جامعة الانبار

كلية التربية للعلوم الانسانية

قسم الجغرافية

مقرر الاحصاء الجغرافي

المرحلة الثالثة

العلم : هو النشاط الذي يحاول جمع الصور التي يمتلكها الناس عن الحقائق ويقربها الى الحقيقة.

اما الادب والفن : هما النشاط الانساني الذي يخلق صورة جديدة عن الحقيقة وعلى هذا الاساس فان الجغرافية تصنف على انها نشاطاً علميا وليس نشاطا ادبيا فوظيفة الجغرافي هي تجميع الصور الموجودة عن الحقيقة وتقريبها الى الحقيقة لذلك فأن تعريفها يكون.

الجغرافية : هو العلم الذي يبحث في جميع الظواهر الطبيعية والبشرية والاقتصادية الناتجة من تفاعل الزمان والمكان على سطح الارض وتفاعل الانسان مع هذه الظواهر والاهتمام بالتوزيع الجغرافي لها و تأثيرها على الانشطة البشرية و تأثرها بالانشطة البشرية.

وان كلمة احصاء تعنى لنا واحدة من اربع مسميات هي:

١-علم الاحصاء : هو فرع من فروع علم الرياضيات الا انه استقل عنه وطور نفسه بصورة متميزة .

٢- عملية جمع البيانات والمعلومات .

٣-طريقة لجمع ومعالجة البيانات الرقمية وتغيرها .

٤-العمليات الحسابية التي تجري على الارقام .

اما الاحصاء الجغرافي : فهو الطريقة لجمع ومعاملة البيانات من خلال العمليات الحسابية اي هو علم الجداول التي تصف ظاهرة معينة والرسوم والاشكال البيانية التي تعرض التغيرات لظاهرة معينة في مدة زمنية معينة ، مع هذا فأن اعتماد الطرق الاحصائية لا يعني ان البحث او الدراسة قد اخذت الصفة العلمية وانها تأخذ هذه الصفة فقط عندما تستخدم الطرق الاحصائية بصوره صحيحة وعندما تفسر النتائج وفق فلسفة الجغرافية ومنطقها بصورة صحيحة.

فان الاحصاء ليس بديلا عن التفكير النظري ولا هو بديل عن عملية الاختبار والفحص فهو طريقة من الطرق وليس كل الطرق وليس كل الطرق التحليلية وهو مكمل للطريقة الوصفية .

الاسباب التي دعت الى استخدام الاحصاء في الجغرافية .

لما كانت الجغرافية معنية بدراسة الانسان والبيئة التي يعيش فيها فأنها وبدون استثناء لا فرع من فروعها تعتمد على الارقام والمعلومات الرقمية عند شرح الخصائص المكانية والعلاقات المتداخلة للظواهر الجغرافية وان استخدام الجغرافيين للإحصاء ليس حديثاً انما بدأت موجة استخدام الطرق الكمية منذ ستينيات القرن الماضي .

ويمكن بيان الاسباب التي دعت الجغرافيين الى استخدام الطرق الاحصائية بما يأتي:

١-تسمى الجغرافية علم التصانيف ويوفر علم الاحصاء طرقاً عديدة للقيام بمهمة التصنيف فقد استفاد الجغرافيون من مختلف الطرق الاحصائية والرياضية لتقسم العالم الى اقاليم مناخية ومجموعات واصناف .

٢-في كثير من الاحيان لا يستطيع الجغرافي تغطية منطقة دراسته كلياً لذا يلجأ الى استخدام العينات في البحث وهنا فانه سيحتاج الى الطرق الاحصائية لتعميم الانماط المستخرجة من انماط جزئية الى انماط شمولية .

٣-الجغرافي معني بصورة مباشرة بملاحظة البيئة التي يعيش وينشط عليها سواء كانت البيئة طبيعية او اقتصادية او اجتماعية وغالباً ما تقوده ملاحظاته الى فرضيات وهذه الفرضيات تحتاج الى اختبار ولا يتم هذا علميا الا من خلال اعتماد طرق احصائية معينة .

٤-يسعى الجغرافيون الى وضع نظريات في تخصصهم كباقي العلوم الاخرى واختبارها ولا يمكن ان يتم الا من خلال اعتماد طرق احصائية .

٥- عند دراسة الواقع الجغرافي وحركته ليس هناك شيء ثابت او شيء محدد وانما احتمالية الحدوث واحتمالية التغيير شيء متوقع وهذه النقطة مهمه جعلت الجغرافيين يميلون الى استخدام الاحصاء .

٦- عندما يدرس الجغرافي الظواهر فانه يميل الى دراسة العلاقات والعوامل التي تشكل الظاهرة والعوامل التي تؤثر عليها لذلك فانه يتجه الى معرفة التباين بين الظواهر او بين الاماكن وليس بإمكانه القيام بمثل هذه الدراسات من دون استخدام الطرق الاحصائية .

٧- عندما يريد الجغرافي وصف ظاهرة معينة او اقليم محدود فانه بالتأكيد سيحتاج الى اعتماد ارقام تمثل الظاهرة حيث يعتمد الطرق الاحصائية لوصف وتمثيل مسوحات التمركز والتشتت .

<u>يمكن تقسيم الاحصاء الى قسمين رئيسيين هما:</u>

١- الاحصاء الوصفي : ويمثل الطرق الرقمية والحسابية لجمع البيانات والمعلومات وتنظيمها واختصارها ثم عرض المعلومات عن طريق الجداول والرسوم البيانية والخرائط ومن ثم تفسير وتحليل النتائج .

٢- الاحصاء الاستدلالي : يعني تحليل البيانات المتوفرة في العينة كأساس لتحليل البيانات الموجودة في المجتمع الى اساليب التقدير والاختبار ومن ثم اتخاذ القرار .

البيانات الجغرافية . (طبيعتها ، انواعها ، مصادرها)

تتميز الجغرافية بتنوع بياناتها وبتعدد مصادر هذه البيانات والسبب في ذلك يعود الى الحيز الكبير الذي تهتم به الجغرافية والى تنوع المفردات الطبيعية والبشرية الموجودة داخل هذا الحيز فقد اشتركت الجغرافية مع العلوم الاخرى في البيانات والمعلومات فلما كانت الجغرافية تهتم بالطبيعة ومشكلاتها فقد ادى الى استفادة الجغرافية من البيانات والمعلومات التي تتوفر للعاملين في مجال الجيولوجيا والتربة والانواء الجوية وغيرها من العلوم الطبيعية . والبيانات التي تتوفر للعاملين في مجال دراسات الانسان كعلم الاجتماع والاقتصاد والسياسية والتخطيط وغيرها من العلوم البشرية .

وعندما كانت الجغرافية تتميز عن العلوم الاخرى في انها تدرس المكان ويذهب البعض لتسميتها بعلم المكان . وتجتهد في دراسة اثر التغير المكاني على الظواهر الجغرافية المدروسة لذا فان اختلاف المكان يعد سببا اخر في تنوع البيانات الجغرافية .

ولما كانت الجغرافية باحثة في اثر تفاعل المكان والزمان فان التغير الزماني هو الاخر يقود الى تنوع البيانات الجغرافية ولكن بدرجة متفاوتة اذ تعتبر الظواهر الطبيعية بطيئة التغير قياسياً بالظواهر البشرية والاقتصادية دائمة التغير .

اصناف وانواع البيانات الجغرافية :

تصنف البيانات الجغرافية وفقاً لاعتبارات مختلفة .

على اساس حالة التجمع	البيانات المنفردة
	البيانات المتجمعة (المبوبة)
على اساس النوع	البيانات العددية الحقيقية
	البيانات النسبية
	البيانات الاسمية
	البيانات الرتبية

أ- على اساس حالة التجميع:

١-بيانات منفردة : ويقصد بها البيانات التي تم الحصول عليها بشكلها الخام سواء من الدراسة المكتبية او من الدراسة الميدانية وهذه البيانات ترتبط القيم فيها بمتغير واحد او خاصية واحدة من خواص المتغير وتمتاز هذه البيانات بدقة القيم فيها .

٢-البيانات المتجمعة (المبوبة) : وتشمل البيانات المنفردة التي اجريت عليها عملية التصنيف والجدولة وفقاً لاعتبارات جغرافية او رياضية وتشمل جميع الجداول التي تظهر الظاهرة الجغرافية على اساس حالات التكرار وتجميع البيانات فيها على اساس زماني وتسمى بيانات زمنية او على اساس مكاني وتسمى بيانات مكانية.

ب-على اساس النوع :

١-البيانات العددية الحقيقة : وهذه شائعة في الجغرافية ويكون الصفر فيها غير اعتباري بمعنى اننا عندما نقول ان قيمة الظاهرة (س = صفر) فلا يعني هذا عدم وجوده الظاهرة، عندما نقول ان (درجة الحرارة = صفر) هذا لا يعني عدم وجود الظاهرة.

٢-البيانات النسبية : وتعني نسبة توفر الظاهرة الى المجموع الكلي للظواهر الموجودة في هذه الحالة يكون للصفر قيمة فعندما نقول ان نسبة ظاهرة معينة في مكان ما تساوي صفر فمعنى هذا ان الظاهرة غير موجودة اساسا (نسبة الطالبات الغير محجبات في شعبة ج = صفر) . دل هذا على عدم وجود الظاهرة.

٣-البيانات الاسمية : وتستخدم في الجغرافية للإشارة الى النوع والصفة فقال مثلاً ترب حمراء او بنية او يقال اقليم جاف او اقليم شبة جاف.

٤-بيانات رتبيه : وهي بيانات ترتب القيم فيها على اساس تسلسلها اما تصاعدياً او تنازلياً مثل ترتيب المدن وفقاً لأحجامها فيقال المدينة الاولى والمدينة الثانية ...الخ .

مصادر المعلومات الجغرافية

ان تداخل الجغرافية مع العلوم الاخرى جعلها توفر عدد كبير من المعلومات والمصادر يمكن تصنفيها الى ثلاث انواع رئيسية:

مصادر مكتبية	الدراسة الميدانية	نظام المحاكاة

١-مصادر مكتبية :

يقصد بها المعلومات والبيانات المنشورة وغير المنشورة وبأشكال متعددة وتنقسم الى نوعيين:

ا- بيانات ومعلومات مساحية : تظم هذه المجموعة الخرائط والصور الجوية والفضائية والمعلومات المساحية المأخوذة من نظم المعلومات الجغرافية وتتميز هذه المجموعة من البيانات بشموليتها وانفراد علم الجغرافية بها وامكانية تحويلها الى بيانات غير مساحية اذ ممكن ان تفرغ معلومات

الخريطة والصورة في جداول إحصائية تتلاءم مع الكيفية التي يرغب الباحث الجغرافي وضعها فيها .

ب-معلومات غير مساحية : ويقصد بها المعلومات والبيانات غير الممثلة على الخرائط والصور الجوية وهذه تتوفر في النشرات والدوريات التي تصدرها الامم المتحدة والمنظمات التابعة لها وكذلك في التعدادات السكانية والمجاميع الاحصائية التي تنشرها الجهات الحكومية والسجلات والوثائق غير المنشورة التي تحتفظ بها الدوائر وذات الصلة بالظاهرة الجغرافية المدروسة .

٢- **الدراسة الميدانية** : ويقصد بها قيام الباحث بجمع المعلومات عن الظاهرة بشكل مباشر وهي مكلمة للدراسة المكتبية اذ يقوم الباحث خلالها بالبحث والتقصي عن موضوع الدراسة في الحقل بشكل مباشر.

<u>للدراسة الميدانية عدة فوائد هي:</u>

١-تحديث البيانات والمعلومات التي يتم الحصول عليها من الدراسة المكتبية .

٢-استفتاء راي الجمهور عما قد يكون خافياً عن الباحث من اقتراحات وحلول للمشاكل التي تواجه الظاهرة المدروسة.

٣-استكمال النقص الموجود في الدراسات والبيانات السابقة التي تم الحصول عليها من المكتبة . وذلك كون هذه الدراسات غالباً ما تكون قد تناولت جزءاً من موضوعات الدراسة الحالية وتركت جزءاً منها يحتاج الى دراسة.

٤-اختبار الفرضيات التي بنيت عليها الدراسة والتحق ميدانياً من الصعوبات التي تواجهها اقتراحات الباحث وفي خطوة لاحقة تعديل الفرضيات.

اساليب الدراسة الميدانية

تتطلب عملية جمع البيانات تحديد الأسلوب المناسب لجمعها، وتحديد هذا الأسلوب ليس بالأمر السهل وهي مشكلة حقيقية يواجهها الباحث، وعلى العموم **هناك بعض المعايير التي يجب أخذها بالحسبان لاختيار الأسلوب المناسب، وهي:**

١- درجة الدقة المطلوبة: في بعض الأحيان يستعمل أسلوب الحصر الشامل عندما نريد بيانات دقيقة وشاملة، مثل حالة البحوث التي تتعلق بحياة الأفراد (أجهزة الغواصين وسلامتها).

٢- مدى تجانس الوحدات الإحصائية: فكلما كانت درجة التجانس عالية وخاصة في المجتمعات الكبيرة عندئذ يفضل استعمال أسلوب العينة.

٣- المدة الزمنية المخصصة للبحث: فكلما كانت المدة الزمنية طويلة أمكن استخدام أسلوب الحصر الشامل أو الجزئي وبالضد من ذلك يستعمل أسلوب العينة في حالة قصر المدة الزمنية.

٤- مدى توفر الامكانيات المالية والبشرية: فعند توافرها بشكل كافٍ يمكن استعمال أسلوب الحصر وعند قلتها يمكن اعتماد أسلوب العينة. وبشكل عام وفضلاً عن المعايير السابقة يعتمد الأسلوب الذي يستعمل في جمع البيانات الخاصة بالبحث على طبيعة الهدف من البحث من جهة، وعلى حجم المجتمع الإحصائي محل البحث من جهة أخرى

يمكن اجراء الدراسة الميدانية بأحد الاسلوبين .

اسلوب المسح الشامل اسلوب المسح بالعينات

١-اسلوب المسح الشامل : ويتم فيه تناول جميع مفردات الظاهرة بالبحث اذ يقوم الباحث بتحديد جميع مفردات الظاهرة تحديداً دقيقاً وجمع كافة المعلومات والبيانات المطلوبة عنها.

اي هو أسلوب لجمع البيانات من جميع وحدات المجتمع الإحصائي محل البحث والدراسة دون استثناء. ويلجأ الباحث إلى هذا الأسلوب عندما يرغب في الحصول على بيانات تفصيلية عن جميع وحدات المجتمع وعندما يجهل الباحث طبيعة المجتمع لاسيما إذا لم تنفذ عليه دراسات سابقة.

ويمتاز هذا الأسلوب بدقة النتائج والشمولية وعدم التحيز، ولكن الاحتياجات المادية والبشرية والزمنية التي يتطلبها قد تحول دون إمكانية استعماله.

٢-اسلوب المسح بالعينة : يعتبر هذا الاسلوب اقل دقة من الاسلوب السابق اذ يقوم الباحث باختيار جزء من المجتمع بأساليب رياضية محددة يدرس فيها الظاهرة الجغرافية المطلوبة ثم يعمم النتائج التي توصل اليها على انها نتائج شمولية.

و سواء كانت الدراسة باستخدام المسح الشامل او بأسلوب العينة . فان الباحث سيحتاج الى ورقة يضع عليها الاسئلة التي من خلالها يستطيع استكمال المعلومات التي لم يحصل عليها في الدراسة المكتبية وتمسى هذه الورقة استمارة الاستبيان وتصمم وفقا لمجموعة من الشروط والمواصفات التي يجب مراعاتها وهي:

١-يجب ان يكون الباحث على دراية تامة بموضوع البحث والهدف منه وكذلك المعلومات التي ينبغي الحصول عليها اذ ان عدم وضوح الهدف سيؤدي الى ضياع في الوقت والجهد ومن دون نتيجة لذا فان الباحث سيحتاج الى اعداد استبيان اولى لتحقيق ما يأتي:

أ- التعمق في فرضيات الدراسة

ب-الحصول على فكرة اولية عن النتائج المتوقعة .

ت-تحديد المشاكل التي تواجه استمارة الاستبيان ومعالجتها .

٢-يفترض ان يكون الباحث على دراية تامة بالأساليب الاحصائية التي سيلجأ اليها عند إجراء التحليل وان يكون على دراية بالحاسبة الالكترونية والبرنامج الاحصائي الذي سيستخدم بالدراسة .

اما في عملية تنظيم الاستمارة فيجب مراعاة ما يأتي:

أ- تقديم الاستمارة بفقرة يوضح فيها للشخص المبحوث الهدف من الدراسة والتأكيد له بان المعلومات ستستخدم لأغراض البحث العلمي فقط مع تقديم كلمة شكر له .

ب-ان يكون عدد الاسئلة قليل مع تجنب الاسئلة المزدوجة والاسئلة التي يمكن ان يجاب عليها بشكل مقالة . مثل (ما هو رأيك بالخدمة الصحية في مدينتك) .

ج- محاولة دفع الشخص المبحوث الى الاجابة بشكل رقمي عن الاسئلة فمثلاً السؤال السابق عن الخدمة الصحية يمكن ان يصاغ بالشكل الاتي:

ماهي الدرجة التي تمنحها لكفاءة الخدمات الصحية في منطقتك ؟

١	٢	٣	٤	٥	٦	٧	٨	٩	١٠

د- حتى في الاسئلة التي تكون اجابتها كيفية حتما يجب تحديد الشخص المبحوث بالإجابة. فالسؤال عن المستوى التعليمي لا يصاغ بالشكل الاتي (ما هو مستواك التعليمي) . وانما يصاغ بالشكل الاتي (ما هو مستواك التعليمي).

امي	ابتدائي	متوسط	اعدادي	جامعي

هـ - تجنب الاسئلة التي تناول الجوانب الاقتصادية والصحية بشكل مباشر ويمكن الحصول على معلومات عنها بأسئلة غير مباشرة .

و- ان تكون الاستمارة مطبوعة بلغة مفهومة في منطقة الدراسة مع استخدام مصطلحات قريبة من اللهجة المحلية .

ز- في حال استخدام البريد يفضل وضع ضرف فارغ داخل الظرف المرسل يثبت فيه الباحث عنوانه مع طابع بريدي لضمان رجوع الاستمارة .

٣- نظام المحاكاة

يقصد بها تمثيل حالة معينة مختبرياً بتجميع اجزاءها بعد توليدها باستخدام الحاسبة الالكترونية وفقاً لمعطيات معلومة عن هذه الحالة وقد استخدمت نظم المحاكاة في الجغرافية كمصدر للمعلومات وبأكثر من شكل حيث استخدمت للتعرف على ماضي الظاهرة الجغرافية وعن مستقبلها واستخدامت لاختبار دقة الحلول والمقترحات للمشكلات المدروسة .

يمكن استخدام نظم المحاكاة في الجغرافية من خلال ثلاث محاور:

١-التعرف على ماضي الظاهرة الجغرافية وذلك من خلال اعطاء الحاسبة المعلومات المتوفرة عن فترات متعددة لماضي الظاهرة لتستدل الحاسبة من خلالها على الكيفية التي نمت بها الظاهرة ويمكنها بعد ذلك الوصول الى معلومات عن الفترات السابقة التي لا نمتلك عنها معلومات.

٢-التنبؤ بمستقبل الظاهرة اذ يمكن من خلال نظم المحاكاة وبعد تلقين الحاسبة معلومات عن ماضي الظاهرة الجغرافية الحصول على تنبؤات متعددة لمستقبل الظاهرة المدروسة وبمستويات متعددة ممن الثقة .

٣-محاكاة الاقتراحات والتوصيات التي تطرحها الدراسة بهدف التعرف على صلاحيتها والمشكلات التي يمكن ان تعترضها عند التنفيذ ومن ثم اختبار الافضل و الاكثر واقعية .

العينـــــــات

قد يكون من الصعب جداً اتباع أسلوب الحصر الشامل في دراسة العديد من الظواهر؛ لذا يمكن للباحث أن يكتفي باختيار عينة من وحدات المجتمع الإحصائي لتحليلها واستخلاص نتائجها وتعميمها على المجتمع الأصلي.

العينة : هي جزء صغير من المجتمع يتم اختياره وفقاً لضوابط محددة بهدف الكشف عن الصفات الكمية والنوعية لذلك المجتمع الى ذلك فان اختيار العينة يجب ان يكون مراعياً لطبيعة المجتمع وتوزيعه الاحصائي . والحيادية التي تضمن تمثيل العينة افضل تمثيل.

يعتبر اسلوب العينة اقل دقة من اسلوب المسح الشامل في الكشف عن صفات المجتمع إلا إذا أحسن استعماله على أسس علمية سليمة ودقيقة. غير ان هناك ضرورات تدفع الباحث الجغرافي الى اختيار اسلوب المسح بالعينات هي:

١-الكلفة العالية والادارة التي تحتاجها عمليات المسح الشامل عندما يكون المجتمع كبير ، غالباً ما يواجه الجغرافي كبر حجم المجتمع الذي يقوم بدراسته وهذا يعني انه بحاجة الى اموال كبيرة وكادر مدرب لإجراء المسح الشامل.

٢-في معظم الحالات تكون مفردات الظاهرة الجغرافية موزعة على مساحات واسعة واحيانا تكون موزعة في مناطق نائية يصعب الوصول اليها . مثال / لو اردنا اجراء دراسة ميدانية للآبار في العراق فإننا سنواجه صعوبة

بالوصول الى الكثير من الابار في المنطقة الغربية وذلك كونها مناطق نائية يصعب الوصول اليها.

٣-عدم محدودية مجتمع الدراسة وعدم امكانية حصرة ضمن نطاق محدود لأجراء الفحص الشامل عليه. مثال/ المدن الكبيرة ومجتمعات الثروة الحيوانية في اي مكان في العالم.

٤-التغير السريع للظاهرة الجغرافية ففي احيان كثيرة تتغير الظاهرة الجغرافية قبل توفر الفرصة الكافية لأجراء المسح الشامل . مثال / قيام الباحث الجغرافي بدراسة استعمالات الارض الزراعية في على مستوى البلد، ان مثل هذه الدراسة قد لاتكون ممكنة بسبب تغير استعمالات الارض في حال استخدام المسح الشامل.

٥-في حالات محددة و تحديداً عندما يكون حجم المجتمع صغير والدراسة تتطلب تحليل احصائي لمجتمع كبير في هذه الحالة فقط يتم اختيار عينه لتعظيم حجم المجتمع وذلك باستخدام اسلوب السحب بإرجاع / مثال / مرض الايدز ففي هذا النوع من الدراسات قد لا نحصل على عدد كافي من المفردات لأغراض الدراسة فنلجأ الى اسلوب العينات لتعظيم حجم المجتمع.

إن نجاح استخدام هذا الأسلوب يتوقف على عوامل عديدة أهمها:

١- تحديد إطار العينة.

٢- تحديد حجم العينة.

٣- تحديد نوع العينة التي سيتم استعمالها.

٢-٣-٣-١ تحديد إطار العينة ***the Sample Frame Specify:***

من الضروري قبل البدء باختيار وحدات العينة عمل مخطط يحتوي على جميع الوحدات التي ستتم عملية الاختيار من بينها، وتحديد مكان تواجد تلك الوحدات لتعيين السبل الكفيلة للوصول إليها في جمع البيانات والمعلومات المطلوبة عنها.

وللإطار أشكال متعددة، فقد يكون على شكل خريطة تضم جميع المواقع المطلوب دراستها أو على شكل قائمة أو مجموعة قوائم تحتوي على أسماء،

وعناوين، ومواقع وحدات المجتمع المطلوب دراسته، ويجب على الباحث التحقق من دقة إطار العينة وصلاحيته وشموليته قبل اختيارها؛ وللحصول على عينة ممثلة للمجتمع تمثيلاً صادقاً من جهة ومطابقتها لهدف البحث ومنهجيته من جهة أخرى.

٢-٣-٣-٢ تحديد حجم العينة ***Determine of sample size:***

بعد تحديد الإطار العام للمجتمع الذي منه سيتم اختيار العينة يجب أن يكون الباحث جاهزاً لاختيار وحداتها. ولكن قبل عملية الشروع في الاختيار يحتاج الباحث إلى تحديد الحجم الأمثل للعينة المطلوبة؛ لأن صغرها قد يجعلها غير ممثلة لمجتمع الدراسة وبالمقابل فإن كبرها قد يكلف الباحث كثيراً من المال والوقت والجهد، وبشكل عام لا يوجد حجم محدد أو نسبة مئوية محددة من حجم المجتمع الإحصائي يمكن اعتمادها في جميع الدراسات والبحوث. إلا أن هناك مؤشرات يمكن الاهتداء بها لتحديد حجم العينة منها:

١- إذا كان مجتمع الدراسة متجانساً فإن حجم العينة يمكن أن يكون صغيراً، أما إذا كان غير متجانس فلا بد أن يكون حجم العينة كبيراً.

٢- درجة الدقة المطلوبة من البحث، إذ يكبر حجم العينة كلما كانت درجة الدقة المطلوبة عالية وبالعكس.

٣- توافر الإمكانيات المادية والبشرية، فكلما كان كلاهما متوفرين سيتمكن الباحث من اختيار عينة كبيرة وبالعكس.

٤- طبيعة المنهج البحثي الذي سيتبع يؤثر بشكل مباشر على حجم العينة، فإن المنهج التجريبي مثلا يتطلب حجماً كبيراً للعينة.

وبشكل عام في حالة العينات العشوائية يمكن اعتماد الصيغة الآتية لتحديد حجم العينة

$$n = \frac{t^2}{r^2 + \frac{1}{N} t^2}$$

إذ أن:

n = حجم العينة المطلوبة.

t = قيمة t المجدولة التي تقابل الخطأ المسموح به.

r = احتمال الخطأ.

N= عدد وحدات المجتمع الإحصائي.

مثال:

جد حجم العينة المطلوب استخدامها في دراسة إحصائية للعوامل المؤثرة على الخصوبة السكانية لمحافظة ما، يبلغ عدد أسرها ٢٤٩٣٦١ أسرة، إذا أردنا أن يكون مقدار الخطأ بالتقدير ٠.٠٤ من الانحراف المعياري واحتمال هذا الخطأ ٠.٠٠٥.

الحل:

$$n = \frac{t^2}{r^2 + \frac{1}{N}t^2}$$

$$n = \frac{(1.96)^2}{(0.05)^2 + \frac{1}{249361}(1.96)^2}$$

$$n = \frac{3.841}{0.0025154} = 1526$$ عدد وحدات العينة التي سيتم سحبها من المجتمع

انواع العينات

هناك انواع عديدة من العينات يلائم كل منها نوع محدد من المجتمع

العينة العشوائية البسيطة	العينة العشوائية الطبقية	العينة العشوائية الاسلوبية

١-العينة العشوائية البسيطة:

هي اكثر العينات شيوعا بسبب بساطتها و سهولة استخدامها. وتقوم فلسفتها على اعطاء فرصة متساوي لكل مفردة من مفردات المجتمع للظهور في العينة ، ولهذا فان هذا النوع يكون مناسبا في الحالات الاتية:

أ- عندما يكون المجتمع متجانس اي ان تكون مفردات المجتمع متقاربة القيم في الظاهرة المطلوب دراستها.

ب-في المجتمع غير المحدود كما هو الحال في دراسات الموارد الطبيعية.

ج- في المجتمعات الصغيرة التي لا يتجاوز عدد مفرداتها ٢٥ مفردة.

مثال / اذا اردنا اظهار نجاح العينة العشوائية البسيطة مع المجتمعات المتجانسة وعدم نجاحها مع المجتمعات غير المتجانسة نقوم بسحب ثلاث عينات من مجتمع متجانس والذي هو اعمار الطلبة في شعبة ج المرحلة الثالثة ونقوم بسحب ثلاث عينات من مجتمع غير متجانس والذي هو المصروف اليومي لطلبة تلك الشعبة و كالاتي:

مجتمع متجانس			مجتمع غير متجانس		
عينة اولى	عينة ثانية	عينة ثالثة	عينة اولى	عينة ثانية	عينة ثالثة
٢٢	٢١	٢١	١.٠٠٠	١٥.٠٠٠	٧.٠٠٠
٢٢	١٩	٢٢	٢.٠٠٠	٧.٠٠٠	٦.٠٠٠
١٩	٢٢	٢٣	٢.٠٠٠	٢٠.٠٠٠	٤.٠٠٠
٢٠	٢٣	٢١	٤.٠٠٠	١٨.٠٠٠	٨.٠٠٠
٢٤	٢١	٢٢	١.٠٠٠	٥.٠٠٠	٣.٠٠٠
$X1=\frac{107}{5}=21.5$	$X2=\frac{106}{5}=21.2$	$X3=\frac{109}{5}=21.8$	$X1=\frac{10.000}{5}=2000$	$X2=\frac{65.000}{5}$	$X3=\frac{27000}{5}=5.250$

٢-العينة العشوائية الطبقية:

تقوم فلسفة هذا الاسلوب على اعطاء فرصة متباينة لكل مفردة من مفردات المجتمع وزن هذه الفرصة مساوي لوزن الطبقة التي تنتمي اليها المفردة. وقد جعل ذلك هذا النوع من العينات اكثر دقة كونها تشمل جميع مفردات المجتمع على اساس حجم تمثيلها في المجتمع وتنحصر معوقات استخدامها في وجوب معرفة حجم المجتمع وتوزيع مفرداته.

خطوات اختيار العينة لاستخدام هذا الاسلوب (العينة العشوائية الطبقية)

١-نقوم بتصنيف المجتمع الى طبقات ومن ثم احتساب تكرار كل طبقة.

٢-نقوم باحتساب وزن الطبقة في المجتمع وذلك بقسمة عدد المفردات في كل طبقة على عدد مفردات المجتمع ومن خلال المعادلة الاتية $Wi=\frac{Ni}{n}$

Wi	وزن الطبقة
Ni	عدد المفردات
N	عدد مفردات المجتمع

٣-نقوم باحتساب عدد المفردات المطلوب اختيارها من كل طبقة في حجم العينة وذلك بضرب وزن كل طبقة x حجم العينة ومن ثم تعديل النتائج الى اعداد صحيحة .

٤-نقوم بعد ذلك باختيار المفردات من المجتمع بعدد مساوي لعدد المفردات التي حصلنا عليها في الخطوة السابقة وبطريقة عشوائية .

مثال: البيانات التالية تمثل ظاهرة معينة المطلوب سحب (٣) عينات بحجم (٥) مفردات للعينة

١	٢٥	٢٩	٨٠	٧٩
١٢	٣٦	٤٢	٧٤	٦٣
١٨	٣٤	٥٦	٦١	٧٦
٢٠	٢٣	٤٣	٥٧	٦٩

الحل

الطبقات	التكرار	وزن الطبقة	عدد المفردات وزن الطبقة*حجم الطبقة	≅
١-٢٠	٤	$0.2=\frac{4}{20}$	١=٥*٢.٠	١
٢١-٤٠	٥	$0.25=\frac{5}{20}$	١.٢٥ = ٥*٠.٢٥	١
٤١-٦٠	٤	$0.2=\frac{4}{20}$	١= ٥ *٠.٢	١
٦١ – ٨٠	٧	$0.35=\frac{7}{20}$	١.٧٥ = ٥*٠.٣٥	٢
	٢٠			٥

ع ١ = (٨٠ . ٦٩ . ٥٧ . ٣٤ . ١٨) ٥١.٦ = $x1 = \frac{258}{5}$

ع ٢ = (٧٩ . ٧٤ . ٥٦ . ٢٥ . ٢٠) ٤٨.٢ = $x2 = \frac{241}{5}$

ع٣= (٧٩ . ٧٤ . ٥٦ . ٣٤ . ٢٠) ٥٢.٦ = $3x = \frac{263}{5}$

٣-**العينة العشوائية الاسلوبية** .

تشابه فلسفة هذا النوع من المعاينة فلسفة العينة العشوائية البسيطة بانها تعطي فرصة متساوية لكل مفردة من مفردات المجتمع بالظهور في العينة غير انها تمتلك اليه محددة لاختيار المفردات في العينة فضلا على ان هذا

النوع من العينات لا يمكن استخدامه مع المجتمعات التي تتعرض للتأثيرات الدورية والموسمية مثل البيانات المناخية والاقتصادية التي تتأثر بالدورات الاقتصادية والسبب في ذلك هو الخوف او الحذر من ان تتطابق الدورة التي تؤثر على الظاهرة مع قيمة (K) اما كيفية اختيار المفردات في العينة فيمكن ان يتم بالخطوات الاتية:

أ- وضع تسلسل على مفردات المجتمع لمعرفة المفردات الكلية للمجتمع.

ب-نقوم باحتساب قيمة K وهي المسافة بين اي مفردتين في المجتمع يتم اختيارهم في العينة وذلك من خلال المعادلة التالية $K=\frac{N}{n}$

المسافة بين مفردتين	K
حجم المجتمع	N
حجم العينة	n

ج- من اول (K) من المفردات يتم اختيار مفردة عشوائية ولتكم المفردة التي تسلسلها (i) حيث ان (i اصغر من او تساوي قيمة K)

د- المفردات التالية التي يتم اختيارها بالعينة هي المفردات التي تسلسلها

K + i

K2 + i

K3 + i

مثال / اسحب عينتين اسلوبيتين بحجم (٥) مفردات للعينة الواحدة

٢	١٢	٢٦	٤٣	٧٦
٦٨	٥٩	٦٤	١٠٠	٨٤
١٨	٢٣	٣٩	٧	٤٤
٥١	٦٦	٣٤	٩٧	٨٦

١-وضع تسلسل للمفردات.

٢-ايجاد قيمة (K) $K=\frac{20}{5}=4$

٣-نأخذ تسلسل (i) ويكون اقل من او يساوي قيمة (K) المستخرجة وتكون افتراضية على ان لا تتشابه بالعينة الاولى والثانية .

ع١	
تسلسل المفردة	قيمة المفردة
I=2	١٢
I+k=2+4=6	٦٨
I+k2=2+8=10	٨٤
I+k3= 2+12=14	٧
I+k4=2+16=18	٣٤
	X1= $\frac{205}{5}=41$

ع٢	
تسلسل المفردة	قيمة المفردة
I=4	٤٣
I+k = 4+4 = 8	٦٤
I+k2 = 4+8 =12	٢٣
I+k3 = 4+12 =16	٣١

$l + k4 = 4+16 = 20$	٨٦
	X2 = $\frac{247}{5} = 49.4$

تصنيف وجدولة البيانات

وهي عملية جمع المتشابهات في مجموعة تشترك عناصرها في صفة واحدة او اكثر.

مثال علي ذلك التصنيف الجغرافي للظواهر حيث تصنف على اساس خواص مشتركة مثل تصنيف التربة الى تربة بنية وترب حمراء.... الخ ... ويصنف المناخ الى مناخ البحر المتوسط ومناخ التندرا.

كذلك قد يكون التصنيف على اساس زمني فتسمى النتيجة في هذه الحالة بالسلسلة الزمنية.

مثال : على ذلك تصنف كميات الامطار النازلة حسب السنوات والاشهر اما اذا كان التصنيف على اساس اماكن سقوط الامطار فتسمى بيانات مكانية. اما اذا كان التصنيف على اساس عددي فان التصنيف في مثل هذه الحالة يتحول الى عملية جدولة ويسمى بالناتج في هذه الحالة بجدولة التوزيع التكراري. وقد لاحظ علماء الرياضيات ان الجداول تختلف فيما بينها في عدد الفئات وطولها ولهذا فقد تباينت وجهات النظر للعلماء في العدد المناسب من الفئات و الاطوال الملائمة بغية تحقيق اعلى دقة وذلك كون عملية الجدولة للبيانات تتأرجح بين متغيرين هما (الدقة ، الاختصار) فكلما ازداد عدد الفئات ازدادت الدقة وقلت درجة الاختصارات وكلما قل عدد الفئات قلت الدقة

وازدياد الاختصار . وقد نالت عملية جدولة البيانات اهتماماً واسعاً من قبل الجغرافيين وقد ظهر في هذا الجانب اتجاهين:

١-الاتجاه الاول : يعتقد اصحاب هذا الاتجاه بان افضل طريقة للحصول على اعلى دقة واعلى درجة من الاختصار يكون بوضع قوانين تقوم باحتساب اعداد الفئات واطوالها من اجل تحقيق الهدف المطلوب و من اهم اصحاب ذا الاتجاه (يول ، سكربتر) Yule scripter.

٢-الاتجاه الثاني : يعتقد اصحاب هذا الاتجاه بان القوانين الرياضية التي وضعها اصحاب الاتجاه الاول تتجاوز خبرة الجغرافي وطبيعة البيانات وقدرة العين على التمييز بين الالوان في حالة استخدام هذا التصنيف في خرائط التضليل .اذ يقوم هذا الاتجاه على مبدأ ترك حرية الاختيار في عدد الفئات واطولها للباحث ...

الاتجاه الاول والقوانين التي سنستخدمها .

اولا: احتساب اعداد الفئات:

أ- طريقة Yule : بموجب هذا القانون فان عدد الفئات يساوي ضعفين ونصف الجذر الرابع لعدد المفردات ويحتسب باستخدام القانون الاتي:

$$F=2.5 * \sqrt[4]{N}$$

F	عدد الفئات
N	حجم المجتمع

استخدام الحاسبة للتخلص من الجذر الرابع ((العدد المتمثل ب N ثم $\sqrt{\quad}$ ثم نكرر الضغط على $\sqrt{\quad}$ ثم * ٢.٥ = ثم نقرب النتيجة))

مثال : استخرج عدد الفئات اذا علمت بان حجم المجتمع = ٢٥

$$F = 2.5 * \sqrt[4]{N}$$

$$F= 2.5 * \sqrt[4]{25}$$

$F= 5.5 \simeq 6$

ب: طريقة الدليل العام: حسب هذه الطريقة فأن عدد الفئات يساوي خمسة اضعاف لوغارتم عدد مفردات المجتمع و تحسب من القانون الاتي:

$F =5* Log (N)$

مثال:حجم المجتمع (٢٤) مفردة عدد الفئات يكون.

$F =5* Log (24)$

$F =6.9 \simeq 7$

ثانيا: احتساب اطوال الفئات.

أ: الفئات متساوية الطول: تلائم هذه الطريقة البيانات المتجانسة . و تكون البشكل الاتي:

ايجاد المدى	**R= Max (xi) – Min (xi) =** **المدى= اكبر قيمة بالمجتمع – اصغر قيمة في المجتمع**
ايجاد عدد الفئات	**اما باستخدام طريقة يول او طريقة الدليل العام**
ايجاد طول الفئة	$L = \frac{R}{F}$

مثال / كون جدول توزيع تكراري بفئات متساوية الطول اذا علمت بان اصغر قيمة في المجتمع (٢) واكبر قيمة (١٠٠) وحجم المجتمع (٢٠) استخدم طريقة يول Yule لاحتساب عدد الفئات .

ايجاد المدى	**R= Max (xi) – Min (xi) =**
ايجاد عدد الفئات	$F = 2.5 * \sqrt[4]{N}$
ايجاد طول الفئة	$L = \frac{R}{F}$

R= Max (xi) – Min (xi) =

$R = 100-2$

$R= 98$

$$F=2.5 * \sqrt[4]{N}$$

$$F = 2.5 * \sqrt[4]{20}$$

$$F = 5.2 \simeq 5$$

$$L= \frac{R}{F}$$

$$L = \frac{98}{5} = 19.6$$

نبدأ من اصغر فئة بالمجتمع ثم نقوم بالزيادة على الفئة الاصغر منها بالرقم المستخرج في الاطول

فئات	تكرارات
٢ ـــ ٢١.٥	-
٢١.٦-٤١.١	-
٤١.٢- ٦٠.٧	-
٦٠.٨ - ٨٠.٣	-
٨٠.٤ ـــ ١٠٠	-

ب- فئات غير متساوية الطول

يستخدم هذا النوع من التصنيف عندما يكون المجتمع غير متجانس و هي على انواع:

- الفئات التربيعية :
- الفئات اللوغاريتمية الهندسية:

١- الفئات التربيعية :

تستخدم هذه الفئات عندما يكون المجتمع غير متجانس بسبب شذوذ عدد كبير من المفردات ويحسب اطوال الفئات التربيعية وبالشكل التالي .

١-نقوم باحتساب المدى التربيعي

QR = $\sqrt{Max(xi)} - \sqrt{(Min(xi)}$

٢-نقوم بأحتساب عدد الفئات باستخدام القوانين التي اشرنا اليها سابقاً وهي طريقة (Yule) وطريقة (الدليل العام) .

٣-نقوم بأحتساب طول الفئة التربيعية من القانون التالي . QL=$\frac{QR}{F}$

٤-نقوم باحتساب حدود الفئات التربيعية وذلك بإضافة طول الفئة التربيعية الى الجذر التربيعي لأصغر قيمة في التوزيع لنحصل في كل مرة على الحد الادنى الفئة التي تليها وهكذا وصولاً الى الفئة الاخيرة التي تغلق بالجذر التربيعي لأكبر قيمة في التوزيع وكما يأتي:

حدود الفئات التربيعية

$\sqrt{\min(xi)}$

$\sqrt{\min(xi)} + QL$

$\sqrt{\min(xi)} + 2QL$

...الخ $\sqrt{Max(xi)}$

٥-نقوم بإيجاد حدود الفئات الحقيقة وذلك بتربيع حدود الفئات التي حصلنا عليها في الخطوة السابقة واحتساب التكرارات في كل منها.

مثال / ضع البيانات التالية في جدول توزيع تكراري بفئات تربيعية وبأستخدام طريقة الدليل العام لاحتساب عدد الفئات .

٩	١٢	١٦	٢٥	٤٩
٣٧	١٨	١٤	٣١	٥٨
٥٤	٦٦	٨٧	٩٤	٢٧
١٠٥	٧١	٩٣	١٢٠	٤٤
٢٤	١٣	١٧	٣٥	

أولا: نقوم بإيجاد المدى التربيعي :

$$QR = \sqrt{Max(xi)} - \sqrt{Min(xi)}$$

$$=\sqrt{120} - \sqrt{9}$$

$$= 10.95 – 3$$

$$= 7.95$$

ثانياً : نقوم بإيجاد عدد الفئات باستخدام طريقة الدليل العام

$$F= 5 * Log(N)$$

$$= 5 * Log (24)$$

$$= 6.9 \simeq 7$$

ثالثاً نقوم بإيجاد طول الفئة التربيعية

$QL= \frac{QR}{F} = \frac{2.95}{7} = 1.135$ طول الفئة التربيعية

رابعاً : نجد الجذر التربيعي لأصغر فئة بالمجتمع المتمثلة ب (9). ثم نقوم بالزيادة في كل مرة على الرقم المستخرج بطول الفئة .

حدود الفئات التربيعية
٣
٤.١٣٥
٥.٢٧
٦.٤
٧.٥٣
٨.٦
٩.٧٣ – ١٠.٩٥

خامساً : نقوم بإيجاد حدود الفئات الحقيقة وذلك بتربيع حدود الفئات التي حصلنا عليها . حيث نقوم بضرب الرقم في نفسة

فئات	تكرارات
٩ – ١٧.٠٨	٦

١٧.٠٩ – ٢٧.٧٦	٤
٢٧.٧٧ – ٤٠.٩٥	٣
٤٠.٩٦ – ٥٦.٦	٣
٥٦.٧ – ٧٣.٩٥	٣
٧٣.٩٦ – ٩٤.٦٦	٣
٩٤.٦٧ – ١٠٠	٢
	٢٤

٢- الفئات اللوغاريتمية (الهندسية) :

يستخدم هذا النوع من الفئات عندما يكون المجتمع غير متجانس بسبب شذوذ عدد قليل من المفردات

اما طريقة احتساب فتكون كما يأتي :

١-نقوم باحتساب المدى اللوغاريتمي من المعادلة الاتية:

LR= Lin max (xi) – Lin min (xi)

حيث ان:

المدى اللوغاريتمي	**LR**
اللوغاريتم الطبيعي لاكبر قيمة	**Lin Max**
اللوغاريتم الطبيعي لاصغر قيمة	**Lin Min**

٢-نقوم باحتساب عدد الفئات المناسبة باستخدام احدى الطرق السابقة (Yule) او (الدليل العام).

٣-نقوم باحتساب طول الفئة اللوغاريتمية على عدد الفئات: $LL=\frac{LR}{F}$

٤-نقوم بإيجاد حدود الفئات اللوغاريتمية وذلك بوضع لوغاريتم اصغر قيمة في التوزيع كحد ادنى للفئة اللوغاريتمية الاولى ثم نبدأ بإضافة طول الفئة اللوغاريتمية الى هذا الحد في كل مرة لنحصل على الحد الادنى للفئة

اللوغاريتمية التي تليها وهكذا وصولاً الى الفئة الأخيرة وتغلق بـ لوغارتم اكبر قيمة في المجتمع وتكون حدود الفئات اللوغاريتمية كما يأتي:

Lin min (xi)

Lin min (xi) + LL

Lin min (xi) + 2 LL

Lin max (xi).............

٥-نستخرج حدود فئات الحقيقية وذلك بإيجاد معكوس (القيمة الاسية) للحدود التي حصلنا عليها من الخطوة السابقة فنحصل على الحدود للفئات ثم نقوم بأحتساب عدد التكرارات المقابلة لكل فئة

مثال / كون جدول توزيع تكراري بفئات لوغارتمية اذا علمت ان اصغر قيمة بالمجتمع (١) واكبر قيمة (١٠٠) وعدد مفردات المجتمع (٢٠)

اولاً : نقوم احتساب المدى اللوغاريتمي

LR= Lin max (xi) – Lin min (xi)

= Lin (100) – Lin (1)

=4.6 – 0 = 4.6

ثانياً : احتساب عدد الفئات باي من الطرق سواء اكانت يول او طريقة الدليل العام:

F= 2.5 * $\sqrt[4]{N}$

F= 2.5 * $\sqrt[4]{20}$

F= 5.2 ≃ 5

ثالثاً : نقوم باحتساب اطوال الفئات.

LL= $\frac{LR}{F} = \frac{4.6}{5} = 0.92$

رابعاً : نقوم بإيجاد حدود الفئات اللوغاريتمية وذلك بوضع لوغاريتم اصغر قيمة في التوزيع كحد ادنى للفئة اللوغاريتمية الاولى ثم نبدأ بإضافة طول الفئة اللوغاريتمية الى هذا الحد في كل مرة لنحصل على الحد الادنى للفئة اللوغاريتمية التي تليها وهكذا وصولاً الى الفئة الأخيرة وتغلق بـ لوغارتم اكبر قيمة في المجتمع وتكون حدود الفئات اللوغاريتمية كما يأتي:

حدود الفئات اللوغاريتمية
٠ - ٠.٩١
٠.٩٢ - ١.٨٣
١.٨٤ – ٢.٧٥
٢.٧٦ – ٣.٦٨
٣.٦٨-٤.٦

خامساً : ايجاد اعداد الفئات الحقيقة والتكرارات من خلال الضغط على الرقم في الفئات اللوغارتيمية ومن ثم 2ndf ومن ثم in

فئات	تكرارات
١ - ٢.٤	
٢.٥ - ٦.٢٨	
٦.٢٩ – ١٥.٧٨	
١٥.٧٩ – ٣٩.٦٥	
٣٩.٦٤ - ١٠٠	

طريقة سكربتر (Scripter) :

وضعت هذه الطريقة عام (١٩٦٩) من قبل هذا العالم (سكربتر) ولهذا سميت باسمه وتقوم فكرته على اعتماد الوسط الحسابي (المعدل) كنقطة تقسم الى فئات على اساس ان الوسط الحسابي غير متحيز واكثر المقاييس كفاءة . يعاب على هذه الطريقة ان عدد الفئات فيها محدده ب (٤ ، ٨ ، ١٦ ، ٣٢) وهكذا اي ان هذه الطريقة لا تحتاج الى احتساب اعداد الفئات التي تعودنا عليها في الطرق السابقة ..

اما خطوات احتساب عدد الفئات في هذه الطريقة فهي:

١-نقوم اولا باحتساب الوسط الحسابي للمجتمع المطلوب ووضعه في جدول توزيع تكراري نسمية الوسط العام ونرمز له بالرمز (M)

٢-تقسم مفردات المجتمع الى مجموعتين تضم المجموعة الاولى جميع المفردات التي تقل قيمتها عن الوسط الحسابي العام فيما تظم المجموعة الثانية جميع مفردات التي تساوي قيمتها او تزيد على الوسط الحسابي العام .

٣-نقوم باحتساب الوسط الحسابي للمجموعة الاولى ويرمز له بالرمز (x1) والوسط الحسابي للمجموعة الثانية يرمز له بالرمز (x2)

٤-نكون جدول التوزيع التكراري بأربع فئات وذلك بوضع اصغر قيمة في التوزيع كحد ادني للفئة الاولى والوسط الحسابي للمجموعة الاولى (x1)كحد ادنى للفئة الثانية و الوسط الحسابي العام (M) كحد ادنى للفئة الثالثة و الوسط الحسابي للمجموعة الثانية (x2) كحد ادنى للفئة الرابعة والتي تغلق بأكبر قيمة في التوزيع ثم نجد بعد ذلك التكرارات . ويكون جدول التوزيع التكراري كما يأتي :

الفئة	التكرارات
- Min (xi)	**X**
- X1	**X**

- M	X
- X2	X

مثال : ضع البيانات التالية في جدول توزيع تكراري باستخدام طريقة سكربتر .

٥	١٢	١٧	٤٣
٩٢	٨٤	١٠٠	٦٥
٣٦	٢٧	٤٩	٦١
١٥	٢٤	٥٥	٧٣

الحل :

أ- ايجاد الوسط الحسابي العام من خلال $M = \frac{\text{مجموع المفردات}}{\text{عدد المفردات}}$

$$M = \frac{758}{16} = 47.4$$

ب- تقسم المجتمع الى قسمين المجموعة الاولى والمجموعة الثانية

- المجموعة الاولى : تأخذ الارقام التي تقل عن الوسط الحسابي
- المجموعة الثانية : تأخذ الارقام التي تزيد او تساوي الوسط الحسابي

المجموعة الاولى			المجموعة الثانية		
٥	١٢	١٧	٩٢	٨٤	١٠٠
٤٣	٣٦	٢٧	٦٣	٤٩	٦١
١٥	٢٤		٥٥	٧٣	
$X1 = \frac{179}{8} = 22.4$			$X2 = \frac{579}{8} = 72.4$		

تكوين جدول تكراري

	الفئات	التكرارات
Min	**٥ - ٢٢.٣**	**٤**
Xi	**٢٢.٤ - ٤٧.٣**	**٤**
M	**٤٧.٤ — ٧٢.٣**	**٥**
X2	**٧٢.٤ - ١٠٠**	**٣**

مقايس النزعة المركزية

إن التمثيل الجدولي والرسوم البيانية للبيانات تعد من المؤشرات الإحصائية التي يمكن الاعتماد عليها في وصف الظاهرة، وتقديمها بشكل مختصر وبسيط، إلا أننا نفضل دائماً استعمال طرائق القياس الكمي "العددي" لقياس تجمع البيانات حول قيمة معينة بحيث تمثلها أفضل تمثيل ويحدث هذا عندما تكون تلك القيمة مركز ثقل حقيقي تجذب إليها أكبر عدد من قيم بيانات الظاهرة وبالعكس تفقد تلك القيمة أهميتها إذا ما ابتعد كثير من البيانات عنها.

وتعد مقاييس النزعة المركزية "المتوسطات" من أهم تلك المقاييس العددية استعمالاً لهذا الغرض، ومن أهمها وأكثرها شيوعاً الوسط الحسابي، والوسيط، والمنوال.

الوسط الحسابي :

يعد الوسط الحسابي أبرز مقاييس النزعة المركزية شهرة وأكثرها استعمالاً، بل لعله من أول المقاييس الإحصائية على الاطلاق وأهمها؛ لما يتمتع به من مزايا وخواص، ولدخوله في حساب كثير من المقاييس الإحصائية الأخرى.

ويعرف الوسط الحسابي بشكل عام بأنه "القيمة التي تساوي مجموع قيم المشاهدات مقسوماً على عددها" ويتم حسابه من البيانات المبوبة وغير المبوبة.

أولاً– الوسط الحسابي للبيانات غير المبوبة ويحسب باستعمال الصيغة الآتية:

$$\overline{X} =$$

ثانياً– الوسط الحسابي للبيانات المبوبة ويحسب باستعمال الصيغة الآتية:

$$\frac{\sum mi * fi}{\sum fi}$$

اذ ان:

تكرارات : fi	**مركز الفئة : Mi**

مثال : احسب الوسط الحسابي من خلال جدول التكرارات الاتي:

الفئات	تكرارات / fi	مركز الفئة / Mi	Mi * Fi
١ – ١٠	٦	$\frac{1+11}{2}$=٦	٣٦
١١ – ٢٠	٥	$\frac{11+21}{2}$=١٦	٨٠
٢١ – ٣٠	٢	$\frac{21+31}{2}$=٢٦	٥٢
٣١ – Max	٢	$\frac{31+51}{2}$=٤١	٨٢
	١٥		$\sum = 250$

١-ايجاد مركز الفئة (Mi) من خلال جمعالحد الادنى للفئة مع الحد الادنى للفئة التي تليها ثم تقسم على (٢).

٢-ايجاد (mi*fi) من خلال ضرب مركز الفئة مع تكرارها.

٣-تقسم مجموع (mi*fi) على (fi): $\frac{\sum mi*fi}{\sum fi}$

$$\frac{250}{15} = 16.6$$

الوسيط الحسابي:

ويرمز له بالرمز (me) وهي احد مقاييس النزعة المركزية ويمثل قيمة المفردة التي تقع في منتصف التكرار المتجمع الصاعد او التكرار المتجمع النازل للمفردات

ويمتاز هذا المقياس بعدم تأثره بالقيم الشاذة اما طريقة احتسابه فتكون كما ياتي:

أ- اذا كانت البيانات غير موضوعة في جدول توزيع تكراري فان قيمة المفردة او الوسيط يتم احتسابها بعد ترتيب البيانات ترتيباً تصاعدياً او تنازلياً كما يأتي:

١-اذا كان عدد البيانات المفردات فردياً فان نسبة الوسيط هي المفردة التي تسلسلها $\frac{N+1}{2}$

٢-اذا كان عدد البيانات (المفردات) زوجياً فان قيمة الوسيط تساوي قيمة الوسط الحسابي للمفردتين التي تسلسلها ($\frac{N}{2}+1$ و $\frac{N}{2}$)

مثال / البيانات التالية تمثل توزيع جغرافي لظاهرة معينة اوجد قيمة الوسط الحسابي

(٦ ، ٢ ، ٨ ، ٧ ، ٤ ، ١٠ ، ١١ ، ١٥ ، ١٨)

١٨	١٥	١١	١٠	٨	٧	٦	٤	٢

Me = $\frac{9+1}{2} = 5$

الوسيط Me = 8

مثال ٢ : جد الوسيط الحسابي للمفردات الاتية:

(١٢ ، ١٠ ، ٩ ، ٧ ، ٧ ، ٥ ، ٤ ، ٣ ، ٣ ، ١)

الحل:

١٢	١٠	٩	٧	٧	٥	٤	٣	٣	١

me= $\frac{N}{2}$ و $\frac{N}{2}+1$

=5 , 6

me= $\frac{5+7}{2} = 6$ الوسيط

٣-اذا كانت البيانات موضوعة في جدول توزيع تكراري فان قيمة الوسيط يتم احتسابها من الخطوات الاتية:

أ- نقوم اولاً بإيجاد الفئة الوسيطة وذلك بقسمة عدد المفردات على (٢)

ب- نقوم بعد ذلك باحتساب قيمة الوسيط من المعادلة التالية

$$me = A + \left(\frac{\frac{N}{2} - F}{f}\right) * L$$

حيث ان

قيمة الوسيط	Me
الحد الادنى للفئة الوسيطة	A
عدد المفردات	N
تكرار المجتمع في بداية الفئة الوسيطة	F
تكرار الفئة الوسيطة	f
طول الفئة الوسيطة	L

مثال : احسب الوسيط الحسابي من جدول التوزيع التكراري الاتي:

الفئات	التكرارات	
١ ــ ٩	٦	F
١٠ ــ ١٩	٤	
٢٠ ــ ٢٩	٨	الفئة الوسطية f
٣٠- ٤٠	٦	
	٢٤	

الحل : نقوم اولاً بايجاد بتحديد الفئة الوسيطة وذلك بقسمة عدد المفردات على (٢)

ترتيب الوسيط $\frac{24}{2} = 12$

$$me = A + \left(\frac{\frac{N}{2} - F}{f}\right) * L$$

$$me = 20 + \left(\frac{\frac{24}{2} - 10}{8}\right) * (30 - 20)$$

$$me = 20 + \frac{2}{8} * 10$$

me = 20+0.25 *10

me = 20+2.5 = 22.5

المنوال :

وهو القيمة الاكثر انتشاراً في التوزيع ويتم احتسابها وفقاً لما يأتي:

١- اذا كانت البيانات غير موضوعة في جدول توزيع تكراري فان قيمة المنوال تساوي قيمة المفردة الاكثر تكراراً

مثال ١ : (٨ . ٥ . ١٠ . ٤ . ٤ . ١ . ١ . ١ . ٢ . ٦) mo = 1

مثال ٢ : (٣ . ٧ . ١ . ١ . ٣ . ٢ . ٣ . ٤ . ٦ . ٦ . ٦) mo= 6 , 3

مثال ٣ : (٩ . ١٧ . ٢٣ . ٨ . ٣ . ١ . ٥ . ٢) لا يوجد = mo

٢- اذا كانت البيانات موضوعة في جدول توزيع تكراري فان قيمة المنوال يتم احتسابها من القانون التالي

$$mo = A + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) * L$$

المنوال	Mo
الحد الادنى للفئة المنوالية	A
هو الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها	Δ1
هو الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها	Δ2
هو طول الفئة المنوالية	L

مثال : جد المنوال من جدول التوزيع التكراري

	الفئات	التكرارات
	١٠ ــ ١٩	٢
A	٢٠ ــ ٢٩	٨
	٣٠ ــ ٣٩	٤
	٤٠ ــ ٥٠	١
		١٥

الحل : أولاً نقوم بتحديد الفئة المنوالية وهي الفئة التي لها اكبر تكرار كما وضح في الجدول السابق

$$mo = 20 + \left(\frac{8-2}{(8-2)+(8-4)}\right) * (30 - 20)$$

$$= 20 + \left(\frac{6}{10}\right) * 10$$

$$=20+(0.6*10)$$

$$=20+6= 26$$

مقاييس التشتت

كما هو واضح في المقاييس السابقة (مقاييس النزعة المركزية) عندما تستخدم في تمثيل مجتمع معين فانها تعاني من نقص في تمثيلة وكما هو واضح في المثال التالي

المجموعة الاولى	المجموعة الثانية	المجموعة الثالثة
٢	٥	٤
٥	٥	٦
١	٥	٥
٥	٥	٥
١٢	٥	٥
X1=5	X2=5	X3=5
Me1=5	Me2=5	Me3=5
Mo1=5	Mo2=5	Mo3=5
R1=11	R2 = 0	R3=3

نلاحظ ان المجتمعات الثلاثة تتباين في قيم مفرداتها ولكن رغم ذلك تشترك في قسم الوسط الحسابي والوسيط والمنوال وهذا يعني اننا بحاجة الى مقاييس جديد للكشف عن هذا التباين وهذه المقاييس تسمى (مقاييس التشتت)

المدى :

هو احد هذه المقاييس وهو ابسط مقاييس التشتت ويعرف بانه الفرق بين اعلى قيمة وبين اصغر قيمة في التوزيع ويرمز له (R) و يحسب :

$$R= \max(x_i) - \min(x_i)$$

وفي حالة البيانات المبوبة فان المدى يساوي الفرق بين الحد الاعلى لاكبر للفئات والحد الادنى لأصغر الفئات او الفرق بين مركز اكبر الفئات ومركز اصغر الفئات ويمتاز بالخصائص الاتية:

أ- انه مقياس سهل وسريع الحساب .

ب- لا يمكن احتسابة من جدول توزيع تكراري مفتوح

ت- اعتماده على اعلى قيمة واقل قيمه فهو يتأثر بالقيم الشاذة

ثانياً : **الانحراف المتوسط** ***The Mean Deviation***:

يعرف الانحراف المتوسط بانه الوسط الحسابي للقيم المطلقة للانحرافات عن المتوسط. علماً ان القيمة المطلقة للانحراف هي انحراف القيمة عن متوسط المجموعة بإهمال الاشارة الجبرية المرافقة لها، ويعبر عن ذلك بخطين رأسيين vertical lines يوضعان حول الرقم. ويحسب بالصيغة الآتية:

أولاً- للبيانات غير المبوبة:

$$M.D = \frac{\sum |x_i - \bar{x}|}{n}$$

وبشكل عام يحسب الانحراف المتوسط باتباع الخطوات الآتية:

١- استخراج المتوسط الحسابي $\bar{X}$.

٢- نحسب انحراف القيم عن وسطها الحسابي باهمال الاشارة.

٣- جمع الانحرافات وتقسيمها على عددها (n) لنحصل على الانحراف المتوسط.

مثال:

اوجد الانحراف المتوسط من البيانات الآتية:

٨، ١٢، ٩، ٦، ١٥، ٧، ١٣

١- نستخرج الوسط الحسابي $\bar{x} = \frac{\sum x_i}{n} = \frac{70}{7} = 10$

٢- نحسب انحراف القيم عن وسطها الحسابي

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
8	-2	2
12	2	2
9	-1	1
6	-4	4
15	5	5
7	-3	3
13	3	3
$\sum 70$	0	20

قيمة الانحراف المتوسط

$$\therefore M.D = \frac{\sum |x_i - \bar{x}|}{n} = \frac{20}{7} = 2.85$$

ثانياً: للبيانات المبوبة:

للعينة $$M.D = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$$

مثال:

احسب الانحراف المتوسط من التوزيع التكراري الآتي:

الفئات	٤-٦	٦-٨	٨-١٠	١٠-١٢	١٢-١٤	١٤-١٦
التكرارات	٣	٥	٨	٦	٧	2

الحل:

الفئات C	التكرارات F_i	مركز الفئة x_i	$F_i x_i$	$\bar{x} - x_i$	$\lvert\bar{x} - x_i\rvert$	$F_i \lvert\bar{x} - x_i\rvert$
-٤	٣	٥=٢/٦+٤	١٥=٣*٥	٥- =١٠ -٥	٥	١٥=٥*٣
-٦	٥	٧=٢/٨+٦	٣٥=٥*٧	٣- =١٠ -٧	٣	١٥=٣*٥
-٨	٨	٩=٢/١٠+٨	٧٢=٨*٩	١- =١٠ -٩	١	٨=٨*١
-١٠	٦	١١=٢/١٢+١٠	٦٦=٦*١١	١=١٠ -١١	١	٦=٦*١
-١٢	٧	١٣=٢/١٤+١٢	٩١=٧*١٣	٣= ١٠ -١٣	٣	٢١=٧*٣
١٤-١٦	٢	١٥=٢/١٦+١٤	٣٠=٢*١٥	٥= ١٠ -١٥	٥	١٠=٢*٥
المجموع	٣١		٣٠٩			٧٥

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{309}{31} = 10$$

قيمة الانحراف المتوسط $$M.D = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{75}{31} = 2.4$$

ويعاب على الانحراف المتوسط بأنه غير قابل للمعالجة الجبرية؛ لكونه يهمل الإشارات الجبرية، لذا فانه لا يميز بين الانحرافات السالبة والموجبة، لأنه عديم الفائدة على الإطلاق من الناحية الإحصائية.

٤–٢–٥ التباين *Variance*:

يعد التباين من افضل مقاييس التشتت وأدقها، وأكثرها استعمالاً، ولاسيما في المجالات التطبيقية، ويعمل التباين على قياس متوسط تشتت قيم المجموعة حول وسطها الحسابي.

ويعرف التباين بانه مجموع مربعات انحرافات القيم مقسوماً على عددها، ويرمز لتباين المجتمع بـ σ^2 ولتباين العينة بـ S^2، ويحسب على وفق طبيعة البيانات وبحسب ما يأتي:

أولاً: للبيانات غير المبوبة:

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

مثال:

جد التباين من البيانات الآتية:

٣، ٧، ٩، ٥، ٨، ٤، ٦

الحل:

x_i	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
٣	٣–	٩
٧	١	١
٩	٣	٩
٥	١–	١
٨	٢	٤
٤	٢–	٤
٦	٠	٠
$\sum 42$	٠	٢٨

$$\bar{x} = \frac{\sum x_i}{n} = \frac{42}{7} = 6$$

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

$$S^2 = \frac{28}{6} = 4.66$$ قيمة التباين

ويمكن بجذر قيمة التباين ان نحصل على قيمة الانحراف المعياري أي $\sqrt{4.66}$ التي تساوي ٢.١٦.

ثانياً: للبيانات المبوبة:

$$S^2 = \frac{\sum f_i \left(x_i - \bar{x}\right)^2}{\sum f_i - 1}$$

مثال:

جد التباين من التوزيع التكراري الآتي:

الفئات	٤-٨	٨-١٢	١٢-١٦	١٦-٢٠	٢٠-٢٤
التكرارات	٣	٢	٢	٤	٢

الحل:

الفئات	التكرارات F_i	مركز الفئة x_i	$F_i x_i$	$\bar{X} - X_i$	$\left(\bar{x} - x_i\right)^2$	$f_i\left(\bar{x} - x_i\right)^2$
٤-	٣	٤+٨/٢=٦	٦*٣=١٨	-٨	٦٤	٣*٦٤=١٩٢
٨-	٢	٨+١٢/٢=١٠	١٠*٢=٢٠	-٤	١٦	٢*١٦=٣٢
١٢-	٢	١٢+١٦/٢=١٤	١٤*٢=٢٨	٠	٠	٢*٠=٠
١٦-	٤	١٦+٢٠/٢=١٨	١٨*٤=٧٢	٤	١٦	٤*١٦=٦٤
٢٠-٢٤	٢	٢٠+٢٤/٢=٢٢	٢٢*٢=٤٤	٨	٦٤	٢*٦٤=١٢٨
المجموع	١٣		١٨٢	٠		٤١٦

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{182}{13} = 14$$

$$S^2 = \frac{\sum f_i \left(x_i - \bar{x}\right)^2}{\sum f_i - 1} = \frac{416}{12} = 34.66$$

* **الانحراف المعياري** ***Standard Deviation***:

إن هذا المقياس يعد من أهم مقاييس التشتت وأكثرها شيوعاً واستعمالاً؛ لدقته وقابليته للعمليات الجبرية فضلاً عن انه يدخل في حساب كثير من المقاييس الإحصائية الأخرى.

والانحراف المعياري هو الجذر التربيعي الموجب لمتوسط مربعات انحرافات القيم عن وسطها الحسابي، ويرمز له بـ (σ) في حالة المجتمع و(S) في حالة العينة.

ويحسب بشكل عام بإتباع الخطوات الآتية:

١- استخراج الوسط الحسابي للمجتمع أو العينة.

٢- إيجاد انحرافات القيم عن وسطها الحسابي.

٣- تربيع انحرافات القيم عن الوسط الحسابي.

٤- جمع مربعات الانحرافات وإيجاد متوسطها ثم جذرها للحصول على الانحراف المعياري.

والانحراف المعياري يحسب للبيانات غير المبوبة والمبوبة على وفق الصيغ الآتية:

أولاً: للبيانات غير المبوبة:

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

مثال:

احسب الانحراف المعياري من البيانات الآتية:

٣، ٤، ٩، ٥، ٧، ٨، ٦الحل: بالطريقة المطولة

$(\bar{x}-x_i)^2$	$\bar{x}-x_i$	x_i
9	3-	3
4	2-	4
9	3	9
1	1-	5
1	1	7
4	2	8
0	0	6
28	0	المجموع ٤٢

$$\overline{X} = \frac{\sum x_i}{n} = \frac{42}{7} = 6$$

$$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{28}{6}} = \sqrt{4.66}$$

$$\therefore S = 2.16$$

ثانياً: للبيانات المبوبة:

$$S = \sqrt{\frac{\sum f_i (x_i - \overline{x})^2}{\sum f_i - 1}}$$

مثال:

احسب الانحراف المعياري من جدول التوزيع التكراري الآتي:

الفئات	٦٢-٦٠	٦٥-٦٣	٦٨-٦٦	١-٦٩	٧٤-٧٢
التكرارات	٥	١٨	٤٢	٢٧	٨

الحل بالطريقة المطولة:

الفئات	التكرارات F_i	مركز الفئة x_i	$F_i x_i$	$\overline{x} - x_i$	$(\overline{x} - x_i)^2$	$f_i(\overline{x} - x_i)^2$
٦٢-٦٠	٥	٦١	٣٠٥	٦.٤٥-	٤١.٦٠٢	٢٠٨.٠١٢
٦٥-٦٣	١٨	٦٤	١١٥٢	٣.٤٥-	١١.٩٠٢	٢١٤.٢٤٥
٦٨-٦٦	٤٢	٦٧	٢٨١٤	٠.٤٥-	٠.٢٠٢	٨.٥٠٥
٧١-٦٩	٢٧	٧٠	١٨٩٠	٢.٥٥	٦.٥٠٢	١٧٥.٥٦٧
٧٤-٧٢	٨	٧٣	٥٨٤	٥.٥٥	٣٠.٨٠٢	٢٤٦.٤٢٠
المجموع	١٠٠		٦٧٤٥			٨٥٢.٧٥٠

$$\overline{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{6745}{100} = 67.45$$

$$S = \sqrt{\frac{\sum f (x - \overline{x})^2}{\sum f_i - 1}} = \sqrt{\frac{852.750}{99}}$$

$$S = \sqrt{8.613}$$

قيمة الانحراف المعياري $\therefore S = 2.93$

تحليل الارتباط والانحدار ***Correlation Analysis & Regression***

تم التطرق في الفصول السابقة إلى الطرائق والأساليب الإحصائية المختلفة لجمع البيانات وتصنيفها وتبويبها واستخراج بعض المقاييس الإحصائية الوصفية التي يمكن من خلالها إعطاء فكرة واضحة عن طبيعة تلك البيانات ومنها االمتوسطات ومقاييس التشتت المختلفة.

إن جميع هذه الطرق قد اعتمدت على البيانات المجمعة من متغير واحد فقط (Y) أو (X)، ولكن في كثير من الأحيان يواجه الباحث حالات تتطلب دراسة متغيرين أو أكثر في آن واحد للتعرف على طبيعة والعلاقة التي ترتبط بها تلك المتغيرات ونوعها وقوتها.

وبشكل عام تقاس تلك العلاقة إحصائياً باستعمال أسلوبين رئيسين هما: الارتباط (Correlation) والانحدار (Regression) وهما أسلوبان متشابهان في نواحٍ كثيرة جداً ولكنهما مختلفان في عدة نواح. فالارتباط يقيس درجة العلاقة واتجاهها بين المتغيرين، بينما الانحدار يبحث في العلاقة بين المتغيرات من خلال معادلة رياضية يمكن بوساطتها تفسير أو تقدير أو التنبؤ بأحد المتغيرين من خلال المتغير الآخر.

اولا: تحليل الارتباط ***Correlation Analysis:***

الارتباط هو أداة من أدوات التحليل الوصفي لمعرفة العلاقة بين متغيرين مستقلين (x_2, x_1) يمثل كل منهما ظاهرة معينة، أو بين متغير مستقل واحد (X) ومتغير معتمد (Y). أو بين (Y) ومجموعة من المتغيرات المستقلة $(x_n .....x_1)$.

والارتباط هو وصف قوة العلاقة بين المتغيرات المتعددة في تفسير بعضها، لتحديد مدى تأثر هذه المتغيرات بعضها ببعض ليحدد بذلك أو ليصف العلاقة (الترابطية) بين المتغيرات.

لذا فالارتباط هو وصف درجة تأثر احد المتغيرين بالاخر وبيان مدى العلاقة الواقعة بين هذين المتغيرين.

والمقياس الأساس لهذه العلاقة يدعى معامل الارتباط Correlation Coefficient وهو مقياس كمي لقياس قوة هذه العلاقة ويرمز له في حالة المجتمع بـ(ρ) "رو" وفي حالة العينة بـ"r" الذي تتراوح قيمته بين (الصفر و±١). وبما اننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع الإحصائي؛ لذا فاننا سنركز على حساب معامل ارتباط العينة "r" كونه قيمة تقديرية لمعامل ارتباط المجتمع المأخوذة منه العينة.

ان معامل الارتباط بشكل عام يركز على نقطتين اساسيتين هما:

١- نوع العلاقة التي تأخذ ثلاثة اشكال اعتماداً على اشارة معامل الارتباط وهي:

أ- العلاقة الطردية "الموجبة" بين المتغيرين أي " $r > 0$ " إذ أن زيادة احد المتغيرين (x) يصاحبه زيادة في المتغير الثاني (y) أو العكس يصاحبه نقص في المتغير الأول (x) يقابله نقص في المتغير الثاني (y).

ب- العلاقة العكسية "السالبة" بين المتغيرين أي " $r < 0$ " إذ أن أي زيادة في المتغير الأول (x) يقابلها نقص في المتغير الثاني (y) أو العكس من ذلك، أن أي نقص في المتغير الأول (x) يقابله زيادة في المتغير الثاني (y).

جـ- انعدام العلاقة بين المتغيرين أي " $r = 0$ " إذ أن أي زيادة أو نقص في المتغير الأول "x" لا يؤدي إلى أي تغير في المتغير الثاني "y".

٢- قوة العلاقة التي يمكن الحكم عليها من خلال درجة قرب أو بعد قيمة r عن (±١)، إذ أن قيمة معامل الارتباط تقع في المدى $(-1 < r < 1)$. وكلما اقتربت قيمة معامل الارتباط من الصفر دلت على عدم وجود علاقة خطية بينهما، بمعنى قد

تكون هناك علاقة الا انها غير خطية. ولقد صنف بعض الإحصائيين درجات قوة العلاقة بالشكل الآتي:– **درجات قوة معامل الارتباط**

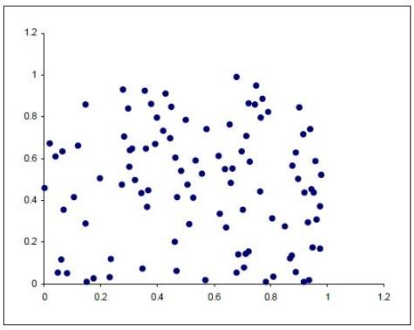
ارتباط موجب					ارتباط سالب				
قوي جداً	قوي	متوسط	ضعيف	ضعيف جداً	ضعيف جداً	ضعيف	متوسط	قوي	قوي جداً
	٠.٩	٠.٧	٠.٥	٠.٣	-٠.٣	-٠.٥	-٠.٧	-٠.٩	

+١ ٠ -١

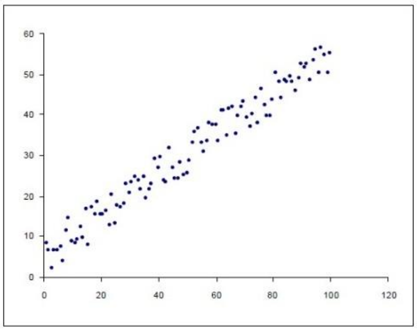
ارتباط تام **عدم وجود ارتباط** **ارتباط تام**

١–١ الشكل الانتشاري لتحديد طبيعة الاتجاه العام للارتباط:

ان تحديد الارتباط ورسم الشكل الانتشاري يعتمدان على وجود متغيرين يمكن ان يكون احدهما مستقلاً X والاخر معتمداً "تابع" Y، ولكل قيمة من المتغير X توجد قيمة تقابلها من المتغير Y. فاذا ما تم جمع البيانات عن ازواج قيم هذين المتغيرين، وتم تمثيلهما بيانياً فيما يسمى بشكل الانتشار Scatter diagram الذي يمكن من خلاله تكوين فكرة عامة تساعد الفحص البصري للباحث للتعرف على طبيعة العلاقة التي تربط بين هذين المتغيرين، فاذا كانت تلك القيم مبعثرة "متباعدة" دل ذلك على ضعف العلاقة التي تربط بين المتغيرين، اما اذا كانت القيم محددة "متقاربة" ويمكن تمثيلها بخط مستقيم دل ذلك على وجود علاقة بين المتغيرين، ان هذه المعلومات ذات درجة عالية من الاهمية في التحليل الإحصائي واتخاذ القرارات، عندما تظهر المتغيرات درجة عالية من الترابط، فيمكن الافتراض بوجود علاقة بين المتغيرات، وبشكل عام فان هذه العلاقة يمكن ان تأخذ واحداً من الأشكال الآتية:



عدم وجود ارتباط



ارتباط خطي موجب

أمثلة من أشكال الانتشار

٢- أنواع الارتباط:

إن الارتباط الذي يمثل الظواهر التي يمكن التعبير عنها كمياً "عددياً" يمكن تقسيمه إلى ثلاثة أنواع تبعاً لعدد المتغيرات التي يتضمنها وهي:

١- الارتباط البسيط ***Simple correlation*:**

إن هذا النوع من الارتباط يهتم بدراسة العلاقة بين متغيرين احدهما مستقل (x) والأخر معتمد (y).

٢- الارتباط المتعدد ***Multiple correlation*:**

وهذا النوع يهتم بدراسة العلاقة بين اكثر من متغيرين مستقلين ومتغير معتمد.

٣- الارتباط الجزئي ***Partial correlation*:**

وهو الذي يهتم بدراسة العلاقة بين زوج من المتغيرات فقط من بين مجموعة من المتغيرات الاخرى التي يتم تثبيت تأثيرها عن طريق استبعادها أو عزلها. ان الفرق بين الارتباط البسيط والارتباط الجزئي هو ان الاول يقيس قوة العلاقة واتجاهها بين متغيرين ضمن تأثيرات المتغيرات الاخرى، في حين يقيس الثاني قوة العلاقة واتجاهها بين متغيرين بعد استبعاد تأثير المتغيرات الاخرى.

فمثلاً اذا كان لدينا ثلاثة متغيرات Y, X_2, X_1 فمن الممكن قياس الارتباط الجزئي بين أي اثنين منها وعزل اثر المتغير الثالث باستعمال معامل الارتباط الجزئي.

٣ – مقاييس الارتباط:

بعد الاستعانة بالشكل الانتشاري يمكن للباحث من التحديد المبدئي "الاولي" لنوع الارتباط بين المتغيرات "ارتباط خطي موجب أو سالب أو ارتباط غير خطي" ولقياس نوع تلك العلاقة وقوتها نستعمل مقاييس خاصة تسمى بمقاييس الارتباط التي تقسم إلى نوعين أساسيين هما:

٣–١ معامل الارتباط للظواهر المقيسة:

ويشمل دراسة العلاقة بين الظواهر القابلة للقياس الكمي "الرقمي" وهذا يشمل جميع الظواهر التي يمكن التعبير عنها بصورة رقمية ومنها الطول والوزن، وكمية الأمطار المتساقطة، والإنتاج الزراعي، وغيرها من الظواهر التي يمكن التعبير عنها رقمياً.

وتقسم إلى الأنواع الآتية:

١– معامل الارتباط البسيط.

٢– معامل الارتباط المتعدد.

٣– معامل الارتباط الجزئي.

٣–١–١ معامل الارتباط البسيط *Simple correlation coefficient*

يستعمل لقياس العلاقة بين متغيرين ذوي قيم مستمرة Contiunous variables، ويعد معامل ارتباط بيرسون Pearson من اهم مقاييس الارتباط وأقواها ولاسيما عندما تكون العلاقة بين المتغيرين خطية Linear، وكثيراً ما يستعمل هذا المعامل في المجالات التطبيقية ومنها العلاقة بين الانتاج والكلفة، والاستهلاك والدخل، والطول والوزن، والانتاج الزراعي والمطر، والعلاج والمرض وغيرها.

ان الصيغة الاساسية لمعامل بيرسون لحساب ارتباط العينة هي:

$$r = \frac{\sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{n}}{\sqrt{\sum Y_i^2 - \frac{\sum (Y_i)^2}{n}}\sqrt{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}}$$

مثال:

البيانات الآتية تمثل الكمية المعروضة من سلعة ما وسعر الوحدة الواحدة منها.

الكمية المعروضة (Y):	٣	٥	٧	٢	٩	١١	٦	٨	٦
سعر الوحدة (X):	٢	٢	٥	٤	٥	٦	٣	٦	٤

المطلوب:

احسب معامل الارتباط البسيط (بيرسون) بين الكمية المعروضة والسعر.

الحل:

Y_i	X_i	Y_iX_i	Y_i^2	X_i^2
3	2	6	9	4
5	2	10	25	4
7	5	35	49	25
8	4	32	64	16
9	5	45	81	25
11	6	66	121	36
6	3	18	36	9
8	6	48	64	36
6	4	24	36	16
$\sum 63$	37	284	485	171

$$r = \frac{\sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{n}}{\sqrt{\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}}\sqrt{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}}$$

$$r = \frac{284 - \frac{(63)(37)}{9}}{\sqrt{485 - \frac{(63)^2}{9}}\sqrt{171 - \frac{(37)^2}{9}}}$$

$$r = \frac{284 - 259}{\sqrt{485 - 441}\sqrt{171 - 152}}$$

$$r = \frac{25}{\sqrt{44}\sqrt{19}} = \frac{25}{(6.63)(4.35)} = \frac{25}{28.59} = 0.86$$

وهذا يدل على ان العلاقة بين الكمية المعروضة والسعر هي علاقة موجبة وقوية في الوقت نفسه.

٦-١-٣-١-٢ معامل الارتباط المتعدد

Multiple correlation coefficient

في معظم الدراسات العلمية النظرية والتطبيقية والتخطيطية لا تعتمد علاقة الارتباط على متغيرين احدهما يمثل المتغير المعتمد "التابع" (Y) والاخر يمثل المتغير المستقل (X)، بل يمتد ليشمل عدداً من المتغيرات المستقلة التي تؤثر بشكل أو باخر على المتغير المعتمد. فمثلاً إنتاج محصول زراعي معين يعتمد على (خصوبة التربة، وكمية المياه، وكمية الاسمدة، ودرجات الحرارة، والسطوع الشمسي) وغيرها، فضلاً عن أن الانتاج الصناعي لسلعة ما يتوقف على (حجم الاستثمار، والمواد الاولية، والأيدي العاملة، والطاقة المستعملة) وغيرها.

لذا فان انتاج المحصول الزراعي أو الانتاج الصناعي لاية سلعة هو المتغير المعتمد (Y) بالنسبة الى المتغيرات المستقلة الاخرى التي تؤثر عليه، لذا فان دراسة العلاقة بين المتغير المعتمد (Y) ومجموعة المتغيرات المستقلة $(X_n ... X_3, X_2, X_1)$ في آن واحد هو ما يطلق عليه بالارتباط المتعدد الذي يمكن قياسه بمعامل الارتباط المتعدد الذي يقيس قوة العلاقة بين اكثر من متغيرين من المتغيرات العشوائية المتصلة التوزيع "توزيع متعدد Multivariate distribution". ويرمز لهذا العامل بالحرف R. ان حساب قيمة هذا المعامل ما هو الا امتداد لحساب قيمة معامل الارتباط البسيط (r) بإضافة متغيرات مستقلة اخرى له. وقيمة معامل الارتباط المتعدد تقع ايضاً بين الصفر و ± 1 (الصفاوي، ٢٠٠٨، ص٣١٥).

وكلما كانت قيمته قريبة من الواحد دل ذلك على قوة العلاقة بين المتغير المعتمد والمتغيرات المستقلة وبالعكس اذا كانت قيمته قريبة من الصفر. سوف نقتصر في دراستنا لهذا المعامل على العلاقة الخطية بين ثلاثة متغيرات Y و X_1 و X_2 لنحصل على الصيغ الآتية:

$$r_{YX_1} = \frac{n\sum YX_1 - \sum Y \sum X_1}{\sqrt{N\sum Y_I^2 - \left(\sum Y_I\right)^2}\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}}$$

$$r_{YX_2} = \frac{n\sum YX_2 - \sum Y \sum X_2}{\sqrt{n\sum Y_1^2 - \left(\sum Y_1\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$r_{X_1X_2} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

لذا فان الصيغة النهائية لحساب معامل الارتباط المتعدد هي:

$$R_{YX_1X_2} = \sqrt{\frac{r^2{}_{YX_1} + r^2{}_{YX_2} - 2r_{YX_1} r_{YX_2} r_{X_1X_2}}{1 - r^2{}_{X_1X_2}}}$$

مثال:

جد قيمة معامل الارتباط المتعدد من الجدول الآتي:

Y	**X_1**	**X_2**
2	1	1
3	6	8
2	5	2
1	7	6
4	10	8

الحل:

العمليات اللازمة للحصول على معامل الارتباط المتعدد بين Y والمتغيرين X_1 وX_2.

Y	X_1	X_2	YX_1	YX_2	X_1X_2	Y_i^2	X_1^2	X_2^2
2	1	1	2	2	1	4	1	1
3	6	8	18	24	48	9	36	64
2	5	2	10	4	10	4	25	4
1	7	6	7	6	42	1	49	36
4	10	8	40	32	80	16	100	64
$\sum$12	29	25	77	68	181	34	211	169

$$r_{YX_1} = \frac{n\sum YX_1 - \sum Y \sum X_1}{\sqrt{n\sum Y_i^2 - \left(\sum Y_i\right)^2}\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}}$$

$$= \frac{5(77)-(12)(29)}{\sqrt{5(34)-(12^2)}\sqrt{5(211)-(29)^2}}$$

$$= \frac{383-348}{\sqrt{(170-144)}\sqrt{(1055-841)}} = \frac{37}{\sqrt{26}\sqrt{214}} = \frac{37}{74.5} = 0.49$$

$$r_{YX_2} = \frac{n\sum YX_2 - \sum Y \sum X_2}{\sqrt{n\sum Y_i^2 - \left(\sum Y_i\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$= \frac{5(68)-(12)(25)}{\sqrt{5(34)-(12)^2}\sqrt{5(169)-(25)^2}} = \frac{40}{}$$

$$\frac{340-300}{\sqrt{(170-144)}\sqrt{(845-625)}} = \frac{40}{\sqrt{26}\sqrt{220}} = \frac{40}{75.6} = 0.53$$

$$r_{X_1X_2} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_2^2 - (\sum X_2)^2}}$$

$$r_{X_1X_2} = \frac{5(181)-(29)(25)}{\sqrt{\sum 5(211)-(29)^2}\sqrt{5(169)-(25)^2}}$$

$$= \frac{905-725}{\sqrt{(1055-841)}\sqrt{842-625}} = \frac{180}{\sqrt{214}\sqrt{220}} = \frac{180}{217} = 0.83$$

$$R_{YX_1X_2} = \sqrt{\frac{r^2{}_{YX_1} + r^2{}_{YX_2} - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1-r^2{}_{X_1X_2}}}$$

$$= \sqrt{\frac{(0.49)^2 + (0.53)^2 - 2(0.49)(0.53)(0.83)}{1-(0.83)^2}}$$

$$= \sqrt{\frac{0.240 + 0.280 - 2(0.431)}{0.312}}$$

$$= \sqrt{\frac{0.52 - 0.431}{0.312}}$$

$$= \sqrt{\frac{0.089}{0.312}} = \sqrt{0.285} = 0.53$$

من قيمة معامل الارتباط المتعدد اعلاه البالغة ٠.٥٣ يمكن القول بأنه توجد علاقة موجبة بين المتغير المعتمد (Y) والمتغيرات المستقلة (X_1 و X_2) ولكنها علاقة ضعيفة.

٣–١–٣ الارتباط الجزئي *Partial Correlation*:

لقد عرفنا سابقاً أن الارتباط المتعدد يدرس العلاقة بين المتغير المعتمد "التابع" ومجموعة من المتغيرات المستقلة في آن واحد. أما الارتباط الجزئي فانه يدرس العلاقة بين متغيرين اثنين فقط من تلك المتغيرات عندما تكون باقي المتغيرات الأخرى ثابتة بعزلها أو استبعاد أثرها إحصائياً.

فمثلاً إذا كان لدينا ظاهرة معينة (Y) وهناك مجموعة من المتغيرات المستقلة (X_4,X_3,X_2,X_1) تؤثر بشكل متفاوت على هذه الظاهرة فإيجاد معامل الارتباط الجزئي بين المتغيرين X_3,X_1 سيتم مع استبعاد اثر المتغيرين X_4,X_2، والغرض الأساس من هذا الاستبعاد هو لمعرفة طبيعة العلاقة بينهما ومدى جدوى بقاء احدهما أو كليهما على وفق درجة تأثيرهما على المتغير المعتمد (Y).

إن حساب قيمة معامل الارتباط الجزئي يتم على وفق الخطوات الآتية:

١- ايجاد معاملات الارتباط البسيط (r) بين المتغيرات المدروسة.

٢- ايجاد معامل الارتباط الجزئي من الصيغة الآتية:

أ- صيغة معامل الارتباط الجزئي بين Y وx_1 بثبات X_2

$$R_{YX_1X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r^2{}_{X_1X_2}}\sqrt{1 - r^2{}_{YX}}}$$

ب- صيغة معامل الارتباط الجزئي بين Y وX_2 بثبات X_1.

$$R_{YX_2X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{1 - r^2{}_{X\ X_2}}\sqrt{1 - r^2{}_{YX_1}}}$$

مثال:

قام باحث باختيار عينة من ستة عمال يعملون في احد مصانع الاغذية وقد جمع معلومات عن مقدار الانتاج الشهري لهؤلاء العمال الستة، ومدة الخدمة الفعلية لهم في المصنع، وكانت بيانات العينة بحسب الآتي:

مقدار الانتاج (بالطن) لكل عامل Y: ١ ٤ ٤ ٥ ٨ ٩

الطاقة التصميمية للانتاج (بالطن) X_1: ٣ ٥ ٨ ١٢ ١٢ ١٥

مدة الخدمة الفعلية لكل عامل (بالسنة) X_2: ٢ ٤ ٦ ٨ ٩ ٩

المطلوب: ايجاد معاملات الارتباط الجزئي وبيان ايهما اكثر تأثيراً على الانتاج.

الحل:

ان حساب معامل الارتباط الجزئي يتطلب اولاً حساب معاملات الارتباط البسيط بين المتغيرات الثلاث، وهذا يقتضي تكوين الجدول الاتي:

Y	X_1	X_2	YX_1	YX_2	X_1X_2	Y^2	X_1^2	Y_2^2
1	3	2	3	2	6	1	9	4
4	5	4	20	16	20	16	25	16
4	8	6	32	24	48	16	64	36
5	12	8	60	40	96	25	144	64
8	12	9	96	72	108	64	144	81
9	15	9	135	81	135	81	225	81
$\sum 31$	55	38	346	235	413	203	611	282

$$r_{YX_1} = \frac{\sum X_1 Y - \frac{\sum X_1 \sum Y}{n}}{\sqrt{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

$$= \frac{346 - \frac{(55)(31)}{6}}{\sqrt{611 - \frac{(55)^2}{6}}\sqrt{203 - \frac{(31)^2}{6}}} = \frac{62}{67.8} = 0.91$$

$$r_{YX_2} = \frac{\sum X_2 Y - \frac{\sum X_2 \sum Y}{n}}{\sqrt{\sum X_2^2 - \frac{(\sum X_2)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

$$r_{YX_2} = \frac{235 - \frac{(38)(31)}{6}}{\sqrt{282 - \frac{(38)^2}{6}}\sqrt{203 - \frac{(31)^2}{6}}} = \frac{38.67}{42} = 0.92$$

$$r_{X_1X_2} = \frac{\sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n}}{\sqrt{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}\sqrt{\sum X_2^2 - \frac{(\sum X_2)^2}{n}}}$$

$$= \frac{413 - \frac{(55)(38)}{6}}{\sqrt{611 - \frac{(55)^2}{6}}\sqrt{282 - \frac{(38)^2}{6}}} = \frac{64.7}{66.4} = 0.97$$

بعد استخراجنا قيم معامل الارتباط البسيط للمتغيرات الثلاث البالغة $r_{X_1} = 0.91$، $r_{X_2} = 0.92$، $r_{X_1X_2} = 0.97$ نبدأ بتطبيق قانون الارتباط الجزئي $YX_1.X_2$ بتثبيت X_2 و $YX_2.X_1$ بتثبيت X_1.

$$r_{YX_1X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1-(r_{X_1X_2})^2}\sqrt{1-(r_{YX_2})^2}}$$

$$= \frac{0.91 - (0.92)(0.97)}{\sqrt{1-(0.97)^2}\sqrt{1-(0.92)^2}} = \frac{0.02}{0.097} = 0.20$$

$$R_{YX_2X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{1-\left(r_{X_1X_2}\right)^2}\sqrt{1-\left(r_{YX_1}\right)^2}}$$

$$= \frac{0.92-(0.91)(0.97)}{\sqrt{1-(0.97)^2}\sqrt{1-(0.91)^2}} = \frac{0.04}{0.100} = 0.4$$

من نتائج الارتباط الجزئي اعلاه يمكن القول بأن المتغير الثاني "مدة الخدمة الفعلية للعمال" اكثر اهمية في التأثير على كمية الانتاج (Y) من متغير الطاقة التصميمية للمصنع.

٣–٢ معامل الارتباط للظواهر غير المقيسة:

توجد بعض الظواهر لا يمكن قياسها كمياً "رقمياً" وقد تكون على شكل صفات أو على شكل رتب ومن هذه الظواهر، الحالة الاجتماعية، والحالة الصحية لافراد المجتمع، والتدخين، والإشباع، والذكاء وغيرها، وهذه الظواهر لا يوجد مقياس رقمي لقياسها وكل ما نستطيع ان نقوم به هو تصنيف افراد المجتمع من حيث الحالة الصحية مثلاً الى اصناف متدرجة، اما من الاعلى الى الادنى أو بالعكس. مثلاً ابتداءً من الحالة الصحية الممتازة وانتهاء بالحالة الصحية السيئة، وهكذا نطبقه على بقية الظواهر المماثلة الاخرى. ومن اهم تلك المقاييس هي:

١– معامل ارتباط الرتب لسبيرمان.

٢– معامل ارتباط الرتب لكيندل.

٣–٢–١ معامل ارتباط الرتب لسبيرمان

Spearmiain's Rank correlation coefficient

وهو من المقاييس المهمة والشائعة الاستعمال في الحالات الآتية:

١– اذا كان كلا المتغيرين أو احدهما من النوع الترتيبي ordinal variables القابل للترتيب التصاعدي والتنازلي.

٢– إذا كان كلا المتغيرين أو أحدهما لا يتبع التوزيع الطبيعي، أو في حالة البيانات اللامعلمية ويعتبر كمعامل بديل لمعامل ارتباط بيرسون.

وقد اكتسب معامل سبيرمان في الدراسات الجغرافية مكانة منمازة نظراً لكون كثير من البيانات الجغرافية هي بيانات غير كمية، وامتد استعماله الى البيانات الكمية، وان قوته في قياس الارتباط لا تقل عن ٠.٩١ من قوة معامل بيرسون. وقد استنبط سبيرمان معادلة لحساب معامل ارتباط الرتب الذي يرمز له بالرمز (r_s) وعلى وفق ما يأتي:

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

إذ ان:

n تمثل عدد ازواج الرتب.

di تمثل الفرق بين رتب مستويات المتغير الاول (Y) ورتب مستويات المتغير الثاني (X).

١، ٦ تمثل ثوابت مبرهنة بصرف النظر عن n و di.

مثال A:

الجدول الآتي تقديرات لكفاءة اداء خمسة من العاملين في مصنع ما وتحصيلهم الدراسي:

كفاءة الاداء (y)	جيد جداً	ضعيف	ممتاز	جيد	متوسط	مقبول
التحصيل الدراسي (X)	دبلوم	متوسطة	بكالوريوس	يقرأ ويكتب	ثانوية	ابتدائية

المطلوب:

احسب قيمة معامل الارتباط بين كفاءة الاداء والتحصيل الدراسي وما هي مدلولاته؟

الحل:

ترتب التقديرات تصاعدياً بالشكل كالآتي:

كفاءة الاداء Y	ضعيف	مقبول	متوسط	جيد	جيد جداً	ممتاز
الرتب	١	٢	٣	٤	٥	٦
التحصيل الدراسي X	يقرأ ويكتب	ابتدائية	متوسطة	ثانوية	دبلوم	بكالوريوس
الرتب	١	٢	٣	٤	٥	٦

di^2	$di=y_i-x_i$	رتب x	رتب y	X	y
صفر	صفر	٥	٥	دبلوم	جيد جداً
٤	−٢	٣	١	متوسط	ضعيف
صفر	صفر	٦	٦	بكالوريوس	ممتاز
٩	٣	١	٤	يقرأ ويكتب	جيد
١	−١	٤	٣	ثانوية	متوسط
صفر	صفر	٢	٢	ابتدائية	مقبول
١٤	صفر				$\sum$

$$r_s = 1-\frac{6\sum di^2}{n(n^2-1)} = 1-\frac{6(14)}{6(36-1)} = 1-\frac{84}{210} = 1-0.4 = 0.6$$

أي ان هناك ارتباط موجب ومتوسط بين كفاءة الاداء والتحصيل الدراسي.

مثال B:

البيانات الآتية تمثل معدل انتاج الدونم من محصول الحنطة في المناطق شبه المطرية وكمية الامطار المتساقطة للمدة من ٢٠٠٥–٢٠٠٩.

السنة:	٢٠٠٥	٢٠٠٦	٢٠٠٧	٢٠٠٨	٢٠٠٩
معدل انتاج الدونم/ كغم:	١٥٠	٢٠٠	١٢٠	١٨٠	٩٠
كمية الامطار المتساقطة/ كم:	٢٥	٣٠	٢٠	٢٨	٢٦

المطلوب:

اوجد معامل ارتباط الرتب (سبيرمان) بين معدل انتاج الدونم من محصول الحنطة وكمية الامطار المتساقطة.

الحل:

نقوم باعداد الجدول الآتي لتسهيل العمليات الحسابية لمعامل ارتباط سبيرمان.

Y معامل الانتاج	X كمية الامطار	ترتيب Y	ترتيب X	$d_i=y_i-y_i$	d_i^2
١٥٠	٢٥	٣	٤	1-	1
٢٠٠	٣٠	١	١	صفر	صفر
١٢٠	٢٠	٤	٥	1-	1
١٨٠	٢٨	٢	٢	صفر	صفر
٩٠	٢٦	٥	٣	2	4
				صفر	6

$$r_s = 1 - \frac{6\sum di^2}{n(n^2-1)}$$

$$= 1 - \frac{6(6)}{5(25-1)}$$

$$= 1 - \frac{36}{120}$$

$$= 1 - 0.3$$

$$= 0.7$$

وتدل قيمة معامل ارتباط سبيرمان اعلاه على وجود علاقة موجبة أو متوسطة بين انتاتج القمح وكمية الامطار المتساقطة في تلك السنوات.