ، ونحدد مركز كل منهما ونصلهما بطرفي المضلع فيتم إقفاله.

ويمكن رسم المضلع التكراري من المدرج التكراري، وذلك بتنصيف قمم المستطيلات وتوصيل منتصفات هذه القمم بخطوط مستقيمة، مع إقفال الشكل عند بدايته ونهايته بالطريقة السابقة نفسها، ويلاحظ أن المساحة المحصورة أسفل المضلع التكراري تساوي تمام المساحة المحصورة تحت المدرج التكراري.

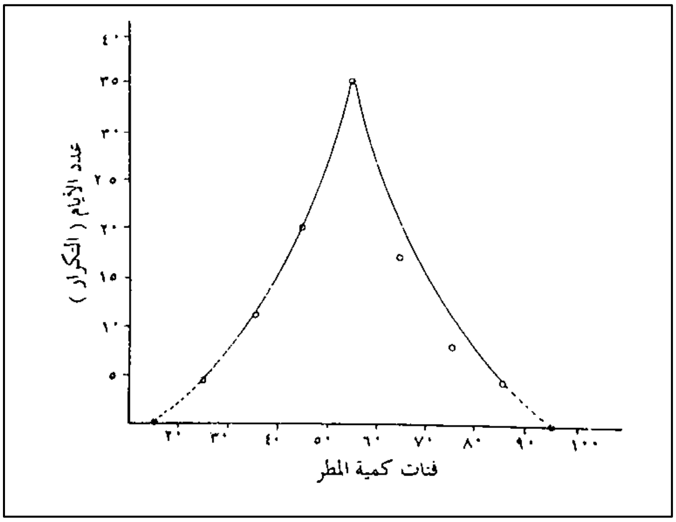


شكل يوضح المدرج والمضلع التكراري لتوزيع كمية المطر

وفي حالة الفئات غير المتساوية، تتبع طريقة التكرارات المعدلة بدلا من التكرارات الأصلية على المحور الرأسي، وذلك بعد تعديل التكرارات، كما سبق شرحه في حالة المدرج التكراري.

المنحنى التكراري Frequency Curve :

بتمثيل مراكز الفئات على المحور الأفقي، والتكرارات على المحور الرأسي نعين عددًا من النقط تمثل رؤوس المضلع التكراري، ونحاول أن نرسم منحنى ممهد يمر بمعظم النقط ويتوسط بقيمتها خير توسط ، ويسمى الشكل الناتج : بالمنحنى التكراري، والملاحظة أنه نتيجة لعملية التمهيد، فإن المساحة الواقعة تحت المنحنى لا تساوي بالضرورة المساحة الواقعة تحت المضلع التكراري .



شكل يوضح المنحنى التكراري لتوزيع كمية الأمطار

إن شكل المنحنى يتوقف على التوزيع التكراري الذي يمثله ، ولقد سبق أن بينا كيفية استخدام المضلع التكراري لمقارنة توزيعين تكراريين أو أكثر وإن كنا لم نتعرض بعد إلى الأسس التي ستتم على أساسها المقارنة بين التوزيعات التكرارية المختلفة، وسنستخدم هنا خصائص أربعة محددة يمكن على أساسها تعريف التوزيعات التكرارية، والتفريق بينها، وهذه الخصائص هي:

١- القيمة الوسطى أو القيمة المركزية Central Location:

وتشير إلى قيمة محددة تقع عند وسط التوزيع ، والمقاييس التي تستخدم لحساب هذه القيمة تعرف بالمتوسطات Averages ومنها الوسط الحسابي، والوسيط، والمنوال .

٢- التشتت أو الاختلاف Variation :

ونعني بذلك درجة تركز أو عدم تركز القيم المختلفة للظاهرة المدروسة حول قيمتها الوسطى، وكلما ازداد تركز المفردات حول القيمة الوسطى كلما قل تشتتها ، وكلما كانت مفردات الظاهرة أكثر تجانسًا، وهناك عدد من مقاييس التشتت مثل : المدي، والانحراف المتوسط، والانحراف المعياري .

٣- الالتواء Skewness :

وتشير هذه الخاصية إلى مدى تماثل التوزيع أو بعده عن التماثل والتوزيع المتماثل Symmetrical هو التوزيع الذي تقسمه القيمة الوسطى إلى قسمين متطابقين، ويكون تزايد أو تناقص التكرارات متشابهًا ومنتظمًا على جانبي المحور المقام عند وسط التوزيع .

أما التوزيعات غير المتماثلة Asymmetrical Distributions فهي التي تتزايد أو تتناقص فيها التكرارات بشكل غير منتظم على جانبي المحور المقام عند وسط التوزيع، وإذا كانت التكرارات الكبيرة فيه تميل إلى التركز عند فئاته الدنيا، قيل: إن هذا التوزيع غير متماثل، وبالتالي فإن الظاهرة التي يمثلها هذا التوزيع موجبة الالتواء Positive Skewness، وعلى العكس إذا كانت التكرارات الكبيرة تميل إلى التركيز عند فئات التوزيع العليا فإن هذا التوزيع يكون سالب الالتواء Negative Skewness .

٤- التفرطع Kurtosis:

ونعني بذلك مدى اختلاف التوزيع التكراري للظاهرة عن التوزيع المعروف : بالتوزيع العادي Normal Distributions ، فإذا كان التوزيع التكراري أكثر تحدبًا عند قمته أو قيمته المركزية وكانت تلك القيمة أعلى منها للتوزيع المعتاد قيل : إن هذا التوزيع مدببًا Leptokurtic، وعلى العكس إذا كانت قيمة التوزيع التكراري للظاهرة أكثر استقامة وأدنى قيمة منها للتوزيع العادي قيل : إن هذا التوزيع مفرطحًا Platykurtis

سنستخدم المنحنى التكراري كشكل بياني لعرض نماذج مختلفة من التوزيعات التكرارية التي تختلف فيما بينها على أساس خاصية أو أكثر من هذه الخصائص الأربعة

جدول يوضح توزيعات تكرارية فرضية

١الفئات	٢متماثل	٣مدبب	٤مفرطح	٥ذوقمتين	٦نوني	٧موجب الالتواء	٨سالب الالتواء	٩أسي موجب	١٠أسي سالب
صفر-	١	٣	٥	٥	٣٠	١٠	٢	٢	٥٠
١٠-	٧	٨	١٤	١٠	٢٠	٢٥	٦	٤	٣٠
٢٠-	٢١	١٣	٢٠	٣٥	١٠	٤٠	١٠	٥	٢٠
٣٠-	٣٥	٤٠	٢٥	١٤	٤	٢٠	١٥	٧	١٠
٤٠-	٣٥	٤٠	٢٥	١٤	٤	١٥	٢٠	١٠	٧
٥٠-	٢١	١٣	٢٠	٣٥	١٠	١٠	٤٠	٢٠	٥
٦٠-	٧	٨	١٤	١٠	٢٠	٦	٢٥	٣٠	٤
٧٠-	١	٣	٥	٥	٣٠	٢	١٠	٥٠	٢
المجموع	١٢٨	١٢٨	١٢٨	١٢٨	١٢٨	١٢٨	١٢٨	١٢٨	١٢٨

يحتوي الجدول السابق على مجموعة من البيانات الفرضية التي تمثل أنواع مختلفة من التوزيعات التكرارية . وبمقارنة هذه التوزيعات يتضح لنا ما يلي:

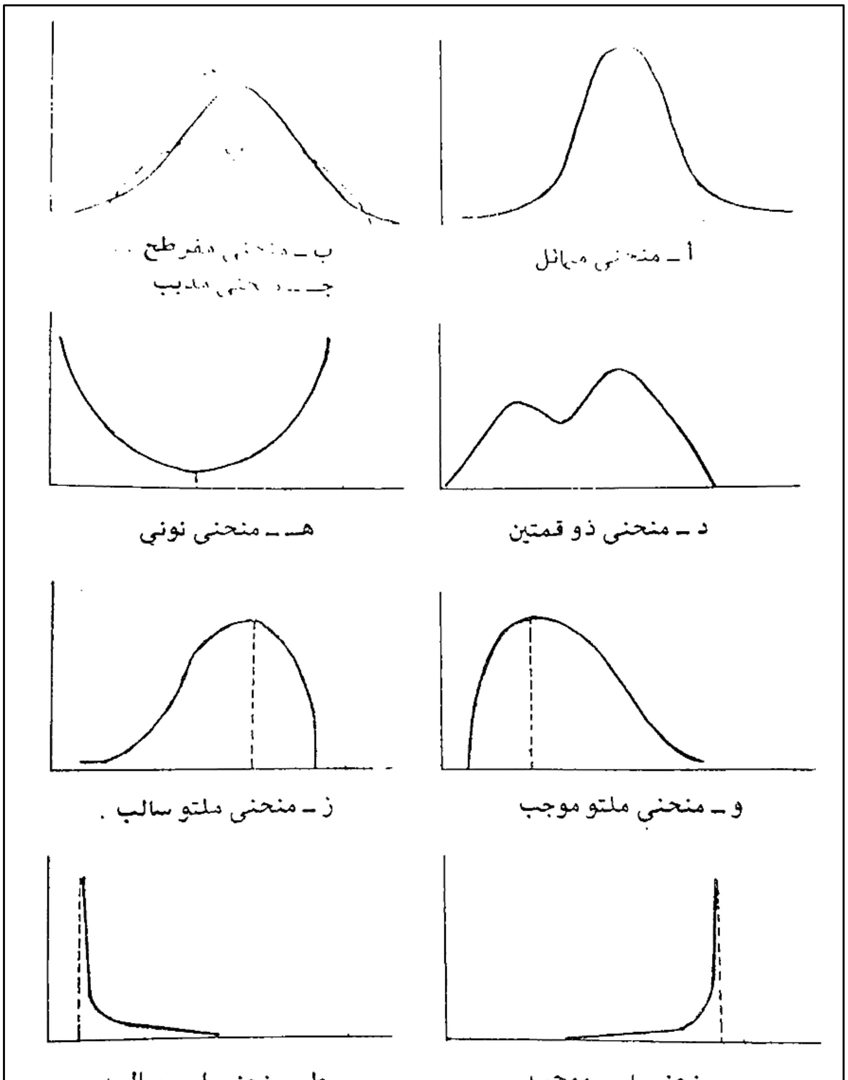
١- إن التوزيع الذي أدرجت تكراراته بالعمود رقم (٢) هو توزيع متماثل؛ لأن توزيع التكرارات منتظمة على جانبي المحور المار من وسط المنحنى. أما التوزيع الذي أدرجت تكراراته بالعمود رقم (٣) فله تكرارات مركزية أكبر من التوزيع السابق. ويزداد تركز التكرارات قريبًا من الفئات الوسطى؛ لذلك فهو توزيع مدبب، في حين أن التوزيع الذي أدرجت تكراراته بالعمود رقم (٤) تقل تكراراته المركزية عن تلك التي في العمود (٢) وتنتشر تكراراته على مدى أكبر حول الفئات الوسطى؛ لذلك فهو توزيع مفرطح، ومن

المعلوم أن الأشكال الثلاثة السابقة هي متماثلة في الشكل حول وسطها، ويبين الشكل التالي بعض هذه التوزيعات التكرارية وهي كالتالي :

(أ) هو لتوزيع متماثل من النوع المعروف بالتوزيع العادي Normal Distribution

أما التوزيع (ب) فهو أكثر استواء عند قمته من التوزيع العادي، وهو مثال لتوزيع مفرطح

وأما التوزيع الثالث (ج) فقمته أكثر تحدبًا من التوزيع العادي (أ) فهو توزيع مدبب .



٢- إن التوزيع الذي أدرجت تكراراته بالعمود رقم (٥) من الجدول توزيع له قمتان : إحداهما داخل الفئة ٢٠ - والثانية داخل الفئة ٥٠ - وتوزيع كهذا له أكثر من قمة في منحناه التكراري يدل على عدم تجانس مفردات الظاهرة المدروسة، أي أن البيانات قد تمثل ظاهرتين متداخلتي التوزيع.

مثال ذلك : أننا لو قمنا بدراسة أجور العمال في مصنع وكانت أجور العمال الذكور تتركز حول ٥٠ جنيه بينما تتركز أجور العمال الإناث حول ٣٠ جنيه ، فإذا ما رسمنا المنحنى لمثل هذا التوزيع نجد له قمتين : إحداهما بالقرب من ٥٠، والأخرى بالقرب من ٣٠؛ وذلك لعدم تجانس الأجور كلها، ويمكن في هذه الحالة فصل التوزيع إلى توزيعين مختلفين (انظر شكل د).

٣- إن العمود رقم (٦) يبين توزيعًا تتركز فيه التكرارات الكبيرة عند فئاته الدنيا و العليا أي عند طرفيه، وأما تكراراته المركزية وهي التي تتوسط التوزيع فهي أقلها. ويعرف هذا المنحنى بالمنحنى النوني، أو على شكل حرف U - Shape, U Curve ، وهو على عكس المنحني المعتدل (انظر شكل هـ)

إن توزيع الوفيات حسب السن يكون قريبًا جدًا من هذا الشكل. فمن المعروف أن عدد المتوفين من الأطفال كبير جدًا، وكذلك عدد المتوفين من الشيوخ، وأما المتوفون من الشبان ذوي الأعمار المتوسطة فعددهم قليل ؛ ولهذا فيكون توزيع الوفيات مشابهة للمنحنى النوني.

٤- إن التوزيع المبين تكراراته في العمود رقم (٧) تزداد تكراراته عند فئاته الدنيا وتقل عند فئاته العليا، ويعرف هذا التوزيع بأنه موجب الالتواء، وعلى العكس فإن التوزيع الذي تزداد تكراراته عند فئاته العليا كما هو مبين بالعمود رقم (٨) وتقل عند حدوده الدنيا فهو توزيع سالب الالتواء

و إذا نظرنا إلى الشكل (أ) تبين لنا أن التوزيع (أ) متماثل حول قمته الوسطى، في حين (و)، (ز) غير متماثلين في التوزيع ، فإذا كان الفرع الأيمن للمنحني أطول من الأيسر، كالمنحنى (و) سمي المنحنى موجب الالتواء أو ملتوية جهة اليمين، وفي هذه الحالة تكون تكرارات القيم الصغيرة كثيرة بينما تكرارات القيم الكبيرة قليلة . وعلى العكس نجد أن المنحنى (ز) الفرع الأيسر فيه أطول من الأيمن، أي أن تكرارات القيم الصغيرة قليلة بينما تكرارات القيم الكبيرة كثيرة ؛ ولذا يسمى منحنى سالب الالتواء، أو ملتوية جهة اليسار.

ومن أمثلة المنحنيات الملتوية ، المنحنيات التكرارية التي تمثل دخل الفرد في أي دولة غالبية أفرادها فقراء، وفي هذه الحالة فالالتواء موجب لتركز التكرارات في الفئات

الدنيا، أو غالبية أفرادها أغنياء، وفي هذه الحالة فالالتواء سالب لتركز التكرارات في الفئات العليا .

٥- إن العمودين ٩، ١٠ في الجدول يظهران تكرارات التوزيعين شديدي الالتواء، وفيهما تكون أكبر التكرارات عند أحد طرفي التوزيع، وتقل التكرارات في اتجاه الطرف الآخر، وبذلك يكون لها فرع واحد ؛ (ولهذا تسمى أحيانا بالمنحنيات ذات الفرع الواحد) وقد يصعد المنحني من اليسار إلى اليمين فتكون القيم الصغيرة قليلة، والقيم الكبيرة كثيرة، والمنحنى في هذه الحالة يشبه حرف (ر)؛ ولهذا يسمى: بالمنحنى الرائي J - Shaped Curve (عمود ٩)، وقد ينزل المنحني من اليسار إلى اليمين حيث تكون التكرارات كبيرة عند القيم الصغيرة، وصغيرة عند القيم الكبيرة، وشكل هذا المنحنى عكس المنحني السابق ؛ ولهذا فيسمى أحيانا : (منحنى رائي مقلوب).

وهذان النوعان يسميان بالمنحنى الأسى Exponential الموجب والسالب على الترتيب انظر شكل (ح، ط)، وهذه المنحنيات توجد في ميادين مختلفة في الحياة العملية . ففي النواحي الاقتصادية نجده يمثل توزيع الثروة في المجتمع مثل : توزيع الملكية، توزيع الدخل، توزيع الضرائب.. إلخ. ففي توزيع المجتمع مثلا يزداد عدد الملاك الزراعيين ذوي الملكيات الصغيرة ويقل عددهم عند فئات الملكية الزراعية الكبيرة .إن المنحنى التكراري لمثل هذه الظاهرة يمثل توزيع أسي سالب.

المنحنيات التكرارية المتجمعة Cummulative Frequency Curves:

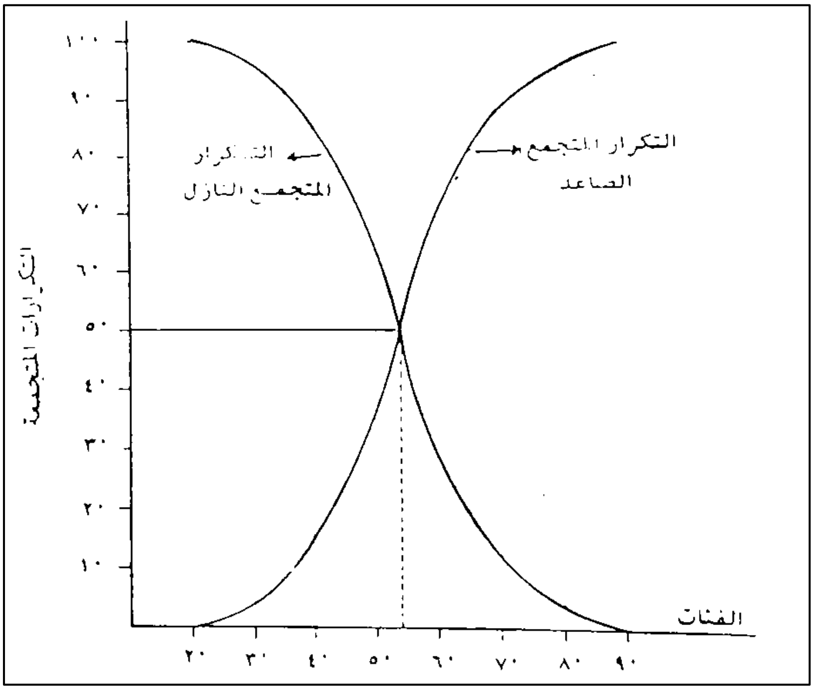
تتطلب بعض التحليلات تحويل أرقام جداول التوزيع التكراري إلى أرقام متجمعة تصاعديًا أو تنازليًا، فينشأ عن ذلك جدول التوزيع التكراري المتجمع الصاعد، ويمكن تمثيله بالمنحنى المتجمع الصاعد Ascending Ogive أو جدول التوزيع التكراري المتجمع النازل ويمكن تمثيله بالمنحنى المتجمع النازل Descending Ogive.

فإذا أردنا عمل الجداول التكرارية المتجمعة للبيانات المتعلقة بكميات الأمطار الساقطة على مدينة ما، نقوم بعمل جدولين على النمط الآتي من واقع التوزيع التكراري البسيط .

جدول التوزيعات التكرارية البسيطة والمتجمعة لكميات الأمطار الساقطة على مدينة ما

جدول التوزيع		**جدول التوزيع التكراري المتجمع**			
الفئات	**التكرارات**	**الصاعدة**		**النازلة**	
		الحدود الدنيا للفئات	**تكرارات صاعدة**	**الحدود العليا للفئات**	**تكرارات نازلة**
٢٠-	٤	أقل من ٢٠	٠٠	٢٠ فأكثر	١٠٠
٣٠-	١١	أقل من ٣٠	٤	٣٠ فأكثر	٩٦
٤٠-	٢٠	أقل من ٤٠	١٥	٤٠ فأكثر	٨٥
٥٠-	٣٦	أقل من ٥٠	٣٥	٥٠ فأكثر	٦٥
٦٠-	١٧	أقل من ٦٠	٧١	٦٠ فأكثر	٢٩
٧٠-	٨	أقل من ٧٠	٨٨	٧٠ فأكثر	١٢
٨٠-٩٠	٤	أقل من ٨٠	٩٦	٨٠ فأكثر	٤
		أقل من ٩٠	١٠٠	٩٠ فأكثر	٠٠
المجموع	١٠٠				

ويمكن تمثيل أي من هذين التوزيعين المتجمعين ببساطة في الرسم ، وذلك أن يخصص المحور الأفقي للفئات كما هو الحال في التوزيع التكراري، وكذلك الإحداثي الرأسي للتكرارات بحيث يتسع لمجموع التكرارات ثم نرصد النقط على الرسم كالمعتاد أمام الحدود العليا أو الدنيا للفئات حسب نوع المنحني الذي نريد الحصول عليه صاعدًا أو نازلًا ، ونصل بين هذه النقط بمنحنى ممهد فنحصل على المنحنى المتجمع الصاعد والنازل على النحو الموجود في الشكل ونلاحظ أن نغلق الرسم في طرفي المنحنى وذلك على اعتبار أن التكرار يساوي صفرًا لكل من الفئتين أقل من ٢٠ (في حالة المنحنى الصاعد) و ٩٠ فأكثر (في حالة المنحنى النازل) ومعنى هذا أن المنحنى الصاعد لا بد أن يلامس المحور الأفقي عند ٢٠، وأن المنحنى النازل يلامسه عند ٩٠.



شكل يوضح التكرار المتجمع الصاعد والنازل لكمية الأمطار الساقطة في ١٠٠ يوم

ولمقارنة منحنيين متجمعين على الرسم نفسه لا يشترط أن تكون الفئات متساوية في التوزيعين على غير الحال عند مقارنة مضلعين تكراريين ولكن يشترط تحويل القيم إلى نسب مئوية ؛ وذلك لإزالة أثر الاختلاف في المجموع الكلي بتوحيده في التوزيعين ؛ وذلك بمساواته بمائة .

استخدامات المنحنى المتجمع الصاعد أو النازل:

١- الحصول على تقديرات معينة: من فوائد المنحنى المتجمع الصاعد والنازل في أي توزيع أنه يمكن معرفة عدد التكرارات التي تقل قيمتها عن حد معين في حالة المنحني الصاعد أو تزيد عن قيمة معينة في حالة المنحنى النازل. فإذا أردنا معرفة عدد الأيام التي سقطت فيها كمية من المطر أقل من ٤٣ ملم، وهي قيمة ليست في الجدول الأصلي، أقمنا عمود على المحور الأفقي عند ٤٢ إلى أن يلتقي بالمنحنى الصاعد في نقطة نرسم منها عمودا على المحور الرأسي ونقرأ التكرار المواجه لهذه النقطة على هذا المحور . وكذلك يمكن بالطريقة نفسها وباستخدام المنحنى النازل معرفة عدد الأيام التي سقطت فيها كمية من المطر أكثر من ٤٢ ملم. ويمكن استخدام المنحنيين بطريقة عكسية لمعرفة كمية المطر التي سقطت في عدد معين من الأيام، وفي هذه الحالة نقيم العمود من المحور الرأسي عند التكرار الذي نختاره فيقابل المنحني عند نقطة نسقط منها عمودا على المحور الأفقي فتدلنا على كمية المطر التي نبحث عنها.

٢- إيجاد الوسيط : إذا رسمنا المنحنيين التكراريين المتجمعين الصاعد والنازل في شكل واحد وبمقياس الرسم نفسه على المحورين نفسيهما فإنهما يتلاقيان في نقطة يكون بعدها عن المحور الرأسي يساوي نصف مجموع التكرارات كلها، أما بعدها عن المحور الأفقي فيساوي قيمة الوسيط .

٣-إيجاد منحنى لورنز: و هو المنحني الذي يبين مدى العدالة أو المساواة في توزيع بعض الظواهر الاقتصادية ، ويمكن إيجاد منحنى لورنز باستخدام المنحنى المتجمع الصاعد .

الهرم السكاني:

يحتاج الباحث في الجغرافيا البشرية إلى تمثيل التوزيع العمري لسكان بلد ما أو منطقة جغرافية معينة بحسب النوع ؛ وذلك إما لمقارنة الاختلاف في التوزيع العمري بين الذكور والإناث للبلد نفسه أو المنطقة الجغرافية، أو لمقارنته بالتوزيع العمري لسكان بلد آخر أو منطقة جغرافية أخرى. ويستخدم الشكل المعروف بالهرم السكاني لهذا الغرض، وذلك بأن يخصص المحور الرأسي لبيان فئات العمر، وأما المحور الأفقي على يمين ويسار المحور الرأسي فيخصص لبيان عدد أو نسبة الأفراد في كل فئة عمرية، ويخصص أحد جانبي الشكل للذكور والجانب الآخر للإناث ، وبإقامة مستطيلات متلاصقة على جانبي المحور الرأسي المخصص لفئات الأعمار المختلفة يكون الشكل الناتج هو الهرم السكاني الذي لا يخرج عن كونه مدرجين تكرارين متقابلين عند قاعدتيهما مع ملاحظة أن طول كل من هذه المستطيلات يمثل الذكور والإناث أو نسبة الذكور والإناث.

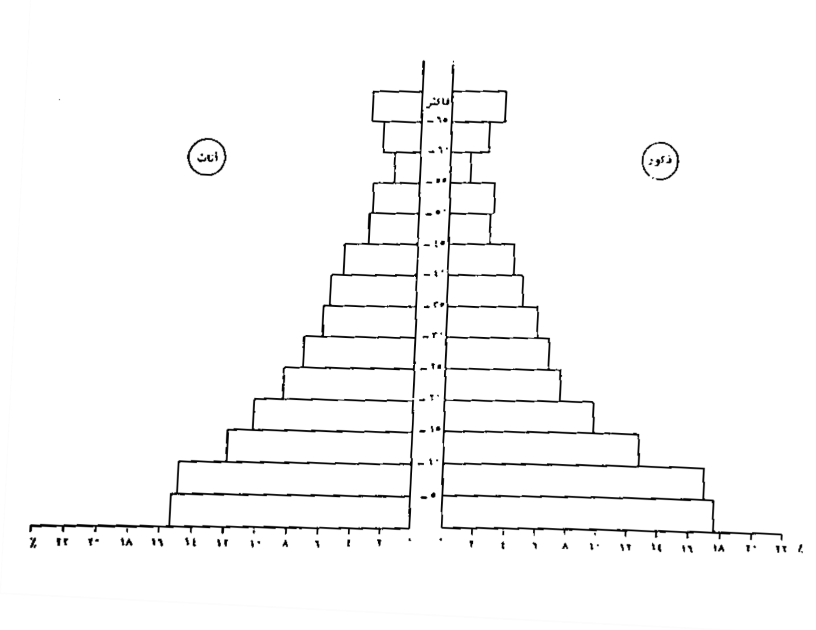
وبطبيعة الحال فإننا نستطيع استخدام التكرارات أي أعداد الذكور والإناث في كل فئة عمرية، كما نستطيع استخدام النسب المئوية بدلا من التكرارات، مما يسهل عملية المقارنة خاصة بين سكان دول مختلفة أو مناطق جغرافية مختلفة. والمثال التالي يبين استخدام شكل الهرم السكاني التمثيل التوزيع العمري لمنطقة مكة المكرمة عام ١٩٧٤م، وكذلك دراسة هذا التوزيع بالنسبة للنوع.

يبين الجدول التالي فئات السن المختلفة مع ما يقابلها من تكرارات للذكور والإناث، مع تحويل الأرقام الفعلية إلى نسب مئوية لتسهيل عملية المقارنة.

فئات السن	الذكور	الإناث	% ذكور	% إناث
٠٠-٤	١٤٦٤٣٠	١٤٢١٦٣	١٥.٤	١٧.٦
٥-٩	١٤٢٤٠٢	١٣٦٦٠٠	١٥	١٦.٩
١٠-١٤	١١٤٤٨٨	١٠٢١٦٢	١٢	١٢.٧
١٥-١٩	٩٧٩١١	٧٩١٠٤	١٠.٣	٩.٨
٢٠-٢٤	٨٢٢٦٥	٦٠٤٩٩	٨.٦	٧.٥
٢٥-٢٩	٧٠٤٧١	٥٥٠٧٣	٧.٤	٦.٨
٣٠-٣٤	٥٩٤٢٩	٤٨٧٠٢	٦.٢	٦
٣٥-٣٩	٥٤٧٨٨	٤١٠٠٧	٥.٨	٥
٤٠-٤٤	٤٦٣١٧	٣٥٧٢٩	٤.٩	٤.٤
٤٥-٤٩	٣٣٣٧٦	٢٢٠٥٠	٣.٥	٢.٧
٥٠-٥٤	٣٠٢٢٦	٢٤٣٤٨	٣.٢	٣
٥٥-٥٩	١٧٣٢٧	١١١٦٩	١.٨	١.٥
٦٠-٦٤	٢٤٧٦٧	١٩٨٥٦	٢.٥	٢.٥
٦٥- فأكثر	٣٢٢١٣	٢٩٠٢٠	٣.٤	٣.٦
غير مبينين	٢٥٤	٧٠	٠.٠٣	٠.٠٠٨
الإجمالي	٩٥٢٦٦٤	٨٠٧٥٥٢	١٠٠%	١٠٠%

وقد رسمت المعلومات الواردة في هذا الجدول بالشكل التالي الذي خصص فيه الجانب الأيمن من المحور الأفقي لنسب الذكور، والجانب الأيسر لنسب الإناث، كما خصص المحور الرأسي لفئات السن المختلفة ، ويلاحظ من الشكل أنه عبارة عن شكل هرمي تتناقص أطواله ودرجاته تدريجيًا تجاه فئات العمر العليا، وأن معدل التناقص يختلف باختلاف معدلات الوفاة من فئة عمر إلى فئة العمر التالية، ويلاحظ على الشكل أن قاعدته عريضة ، وهذا يعني وجود نسبة كبيرة من السكان في سن الشباب ، كما يلاحظ أيضًا وجود

عدم انتظام في بعض الفئات، وعادة تفسر هذه الملاحظة إما بسبب أخطاء في التبليغ عن الأعمار أو عوامل أخرى تؤثر على عدد السكان كالحروب والأوبئة الشديدة والهجرة، أو تغيرات ملموسة في المعدلات الحيوية (معدلات المواليد ومعدلات الوفيات).



شكل يوضح الهرم السكاني لمنطقة مكة المكرمة

المراجع

- سمير محمد على حسن ، الإحصاء في الجغرافيا ، جامعة الخرطوم ، ٢٠١٢.
- عيسى على إبراهيم ، الأساليب الإحصائية والجغرافيا ، دار المعرفة الجامعية ، الإسكندرية ، ٢٠١٧.
- فتحي محمد أبو عيانة ، مدخل إلى التحليل الإحصائي ، دار المعرفة الجامعية ، الإسكندرية ، ١٩٨٧.
- فتحي عبدالله فياض ، التحليل الإحصائي للبيانات الجغرافية ، دار الفكر العربي ، ١٩٩١.
- محمد خميس الزوكة ، ومحمد إبراهيم رمضان ، الإحصاء والأساليب الكمية في العلوم الإنسانية ، دار المعرفة الجامعية ، ٢٠٠٠.
- ناصر عبدالله الصالح ، ومحمد محمود السرياني ، الجغرافيا الكمية والإحصائية أسس وتطبيقات بالأساليب الحاسوبية الحديثة ، مكتبة العبيكان، الرياض ،٢٠٠٠.
- ياسر أحمد السيد ، الإحصاء التطبيقي لوصف وتحليل وتفسير الظاهرات الجغرافية ، مكتبة بستان المعرفة ، كفر الدوار ، ٢٠٠٨.