وزارة التعليم العالي والبحث العلمي

جامعة تكريت

كلية الإدارة والاقتصاد

قسم ادارة الاعمال

المرحلة الاولى

محاضرات في مادة الاحصاء

مدرس المادة

م.د ثائر جاسم مُحَمَّد

1445 هـ **2023 مـ**

الفصـــــل الأول

المبحث الاول

((معنى الإحصاء وأهميته في البحوث والدراسات العلمية))

1- تعريف الإحصاء.

علم الإحصاء :هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهره او فرضيه معينه وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج باتخاذ القرار على ضوء ذلك ، وبشكل عام فأن علم الإحصاء يقسم الى القسمين الرئيسيين الآتيين :

1- **الإحصاء الوصفي** : وهو الفرع الذي يتضمن الطرق والأساليب المستخدمة في جمع البيانات عن ظاهرة معينة او مجموعه ظواهر وكيفية تنظيمها وتصنيفها وتبويبها وعرضها وحساب المؤشرات الإحصائية منها.

2- **الإحصاء الاستدلالي** : وهو الفرع الآخر لعلم الإحصاء الذي يهتم بموضوع التقدير واختبار الفرضيات.

2- اهميه علم الاحصاء ومجالات تطبيقاته

يعتبر علم الاحصاء احد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه وطرقه في جمع البيانات اللازمة للبحث العلمي وتحليل هذه البيانات بغيه الوصول الى النتائج التي يهدف لها البحث . ويمكن تمثيل مجالات تطبيق علم الاحصاء حسب ما هو موضح ادناه:

1- البحوث الزراعية التطبيقية.

2- البحوث الصناعية التطبيقية.

3- البحوث النفسية والتربوية.

4- البحوث الطبية.

5- بحوث التربية الرياضية.

6- البحوث الاقتصادية والإدارية.

7- البحوث الهندسية.

8- بحوث علم الأرض ، علم الأحياء (المايولوجيا) ، الصيدلة ، علم الفيزياء ، معالجة الصور الرقمية ... الخ.

3- الطرق الإحصائية في البحث العلمي

ان استخدام الاسلوب الإحصائي في البحث العلمي يعني توفر بيانات ومعلومات عن الظاهرة أو الظواهر المطلوب دراستها في البحث العلمي ، لذلك فأن إمكانية تطبيق الطرق الإحصائية مرهون بإمكانية التعبير عن هذه الظاهرة تعبيراً كمياً.

4- المراحل الرئيسية للطرق الإحصائية في البحث العلمي :

أ- تحديد مشكلة البحث أو فرضية البحث أو الدراسة.

ب- جمع البيانات والمعلومات عن الظاهرة أو الظواهر ذات العلاقة بالبحث أو الدراسة.

ج- تصنيف البيانات وتبويبها وعرضها.

د- تحليل معطيات الدراسة للوصول إلى النتائج في ضوء فرضية البحث أو الدراسة.

هـ- حساب المؤشرات الإحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.

و- تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث.

5- أسلوب تصميم البحث :

توجد اعتبارات كثيرة يتوقف عليها تصميم البحث العلمي يتوجب على الباحث اخذها بنظر الاعتبار وأهمها:

أ- تحديد الغرض من البحث : يجب تحديد الغرض من البحث بشكل واضح ودقيق لغرض الاستفادة من النتائج التي يتوصل إليها الباحث ، فمثلاً إذا كان البحث يدرس نمط استهلاك الفرد من سلعة ما مثلاً(لحم الغنم) فأنه يجب دراسة الطلب على لحم الغنم من قبل المستهلكين.

ب- تحديد اطار البحث: احد الامور المهمة قبل البدء بتنفيذ البحث هو تحديد نوع وطبيعة مجال ذلك البحث ، اي ما نعنيه المجتمع الاحصائي على نحو واضح ودقيق، *والمجتمع الاحصائي* هو عباره عن مجموعه من الوحدات او المفردات التي تشترك بصفه او صفات معينه والتي غالبا ما يتم الحصول منها على البيانات والمعلومات المطلوبه . فمثلا اذا كان البحث يهدف الى التعرف على نسبة المرضى المصابين بمرض معين في قرية ما ، فان المجتمع الاحصائي هو افراد القرية ، اما ما يسمى بالمفردة الاحصائية فهي تمثل الفرد الذي يمكن فحصة من افراد القرية . والمجتمع الاحصائي يمكن ان يكون محدد او غير محدد. *فالمجتمع الاحصائي المحدد* هو الذي يمكن ملاحظة او مواجهة كل مفرده من مفرداته اي يمكن حصر كافة المفردات التي تنتمي له . مثلا طلبه جامعه تكريت .اما *المجتمع الاحصائي الغير محدد* هو الذي لا يمكن ملاحظة او مواجهة كل مفرداته اي لا يمكن حصر كافة مفرداته. مثلا كريات الدم الحمراء في جسم الانسان.

ت- تحديد امكانية التنفيذ الفعلي للبحث: من الضروري جدا تحديد المتطلبات التي تستلزمها عملية تنفيذ البحث وبشكل واضح ودقيق كالموارد المالية المطلوبة عند التنفيذ والامكانات البشرية المتاحة

المطلوبة لتحقيق بعض فقرات البحث كذلك التأكد من مدى توفر البيانات والمعلومات الدقيقة عن مشكلة البحث.

ث- تحديد اسلوب جمع البيانات والمعلومات : لغرض الحصول على البيانات والمعلومات التي يتطلبها البحث او الدراسة يجب تحديد نوع او اسلوب جمع هذه البيانات و المعلومات .

<u>6- أساليب جمع البيانات : هنالك نوعان من اساليب جمع البيانات هي :</u>

أ-اسلوب التسجيل الشامل لكافة مفردات المجتمع الاحصائي/

وهو جمع البيانات والمعلومات عن كافة المفردات التي تؤلف المجتمع الإحصائي الظاهرة (أو الظواهر) قيد البحث ، ويجب أن يكون المجتمع محدد مثل عملية التعداد العام للسكان في العراق ، وهو أفضل أسلوب لكونه يجهز الباحث ببيانات كاملة عن مفردات المجتمع والدراسة ، إلا أنه يحتاج إلى موارد مادية وبشرية كبيرة ، وكذلك احتمال الوقوع في أخطاء أكبر لكبر العينة وضخامة البيانات.

ب- أسلوب التسجيل عن طريق العينات /

يقصد بأسلوب العينات عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات مجتمع الدراسة ، وهذه المجموعة من المفردات تسمى (عينة) مثل استفتاء رأي ، دراسة فاعلية علاج معين للمصابين بمرضى القلب ، إن هذا الأسلوب يحتاج إلى جهد وموارد مالية وبشرية أقل من التسجيل الشامل ، وكذلك فأنه مفيد عند دراسة المجتمعات غير المحددة ، إلا أن دقته تعتمد على النتائج المستخلصة من بيانات العينة.

7- <u>وسائل جمع البيانات</u> :

بعد تحديد حجم العينة وأسلوب المعاينة الملائم في اختيار مفردات هذه العينة من مجتمع ما ، يتطلب اختيار الوسيلة الملائمة في جمع البيانات وهنالك عدة وسائل لجمع البيانات أهمها :

أ- أسلوب الجمع المباشر :

وفق هذا الاسلوب يتم جمع البيانات والمعلومات المتوفرة لدى جهات معينة كأجهزة الدولة أو الهيئات ذات العلاقة مثل الجهاز المركزي للإحصاء ، وكذلك البيانات ذات الطابع المختبري أو الحقلي التي يتم من قبل الباحث.

ب- الاستبيان :

هو عبارة عن استمارة تحتوي على مجموعة من الأسئلة يتم من خلالها جمع البيانات أو المعلومات من مفردات (أو بعض مفردات) مجتمع الدراسة من خلال مواجهة الباحث شخصياً للمفردة الإحصائية أو عن طريق المراسلة ، كما في التعداد العام للسكان.

لغرض تصميم استمارة الاستبيان يتوجب على الباحث مراعاة ما يلي :

أ- إعداد مقدمة إيضاحية يتم من خلالها الإطلاع على هدف الاستمارة والبحث لغرض تسهيل الإجابة على أسئلة الاستمارة وبصورة صحيحة.

ب- أن تكون فقرات الاستمارة متسلسلة ومتكاملة ويكون كل جزء منها يحقق غرض معين . ويجب عند وضع الاسئلة مراعاة ما يلي:

1- أن تكون الأسئلة متوسطة العدد.

2- أن تكون الأسئلة واضحة المعنى ليس فيها غموض وقصور في جانب معين.

3- أن تكون الاجابة على الاسئلة بـ نعم أو لا أو إشارة وأن تكون الأجوبة أمام السؤال لكي يجيب عليها الفرد بسهولة.

4- أن يؤخذ بنظر الاعتبار عند وضع السؤال ظروف تفريغ وتصنيف وتبويب وترميز الأجوبة لتسهيل وسرعة إنجاز البحث.

8- الأخطاء الشائعة في جمع البيانات :

يحدث في بعض الأحيان أخطاء يقع بها الباحث عند جمعه للبيانات والمعلومات التي يتطلبها البحث وهذه الأخطاء تحدث نتيجة سوء استخدام الطريقة الإحصائية أهمها :

أ- **خطأ التحيز:** وهو الخطأ الذي يحدث عندما يأخذ الباحث المعلومات أو البيانات من مصادر ثانوية وليست المصادر الاصلية.

ب- **أخطاء الصدفة** : يحدث هذا الخطأ بسبب الباحث نفسه كأن يقوم الباحث باستيفاء بعض المعلومات والبيانات بالاعتماد على معلوماته الشخصية أو التعمد في جمع البيانات من بعض المفردات دون الأخرى المحددة ، أو جمع بيانات ناقصة لسبب أو آخر ، هذه الأخطاء وغيرها لها أثر في الحصول على نتائج غير دقيقة لتلك الدراسة.

9- النقاط التي يجب تحديدها قبل الشروع بعملية جمع البيانات هي :

1- تحديد الغرض من البحث.

2- تحديد مجال البحث.

3- تحديد المصدر.

4- تعيين الطريقة التي تتبع في الحصول على البيانات الإحصائية.

5- تحديد الوحدات القياسية المناسبة المستخدمة في عملية القياس.

6- تحديد الميزانية اللازمة للبحث.

7- تحديد الوقت

10- المصادر الإحصائية في جمع البيانات :

1- المصادر التاريخية : وهي البيانات والمعلومات المحفوظة والمتجمعة لدى أجهزة ومؤسسات دوائر الدولة المختلفة نتيجة الاستقصاءات أو المسوحات التي قامت بها هذه الجهات أو الهيئات مثال ذلك الجهاز المركزي للإحصاء مثال ذلك، التعداد العام للسكان في العراق.

2- المصادر الميدانية : وهي المصادر التي يقوم الباحث بجمعها والحصول عليها بنفسه من مصادرها الأصلية أما عن طريق المقابلة الشخصية أو المراسلة بالبريد أو محاكاة الحاسوب والانترنت.

أسلوب العينات

أ- تعريف المجتمع الاحصائي : هو عبارة عن جميع القيم أو المفردات التي يأخذها المتغير ، فمثلاً إذا كانت الصفات متعلقة بعمر الموظفين العاملين بإحدى الوزارات فأن المجتمع في هذه الحالة هو جميع الموظفين في تلك الوزارة والمجتمع أما أن يكون :

1- مجتمعاً محدداً : أي من الممكن حصر عدد مفرداته كما هو الحال في حصر عمر عدد من الموظفين في وزارة الصناعة مثلاً أو عدد الوحدات الإنتاجية من نوع معين في يوم معين.

2- المجتمعات غير المحددة : وهي المجتمعات التي يتعذر حصرها أو عدها مثل عدد البكتريا في ماده معينه.

ملاحظة : المجتمع ليس بالضرورة أن يكون مجتمعاً بشرياً.

ب- العينة : تعرّف العينة بأنها مجموعه معينة من مفردات مجتمع الدراسة ندرس صفاتها وخواصها وتعتبر ممثلة للمجتمع الأصلي وتعمم النتائج على المجتمع عند الحصول عليها ، حيث تكون ممثلة ومعبرة له ، وذلك أن دراسة المجتمع ككل قد يكون ضخماً يحتاج إلى وقت وجهد ومال ، لذا فقد تم الاستعانة عن دراسة المجتمع بدراسة العينة .

ان اسلوب العينات : هو عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات مجتمع معين للدراسة (العينة) مثل استفتاء رأي ، دراسة فاعلية علاج جديد معين لمرضى القلب، هذا الاسلوب يحتاج الى جهد وموارد مالية وبشرية اقل من التسجيل الشامل ، وهو ايضا مفيد عند دراسة المجتمعات غير المحددة الا ان دقته تعتمد على النتائج المستخلصة من البيانات.

تقسم العينات إلى قسمين

اولاً : العينات الإحصائية (العشوائية) : ويقصد بها تلك المجموعة من المفردات المختارة من مجتمع الدراسة ، بحيث أنه ليس للباحث أي دخل في اختيار هذه المفردة دون الاخرى اي أن هناك مبدأ تساوي الفرصة لظهور أية مفردة من مفردات المجتمع ضمن هذه العينة ، وتقسم العينات العشوائية (الاحتمالية) إلى ما يلي وحسب طريقة الاختيار :

1- **العينة العشوائية البسيطة :** وهي عملية اختيار عينة عشوائية من مجتمع الدراسة ، بحيث تملك كل مفردة من مفردات المجتمع نفس الفرصة والاحتمال في الظهور ضمن مفردات العينة ويتوجب عند اتباع هذا الأسلوب ملاحظة مسألة تجانس مفردات المجتمع من حيث الصفة والصفات ذات العلاقة بموضوع البحث ، ويتم اختيار العينة العشوائية البسيطة كما يلي :

أفرض أن مجتمع الدراسة متجانس ومحدود وأن عدد مفرداته N، وأن هناك دراسة تتطلب اختيار عينة من هذا المجتمع قوامها n ، بمعنى أن عدد العينات ذات الحجم n تمثل عدد مفردات العينة المختارة من مجتمع حجمه N ، وأن احتمال الظهور أي مفردة ضمن العينة هو $\frac{1}{N}$ وأن عدد العينات الممكنة الاختيار من هذا المجتمع هي r وتكون وفق القانون الاحتمالي التالي :

$$r=C_n{}^N = \frac{N!}{n1(\,N-n\,)1}$$

$$n! = n(\,n -1\,)\,(\,n – 2\,) \ldots 1$$

حيث أن n! تسمى المفكوك النوني او مضروب n
على سبيل المثال

$$4! = 4\,(\,4 – 1\,)\,(\,4 – 2\,)\,(\,4 – 3\,)+$$
$$= 4\,(3)\,(2)\,(1) = 24$$

وإن 1 = 1! و 1 = 0!

مثال : أفرض أنه تم اختيار ثلاثة منشآت من أربعة منشآت في قطاع وزارة الصناعة ، المطلوب : ما هو عدد العينات الممكنة الاختيار من هذا القطاع ؟ بين توزيعها بدون تكرار ؟

الحل : $N = 4\,,\, n = 3$

$$r=C_3{}^4 = \frac{4!}{3!(\,4-3\,)!} = \frac{4*3*2*1}{3*2*1*1} = 4$$

التوزيع : نفرض أن رموز هذه المنشآت هي (A , B , C , D) . اذا يكون توزيع هذه المنشاة حسب ما هو موضح:

عدد العينات (4)

التوزيع الأول→ (A , B , C)

التوزيع الثاني→ (A , B , D)

التوزيع الثالث→ (A , C , D)

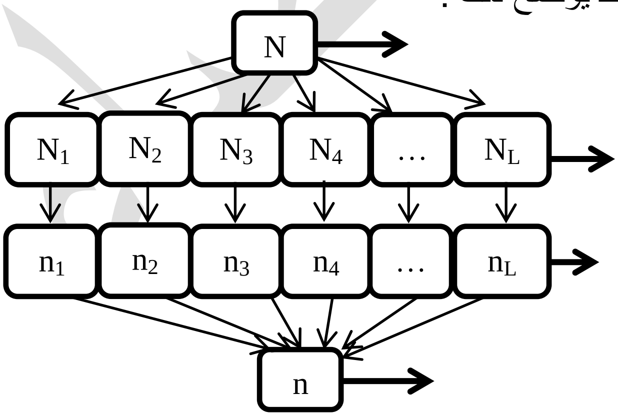
التوزيع الرابع→ (B , C , D)

2- العينة الطبقية العشوائية : تعتبر العينات المختارة وفق هذا الأسلوب أفضل أنواع العينات وأكثرها دقة في تمثيل المجتمع ، حيث أنه وفي أحوال كثيرة يلاحظ أن مفردات المجتمع الإحصائي غير متجانسة ، أي متكونه من طبقات تختلف فيما بينها مثال لو كنا بصدد دراسة المستوى العلمي لطلبة كلية الإدارة والاقتصاد (مجتمع احصائي فيه الطلبة هم المفردة الإحصائية) حيث ينقسم المجتمع الاحصائي الى طبقة قسم الإدارة وطبقه قسم المحاسبة وطبقه قسم الاقتصاد.

إن اختيار هذه العينة يتم على مرحلتين :

المرحلة الأولى : يتم في هذه المرحلة تقسيم المجتمع إلى طبقات متجانسة (مجتمعات جزئية).

المرحلة الثانية : يتم في هذه المرحلة أخذ عينة من كل طبقة بطريقة عشوائية ، أي أن كل طبقة تكون مجتمعاً بحد ذاته والشرط هو يجب ملاحظة التركيب النسبي ، أي يجب أن يتناسب حجم العينة مع حجم الطبقة في المجتمع النسبي ، وفيما يلي مخطط يوضح ذلك :



المجتمع الإحصائي

التقسيم إلى طبقات

تحديد حجوم العينات

حجم العينة المط

مخطط المعاينة العشوائية البسيطة

مثال 2 : يراد اختيار عينة عشوائية بحجم 20 مهني من مجتمع حجمه 100 مهني موزعين على النحو التالي 25 مهندس ، 35 موظف ، 40 عامل.

المطلوب : بيان عدد المهندسين والموظفين والعمال المكونين لهذه العينة.

الحل :

$$n = 20 \qquad\qquad N = 100$$

نستخرج كسر المعاينة f وهي نسبه العينة من التجمع

$$f = \frac{n}{N} = \frac{20}{100} = \frac{2}{10} = \frac{1}{5}$$

المهندسين→5 = $n1 = (25)\left(\frac{1}{5}\right) = \frac{25}{5} = 5$ العينة الأولى

الموظفين→ 7 = $n2 = (35)\left(\frac{1}{5}\right) = \frac{35}{5} = 7$ العينة الثانية

العمال→ 8 = $n3 = (40)\left(\frac{1}{5}\right) = \frac{40}{5} = 8$ العينة الثالثة

$$n = n_1 + n_2 + n_3$$

$$= 5 + 7 + 8 = 20$$

3- العينة العشوائية المنتظمة : وهي أقل استخداماً من النوعين الأولين وعند اختيارها يتبع أسلوب معين في الاختيار ، وفي هذه الحالة يكون على الأغلب المجتمع متجانس ، أفرض أن مفردات مجتمع الدراسة البالغ حجمها N مفردة مرتبة وفق ترتيب معين ، أما تصاعدي أو تنازلي أو وفق أي معيار آخر للترتيب (مثلاً درجات طلبة من أدنى درجة لأعلى درجة) ، وأفرض أن دراسة معينة تتطلب اختيار عينة من المفردات حجمها n مفردة ، عندئذ فأن ذلك يتم على النحو التالي :

1- يتم تقسيم مفردات المجتمع المرتبة أما تصاعدياً أو تنازلياً إلى عدد من المجاميع كل مجموعة منها تضم عدد من المفردات وفق القانون $k = \frac{N}{n}$.

2- بافتراض أن مفردات المجتمع متسلسلة في الترتيب فأن هذه المجاميع ستكون وفق السلسلة

1،2 , 3 … k   k+1 , k+2 ,k+3   (n-1)k.... ،nk

3- من تقسيم السلسلة في الخطوة (2) ولغرض تحديد مفردات العينة المطلوبة للدراسة فأنه يتم وبشكل عشوائي اختيار مفردة واحدة من المجموعة الأولى.

4- بعد معرفة تسلسل هذه المفردة يتم تلقائياً تحديد بقية المفردات الأخرى من خلال إضافة العدد k إلى تسلسل المفردة الأولى لنحصل على تسلسل المفردة الثانية ويضاف العدد k لتسلسل المفردة الثانية لتحصل على تسلسل المفردة الثالثة وهكذا.

وبذلك نضمن أن تكون كل مجموعة قد ساهمت بمفردة واحدة من إجمالي حجم العينة.

مثال 3 : في امتحان الطلبة صف معين (مجتمع إحصائي) عددهم 24 طالب رتبت أسماؤهم حسب تسلسل درجاتهم تنازلياً وبهدف التعرف على أسباب انخفاض مستواهم في الامتحان تطلب استقراء رأي ستة طلاب منهم أي حجم العينة = 6 ، المطلوب تحديد تسلسل هؤلاء الطلاب وبشكل عشوائي.

الحل : واضح في هذا المثال أن الطالب الأول في القائمة يحمل أعلى درجة والطالب الأخير فيها يحمل أوطأ درجة وعليه فأن هناك ستة مجاميع كل منها تضم أربعة طلاب.

1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24
المجموعة الأولى	المجموعة الثانية	المجموعة الثالثة	المجموعة الرابعة	المجموعة الخامسة	المجموعة السادسة

$$K = \frac{N}{n} = \frac{24}{6} = 4$$

الآن وبشكل عشوائي تم اختيار مفردة واحدة من المجموعة الأولى ولنفرض أنه الطالب الذي يحمل التسلسل 2 على ضوء ذلك تتحدد بقية تسلسلات مفردات العينة من خلال إضافة العدد $k = 4$ وفقاً ما يلي :

اختيار عشوائي المفردة الأولى =2، 6 = 4 +2 ، المفردة الثانية من المجموعة الثانية

المفردة الثالثة 10=4+6

14 = 4 + 10 المفردة الرابعة ، 18 = 4 + 14 المفردة الخامسة ، 22 = 4 + 18 المفردة السادسة

وبذلك فأن العينة المختارة في هذا الصف تمثل الطلبة الذي تسلسلاتهم 22 , 18 , 14 , 10 , 6 , 2

حجم العينة 6

ومثال ذلك عدد الموظفين والعمال في إحدى الوزارات.

4- **العينة أو المعاينة متعددة المراحل**: يتم اختيار هذا النوع من العينة عن طريق تقسيم المجتمع إلى وحدات تدعى بالوحدات الأولية يتم اختيار عينة عشوائية من هذه الوحدات الأولية كمرحلة أولى ثم تقسم كل وحدة أولية إلى وحدات أصغر تدعى وحدات ثانوية وهكذا يتم اختيار المفردات المطلوبة في هذا النوع.

ثانياً : العينة غير العشوائية : وهي تلك المجموعة من المفردات المختارة من المجتمع التي يكون للباحث دور أو دخل في اختيارها أي (يتدخل الباحث باختيارها) ، وذلك لاعتبارات تتعلق بطبيعة تلك الدراسة (وليس بطريقة عشوائية).

وتقسم هذه العينات إلى :

1- **العينة الحصصية** : يتم حسب هذا النوع تقسيم مجتمع الدراسة إلى عدة طبقات استناداً إلى معايير معينة تتعلق بطبيعة الدراسة ، فمثلاً استطلاع رأي على برنامج تلفزيوني معين يتم تقسيم المجتمع إلى ذكور وإناث ثم اختيار عينة من الذكور وعينة من الإناث.

2- **العينة العمدية :** وهو أسلوب اختيار عينة من المجتمع بشكل متعمد أي تعتمد مسبقاً على أن المفردات المختارة من هذه العينة هي خير من يمثل مجتمع الدراسة ، فمثلاً عند دراسة واقع التعليم الجامعي في العراق يجب أن تكون العينة المختارة هي أساتذة الجامعة وطلبة الجامعة.

عرض البيانات الإحصائية : عند جمع البيانات الأولية الخاصة بظاهرة ما فأنه عادة لا يمكن الاستفادة منها وهي بهذه الصورة لذلك غالباً ما توضع في جداول مبسطة أو يعبر عنها في صور أشكال ورسوم بيانية لكي تسهل دراستها وتحليلها ولذلك هناك وسائل تعرف البيانات هي :

أ- وضع البيانات بشكل تقديري.

ب- تقديم البيانات في جداول.

ج- رسمها في صور وأشكال هندسية لتساعد على فهمها.

د- تمثيل البيانات وتلخيصها بمؤشرات ومقاييس إحصائية معروفة.

العرض الجدولي : سوف تخصص هذه الفقرة لدراسة أساليب عرض البيانات المصنفة في جداول خاصة تدعى بجداول التوزيعات التكرارية التي تتخذ أشكالاً متعددة حسب نوع المتغير العشوائي الذي صنفت على أساسه البيانات الأولية (الخام).

بعض التعاريف :

1- البيانات غير المبوبة : وهي البيانات الأولية أو الأصلية التي جمعت ولم تبوب.

2- البيانات المبوبة : وهي البيانات التي نظمت وبوبت ضمن جداول توزيع تكراري.

3- التوزيع التكراري: هو عبارة عن تلخيص وترتيب لبيانات المتغير العشوائي التي سبق ان جمعت وصنفت ، مقسمة إلى عدد من المجاميع كل منها تسمى بـ (الفئة class) ، هذه الفئات قد تكون مرتبة تصاعدياً أو تنازلياً حسب طبيعة البيانات ، ويسمى توزيع عدد قيم (x) حسب الفئات بـ (التوزيع التكراري) ، وقد تكون فئات التوزيع التكراري متساوية في الطول أو غير متساوية وذلك يعتمد على طبيعة الدراسة ومتطلباتها.

- الفئات : وهي المجاميع التي قسمت إليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير.

- حدود الفئات : لكل فئة حدان حد أعلى وحد أدنى مثلاً الفئة من خمسين إلى ستين ، الخمسين تمثل الحد الأدنى والستين تمثل الحد الأعلى.

- الحدود الحقيقية للفئات : لكل فئة عددان حقيقيان حد أعلى وحد أدنى.

- طول الفئة : وهو مقدار المدى بين حدي الفئة ، هذا ويستحسن أن تكون أطوال الفئات متساوية لتسهيل العمليات الحسابية وسنرمز لطول الفئة بالرمز .L.

- مركز الفئة : لكل فئة مركز وسنرمز له (xi)وهو عبارة عن منتصف المدى بين حدي الفئة ، فمثلاً الفئة من 50 إلى 60 مركزها = الحد الأعلى + الحد الأدنى تقسيم (2 ′

Xl x_i x_s

أكبر قيمة   أصغر قيمة

$$X_i = \frac{\mathbf{50+60}}{\mathbf{2}} = 55$$

- تكرار الفئة : يمثل تكرار الفئة جزء من مفردات العينة التي تتصف بكونها تقع من حيث القيمة العددية ما بين حدي الفئة ، بحيث أن مجموع هذه الأجزاء يشكل عدد مفردات العينة n فإذا رمزنا لتكرارات الفئات بالرمز $f_{na}f_1 , f_2 \ldots f_m$ فأن :

$$f_1 + f_2 + \ldots + f_m = n$$

الفصل الثاني

التوزيعات التكرارية وأساليب عرض البيانات

تصنيف وتبويب البيانات :

بعد جمع البيانات فأن هذه البيانات التي تم الحصول عليها بخصوص ظاهرة معينة بشكلها الاولي تسمى بيانات خام raw data أو البيانات الأولية elementary data أو البيانات غير المصنفة ، وتكون هذه البيانات في الغالب غير منظمة وبالتالي يتعذر الاعتماد عليها لأغراض التحليل الإحصائي للوصول إلى النتائج المطلوبة ، لذلك فأن أولى الخطوات بعد جمع البيانات هي :

أ- مراجعة البيانات : وهي العملية التي تأتي بعد عملية جمع البيانات حيث يتم مراجعة وتدقيق البيانات.

ب- تصنيف البيانات : تبدأ هذه العملية بتصنيف البيانات على أساس الظواهر التي جمعت عنها فقد يكون للتصنيف على أساس العمر ، الجنس ، الوزن ... الخ.

ج- تبويب البيانات : ويقصد بالتبويب عملية تفريغ البيانات المصنفة في جداول خاصة بحيث أن كل جزء من البيانات المصنفة عن الظاهرة المعينة يعود إلى مستوى معين لتلك الظاهرة.

طبيعة البيانات الإحصائية والرموز الإحصائية :

عند توفر بيانات حول ظاهرة ما فأننا نرمز للظاهرة بالرمز (Y) وكل مفردة أو مشاهدة منها نرمز لها بـ (y) فمثلاً عند دراسة عدد الشركات في القطاع الصناعي فأننا نرمز لصفة القطاع الصناعي بالرمز (Y) وأي شركة تابعة له بالرمز (y_i) وتسمى y_i المشاهدة أو المفردة ، هذا وأن قيمة (y_i) قد تختلف من منشأة إلى أخرى ولهذا نقول أن (Y) متغير.

المتغيرات العشوائية Random variable :

يعرف المتغير العشوائي بأنه دالة ذات قيمة حقيقية real – valued function معرفة على مجال العينة sample space وغالباً ما يرمز للمتغير العشوائي بأحد الاحرف الكبيرة مثل X , Y , Z ... الخ ، ولقيم المتغير عند تنفيذ التجربة بأحد الأحرف الصغيرة مثل x , y , z ... الخ.

<u>وتقسم المتغيرات إلى :</u>

1- المتغيرات النوعية (الوصفية): وهي المتغيرات التي لا يمكن قياسها بوسائل قياس مألوفة كالعدد أو التقييس أنما تشكل صفات لذلك المتغير مثل الحالة الإنتاجية لإحدى الشركات (رديئة ، متوسطة ، جيدة ، جيدة جداً) وغيرها من الأمثلة.

2- المتغيرات الكمية: وهي المتغيرات التي يمكن قياسها بوسائل قياس مثل الموظفين في وزارة معينة وهذا النوع من المتغيرات يكون على نوعين أهمها :

أ- المتغيرات المستمرة:

إذا كانت مجموع القيم الممكنة للمتغير (x) مجموعة غير قابلة للعد سواء كانت مجموعة محدودة أو غير محدودة عندئذ يقال أن (x) متغير عشوائي مستمر ، بمعنى أن المتغير المستمر هو المتغير الذي تأخذ لمشاهدة أو المفردة فيه أي قيمة رقمية في مدى وجيز ، فلو فرضنا أطوال طلبة كلية معينة تتراوح بين 130,5–180,5 فعندئذ يقول بأن $130,5 \leq x \leq 180,5$ أي أن المتغير (x) يمكن ان يأخذ أي قيمة بين 130,5 و 180,5 وهنالك أمثلة أخرى مثل الوزن وكمية المحصول والزمن لأنها ممكن قياسها بأجزاء صغيرة جداً.

ب- المتغيرات المتقطعة أو المنفصلة:

إذا كانت مجموعة القيم الممكنة للمتغير (x) مجموعة قابلة للعد سواء كانت مجموعة محددة أو غير محددة عندئذ يقال أن (x) متغير عشوائي متقطع.

الخطوات العامة في إنشاء جدول التوزيع التكراري :

لتكوين جدول التوزيع التكراري يجب إتباع ما يلي :

1- استخراج المدى :

المدى = أكبر قيمة – أقل قيمة + 1

$$T.R = x_L - x_S + 1$$

2- اختيار وتحديد عدد الفئات : تمثل عدد الفئات عدد المجاميع التي يتألف منها التوزيع التكراري ، وهناك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع أهمها :

أ- صيغة يول Yule وهي :

$$m = 2.5 \sqrt[4]{n}$$

حيث أن n تمثل عدد المفردات.

ب- صيغة سترجس وهي :

m = عدد الفئات

$$m = 1 + 3 . 322 \log_{10}(n)$$

وعند التطبيق يتم تقريب الناتج لأقرب عدد صحيح.

لكل من الطريقتين مزايا وعيوب ولكن سوف تستخدم في العمل الحسابي عدداً من الفئات لا يقل عن 5 ولا يزيد عن 15.

3- إيجاد طول الفئة :

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

$$L = \frac{\boldsymbol{T.R}}{\boldsymbol{m}}$$

ويستحسن أن تكون أطوال الفئات متساوية.

مثال : البيانات التالية تمثل أعمار 40 موظف في إحدى الشركات التابعة لوزارة التجارة ، المطلوب : تكوين جدول التوزيع التكراري لهذه البيانات :

19 , 31 , 22 , 22 , 27 , 28 , 28 , 28 , 26 , 25 , 26 , 25 , 26 , 26 , 24 , 22 , 25 ,
24 , 24 , 22 , 24 , 24 , 24 , 23 , 21, 22 , 22 , 29 , 29 , 24 , 24 , 19 , 20 ,24 ,
24 , 24 , 24 24 , 24 , 24

الحل :/ الخطوة الأولى للحل نرتب القيم تصاعديا لسهولة الحل :

19 , 19 , 20 , 21 , 22 , 22 , 22 , 22 , 22 , 22 , 23 , 24 , 24 ,24 ,24 ,24 ,24 ,24
,24 ,24 ,24 ,24 ,24 ,24 ,24 ,24 , 25 , 25 , 25 , 26 , 26 , 26 , 26 , 27 , 28 , 28 ,
28 , 29 , 29 , 31

2- المدى : $T.R = x_L - x_S + 1$

$= 31 – 19 + 1 = 13$

3- عدد الفئات : $m = 1 + 3.322 \log(n)$

$= 1 + 3.322 \log(40) = 1+3.322 (1.6) = 1 + 5.322 = 6.322 = 6$ بالتقريب

4- طول الفئة :

$$L = \frac{\boldsymbol{T.R}}{\boldsymbol{m}} = \frac{\mathbf{13}}{\mathbf{6}} = 2.16 = 2$$ بالتقريب

1- الحد الأدنى للفئة الأولى يمثل أصغر قيمة حسب المثال هنا = 19

2- يضاف إلى الحد الأدنى طول الفئة 21 = 2 + 19

3- ويستمر الجدول الحد الأعلى للفئة الأولى يكون الحد الأدنى للفئة الثانية يضاف إليه طول الفئة يكون الحد الأعلى.

f_i	التكرار	عدد الفئات
3	III	19 – 21
7	IIIIIII	21 – 23
16	IIIIIIIIIIIIIIII	23 – 25
7	IIIIIII	25 – 27
4	IIII	27 – 29
3	III	29 – 31
40		المجموع

<u>الطرق المختلفة في كتابة حدود الفئات :</u>

نفرض لدينا الجدول التالي :

الفئات	الفئات	الفئات
40 – 49	40 -	40 – 50
50 – 59	50 -	50 – 60
60 – 69	60 – 70	60 – 70

جدول توزيع تكراري متقطع — جدول توزيع مستمر — جدول توزيع تكراري متصل أو مستمر

جدول التوزيع التكراري حسب نوع المتغير :

1- جدول التوزيع التكراري في حالة المتغيرات العشوائية المستمرة :

في هذه الحالة تكتب حدود الفئات كما هو موضح في المثال التالي :

مثال 2 : البيانات التالية تمثل أوزان (كغم) عينة من طلبة إحدى الكليات قوامها (37) طالب ، المطلوب : كون جدول التوزيع التكراري لهذه البيانات :

100 , 65 , 70 , 83 , 94 , 55 , 46 , 47.8 , 62.3 , 77.2 , 61.3 , 62.9 , 66.5 , 66.9 , 68.3 , 70 , 73.2 , 73.8 , 82.4 , 58.2 , 59.9 , 54.4 , 58.7 , 52.6 , 51.8 , 74.1 , 80.2 , 75.1 , 78.3 , 58.5 , 62 , 66.3 , 65 , 67.1 , 88.2 , 89.1 , 80.1 , 76.3

الحل :

1- نجد المدى $T.R = x_L - x_s + 1$

$= 100 - 46 + 1 = 55$

2- نجد عدد الفئات $m = 2.5\sqrt[4]{\boldsymbol{n}}$

3- طول الفئه $= 2.5\sqrt[4]{\boldsymbol{38}} = 6.16 = 6$

$L = \frac{\boldsymbol{T.R}}{\boldsymbol{m}} = \frac{\boldsymbol{55}}{\boldsymbol{6}} = 9$

الفئات	التكرار
46 – 55	5
55 – 64	9
64 – 73	8
73 – 82	9
82 – 91	4
91 – 100	2

2- في حالة المتغيرات العشوائية المتقطعة :

في هذه الحالة تكتب حدود الفئات حسب التوزيع التكراري التالي ، بحيث تضمن أن كل قيمة من قيم البيانات تضمن في فئة واحدة من فئات التوزيع دون أي تكرار قد يحصل في هذه الفئات أو تلك ، بحيث أن طول الفئة L يكون مساوياً للفرق ما بين الحد الأعلى للفئة وحدها الأدنى مضاف للفرق العدد واحد.

مثال 3 : البيانات التالية تمثل عدد الموظفين الفنيين والإداريين والخدميين لـ 60 منشأة تابعة لوزارة الصناعة والتجارة.

المطلوب : تبويب هذه البيانات في جدول تفريغ تكراري :

60 , 76 , 80 , 120 , 132 , 82 , 90 , 65 , 68 , 73 , 150 , 142 , 157 , 164 , 88 , 90 , 98 , 101 , 103 , 110 , 119 , 116 , 120 , 126 , 109 , 114 , 120 , 122 , 111 , 116 , 90 , 78 , 93 , 95 , 98 , 104 , 120 , 113 , 121 , 119 , 125 , 126 , 130 , 131 , 136 , 118 , 142 , 148 , 150 , 154 , 122 , 123 , 139 , 125 , 156 , 154 , 136 , 137 , 110 , 136

1- T.R = 164 – 60 + 1 = 105 المدى

2- $m = 2.5 \sqrt[4]{\boldsymbol{n}} = 2.5 \sqrt[4]{\boldsymbol{60}} = 2.5\ (2.783)$ عدد الفئات

m = 6.958 = 7

3- $L = \frac{\boldsymbol{T.R}}{\boldsymbol{m}} = \frac{\boldsymbol{105}}{\boldsymbol{7}} = 15$ طول الفئة

ترتيب القيم تصاعدياً :

60 , 65 , 68 , 73 , 76 , 78 , 80 , 82 , 88 , 90 , 90 , 90 , 93 , 95 , 98 , 98 , 101 , 103 , 104 , 109 , 110 , 110 , 111 , 113 , 114 , 116 , 116 , 118 , 119 , 119 , 120 , 120 ,120 ,120 , 121 , 122 , 122 , 123 , 125 , 125 , 126 , 126 , 130 , 131 , 132 , 136 ,136,136, 137 , 139 , 142 , 142 ,148,150, 150 , 154 , 154 , 156 , 157 , 164

التكرار	التكرار	الفئات
4	IIII	60 – 74
5	IIIII	75 – 89
10	IIIIIIIIII	90 – 104
11	IIIIIIIIIII	105 – 119
15	IIIIIIIIIIIIIII	120 – 134
8	IIIIIIII	135 – 149
7	IIIIIII	150 – 164
60		

ملاحظات على تكوين الجداول التكرارية :

1- عند تحديد عدد الفئات يفضل أن يكون العدد مناسب وعلى هذا يجب أن توازن دائماً عند تحديد الفئات بين الدقة بالنتائج والسهولة بالعمل الحسابي.

2- يفضل جعل فئات الجداول متساوية في الطول.

3- عند كتابة الفئات في الغالب نبدأ بأقل مفردة لدينا كحد أدنى للفئة الأولى ويجوز في بعض الأحيان أن نبدأ بقيمة أقل عن قيمة أقل مفردة كحد أدنى للفئة الأولى ، وينتهي الجدول بقيمة أكبر مفردة كحد أعلى للفئة الأخيرة ويجوز أن ينتهي بقيمة أكبر من هذه القيمة ولا يجوز أن ينتهي بقيمة أقل من قيمة أكبر مفردة.

4- يفضل قدر الإمكان تجنب الجداول المفتوحة حيث يفضل عليها الجداول المغلقة.

5- يسمى الجدول مقفلاً إذا كان الحد الأعلى للفئة الأولى معروفاً وكذلك الحد الأعلى للفئة الأخيرة معروف أيضاً أما الجدول المفتوح فيكون على الحالات التالية :

أ- مفتوحاً من طرفين إذا كان الحد الأدنى للفئة الأولى غير معلوم وكذلك الحد الأعلى للفئة الأخيرة غير معلوم أيضاً.

ب- يكون مفتوحاً من طرفه الادنى فقط إذا كان الحد الأدنى للفئة الأولى غير معلوم فقط.

ج- يكون مفتوحاً من طرفه الاعلى فقط إذا كان الحد الاعلى للفئة الاخيره غير معلوم فقط.

الفئات	الفئات	الفئات	الفئات	الفئات
30 – 40	أقل من 40	أقل من 40	40 – 50	40 - 50
40 – 50	40 – 50	40 – 50	50 – 70	50 – 60
50 – 60	50 – 60	50 – 60	70 – 80	60 – 70
60 فأكثر	60 – 70	60 – 70	80 – 100	70 – 80
مفتوح في الطرف الاعلى فقط	مفتوح من الطرف الأدنى فقط	70 فأكثر مفتوح من الطرفين	جدول غير منتظم	جدول منتظم

التكرار النسي والمئوي

ان جدول التوزيع التكراري النسبي هو جدول يبين الأهمية النسبية لكل فئة حيث :

التكرار النسبي = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي التكرارات}}$

التكرار النسبي = $\frac{fi}{\sum_{i=1}^{m} fi}$

اما التكرار المئوي فانه يبين الاهمية المئوية لكل فئة :

التكرار المئوي = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي التكرارات}} \times 100$

التكرار المئوي = $\frac{fi}{\sum_{i=1}^{m} fi} * \mathbf{100}\%$

وعادة يصبح التكرار النسبي تكرار مئوي وذلك بضربه في 100%

$\frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}} \times 100$ = التكرارات المئوي

مثال 5 :

للجدول التالي جد التكرار النسبي والمئوي لكل فئة :

الفئات	Fi	التكرار النسبي	التكرار المئوي
31 – 40	1	$\frac{1}{80}$ = 0.0125	$\frac{1}{80}$ = 0.0125 * 100 = 1.25
41 – 50	2	$\frac{2}{80}$ = 0.0250	$\frac{2}{80}$ = 0.025 * **100** = 2.5
51 – 60	5	$\frac{5}{80}$ = 0.0625	0.06 25* **100**= 6.25
61 – 70	15	0.187	18.7
71 – 80	25	0.3125	31.25
81 – 90	20	0.25	25
91 – 100	12	0.15	15
المجموع	80	1	100

التكرار النسبي = $\frac{fi}{\sum_{i=1}^{m} fi}$

ملاحظة الرمز $\sum$ يعني المجموع الكلي.

على سبيل المثال $\frac{1}{80} = \frac{fi}{\sum_{i=1}^{m} fi}$

التكرار المئوي = $\frac{\text{تكرار كل فئة}}{\text{المجموع الكلي التكرارات}}$ × 100%

= $\frac{1}{80}$ * 100% = = 0.0125* 100% = 1.25%

<u>التوزيع التكراري المتجمع :</u> إن جدول التوزيع التكراري العادي الذي سبق شرحه يبين توزيع قيم المتغير على الفئات المختلفة ولكن في بعض الأحيان قد تكون هناك حاجة إلى معرفة عدد القيم أو المفردات التي تقل أو تزيد من قيمة معينة والجداول التي تحتوي على مثل هذه المعلومات تدعى بالجداول التكرارية المتجمعة ، وهناك نوعان من هذه الجداول :

1- جدول التوزيع التكراري المتجمع الصاعد : ولتكوين هذا الجدول نبدأ بتجميع التكرارات من الفئة الأولى إلى أن نصل إلى الفئة العليا ، وسنرمز للتكرار المتجمع لأي فئة بالرمز (Fi)، حيث أن الرمز (fi) يكون للتكرار العادي ، وفيما يلي توضيح لذلك :

التكرار المتجمع الصاعد للفئة الأولى $F_1 = f_1$

التكرار المتجمع الصاعد للفئة الثانية $F_2 = f_1 + f_2 = F_1 + f_2$

التكرار المتجمع الصاعد للفئة الثالثة $F_3 = f_1 + f_2 + f_3 = F_2 + f_3$

التكرار المتجمع الصاعد للفئة الأخيرة = $F_n = f_1 + f_2 + \dots f_n = F_{n-1} + f_n$

2<u>- التكرار المتجمع النازل :</u> هي هذا النوع من التكرار نضع التكرارات المتجمعة مبتدئين من الفئة العليا إلى أن نصل إلى الفئة الدنيا وكما موضح أدناه :

التكرار المتجمع النازل للفئة الأولى $F_1 = F_n = \sum \boldsymbol{fi}$

التكرار المتجمع النازل للفئة الثانية $F_2 = F_n - f_1 = F_1 - f_1$

التكرار المتجمع النازل للفئة الثالثة $F_3 = F_n - f_1 - f_2 = F_2 - f_2$

التكرار المتجمع النازل للفئة الأخيرة $F_n = F_1 - F_2 \dots\dots F_{n-1} = F_1 - \sum_{i=1}^{n} f_i$

ملاحظة :

لإيجاد التكرار المتجمع الصاعد أو النازل النسبي أو المئوي نتبع ما يلي :

$$\text{التكرار المتجمع الصاعد أو النازل النسبي} = \frac{\text{التكرار المتجمع الصاعد أو النازل للفئة}}{\text{المجموع الكلي التكرارات}}$$

أما التكرار المئوي المتجمع الصاعد أو النازل فهو عبارة عن التكرار النسبي الصاعد أو النازل مضروب × 100%

مثال 6 : البيانات التالية تمثل توزيع 80 عائلة حسب ملكيتها من الدخل الشهري.

المطلوب :

1- أوجد التكرار المتجمع الصاعد والنازل.

2- إيجاد عدد العوائل التي دخلها الشهري تقل عن 80.

3- إيجاد عدد العوائل التي دخلها الشهري يزيد عن 60.

4- التكرار النسبي والمئوي المتجمع الصاعد والنازل.

F_i	التكرار المتجمع الصاعد / الحدود العليا للفئات	عدد العوائل f_i	الفئات
1	40 وأقل من 40	1	30 – 40
3	50 وأقل من 50	2	40 – 50
8	60 وأقل من 60	5	50 – 60
23	70 وأقل من 70	15	60 – 70
48	80 وأقل من 80	25	70 – 80
68	90 وأقل من 90	20	80 – 90
80	100 وأقل من 100	12	90 – 100

$F_1 = f_1 = 1$

$F_2 = F_1 + f_2 = f_1 + f_2 = 1 + 2 = 3$

$F_3 = F_2 + f_3 = f_1 + f_2 + f_3 = 1 + 2 + 5 = 8$

$F_4 = F_3 + f_4 = f_1 + f_2 + f_3 + f_4 = 1 + 2 + 5 + 8 = 23$

$F_5 = F_4 + f_5 = 23 + 25 = 48$

$F_5 = f_1 + f_2 + f_3 + f_4 + f_5 = 1 + 2 + 5 + 15 + 25 = 48$

$F_6 = F_5 + f_6 = 48 + 20 = 68$

$F_6 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 1 + 2 + 5 + 15 + 25 + 20 = 68$

$F_7 = F_6 + f_7 = 68 + 12 = 80$

$F_7 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 = 1 + 2 + 5 + 15 + 25 + 20 + 12 = 80$

الفئات	عدد العوائل f_i	التكرار المتجمع النازل / الحدود الدنيا للفئات	F_i
30 – 40	1	30 فأكثر	80
40 – 50	2	40 فأكثر	79
50 – 60	5	50 فأكثر	77
60 – 70	15	60 فأكثر	72
70 – 80	25	70 فأكثر	57
80 – 90	20	80 فأكثر	32
90 – 100	12	90 فأكثر	12
		100 فأكثر	0

$F_1 = f_n = \frac{fi}{\sum_{i=1}^{m} fi} = 80$

$F_2 = F_n - f_1 = 80 - 1 = 79$

$F_3 = F_n - f_1 - f_2 = F_2 - f_2 = 80 - 1 - 2 = 77$

$F_4 = F_n - f_1 - f_2 - f_3 = 80 - 1 - 2 - 5 = 72$

$F_5 = F_n - f_1 - f_2 - f_3 - f_4 = 80 - 1 - 2 - 5 - 15 = 57$

$F_6 = f_n - f_1 - f_2 - f_3 - f_4 - f_5 = 80 - 1 - 2 - 5 - 15 - 25 = 32$

$F_7 = f_n – f_1 - f_2 - f_3 - f_4 - f_5 - f_6 - f_7 = 80 – 1 – 2 – 5 – 15 – 25 – 20 – 12 = 0$

2- عدد العوائل التي دخلها الشهري يقل عن 80وأقل هو 48.

3- عدد العوائل التي دخلها الشهري يزيد عن 60فأكثر هو 72.

4- التكرار النسبي والمئوي :

التكرار المتجمع النازل المئوي F_i	التكرار المتجمع النازل النسبي F_i	التكرار المتجمع الصاعد المئوي F_i	التكرار المتجمع الصاعد النسبي F_i
(1 * 100%) =100	80 \ 80 = 1	(0.125 *100%) = 1.25	1\80 = 0.125
(0.9875 * 100%) = 98.75	79 \ 80 = 0.9875	3.75	3\80 = 0.0375
96.25	77 \ 80 = 0.9625	10	8\80 = 0.1
90	0.9	28.75	0.2875
71.25	0.7195	60	0.6
40	0.4	85	0.85
15	0.15	100	1

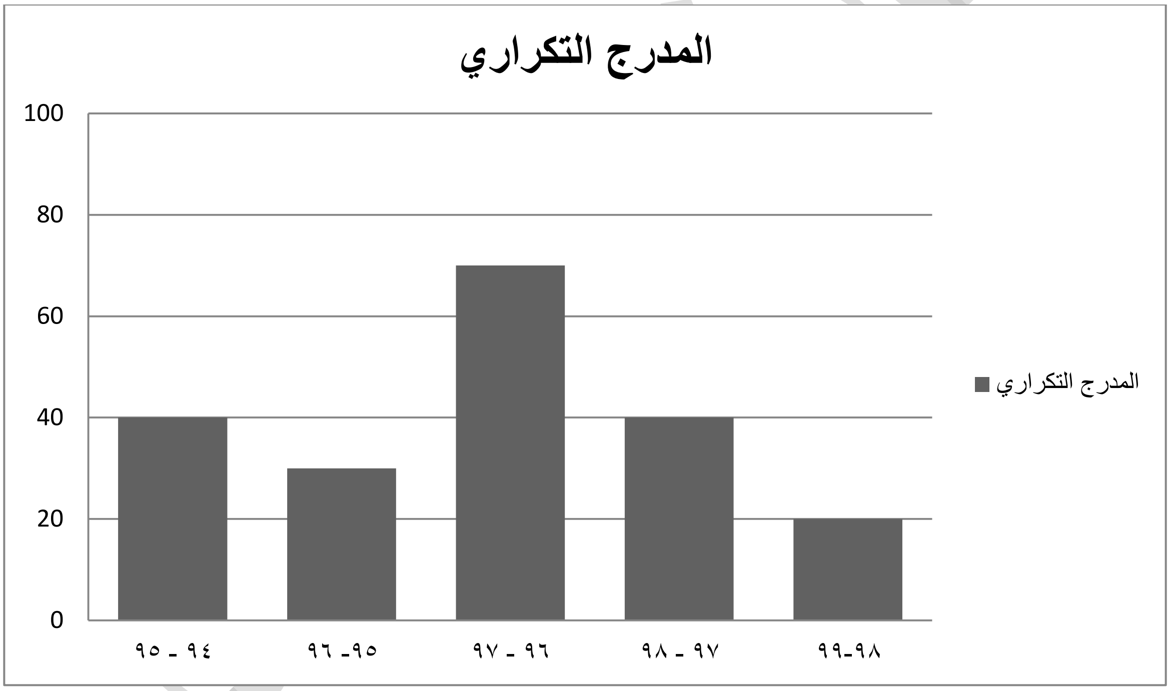
<u>العرض الهندسي للبيانات :</u> بغية إعطاء فكرة واضحة وسريعة عن البيانات المبوبة في جداول التوزيعات التكرارية فأنه يتم عرض هذه البيانات بهيئة رسوم بيانية وأشكال هندسية متعددة الأشكال والتصاميم والبعض منها بهيئة رسوم تصورية ، وأن هذه الرسوم والصور والأشكال الهندسية ما هي إلا تعيين وتوضيح للبيانات بطريقة سهلة وفعالة تساعد القارئ على فهم واستيعاب قيم الظاهرة ومقارنتها مع بعضها ، هناك عدة طرق تستخدم لتمثيل التوزيعات التكرارية بالرسم منها :

1- المدرج التكراري : وهو عبارة عن مستطيلات رأسية تمتد قواعدها على المحور الأفقي لتمثيل أطوال الفئات بينما ارتفاعها تمثل تكرارات الفئات ، وهذه المستطيلات تكون منفصلة عن بعضها في حالة المتغيرات المتقطعة وتكون متصلة مع بعضها في حالة المتغيرات المستمرة وحسب تسلسل فئات التوزيع.

مثال 7 :بلغت عدد الدورات للموظفين التي نفذت من قبل مركز التدريب والتعليم المستمر في جامعة بغداد للسنوات في عام 1994ولغاية عام 1999كما يلي :

الفئات (السنوات)	عدد الدورات fi
94 – 95	40
95 – 96	30
96 – 97	70
97 – 98	40
98 - 99	20
	200

المطلوب : أرسم المدرج التكراري.



2- المضلع التكراري :

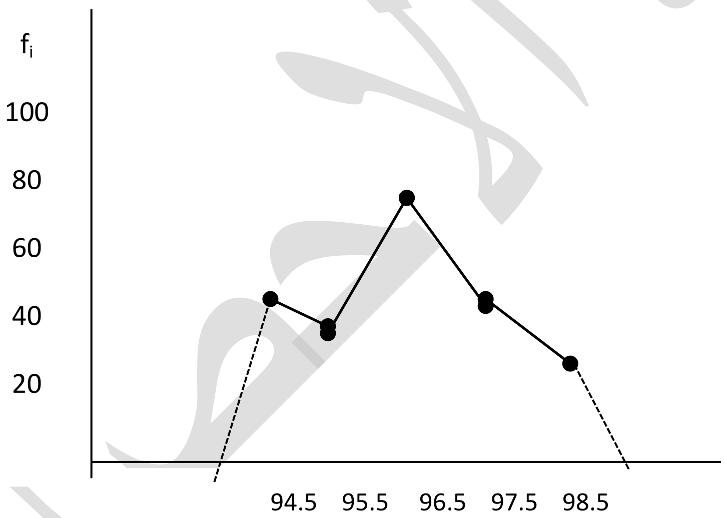
وهو عبارة عن خطوط مستقيمة ومتكسرة تصل بين نقاط كل منها. واقعة فوق مركز فئة على ارتفاع يمثل تكرار تلك الفئة وعادة يفضل في المضلع التكراري بأن نصل بداية المضلع بالمحور الافقي بمركز فئة (خيالية) واقعة إلى يسار أول فئة تكرارها صفر وتصل نهاية المضلع الأفقي بمركز فئة خيالية واقعة إلى يمين آخر فئة تكرارها أيضاً صفر هذا ويمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تصنيف القواعد العليا للمستطيلات والتي تمثل مراكز الفئات بنقاط ثم نوصل هذه النقاط بمستقيمات.

مثال 8 / لنفس البيانات في المثال رقم (7) ارسم المضلع التكراري.

مركز الفئة الأولى = $\frac{\text{الحد الأعلى+الحد الأدنى}}{2}$

$= \frac{95+94}{2} = \frac{189}{2} = 94.5$

مركز الفئة x_i	f_i	الفئات
94.5	40	94 – 95
95.5	30	95 – 96
96.5	70	96 – 97
97.5	40	97 – 98
98.5	20	98 – 99



مركز الفئة

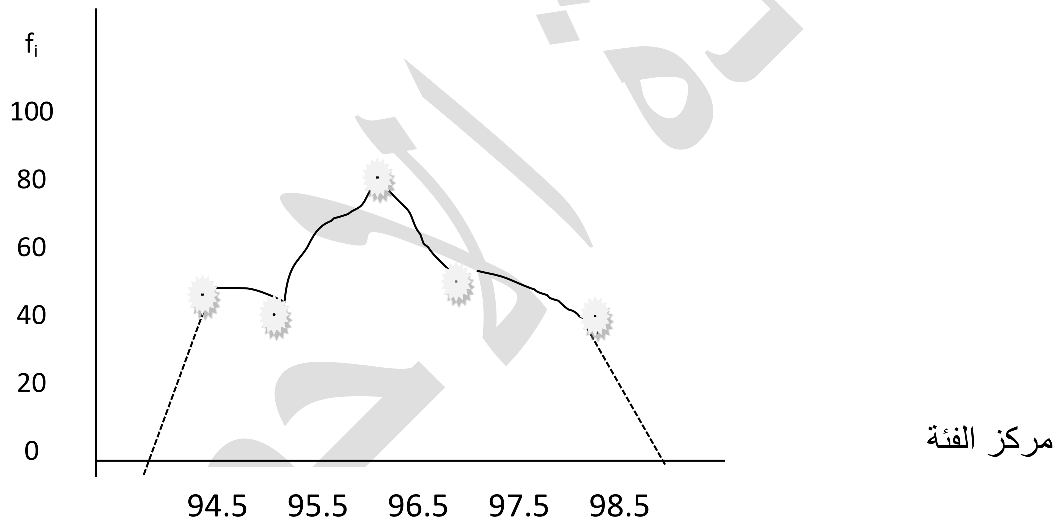
ملاحظة : في حالة رسم المضلع يتم توصيل الإحداثي بمركز الفئة الخيالي مثل 35و 99كما في المثال أعلاه لغلق المضلع وتكرار هذه المراكز يكون مساوي إلى الصفر.

3- المنحنى التكراري : لا تختلف فكرة رسم المنحنى التكراري في المضلع التكراري من حيث الأسلوب لكن الفرق الوحيد ما بينهما هو أنه بدلاً من توصيل النقاط (مركز الفئة) والتكرار بمستقيمات فأنه يتم تمرير منحني ما بين هذه النقاط هذا المنحني يمثل المنحنى التكراري لذلك التوزيع وعادة يرسم المنحنى التكراري بأن نصل بدايته بالحد الأدنى للفئة الأولى ونهايته بالحد الأعلى للفئة الأخيرة وبالإمكان رسم المنحى التكراري من المضلع التكراري.

مثال 9 :

لنفس البيانات السابقة في المثال رقم (8) أرسم المنحنى التكراري.

الفئات	f_i	مركز الفئة x_i
94 – 95	40	94.5
95 – 96	30	95.5
96 – 97	70	96.5
97 – 98	40	97.5
98 – 99	30	98.5



ملاحظة : بالإمكان رسم المضلع التكراري والمنحنى المدرج التكراري سوية وذلك بأن تكون رؤوس المنحني والمضلع في نصف ارتفاع المستطيل.

4 - المستطيل البياني :

وهو عبارة عن شكل هندسي يستخدم في تمثيل بيانات ظاهرة معينة بكل أجزاءها الى عدد من الأصناف القابلة للتجميع ، مثل عدد الطلبة الموزعين حسب المراحل الدراسية ، ان فكرة هذا الشكل بسيطة تتلخص باختيار مستطيل ذو قاعدة مناسبة بقياس مناسب ، ويمثل هذا المستطيل الكبير مجموع البيانات الكلية عن الظاهرة (عدد الطلبة) مثلا ، وبعد ذلك يتم تمثيل كل صنف من الاصناف بمستطيل جزئي داخل المستطيل الكبير بحيث ان مجموع المستطيلات الجزئية تمثل مساحة المستطيل الكبير ، ويتم اختيار قواعد المستطيلات الجزئية وفق القانون التالي :

طول قاعدة المستطيل الجزئي = عدد بيانات الصنف / مجموع البيانات الكلية * طول قاعدة المستطيل الكبير

مثال / بلغ عدد الطلبة في احدى الكليات 2000 طالب وطالبة منهم 800 في المرحلة الاولى ، 500 في المرحلة الثانية ، 400 في المرحلة الثالثة ، 300 في المرحلة الرابعة ، يطلب تمثيل هذه البيانات بمستطيل بياني .

الحل / نختار مستطيل ذو قاعده مساويه الى 10 سم وعندئذ فأن طول قاعدة كل مستطيل جزئي لكل مرحلة هو

للمرحلة الاولى ________________ $\frac{800}{2000}*10 = 4cm$

للمرحلة الثانية ________________ $\frac{500}{2000}*10 = 2.5cm$

للمرحلة الثالثة ________________ $\frac{400}{2000}*10 = 3cm$

للمرحلة الرابعة ________________ $\frac{300}{2000}*10 = 1.5cm$

المرحلة الرابعة	**المرحلة الثالثة**	**المرحلة الثانية**	**المرحلة الاولى**

- الدائرة البيانية :

وهذه عباره عن شكل هندسي مثلها مثل المستطيل البياني حيث انه هنا وبدلا من تمثيل الاصناف بمستطيلات جزئية فأنه يتم تمثيلها بقطاعات داخل دائرة بحيث ان مجموع مساحات القطاعات تمثل مساحة الدائرة وبهدف تحديد كل قطاع فأنه يتوجب تحديد زاوية كل منها ان ذلك يتم وفق مايلي :

زاوية القطاع = عدد بيانات اصنف /مجموع البيانات الكلية * 360°

مثال / بلغ عدد المهمات القتالية التي كلفت بها طائرات احدى القواعد الجوية خلال اسبوع معين 1600 مهمة قتالية موزعة على النحو التالي :

يوم السبت 250 ، الاحد 370 ، الاثنين 400 ، الثلاثاء 280 ، الاربعاء 150 ، الخميس 100 ، الجمعة 50 ،

يطلب تمثيل هذه البيانات بدائرة بيانية ؟

الحل / نختار دائرة ذات نصف قطر (5 سم) ونبدأ بتحديد زاوية كل قطاع الذي يمثل يوم معين وعلى النحو التالي :

يوم السبت $\frac{250}{1600} * 360^o = 56.25^o$

يوم الاحد $\frac{370}{1600} * 360^o = 83.25^o$

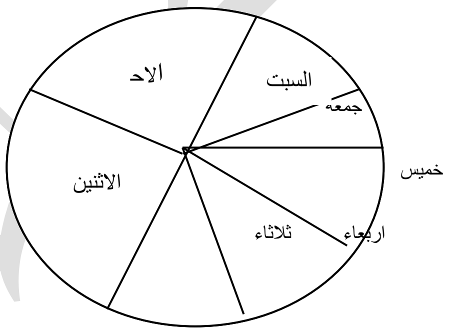
يوم الاثنين $\frac{400}{1600} * 360^o = 90^o$

يوم الثلاثاء $\frac{280}{1600} * 360^o = 63^o$

يوم الاربعاء $\frac{150}{1600} * 360^o = 33.75^o$

يوم الخميس $\frac{100}{1600} * 360^o = 22.5^o$

يوم الجمعة $\frac{50}{1600} * 360^o = 11.25^o$



التوزيع التكراري المزدوج :

يلاحظ في أحوال كثيرة وجود علاقة ما بين متغيرين (ظاهرتين) أو أكثر مثل العلاقة بين وزن الشخص وطوله ، سعر سلعة معينة والكمية المطلوبة منها ، الدخل الشهري والانفاق الشهري على السلع الغذائية وغيرها من الأمثلة الأخرى وهذا يعني أنه على أساس عينة عشوائية من المفردات وقوامها (n) مفرده سوف نحصل على زوج من القيم عن كل مفردة ، فإذا فرضنا أن xيمثل المتغير الأول وy المتغير الثاني فأن البيانات التي سيتم الحصول عليها من هذين المتغيرين هي أزواج القيم التالية :

$$(x , y) (x , x) (x_1 , y_2) (x_n , y_n)$$

وبغية تبويب هذه البيانات في جدول موحد من أجل إعطاء ملخص واضح عنها فأن ذلك يتم من خلال تكوين فئات توزيع كل متغير بشكل منفصل ومن ثم دمج التوزيعين في جدول واحد (ذات اتجاهين) صفوفه تمثل أحد المتغيرين واعمدته تمثل فئات المتغير الآخر ، بعد اكمال هذه العملية نبدأ بتفريغ البيانات في خلايا هذا الجدول بحيث نضمن أن كل زوج من أزواج القيم يؤشر في خلية واحدة فقط من خلايا الجدول التي تمثل منطقة التقاء فئة صفية مع فئة عمودية ، وعندئذ فأن مجموعي التأشيرات في الخلية الواحدة سوف يمثل تكرار تلك الفئة الخلية الذي يمثل عدد المفردات التي تتصف بكونها تنتمي لتلك الفئة الصفية وفي نفس الوقت لتلك الفئة العمودية ، والجدول التالي يمثل نموذج لجدول توزيع تكراري مزدوج لافتراض أن عدد فئات المتغير xهوm وأن عدد فئات المتغير yهو k.

فئات *x* y	1 2 … m	المجموع
1	$f_{11}f_{12} \dots f_{1m}$	f.1
2	$f_{21}f_{22} \dots f_{2m}$	f.2
:	: : :	:
K	$f_{k1}f_{k2} \dots f_{km}$	f.k
	$f_{1.}$ $F_{2.}$ … $f_{m.}$	$f_{rm} \rightarrow n$

فمثلاً f_{21}يمثل تكرار الخلية الناتجة من التقاء الفئة الثانية من فئات (x) مع الفئة الأولى من فئات (y) كذلك فأن المجاميع المقابلة لفئات (x) أسفل الجدول تمثل تكرارات مخططات (x) فقط ، في حين أن المجاميع المقابلة الفئات (y) تمثل تكراراتها فقط ، وإن fيمثل مجموع التكرارات الكلية ، أي عدد مفردات العينة ، وهذا يعني أن :

$$f.. = f_1 + f_2 + \dots f_m = f_1 + f_2 + \dots f_k = n$$

مثال :

في دراسة لبيان العلاقة ما بين وزن الشخص (كغم) وطوله (سم) ، اختيرت عينة عشوائية قوامها (12) شخص وتم قياس وزن كل منهم (x) وطوله (y) وكانت نتائج القياس ما يلي : المطلوب : تفريغ هذه البيانات في جدول توزيع تكراري مزدوج.

X	*50*	*53*	*55*	*62*	*67*	*70*	*71*	*77*	*80*	*82*	*90*	*90*
Y	*163*	*168*	*166*	*164*	*166*	*172*	*174*	*180*	*176*	*185*	*174*	*188*

1- المدى الكلي للمتغير x هو :

$$T.R_x = x_L - x_s + 1 = 90 - 50 + 1 = 41$$

المدى الكلي إلى y هو :

$$T.R_y = x_L - x_s + 1 = 188 - 163 + 1 = 26$$

2- عدد الفئات لكل متغير هي :

$$m = k = 1 + 3.322 \log (12) = 1 + 3.322\ (1.08) = 1 + 3.58 = 4.58=5$$

3- طول الفئة x_L

$$L_x = \frac{T.Rx}{m} = \frac{41}{5} = 8.2 = 8$$

طول الفئة y_L

$$L_y = \frac{T.Rx}{k} = \frac{26}{5} = 5.2 = 5$$

المجموع	90 - 82	82 - 74	74 - 66	66 - 58	58 – 50	فئات Y/ فئات x
4	-	-	I	I	II	168 – 163
2	-	-	I	-	I	173 – 168
3	I	I	I	-	-	178 – 173
1	-	I	-	-	-	183 – 178
2	II	-	-	-	-	188 – 183
12	3	2	3	1	3	المجموع

X	*50*	*53*	*55*	*62*	*67*	*70*	*71*	*77*	*80*	*82*	*90*	*90*
Y	*163*	*168*	*166*	*164*	*166*	*172*	*174*	*180*	*176*	*185*	*174*	*188*

تمارين الفصل الاول والثاني

س1 / البيانات التالية تمثل التوزيع التكراري لـ 60 عائلة حسب دخلها الشهري : المطلوب 1- أوجد التكرار المتجمع الصاعد ، 2- إيجاد عدد العوائل التي دخلها الشهري يقل عن 120 3- إيجاد عدد العوائل التي دخلها الشهري يزيد عن 104 4- التكرار النسبي والمئوي ، 5- التكرار النسبي الصاعد.

الفئات : 164 - 50 ، 135 – 149 ، 120 – 134 ، 105 – 119 ، 90 – 104 ، 75 – 89 ، 60 – 74

عدد العوائل / التكرار f_i : 7 7 16 11 10 5 4

س2 / البيانات التالية تمثل عدد الشركات التابعة لوزارة الصناعة.

المطلوب : كون جدول توزيع تكراري لهذه البيانات :

البيانات 3 , 4 , 7 , 8 , 10 , 9 , 2 , 5 , 6 , 9 , 7 , 5

س3 / إذا كان لديك التفريغ التكراري :

المطلوب : 1- مراكز الفئات. 2- التكرار المتجمع النازل. 3- التكرار النسبي النازل. 4- التكرار المئوي النازل.

الفئات : 51 – 4748 – 4344 – 3940 – 3536 – 32

10 6 8 4 8

س4 / مصنع لإنتاج المصابيح الكهربائية مؤلف من 10 مكائن متماثلة.

المطلوب : اختيار عينة عشوائية مؤلفة من 2 من المكائن لغرض تقييم كفاءة الأداء في المصنع ، ما عدد العينات الممكنة الاختيار من هذا المصنع ؟

س5 / أفرض أن لدينا مجتمع حجمه N = 5 بالقياسات 5 , 4 , 3 , 2 , 1.

س6 / البيانات التالية تمثل عدد أفراد عينة من الأسر قوامها 170 أسرة , 4 , 13 , 9 , 6 , 4 , 3 , 2 , 9 , 18 , 15 , 14 , 16 , 14 , 12 , 13 , 11 , 6 , 22 , 8 , 7 , 6 , 5 , 9 , 2 , 2 , 6 , 6 9 , 12 , 15 , 17 , 18, 9 , 9 8 , 9 , 13 , 12 , 14 , 15 , 11 , 8 , 10 , 9 , 15 , 12 , 10

, 8 , 15 , 13 , 10 , 18 , 17 , 12 , 11 , 9 , 9 , 8 10 , 8 15 , 14 , 10 8 , 12 , 11 ,
10 , 20.

المطلوب : 1- كون جدول التوزيع التكراري. 2- أحسب التوزيع التكراري المتجمع الصاعد والنازل وكذلك التوزيع التكراري المتجمع الصاعد والنازل النسبي.

3- إيجاد عدد الأسر التي عدد أفرادها يزيد عن 8 أفراد. 4- عدد الأسر التي عدد أفرادها يقل عن 13 فرد. 5- عدد الأسر التي يتراوح عدد أفرادها بين 10 و 17 فرد. 6- إيجاد نسبة عدد الأسر التي عدد أفرادها يزيد عن 7 أفراد. 7- نسبة عدد الأسر التي يتراوح عدد أفرادها بين 5 إلى 16.

س7 / البيانات التالية تمثل التوزيع التكراري لعدد من المصانع الصغيرة وكلفها التقديرية.

المطلوب : 1- أرسم المدرج التكراري. 2- المضلع التكراري. 3- المنحني التكراري.

عدد الفئات :	2 - 4	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14	14 - 16
عدد المصانع f_i :	50	20	15	11	9	6	4

الفصل الثالث

مقاييس النزعة المركزية (مقاييس التوسط)

مقدمة :

إن معظم القيم لمختلف الظواهر تتحرك عادة في الوسط أو بالقرب منه (ومقاييس) التوسط لأي مجموعة من البيانات التابعة لظاهرة ما ، هي تلك المقاييس التي تبحث في تقوية قيمة تتمركز حولها أغلبية هذه البيانات وأن هذه القيمة المتوسطة أو المتمركزة هي عبارة عن رقم واحد يعبر عن أو يمثل جميع بيانات لتلك المجموعة ، هذا العدد يميل لأن يقع في وسط تلك المجموعة من البيانات في حالة ترتيبها حسب صغرها أو كبرها لذا أطلق عليها مقاييس النزعة المركزية.

أهم مقاييس النزعة المركزية :

1- الوسط الحسابي.

2- الوسط الهندسي.

3- الوسط التوافقي.

4- الوسط التربيعي.

5- الوسيط.

6- المنوال.

1- الوسط الحسابي :

ويسمى في بعض الأحيان الوسط أو المتوسط أو المعدل الحسابي لقيم متغير ما ، وهو أهم مقاييس النزعة المركزية على الإطلاق لما يمتاز بخصائص جيدة بسهولة في الحساب ، وهو عبارة عن القيمة الناتجة من قسمة مجموع المشاهدات على عددها ويرمز له بالرمز $\bar{x}$.

طرق حساب الوسط الحسابي :

أ- في حالة البيانات الغير المبوبة :

الوسط الحسابي يكون مساوي إلى :

$$\bar{x} = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{\sum_{i=1}^{n} xi}{n}$$

حيث أن $\bar{x}$ يمثل تقدير الوسط الحسابي لقياسات مفردات المجتمع الذي اختيرت منه العينة.

مثال:

البيانات التالية أوجد الوسط الحسابي لأوزان عمال لعينة حجمها 10عمال.

وزن العمال

$$(50.2, 52.9, 58.1, 59.2, 60.9, 61.9, 62.3, 63.2, 65.3, 68.3)$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(50.2 + 52.9 + \dots\dots 68.3)}{10}$$

$$= \frac{601.9}{10}$$

$$= 60.19$$

يلاحظ تمركز قيمة $\bar{x}$ وسط هذه المجموعه لذا سمي مقياس نزعة مركزية

ب- حساب الوسط الحسابي في حالة البيانات المبوبة :

لتكن $x_1 , x_2 \dots , x_n$ تمثل مراكز الفئات لتوزيع تكراري عدد فئاته n وبأن $f_1 , f_2 , \dots , f_n$ تمثل التكرارات المقابلة لهذه الفئات يستخرج الوسط الحسابي ، سواء كانت أطوال الفئات متساوية أم لا حيث يعبر عن صيغة الوسط الحسابي للبيانات المبوبة بالشكل التالي :

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

حيث أن :

Xi= مركز الفئة.

$\sum_{i=1}^{n} f_i x_i$: تمثل حاصل ضرب مراكز الفئات بالتكرارات.

مثال 2 :

الآتي توزيع تكراري لدرجات الحرارة في مدينة معينة لمدة 95 يوماً يطلب إيجاد متوسط درجة الحرارة خلال هذه الفترة.

درجات الحرارة (الفئات)	عدد الأيام fi	مركز الفئة xi	حاصل ضرب fi xi
0 – 1	4	0.5	(4) (0.5) = 2
1 – 2	8	1.5	(8) (1.5) = 12
2 – 3	12	2.5	(12) (2.5) = 30
3 – 4	16	3.5	56
4 – 5	20	4.5	90
5 – 6	25	5.5	137.5
6 – 7	6	6.5	39
7 – 8	4	7.5	30
	95		396.5

مركز الفئة = $\frac{0+1}{2} = 0.5$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{396.5}{95} = 4.174$$

<u>تمرين :</u> الآتي توزيع تكراري لأطوال عينة من الأشخاص قوامها (69) شخص يطلب إيجاد متوسط طول كل شخص في هذه العينة.

فئات الأطوال	عدد الأشخاصfi	مركز الفئة xi	fi xi
140 – 148	4	144	576
148 – 156	6	152	912
156 – 164	15	160	2400
164 – 172	20	168	3360
172 – 180	17	176	2992
180 – 188	7	184	1288

188 - 196	1	192	192
	69		11720

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\mathbf{11720}}{\mathbf{69}} = 169.85$$

مزايا وعيوب الوسط الحسابي :

أ- المزايا :

1- بساطة وسهولة فكرته.

2- أن حسابه يستند إلى كافة البيانات المتاحة.

3- أن حسابه يخضع للعمليات الجبرية.

ب- العيوب :

1- لا يمكن إيجاد الوسط الحسابي وبيانات وصفية كالجنس واللون وصنف الدم ما عدا الحالات التي يتم فيها إعادة تقييس الصفات.

2- لا يمكن حساب الوسط الحسابي في حالة فقدان قيمة أو أكثر من قيم العينة.

3- أن قيمة الوسط الحسابي تتأثر بالقيم الشاذة (المتطرفة) لذا لا يمكن الاعتماد على هذا القياس في هذه الحالة.

4- لا يمكن إيجاده في حالة التوزيعات التكرارية المفتوحة من طرف أو طرفين.

5- لا يجوز الاعتماد على قيم الوسط الحسابي في حالة التوزيعات ذات التواء حاد سواء أكانت التواء سالب أو موجب بسبب انحياز قيمة الوسط الحسابي نحو جهة الالتقاء.

خصائص الوسط الحسابي :

1- مجموع انحرافات قيم المتغير x من وسطها الحسابي الذي احتسبت منه يساوي صفر :

أ- في حالة البيانات غير المبوبة.

$$\sum_{i=1}^{n}(x_i-\bar{x})$$

البرهان : نفرض ان

$d_i = x_i - \bar{x}$, *i= 1,2,……,n*

بإضافة المجموع للطرفين $\sum_{i=1}^{n} d_i = \sum_{i=1}^{n}(x_i - \bar{x})$

$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x} \longrightarrow \bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} \longrightarrow n\bar{x} = \sum_{i=1}^{n} x_i$

نعوض عن $\sum xi$ بما يساوي في المعادلة يكون :

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

ب- في حالة البيانات المبوبة :

$$\sum_{i=1}^{n} f_i(x_i - \bar{x}) = 0$$

البرهان :

$$\sum_{i=1}^{n} f_i x_i - \bar{x}\sum_{i=1}^{n} f_i \rightarrow \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$\bar{x}\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i x_i$$

نعوض عن $\sum_{i=1}^{n} f_i x_i$ بما يساويها في المعادله اعلاه

$$\bar{x}\sum_{i=1}^{n} f_i - \bar{x}\sum_{i=1}^{n} f_i = 0$$

الوسط التوافقيHarmonic mean *:*

يعتبر الوسط التوافقي أحد مقاييس النزعة المركزية ذات الاستخدامات القليلة في التطبيقات الإحصائية ويرمز له بالرمز H نسبة إلى كلمة Hormonic.

أ- في حالة البيانات غير المبوبة :

يأخذ الوسط التوافقي الصيغة التالية بعد أخذ مقلوب القيم كما يلي :

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

حيث أن : $\frac{1}{xi}$ تمثل مقلوب قيم المشاهدات.

مثال 3 :

جد الوسط التوافقي للبيانات التالية علماً أنها تمثل عدد الموظفين في بعض أقسام الشركة العامة للنقل البري :

8 , 8 , 7 , 4 , 3 , 5 , 2 قيم المشاهدات

الحل :

$$\sum_{i=1}^{n} \frac{1}{x_i} = \frac{1}{2} + \frac{1}{5} + \frac{1}{3} + \frac{1}{4} + \frac{1}{7} + \frac{1}{8} + \frac{1}{8}$$

$$= 0.5 + 0.2 + 0.33 + 0.25 + 0.14 + 0.13 + 0.13 = 1.68$$

$$H = \frac{7}{1.68} = 4.17$$

ب- في حالة البيانات المبوبة :

يأخذ الوسط التوافقي الصيغة التالية بالاعتماد على مقلوب قيم مركز الفئة :

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}}$$

علماً أن fi تمثل التكرار xi مراكز الفئات.

مثال 4 :

البيانات التالية تمثل عدد العاملين f أحد المصانع التابعة لوزارة الصناعة.

المطلوب : احسب الوسط التوافقي ؟

فئات الأجر	عدد العاملين fi	مركز الفئة xi	$\frac{fi}{xi}$
50 – 60	8	55	$\frac{8}{55}$ = 0.145
60 – 70	10	65	$\frac{10}{165}$ = 0.154
70 – 80	16	75	$\frac{16}{75}$ = 0.213
80 – 90	14	85	0.165
90 – 100	10	95	0.105
100 – 110	5	105	0.048
110 - 120	2	115	0.017
	65		0.874

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}} = \frac{65}{0.874} = 74.741$$

عيوب الوسط التوافقي :

1- لا يمكن تحديد قيمة الوسط التوافقي إذا كانت إحدى قيم المتغير العشوائي مساوية للصفر أو إحدى مراكز فئات التوزيع التكراري مساوية للصفر.

2- لا يمكن إيجاد قيمته في حالة التوزيعات التكرارية المفتوحة من طرف واحد أو كلا الطرفين.

3- لا يمكن حساب قيمته في حالة فقدان قيمة أو أكثر من قيم العينة.

4- لا يمكن حساب قيمته في حالة فقدان البيانات النوعية.

3- الوسط الهندسي ***Geometric mean*** :

أ- في حالة البيانات غير المبوبة :

لو فرضنا لدينا القيم أو المشاهدات التالية $x_1 \; , \; x_2 \; , \; \dots \; , \; x_n$، فالوسط الهندسي هو عبارة عن الجذر النوني الحاصل ضرب تلك القيم ويرمز له بالحرف G نسبة إلى كلمة Geometric وكما يلي :

$$G = \sqrt[n]{x_1 . x_2 . \dots . x_n}$$

يرفع الجذر : $G = (x_1 . x_2 . \dots . x_n)^{\frac{1}{n}}$

بأخذ لوغاريتم الطرفين :

$$\log G = \frac{1}{n} \log (x_1 . x_2 . \dots . x_n)$$

$$\text{Log G} = \frac{\mathbf{1}}{\boldsymbol{n}} [\log \boldsymbol{x}\mathbf{1} + \log \boldsymbol{x}\mathbf{2} + \dots + \log \boldsymbol{n}]$$

الصيغة اللوغارتمية للوسط الهندسي في حالة البيانات غير المبوبة

$$\log G = \frac{x \sum_{i=1}^{n} \log x_i}{n}$$

مثال 5 :

القيم التالية تمثل عدد الوحدات المنتجة في أحد المصانع الإنتاجية.

المطلوب : أوجد الوسط الهندسي لهذه الوحدات :

Xi : 3 , 5 , 7 , 8 , 3 , 7 , 2

الحل :

- الطريقة الأولى :

$$G = \sqrt[6]{(\mathbf{3})(\mathbf{5})(\mathbf{7})(\mathbf{3})(\;\mathbf{7})(\;\mathbf{2})} = 4.14$$

- الطريقة الثانية : باستخدام اللوغاريتم :

$$\text{Log G} = \frac{\mathbf{X} \sum_{i=1}^{6} \log xi}{\mathbf{6}} = \frac{\log(3)+\log(5)+\log(7)+\log(3)+\log(7)+\log(2)}{\mathbf{6}}$$

$$= \frac{\mathbf{0.477+0.698+0.903+0.477+0.845+0.301}}{\mathbf{6}}$$

$$\text{Log } G = 0.617 \rightarrow G = 4.14$$

ملاحظة : في المثال رقم 5 أعلاه في حالة احتساب الوسط الحسابي للبيانات فأنه يكون مساوي إلى $\overline{x} = 4.64$، ومن هذه النتيجة يتضح أن الوسط الحسابي لمجموعة من القيم أكبر من الوسط الهندسي لتلك القيم ، كما أنه إذا كان أحد القيم سالباً فلا يمكن إيجاد الوسط الهندسي لها.

ب- الوسط الهندسي في حالة البيانات المبوبة :

الوسط الهندسي في حالة البيانات المبوبة يمثل وسط هندسي مدون ، حيث أن التكرارات في الواقع تمثل الأوزان التي تبين أهمية كل فئة ، فلو كانت $x_1 , x_2 , \dots , x_n$ تمثل مراكز الفئات في جدول التوزيع التكراري وتكراراتها هي $f_1 , f_2 , \dots , f_m$ فأن الوسط الهندسي يكون مساوي للصيغة التالية :

$$G = \sqrt[\sum_{i=1}^{m} f_i]{\genfrac{}{}{0pt}{}{f_1, f_2, \dots . f_m}{x_1 . x_2 \dots . x_m}}$$

ونظراً لصعوبة حل الصيغة أعلاه نأخذ لوغاريتم الطرفين بعد رفع الجذر كما يلي :

$$G = \begin{pmatrix} f \;,, f \;,,,, f \\ x_1 . x_2 \dots . x_m \end{pmatrix}^{\frac{1}{\sum_{i=1}^{m} f_i}}$$

$$\log G = \frac{1}{\sum_{i=1}^{m} f_i} \log \begin{pmatrix} f \;,, f \;,,,, f \\ x_1 . x_2 \dots . x_m \end{pmatrix}$$

$$= \frac{1}{\sum_{i=1}^{m} f_i} \; (f_1 \log x_1 + f_2 \log x_2 + \dots + f_m \log x_m)$$

الصيغة النهائية باستخدام اللوغاريتم

$$\log G = \frac{\sum_{i=1}^{m} f_i \log x_i}{\sum_{i=1}^{m} f_i}$$

مثال // البيانات التالية تمثل عدد الموظفين المحالين على التقاعد والاستقالة هي إحدى الوزارات موزعين حسب الفئات العمرية لهم .

المطلوب : أوجد الوسط الهندسي لهذه البيانات :

الفئات العمرية	عدد العاملين fi	مركز الفئة xi	log xi	fi log xi
60 – 62	5	61	1.79	(5) (1.79) = 8.8910
63 – 65	18	64	1.81	(18) (1.81) = 32.5116
66 - 68	42	67	1.81	76.6962
69 – 71	27	70	1.85	49.8177
72 – 74	8	73	1.86	14.9064
	100			182.8229 $\rightarrow \sum_{i=1}^{m} fi \log xi$

$$\log G = \frac{\sum_{i=1}^{m} f_i \log x_i}{\sum_{i=1}^{m} f_i} = \frac{\mathbf{182.8229}}{\mathbf{100}} = 1.828229 = 67.3$$

عيوب الوسط الهندسي :

1- لا يمكن استخدامه في حالة الجداول المفتوحة.

2- لا يمكن استخدامه إذا كانت ‘احدى قيم المجموعة صفر.

3- لا يمكن استخدامه إذا كانت قيم المجموعة كمية سالبة.

4- الوسط الحسابي الموزون (المرجح) :

في بعض الأحيان تكون إحدى المفردات أكثر أهمية من الاخرى لذلك يتوجب أخذ ذلك بنظر الاعتبار عند حساب الوسط الحسابي الموزون.

أ- في حالة البيانات غير المبوبة :

افترض أن $x_1 , x_2 , \ldots , x_n$ تمثل قياسات عينة من المفردات قوامها n وأن $w_1 , w_2 , \ldots , w_n$ أوزان هذه المفردات يعرف الوسط الحسابي المرجح على النحو التالي :

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

مثال 7 : لديك القيم التالية وأوزانها ، المطلوب إيجاد الوسط الحسابي الموزون :

Xi	Wi	xi wi
70	10	(70) (10) = 700
60	30	(60) (30) = 1800
75	10	(75) (10) = 750
55	50	(55) (50) = 2750
المجموع	100	6000

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{\mathbf{6000}}{\mathbf{100}} = 60$$

ب- الوسط الحسابي الموزون في حالة البيانات المبوبة :

أما في حالة التوزيعات التكرارية فأن الوسط الحسابي الموزون يكون وفق الصيغة التالية :

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i f_i x_i}{\sum_{i=1}^{n} w_i f_i}$$

حيث أن fi تمثل التكرار

Xi = مركز الفئة

wi = الأوزان

مثال 8 : الآتي توزيع تكراري لإنتاج مصنع معين (طن) من سلعة معينة لأحد الأيام موزع حسب عدد المكائن العاملة في ذلك اليوم وعدد ساعات العمل المحددة لإشغال كل ماكنة حسب مواصفات المنشأ ، يطلب حساب متوسط انتاجية الماكنة الواحدة في هذا المصنع.

فئات الانتاج	عدد المكائن Fi	عدد الساعات الاوزان wi	مركز الفئة xi	wi fi	wi fi xi
2 – 4	4	6	3	(6) (4) = 24	(24) (3) = 72
4 – 6	5	5	5	(5) (5) = 25	(25) (5) = 125
6 – 8	6	6	7	(6) (6) = 36	(36) (7) = 252
8 – 10	3	4	9	(3) (4) = 12	(12) (9) = 108
10 – 12	2	4	11	(2) (4) = 8	(8) (11) = 88
	20			105	645

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i f_i x_i}{\sum_{i=1}^{n} w_i f_i} = \frac{\mathbf{645}}{\mathbf{105}} = 6.143$$

5- الوسط التربيعي Quadratic mean :

يعرف الوسط التربيعي بأنه الجذر التربيعي الموجب لمتوسط مربعات قيم المتغير العشوائي (x) كما أن قيمة الوسط التربيعي هي قيمة موجبة دائماً استناداً إلى نوع البيانات ، ويرمز له بالرمز (Q) نسبة إلى Quadratic.

أ- حساب الوسط التربيعي في حالة البيانات غير المبوبة :

يحسب الوسط التوافقي وفق الصيغة التالية بعد تربيع قيم المشاهدات وأخذ مجموعها وكما يلي :

$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

مثال 9 : جد الوسط الحسابي التربيعي للبيانات التالية :

2 , 3 , 4 , 5 , 6

الحل :

$$Q = \sqrt{\frac{(2)^2 + (3)^2 + (4)^2 + (5)^2 + (6)^2}{5}} = \sqrt{\frac{\mathbf{4+9+16+25+36}}{\mathbf{5}}} = \sqrt{\frac{\mathbf{8}}{\mathbf{5}}} = \sqrt{\mathbf{1.6}} = 1.2$$

ب- الوسط التربيعي في حالة البيانات المبوبة :

ويحسب ووفق الصيغة التالية ، حيث ان fi تمثل عدد التكرارات

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}}$$

مثال 10 :

التوزيع تكراري التالي يمثل عدد المشاركين في دورات التعليم المهني التابع لوزارة الصناعة موزعين حسب عدد أيام المشاركة لكل دورة.

المطلوب : أوجد الوسط التربيعي لعدد أيام المشاركة.

فئات الانتاج	عدد المشاركين fi	مركز الفئة xi	xi^2	$fi\ xi^2$
10 – 20	2	15	$(15)^2 = 225$	(2) (225) = 450
20 – 30	3	25	$(25)^2 = 625$	(3) (625) = 1875
30 – 40	4	35	$(35)^2 = 1225$	(4) (1225) = 4900
40 – 50	10	45	$(45)^2 = 2025$	(10) (2025) = 20250
50 - 60	36	55	$(55)^2 = 3025$	(36) (3025) = 108900
60 - 70	14	65	$(65)^2 =$	(14) (4225) = 59150

			4225	
المجموع	69			195525

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}} = \sqrt{\frac{\mathbf{195525}}{\mathbf{69}}} = \sqrt{\mathbf{2833.7}} \quad Q = 53.23$$

مميزات الوسط التربيعي :

1- يمكن إيجاده إذا كانت أطوال الفئات متساوية أو غير متساوية.

2- بساطة فكرته وخضوعه للعمليات الجبرية كما أن عملية حسابه تستند إلى كلفة البيانات المتاحه

عيوب الوسط التربيعي :

1- لا يمكن إيجاده في حالة التوزيعات التكرارية المفتوحة من طرف واحد أو طرفين.

2- لا يمكن حساب قيمته لبيانات متغير وصفي.

3- لا يمكن تعيينه هندسياً.

4- لا يمكن تحديد قيمته في حالة فقدان إحدى قيم العينة.

2- مقاييس توسط أخرى :

- مقاييس توسط أخرى :

1- المنوال (MO) the mode

المنوال لمجموعة من القيم هو القيمة المفردة التي تتكرر أكثر من بقية المفردات في مجموعة القيم ، أي أن المنوال هو القيمة الأكثر تكراراً بين المفردات.

طرق إيجاد المنوال :

أ- في حالة البيانات غير المبوبة :

لتكن $x_1 , x_2 , \ldots x_n$ قياسات عينة من المفردات حجمها n وافترض أن x_i قيمة من قيم المجموعة تتكرر أكثر من غيرها ، عندئذ وحسب تعريف المنوال فإن x_i تمثل المنوال لهذه المجموعة.

مثال 12 :

للبيانات التالية جد المنوال :

20 , 40 , 60 , 80 , 20 , 90 , 20 , 50

واضح أن العدد 20 تكرر ثلاث مرات فالمنوال هو العدد (20) أي أن $x_i = 20$.

مثال 13 :

للبيانات التالية في المجاميع أدناه أوجد المنوال :

$X_i = 20 , 40 , 60 , 80$

لا يوجد قيمة للمنوال.

$X_i = 20 , 40 , 40 , 60 , 60, 80 , 90$

$Mo = 40 , Mo = 60$

$X_i = 20 , 20 , 20 , 40 , 40 , 60 , 80 , 90 , 90$

$Mo = 20$

ب- إيجاد المنوال لبيانات مبوبة :

1- في حالة المتغير المتقطع :

المنوال في هذه الحالة يمثل مركز تلك الفئة التي تقابل أكبر تكرار في التوزيع.

ملاحظة :

في حالة وجود فئتين أو أكثر تقابل نفس التكرار فأن توزيع من هذا النوع سوف يملك أكثر من قيمة للمنوال ، وكل منها تمثل مركز لتلك الفئة التي تقابل ذلك التكرار.

مثال 14 :

البيانات التالية توزيع 40 عائلة حسب الدخل الشهري لها محسوب بالآلاف الدنانير.

المطلوب : تحديد المنوال :

الحل :

نبحث عن أكبر تكرار في هذا التوزيع وهو العدد 14 المقابل للفئة 104 – 90 وعليه فأن المنوال في هذا التوزيع يكون مساوي إلى :

Mo = $\frac{\mathbf{104+90}}{\mathbf{2}}$ =97

عدد الفئات والدخل بالآلاف الدنانير	التكرار fi عدد العوائل
60 – 74	2
75 – 89	6
90 – 104	14
105 – 119	10
120 – 134	8

مع ملاحظة أنه يتم تقريب الناتج في حالة وجود كسور إلى أقرب عدد صحيح.

2- في حالة المتغير المستمر :

أفرض أن هناك توزيعاً تكرارياً عدد فئاته m مرات f_K تمثل أكبر تكرار هذا يعني أن الفئة التي تحتوي على المنوال هي الفئة المقابلة إلى f_k وأن f_{k-1} يمثل التكرارات السابقة لتكرار فئة المنوال و f_{k+1} تمثل التكرار اللاحق لتكرار فئة المنوال ، أي أن :

$F_{k-1} < f_k < f_{k+1}$

MO =a+ $\frac{d1}{d1+d2} * \mathrm{L}$

حيث ان : d_1 تكرار الفئة المنوالية – تكرار الفئة التي قبلها

d_2 تكرار الفئة المنوالية – تكرار الفئة التي بعدها

a تمثل الحد الأدنى لفئة المنوال.

L تمثل طول فئة المنوال.

مثال 15 :

الآتي توزيع تكراري لأعمار عدد من المرضى الراقدين في إحدى المستشفيات ، يطلب إيجاد العمر الشائع للمرضى في هذه المجموعة.

الحل :

إن أكبر تكرار هو 23 فأن فئة المنوال هي الفئة المقابلة لأكبر تكرار أي أن 60 – 50.

الفئات	عدد المرضى fi
أقل من 20	2
20 – 30	8
30 – 40	16
40 – 50	17
50 – 60	23
60 فأكثر	14

L = 10 , d1= 6 , d2 = 9 , a = 50

MO $=a+ \frac{d1}{d1+d2} * \mathrm{L}$

Mo $= 50 + \frac{(6)}{(6)+(9)} * 10$

$= 50 + \frac{6}{15} * 10$

= 50+ (0.4) 10 =54

يلاحظ على الرغم من كون التوزيع مفتوح من كلال الطرفين إلا أنه أمكن إيجاد المنوال وهذه إحدى مزايا المقياس.

2- الوسيط The Median :

يعتبر الوسيط أحد مقاييس النزعة المركزية المهمه في التطبيقات الإحصائية ويعرف الوسيط بأنه تلك القيمة من قيم المتغير العشوائي (x) التي تقسم مجموعة المتغيرين إلى قسمين متساويين أي أنها قيمة (x) التي تجعل عدد القيم قبلها مساوي لعدد القيم بعدها.

<u>طرق إيجاد الوسيط :</u>

أ- إيجاد الوسيط لبيانات غير مبوبة :

لتكن x_n , ... , x_2 , x_1 تمثل قياسات مفردات عينة حجمها n ورتبت ترتيباً تصاعدياً أو تنازلياً فأن :

1- إذا كان عدد القيم فردي (n) فأن قيمة الوسيط تمثل قيمة x بعد الترتيب ، اي : $\frac{n+1}{2}$.

مثال 16 :

الآتي درجات عينة عددها تسعة من الطلبة ، جد الوسيط لهذه المجموعة :

55 , 62 , 53 , 70 , 68 , 65 , 63 , 79 , 80

الحل :

- ترتب القيم ترتيبا تصاعديا كما يلي :

الترتيب التصاعدي للقيم
53
55
62
63
الوسيط 65
68
70
79
80

- ترتيب الوسيط هو 5 = $\frac{9+1}{2}$ = $\frac{n+1}{2}$.

هذا يعني أن القيمة الخامسة هي :

قيمة الوسيط أي الدرجة Me = 65

2- إذا كان عدد القيم n زوجي فأن قيمة الوسيط تمثل الوسط الـ

$$Me = \left[x_{\frac{n}{2}+1} + x_{\frac{n}{2}} \right] / 2$$

<u>مثال 17</u> :الآتي اعمار عينة من الأفراد حجمها 12 فرد جد الوسيط لعمر الفرد في المجموعة :

الحل :25 , 23, 20 , 18 , 29 , 28 , 27 , 24.5 , 26 , 19.5 , 22 , 20

الحل: نرتب هذه القيم تنازلياً

29 , 28 , 27 , 26 , 25 , 24.5 , 23 , 22 , 20 , 20 , 19.5 , 18

القيمتان اللتان تحددان الوسيط هما :

$\frac{12}{2}+1 = 7 + \frac{12}{2} = 6$

أي القيمة السادسة والسابعة بعد الترتيب ، إذ الوسيط هو الوسط الحسابي لهاتين القيمتين :

$Me = \frac{24.5+23}{2} = 23.75$

ب- إيجاد الوسيط لبيانات مبوبة :

1- في حالة المتغير المتقطع :

أفرض وجود توزيع تكراري لبيانات متغير متقطع عدد فئاته m فإن $f_1 , f_2 , \dots , f_m$ تمثل التكرارات المقابلة لهذا التوزيع إن $f_1 , f_2 , \dots , f_m$ تمثل التكرارات المتجمعة الصاعدة المقابلة للحدود العليا لفئات التوزيع . نحدد الوسيط وهو الذي يترك نصف مجموع التكرارات قبلها والنصف الآخر بعدها ولتحديد الوسيط نقارن الوسيط مع التكرار المتجمع الصاعد ، فإذا كان :

$$F_{K-1} < \frac{\sum_{i=1}^{m} f_i}{2} < F_{k+1}$$

يقال أن فئة الوسيط هي التي تسلسلها (k+1) وإن :

$$Me = \frac{\text{الحد الأدنى لفئة الوسيط+الحد الأعلى لفئة الوسيط}}{2}$$

مثال :

الآتي توزيع تكراري لعينة من الأسر حسب عدد أفراد الاسرة.

المطلوب : حساب الوسيط لعدد أفراد الاسرة.

الحل :

إن ترتيب الوسيط :

$$\frac{\sum_{i=1}^{m} f_i}{n} = \frac{80}{2} = 40$$

$$F_{k-1} < \frac{\sum_{i=1}^{m} f_i}{n} < F_{k+1}$$

$$27 < 40 < 47$$

فئات عدد الافراد	عدد الأسر f	تكرار متجمع صاعد F_i
2 – 4	6	6
5 – 7	9	15
8 – 10	12	27
11 – 13	20	47
14 – 16	14	61
17 – 19	11	72
20 - 22	8	80
	80	

عليه فأن فئة الوسيط هي الرابعة اي التي تسلسلها (k+1) اي 11-13

$$Me = \frac{11+13}{2} = 12$$

2- إيجاد الوسيط لبيانات مبوبة / في حالة متغير مستمر :

أفرض f_1 , f_2 , ... ,f_m ،التكرارات المقابلة للتوزيع ، f_1 , f_2 , ... ,f_m التكرارات المتجمعة الصاعدة المقابلة للحدود العليا لفئات التوزيع.

$\frac{\sum_{i=1}^{m} fi}{2}$ ترتيب الوسيط في التوزيع ، فإذا كان :

$$F_{K-1} < \frac{\sum_{i=1}^{m} f_i}{2} < F_{k+1}$$

يقال أن فئة الوسيط هي التي تسلسلها $_{k+1}$ وأن الوسيط يكون مساوي إلى القانون التالي :

$$Me = a + \frac{Lk}{fk}\left(\frac{\sum_{i=1}^{m} fi}{2} - F\right)$$

حيث أن :

a: تمثل الحد الأدنى لفئة الوسيط

f_k = تكرار فئة الوسيط

L_k = طول فئة الوسيط

F = التكرار المتجمع الصاعد السابق لفئة الوسيط

مثال 19 :

الآتي توزيع تكراري للدخول الشهرية لعينة من الأسر حجمها (80) أسرة ، جد الوسيط للدخل الشهري للأسر في هذه العينة :

الفئات: 100 120 140 160 180 200 220 240

عدد الأسر: 3 7 14 20 18 12 6

الحل :

الفئات	Fi	التكرار المتجمع الصاعد Fi
100 – 120	3	3
120 – 140	7	10
140 – 160	14	24
160 – 180	20	44
180 – 200	18	62
200 – 220	12	74
220 – 240	6	80
	80	

$$\frac{\sum_{i=1}^{m} fi}{2} = \frac{80}{2} = 40$$

$$24 < 40 < 44$$

ترتيب الوسيط هو الفئة الرابعة أي الفئة $160-180$

$$Me = a + \frac{bk}{fk}\left(\frac{\sum fi}{2} - F\right)$$

$a = 160$, $b_k = 20$, $f_k = 20$, $F = 24$, $\sum fi = 80$

$$Me = 160 + \frac{\mathbf{20}}{\mathbf{20}}\left(\frac{\mathbf{80}}{\mathbf{2}} - 24 \right)$$

= 160 + 1 (40 - 24) = 176 دينار

العلاقة بين بعض مقاييس النزعة المركزية :

ترتبط بعض مقاييس النزعة المركزية مع بعضها بعلاقات معينة منها العلاقة بين الوسط الحسابي ، الوسيط ، المنوال ، إذ لوحظ في حالة المنحنيات المتماثلة تكون قيم المتوسطات الثلاثة متساوية ويزداد الاختلاف بين هذه القيم كلما بعد التوزيع عن التماثل ، ولكن هناك علاقة تربط بين هذه المتوسطات في حالة المنحنيات القريبة من التماثل وهي :

$$\bar{x} - Me = \frac{1}{3}(\bar{x} - Mo)$$

هذه العلاقة مهمة جداً أثناء التطبيق حيث يتعذر حساب الوسط من توزيع تكراري مفتوح من طرف أو طرفين ويمكن إيجاد الوسيط والمنوال فأن :

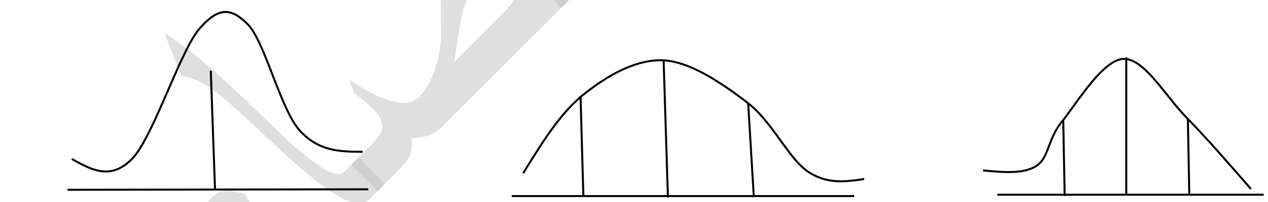
$$\bar{x} = \frac{\mathbf{3Me - Mo}}{\mathbf{3}}$$

وإن :

$$Me = \frac{\mathbf{1}}{\mathbf{3}}(2\bar{x} + Mo)$$

$$Mo = 3Me - 2\bar{x}$$

ويمكن توضيح العلاقة بين هذه الأوساط بالرسوم التالية :



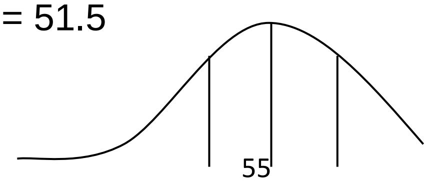
$Mo = \bar{x} = Me$ — الرسم متماثل

Mo Me $\bar{x}$ — $Mo < Me < \bar{x}$ — منحنى ملتوي التواء موجب من جهة اليمين

$\bar{x}$ Me Mo — $\bar{x} < Me < Mo$ — منحنى ملتوي التواء سالب من جهة اليسار

مثال 20: تعذر في أحد التوزيعات القريبة من حالة التماثل الحصول على قيمة الوسط الحسابي ، حيث أمكن الحصول على قيمة Me , Mo ، حيث Me = 52 ، Mo = 53 ، جد الوسط الحسابي مبين نوع الالتقاء.

$$\bar{x} = \frac{\mathbf{3\,Me - Mo}}{\mathbf{2}} = \frac{\mathbf{3(52) - 53}}{\mathbf{2}} = 51.5$$



$$\bar{x} < Me < Mo$$

منحنى ملتوي التواء سالب من جهة اليسار

الفصل الرابع

مقاييس التشتت

مقدمة :

وهي نوع آخر من المقاييس الإحصائية التي تصف المجموعات الإحصائية أو التوزيعات التكرارية في المقاييس التي سبق ذكرها ، فلو فرضنا لدينا الراتب الشهري لخمسة موظفين وكان بالصورة التالية :

100 , 17 , 21 , 22 , 25 فالوسط الحسابي لهذه الرواتب والذي هو عبارة عن :

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25+22+\cdots+100}{5} = 37$$

هذا المتوسط الذي يستخدم لتمثيل الرواتب لا يمثلها تمثيلاً صحيحاً لهذا السبب سوف نستخدم مقاييس أخرى ، وهي مقايس التشتت والهدف من دراسة التشتت هو تكوين فكرة عن مدى تجانس قيم مجموعة من المفردات ، وهذا يعني أن دراسة التشتت أمر مفيد في إجراء المقارنة بين مجموعتين أو أكثر من البيانات ، وكلما كان مقياس التشتت كبيراً دل ذلك على عدم التجانس بين القيم ويكون مقياس التشتت صغيرا عندما يكون الاختلاف من بين قيم المشاهدات قليل ، ولمقياس التشتت أهمية أيضاً في وصف التوزيعات ومقارنتها مع بعضها.

ملاحظة :

قد يتساوى الوسط الحسابي لمجموعتين من القيم بينما يختلف مدى انتشار القيم في المجموعة الأولى عن المجموعة الثانية كما مبين أدناه :

إذا كانت قيم المجموعة الأولى هي : 17 , 20 , 23 , 18 , 12 , 21 , 22

وكانت قيم المجموعة الثانية هي : 35 , 10 , 7 , 5 , 45 , 20 , 13

فالوسط الحسابي لكلا المجموعتين مساوي(20) ولكن المجموعة الأولى متجانسة أكثر من الثانية.

أهم مقاييس التشتت :

أولاً : مقاييس التشتت المطلقة :

وهي التي تقيس تشتت المفردات بدلالة الوحدات المطلقة المستخدمة في قياس الظاهرة مثل فلس ، دينار ، درجة ، وزن ، كثافة ... الخ ، ومنها :

أ- مقاييس تقيس تشتت المفردات بينها ومنها على سبيل المثال :

1- المدى.

ب- مقاييس تقيس تشتت المفردات حول الوسط الحسابي ومنها :

1- الانحراف المتوسط. 2- التباين والانحراف المعياري.

ثانياً : مقاييس التشتت النسبية :

وهي التي تقيس تشتت المفردات بدلالة النسب المئوية وأهمها ما يسمى بمعامل الاختلاف.

أولاً : مقاييس التشتت المطلقة :

1- المدى : في حالة البيانات غير المبوبة :

يعتبر المدى أبسط أنواع مقاييس التشتت المطلقة ويعرف بأنه الفرق بين أكبر قيمة في مجموعة البيانات وأصغر قيمة حيث :

$R = x_L - x_s$

R = يمثل مقياس المدى

X_L:يمثل أكبر قيمة في مجموعة البيانات.

X_s : يمثل أصغر قيمة في مجموعة البيانات.

مثال 1 : أوجد المدى للقيم التالية :

$X_i : 2 , 10 , 12 , 15 , 22$

$R = x_L - x_s \rightarrow 22 - 2 = 20$

المدى في حالة البيانات المبوبة :

ويتم حساب المدى للبيانات المبوبة بإيجاد حاصل الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الاولى ، أي أن :

المدى = الحد الاعلى للفئة الاخيرة – الحد الأدنى للفئة الاولى :

مثال 2 : أوجد المدى للجدول التكراري التالي :

الحل :

الحد الأدنى للفئة الأولى – الحد الاعلى للفئة الأخيرة =R

$R = 80 - 40 = 40$

الفئات	f_i
40 – 50	2
50 – 60	10
60 – 70	15
70 – 80	3

ب– مقاييس تشتت المفردات حول الوسط الحسابي :

1– الانحراف المتوسط : وهو من مقاييس التشتت التي تقيس تشتت المفردات حول وسطها الحسابي الحقيقي ويعتمد على القيمة المطلقة للانحرافات.

أ– في حالة البيانات الغير مبوبة : إذا كان $x_1 , x_2 , \ldots , x_n$ مجموعة من المشاهدات أو القيم فالانحراف المتوسط لها هو متوسط الانحرافات المطلقة (بإهمال الإشارة) عن وسطها الحسابي الحقيقي ويرمز له بالرمز M.Dوعليه فأن الانحراف المتوسط هو :

$$M.D. = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

مثال 3 :

أوجد الانحراف المتوسط للبيانات التالية :

x_i	$x_i - x$	$\|xi - x\|$
9	9 − 7 = 2	$\|2\|$ = 2
8	8 − 7 = 1	1 = 1
6	6 − 7 = −1	$\|-1\|$ = 1
5	5 − 7 = −2	$\|-2\|$ = 2
7	7 − 7 = 0	$\|0\|$ = 2
35		6

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{35}{5} = 7$$

$$M.D. = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n} = \frac{6}{5} = 1.2$$

ب− في حالة البيانات المبوبة :

إذا كان x_n , ... , x_2 , x_1تمثل مراكز الفئات في جدول التوزيع التكراري وتكراراتها , ... , f_2 , f_1

$x_=f_m$ فأن الانحراف المتوسط هو :

$$M.D. = \frac{\sum_{i=1}^{m} f_i \left| x_i - \bar{x} \right|}{\sum_{i=1}^{m} f_i}$$

مثال 4: أوجد الانحراف المتوسط لجدول التوزيع التكراري التالي والذي يمثل عدد الموظفين المحالين على التقاعد موزعين حسب الفئات العمرية.

الحل :

الفئات	fi	xi	fi xi	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$fi\lvert x_i - \bar{x} \rvert$
60 – 62	5	61	(5) (61) = 305	61 – 67.45 = 6.45	6.45	(5) (6.45) = 32.25
63 – 65	8	64	1152	64 – 67.45 = 3.45	3.45	(8) (3.45) = 68.1
66 – 68	42	67	2814	67 – 67.45 = 0.45	0.45	18.9
69 – 71	27	70	1890	3.45	3.45	68.85
73 – 74	8	73	584	5.55	5.55	44.40
	100		6745			226.5

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} = \frac{6745}{100} = 67.45$$

$$M.D. = \frac{\sum_{i=1}^{m} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{m} f_i} = \frac{226.5}{100} = 2.265$$

2– مقاييس تقيس تشتت المفردات حول الوسط الحسابي :

التباين والانحراف المعياري The variance and standard Deviations:

- **التباين :** يعرف التباين بأنه مربع الانحراف المعياري ويرمز له بالرمز *S2*وعليه فأن وحدات قياس التباين تمثل مربع وحدات قياس المتغير الأصلي.
- **الانحراف المعياري :** من أهم مقاييس التشتت وأكثرها استخداماً وشيوعاً وأهمية الانحراف المعياري كمقياس تأتي كأهمية الوسط الحسابي بالنسبة للمتوسطات.

ويعرف الانحراف المعياري لمجموعة من القيم بأنه الجذر التربيعي لمتوسط مربعات انحرافات القيم من وسطها الحسابي الحقيقي.

طرق حساب التباين والانحراف المعياري :

1- في حالة البيانات غير المبوبة :

مثال : أوجد التباين والانحراف المعياري للبيانات التالية والتي تمثل عدد المنتوجات اليومية من المصابيح الكهربائية لمصنع انتاج المصابيح الكهربائية :

x_i	$x_i - \bar{x}$	$(xi - x)^2$
3	$3 - 11 = -8$	$(-8)^2 = 64$
7	$7 - 11 = -4$	$(-4)^2 = 16$
10	-1	1
15	4	16
20	9	81
55		178

$$\bar{x} = \frac{\sum xi}{n} = \frac{55}{5} = 11$$

$$s^2 = \frac{\sum (xi - \bar{x})^2}{n}$$

$$s^2 = \frac{178}{5} = 35.6$$

ويمكن ايجاد الانحراف المعياري عن طريق جذر التباين

$$S = \sqrt{35.6} = 5.9$$

$$S = \sqrt[2]{\frac{\sum (xi - \bar{x})^2}{n}}$$

$$S = \sqrt{\frac{178}{5}} = \sqrt{35.6} = \mp 5.9 = \mp 6$$

ب- حساب التباين والانحراف المعياري في حالة البيانات المبوبة :

لتكن $x_1 , x_2 , \ldots , x_m$ تمثل مراكز الفئات لتوزيع تكراري عدد فئاته $f_1 , f_2 , \ldots , f_m$ تمثل التكرارات المقابلة لهذه الفئات ، عندئذ وحسب التعريف لهذا المقياس يمكن حساب قيمته وفق الصيغة التالية :

$$s^2 = \frac{\sum fi\,(xi-\bar{x})^2}{\sum fi}$$

حيث ان : $\sum_{i=1}^{n} f_i = n$ و $\bar{x}$ تمثل الوسط الحسابي في حالة البيانات المبوبة ويكون

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

مثال 6 : أوجد التباين والانحراف المعياري للجدول التالي ، علماً أن البيانات التالية تمثل عدد الموظفين المحالين على التقاعد حسب فئاتهم العمرية.

الفئات	Fi	Xi	fi xi	(xi – x)	$(x_i - \bar{x})^2$	$fi(x_i - \bar{x})^2$
60 – 62	5	61	(5) (61) = 305	−5.45	29.70	5 (29.70) = 148.5
63 – 65	18	64	1152	−2.45	6	18 (6) = 108
66 – 68	42	67	2814	−0.55	0.30	42 (0.30) = 12.6
69 – 71	27	70	1790	3.55	12.60	27 (12.60) = 340.2
72 – 74	8	73	584	6.55	42.90	8 (42.90) = 343.2
المجموع	100		66.45			952.5

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{6645}{100} = 66.45$$

$$s^2 = \frac{\sum fi\,(xi-\bar{x})^2}{\sum fi}$$

$= \frac{952.5}{100} = 9.52$

وعلية فان الانحراف المعياري هو :

$S = \sqrt{s^2} = \sqrt{9.52}$

$S = 3.08$

$$S = \sqrt[2]{\frac{\sum fi\,(xi - \bar{x})^2}{\sum fi}} = \sqrt{\frac{952.5}{100}} = \sqrt{9.52}$$

ثانياً : مقاييس التشتت النسبية : من أهم مقاييس التشتت النسبية هو ما يسمى بمعامل الاختلاف وهناك عدة أنواع من معاملات الاختلاف كل نوع يستخدم حسب التجربة ونوع الجدول وأهم أنواعها هو :

معامل الاختلاف القياسي ويعرف بالقانون :

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} * 100\%$$

وباللغة الانكليزية يكتب :

$$C.V = \frac{S}{\bar{x}} * 100\%$$

مثال 6 :

أجريت دراسة لطول النباتات بالسم لأحد المحاصيل وكمية المحصول بالكغم لـ 150 نباتاً من نباتات الذرة وكانت النتائج مقسمة بالصورة التالية بالنسبة لصفة الطول وبكمية المحصول والمطلوب المقارنة بين تشتت الصيغتين :

الطول	كمية المحصول
$200 \rightarrow \bar{x}$	800
$16 \rightarrow S$	36

$$C.V. = \frac{S}{\bar{x}} * 100\%$$

بالنسبة للطول $= \frac{16}{200} * 100\% = 8\%$

بالنسبة لكمية المحصول $C.V. = \frac{36}{800} * 100\% = 4.5\%$

نلاحظ تشتت الطول أكبر مما عليه بالنسبة لكمية المحصول.

الفصل الخامس

الارتباط والانحدار

الارتباط الخطي :

إن مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين أو أكثر يرتبطان مع بعضهما بعلاقات خطية معينة ، على سبيل المثال العلاقة بين طول الشخص ووزنه ، فإذا كان التغيير في إحدى المتغيرات يؤثر في تغيير متغير آخر أو مجموعة متغيرات عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها ، وإذا كان المتغيرين يتغيران بنفس الاتجاه أي زيادة (نقصان) في أحدها تؤدي إلى زيادة (نقصان) في الآخر عندئذ يقال أن الارتباط فيما بينهما هو موجب ، أما إذا كان المتغيرين المرتبطين أو مجموعة المتغيرات يتغيران باتجاه معاكس أي زيادة (نقصان) في أحدهما يؤدي إلى نقصان (زيادة) في الآخر عندئذ يقال الارتباط هو سالب.

الارتباط الخطي البسيط : يعرف بأنه درجة أو القيمة العددية للعلاقة بين متغيرين وفيما يلي المفاهيم الأساسية التي توضح فكرة الارتباط البسيط.

اولاً : معامل الارتباط الخطي (بيرسون)

حساب معامل الارتباط البسيط للبيانات: وعليه فأن معامل الارتباط يكون مساوي للقانون التالي :

$$r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2 (n\sum y^2 - (\sum y)^2\,)}}$$

مثال 1 : البيانات التالية تمثل الكمية المعروضة من سلعة معينة وسعر الوحدة الواحدة من هذه السلعة.
المطلوب : حساب معامل الارتباط البسيط بين الكمية المعروضة والسعر.

السعر x	الكمية المعروضة y	Xy	x^2	y^2
2	3	6	4	9
2	5	10	4	25
5	7	35	25	49

4	8	32	16	64
5	9	45	25	81
6	11	66	36	121
3	6	18	9	36
5	8	40	25	64
4	6	24	16	36
36	63	276	160	485

$$r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2 (n\sum y^2 - (\sum y)^2)}}$$

$$= r_p = \frac{9\,(276) - (36)(63)}{\sqrt{(9(160) - 36^2)\,(9\,(485) - 63^2)}}$$

$$= \frac{2484 - 2268}{\sqrt{(1440 - 1296)\,(4365 - 3969)}}$$

$$= \frac{216}{\sqrt{(144)\,(396)}}$$

$$= \frac{216}{\sqrt{(144)\,(396)}} = \frac{216}{\sqrt{57024}}$$

$$= \frac{216}{238.79} = 0.90$$

الارتباط طردي قوي

وبالإمكان تقسيم الارتباط الخطي إلى :

أ- الارتباط الخطي البسيط : وهو الارتباط بين الظاهرتين كما ذكرنا.

ب- الارتباط الجزئي : وهو الارتباط بين ظاهرتين بثبوت الظاهرة الثالثة.

ج- الارتباط المتعدد : وهو الارتباط بين ثلاث ظواهر.

من المعلوم أن مقياس الارتباط يختلف حسب اختلاف الظواهر المقاسة والتي يراد إيجاد العلاقة بينهما فإذا كانت لدينا الحالات التالية :

1- إذا كانت الظاهرتين مقاسة قياساً كمياً أي بالإمكان التعبير عنها بالأرقام نستخدم صيغة بيرسون للارتباط كما موضح سابقاً.

2- إذا كانت إحدى الظواهر المقاسة لا يمكن التعبير عنها كمياً مثل الحالة العلمية مع الدخل سوف نستخدم صيغة (سبيرمان أو كندال).

ثانياً : معامل ارتباط الرتب لسبيرمان Spearman Rank correlation coff. :

أفرض أن y , x متغيرين من النوع الوصفي وأفرض أن البيانات المستحصل عليها على أساس عينة عشوائية من المفردات قوامها n هي بهيئة صفات غير قابلة للقياس الكمي ويسمى الارتباط لهذه الحالة كما تكرر بارتباط الرتب والصيغة المستخدمة لإيجاد معامل الارتباط هي :

معامل ارتباط الرتب

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

حيث أن : r تمثل معامل الارتباط (ارتباط الرتب).

n : عدد القيم لكلا الظاهرتين.

$\sum_{i=1}^{n} d_i^2$: عبارة عن مجموع مربعات الفروق بين رتب الظاهرة الأولى ورتب الظاهرة الثانية المناظرة لها.

أما الخطوات الواجب اتباعها لتطبيق هذه الصيغة هي كما يلي :

1- نعطي قيم من الأعداد الطبيعية على قيم الظاهرتين مباشرة بترتيب تصاعدي.

2- نعطي القيم المتكررة معدل من الرتب التي تناظرها.

3- نرتب قيم الظاهرة الأولى بتسلسل الأعداد الطبيعية وتذكر الرتب التي تناظرها من رتب الظاهرة الثانية.

4- تحسب الفروق بين الرتب والرتب المناظرة فنحصل على $\sum di$ ثم نربع هذه الفروق ثم نجمعها والخطوة الأخيرة نطبق الصيغة المذكورة أعلاه.

ملاحظة : يمكن استخدام صيغة (سبيرمان) لإيجاد معامل ارتباط الرتب بين ظاهرتين غير مقاسة كمياً ، وذلك بذكر تسلسل الظاهرتين تصاعدياً وإعطاءهما رتب من الاعداد الطبيعية ثم نكمل الحل بإتباع الخطوات المذكورة أعلاه.

مثال 2 :

لديك البيانات التالية والمطلوب معرفة العلاقة بين الظاهرتين (y , x) وتفسير النتيجة ، علماً أن البيانات تمثل الحالة الانتاجية لأحد المصانع وكمية الإنتاج لإحدى السلع في هذا المصنع.

x_i	y_i	الرتب x_i والرتب المناظرة y_i	d_i	d_i^2
$^{3}76$	متوسط^{2}	3 2	1	1
$^{4}82$	جيد^{3}	4 3	1	1
$^{1}63$	ضعيف^{1}	1 1	0	0
$^{2}70$	جيد جداً^{4}	2 4	2	4
				$6 \rightarrow \sum_{i=1}^{n} d_i^2$

$$r_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6\,(6)}{4\,(16-1)} = 1 - \frac{36}{(4)(15)} = 1 - \frac{36}{60} = 1 - 0.60 = 0.40$$

مثال 3 : أجاب عشرة موظفين عن حالتهم العلمية ومقدار دخلهم الشهري بالدينار فكانت إجاباتهم كما في الجدول أدناه ، المطلوب إيجاد معامل ارتباط الرتب باستخدام صيغة (سبيرمان) بين الدخل الشهري والحالة العلمية وتفسير النتائج.

الحالة العلمية xi	الدخل الشهري y_i	الرتب x_iوالرتب المناظرة إلى y_i	d^i	d^{i2}
يقرأ ويكتب ^{4}	$^{7}43$	4 − 7	−3	9
متوسطة ^{7}	$^{6}39$	7 − 6	1	1
أمي$^{1.5}$	$^{2}11$	1.5 − 2	−0.5	0.25
يقرأ ويكتب ^{4}	$^{3}16$	4 − 3	1	1
عليا $^{9.5}$	$^{1}10$	9.5 − 1	8.5	2.25
عليا $^{9.5}$	$^{9}78$	9.5 − 9	0.5	0.25
متوسطة ^{7}	$^{10}87$	7 − 10	−3	9
يقرأ ويكتب ^{4}	$^{5}32$	4 − 5	−1	1
أمي$^{1.5}$	$^{4}19$	1.5 − 4	−2.5	6.25
متوسطة ^{7}	$^{8}50$	7 − 8	−1	1
المجموع				101

$$r = 1 - \frac{6\sum_{i=1}^{n} di^2}{n\,(n^2-1)} = 1 - \frac{6\,(101)}{10\,(10-1)} = 1 - \frac{606}{990} = 0.39$$

العلاقة ضعيفة بين الظاهرتين.

الانحدار The Regression:

يعتبر العالم الانكليزي Francis Galton (1822 – 1911) أول من استخدم مفهوم الانحدار في التطبيقات البيولوجية بهدف اكتشاف بعض العلاقات بين بعض المتغيرات البيولوجية.

يعرف الانحدار أو تحليل الانحدار بشكل عام بأنه مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات المعتمدة في العلاقة وغالباً ما تسمى العلاقات من هذا النوع بنماذج الانحدار.

ويختص الانحدار الخطي بتحليل العلاقة بين متغيرين أو أكثر يظهر كل منهما ب (أس مساو للواحد) في العلاقة المعتمدة للتحليل وتكون المعادلة بالشكل التالي :

$$y_i = a + bx_i + e_i \rightarrow \text{نموذج معادلة الانحدار}$$

حيث أن a , b معالم النموذج ، y يمثل المتغير المعتمد للنموذج ، علماً أن كل من x , y متغيرات من الدرجة الأولى (أس واحد) ولتقدير معادلة نموذج الانحدار أعلاه باستخدام أحد طرق التقدير نحصل على :

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

حيث أن $\hat{a},\hat{b}$ مقدرات المعالم a , b وتلفظ (a hat) ، (b hat) أما هذا الخطأ العشوائي e_i فأنه يكون مساوي إلى :

متغير الخطأ العشوائي $e_i = y_i - \hat{y}_i$

بتربيع المعادلة أعلاه ينتج :

$$e_i^2 = (y_i - \hat{y}_i)^2$$

بإدخال المجموع نحصل على :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

حيث ان:

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

$$\therefore \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 \qquad ,,,,,,,,,,(1)$$

الهدف هنا إيجاد مقدرات $\hat{a},\hat{b}$ التي تجعل $\sum_{i=1}^{n} e_i^2$ أقل ما يمكن ، وعليه فأن :

$\bar{y} = \hat{a} - \hat{b}\bar{x}$ ويسمى انحدار y / x

وبعد إجراء عدة اشتقاقات على المعادلة (1) نحصل على تقديرات $\hat{a},\hat{b}$ مساوية أي :

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{y} = \hat{a} + \hat{b}x$$

مثال 7 :

البيانات التالية تمثل الكمية المطلوبة (y) من سلعة معينة وسعر الوحدة الواحدة منها x .

المطلوب:

1- تقدير معادلة انحدار الكمية المعروضة على السعر ؟

2- تقدير الكمية المعروضة عند سعر 20؟

X	Y	X y	x^2
7	10	70	49
10	15	150	100
4	5	20	16
3	4	12	9
2	3	6	4
9	13	117	81
35	50	375	259

$$\bar{X} = \frac{35}{6} = 5.83$$

$\bar{Y} = \frac{50}{6} = 8.33$

$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{6(375)-(35)(50)}{6(259)-35^2} = \frac{2250-1750}{1554-1225} = \frac{500}{329} = 1.51$

$\hat{a} = \bar{y} - \hat{b}\bar{x}$

$= 8.33 - 1.51\ (5.83)$

$= 8.33 - 8.80 = -0.47$

معادلة الانحدار المقدرة هي : $\hat{y} = \hat{a} + \hat{b}x$

وتقدير الكمية المعروضة عند سعر (20) يكون عن طريق تعويض معادلة الانحدار المقدرة ($x = 20$)

$\hat{y} = -0.47 + 1.51\ (20)$

$= -0.47 + 30.2 = 27.73$