أهمية علم الإحصاء ومجالات تطبيقاته

1-علم الإحصاء

ان علم الإحصاء من العلوم القديمة والتي تطورت مع تطور البشرية. حيث ان أصل الإحصاء يمكن ان ينسب إلى الأزمنة السابقة عندما كان يعامل كنظام للعد والترقيم لأنشطة الدولة المختلفة. لقد اهتم البابليين واليونانيين والفراعنة بالإحصاء من أجل الحصول على معلومات حول عدد القادرين على حمل السلاح ومعلومات عن الانتاج الزراعي لغايات الضرائب. ويمكن تقسيم علم الإحصاء إلى قسمين:

1-الإحصاء الوصفي: ويتضمن هذا الفرع الطرق والأساليب المستخدمة في جمع البيانات والمعلومات لظاهرة معينة وكيفية تنظيم وتصنيف وتبويب هذه البيانات وعرضها في جداول ورسومات بيانية وحساب بعض المؤشرات الإحصائية.

2-الإحصاء الاستدلالي أو التحليلي: وهو العلم الذي يشتمل على الطرق الإحصائية التي تهدف ألي تحليل البيانات وعرضها بهدف الوصول إلى نتائج تفيد في اتخاذ القرارات ويتم على مرحلتين:

أ. التقدير ب. اختبار الفرضيات.

وهناك عوامل ساعدت إلى تطور علم الإحصاء عبر مرور حقب زمنية مختلفة ومن هذه العوامل:

1-الحاجة المتزايدة للبيانات الإحصائية ومن قبل مختلف العلوم.

2-قلة تكاليف الدراسات الإحصائية مقارنة بغيرها من الدراسات في العلوم الأخرى.

بذلك ومن خلال العرض السابق يمكن تعريف علم الإحصاء بالشكل التالي:

علم الإحصاء: هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهرة أو فرضية معينة وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

1)أهمية علم الإحصاء ومجالات تطبيقاته.

ان علم الإحصاء أحد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه وطرقه في جمع البيانات والمعلومات اللازمة للبحث العلمي وتحليل هذه البيانات والمعلومات بغية الوصول إلى النتائج التي يهدف لها الباحث.

كما وان للإحصاء دورا في وضع الخطط المستقبلية عن طريق التنبؤ بالنتائج ولكافة القطاعات سواء كانت انتاجية أم خدمية.

2) الطريقة الإحصائية في البحث العلمي.

لكي يمكننا استخدام الأسلوب الإحصائي في البحث العلمي يتوجب توفر بيانات ومعلومات عن الظاهرة أو الظواهر المطلوب دراساتها في ذلك البحث. وفيما يلي المراحل الرئيسية للطريقة الإحصائية في البحث العلمي:

1- تحديد مشكلة أو فرضية البحث أو الدراسة.

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

3- تصنيف البيانات وتبويبها وعرضها.
4- حساب المؤشرات الإحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.
5- تحليل معطيات الدراسة والتوصل للنتائج على ضوء فرضية أو فرضيات البحث أو الدراسة.
6- تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث.

3) جمع البيانات.

في أي بحث علمي يعتمد في تحليله بالطرق الإحصائية فانه يحتاج إلى بيانات ومعلومات حول موضوع البحث قيد الدراسة.
ويمكن للباحث الحصول على هذه البيانات والمعلومات من أحد المصدرين الآتيين:
أ. المصادر التاريخية: وهي عبارة عن سجلات أو جداول إحصائية مثال ذلك السجلات الموجودة في الدوائر والسجلات والتي قامت بجمعها وتنظيمها في جداول في وقت سابق.
ب. المصادر الميدانية: وهي جمع البيانات عن مفردات مجتمع الدراسة مباشرة وتتم هذه العملية عن طريق المقابلة أو البريد أو التلفون في حالة الأفراد، وفي حالة المقابلة فان الباحث يتمكن من شرح وتوضيح بعض الأسئلة الغامضة في استمارة البحث.
وهناك أسلوبان ممكن من خلالهما جمع البيانات والمعلومات آيا كان مصدرهما، وهذان الأسلوبان هما:
1-التسجيل الشامل.
ويقصد بهذا أسلوب جمع كافة المعلومات لكل مفردة من مفردات المجتمع وبذلك فانه سوف يكون على الباحث عدم ترك أي مفردة دون المرور بها، لكن من عيوب هذه الطريقة هي الجهد الكبير الذي سيبذله الباحث في الجمع والمرور بكل مفردة كذلك التكاليف العالية التي ممكن ان تقع على الباحث.
2-العينة.
هي عبارة عن عملية اخذ مجموعة معينة من مجتمع المراد الدراسة عليه بحيث ان هذه العينة تمثل المجتمع تمثيل دقيق بهدف الوصول إلى نتائج قابلة للتقييم واتخاذ القرار. ويمتاز أسلوب العينات بانه يحتاج إلى وقت وجهد وموارد مادية وبشرية اقل مما يحتاجه أسلوب التسجيل الشامل. أما ما يقصد بالمعاينة فهي أسلوب لاختيار مفردات من مجتمع الدراسة تؤلف العينة. وتنقسم العينات بشكل عام إلى قسمين رئيسيين وهما عينات عشوائية وعينات غير عشوائية.

أولا: العينات العشوائية.

يقصد بالعينة العشوائية بانها تلك المجموعة من المفردات التي تم اختيارها من مجتمع معين بحيث تكون الفرص متساوية في ظهور جميع المفردات أي ان ليس للباحث أي دخل في اختيار هذه المفردات. والعينات العشوائية على انواع عديدة أهمها:
1-المعاينة العشوائية البسيطة.
ويقصد بالمعاينة العشوائية البسيطة هو عملية اختيار لمفردات من مجتمع بحيث ان كل المفردات لها نفس الفرصة في الظهور بحيث يجب ان تكون هذه المفردات من نفس الجنس أي تمتلك نفس الصفات.
2-المعاينة الطبقية العشوائية.
يقصد بأسلوب العينة الطبقية هو تقسيم مجتمع الدراسة إلى طبقات متجانسة ومن ثم يتم اختيار عينة جزئية من كل قسم أو طبقة بحيث ان مجموع العينات الطبقية يكون مساوي إلى حجم العينة المطلوبة للدراسة.

مثال: أذا علمت ان في احد الدراسات البحثية تطلب الأمر اختيار عينة قوامها 110 وحدة من مصنع يحتوي على أربعة أقسام انتاجية حيث ان القسم الأول يحتوي 600 وحدة والقسم الثاني 800 وحدة والقسم الثالث

500 وحدة والقسم الرابع 300 وحدة . يطلب تحديد حجم العينات التي يتطلب اختيارها من كل قسم من الأقسام الأربعة.

الحل: واضح ان

$N=2200$ وان $n=110$

وان $N_1=600 \quad N_2=800 \quad N_3=500 \quad N_4=300$

$W_1=600/2200=6/22 \qquad n_1=6/22*110=30$

$W_2=800/2200=8/22 \qquad n_2=8/22*110=40$

$W_3=500/2200=5/22 \qquad n_3=5/22*110=25$

$W_4=300/2200=3/22 \qquad n_4=3/00*110=15$

حيث ان

$n = n_1 + n_2 + n_3 + n_4$

$30+40+25+15=11$

3-المعاينة العشوائية المنتظمة.
يتم اختيار المفردة الأولى بطريقة الاختيار العشوائي وباقي المفردات يتم اختيارها على النحو التالي:

نوجد كسر المعاينة = $\frac{\text{حجم المجتمع}}{\text{حجم العينة}}$

ثم نقسم مفردات المجتمع إلى مجموعات عددها يساوي حجم العينة وعدد المفردات داخل كل مجموعة يساوي كسر المعاينة.
وبإضافة كسر المعاينة بالتتابع إلى رقم أول مفردة ثم التالية لها إلى ان نحصل على العدد المطلوب.
4-المعاينة المتعددة المراحل.
ويقصد بها تقسيم المجتمع الإحصائي إلى وحدات تدعى بالوحدات الأولية ثم يتم اختيار عينة عشوائية من هذه الوحدات الأولية كمرحلة أولى ثم يتم تقسيم كل وحدة أولية إلى وحدات اصغر تدعى بالوحدات الثانوية ومن ثم يتم اختيار عينة عشوائية من الوحدات الثانوية وهكذا تستمر العملية لحين الوصول إلى المفردات التي يتم جمع البيانات منها والتي تؤلف عينة البحث.

ثانيا: العينات غير العشوائية.

يقصد بالعينات غير العشوائية هي تلك المجموعة من المفردات او الوحدات التي يقوم الباحث باختيارها ضمن اعتبارات معينة تتعلق بطبيعة الدراسة او الباحث وبذلك سميت بالعينات غير العشوائية ومن هذه العينات.
1-المعاينة الحصصية.
يقصد بالمعاينة الحصصية هي عملية تقسيم مجتمع الدراسة إلى عدة طبقات استنادا لمعايير تقسيم معينة تتعلق بطبيعة الدراسة، ثم يتم اختيار عدد من المفردات من كل طبقة وبشكل شخصي من قبل الباحث بحيث ان أجمالي عدد المفردات لهذه العينات يشكل حجم العينة المطلوبة لتلك الدراسة.
2-المعاينة العمدية.

وهو أسلوب لاختيار عينة من مجتمع بشكل متعمد نعتقد مسبقا ان مفردات هذه العينة هي خير من يمثل مجتمع الدراسة.

وسائل جمع البيانات.

بعد ان حددنا حجم العينة وأسلوب العينة الملائم في اختيار مفردات المجتمع بعد ذلك يتم اختيار الوسيلة أو الطريقة الملائمة في جمع البيانات عن الظاهرة المتعلقة بالدراسة. وهنالك وسائل عديدة في جمع البيانات أهمها:

1-أسلوب الجمع المباشر:

ويقصد بهذا الأسلوب هو جمع البيانات المسجلة في سجلات مثال ذلك البيانات او المعلومات المسجلة في أجهزة الدولة أو هيئات معينة ذات علاقة بالدراسة.

2-الاستبيان:

عبارة عن استمارة يتم من خلالهما جمع البيانات والمعلومات من مفردات مجتمع الدراسة وذلك عن طريق مواجهة الباحث الشخصية للمفردة الإحصائية أو عن طريق المراسلة، كما هي الحال في التعدادات السكانية مثلا حيث يتم مواجهة رب الأسرة أو من ينوب عنه لغرض ملئ استمارة التعداد بالبيانات والمعلومات اللازمة.

وهناك أمور كثيرة يجب مراعاتها عند تصميم هذه الاستمارة منها:

1-شمول الأسئلة لكي تغطي الجوانب المختلفة للباحث.

2-سهولة ووضوح هذه الأسئلة وتسلسلها المنطقي.

3-ان تكون الإجابات رقمية للمتغيرات الكمية وان تحدد الاستمارة الأوجه المختلفة للمغيرات النوعية.

4-ان تكون الأسئلة التي تتضمنها استمارة البحث غير محرجة مع ضرورة التأكيد على سرية البيانات المعطاة.

4) تصنيف وتبويب البيانات.

بعد جمع البيانات بالأساليب التي تم إيضاحها في المحاضرات السابقة تكون هذه البيانات خام أي غير منطقية و لا يمكن الاستفادة منها لذلك فان أولى الخطوات الهامة بعد عملية جمع البيانات هي عملية مراجعة وتصنيف وتبويب البيانات.

سوف بتم توضيح ما ذكرى أعلاه في الخطوات التالية:

1-مراجعة البيانات: بعد جمع البيانات وفق أسلوب معين يقوم الباحث بمراجعة هذه البيانات والتأكد منها.

2-تصنيف البيانات: بعد التأكد من عملية جمع البيانات ومراجعة هذه البيانات يتم عملية تصنيف هذه البيانات وفق ظواهر معينة تعتمد على نوع الدراسة فقد يكون التصنيف على أساس ظاهرة العمر آو ظاهرة الجنس أو غيرها .

3-تبويب البيانات: بعد تصنيف البيانات نقوم بعملية تبويب هذه البيانات في جداول خاصة بحيث ان كل جزء من البيانات المصنفة عن الظاهرة المعينة يعود إلى مستوى معين لتلك الظاهرة.

وهناك أربع انواع من التبويب:

1-التبويب الزمني: وهو عبارة عن تجميع البيانات المصنفة وترتيبه في جداول على أساس ان كل جمع منها يعود لوحدة زمنية كاليوم، الأسبوع، الشهر، السنة.

2-التبويب الجغرافي: وهو عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على أساس ان كل جمع منها يعود بوحدة جغرافية معينة أو تقسيم أداري معين كالنواحي، الا قضية، المحافظات، البلدان.
3-التبويب الكمي: وهو عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على أساس ان كل جمع منها خاص بوحدة كمية معينة كوحدات الوزن، الطول، المسافة، الحجم.
4-التبويب على أساس صفة معينة: وهو عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على أساس ان كل جمع منها يشترك بصفة معينة كالجنس، الحالة الاجتماعية، عنوان السكن.

5)الأخطاء الشائعة في جمع البيانات.

أثناء جمع البيانات فان الباحث يقع في بعض الأخطاء، هذه الأخطاء تحدث نتيجة سوء استخدام الطريقة الإحصائية، وفيما يلي عرض موجز لأهم هذه الأخطاء.

1- خطأ التحيز: في حالة تحديد أسلوب الجمع ووسيلة الجمع يجب ان تتم هذه العملية من مصادر البيانات الأصلية. إلا انه يحصل في بعض الأحيان ان تتم هذه العملية من مصادر أخرى غير المصادر الأصلية.
2- خطأ الصدفة: في بعض الأحيان يقع الباحث نفسه في أخطاء معينة وهذه الأخطاء تسمى بالصدفة كان يقوم بالاستيفاء لبعض البيانات والمعلومات بالاعتماد على معلوماته الشخصية او التعمد في جمع هذه البيانات من المفردات دون الأخرى المحددة أو ان يقوم بجمع بيانات ناقصة لسبب أو لأخر.

مثال / عن المعاينة العشوائية المنتظمة

في امتحان لطلبة صف معين عددهم 24 طالب رتبت اسمائهم حسب تسلسل درجاتهم تنازليا وبهدف التعرف على اسباب انخفاض مستواهم في الامتحان تطلب الامر استقراء رأي ستة طلاب منهم. يطلب تحديد تسلسل هؤلاء الطلبة وبشكل عشوائي

الحل/

من معطيات السؤال ان حجم العينة n=6 اي سيتم تقسيم البيانات الى ست مجاميع , وباحتساب كسر المعاينة $k = \frac{N}{n} = \frac{24}{6} = 4$ حيث ان N تمثل حجم المجتمع population size وان n تمثل حجم العينة . sample size وهذا يعني ان كل مجموعه تحتوي على 4 افراد حسب التسلسل كالاتي

2 3 4 , 5 6 7 8 ,...., 21 22 23 24

وبعدها بشكل عشوائي نختار مفردة من المجموعة الاولى ولنفرض انه الطالب الذي يحمل التسلسل رقم 3 وعلى اساسه يتم تحديد بقية تسلسلات مفردات العينة من خلال اضافة العدد k=4 وهي التسلسلات

7,11,15,19,20 . وبذلك فان العينة المختارة من هذا الصف تمثل الطلبة الذين تسلسلهم 3,7,11,15,19,20 والذين يتم استقراء رأيهم حول اسباب انخفاض مستواهم.

التوزيعات التكرارية وأساليب عرض البيانات

بعض التعاريف والمفاهيم الاساسية المطلوبة

1- المتغيرات العشوائية وانواعها Random variables .

يعرف المتغير العشوائي بانه دالة ذات قيمة حقيقية معرفة على فضاء يدعى فضاء العينة وغالبا ما يرمز للمتغير العشوائي بأحد الأحرف الكبيرة مثال X،Y،Z وتنقسم المتغيرات العشوائية إلى قسمين وهما:

1-المتغيرات النوعية (الوصفية) Qualitative Variables.

وهي المتغيرات التي لا يمكن قياسها بوحدات قياس محددة وانما تظهر على شكل صفات لذلك المتغير مثل لون الشعر أو الحالة الاجتماعية (أعزب، متزوج، مطلق، أرمل) أو الجنس (ذكر أو انثى).

2-المتغيرات الكمية Quantitative variables:

وهي المتغيرات التي يمكن قياسها بوحدات معينة مثل الطول، الوزن، المسافة او عدد الطلاب في صف معين ...الخ وهي على نوعين:

أ -المتغيرات المتقطعة discrete variable :

إذا كان مجموعة القيم الممكنة للمتغير x مجموعة قابلة للعد سواء كانت مجموعة محددة ام غير محددة عندئذ يقال ان x متغير عشوائي متقطع.

مثال ان مجموعة القيم الممكنة إلى x في تجربة رمي الزهر هي المجموعة $\Omega=[X:x=1,2,3,4,5,6]$ وحيث انه من الممكن عد عناصر هذه المجموعة (أي انها مجموعة قابلة للعد) بالرغم من كونها محدودة (أي لها بداية العدد 1 ونهاية العدد 6) عليه فان x متغير عشوائي متقطع. وأمثلة أخرى على المتغيرات المتقطعة هي: عدد أفراد الأسرة، عدد الأقسام في مصنع، الجنس ذكر وانثى، عدد الطلاب، عدد النداءات الهاتفية.

ب. المتغيرات المستمرة continuous variable:

إذا كانت مجموعة القيم الممكنة للمتغير x مجموعة غير قابلة للعد سواء كانت مجموعة محددة أم غير محددة يقال ان x متغير عشوائي مستمر.

مثال: افرض ان x متغير عشوائي يشير إلى الزمن المستغرق لقطع المسافة بين بغداد ونينوى 400كم /ساعة. واضح وفق قانون السرعة ان الزمن المستغرق لقطع هذه المسافة يتراوح ما بين 4 ساعات إلى 5 ساعات .عليه فان مجموعة القيم الممكنة إلى x هي مجموعة الأعداد الحقيقية

$\Omega = [x:4 < x < 5]$ وحيث انه لا يمكن عد عناصر هذه المجموعة كونها واقعة ضمن فئة مستمرة أي وجود عدد غير منتهي من القيم الواقعة ضمن هذه الفترة على الرغم من كونها مجموعة محددة (لها بداية ونهاية) فأذن x متغير عشوائي مستمر.

أمثلة على المتغيرات المستمرة: الوزن، درجة الحرارة، الزمن، الطول، العمر، الأجر، المبيعات، كميات الانتاج.

الثوابت: فهي السمات والخواص التي لا تتغير وهي تصف ماهية المواد في ظروف معينة مثل الكثافة النوعية لعنصر ما في ظرف محدد فمثلا الكثافة النوعية للماء النقي في درجة الحرارة العادية.

معلمة المجتمع parameter of population: هو الثابت الذي يصف المجتمع وهو عبارة عن مقياس سمة مثل معدل المجتمع.

2-العرض الجدولي للبيانات:

التوزيعات التكرارية:

التوزيع التكراري: هو عبارة عن توزيع البيانات المأخوذة عن ظاهرة معينة على الفئات بحيث تقع كل مفردة في فئة واحدة فقط، والمفردات التي تقع في فئة واحدة تكون متجانسة ثم نضعها في جدول يسمى جدول توزيع تكراري.

او هو عبارة عن ترتيب البيانات التي جمعت وصنفت في جداول بعد تقسيمها إلى عدد من المجاميع والتي تسمى بالفئات، هذه الفئات قد تكون مرتبة تصاعديا أو تنازليا حسب طبيعة البيانات. وان هذا التوزيع للقيم ل x يسمى بالتوزيع التكراري. وقد تكون عدد الفئات للتوزيع التكراري متساوية في الطول أو غير متساوية في الطول وذلك يعتمد على طبيعة البيانات.

بناء جداول التوزيعات التكرارية:

تقسم جداول التوزيع التكراري الى نوعين وهما :

1- جداول التوزيع التكراري للبيانات الوصفية (الأسمية) Qualitative frequency distribution

يحوي الجدول التكراري للبيانات الوصفية عمودين، يتضمن الأول قائمة الفئات، وهي مجموعة كل الحالات(الصفات)التي تكون البيانات، ويتضمن العمود الثاني عمود التكرارات الذي يمثل عدد عناصر العينة لكل حالة(تكرارات تلك الحالة في العينة.)

مثال :

البيانات التالية توضح نتائج اختبار فحص فصيلة الدم لـ25 شخص المطلوب كون جدول توزيع تكراري

[A, B ,B, AB, O,O,O,B ,AB, B, B, B, O, A, O, A, O, O, O, AB, AB, A, O, B, A]

classes الفئات	التكرار بالإشارة	Frequency التكرارات
A	//// /	5
B	//// ///	7
O	////	4
AB	//// //// /	9

2–جدول التوزيع التكراري للبيانات الكمية

1- أذا كان مدى البيانات صغيرا فانه يمكننا بناء الجدول التكراري بترتيب البيانات ترتيبا تصاعديا حتى نصل إلى أعلى قيمة وهذا يمثل العمود الأول في الجدول التكراري . أما العمود الثاني فيمثل عدد المرات التي تكررت فيها كل مفردة.(كما في الحالة السابقه)

مثال: تبين البيانات التالية درجات 25 طالبا في امتحان معين وان درجة الامتحان من 10. وكما يلي:

6، 5، 8، 7، 5 , 8، 9، 8، 6، 8 , 5، 10، 7، 9، 7 , 7، 3، 9، 8،10, 3، 7، 7، 5، 6

المطلوب: عرض هذه البيانات في توزيع تكراري.

الحل: نحدد المدى Range=أعلى قيمة – اقل قيمة

المدى= 10-3 =8 نلاحظ ان قيمة المدى صغيرة

نبدأ من اقل قيمة ثم نرتب القيم تصاعديا إلى ان نصل إلى أعلى قيمة

الدرجة	التكرار	التكرار بالاشارة
3	2	//
4	0	
5	4	////
6	3	///
7	6	//// //
8	5	//// /
9	3	///
10	2	//

2- في حالة كون مدى البيانات كبير عندها سيتم تقسيم البيانات الى مجاميع تسمى الفئات يكون طولها اكثر من قيمة واحدة.

ويمكن توضيح عمل جدول توزيع تكراري بالخطوات التالية

1- المدى الكلي للتوزيع: total range

إذا كان لدينا على سبيل المثال $x_1, x_2, \ldots, x_n$ تمثل بيانات المتغير العشوائي X من عينة عشوائية من المفردات قوامها n مفردة ونرغب في تلخيص هذه البيانات في توزيع تكراري عدد فئاته هو m.

لنفرض ان X_S تمثل أصغر قيمة وX_L تمثل أكبر قيمة في مجموعة البيانات هذه عندئذ نقوم بإيجاد المدى الكلي TR والذي يعرف بانه الفرق بين أعلى قيمة واقل قيمة أي ان:

$$TR = x_l - x_s$$

2-عدد فئات التوزيع number of classes

نقصد بعدد فئات التوزيع هي عدد المجاميع التي يتألف منها التوزيع التكراري وهنالك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع أهمها:

1- صيغة YULE وهي:

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

وحيث ان n عدد المشاهدات.

2- صيغة سترجس sturges وهي:

M=1+3.322 LOG10 n

وعند التطبيق يتم تقريب النتائج لأقرب عدد صحيح.

3- طول الفئة length of class.

ويمثل مقدار سعة الفئة أي مقدار المسافة بين الحد الأدنى للفئة وحدها الأعلى، ويتناسب طول الفئة عكسيا مع عدد فئات التوزيع. ويرمز لطول الفئة بالرمز L. ويمكن تحديد قيمة L(في حالة التوزيعات ذات الفئات المتساوية في الطول) من خلال الصيغة التالية

$$L = \frac{TR}{m}$$

حيث ان TR : تمثل المدى , m : تمثل عدد الفئات

بطريقة أخرى

طول الفئة=الحد الأعلى - الحد الأدنى

طول الفئة = الفرق بين الحدين الأدنى أو الأعلى لفئتين متتاليتين

طول الفئة = الفرق بين مركزي فئتين متتاليتين.

4- تحديد الحد الأدنى والحد الأعلى للفئة: Lower and upper bound of class

لكل فئة في جدول التوزيع التكراري لها حدين أعلى وأدنى، وفي حالة تساوي أطوال الفئات فانه يمكن تكوين حدود الفئات على النحو التالي:

هذا الجدول خاص **بالمتغيرات المتقطعة**

تسلسل الفئة	الحد الأدنى للفئة	الحد الأعلى للفئة
1	Xs	Xs+L-1
2	Xs+L	Xs+2L-1
3	Xs+2L	Xs+3L-1
.	.	.
M	Xs+(m-1)L	Xs+ML-1

بطريقة اخرى/ فان الفئة الاولى وحدودها هي $\left(x_s-\left(x_s+l-1\right)\right)$ فان الحد الادنى لها هو اصغر قيمة في البيانات اما الحد الاعلى لها فهو (الحد الادنى + طول الفئة – 1) , اما بالنسبة الى الفئة الثانيه فان الحد الادنى لها يمكن ايجاده باضافة طول الفئة الى الحد الادنى للفئة الاولى , وهكذا للفئة الثالثه فان الحد الادنى لها هو (الحد الادنى للفئة الثانية + طول الفئة).

وكذلك بالنسبة للحد الاعلى للفئات بعد حساب الحد الاعلى للفئة الاولى ممكن حساب حدود الفئات باضافة طول الفئة الى حد الاعلى للفئة السابقة

5- تحديد مركز الفئة center of class

ان مركز الفئة يمثل قيمة من قيم المتغير العشوائي X التي تتوسط المسافة بين الحد الأدنى والحد الأعلى للفئة، ويرمز بالرمز X لمركز الفئة حيث ان يمكن حساب مركز الفئة من خلال الصيغة التالية:

$$X=\frac{L.L+U.L}{2}$$

حيث ان L.L تمثل الحد الأدنى للفئة

وانU.L تمثل الحد الأعلى للفئة

6- تكرار الفئة Class frequency

يرمز لتكرار الفئة بالرمز f_i والذي يمثل جزء من مفردات العينة التي تتصف بكونها تقع من حيث القيمة العددية ما بين حدي الفئة **بحيث ان مجموع هذه الأجزاء يشكل عدد مفردات العينة n**.

مع ملاحظة انه ليس من الضروري عند التطبيق ان يكون الحد الأدنى للفئة الأولى مساو تماما لأصغر قيمة في المجموعة بل قد يكون اقل منها لاعتبارات تتعلق بتسهيل العمليات الحسابية اللاحقة.

بالإضافة إلى ذلك يمكن ان تكون الجداول التكرارية مغلقة أو مفتوحة والمقصود بالمغلقة انه لها حد أدنى وحد أعلى أم المفتوحة فلها حد أدنى للفئة الأولى وليس لها حد أعلى للفئة الأخيرة.

وبشكل عام يفضل ان يمتاز التوزيع التكراري بما يلي:

1-يجب ان تكون الفئات منفصلة عن بعضها البعض (غير متداخلة فيما بينها).

2-ان تكون الفئات متساوية في الطول كلما أمكن ذلك.

3-يجب ان تكون الفئات كافية لاحتواء جميع البيانات .

4-ان يكون التوزيع التكراري توزيع مغلق وفي ذلك أهمية كبيرة في تسهيل عملية حساب بعض المؤشرات الإحصائية.

5-ان تبدأ فئات التوزيع وتنتهي (حدود الفئات) بأعداد صحيحة لما في ذلك من أثر في تسهيل العمليات الحسابية.

6-ان لا يقل عدد فئات التوزيع عن خمسة ولا يزيد عن خمسة عشر.

مثال:

البيانات التالية تمثل درجات (40) طالباً في امتحان مادة الإحصاء حيث كانت درجة النجاح 50%

25 35 22 33 47 50 45 42 21 15 17 41 19 27 29 31 32 37 43 27

13 10 26 25 38 31 40 39 30 20 21 22 27 28 29 39 33 38 37 27

المطلوب: تكوين جدول توزيع تكراري .

الحل:

1- إيجاد المدى الكلي:

$$T.R.=X_L-X_S$$

$$=50-10=40$$

2- تحديد عدد الفئات ويفضل ان لا يقل عن خمسة ولا يزيد عن خمسة عشر ويتوقف ذلك على حجم البيانات وسنتبع طريقة سترجس في حساب عدد الفئات وكما يلي:

$$M=1+3.322\log(n)$$

$$=1+3.322\log(40)$$

$$=1+3.322(1.6) = 1+5.28 = 6.28 \approx 6$$

3- ايجاد طول الفئة

$$\frac{T.R}{m} = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{40}{6} = 6.66 \approx 7$$

4- كتابة الحدود

الحد الأدنى الأول $X_S=10$

الحد الأعلى الأول $X_S+L-1 = 10+7-1 = 16$

الحد الادنى الثاني $X_S+L = 10+7 = 17$

الحد الأعلى الثاني $X_S+2L-1 = 10+2(7) -1 = 23$

او الحد الاعلى للفئة الاولى +طول الفئه (16 + 7=23)

وبنفس الطريقة نكمل باقي الفئات وستكون الفئات كما يلي:

الفئات	مركز الفئة	تفريغ التكرارات	التكرارات
16-10	13	///	3
23-17	20	// /////	7

11	/ ///// /////	27	30-24
9	//// /////	34	37-31
7	// /////	41	44-38
3	///	48	51-45

الحد الاعلى والادنى للفئات فى حالة المتغيرات المستمرة

يمكن ايجاد حدود الفئات للمتغيرات المستمرة باستخدام الجدول التالي

تسلسل الفئة	الحد الأدنى	الحد الأعلى
1	Xs	Xs+L
2	Xs+L	Xs+2L
3	Xs+2L	Xs+3L
.	.	.
m	Xs+(m-1)L	Xs+mL

طرق مختلفة لكتابة الفئات:

توجد عدة طرق لكتابة حدود الفئات منها:

الطريقة (1)	الطريقة (2)
19-10	20-10
29-20	30-20
39-30	40-30

الطريقة (3) (متغير مستمر)	الطريقة(4) ويمكن ان تكتب بشكل مختصر(للمتغيرات المستمرة)
10 إلى اقل من 20	10-
20 إلى اقل من 30	20-
30 إلى اقل من40	30-
40 الى اقل من 50	40-

سيتم استخدام الطريق الاولى للمتغيرات المتقطعة والطريقة الرابعة للبيانات المستمرة

الجداول المفتوحة:

1-الجدول المفتوح من الطرفين، مثال ذلك: (اي لا يمتلك حد ادنى للفئه الاولى ولا يمتلك حد اعلى للفئة الاخيرة)

اقل من-10

20-

30-

40- فأكثر

2-جدول مفتوح من طرف الحد الأدنى للفئة الأولى: اي (اي لا يمتلك حد ادنى للفئه الاولى)

اقل من-10

20-

30-

40-50

3- جدول مفتوح من طرف الحد الأعلى للفئة الأخيرة: (ولا يمتلك حد اعلى للفئة الاخيرة)

10-

20-

30-

40- فأكثر

مثال: عن المتغير المستمر

البيانات التالية تمثل اوزان (50) طالبا في احدى المدارس والمطلوب انشاء جدول توزيع تكراري؟

57 43 46 24 44 38 19.33 54 49 57 29.41 53 45 47 41 37 49 56 47 29

31 32 51 52 42 45 28 43 49 34 22.32 42 45 28 39 21.5 18 37 34 29

23 35.32 43 39 41 26.6 27 32.8 37 28 .

ملاحظة / البيانات العشرية تعتبر تابعه للمتغيرات المستمرة

الحل/

1- إيجاد المدى الكلي

$$T.R = X_L - X_S = 57-18=39$$

2- تحديد عدد الفئات حسب طريقة سترجس

$$M=1+3.322\log(n)$$

$$=1+3.322\log(50)$$

$$=1+3.322(1.69) = 1+5.614 = 6.614 \approx 7$$

3- تحديد طول الفئة

$$L = \frac{T.R}{m} = \frac{39}{7} = 5.57 \approx 6$$

4- نحدد الحدود الدنيا والعليا للفئات، وبما ان المتغير من النوع المستمر بذلك سوف نستخدم الجدول الثاني الخاص بالمتغيرات المستمرة لتحديد حدود الفئات:

الحد الأدنى الأول $X_S = 18$

الحد الأعلى الأول $X_S+L = 18+6 = 24$

الحد الأدنى الثاني $X_S+L = 18+6 = 24$

الحد الأعلى الثاني $X_S+2L = 18+2*6 = 30$,

او باستخدام الطريقة الثانية (الحد الاعلى للفئة الاولى +طول الفئة 24+6=30)

الحد الأدنى الثالث $X_S+2L = 18+2*6 = 30$, (الحد الادنى للفئة الثانية +طول الفئة 24+6=30)

الحد الأعلى الثالث $X_S+3L = 18+18 = 36$

الحد الأدنى الرابع $X_S+3L = 18+18 = 36$

الحد الأعلى الرابع $X_S+4L = 18+24 = 42$

وهكذا لبقية الحدود

وبذلك سيكون الجدول التكراري بالشكل التالي

الفئات	تفريغ التكرارات	f_i	طريقة ثانية للفئات	مركز الفئة
24-18	I I I I I	5	-18	21
30-24	I I I I I I I I I	9	-24	27
36-30	I I I I I I	6	-30	33
42-36	I I I I I I I I	8	-36	39
48-42	I I I I I I I I I I I I	12	-42	45
52-48	I I I I I I	6	-48	51
60-54	I I I I	4	-54	57

نلاحظ في هذا الجدول ان الحد الادنى للفئة الثانية يبدأ بنفس قيمة الحد الاعلى للفئة الاولى وهكذا بالنسبة لبقية الفئات وذلك لان المتغيرات المستمرة تحتوي على قيم عشرية وفي هذه الحالة ستحتوي الفئة الاولى على الارقام **(من 18 الى اقل من 24**) , اما الفئة الثانية ستحتوي على الارقام **(من 24 الى اقل من 30)**

ملاحظة : اذا كان الحد الادنى في البيانات رقم عشري نستطيع لتسهيل الحسابات ان تكون الحد الادنى للفئة الاولى قيمه صحيحه اقل من القيمه الصغرى للبيانات وفي هذه الحاله ستكون حدود الفئات اعداد صحيحة للسهولة في الحسابات

التوزيع التكراري النسبي *proportionate frequency distribution*:

ان التوزيع التكراري النسبي هو توزيع تكراري اعتيادي فيه التكرارات معبَر عنها بنسب مئوية يمكن الحصول عليها من قسمة تكرار كل فئة على مجموع التكرارات الكلية.

فعلى فرض انه لدينا توزيع تكراري عدد فئاته (M) وان التكرارات المقابلة لها هي $(f_1, f_2, \ldots, f_m)$ فان التكرار النسبي المقابل للفئات هو $f_1^*, f_2^*, f_3^* \ldots f_m^*$

حيث ان:
$$f_i^* = \frac{f_i}{\sum f_i}$$

وان $\sum f_i$ يمثل مجموع التكرارات الكلي والذي يساوي حجم العينه n

وإذا ما ضربنا النتيجة ب 100 يتكون لدينا التكرار النسبي المئوي وكما يلي:

$$Pf_i^* = \frac{f_i}{\sum f_i} * 100$$

مثال:

الاتي توزيع تكراري لأطوال (90) نبتة، يطلب تكوين التوزيع التكراري النسبي المئوي:

الفئات	-10	-20	-30	-40	-50	-60	-70	90-80
f_i	2	3	5	10	18	25	15	12
f_i^*	0.0222	0.0333	0.0555	0.1111	0.2	0.277	0.1666	0.1333
التكرار النسبي المئوي	2.22	3.33	5.55	11.11	20	0.277	16.66	13.33

التكرار النسبي الاول هو $f_1^* = \frac{2}{90} = 0.0222$ وهكذا لباقي الفئات

اما التكرار النسبي المئوي الاول هو $pf_1^* = f_1^* \times 100$ وهكذا لباقي الفئات

التوزيع التكراري المتجمع: cumulative frequency distribution

وهو عبارة عن توزيع يبين فيه كمية التكرار المتجمع عند قيمة معينة من قيم المتغير العشوائي . وهذا التوزيع على نوعين هما:

1- التوزيع التكراري المتجمع الصاعد:

وهو التوزيع الذي يبين تراكم التكرارات ابتداءً من الفئة الأولى في التوزيع وانتهاءً بالفئة الأخيرة منه، ويتم حساب التكرارات المتجمعة على أساس الحدود العليا للفئات. ويرمز للتكرار المتجمع الصاعد بـ F_i .

مثال: الجدول التالي يبين توزيع تكراري حيث يمثل توزيع (60) عائلة فلاحية حسب ملكيتها من عدد أشجار البرتقال، يطلب تكوين جدول التوزيع التكراري المتجمع الصاعد.

الفئات عدد الاشجار	الحدود العليا للفئات	f_i عدد العوائل	الصاعد F_i	بالرموز F_i	*طريقه ثانية*
74-60	اقل من او يساوي 74	4	4	$F_1=f_1$	F1=f1
89-75	اقل من او يساوي 89	5	9	$F_2=f_1+f_2$	F2= F1+f2
104-90	اقل من او يساوي 104	10	19	$F_3=f_1+f_2+f_3$	F3= F2+f3
119-105	اقل من او يساوي 119	12	31	I	F4= F3+f4
134-120	اقل من او يساوي 134	16	47	I	*I*
149-135	اقل من او يساوي 149	7	54	I	*I*
164-150	اقل من او يساوي 164	6	60	$F_6=f_1+f_2+\text{---}+f_6$	F6=F5+ f6
		60			

من الجدول اعلاه نستطيع القول ان عدد العوائل التي تمتلك 104 شجرة فأقل هي 19 عائله , وان عدد العوائل التي تمتلك 149 شجرة فاقل هي 54 عائلة وهكذا التفسير لباقي القيم .

ويمكن تحويل أي توزيع تكراري متجمع صاعد الى:

1- توزيع تكراري متجمع صاعد نسبي وذلك من خلال قسمة التكرارات المتجمعة على مجموع التكرارات الكلية أي (n) فاذا رمزنا للتكرار المتجمع الصاعد النسبي بالرمز F_i^* فان:

$$F_i^* = \frac{F_i}{\sum f_i} = \frac{F_i}{n}$$

2- التكرار المتجمع الصاعد النسبي المئوي:

$$PF_i^* = \frac{F_i}{n} * 100 = F_i^* * 100$$

وبالتطبيق على المثال اعلاه فان التكرار المتجمع الصاعد النسبي والصاعد النسبي المئوي موضح بالجدول التالي

الصاعد النسبي F_i^*	الصاعد النسبي المئوي PF_i^*
0.06	6
0.15	15
0.316	31.6
0.5166	51.66
0.783	78.3
0.9	90
1	100

اما في حالة كون المتغير العشوائي من النوع المستمر فانه كما في المثال التالي :

مثال: التالي توزيع تكراري **لأوزان** عينة من طلبة احدى الكليات حجمها (100) طالب والمطلوب تكوين جدول التوزيع التكراري المتجمع الصاعد، المتجمع الصاعد النسبي والمتجمع الصاعد النسبي المئوي .

الحل/

الفئات	الحدود العليا	f_i	F_i الصاعد	F_i^* الصاعد	PF_i^* الصاعد
46-	اقل من 53	7	7	0.07	7
53-	اقل من 60	15	22	0.22	22
60-	اقل من 67	27	49	0.49	49
67-	اقل من 74	21	70	0.7	70
74-	اقل من 81	14	84	0.84	84
81-	اقل من 88	8	92	0.92	92
88-	اقل من 95	5	97	0.97	97
95-102	اقل من 102	3	100	1.00	100

2-التوزيع التكراري المتجمع النازل:

وهو التوزيع الذي يبين تناقص التكرارات ابتداءً بالفئة الأولى بالتوزيع وانتهاءً بالفئة الأخيرة منه . ويتم حساب

جدول توزيع التكرارات المتجمعة على أساس الحدود الدنيا للفئات .

مثال: الاتي تكراري يمثل توزيع (60) عائلة فلاحية حسب ملكيتها من عدد أشجار البرتقال، يطلب تكوين التوزيع التكراري المتجمع النازل النسبي .

التكرار النازل بالرموز	النازل $PF_i^{/*}$	النازل $F_i^{/*}$	النازل $F_i^{/}$	f_i	الحدود الدنيا	الفئات
$F_1^{/} = n$	100	1.00	60	4	اكبر من او يساوي 60	**74-60**
$F_2^{/} = n - f_1$	93.33	0.933	56	5	اكبر من او يساوي 75	**89-75**
$F_3^{/} = n - f_1 - f_2$	85	0.85	51	10	اكبر من او يساوي 90	**104-90**
I	68.3	0.683	41	12	اكبر من او يساوي 105	**119-105**
I	48.3	0.483	29	16	اكبر من او يساوي 120	**134-120**
I	21.67	0.216	13	7	اكبر من او يساوي 135	**149-135**
$F_m^{/} = n - f_1 - \cdots - f_m$	10	0.10	6	6	اكبر من او يساوي 150	**164-150**

العرض الهندسي للبيانات

من اجل إعطاء توضيح واسع او صورة موسعة عن البيانات المبوبة في جداول تكرارية فانه يتم عرض هذه البيانات بهيئة رسوم بيانية واشكال هندسية متعددة الاشكال والتصاميم والبعض منها بهيئة رسوم تصويرية، ومن هذه الاشكال:

1- الدائرة البيانية:

وهي عبارة عن شكل هندسي , حيث يتم تقسيم اصناف البيانات الى قطاعات داخل الدائرة , كل قطاع يمثل نسبة مجموعة معينة من البيانات الكلية، ولتحديد كل قطاع فانه يتوجب تحديد زاوية كل منها كما يلي

$$\text{زاوية القطاع} = \frac{\text{عدد البيانات لكل صنف}}{\text{مجموع البيانات الكلية}} * 360$$

<u>مثال:</u> بلغ عدد الطلبة في احدى الكليات 2000 طالب وطالبة، 800 في المرحلة الأول، 500 في المرحلة الثانية، 400 في الثالثة و 300 في الرابعة. يطلب تمثيل البيانات باستخدام الدائرة البيانية

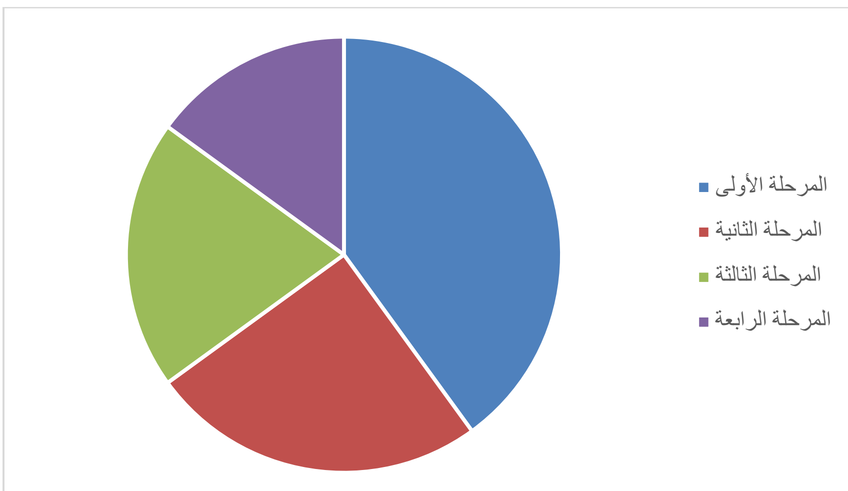
الحل:

زاوية القطاع للمرحلة الاولى= $\frac{800}{2000} * 360 = 144$

زاوية القطاع للمرحلة الثانية= $\frac{500}{2000} * 360 = 90$

زاوية القطاع للمرحلة الثالثة= $\frac{400}{2000} * 360 = 72$

زاوية القطاع للمرحلة الرابعة= $\frac{300}{2000} * 360 = 54$



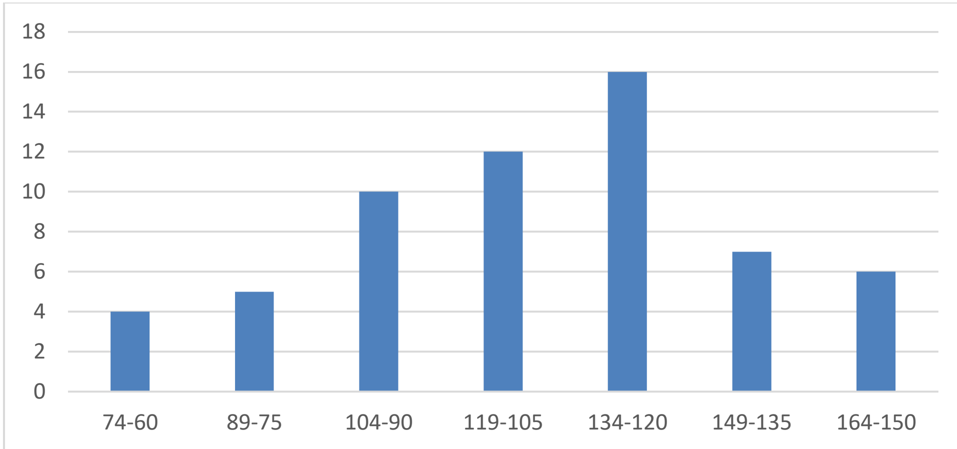
2- المدرج التكراري *Frequency histogram*:

وهو عبارة عن مجموعة من المستطيلات قاعدة كل منها تمثل طول الفئة في التوزيع التكراري ارتفاع كل منها يمثل قيمة التكرار المقابل لتلك الفئة في حالة المتغيرات **المتقطعة** تكون

المستطيلات منفصلة وفي حالة المتغيرات **مستمرة** فان **المستطيلات تكون متصلة**.

<u>مثال:</u> الاتي جدول توزيع تكراري يطلب رسم المدرج التكراري لهذه البيانات.(بيانات متقطعة)

الفئات	التكرار
74-60	4
89-75	5
104-90	10
119-105	12
134-120	16
149-135	7
164-150	6



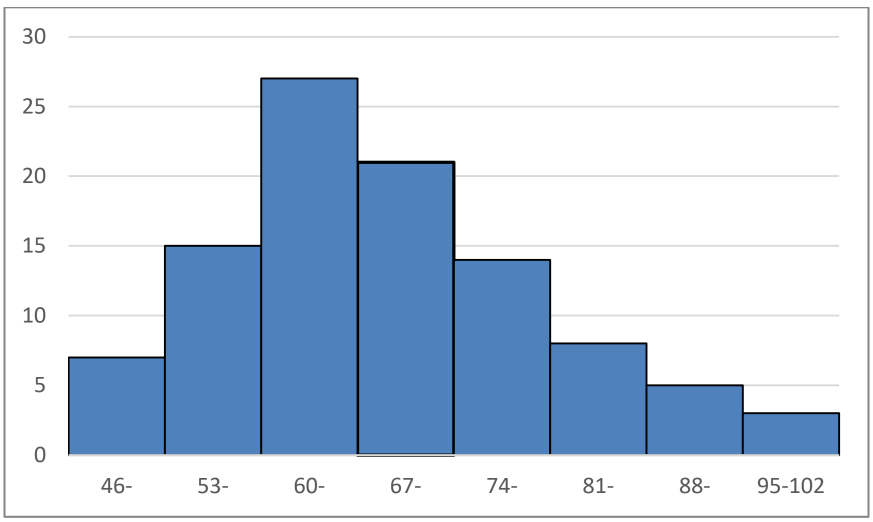
توضيح / ان المحور السيني يمثل الفئات والمحور الصادي يمثل التكرارات فنلاحظ ان اول فئه (74-60) تقابل قيمة تكرارها 4 على المحور الصادي ورسمت وهكذا لباقي الفئات.

مثال: من جدول التوزيع التكراري الاتي، يطلب رسم المدرج التكراري. (متغيرات مستمرة)

الفئات	التكرار
-46	7
-53	15
-60	27
-67	21
-74	14

-81	8
-88	5
102-95	3

<u>الحل:</u>



3- المضلع التكراري:

عبارة عن عدد من المستقيمات المتصلة مع بعضها على شكل سلسلة. ونقطة اتصال المستقيم بالأخر يقابل مركز الفئة. وهذا يعني انه عند رسم المضلع التكراري يستوجب الامر إيجاد مراكز الفئات ومن ثم رسم المضلع على أساس ازواج القيم (مركز الفئة * التكرار). ويفضل غلق المضلع التكراري مع المحور السيني وذلك باختيار مركز

فئة وهمي قبل مركز الفئة الأول، وآخر بعد مركز الفئة الأخير ونفترض ان تكرار هذين المركزين مساوٍ للصفر وتتم عملية الغلق بخط منقط.

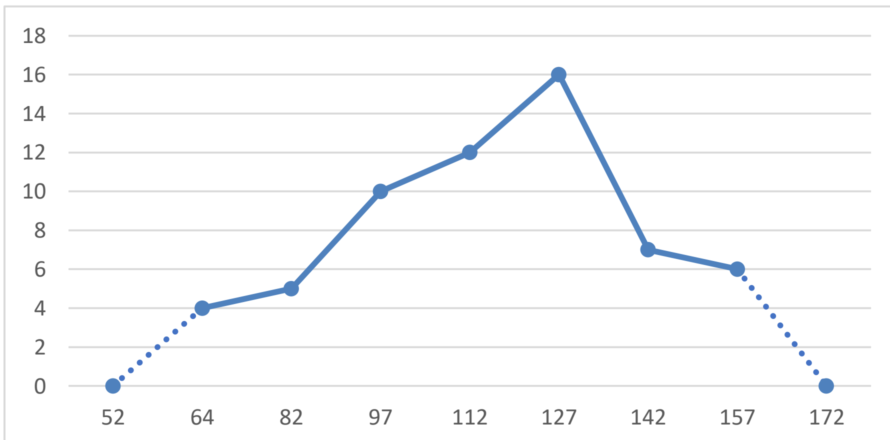
<u>مثال:</u> من جدول التوزيع التكراري التالي، ارسم المضلع التكراري

الفئات	التكرار	مركز الفئة
74-60	4	67
89-75	5	82
104-90	10	97
119-105	12	112
134-120	16	127
149-135	7	142
164-150	6	157

الحل:

مركز الفئة الوهمي الأول=مركز الفئة الأول-طول الفئة

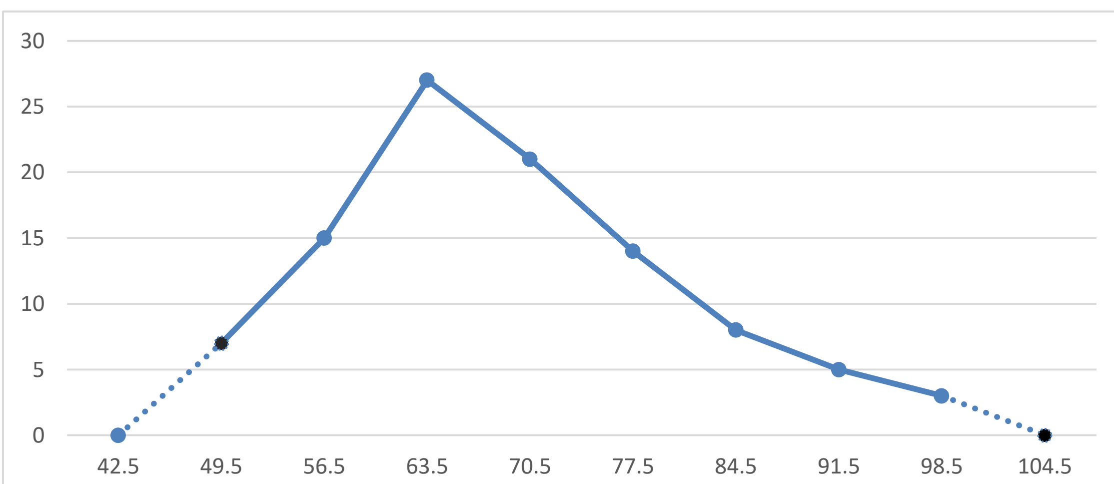
مركز الفئة الوهمي الأخير=مركز الفئة الأخير + طول الفئة



توضيح /نلاحظ هنا ان المحور السيني يمثل مركز الفئه والمحور الصادي يمثل التكرارات وعنده يتم رسم كل مركز فئة مع قيمه التكرار الخاص بتلك الفئة .

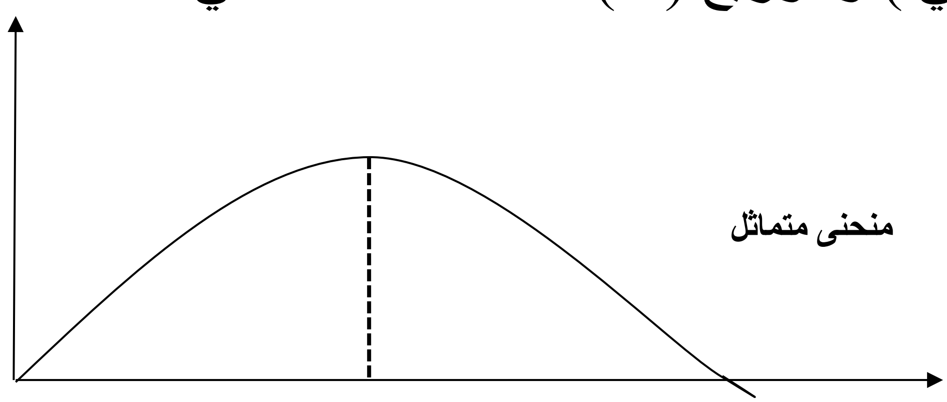
مثال: من جدول التوزيع التكراري التالي، ارسم المضلع التكراري.

الفئات	التكرار	مركز الفئة
-46	7	49.5
-53	15	56.5
-60	27	63.5
-67	21	70.5
-74	14	77.5
-81	8	84.5
-88	5	91.5
102-95	3	98.5



الاشكال البيانية للتوزيعات التكرارية:
تنقسم التوزيعات التكرارية بشكل عام الى قسمين رئيسين:
التوزيعات التكرارية المتماثلة والتوزيعات التكرارية غير المتماثلة (الملتوية) .

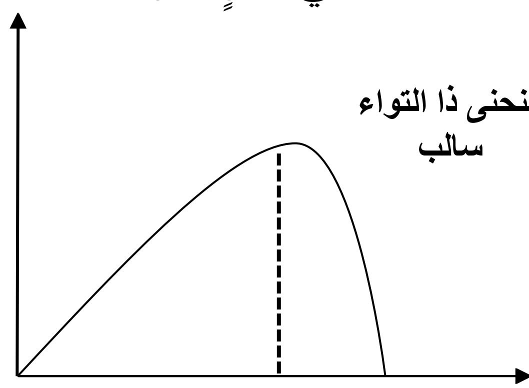
التوزيع التكراري المتماثل:
هو ذلك التوزيع الذي يتصف بان لمنحناه (محور وهمي) يقسم المساحة تحته الى قسمين متشابهين و متساويين، ومثال هذا النوع من التوزيعات هو التوزيع الاحتمالي للـ (التوزيع الطبيعي) و توزيع (t)، لاحظ الشكل التالي:

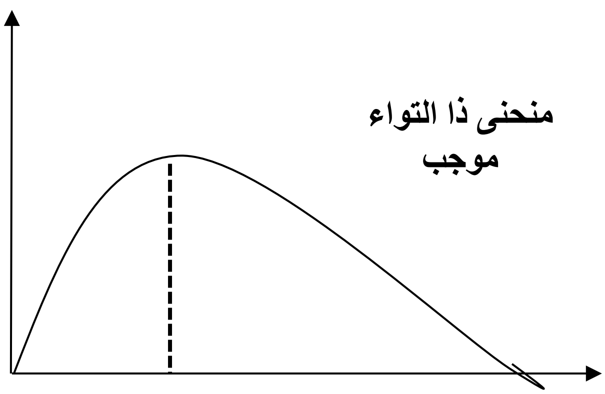


التوزيع التكراري غير المتماثل:
يعرف بانه ذلك التوزيع الذي يتصف بكون المساحة تحت منحناه الى اليمين (المحور الوهمي) لا تساوي المساحة تحت منحناه الى يسار المحور، ومن هذا النوع من التوزيعات، توزيع مربع كاي و توزيع F، وهما على نوعين.

*- **توزيعات ذات التواء موجب**: وهي تلك التوزيعات التي تكون فيها المساحة تحت منحني التوزيع الى يمين المحور أكبر من تلك الى يساره وعليه فان المنحني ملتوٍ التواء موجب.

-توزيعات ذات التواء سالب:** هي تلك التوزيعات التي تكون فيها المساحة تحت منحني التوزيع الى يسار المحور أكبر من تلك التي الى يمينه وعليه فان المنحني ملتوٍ التواء سالب.





رموز ومصطلحات رياضية:

قبل الشروع لدراسة الموضوعات ذات العلاقة بطرق احتساب المؤشرات الإحصائية (مقاييس النزعة المركزية ومقاييس التشتت وغيرها) للبيانات المتاحة فانه من المناسب استعراض اهم الرموز والمصطلحات الرياضية:

1- رمز الجمع(Summation) ∑: غالبا ما نحتاج عند التعامل مع الطريقة الإحصائية في التحليل الى عملية جمع سلسلة من الاعداد والكميات، فعلى سبيل المثال لو كانت لدينا سلسلة من الاعداد وهي X_1،X_2،X_3،---، X_n وعندئذٍ فان المجموع الكلي لهذه الاعداد هي $X_1+X_2+X_3+\text{---}+X_n$ وبهدف تسهيل عملية كتابة هذا المجموع الى نحو اكثر اختصاراً فانه يتم التعبير عنه بالشكل $\sum_{i=1}^{n} X_i$ حيث ان الرمز ∑ يشير الى عملية الجمع وان (i) تمثل دليل لتسلسل العدد عند عملية الجمع فاذا كانت $i=1$ فذلك يعني العدد الأول،

اما اذا كانت $i=5$ فذلك يعني العدد الخامس وهكذا. والجدول التالي يوضح استخدامات هذا الرمز بافتراض وجود سلسلة او كميات او اعداد هي (X_1،X_2،X_3،---، X_n):

العملية المطلوبة	العملية بهيئة رموز	رمز الجمع
مجموع قيم عناصر السلسلة	$x_1 + x_2 + x_3 + \cdots + x_n$	$\sum_{i=1}^{n} x_i$
مجموع مربعات قيم عناصر السلسلة	$x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2$	$\sum_{i=1}^{n} x_i^2$
مربع مجموع قيم عناصر السلسلة	$[x_1 + x_2 + x_3 + \cdots + x_n]^2$	$\left[\sum_{i=1}^{n} x_i\right]^2$
مجموع مقلوب عناصر السلسلة	$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}$	$\sum_{i=1}^{n} \frac{1}{x_i}$
مقلوب مجموع عناصر السلسلة	$\frac{1}{x_1 + x_2 + x_3 + \cdots + x_n}$	$\frac{1}{\sum_{i=1}^{n} x_i}$
مجموع لوغاريتمات عناصر السلسلة	$Logx_1 + Logx_2 + Logx_3 + \cdots + Logx_n$	$\sum_{i=1}^{n} Logx_i$

وفيما يلي خصائص هذا الرمز بافتراض ان X_i تمثل عناصر سلسلة من البيانات عن المتغير X وان Y_i تمثل عناصر سلسلة من البيانات عن المتغير Y حجم كل منها n من العناصر.

*- مجموع حاصل ضرب الثابت(a) بقيم عناصر السلسلة (X_i) يمثل حاصل ضرب الثابت(a) بمجموع قيم عناصر السلسلة أي ان:

$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$

**- مجموع حاصل ضرب عناصر السلسلة (X_i) بعناصر السلسلة (Y_i) المقابلة لها هو:

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + \cdots + x_n y_n$$

***- حاصل جمع الثابت a الى n من المرات هو:

$$\sum_{i=1}^{n} a = n * a$$

****- مجموع عناصر سلسلتين هو

$$\sum_{i=1}^{n}\left(x_i \mp y_i\right) = \sum_{i=1}^{n} x_i \mp \sum_{i=1}^{n} y_i$$

<u>مثال:</u> فيما يلي سلسلتي الاعداد التاليتين عن المتغيرين X و Y

$\boldsymbol{x_i} = 2،4،5،8،3$

$y_i = 1،2،3،2،3$

المطلوب/

$$\sum_{i=1}^{n}\frac{1}{x_i}،\ \sum_{i=1}^{n} x_i y_i،\ \sum_{i=1}^{n} Log x_i،\ \sum_{i=1}^{n} x_i y_i^2،\ \sum_{i=1}^{n} x_i،\ \sum_{i=1}^{n} y_i^2،\ \sum_{i=1}^{n} 3x_i\ ،\ \left(\sum_{i=1}^{n} x_i\right)^2$$

الحل/

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{5} x_i = 2 + 4 + 5 + 8 + 3 = 22$$

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{5} y_i^2 = 1^2 + 2^2 + 3^2 + 2^2 + 3^2 = 27$$

$$\sum_{i=1}^{n} 3x_i \quad = \sum_{i=1}^{5} 3x_i = 3\sum_{i=1}^{5} x_i = 3(22) = 66$$

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \left(\sum_{i=1}^{5} x_i\right)^2 = (22)^2 = 484$$

$$\sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{5} \frac{1}{x_i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{8} + \frac{1}{3} = \frac{169}{120}$$

$$\sum_{i=1}^{n} x_i\, y_i = \sum_{i=1}^{5} x_i\, y_{i=(2*1)+(4*2)+(5*3)+(8*2)+(3*3)=50}$$

$$\sum_{i=1}^{n} Logx_i = \sum_{i=1}^{5} Logx_i = Log2 + Log4 + Log5 + Log8 + Log3 =$$
$$= 0.310 + 0.602 + 0.698 + 0.903 + 0.477 = 2.982$$

$\sum_{i=1}^{n} x_i\, y_i^2 = \sum_{i=1}^{5} x_i\, y_i^2 = (2*1^2) + (4*2^2) + (5*3^2) + (8*2^2) + (3*3^2) = 122$

2- رمز الضرب π (Pi):تصادفنا في بعض الأحيان الى عملية ضرب مجموعة من الكميات ببعضها مما يتطلب الامر الى ترميز هذه العملية برمز، الهدف منه هو تسهيل كتابة العملية هذه في اقل حيز ممكن، والرمز المستخدم في عمليات الضرب هو [**π**] كدليل لوجود عملية ضرب مجموعة من الكميات، وباستخدام رمز الضرب فان التعبير عن هذه العملية هو: $\prod_{i=1}^{n} x_i$ وان هذه العملية تقرأ (حاصل ضرب عناصر سلسلة من الكميات ببعضها ابتداءً بالعنصر الأول) وبافتراض ان x_1،x_2،x_3، ... ،x_n تمثل عناصر سلسلة من البيانات عن المتغير xوان y_1،y_2،y_3، ... ،y_n تمثل عناصر سلسلة من البيانات عن المتغير y وان a،b ثابتين حقيقيين فان:

1- حاصل ضرب عدد حقيقي ل n من المرات هو

$$\prod_{i=1}^{n} a = a^n$$

-2

$$\prod_{i=1}^{n} ax_i = a^n \prod_{i=1}^{n} x_i$$

-3

$$\prod_{i=1}^{n} abx_i\, y_i = (ab)^n \prod_{i=1}^{n} x_i y_i$$

<u>مثال:</u> تأمل سلسلتي الاعداد التاليتين عن المتغيرين x،y :

$$x_i = 2،3،4،5$$

$$y_i = 1،2،3،4$$

1- $\prod_{\mathrm{i}=1}^{6} 3 = 3^6 = 729$

2- $\prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i} = \prod_{\mathrm{i}=1}^{4} \mathrm{x_i} = 2 * 3 * 4 * 5 = 120$

3- $\prod_{\mathrm{i}=1}^{\mathrm{n}} 5\mathrm{x_i} = 5^4 \prod_{\mathrm{i}=1}^{4} \mathrm{x_i} = 5^4 * 120 = 75000$

4- $\prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{x_i} = \frac{1}{\prod_{\mathrm{i}=1}^{4} \mathrm{x_i}} = \frac{1}{120} =$

5-

$$\prod_{\mathrm{i}=1}^{\mathrm{n}} 3\mathrm{x}_{\mathrm{i}} y_i =$$

$$\prod_{\mathrm{i}=1}^{4} 3\mathrm{x}_{\mathrm{i}} y_i = 3^4 \prod_{\mathrm{i}=1}^{4} \mathrm{x}_{\mathrm{i}} \prod_{\mathrm{i}=1}^{4} \mathrm{y}_{\mathrm{i}} = 3^4 * 120 * 24 = 233280$$

مقاييس النزعة المركزية

قبل الدخول في مقاييس النزعة المركزية يجب التذكير بأنواع البيانات لدينا

1- البيانات الخام (الاصلية) غير المبوبة Ungrouped data وهي البيانات الاصلية قبل ترتيبها وعرضها في جداول تكرارية

2- البيانات المبوبة grouped data وهي البيانات التي تم تصنيفها وترتيبها وعرضها في جداول تكرارية

*- مفهوم المتوسطات والهدف من احتسابها:

يمكن تمثيل مجموعة من البيانات بقيمة واحدة فقط، الهدف من ذلك إعطاء صورة سريعة عن ماهيّة تلك المجموعة من خلال إيجاد عدد يمثل قيمها. ان المقياس الذي يختص بتحديد هذا العدد يسمى مقياس نزعة مركزية، هذا العدد يميل لان يقع في وسط تلك المجموعة من البيانات في حال ترتيبها حسب صغرها او كبرها، هذا الامر جعلنا ان نطلق على هذا النوع من المقاييس بـ (مقاييس نزعة مركزية).

ان للمتوسطات أهمية كبيرة في الحياة العملية فهي تستخدم في الجوانب الاقتصادية للدولة كالزراعة والصناعة والانتاج وغيرها، وفيما يلي اهم مقاييس النزعة المركزية:

1- الوسط الحسابي Arithmetic mean:

هو أحد اهم مقاييس النزعة المركزية وأكثرها استخداما وشيوعا وغالبا ما يسمى بالمعدل (Average) ويفضل على جميع مقاييس النزعة المركزية لكونه يستعمل جميع البيانات وباستخدام صيغ مختلفة.

1- الوسط الحسابي للبيانات الخام (البيانات غير المبوبة): إذا كان لدينا (n) من الاعداد (قيم المشاهدات) x_1، x_2، x_3، ... ، x_n فان الوسط الحسابي لها هو حاصل قسمة مجموع البيانات على عددها ويرمز للوسط الحسابي بالرمز $[\bar{X}]$ حيث ان:

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

مثال: جد الوسط الحسابي للمشاهدات التالية: 12، 2، 22، 17، 0، 5، 15، 7

$$\bar{X} = \frac{7+15+5+0+17+22+2+12}{8} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{80}{8} = 10$$

-2 الوسط الحسابي للبيانات المبوبة: اذا كانت لدينا x_1،x_2،x_3، ...،x_n تمثل مراكز الفئات في جدول التوزيع التكراري مع التكرار المقابل لها على التوالي f_1،f_2،f_3، ...،f_n فان الوسط الحسابي لها يكون كما يلي:

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

وان خطوات إيجاد الوسط الحسابي كما يلي:

1- تعيين مراكز الفئات (x_i) .

2- ضرب مركز كل فئة بمقدار تكرارها $(f_i x_i)$.

3- قسمة مجموع (حاصل ضرب مركز كل فئة * تكرارها) على مجموع التكرارات.

<u>مثال:</u> الاتي توزيع تكراري لدرجات الحرارة في مدينة معينة والمسجلة لمدة (95) يوما متتاليا، المطلوب حساب متوسط درجة الحرارة لهذه المدينة خلال هذه الفترة.

الفئات	f_i	x_i	$f_i x_i$
-0	4	0.5	2
-1	8	1.5	12
-2	12	2.5	30
-3	16	3.5	56
-4	20	4.5	90
-5	25	5.5	137.5
-6	6	6.5	39
8-7	4	7.5	30
المجموع	95	-	396.5

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

$$= \frac{396.5}{95} = 4.174$$

وهذا يعني ان متوسط درجة الحرارة في هذه المدينة خلال تلك الفترة كان مساويا الى (4.174).

مثال: الاتي توزيع تكراري لعينة من الاسر حجمها (75) اسرة حسب عدد افراد الاسرة، المطلوب حساب الوسط الحسابي او متوسط عدد افراد الاسرة في هذه العينة.

الفئات	f_i	x_i	$f_i x_i$
4-2	8	3	24
7-5	12	6	72
10-8	20	9	180
13-11	13	12	156
16-14	10	15	150
19-17	8	18	144
22-20	4	21	84
	75		810

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

$$= \frac{810}{75} = 10.8$$

$$\cong 11$$

ملاحظة *: ان عدد افراد الاسرة متغير من النوع المتقطع وانه لا يوجد تقييس لجزء من الفرد، عليه يتم التقريب الى اقرب عدد صحيح، أي ان متوسط عدد افراد الاسرة في هذه العينة هو تقريبا (11) فرد .

ميزات وعيوب الوسط الحسابي:

*- ميزات الوسط الحسابي:

1- بساطة فكرته .

2- سهولة حسابه

3- ان حسابه يستند الى كافة البيانات المتوفرة.

**- عيوب الوسط الحسابي:

1- لا يمكن تحديد الوسط الحسابي في حالة فقدان قيمة او اكثر من قيم العينة الا من خلال تقدير هذه القيم او الاستعاضة عنها بقيم مفردات مكملة لحجم العينة.

2- لا يمكن تحديد الوسط الحسابي لبيانات متغير وصفي (نوعي) كالجنس او صنف الدم وغيرها، ما عدا الحالات التي يمكن فيها إعادة تقييس الصفات مثل تقديرات الامتحان المعتمدة في المرحلة الجامعية وهي: [ضعيف، مقبول، متوسط، جيد، جيد جداً، ممتاز]، والتي تمثل صفات لمتغير نوعي يمكن تقسمها لفئات وهي:

[0-، 50-، 60-، 70-، 80-، 90-، 100].

3- ان قيمة الوسط الحسابي تتأثر بشكل كبير بالقيم الشاذة (المتطرفة).

4- تتأثر قيمة الوسط الحسابي بأخطاء المعاينة أي ان تسجيل خاطئ لقياسات العينة ينعكس على قيمة المتوسط.

خصائص الوسط الحسابي:

1- ان مجموع انحرافات قيم المتغير (X) عن وسطها الحسابي يساوي صفر:

في حالة البيانات غير المبوبة $\sum(x_i - \bar{x}) = 0$

في حالة البيانات المبوبة $\sum f_i(x_i - \bar{x}) = 0$

مثال:

لتكن لدينا القيم المفردة الخمس التالية: $x_i = 7 \quad 4 \quad 3 \quad 9 \quad 2$ ، بين ان انحرافات القيم عن وسطها الحسابي = صفر.

$$\bar{x} = \frac{7+4+3+9+2}{5} = 5$$

ومنه نجد ان انحرافات القيم x_i عن وسطها الحسابي $\bar{x}$ يساوي صفر:

x_i	7	4	3	9	2
$xi - \bar{X}$	2	-1	-2	4	-3

نلاحظ ان مجموع الصف الثاني (2+(-1)+(-2)+4+(-3)) يساوي الصفر وهو المطلوب

2- ان مجموع مربعات انحرافات قيم (X) عن وسطها الحسابي يكون اقل ما يمكن أي اقل من مجموع مربعات أي قيمة أخرى:

$$\sum(x_i - \bar{x})^2 = \text{اقل ما يمكن}$$

مثال:

من القيم التالية:

$$x_i = 9 \quad 8 \quad 6 \quad 5 \quad 7 \text{ ،} \qquad \bar{x} = 7$$

لا ثبات الخاصية الثاني نجد $\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 = (9-7)^2 + (8-7)^2 + (6-7)^2 + (5-7)^2 + (7-7)^2$$

= 4+1+1+4+0= **10**

فاذا طرحنا من القيم هذه أي رقم (غير الوسط الحسابي) وليكن A=10

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 = (9-10)^2 + (8-10)^2 + (6-10)^2 + (5-10)^2 + (7-10)^2$$

= 55

وهذا يعني ان 55 اكبر من 10 وهذا ما يحقق الخاصية الثانية .

3- عند إضافة عدد ثابت مثل (K) الى كل قيمة من قيم المشاهدات فان الوسط الحسابي ف للقيم الجديدة = الوسط الحسابي للقيم الاصلية + العدد الثابت (K) .

مثال: (الخاصية الثالثة)

لنفرض لدينا القيم التالية: $\bar{x} = \frac{35}{5} = 7$ ، $x_i = 8 \quad 3 \quad 2 \quad 12 \quad 10$

فاذا اضفنا لكل من هذه القيم قيمة ثابتة ولتكن (3) فالقيم الجديدة ستصبح: $x_i = 11 \quad 6 \quad 5 \quad 15 \quad 13$

والوسط لحسابي للقيم الجديدة هو:

$$\bar{x} = \frac{50}{5} = 10$$

الذي في الحقيقة هو:

$$\bar{x}new = \bar{x}old + 3 = 7 + 3 = 10$$

4- اذا ضربت كل قيمة من قيم المشاهدات في قيمة ثابتة مثل (K) فان الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الاصلية * العدد الثابت (K) .

مثال: تابعنا دراسة احد الطلاب لخمسة أيام فكانت: $x_i = 4 \quad 3 \quad 5 \quad 3 \quad 5$

$$\bar{x} = \frac{4+3+5+3+5}{5} = 4$$

لو قمنا بضرب كل قيمة من هذه القيم بقيمة ثابته ولتكن (2) فان القيم الجديدة ستكون $x_i = 8 \quad 6 \quad 10 \quad 6 \quad 10$ والوسط الحسابي لها هو

$$\bar{X} = \frac{40}{5} = 8$$

الذي في الحقيقة هو

$$\bar{X}new = \bar{X}old * 2 \rightarrow = 4 * 2 = 8$$

الوسط الحسابي المرجح (الموزون)Weighted mean

الوسط الحسابي المرجح : قد يتطلب الامر في حالات كثيرة ترجيح قيم الظاهرة بأحجام او اوزان مختلفة لبيان الأهمية النسبية للظاهرة.

فاذا كان لدينا القيم x_1،x_2،x_3، ... ،x_n والاوزان المناظرة لها هي w_1،w_2،w_3، ... ،w_n فان الوسط الحسابي المرجح يأخذ الصيغة التالية:

1- بالنسبة للبيانات غير المبوبة

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{(w_1 x_1) + (w_2 x_2) + (w_3 x_3) + \cdots + (w_n x_n)}{w_1 + w_2 + w_3 + \cdots + w_n}$$

2- بالنسبة للبيانات المبوبة

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

مثال:

البيانات التالية تمثل درجات طالب وعدد ساعات الدراسة لمجموعه من المواد الدراسية جد معدل الطالب في هذه المواد

درجات المواد xi	عدد ساعات الدراسة (wi)	Xi wi
75	3	225
87	2	174
63	3	189
94	3	282
61	4	244
المجموع	15	1114

الحل/

باضافة عمود ضر ثالث لحاصل الدرجات في عدد الساعات (xi *wi) فان الوسط الحسابي الموزون

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

$$\bar{x}_w = \frac{1114}{15} = 74.26$$

المنوال The mode

المنوال لمجموعة من القيم هو القيمة الاكثر تكراراً في التوزيع التكراري للقيم، أي ان المنوال هو تلك القيمة التي يقابلها اكبر تكرار.

طرق إيجاد المنوال:

أ- إيجاد المنوال للبيانات غير المبوبة:

هناك عدة حالات لايجاد المنوال للبيانات غير المبوبة:

1- اذا تكرر احد القيم لأكثر من غيره فيكون هنالك منوال واحد.

مثال: أوجد المنوال لقيم المشاهدات التالية:

7،7،11،5،11،7

الحل: القيمة الأكثر تكراراً هي القيمة (7).

2- اذا لم تتكرر أياً من القيم فلا يوجد منوال.

مثال: جد المنوال للقيم التالية:

7،9،11،12،15

الحل: لا يوجد منوال لهذه القيم حيث ان أياً من القيم لم تتكرر.

3- اذا كان لقيمتين نفس العدد من التكرار، فيكون للقيم منوالين.

مثال: أوجد المنوال لقيم المشاهدات التالية:

4،9،17،9،4،11

الحل: يوجد منوالان لهذه القيم وهما 4 و9 لأنه لهما نفس العدد من التكرارات.

ب- إيجاد المنوال للبيانات المبوبة:

لايجاد المنوال للبيانات المبوبة نتبع الخطوات التالية :

1- تحديد الفئة التي تمتلك اكبر تكرار

2-استخدام القانون التالي لايجاد المنوال

$$Mod = A + \frac{d_1}{d_1 + d_2}.L$$

حيث ان :

A : تمثل الحد الادنى لفئة المنوال (الفئة المناظرة لأكبر تكرار)

d_1 : تمثل الفرق الاول = (تكرار فئة المنوال – تكرار السابق لفئة المنوال)

d_2 : تمثل الفرق الثاني = (تكرار فئة المنوال ــ التكرار اللاحق لفئة المنوال)

L : طول فئة المنوال

فئة المنوال = الفئة المناظرة لأكبر تكرار

مثال : الاتي توزيع تكراري لأطوال عينة من الأشخاص البالغين حجمها 50 شخص. يطلب حساب القيمة الشائعة لطول الشخص في هذه العينة (أيجاد المنوال للتوزيع).

فئات الطول	عدد الاشخاص
-150	8
-160	12
-170	15
-180	9
200-190	6

تكرار سابق

تكرار فئة المنوال

تكرار لاحق

<u>الحل:</u> بما ان أكبر تكرار في التوزيع هو 15، فان فئة المنوال هي 170-180، أي الفئة الثالثة.

تكرار فئة المنوال = 15 ,

التكرار السابق لفئة المنوال =12 , التكرار اللاحق = 9

نجد

$$d_1 = 15 - 12 = 3$$

$$d_2 = 15 - 9 = 6$$

طول فئة المنوال L = 10

الحد الادنى لفئة المنوال A =170

$$Mod = A + \frac{d_1}{d_1 + d_2} . L$$

$$= 170 + \frac{3}{3+6} . 10 \ = 173.33$$

<u>ملاحظة:</u> يمكن أيجاد قيمة المنوال إذا كان جدول التوزيع التكراري مفتوح من طرف واحد او طرفين وهذه احدى ميزات المنوال.

الوسيط The Median

يعتبر الوسيط احد مقاييس النزعة المركزية المهمة في التطبيقات الإحصائية. ويعرف الوسيط بانه تلك القيمة من قيم المتغير العشوائي X التي تقسم مجموعة من قيم المتغير الى قسمين متساويين، أي انها قيمة X التي تجعل عدد القيم قبلها مساوٍ لعدد القيم بعدها.

1- **الوسيط للبيانات غير المبوبة**
لحساب الوسيط نتبع الخطوات التالية
1- ترتيب البيانات ترتيبا تصاعديا من اصغر قيمه الى اكبر قيمة (ويمكن تنازليا)

2- اذا كانت عدد البيانات (n) فردي فإن الوسيط هو القيمة التي تقع وسط البيانات بعد ترتيبها والتي يكون تسلسلها $\left(\frac{n+1}{2}\right)$

مثال : الاتي درجات عينة من الطلبة حجمها تسع طلاب في امتحان معين، جد الوسيط لهذه المجموعة:
80، 79،55 ، 63، 65، 68،70, 62،53

الحل:
نرتب القيم ترتيباً تصاعدياً

80، 79، 70، 68، 65، 63، 62، 55،53
وعليه فان ترتيب الوسيط هو:

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

وهذا يعني ان القيمة الخامسة هي قيمة الوسيط أي الدرجة "65".

3- اذا كان عدد البيانات n عدد زوجي فان قيمة الوسيط تمثل الوسط الحسابي لقيمتي x التي تتوسط البيانات بعد الترتيب، أي اللتان تسلسلهما على التوالي هو $\left(\frac{n}{2}, \frac{n}{2}+1\right)$

$$Med = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

مثال:
الاتي اعمار عينة من الافراد قوامها 12 فرد، جد الوسيط لعمر الفرد في هذه المجموعة:
23، 25، 20، 22، 19.5، 26، 24.5، 27، 28، 29، 18،20

الحل:
نرتب القيم وفق ترتيب تصاعدي وكما يلي:

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

عندئذ فان القيمتان اللتان تحددان الوسيط هما:

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

أي القيمتان اللتان تسلسلهما 6 و7 هما على التوالي 23 و24.5، أي القيمة السادسة والسابعة بعد الترتيب، وبذلك فان الوسيط لهذه المجموعة يمثل الوسط الحسابي لهاتين القيمتين وكالتالي:

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

2- الوسيط للبيانات المبوبة
أ- الوسيط لبيانات متغير متقطع
لحساب الوسيط نتبع الخطوات التالية :

1- نحسب التكرار المتجمع الصاعد.
2- الوسيط هنا يمثل القيمة التي تقع نصف مجموع التكرارات قبلها والنصف الآخر بعدها. اي ان تسلسل الوسيط يمثل نصف التكرارات $\frac{\sum_{i=1}^{n} f_i}{2}$
3- نقارن تسلسل الوسيط مع التكرار المتجمع الصاعد فان
$F_{k-1}<\sum\frac{f_i}{2}<F_k$, بحيث ان F_{k-1} يمثل التكرار المتجمع الصاعد السابق , وان F_k تمثل التكرار المتجمع اللاحق ,
4- عندئذ يقال ان فئة الوسيط هي الفئة التي تسلسلها "k"، وبذلك فان قيمة الوسيط تمثل مركز هذه الفئة.

مثال: الآتي توزيع تكراري لعينة من الأسر حسب عدد أفراد الأسرة، يطلب حساب الوسيط لعدد أفراد هذه الأسر.

عدد الفئات	عدد الأسر	الحدود العليا للفئات	التكرار المتجمع الصاعد
4-2	6	**أقل من أو يساوي 4**	**6**
7-5	9	**أقل من أو يساوي 7**	**15**
10-8	12	**أقل من أو يساوي 10**	**27** ← F_{k-1}
13-11	20	**أقل من أو يساوي 13**	**47** ← F_k
16-14	14	**أقل من أو يساوي 16**	**61**
19-17	11	**أقل من أو يساوي 19**	**72**
22-20	8	**أقل من أو يساوي 22**	**80**
المجموع	80		

الحل /

1- نجد التكرار المتجمع الصاعد كما هو في العمود الثالث والرابع

2- ان ترتيب الوسيط هو:

$$\frac{\sum f}{2} = \frac{80}{2} = 40$$

وبملاحظة ترتيب الوسيط ضمن التكرار المتجمع الصاعد، نلاحظ ان 47≥40>27، وعليه فان فئة الوسيط هي الفئة الرابعة من التوزيع أي 11-13 وعندئذ فان الوسيط لهذا التوزيع يمثل مركز هذه الفئة "أي 12" فرد، أي ان:

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

مع ملاحظة انه اذا كان الناتج قيمة كسرية عندئذ يقرب الناتج الى اقرب عدد صحيح.

ب- الوسيط لبيانات المتغير مستمر:

1 حساب التكرار المتجمع الصاعد

2- حساب تسلسل الوسيط $\frac{\sum_{i=1}^{n} f_i}{2}$

3- تحديد فئة الوسيط كما في حالة بيانات المتغير المتقطع

4- ايجاد الوسيط

$$Med = A + \left[\frac{\frac{\sum f_i}{2} - F_{k-1}}{f_k} \right] . L$$

حيث ان :

A : الحد الادنى لفئة الوسيط

F_{k-1} : التكرار المتجمع الصاعد السابق لفئة الوسيط

f_k : تكرار فئة الوسيط

L : طول فئة الوسيط

مثال:

فيما يلي توزيع تكراري للدخل الشهري لعينة من الأسر حجمها 80 أسرة، جد الوسيط للدخل الشهري للأسرة في هذه العينة.

الفئات	عدد الاسر	***الحدود العليا للفئات***	***التكرار المتجمع الصاعد***
-100	3	***أقل من 120***	***3***
-120	7	***أقل من 140***	***10***
-140	14	***أقل من 160***	***24***
-160	20	***أقل من 180***	***44***
-180	18	***أقل من 200***	***62***
-200	12	***أقل من 220***	***74***
240-220	6	***أقل من أو يساوي 240***	***80***

1- نجد التكرار المتجمع الصاعد

2- نجد ترتيب الوسيط وهو مساوي الى نصف التكرارات أي $\frac{\sum f_i}{2} = \frac{80}{2} = 40$ ، نقارن ترتيب الوسيط مع التكرار المتجمع الصاعد حيث نلاحظ ان $24<40<44$ وعليه فان فئة الوسيط هي الفئة الرابعة في التوزيع، أي الفئة 160-180، وبذلك فان:

$F_{k-1}=24$ ، $f_k=20$ ، $L=20$ ، $A=160$

وعندئذ فان:

$$Med = A + \left[\frac{\frac{\sum f_i}{2} - F_{k-1}}{f_k} \right] . L$$

$$= 160 + \left[\frac{40-24}{20} \right] \times 20$$

$$= 160 + 16 = 176$$

<u>مثال في حالة البيانات المفتوحه من البداية والنهاية لملاحظة انها لا تؤثر على حساب الوسيط</u>

<u>مثال:</u> فيما يلي توزيع تكراري لأعمار عينة من تلاميذ احدى المدارس الابتدائية حجمها 90 تلميذ المطلوب إيجاد الوسيط لعمر التلميذ في هذه العينة.

الفئات	التكرار	***الحدود العليا للفئات***	***التكرار المتجمع الصاعد***
أقل من 6	3	***أقل من 6***	***3***
-6	6	***أقل من 7***	***9***
-7	9	***أقل من 8***	***18***
-8	12	***أقل من 9***	***30***
-9	20	***أقل من 10***	***50***
-10	18	***أقل من 11***	***68***
-11	17	***أقل من 12***	***85***
12 فأكثر	5	***أقل من مجهول***	***90***

ترتيب الوسيط = $\frac{\sum_{i=1}^{n} f_i}{2} = \frac{90}{2} = 45$

وعليه فان فئة الوسيط هي الفئة الخامسة من التوزيع، أي الفئة 9-10 وعليه فان:

$F_{k-1}=30$، $f_k=20$ ، $L=1$ ،$A=9$

$$\mu_e = 9 + \frac{(45-30)}{20} * 1 = 9 + \frac{15}{20}$$
$$= 9 + 0.75 = 9.75$$

أهم ميزات وعيوب الوسيط:
أ-ميزات الوسيط

1- بساطة فكرته
2- سهل القيم والحساب
3- يمكن تعيينه هندسياً
4- لا يتأثر اطلاقاً بالقيم الشاذة او المتطرفة
5- يمكن حسابه في حالة التوزيعات التكرارية المفتوحة من طرف واحد او من طرفين
6- يمكن تعيينه في حالة البيانات الوصفية
7- يمكن ايجاده في حالة التوزيعات ذات الفئات غير متساوية الاطوال

ب-عيوب الوسيط

1- لا يستند في حسابه على كافة البيانات المتوفرة، حيث بمجرد معرفة ترتيب الوسيط تحدد قيمته في حالة البيانات غير المصنفة وتهمل بقية القيم الأخرى.

العلاقة بين مقاييس النزعة المركزية:

هناك علاقة رياضية تربط الوسط الحسابي، الوسيط والمنوال في حالة التوزيعات **التكرارية المتماثلة او القريبة من التماثل**، هي:

$$\bar{X} - Med = \frac{1}{3}(\bar{X} - Mod)$$

وان هذه العلاقة مهمة جداً، حيث انهُ يتعذر في بعض الأحيان مثلاً حساب الوسط الحسابي من توزيع تكراري مفتوح من طرف واحد او من طرفين في حين يمكن إيجاد الوسيط والمنوال في هذا التوزيع.

مثال: في احدى التوزيعات القريبة من التماثل تم الحصول على كل من Mod=68 و Med= 67، يطلب حساب القيمة التقريبية من الوسط الحسابي.

الحل:

$$\bar{X} - Med = \frac{1}{3}(\bar{X} - Mod) \Rightarrow \bar{X} - 67 = \frac{1}{3}(\bar{X} - 68)$$

$$3(\bar{X} - 67) = \bar{X} - 68 \Rightarrow 3\bar{X} - \bar{X} = 201 - 68$$

$$2\bar{X} = 133 \Rightarrow \bar{X} = 66.5$$

مقاييس التشتت Measures of Variation

مقاييس التشتت والهدف من احتسابها

في الادبيات الاحصائية ذكر مصطلح التشتت والمقصود به التبعثر Scatteredness أي انتشار قيم مجموعة من المفردات ويمكن تعريف التشتت بانه تباعد او انتشار قيم مجموعة من المفردات عن بعضها البعض أدى الى قيمة معينة ثابتة (كالوسط الحسابي)، ان الهدف من دراسة التشتت هو تكوين فكرة عن مدى تجانس قيم مجموعة من المفردات وهذا يعني ان دراسة التشتت امر مفيد في اجراء مقارنة بين مجموعتين او أكثر من البيانات عن ظاهرة معينة.

فمثلا ان الوسط الحسابي لكل من المجموعات التالية = 9 حيث ان:

المجموعة الأولى: $[11،10،9،8،7،9] \quad \bar{X} = 9$

المجموعة الثانية: $[15،12،9،6،9،3] \quad \bar{X} = 9$

تبدو المجموعة الاولى اكثر تجانسا (اقل انتشارا) وان مقياس التشتت يكون كبير كلما كانت البيانات اقل تجانسا ويكون صغيرا عندما تكون الاختلافات بين القيم قليلة .

هناك عدة مقاييس للتشتت أهمها ما يلي:

1- المدى (Range (R

يعتبر المدى ابسط انواع مقاييس التشتت ، ويعرف بانه الفرق ما بين اكبر قيمة في مجموعة بيانات واصغر قيمة فيها، ويرمز للمدى بالرمز R حيث ان:

$$R = X_L - X_S$$

مثال: جد المدى للبانات التالية:

$$x_i = 2،5،3،8،7،10،9،12،15$$

الحل:

$$R = 15 - 2 = 13$$

اما في حالة البيانات المبوبة في توزيع تكراري فان المدى في هذه الحالة يمثل الفرق ما بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

2- **التباين والانحراف المعياري** variance and standard deviation

يعرف الانحراف لأي قيمة ضمن مجتمع البيانات بانها الفرق بين تلك القيمة ومتوسط مجتمع البيانات ,اي (قيمة x – قيمة المتوسط للبيانات)

Deviation of x = x-µ

يعرف التباين (variance) بانه متوسط مجموع مربعات انحرافات قيم { X } عن وسطها الحسابي , اما الانحراف المعياري (standard deviation) فهو الجذر التربيعي لقيمة التباين .

1- بالنسبة للبيانات غير المبوبة

1- **التباين variance**

- تباين المجتمع : لحساب التباين لجميع مفردات المجتمع نتبع الصيغة التالي

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(x_i - \mu\right)^2}{N}$$

- تباين العينة: غالبا من المستحيل ايجاد التباين لمفردات المجتمع بشكل كامل لذلك يستخدم الاحصائيون تباين العينة والذي يرمز له بالرمز S^2 ويحسب بالصيغة التالية :

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}$$

2- **الانحراف المعياري standard deviation**

- الانحراف المعياري للمجتمع هو الجذر التربيعي للتباين

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \mu\right)^2}{N}}$$

- الانحراف المعياري للعينة

الانحراف المعياري لعينة من البيانات بحجم n هي الجذر التربيعي لتباين العينة ويحسب بالصيغة التالية:

$$s = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}}$$

مثال: جد التباين والانحراف المعياري للبيانات التالية

5 , 10 , 12 , 13 , 20

1- لإيجاد التباين نتبع الخطوات التالية

1- نحسب الوسط الحسابي للبيانات

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{20+13+12+10+5}{5} = 12$$

2- نحسب $x_i - \bar{x}$ لكل قيمة

3- نحسب $(x_i - \bar{x})^2$ وهي مربع القيمة المحسوبة في الفقرة الثانية اعلاه

كما موضح في الجدول التالي

$(x_i - \bar{x})^2$	$x_i - \bar{x}$
$8^2 = 64$	20-12=8
1	13-12=1
0	12-12=0
4	10-12=-2
49	5-12 = -7

التباين يساوي

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \rightarrow= \frac{64+1+0+4+49}{5-1} = \frac{118}{4} = 29.5$$

2- الانحراف المعياري يساوي

$$s = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{29.5} = 5.43$$

2- التباين والانحراف المعياري للبيانات المبوبة

طريقة حساب التباين والانحراف القياسي للبيانات المبوبة يشبه طريقة حساب الوسط الحسابي للبيانات المبوبة وكذلك يستخدم في حسابه مركز الفئة لكل فئة

1- التباين

$$S^2 = \frac{\sum_{i=1}^{m} f_i (X_i - \bar{X})^2}{\sum_{i=1}^{m} f_i}$$

حيث m عدد الفئات في الجدول التكراري

ويمكن تبسيط الصيغة السابقة بالشكل التالي

$$s^2 = \frac{\sum_{i=1}^{m} f_i x_i^2 - \frac{\left(\sum_{i=1}^{m} f_i x_i\right)^2}{\sum_{i=1}^{m} f_i}}{\sum_{i=1}^{m} f_i - 1}$$

2- الانحراف المعياري وهو الجذر التربيعي للتباين

$$s = \sqrt{\frac{\sum_{i=1}^{m} f_i (X_i - \bar{X})^2}{\sum_{i=1}^{m} f_i}}$$

او

$$s = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2 - \frac{\left(\sum_{i=1}^{m} f_i x_i\right)^2}{\sum_{i=1}^{m} f_i}}{\sum_{i=1}^{m} f_i - 1}}$$

مثال من جدول التوزيع التكراري التالي، جد التباين الانحراف المعياري .

الفئات	f_i	X_i	f_iX_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^2$
-30	2	***35***	***70***	***30-***	***900***	***1800***
-40	5	***45***	***225***	***20-***	***400***	***2000***
-50	3	***55***	***165***	***10-***	***100***	***300***
-60	4	***65***	***260***	***0***	***0***	***0***
-70	7	***75***	***525***	***10***	***100***	***700***
90-80	6	***85***	***510***	***20***	***400***	***2400***
	27		***1755***			***7200***

الحل

1- التباين

نجد الوسط الحسابي

$$\bar{X} = \frac{\sum_{i=1}^{6} f_i X_i}{\sum_{i=1}^{6} f_i} = \frac{1755}{27} = 65$$

نحسب التباين حسب القانون وبالجدول اعلاه الاعمدة ذات الخط الغامق جميعها يتم حسابها

$$\frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i - 1} = \frac{1800 + 2000 + \cdots . +2400}{(2 + 5 + \cdots . +6) - 1}$$

$$\frac{7200}{26} = 276.92$$

2- الانحراف المعياري
وهو الجذر التربيعي للتباين وكالاتي

$$S = \sqrt{\frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i - 1}} = \sqrt{\frac{7200}{27-1}} = \sqrt{276.92} = 16.64$$

مثال/ نفس السؤال السابق لكن الحل بالطريقة الثانية

$f_i x_i^2$	$f_i X_i$	X_i	f_i	الفئات
2450	***70***	***35***	2	-30
10125	***225***	***45***	5	-40
9075	***165***	***55***	3	-50
16900	***260***	***65***	4	-60
39375	***525***	***75***	7	-70
43350	***510***	***85***	6	90-80
121275	***1755***		27	

التباين يساوي

=

$$s^2 = \frac{\sum_{i=1}^{m} f_i x_i^2 - \frac{\left(\sum_{i=1}^{m} f_i x_i\right)^2}{\sum_{i=1}^{m} f_i}}{\sum_{i=1}^{m} f_i - 1} = \frac{121275 - \frac{(1755)^2}{27}}{27-1} =$$

مثال واجب

جد التباين والانحراف المعياري للجدول التكراري التالي

الفئات	-30	-34	-38	-42	-46	54-50
fi	10	25	30	20	10	5

3- الانحراف المتوسط Mean Deviation

يعرف الانحراف المتوسط بانه مجموع الانحرافات المطلقة لقيم متغير ما عن وسطها الحسابي مقسوما على عدد هذه القيم ويرمز له بالرمز { MD } وهو يحسب بالطرق التالية:

***- في حالة البيانات غير المبوبة:**

لتكن لدينا X_1، X_2، X_3، ... ، X_n تمثل مجموعة من قيم المتغير { X } عددها { n } وان { $\bar{X}$ } هو الوسط الحسابي لهذه القيم فان الانحراف المتوسط:

$$MD = \frac{\sum \mid \boldsymbol{x_i} - \boldsymbol{\bar{x}} \mid}{\boldsymbol{n}}$$

مثال: من البيانات التالية جد الانحراف المتوسط:

$x_i = 2، 3، 4، 5،5، 6، 7، 10، 13، 14، 19$

الحل:

نجد الوسط الحسابي للبيانات

$$\bar{X} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{2+3+....19}{11} = 8$$

نجد الان القيمة المطلقة للانحرافات

$$\mid x_i - \bar{x} \mid$$

كما موضحه في الجدول

$\mid x_i - \bar{x} \mid$
6
5
4
3
3
2
1
2
5
6
11

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

$$\sum \mid x_i - \bar{x} \mid = 6 + 5 + 4 + \cdots . + 11 = 48$$

$$M.D = \frac{\sum \mid x_i - \bar{x} \mid}{n} = \frac{48}{11} = 4.36$$

- **في حالة البيانات المبوبة**

يمكن حساب الانحراف المتوسط لتوزيع تكراري حسب الصيغة التالية:

- $M.D = \frac{\sum f_i \mid X_i - \bar{X} \mid}{\sum f_i}$

حيث ان X_i = مركز الفئة

f_i = تكرار الفئة

$\bar{X}$ = الوسط الحسابي للتوزيع

$$\sum f_i = n$$

مثال: من جدول التوزيع التكراري التالي، جد الانحراف المتوسط:

الفئات	التكرار	X_i	f_iX_i	$\lvert X_i - \bar{X} \rvert$	$f_i \lvert X_i - \bar{X} \rvert$
-30	2	*35*	*70*	*30*	*60*
-40	5	*45*	*225*	*20*	*100*
-50	3	*55*	*165*	*10*	*30*
-60	4	*65*	*260*	*0*	*0*
-70	7	*75*	*525*	*10*	*70*
90-80	6	*85*	*510*	*20*	*120*
	27		*1755*		*380*

1- -نجد الوسط الحسابي للبيانات المبوبة

$$\bar{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{\mathbf{1755}}{\mathbf{27}} = 65 = \bar{X}$$

2- نحسب انحراف ان المطلقة للقيم عن وسطها الحسابي وكما موضحه في الجدول

$\lvert \mathbf{X_i} - \bar{\mathbf{X}} \rvert$

نحسب الانحراف المتوسط

$$M.D = \frac{\sum f_i \lvert X_i - \bar{X} \rvert}{\sum f_i} = \frac{\mathbf{380}}{\mathbf{27}} = \mathbf{14.07}$$