د. وفاء جعفر حسين

المحاضرة الاولى

اولا : تعريف الاحصاء

هنالك العديد من التعاريف لعلم الاحصاء بما يتوافق مع تطور هذا العلم حيث ان كلمة الاحصاء في الماضي كانت تهدف الى العد والحصر حتى سمي علم الاحصاء بعلم العد وبصورة عامة فان علم الاحصاء يعرف على انه ذلك العلم الذي يهتم بجمع البيانات عن الحقائق والظواهر وتبويبها في جداول وتحليلها او هو الطريقة العلمية التي تختص بجمع البيانات عن الظواهر المعينة وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

وينقسم علم الاحصاء الى فرعين :

١. **الاحصاء الوصفي : ويتضمن هذا الفرع الطرق والاساليب المستخدمة لجمع البيانات عن الظواهر المعينة وكيفية تنظيم وتصنيف وتبويب هذه البيانات وعرضها في جداول ورسوم بيانية وحساب المؤشرات الاحصائية منها .**
٢. **الاحصاء الاستدلالي : وهو الذي يهتم بموضوعي التقدير واختبار الفرضيات**

ثانيا : اهمية علم الاحصاء ومجالات تطبيقه

هنالك العديد من المجالات التي يستخدم فيها علم الاحصاء حيث يستعمل في البحوث الزراعية والهندسية والادارية والاقتصادية والطبية والنفسية والتربية وفي البحوث الصناعية.

ثالثا : الطريقة الاحصائية في البحث العلمي

يمكن تلخيص الطريقة الاحصائية بالخطوات الاتية :

١. **تحديد مشكلة البحث او فرضية البحث او الدراسة**
٢. **جمع البيانات والمعلومات عن الظاهرة او الظواهر ذات العلاقة بالبحث والدراسة**
٣. **تصنيف البيات وتبويبها وعرضها**
٤. **حساب المؤشرات الاحصائية كتقديرات لمعالم مجتمع البحث او الدراسة**
٥. **تحليل معطيات الدراسة والتوصل الى النتائج على ضوء فرضية او فرضيات البحث والدراسة**
٦. **تفسير النتائج وعملية اتخاذ القرارات بشأن فرضيات البحث .**

المجتمع الاحصائي:

هو مجموعه من المفردات او الوحدات الاحصائية التي تشترك بصفة او اكثر من الصفات المعنية وغالبا ما يتم الحصول عليها من البيانات والمعلومات المطلوبة وان المجتمع الاحصائي يمكن تقسيمه بصورة عامة الى

١. **المجتمع الاحصائي المحدد**

٢. المجتمع الاحصائي غير المحدد

رابعا الظواهر المقيسة

١. المتغيرات الكمية المستمرة : يمكن ان تأخذ اي قيمة بين قيمتين متتاليتين مثلا ان وزن الطفل يحسب بالكغم كذلك العمر يمكن ان يحسب بالسنوات والايام
٢. المتغيرات الكمية المتقطعة : وهي المتغيرات التي لا يمكن ان تأخذ اي قيمة بين قيمتين متتاليتين مثلا عدد اولاد عائلة معينة لا يمكن ان يكون خمس اولاد ونصف وهكذا

جمع وتصنيف وتبويب البيانات

اساليب جمع البيانات

١. اسلوب الجمع الشامل : يقصد بأسلوب الجمع الشامل جمع البيانات والمعلومات عن كافة افراد المجتمع قيد الدراسة حيث يجب ان يكون المجتمع محددا في هذه الحالة فيتم جمع البيانات والمعلومات عن جمع افراد المجتمع ومثال على ذلك التعداد السكاني ويتمز هذا الاسلوب بأفضليته لكونه يأخذ بنظر الاعتبار جميع مفردات المجتمع غير انه يحتاج الى الجهد والوقت والتكلفة العالية .
٢. اسلوب العينات : ويقصد به عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات المجتمع هذه المجموعة تسمى العينة ومثال على ذلك دراسة تأثير دواء معين على بعض المرضى المصابين وللحصول على الدقة في هذا الاسلوب يجب اختيار العينة بحيث تلائم جميع مفردات المجتمع كذلك الاخذ بنظر الاعتبار حجم العينة المختارة .

انواع العينات

تنقسم العينات بشكل عام الى قسمين هما :

اولا : العينات العشوائية : ويقصد بالعينة العشوائية بانها تلك المجموعة من المفردات المختارة من مجتمع الدراسة بحيث انه ليس للباحث اي دخل في اختيار هذه المفردة دون تلك اي ان هناك مبدا تساوي الفرص لظهور اي مفردة من مفردات المجتمع ضمن هذه العينة وتنقسم العينات العشوائية الى عدة انواع منها

١. العينة العشوائية البسيطة : على فرض ان مجتمع الدراسة متجانس ومحدد وان عدد مفردات المجتمع N وافرض بان هذه الدراسة تتطلب اختيار عينة حجمها n فأن
   - احتمال ظهور كل مفردة من ضمن العينة $\frac{1}{N}$
   - ان عدد العينات الممكنة من هذا المجتمع R بحيث ان

$$R = \binom{\mathrm{N}}{\mathrm{n}} = \frac{N!}{n!\,(N-n)!}$$

- عدد مرات ظهور اي مفردة من مفردات المجتمع في جميع العينات

$$\binom{\mathrm{N}-1}{\mathrm{n}-1} = \frac{(N-1)!}{(n-1)!\,(N-n)!}$$

ثانيا : العينة الطبقية العشوائية

تعتبر العينات المختارة من هذا الاسلوب من افضل انواع العينات واكثرها دقة في تمثيل المجتمع فكثيرا ما نلاحظ مفردات المجتمع الاحصائي غير متجانس من حيث الصفة ذات العلاقة بموضوع البحث فمثلا لو كنا بصدد دراسة المستوى العلمي لطلبة كلية الادارة والاقتصاد فنلاحظ بان المجتمع غير متجانس من حيث التخصص العلمي فهناك مجموعة من الطلبة من اختصاص الادارة والاقتصاد والمحاسبة والاحصاء وعلى فرض ان مجتمع الدراسة مؤلف من N مفردة وان هذا المجتمع يمكن تجزئته الى L من الطبقات وفق معيار متجانس ملائم لتلك الدراسة بحيث ام في كل طبقة متجانسة من المفردات وكما يلي :

$$N = N_1 + N_2 + \cdots + N_L \quad , L = 1, 2, 3 \ldots.$$

كذلك ان وزن الطبقة

$$W_h = \frac{N_h}{N}$$

حيث $h = 1, 2, \ldots, L$ تسلسل لبطبقة

سيكون حجم العينة المطلوب اختيارها من الطبقة h هي $W_h . n$

ثالثا : العينة العشوائية المنتظمة

على فرض ان مفردات مجتمع الدراسة البالغ عددها N مفردة مرتبة وفق ترتيب معين كان يكون ترتيب تصاعدي او تنازلي او وفق اي معيار اخر للترتيب وافرض ان دراسة معينة تتطلب اختيار عينة من المفردات قوامها n مفردة يتم تقسيم مفردات المجتمع المرتبة الى عدد من المجاميع كل مجموعه منها تضم عدد من مفردات المجتمع وبافتراض ان $k = \frac{N}{n}$ متسلسلة في الترتيب فأن هذه المجاميع ستكون

المجموعة الاولى $1 \quad 2 \quad 3 \ldots k$

المجموعة الثانية $k+1, k+2, \ldots, 2k$

المجموعة الثالثة $(n-1)k, \ldots, nk$

الرابعة : العينة المتعددة المراحل

في هذه العينة يتم تقسيم المجتمع الاحصائي الى وحدات تدعى بالوحدات الاولية ثم يتم اختيار عينة عشوائية من هذه الوحدات الاولية كمرحلة اولى يتم تقسيم كل وحدة اولية مختارة الى وحدات اصغر تدعى بالوحدات الثانوية ويتم اختيار عينة عشوائية من الوحدات الثانوية لكل وحدة اولية كمرحلة ثانية تقسم الوحدات الثانوية المختارة الى وحدات اصغر ويختار من كل منها عينة عشوائية كمرحلة ثالثة ويستمر التقسيم والاختيار على نفس الطريقة لحين الوصول الى المفردات التي يتم جمع البيانات منها والتي تؤلف عينة البحث .

ثانيا: العينات غير العشوائية

ويقصد بالعينة غير العشوائية انها تلك المجموعة من المفردات المختارة من مجتمع الدراسة بطريقة يكون للباحث دخل في اختيار هذه المفردة دون تلك وليس على اساس عشوائي وذلك لاعتبارات تتعلق بطبيعة تلك الدراسة ومن هذه العينات

- **العينة الحصصية**
- **العينة العمدية**
- **العينة العنقودية**
- **العينة المزدوجة**

وسائل جمع البيانات

بعد تحديد حجم العينة واسلوب المعاينة الملائم في اختيار مفردات هذه العينة من مجتمع الدراسة يتطلب الامر اختيار الوسيلة الملائمة في جميع البيانات والمعلومات عن الظاهرة او الظواهر المتعلقة بتلك الدراسة وهناك وسائل عديدة في جمع البيانات اهمها ما يلي

١. اسلوب الجمع الشامل
٢. اسلوب الاستمارة الاحصائية

الاستبيان

وهو عبارة عن استمارة يتم من خلالها جمع البيانات والمعلومات عن مفردات مجتمع الدراسة وذلك عن طريق مواجهة الباحث الشخصية للمفردة الاحصائية او عن طريق المراسلة كما هي الحال في التعدادات السكانية

الاخطاء الشائعة في جمع البيانات

يحدث احيانا ان يقع الباحث في بعض الاخطاء لدى جمعه البيانات والمعلومات التي يتطلبها البحث هذه الاخطاء تحدث نتيجة سوء استخدام الطريقة الاحصائية وفيما يلي موجز اهم هذه الاخطاء

- خطا التحيز
- خطا الصدفة

الجداول الاحصائية

فالجداول على نوعين

- الجداول الاولية : وهي جداول تفصيلية توضح الحقائق وتعرض البيانات المتعلقة بظاهرة معينة
- الجداول الثانوية : وهي جداول مختصرة غرضها عرض نتائج ذات اهمية تتعلق ببحث معين ان الجداول تهدف الى ابراز البيانات وتوضيحها في اضيق حيز

تصنيف وتبويب البيانات

ان البيانات المستحصل عليها بخصوص ظاهرة معينة تسمى بالبيانات الخام او البيانات الاولية او البيانات غير المصنفة وهي غالبا ما تكون غير منظمة وفي هذه الحالة يتعذر على الباحث تكوين فكرة عن هذه الظاهرة او تلك التي جمعت عنها البيانات كذلك يتعذر على الباحث الاعتماد بشكلها الغير منظم لأغراض التحليل الاحصائي للوصول الى النتائج المطلوبة لذلك فأن اولى الخطوات الهامة بعد عملية جمع البيانات هي عملية مراجعة وتصنيف وتبويب البيانات .

- مراجعة البيانات
- تصنيف البيانات
- تبويب البيانات

والهدف من عملية التبويب هو ابراز البيانات وتوضيحها في اضيق حيز ممكن وهناك اربعة اشكال رئيسة هي:

- التبويب الزمني
- التبويب الجغرافي
- التبويب الكمي
- التبويب على اساس صفة معينة

المحاضرة الثانية

التوزيعات التكرارية : وهي عبارة عن تلخيص وترتيب البيانات للمتغير العشوائي التي سبق وان صنفت وجمعت مقسمة عدد من المجاميع كل منها تسمى class بالفئة هذه الفئات قد تكون مرتبة تصاعديا او تنازليا حيث طبيعة البيانات ويسمى توزيع عدد قيم الى فئات بالتوزيع التكراري

ملاحظة: قد تكون فئات التوزيع التكراري متساوية بالطول او غير متساوية وفيما يلي اهم مكونات التوزيع التكراري

- المدى الكلي للتوزيع (T.R) يعرف المدى الكلي بانه الفرق ما بين اكبر قيمة واصغر قيمة من المجموعة فاذا رمزت للمدى الكلي بالرمز T.R فان قانون المدى الكلي $T.R = X_L - X_S$ حيث ان اكبر قيمة $X_L =$ و اقل قيمة $X_S =$
- عدد فئات التوزيع M : يمثل عدد المجاميع التي يتألف منها التوزيع التكراري وهناك صيغ تقريبية يمكن من خلالها تحديد عدد الفئات للتوزيع حيث يرمز لعدد الفئات بالرمز M واهم هذه الصيغ

صيغة يول (yole)

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

حيث ان n تمثل عدد المفردات

صيغة سترجس

$$M = 1 + 3.3LOG_{10}n$$

ملاحظة: عند تطبيق تقريب النتائج الى اقرب عدد صحيح

طول الفئة (L) ويسمى احيانا المدى الفئوي ويتمثل ما بين الحد الادنى للفئة وحدها الاعلى . ان طول الفئة يتناسب عكسيا مع عدد الفئات في التوزيع فكلما كبر طول الفئة قل عدد الفئات والعكس صحيح واذا رمزنا لطول الفئة بالرمز L فيمكن تحديد قيمته كالاتي :

$$L = \frac{T.R}{M}$$

حيث ان $T.R$ المدى الكلي

وان M عدد الفئات

ملاحظة : تستخدم هذه الصيغة في ايجاد طول الفئة عندما تكون اطوال الفئات متساوية في التوزيع التكراري

مركز الفئة: يمثل مركز الفئة قيمة من قيم المتغير العشوائي X التي تتوسط المسافة بين الحد الادنى والحد الاعلى للفئة فاذا رمزنا للحد الادنى بالرمز (L.L) والحد الاعلى والحد الاعلى بالرمز (U.L) يمكن استخراج مركز الفئة

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

ملاحظات مهمة حول التوزيع التكراري

١. ليس من الضروري عند تطبيق ان يكون الحد الادنى للفئة الاولى مساوي تماما لأصغر قيمة في المجموعة بل قد يكون اقل منها
٢. ان التوزيع التكراري قد يكون توزيع مغلق او توزيع مفتوح ويقصد بالتوزيع المغلق بانه ذلك التوزيع الذي يمتلك حدا ادنى للفئة الاولى وحدا اعلى للفئة الاخيرة

تكرار الفئة : يمثل تكرار الفئة جزء من مفردات العينة التي تتصف كونها تقع من القيمة العددية ما بين مدى الفئة بحسب ان مجموع هذه الاجزاء يشكل مفردات العينة n . ويرمز الى تكرار الفئة بالرمز f

مثال :القيم التالية تمثل اطوال 80 نباتا من نباتات القطن مقربة الى اقرب سم المطلوب انشاء جدول توزيع تكراري لأطوال هذه النباتات

جدول رقم (1)

80	87	98	81	74	48	79	80
78	82	93	91	70	90	80	84
73	74	81	56	65	92	70	71
86	83	93	65	51	85	68	72
68	86	43	74	73	83	90	35
75	67	72	90	71	76	92	93
81	88	91	97	72	61	80	91
77	71	59	80	95	99	70	74
63	89	67	60	82	83	63	60
75	79	88	66	70	88	76	63

الحل :نتبع الخطوات الاتية

اولا. استخراج المدى

المدى = اعلى قيمة – اقل قيمة

المدى = 99- 35= 64

ثانيا : اختيار وتحديد عدد الفئات اهمها طريقة سترجس

عدد الفئات =1+ (3.3*لوغارتم عدد المفردات)

او نستخدم طريقة يول

عدد الفئات =عدد المفردات$^{4}\sqrt{}$ *2.5

سوف يكون عدد الفئات يساوي 7

ثالثا : ايجاد طول الفئة

يجب ان لا يقل طول الفئة عن مدى التغير مقسوما على عدد الفئات مقربة الى اقرب عدد صحيح

$$\text{طول الفئة} = \frac{\text{مدى التغير}}{\text{عدد الفئات}}$$

طول الفئة =$\frac{64}{7}$ $= 9\frac{1}{7}$ لذا يستحسن ان يكون طول الفئة يساوي *10*

رابعا كتابة حدود الفئات

يجب كتابة حدود الفئات بحيث ان جميع قيم المتغير تقع بين الحد الادنى للفئة والحد الاعلى للفئة الاخيرة . يستحسن ان نبدأ بكتابة الحد الادنى للفئة الاولى بقيمة اصغر مفردة او اقل من ذلك بقليل وتنتهي بالحد الاعلى للفئة الاخيرة بقيمة اكبر مفردة او اكثر من ذلك بقليل

خامسا : استخراج عدد التكرارات لكل فئة

ويتم ذلك بتسجيل القيم الاصلية واحدة بعد الاخرى في الفئة الخاصة كما مبين في الجدول ادناه

جدول رقم(*1*)

الفئات	التكرارات f_i
40-31	*1*
50-41	*2*
60-51	*5*
70-61	*15*
80-71	*25*
90-81	*20*
100-91	*12*
المجموع	*80*

التوزيع التكراري المتجمع : وهو التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المتغير العشوائي وهو على نوعين هما :

- التوزيع التكراري المتجمع الصاعد : وهو التوزيع الذي يبين كمية تراكم التكرارات ابتداء من الفئة الاولى في التوزيع وانتهاء بالفئة الاخيرة وسنرمز للتكرار المتجمع الصاعد لأي فئة بالرمز $F_{\uparrow}$

فان التكرار المتجمع الصاعد للفئة الاولى هو $F_1 = f_1$

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

التكرار المتجمع الصاعد للفئة الثلثة هو $F_3 = f_1 + f_2 + f_3$

$\vdots$

التكرار المتجمع الصاعد للفئة الاخيرة $F_M = f_1 + f_2 + \cdots F_M = n$ التكرار المتجمع الصاعد

مثال : لبيانات جدول رقم (٢) اوجد التكرار المتجمع الصاعد ومركز الفئة

الحل :

الفئات	التكرارات f_i	مركز الفئة	التكرار الصاعد $F_{\uparrow}$
40-31	1	35.5	1
50-41	2	45.5	3
60-51	5	55.5	8
70-61	15	65.5	23
80-71	25	75.5	48
90-81	20	85.5	68
100-91	12	95.5	80
المجموع	80		

- التوزيع التكراري المتجمع النازل : وهو التوزيع الذي يبين تناقص التكرارات ابتداء بالفئة الاولى في التوزيع وانتهاء بالفئة الاخيرة وسنرمز للتكرار المتجمع النازل $F_{\downarrow}$

التكرار المتجمع النازل للفئة الاولى n

التكرار المتجمع النازل للفئة الثانية $n - f_1$

$\vdots$

التكرار المتجمع النازل للفئة الاخيرة $n - f_1 - f_2 - \cdots - f_{m-1}$

مثال : لبيانات جدول رقم (٢) اوجد التكرار المتجمع الصاعد ومركز الفئة

الحل

الفئات	التكرارات f_i	مركز الفئة	التكرار النازل $F_{\downarrow}$
40-31	1	35.5	80
50-41	2	45.5	79
60-51	5	55.5	77
70-61	15	65.5	72
80-71	25	75.5	57
90-81	20	85.5	32
100-91	12	95.5	12
المجموع	80		0

التوزيع التكراري النسبي : ان التوزيع التكراري النسبي هو توزيع تكراري اعتيادي كما سبق توضيحه فيه تكرارات معبر عنها بنسبة مئوية يمكن الحصول عليها من قسمة تكرار كل فئة على مجموع التكرارات الكلية فعلى افتراض ان هنالك توزيع تكراري عدد فئاته *M*وان التكرارات المقابلة لهذه الفئات وحسب تسلسلها f_1, f_2, f_m عندئذ فان التكرارات النسبية المقابلة لهذه الفئات ستكون

$$f^* = \frac{f_i}{n}.100$$

وان n تمثل مجموع التكرارات

ويمكن حساب التوزيع التكراري المتجمع الصاعد النسبي والتوزيع التكراري المتجمع النازل السلبي وذلك من خلال قسمة التكرارات المتجمعة الصاعدة على مجموع التكرارات الكلية فأن

التكرار المتجمع الصاعد النسبي $F^*_{\uparrow} = \frac{F_i}{n}$

التكرار المتجمع النازل النسبي $F^*_{\downarrow} = \frac{F_{\downarrow}}{n}$

مثال : لبيانات جدول رقم (٢) اوجد التكرار المتجمع الصاعد ومركز الفئة

الفئات	التكرارات f_i	التكرار النازل $F_{\downarrow}$	التكرار المتجمع النازل	التكرار الصاعد $F_{\uparrow}$	التكرار المتجمع الصاعد	التكرار النسبي f^*

40-31	1	80	80/80	1	1/80	1/80*100=0.025
50-41	2	79	79/80	3	3/80	1/80*100=2.5
60-51	5	77	77/80	8	8/80	5/80*100=6.25
70-61	15	72	72/80	23	23/80	18.75
80-71	25	57	57/80	48	48/80	31.25
90-81	20	32	32/80	68	68/80	25
100-91	12	12	12/80	80	80/80	15
المجموع	80	0				

العرض الهندسي للبيانات

يتم عرض البيانات بهيئة رسوم بيانية واشكال هندسية متعددة الاشكال والتصاميم والبعض منها بهيئة رسوم تصويرية وهذه الاشكال هي

١. **الاشرطة البيانية** : هي مجموعة من المستطيلات الراسية او الافقية قواعدها متساوية وتمثل الصفة التي تم على اساسها التبويب (سنة ، شهر ، محافظة ...) وارتفاعاتها تمثل البيانات المقابلة لتلك الصفة (درجات الحرارة حسب الشهر السنة ، اكياس المحصول الحنطة حسب المحافظات وتكون على نوعين :

اولا: الاشرطة البيانية المفردة

وهي اشرطة تخص صنف واحد من البيانات المبوبة مثل عدد الطلبة المقبولين في التعليم العالي للسنوات 2000-2005

ثانيا: الاشرطة البيانية المركبة : وهي اشرطة تخص صنفين او اكثر من البيانات المئوية

المستطيل البياني

عبارة عن شكل هندسي يستخدم في تمثيل بيانات ظاهرة معينة يمكن تجزئتها الى عدد من الاصناف المقابلة للتجميع ويتم باختيار مستطيل ذو قاعدة مناسبة وبعد ذلك يتم تمثيل كل صنف من البيانات بمستطيل جزئي داخل المستطيل الكبير بحيث ان مجموع مساحات المستطيلات الجزئية تمثل مساحة المستطيل الاكبر وتتم عملية اختيار المستطيلات الجزئية وفق ما يلي

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

الدائرة البيانية

شكل هندسي مشابه للمستطيل البياني ولكن يتمثل ببيانات الصف بقطاعات داخل دائرة بحيث ان مجموع مساحات القطاعات تمثل مساحة الدائرة

زاوية القطاع = عدد بيانات الصنف /مجموع البيانات* 360

الخط البياني

شكل يبين التغيرات الحاصلة لظاهرة عبر فترة معينة من الزمن

المدرج والمضلع والمنحنى التكراري

عبارة عن اشكال هندسية الغرض منها عرض البيانات المبوبة في جداول توزيع تكراري

اولا المدرج التكراري

عبارة عن مجموعة من المستطيلات تمثل كل منها طول الفئة في التوزيع التكراري وارتفاعاتها تمثل قيمة التكرارات المقابلة لتلك الفئات تكون المستطيلات منفصلة في حالة البيانات المتقطعة ومتصلة في حالة البيانات المستمرة

ثانيا المضلع التكراري

عبارة عن عدد من المستقيمات المتصلة مع بعضها على شكل سلسلة ونقطة اتصال المستقيم بالأخر

ثالثا المنحنى التكراري

مشابه للمضلع ولكن بدل ان تمرر النقاط بمستقيم تمرر بمنحنى يغلق مع المحور السيني بمركز فئة وهمي

ملاحظة: يمكن رسم المدرج والمنحنى والمضلع التكراري معا فتكون نقاط المنحنى او المضلع هو منتصف رؤوس المستطيلات

د. وفاء جعفر

الرموز الإحصائية

1- يرمز عادة لمجموع قيم المتغير بالرمز $\sum_{i=1}^{n} x_i$ فالرمز $\sum$ يسمى(Summation) أي المجموع.

و. (n,1)هما حدا المجموع وعليه فالرمز $\sum_{i=1}^{n} x_i$ يقرا مجموع القيم x مبتدأ من المشاهدة الأولى وحتى الأخيرة إي أن:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

هناك مجموع جزئي مثل:

$$\sum_{i=3}^{5} x_i = x_3 + x_4 + x_5$$

2- يرمز لمجموع مربعات جميع المشاهدات بالرمز:

$$\sum_{i=1}^{n} x^2{}_i = x^2{}_1 + x^2{}_2 + \cdots + x^2{}_n$$

مثال

اذا كانت $x_i = 1,2,3,4$ اوجد $\sum_{i=1}^{4} x_i^2$

الحل $\sum_{i=1}^{4} x_i^2 = x^2{}_1 + x^2{}_2 + x_3^2 + x^2{}_4$

$$= 1^2 + 2^2 + 3^2 + 4^2$$

$$= 1 + 4 + 9 + 16$$

$$= 30$$

3 - ويرمز لمجموع المربعات بالرمز الاتي:

$$(\sum_{i=1}^{n} x_i)^2 = (x_1 + x_2 + \cdots + x_n)^2$$

$$= (1 + 2 + 3 + 4)^2$$

$$= 100$$

من اعلاه يمكن الاستنتاج بأن $\sum_{i=1}^{n} x_i^2 \neq (\sum_{i=1}^{n} x_i)^2$

4 – يرمز لمجموع حاصل ضرب متغيرين بالرمز الاتي:

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

مثال :اذا كانت $y_i = 1,3,5$ $x_i = 2,4,6$ **اوجد** $\sum_{i=1}^{n} x_i y_i$

الحل $\sum_{i=1}^{3} x_i y_i = y_1 x_1 + y_2 x_2 + y_3 x_3$

$$\sum_{i=1}^{3} x_i y_i = 2 * 1 + 4 * 3 + 6 * 5$$

$$= 44$$

5- ويرمز لحاصل ضرب مجموعين لقيم متغيرين بالرمز

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

مثال : اذا كانت $y_i = 1,3,5$ $x_i = 2,4,6$ **اوجد** $\sum_{i=1}^{3} x_i \sum_{i=1}^{3} y_i$

الحل : $\sum_{i=1}^{3} x_i \sum_{i=1}^{3} y_i = (x_1 + x_2 + x_3)(y_1 + y_2 + y_3)$

$= (2 + 4 + 6)(1 + 3 + 5)$

$= 12 * 8 \ = 108$

من هذا نستنتج ان $\sum_{i=1}^{n} x_i y_i \neq (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)$

بعض القواعد في عملية الجمع:

1 – اذا كان c عدد ثابت فأن

فأن $\sum_{i=1}^{n} c = nc$

البرهان :

$$\sum_{i=1}^{n} c = c_1 + c_2 + \cdots + c_n$$

$$= nc$$

مثال: اوجد: $\sum_{i=1}^{4} 5$

$$\sum_{i=1}^{4} 5 = 4 * 5 = 20$$

2 – اذا كان c عدد ثابت

$$\sum_{i=1}^{n} cx_i = c \sum_{i=1}^{n} x_i$$

البرهان:

$$\sum_{i=1}^{n} cx_i = cx_1 + cx_2 + \cdots + cx_n$$

$$= c(x_1 + x_2 +, \dots, x_n)$$

$$= c \sum_{i=1}^{n} x_i$$

مثال : اذا كانت $x_i = 2, 4, 3, 1$ اوجد ناتج $\sum_{i=1}^{4} 5x_i$

الحل : $\sum_{i=1}^{4} 5x_i = 5(x_1 + x_2 + x_3 + x_4)$

$= 5(2 + 4 + 3 + 1) \quad = \quad 50$

3- جمع متغيرين او اكثر هو مجموع جمعهم اي ان

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

البرهان :

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

$$= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n)$$

$$= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

مثال :اذا كان $\sum_{i=1}^{3} x_i = 4$ **و** $\sum_{i=1}^{3} y_i = 2$ **اوجد** $\sum_{i=1}^{3}(2x_i + 3y_i + 7)$

الحل: $\sum_{i=1}^{3}(2x_i + 3y_i + 7) = 2\sum_{i=1}^{3} x_i + 3\sum_{i=1}^{3} y_i + 3 * 7$

$$= 2 * 4 + 3 * 2 + 21 \quad = \quad 35$$

ملاحظة-:يجب أن نفرق بين بعض الرموز الإحصائية مثل:

1 - $\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} = \frac{(x_1 + x_2 + , \ldots , x_n)}{(y_1 + y_2 + , \ldots , y_n)}$

2- $\sum_{i=1}^{n}(\frac{x_i}{y_i}) = \frac{x_1}{y_1} + \frac{x_2}{y_2} + \cdots + \frac{x_n}{y_n}$

$$\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} \neq \sum_{i=1}^{n}(\frac{x_i}{y_i})$$

3- $\sum_{i=1}^{n}(x_i + 2) \neq \sum_{i=1}^{n} x_i + 2$

مقاييس النزعة المركزية

مقاييس النزعة المركزية : وهي تلك المقاييس التي تبحث عن القيمة المركزية القريبة من الدرجة التي يتجمع عندها اكبر عدد من الدرجات من البيانات التابعة لظاهرة معينة حيث تبحث في تقدير القيمة التي تتمركز حولها اغلبية هذه البيانات حيث ان القيمة المتمركزة هي رقم واحد يعبر عن جميع بيانات تلك المجموعة .

وان من اهم مقاييس النزعة المركزية :

- الوسط الحسابي
- الوسط الهندسي
- الوسط التوافقي
- الوسط التربيعي
- الوسيط
- المنوال

سيتم شرح كل مقياس في حالة البيانات المبوبة وغير المبوبة

الوسط الحسابي أو المتوسط :هو القيمة الناتجة من قسمة مجموع القيم على عددها ويرمز له بالرمز $\bar{x}$.

أ-الوسط الحسابي للبيانات غير مبوبة:

نستخدم القانون التالي لايجاد الوسط الحسابي للبيانات غير المبوبة

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

حيث ان n هي عدد القيم او المشاهدات

مثال: البيانات التالية تمثل كمية الامطار الساقطة سنويا (بالملم) على مدينة الموصل خلال فترة خمس سنوات 520,350,450,380,400 ما هو الوسط الحسابي لسقوط الامطار خلال هذه الفترة

الحل :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{520 + 350 + 450 + 380 + 400}{5}$$

$$= \frac{2100}{5} \quad = \quad 420\ mm$$

مثال 2: اذا كانت درجات الحرارة لأربعة ايام متتالية هي 4,6,8,2,اوجد متوسط درجات الحرارة للأيام الاربعة

الحل :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4 + 6 + 8 + 2}{4} = \quad 5$$

ب -الوسط الحسابي للبيانات مبوبة:

يعرف الوسط الحسابي للبيانات المبوبة كالاتي:

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

حيث ان $x_i , i = 1, 2, \dots, n$ تمثل مراكز الفئات في جدزل التوزيع التكراري مع تكراراتها $f_i , i = 1, 2, \dots, n$

مثال : جد الوسط الحسابي لجدول التوزيع التكراري الاتي :

الفئات	التكرار f_i
31- 40	1
41- 50	2
51 - 60	5
61- 70	15
71- 80	25
81 - 90	20
91 -100	12
	$\sum_{i=1}^{n} f_i = 80$

الحل:

نتبع الخطوات الاتية لإيجاد الوسط الحسابي

1 – تعين مراكز الفئات x_i

2- نضرب كل مركز فئة في التكرار المقابل لها $x_i f_i$

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

الحل في الجدول الاتي:

الفئات	التكرار	مركز الفئة(x_i)	مركز الفئة* التكرار$(x_i f_i)$
31- 40	1	35.5	35.5
41- 50	2	45.5	91.0
51 - 60	5	55.5	277.5
61- 70	15	65.5	982.5
71- 80	25	75.5	1887.5
81 - 90	20	85.5	1710.0
91 -100	12	95.5	1146.0

$\sum_{i=1}^{7} x_i f_i = 6130.0$		$\sum_{i=1}^{7} f_i = 80$	

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{6130.0}{80} = 76.62$$

مثال 2: اوجد الوسط الحسابي للجدول الاتي

الفئات	التكرار	مركز الفئة(x_i)	مركز الفئة* التكرار$(x_i f_i)$
60-62	5	61	305
63-65	18	64	1152
66-68	42	67	2814
69-71	27	70	1890
72-74	8	73	584
	$\sum_{i=1}^{5} f_i = 100$		$\sum_{i=1}^{5} x_i f_i = 6745$

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

الوسط التوافقي the harmonic mean

أ- اذا كانت لدينا n من القيم او المشاهدات $x_1, x_2 \dots x_n$ فان الوسط التوافقي لها والذي يرمز له $\bar{H}$ هو مقلوب الوسط الحسابي لمقلوبات القيم ويكتب كالاتي:

$$\bar{\mathrm{H}} = \frac{1}{(\sum_{i=1}^{n} \frac{1}{x_i})/n} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

مثال(1) : اوجد الوسط التوافقي للبيانات الاتية $x_i = 3, 5, 6, 6, 7, 10, 12$

الحل :

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

$$\bar{H} = \frac{7}{\frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{12}} = 5.87$$

ب- الوسط التوافقي للبيانات مبوبة

اذا كانت $x_1, x_2 \dots x_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2 \dots f_n$ على التوالي فأن الوسط التوافقي هو $\bar{H} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$

مثال(2): اوجد الوسط التوافقي لجدول التوزيع التكراري الاتي :

الفئات	التكرار f_i	مراكز الفئات x_i
62-60	5	61
65-63	18	64
68-66	42	67
71-69	27	70
74-72	8	73
المجموع	100	

الحل :

$$\bar{H} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{\sum_{i=1}^{n} f_i}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \frac{f_4}{x_4} + \frac{f_5}{x_5}}$$

$$\frac{100}{\frac{5}{61} + \frac{18}{64} + \frac{42}{67} + \frac{27}{70} + \frac{8}{73}}$$

$$= \frac{100}{1.4855} = 67.3$$

ميزايا وعيوب الوسط التوافقي

فيما يلي اهم ميزات وعيوب الوسط التوافقي

اولا: ميزات الوسط التوافقي

- ***بساط فكرته وخضوعه للعمليات الجبرية***
- ***ان حسابه يستند الى كافة البيانات المتاحة دون استثناء***
- ***لايتاثر كثيرا باخطاء المعاينة***

ثانيا: عيوب الوسط التوافقي

- ***لايمكن تحديد قيمة الوسط التوافقي اذا كانت احدى المشاهدات تساوي صفر او احدى مراكز فئات التوزيع التكراري تساوي صفر***
- ***لايمكن حسابه في حالة فقدان قيمة او اكثر من قيم العينه***
- ***لايمكن حسابه في حالة البيانات الوصفية***
- ***لايمكن تعينه هندسيا***

الوسط التربيعي *The Quadratic Mean*

أ- الوسط التربيعي لبيانات غير مبوبة

اذا كانت لدينا n من القيم او المشاهدات $x_1, x_2 \dots x_n$ فان الوسط التربيعي لها والذي يرمز له $\overline{G}$ هو الجذر التربيعي للوسط الحسابي لمربعات المشاهدات يكتب كالاتي:

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

مثال(3) : ***اوجد الوسط التربيعي للبيانات التالية***

$$x_i = 1, 3, 4, 5, 7$$

الحل:

$$\overline{G} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{x^2{}_1 + x^2{}_2 + x^2{}_3 + x^2{}_4 + x^2{}_5}{n}}$$

$$= \sqrt{\frac{1^2 + 3^2 + 4^2 + 5^2 + 7^2}{5}}$$

$$\sqrt{20} = 4.47$$

أ- اذا كانت الوسط التربيعي لبيانات مبوبة

$x_1, x_2 \dots x_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2 \dots f_n$ على التوالي فأن الوسط التربيعي هو

$$\overline{G} = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i}}$$

مثال(4) : اوجد الوسط التربيعي للتوزيع التكراري للمثال رقم(2)

الحل:

الفئات	التكرار f_i	مراكز الفئات x_i	x_i^2	$f_i x_i^2$
62-60	*5*	*61*	3721	18605
65-63	*18*	*64*	4096	73728
68-66	*42*	*67*	4489	188538
71-69	*27*	*70*	4900	132300
74-72	*8*	*73*	5329	42632
المجموع	*100*			455803

$$\overline{G} = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i}} = \sqrt{\frac{455803}{100}} = \sqrt{4558.03} = 67$$

ميزايا وعيوب الوسط التربيعي

فيما يلي اهم ميزات وعيوب الوسط التوافقي

اولا: ميزات الوسط التربيعي

- *بساط فكرته وخضوعه للعمليات الجبرية*
- *ان حسابه يستند الى كافة البيانات المتاحة دون استثناء*
- *لايتاثر كثيرا باخطاء المعاينة*

ثانيا: عيوب الوسط الربيعي

- *لايمكن حسابه في حالة فقدان قيمة او اكثر من قيم العينة*

- ***لا يمكن حسابه في حالة البيانات الوصفية***
- ***لايمكن تعينه هندسيا***

الوسط الهندسي Geometric mean

من مقاييس النزعة المركزية المهمة هو الوسط الهندسي الذي لأهميته في الدراسات السكانية وخصوصا عند حساب معدلات النمو للسكان

أ- حساب الوسط الهندسي للبيانات غير المبوبة

لتكن $x_2, x_1, \dots, x_n$ تمثل قياسات عينة من المفردات قوامها n بحيث ان $x_i > 0$ ولجميع القيم عندئذ فان الوسط الهندسي لهذه القيم هو الجذر ذي الرتبة n لحاصل ضرب قياسات هذه المجموعة ببعضها البالغ عددها n كما في القانون الاتي:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$$

مثال اوجد الوسط الهندسي للقيم الاتية :

$$x_i = 3, 5, 8, 3, 7, 2$$

الحل:

$$G = \sqrt[6]{(x_1) * (x_1) * \dots * (x_n)}$$

$$G = \sqrt[6]{(3) * (5) * (8) * (3) * (7) * (2)}$$

$$G = \sqrt[6]{5760} = 4.14$$

ب- الوسط الهندسي للبيانات المبوبة

$x_1, x_2 \dots x_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2 \dots f_n$
على التوالي فأن الوسط التربيعي هو

$$\bar{G} = anti - \log_{10}(\frac{1}{\sum_{i=1}^{n} f_i}\sum_{i=1}^{n} f_i log_{10}\, x_i)$$

مثال: اوجد الوسط الهندسي للتوزيع التكراري الاتي

الفئات	التكرار f_i	مراكز الفئات x_i	$logx_i$	$f_i logx_i$
62-60	5	61	1.7782	8.8910
65-63	18	64	1.8062	32.5116
68-66	42	67	1.8261	76.6962
71-69	27	70	1.8451	49.8177
74-72	8	73	1.8633	14.9064
المجموع	100			182.8229

$$\log\bar{G} = (\frac{1}{n}\sum_{i=1}^{n} f_i log_{10}\, x_i) = \frac{182.8229}{100} = 1.8282$$

$$\bar{G} = 67.3$$

المنوال The Mod

أ- لبيانات غير مبوبة

اذا كان لدينا n من المشاهدات يعبر عنها كالاتي $x_1 x_2, \dots, x_n$ فان المنوال لهذه القيم هي القيمة التي تتكرر اكثر من غيرها ويرمز له بالرمز Mo

من هذا يتضح انه يمكن ان يكون منوال واحد للبيانات وتسمى أحادية المنوال او قد يكون لها موالان او اكثر

مثال 1: اوجد المنوال للبيانات الاتية: ١٠،٥ ,4,٣,9,9 ,5,7,8,9,2

الحل : نلاحظ المفردة ٩ تتكرر اكثر من غيرها اذن فان المنوال هو 9

Mo=9

مثال 2: اوجد المنوال للبيانات الاتية : ١, 4,٣,9, 2, 5,7,8

الحل: نلاحظ انه لا توجد قيمة مكررة فان القيم لايوجد لها منوال

مثال ٣: اوجد المنوال للبيانات الاتية : ١,9,١ ٣,4,4, 2, 5,7,8

الحل: نلاحظ ان القيمة ١ تكررت مرتين كذلك القيمة ٤ أيضا تكررت مرتين فان النوال لهذه البيانات هو ١و٤

ب- إيجاد المنوال لبيانات مبوبة.

اذا كانت $x_1, x_2, \dots, x_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2, \dots, f_n$ على التوالي

فالنوال لهذه البيانات سيعطى بالقانون الاتي:

$$Mo = L_1 + \left[\frac{d_1}{d_1 + d_2}\right] * w$$

حيث ان :

فئة المنوال هي الفئة التي تمتلك اكبر التكرارات وأن:

L_1: الحد الحقيقي الأدنى لفئة النوال

d_1: الفرق بين تكرار فئة المنوال والفئة السابقة لها

d_2: الفرق بين تكرار فئة المنوال والفئة اللاحقة لها

w: طول فئة المنوال

مثال : لجدول التوزيع التكراري الاتي اوجد المنوال

الفئات	التكرار f_i:
62-60	5
65-63	18التكرار السابق
68-66	42
71-69	27 التكرار اللاحق
74-72	8
	$\sum_{i=1}^{n} f_i = 100$

الحل: يمكن ادراج الحل بعدة خطوات

أولا: نجد فئة المنوال وهي الفئة ٦٦-68 حيث انها تقابل اكبر تكرار الذي هو ٤٢

ثانيا: نجد L_1 وهي الحد الحقيقي الأدنى لفئة النوال=٦٥,٥

ثانيا: نجد قيمة d_1 وd_2 وكالاتي :

$$d_1 = 42 - 18 = 24$$

$$d_2 = 42 - 27 = 15$$

ثالثا: w وهو طول فئة النوال يمكن استخراجه كالاتي

w :طول فئة النوال = الحد الأعلى لفئة المنوال – الحد الأدنى لفئة المنوال+١

فأن $w = 68 - 66 + 1 = 3$

رابعا: بعد ان أصبحت كل معالم قانون المنوال جاهزة الان نطبق قانون المنوال

$$Mo = L_1 + \left[\frac{d_1}{d_1 + d_2}\right] * w$$

$$Mo = 65.5 + \left[\frac{24}{24 + 15}\right] * 3$$

$$Mo = 67.35$$

مميزات وعيوب المنوال

مزايا المنوال

١- سهولة حسابة وبساطة فكرته

٢- يمكن تعينه هندسيا

٣- لا يتأثر مطلقا بالقيم الشاذة عكس الوسط الحسابي

4- يمكن ايجاده لبيانات متغير وصفي

عيوب المنوال

١- يتاثر بأخطاء المعاينة

٢- لا يستند في حسابه على كافة البيانات المتاحة حيث انه بمجرد معرفة القيمة التي تتكرر اكثر من غيرها يتم معرفته وعندئذ تهمل كافة القيم الأخرى او الفئات الأخرى

الوسيط

أ- لبيانات غير مبوبة

اذا كان لدينا n من المشاهدات يعبر عنها كالاتي $x_1 x_2, \dots, x_n$ لإيجاد الوسيط لهذه القيم نتبع الاتي:

نرتب البيانات تصاعديا او تنازليا

أولا:اذا كانت n عدد فردي

فان الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$

ثانيا : اذا كانت n عدد زوجي فان الوسيط هوالوسط الحسابي للقيمتين التي ترتيبهما $\frac{n}{2}$ و$\frac{n}{2}+1$

مثال : اوجد الوسيط لدرجات طالب في خمسة امتحانات بدرس الإحصاء اذا كانت الدرجات هي : ٨٤,٨٧,٧٦,٨٢,٨٠

الحل :

نرتب الدرجات تنازليا

القيم	ترتيبها
٨٧	١
٨٤	٢
٨٢	٣
٨٠	٤
٧٦	٥

بما ان عدد الأرقام فردي (n=5) اذن فالوسيط هو القيمة التي ترتيبها

$$\frac{n+1}{2} = \frac{5+1}{2} = 3$$

اذن فالوسيط الذي نرمز له Me هو القيمة التي ترتيبها ٣ أي ان $Me = 82$

مثال: اوجد الوسيط للقيم الاتية :

$$x_i = 5, 4, 8, 7, 3, 12, 9, 2$$

الحل:

نرتب الدرجات تصاعديا

القيم	ترتيبها
2	١
3	٢
4	٣
5	٤
7	٥
8	6
9	7
12	8

نلاحظ ان القيم عددها زوجي (n=8) في هذه الحالة سنجد ترتيب القيمة

$\frac{n}{2} = \frac{8}{2} = 4$ **فالقيمة التي ترتيبها ٤ هي ٥ وكذلك نجد ترتيب القيمة**

$\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$ **اذن فالقيمة التي ترتيبها ٥ هي ٧ فان الوسيط ستكون قيمته كالاتي**

$$Me = \frac{5 + 7}{2} = 6$$

ب- إيجاد الوسيط لبيانات مبوبة.

اذا كانت $x_1, x_2, \dots, x_n$ **تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها** $f_1, f_2, \dots, f_n$ **على التوالي**

فالوسيط لهذه البيانات سيعطى بالقانون الاتي:

$$Me = L_1 + \left[\frac{\left(\frac{\sum_{i=1}^{n} f_i}{2}\right) - F_i}{f_i}\right] * w$$

حيث ان :

L_1: **الحد الحقيقي الأدنى لفئة الوسيط**

F_i: *التكرار المتجمع الصاعد عند بداية فئة الوسيط*

f_i: تكرار فئة الوسيط

w: طول فئة الوسيط

مثال : لجدول التوزيع التكراري الاتي اوجد الوسيط

الفئات	التكرار f_i:	التكرار المتجمع الصاعد
62-60	5	5
65-63	18	23
68-66	42	65
71-69	27	92
74-72	8	100
	$\sum_{i=1}^{n} f_i = 100$	

الحل: يمكن ادراج الحل بعدة خطوات

أولا : إيجاد التكرار المتجمع الصاعد

ثانيا : إيجاد ترتيب وفئة الوسيط وهي تساوي وكالاتي $\frac{\sum_{i=1}^{n} f_i}{2} = \frac{100}{2} = 50$

اذن فان فئة الوسيط هي واقعة بين الرقمين 65, 23 في التكرار المتجمع الصاعد فأن فئة الوسيط هي الفئة 68-66

فان الحد الحقيقي الأدنى لفئة الوسيط $L_1 = 65.5$

ثالثا: F_i وهو التكرار المتجمع الصاعد الذي يقابل الفئة التي تسبق فئة الوسيط وهو ٢٣

رابعا : f_i وهو تكرار فئة الوسيط وهو ٤٢

خامسا : w وهو طول فئة الوسيط يمكن استخراجه كالاتي

w :طول فئة الوسيط = الحد الأعلى لفئة الوسيط – الحد الأدنى لفئة الوسيط +١

فأن $w = 68 - 66 + 1 = 3$

سادسا : بعد ان أصبحت كل معالم قانون الوسيط جاهزة الان نطبق قانون الوسيط

$$Me = L_1 + \left[\frac{\left(\frac{\sum_{i=1}^{n} f_i}{2}\right) - F_i}{f_i}\right] * w$$

$$Me = 6.5 + \left[\frac{50 - 23}{42}\right] * 3$$

$$Me = 67.4$$

مميزات وعيوب الوسيط

مزايا الوسيط

١ـ سهولة حسابة وبساطة فكرته

٢ـ يمكن تعينه هندسيا

٣ـ لايتاثر مطلقا بالقيم الشاذه عكس الوسط الحسابي

عيوب الوسيط

١ـ يتاثر باخطاء المعاينة

٢ـ لايستند في حسابه على كافة البيانات المتاحة حيث انه بمجرد معرفة ترتيب الوسيط تحدد قيمته في حالة البيانات غير المبوبة وتهمل بقية القيم الأخرى وكذلك الحال بالنسبة للتوزيعات التكرارية عند علمنا بفئة الوسيط تهمل بقية الفئات الأخرى .

واجبات المنوال

السؤال الأول : للمجموعات الاتية اوجد المنوال

1 - 4,6,8,6,7,5,5,7,6,6,7,8,6,9,6

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

3- ٢،٥،٣،٥،٢،٤٥،٥،١،٩

السؤال الثاني: اوجد المنوال لجدول التوزيع التكراري الاتي:

الفئات	التكرار f_i
٣٧-٣٣	١٠
٤٢-٣٨	١٢
٤٧-٤٣	٥١
٥٢-٤٨	٣٠

٥٣-٥٧	٨
	$\sum_{i=1}^{n} f_i = 111$

واجبات الوسيط

السؤال الأول : الاتي درجات عينة من الطلبة قوامها ٩ طلاب في امتحان معين جد الوسيط لهذه المجموعة

$$x_i = 55, 62, 53, 70, 68, 65, 63, 79, 80$$

السؤال الثاني: الاتي اعمار عينة من الافراد قوامها ١٢ فرد جد الوسيط لعمر الفرد في هذه المجموعة

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

السؤال الثالث: اوجد الوسيط لجدول التوزيع التكراري الاتي:

الفئات	التكرار f_i
٢-٤	٦
٥-٧	٩
٨-١٠	١٢
١١-١٣	٢٠
١٤-١٦	١٤
١٧-١٩	١١
٢٠-٢٢	٨
	$\sum_{i=1}^{n} f_i = 80$

مقاييس التشتت والاختلاف measurement of dispersion of version

ان التشتت هو التباعد او الاقتراب بين قيم المشاهدات التابعة لمتغير معين اما مقاييس التشتت فهي مقاييس تبين مدى تباعد او اقتراب قيم المشاهدات عن وسطها الحسابي حيث انه كلما كان التشتت كبير دل ذلك على عدم تجانسها .وتنقسم مقاييس التشتت الى نوعان

أولا: مقاييس التشتت المطلق وهي التي تكون وحدات قياسها نفس وحدات قياس قيم المشاهدات الاصلية ومن أهمها

١. المدى The range
٢. الانحراف المتوسط The Mean Deviation
٣. التباين والانحراف القياسي The Variance and the standard deviation

ثانيا مقاييس التشتت النسبي وهي المقاييس التي لا تحتوي على وحدات قياس ومن أهمها معامل الاختلاف

مقاييس التشتت المطلق

١. المدى : يعرف المدى بانه الفرق بين اعلى قيمة واقل قيمة ويرمز له بالرمز R أي ان

$$R = X_{max} - X_{min}$$

مثال : اوجد المدى للقيم الاتية

$$a)\ x_i = 12, 6, 7, 3, 15, 10, 18, 5$$
$$b)\ x_i = 9, 3, 8, 8, 9, 8, 9, 18$$

الحل:

$$a) R = X_{max} - X_{min}$$
$$R = 18 - 3 = 15$$
$$b) R = 18 - 3 = 15$$

نلاحظ ان المدى في كلا المجموعتين متساوي رغم ان المجموعة الثانية اكثر تجانسا من حيث احتوائها على الاعداد ٩و٨ ونلاحظ الاختلاف الكبير في المجموعة الأولى وهذا يدل ان المدى لا يعطي صورة واضحة عن الاختلاف في المجموعات كونه يعتمد على القيمتين العظمى والصغرة واحيا تكون هذه القيم من الشواذ لهذا فان هذا المقياس ليس له القدرة على التمييز بين جميع القيم .

٢. الانحراف المتوسط

أولا:للبيانات غير المبوبة

اذا كانت لدينا n من المشاهدات $x_1, x_2, \dots, x_n$ فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة (أي باهمال الإشارة) عن وسطها الحسابي ويرمز له بـ M.D أي ان :

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

اما السبب في اخذ المطلق لان اخذ الإشارات الموجبة والسالبة سيكون المجموع يساوي صفر ملاحظة مهمة $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

مثال : اوجد الانحراف المتوسط للقيم الاتية :

$$x_i = 9, 8, 6, 5, 7$$

الحل:

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
9	2	2
8	1	1
6	-1	1
5	-2	2
7	0	0
$\sum_{i=1}^{n} 35$ $\bar{x} = 7$	0	6

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

$$M.D = \frac{6}{5} = 1.2$$

ثانيا : الانحراف المتوسط لبيانات مبوبة

اذا كانت $x_1, x_2, \dots, x_n$ تمثل مراكز الفئات في جدول توزيع تكراري و التكرارات المقابلة لها هي $f_1, f_2, \dots, f_n$ على التوالي فأن الانحراف المتوسط هو :

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

مثال : لجدول التوزيع التكراري الاتي اوجد الانحراف المتوسط

الفئات	62-60	65-63	68-66	71-69	74-72	المجموع
التكرارات	5	18	42	27	8	100

الحل :

الفئات	التكرار f_i	مركز الفئة x_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
62-60	5	61	305	6.45	32.25
65-63	18	64	1152	3.45	62.10
68-66	42	67	2814	0.45	18.90
71-69	27	70	1890	2.55	68.85
74-72	8	73	584	5.55	44.40
	$\sum_{i=1}^{n} f_i = 100$		6745		226.50

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

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

$$M.D = \frac{226.50}{100} = 22.650$$

مزيا وعيوب الانحراف المتوسط

أولا : المزايا

١- ان حسابه يعتمد على كافة البيانات المتاحة

٢- انه مقياس سهل الفهم والحساب ويخضع للعمليات الجبرية

ثانيا: العيوب

١- اهمال الإشارة السالبة للفروق عند حسابه

٢- لا يمكن حسابه في حالة البيانات الوصفية

٣- تتأثر قيمته بالقيم الشاذة

٤- يتأثر بأخطاء المعاينة

التباين والانحراف القياسي

لكي يتم التغلب على مشكلة الإشارات عند جمع انحرافات مشاهدات أي عينة عن وسطها الحسابي حيث ان الناتج يساوي صفر يمكننا تربيع مجموع انحرافات القيم عن وسطها الحسابي بدلا من اخذ القيمة المطلقة للانحرافات وبذلك تصبح جميعها موجبة فاذا رمزنا مجموع مربعات الانحرافات بالرمز ss وعلى ذلك فان :

$$ss = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

ولكي نأخذ في الاعتبار حجم العينة حتى يمكن مقارنة العينات المختلفة الاحجام فأننا نقسم مجموع مربعات انحرافات القيم على درجة حرية $n-1$ وبذلك سيتم الحصول على ما يسمى بالتباين والذي يرمو له S^2 .

أولا : التباين والانحراف المعياري للبيانات غير المبوبة

اذا كان لدينا n من المشاهدات $x_i = x_1, x_2, \dots, x_n$ فأن التباين ويرمز له S^2 يكون

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

اما الانحراف القياسي s لعينة هو الجذر التربيعي لتباين تلك العينة :

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

مثال: برهن ان $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$

الحل : $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n-1}$

$$= \frac{\sum_{i=1}^{n} x_i^2 - 2\,\bar{x}\sum_{i=1}^{n} x_i + n\,\bar{x}^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - 2\,(\frac{\sum_{i=1}^{n} x_i}{n})\sum_{i=1}^{n} x_i + n\,(\frac{\sum_{i=1}^{n} x_i}{n})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - 2\,\frac{(\sum_{i=1}^{n} x_i)^2}{n} + \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$$

وهذه الطريقة تسمى الطريقة المختصرة لإيجاد التباين او بأخذ الجذر التربيعي يتم الحصول على الانحراف القياسي بالطريقة المختصرة .

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}}$$

مثال : للبيانات الاتية احسب التباين والانحراف المعياري

$x_i = 9, 8, 6, 5, 7$

أولا: الطريقة المطولة

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
9	2	4
8	1	1
6	-1	1
5	-2	4
7	0	0
$\sum_{i=1}^{n} x_i = 35$ $\overline{x} = 7$	0	10

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{10}{4}$$

الانحراف القياسي $\sqrt{S^2} = \sqrt{\frac{10}{4}}$

ثانيا : التباين والانحراف المعياري اذا كانت البيانات مبوبة

اذا كانت $x_1, x_2, \dots, x_n$ تمثل مراكز الفئات في جدول توزيع تكراري و التكرارات المقابلة لها هي $f_1, f_2, \dots, f_n$ على التوالي فأن التباين لها هو :

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

مطلوب برهانه

اما الانحراف القياسي :

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

حيث ان : $S = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i - 1}}$ هي الطريقة المطولة

$S = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2 - \frac{(\sum_{i=1}^{n} f_i x_i)^2}{\sum_{i=1}^{n} f_i}}{\sum_{i=1}^{n} f_i - 1}}$ هي الطريقة المختصرة

مثال : احسب التباين والانحراف المعياري لجدول التوزيع التكراري الاتي وبالطريقتين المطولة والمختصرة علما ان الوسط الحسابي $\bar{x} = 67.45$:

الفئات	62-60	65-63	68-66	71-69	74-72	المجموع
التكرارات	5	18	42	27	8	100

الحل:

أولا: الطريقة المطولة لحل :

الفئات	التكرار f_i	مركز الفئة x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
62-60	5	61	-6.45	41.6025	208.01225
65-63	18	64	-3.45	11.9025	214.2450
68-66	42	67	-0.45	0.2025	8.5050
71-69	27	70	2.55	6.5025	175.5675
74-72	8	73	5.55	30.8025	246.4200
	$\sum_{i=1}^{n} f_i = 100$				852.7500

$S^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i - 1} = \frac{852.7500}{99} = 8.6$

اما الانحراف القياسي $s = \sqrt{S^2} = \sqrt{8.6} = 2.9$

ثانيا الطريقة المختصرة

الفئات	التكرار f_i	مركز الفئة x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
60-62	5	61	305	3721	1860
63-65	18	64	1152	4096	73728
66-68	42	67	2814	4489	188538

69-71	27	70	1890	4900	132300
72-74	8	73	584	5329	42632
	$\sum_{i=1}^{n} f_i = 100$		6745		455803

$$S^2 = \frac{\sum_{i=1}^{n} f_i x_i^2 - \frac{(\sum_{i=1}^{n} f_i x_i)^2}{\sum_{i=1}^{n} f_i}}{\sum_{i=1}^{n} f_i - 1} = \frac{455803 - \frac{6745^2}{100}}{99} = 8.6$$

اما الانحراف المعياري $S = \sqrt{S^2} = \sqrt{8.6} = 2.9$

خواص التباين والانحراف المعياري (القياسي)

اولا: عند اضافة او طرح ثابت c الى قيم مشاهدات العينة فأن قيمة التباين والانحراف المعياري (القياسي) لا يتغيران اي ان

التباين للقيم الجديدة = التباين للقيم الاصلية

الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية

البرهان :

لتكن $y_i = x_i \mp c$ لبرهنة $S_y^2 = S_x^2$

كما هو معلوم ان : $\bar{y} = \bar{x} \mp c$ خواص الوسط الحسابي

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i \mp c - \bar{x} \mp c)^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = S_x^2$$

مثال: اذا علمت ان $S_x^2 = 7$ جد التباين $y = x - 12$

الحل:

$$S_y^2 = S_x^2 = 7$$

ثانيا : اذا ضربت كل قيمة من قيم المشاهدات بعدد ثابت k فان :

التباين للقيم الجديدة = التباين للقيم الاصلية × مربع الثابت

الانحراف المعياري للقيم الجديدة= الانحراف المعياري للقيم الاصلية × الثابت

البرهان :

لتكن $y_i = kx_i$ **لبرهنة** $S_y^2 = k^2\, S_x^2$ **وكذلك** $S_y = kS_x$ **و كما هو معلوم فأن** $\overline{y} = k\overline{x}$ **خواص الوسط الحسابي**

فأن

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1} = \frac{\sum_{i=1}^{n}(kx_i - k\overline{x})^2}{n-1} = k^2\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = k^2 S_x^2$$

$$S_y = kS_x$$

مثال : لتكن $y = 2x$ **علما ان** $S_x^2 = 7$ **اوجد** S_y^2

الحل :

$$S_y^2 = 4 \times 7 = 28$$

هذا ويمكن التعبير عن الخاصيتين السابقتين كما يلي:

$$y_i = kx_i + c$$

حيث ان c, k **هي ثوابت**

البرهان : كما هو معروف $\overline{y} = k\overline{x} + c$ **خواص الوسط الحسابي**

$$y_i = kx_i + c$$

$$y_i - \overline{y} = kx_i + c - \overline{y}$$

$$y_i - \overline{y} = kx_i + c - k\overline{x} - c$$

$$y_i - \overline{y} = kx_i - k\overline{x}$$

$$y_i - \overline{y} = k(x_i - \overline{x})$$

بتربيع الطرفين وادخال رمز الجمع والقسمة على n-1

$$\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1} = k^2\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$S_y^2 = k^2 S_x^2$$

مثال:

اذا كان $y = 3x + 2$ علما ان $S_x^2 = 6$ اوجد S_y^2

الحل: $S_y^2 = 9 \times 6 = 54$

ثالثا: ان قيمة التباين والانحراف المعياري اكبر من الصفر اما اذا كانت قيمة التباين تساوي صفر دل ذلك على ان جميع قيم العينة لها نفس القيمة اي ان $x = x_i = \bar{x}$ ويعزى سبب ان قيمة التباين والانحراف المعياري موجبة لان مجموع المربعات $\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$

رابعا: اذا كان x, y متغيرين مستقلين وكان المتغير z ياوي مجموعهما اي ان $z_i = x_i + y_i$ فأن تباين z يساوي تباين x مضافا له تباين y اي ان :

$$s_z^2 = s_x^2 + s_y^2$$

واجبات خواص التباين والانحراف المعياري (القياسي)

السؤال الاول : اذا علمت ان $\bar{x} = 6$ و $S_x^2 = 10$ اوجد الوسط الحسابي والتباين لكل مما ياتي :

- $y = x + 3$
- $y = 2x$
- $y = 4x + 8$

السؤال الثاني : اذا علمت بان $y = 3x - 4$ وان $\bar{y} = 90$, $S_y^2 = 40$ اوجد الوسط الحسابي والتباين الى قيم x

السؤال الثالث: للبيانات التالية :

$$y_i = 2, 4, 9, 11, 14$$

$$x_i = 2, , 8, 1, 7, 13$$

وكانت $z_i = x_i + y_i$

اوجد الوسط الحسابي والتباين لقيم z

أسئلة محاضرة اليوم (مقاييس التشتت النسبي)

السؤال الاول :

للبيانات الاتية اوجد معامل الاختلاف $x_i = 2, 3, 6, 12, 13$

السؤال الثاني .: البيانات الاتية تمثل جدول توزيع تكراري للاجور الاسبوعية لعمال معمل الطابوق في محافظة واسط

الفئات	التكرار
21-16	15
27-22	16
33-28	25
39-34	20
45-40	10

اوجد معامل الاختلاف لهذا الجدول

السؤال الثالث :حول القيم التالية الى درجات قياسية :

$$x_i = 5, 6, 6, 10, 16$$

السؤال الرابع : برهن ان الوسط الحسابي للدرجات القياسية يساوي صفر والتباين لها يساوي واحد

السؤال الخامس : حول الدرجات القياسية الاتية الى درجاتها الحقيقية

$$Z_i = 7, 8, 6, 9, 10, 3, 4, 5$$

حل واجبات خواص التباين والانحراف المعياري (القياسي)

السؤال الاول : اذا علمت ان $\bar{x} = 6$ و $S_x^2 = 10$ اوجد الوسط الحسابي والتباين لكل مما ياتي :

- $y = x + 3$

$$\bar{y} = \bar{x} + 3$$

$$\bar{y} = 6 + 3$$

$$\bar{y} = 9$$

$$S_y^2 = S_x^2 + S_3^2$$

$$S_y^2 = 10 + 0 = 10$$

- $y = 2x$

$$\bar{y} = 2\bar{x}$$

$$\bar{y} = 2 * 6$$

$$\bar{y} = 12$$

$$S_y^2 = 4 * S_x^2$$

$$S_y^2 = 40$$

- $y = 4x + 8$

$$\bar{y} = 4 * \bar{x} + 8$$

$$\bar{y} = 4 * 6 + 8$$

$$\bar{y} = 32$$

$$S_y^2 = 16S_x^2 + S_8^2$$

$$S_y^2 = 16 * 10 + 0$$

$$S_y^2 = 160$$

السؤال الثاني : اذا علمت بان $y = 3x - 4$ وان $\bar{y} = 90$, اوجد $S_y^2 = 40$
الوسط الحسابي والتباين الى قيم x

الحل : الوسط الحسابي الى x

$$y = 3x - 4$$

$$\bar{y} = 3\bar{x} - 4$$

$$90 = 3\bar{x} - 4$$

$$3\overline{x} = 90 + 4$$

$$\overline{x} = \frac{94}{3}$$

$$y = 3x - 4$$

$$S_y^2 = 3^2 S_x^2 - S_4^2$$

$$40 = 9S_x^2 - 0$$

$$S_x^2 = \frac{40}{9}$$

السؤال الثالث: للبيانات التالية :

$$y_i = 2, 4, 9, 11, 14$$

$$x_i = 2, , 8, 1, 7, 13$$

وكانت $z_i = x_i + y_i$

اوجد الوسط الحسابي والتباين لقيم z

الحل : (الوسط الحسابي)

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{40}{5} = 8$$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{31}{5} = 6.2$$

$$\overline{z} = \overline{x} + \overline{y}$$

$$\overline{z} = 6.2 + 8$$

$$\overline{z} = 14.2$$

التباين

$$\sum_{i=1}^{n} y_i^2 = 2^2 + 4^2 + 9^2 + 11^2 + 14^2 = 418$$

$$s_y^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}{n-1}$$

$$s_y^2 = \frac{418 - \frac{40^2}{5}}{4}$$

$$s_y^2 = 24.5$$

$$\sum_{i=1}^{n} x_i^2 = 2^2 + 8^2 + 1^2 + 7^2 + 13^2 = 287$$

$$s_x^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$$

$$s_x^2 = \frac{287 - \frac{31^2}{5}}{4}$$

$$s_x^2 = 23.7$$

$$s_z^2 = s_x^2 + s_y^2$$

$$s_z^2 = 23.7 + 24.5$$

$$s_z^2 = 48.2$$

مقاييس التشتت النسبي

ان مقاييس التشتت النسبي هي مقاييس خالية من وحدات القياس لذلك يتم استعمالها للمقارنة بين تشتت مجموعتين او اكثر تختلف في وحدات قياسها. ومن اهم مقاييس التشتت النسبي هو :

معامل الاختلاف coefficient of variation

يعبر عن معامل الاختلاف كالاتي :

$$\mathbf{c.v} = \frac{S}{\overline{x}} \times 100$$

حيث ان :

S: الانحراف المعياري للعينة

$\overline{x}$: الوسط الحسابي للعينه

مثال : نتائج الامتحانات النهائية لدرسي الاحصاء والرياضيات كانت كالاتي :

	الاحصاء	الرياضيات
الوسط الحسابي	78	73
الانحراف القياسي	8	76

ففي اي الموضوعين كانت تشتت الدرجات ؟

الحل :

$$\mathbf{c.v} = \frac{S}{\overline{x}} \times 100$$

$$\mathbf{c.v} = \frac{8}{78} \times 100 = 10.25\%$$

$$\mathbf{c.v} = \frac{76}{73} \times 100 = 10.41\%$$

نلاحظ ان التشتت لدرجات الاحصاء كان اقل لذلك فان اداء الطلبة في امتحان الاحصاء كان اكثر تقاربا.

الدرجة المعيارية

في كثير من الاحيان نحتاج تحويل المتغير العشوائي x الى شكل اخر يسمى بالشكل المعياري لهذه القيم وتستعمل للمقارنة بين مجموعتين او اكثر من القيم ذات الاوساط الحسابية والانحرافات المعيارية المختلفة اقيامها

فاذا كانت $x_1, x_2, \dots, x_n$ تمثل قياسات المفردات لعينة قوامها n

مفردات لعينة وان $\overline{x}$ هو الوسط الحسابي للعينة وS هو الانحراف المعياري لها عندئذ يمكن تعريف الدرجة المعيارية zلآية قيمة من قيمة من قيم x كالاتي :

$$z_i = \frac{x_i - \bar{x}}{s}$$

حيث يلاحظ ان المتغير x خالي من وحدات القياس

خواص الدرجة المعيارية :

اولا : الوسط الحسابي للدرجة المعيارية يساوي صفر

البرهان:

$$z_i = \frac{x_i - \bar{x}}{s}$$

$\sum_{i=1}^{n} z_i = \frac{\sum_{i=1}^{n} x_i - \bar{x}}{s}$ بادخال رمز الجمع على الطرفين

$\frac{\sum_{i=1}^{n} z_i}{n} = \frac{\frac{\sum_{i=1}^{n} x_i - \bar{x}}{s}}{n}$ بقمة الطرفين على n

$\bar{z} = 0$ لان مجموع انحرافات القيم عن وطها الحسابي تساوي صفر

ثانيا : التباين للدرجات المعيارية يساوي واحد

البرهان:

$$S_Z^2 = \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n-1} = \frac{\sum_{i=1}^{n} z_i^2}{n-1} \qquad \bar{z} = 0$$

$$z_i = \frac{x_i - \bar{x}}{s} \rightarrow z_i^2 = \frac{(x_i - \bar{x})^2}{s^2}$$

$$\therefore S_Z^2 = \frac{1}{n-1} * \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{s^2} = \frac{(n-1)s^2}{(n-1)s^2} = 1$$

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

حيث ان $\sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)s^2$

مثال :كانت درجات خمسة طلاب في مادتي الاحصاء والرياضيات كالاتي :

مادة الرياضيات x : 81, 72, 69, 52, 50

مادة الاحصاء y : 86, 85, 82, 77, 62

يطلب حساب الدرجات المعيارية لكل موضوع علما ان لديك المعلومات الاتية $\bar{x} =$

$$64.8 \quad ,S_x = 11.9 \quad , \quad \bar{y} = 78.4 \quad , \quad S_y = 8.78$$

الحل:

$$z_x = \frac{x_i - \bar{x}}{s_x}, \quad z_y = \frac{y_i - \bar{y}}{s_y}$$

$$z_x = 1.36 \quad , \quad 0.60 \quad ,0.35, \quad -1.07 \, , -1.24$$

$$z_y = 0.87 \quad , \quad 0.75 \quad ,0.41, \quad -0.16 \, , -1.08$$

نلاحظ ولاغراض المقارنة ان الطالب الاول كانت درجته في مادة الاحصاء 86 افضل منها في الرياضيات 81 كمقارنة مطلقة الا انه من حيث الاهمية فأن درجته في الرياضيات هي في الحقيقة افضل منها في الاحصاء حيث ان الدرجة المعيارية في الرياضيات كانت اعلى من الدرجة المعيارية في الاحصاء .

السؤال الثاني :

احسب العزم الرابع حول الوسط الحسابي لجدول التوزيع التكراري الاتي

الفئات	التكرار
62-60	5
65-63	18
68-66	43
71-69	27
74-72	8

العزوم Moments

العزم الرائي حول الصفر The rth moment about zero

اولا: بيانات غير مبوبة

اذا كانت لدينا n من المشاهدات التابعة للمتغير x فأن العزم الرائي حول الصفر هو

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

فالعزم الاول حول الصفر هو الوسط الحسابي

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

فالعزم الثاني حول الصفر هو

$$\overline{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

فالعزم الثالث حول الصفر هو

$$\overline{x}^3 = \frac{\sum_{i=1}^{n} x_i^3}{n}$$

$$\overline{x}^3 = \frac{\sum_{i=1}^{n} x_i^3}{n}$$

وهكذا

ثانيا :بيانات مبوبة

اذا كانت $x_1, x_2, \dots, x_n$ تمثل مراكز الفئات في جدول توزيع تكراري و التكرارات المقابلة لها هي $f_1, f_2, \dots, f_n$ على التوالي فأن العزم الرائي حول الصفر لها هو :

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

العزم الرائي حول الوسط الحسابي

اولا : بيانات غير مبوبة

العزم الرائي حول الوسط الحسابي هو :

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

فاذا كانت $r = 1$ فأن

$$m_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^1}{n} = 0$$

اما اذا كانت $r = 2$ فأن

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

وهكذا

ثانيا : بيانات مبوبة

العزم الرائي حول الوسط الحسابي هو :

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

مثال (1)

اوجد العزم الاول والثاني والثالث حول الصفر للبيانات التالية

$$x_i = 4, 7, 5, 9, 8, 3, 6$$

الحل :

العزم الاول حول الصفر

$$\bar{x}^1 = \frac{\sum_{i=1}^{n} x_i^1}{n} = \frac{4+7+5+9+8+3+9}{7} = \frac{42}{7} = 6$$

العزم الثاني حول الصفر

$$\bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{4^2+7^2+5^2+9^2+8^2+3^2+9^2}{7} = 40$$

العزم الثالث حول الصفر

$$\overline{x}^3 = \frac{\sum_{i=1}^{n} x_i^3}{n} = \frac{4^3 + 7^3 + 5^3 + 9^3 + 8^3 + 3^3 + 9^3}{7} = 288$$

مثال (2)

اوجد العزم الاول والثاني والرابع حول الوسط الحسابي لنفس البيانات

الحل:

$$m_r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^r}{n}$$

فالعزم الاول حول الوسط الحسابي يساوي صفر

$$m_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^1}{n} = \frac{0}{n} = 0$$

اما العزم الثاني حول الوسط الحسابي فهو

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$
$$= \frac{(4-6)^2 + (7-6)^2 + (5-6)^2 + (9-6)^2 + (8-6)^2 + (3-6)^2 + (9-6)^2}{7}$$
$$= 4$$

اما العزم الرابع حول الوسط الحسابي فهو

$$m_4 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{n}$$
$$= \frac{(4-6)^4 + (7-6)^4 + (5-6)^4 + (9-6)^4 + (8-6)^4 + (3-6)^4 + (9-6)^4}{7}$$
$$= 25.8$$

الواجب البيتي

السؤال الاول : من البيانات الاتية

$$x_i = 5, 8, 3, 2, 6, 4, 9$$

اوجد

- **العزم الاول والثاني والثالث والخامس حول الصفر**
- **العزم الاول والثالث والرابع حول الوسط الحسابي**

السؤال الثاني :

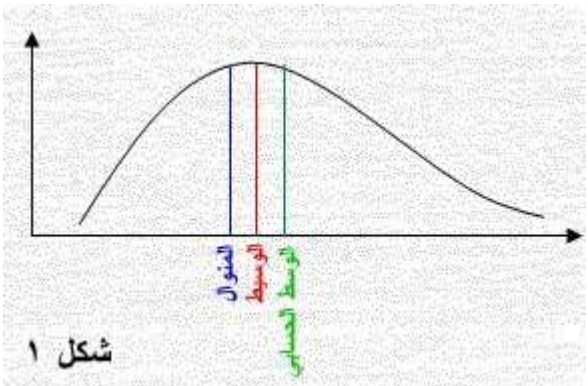
احسب العزم الرابع حول الوسط الحسابي لجدول التوزيع التكراري الاتي

الفئات	التكرار
62-60	5
65-63	18
68-66	43
71-69	27
74-72	8

وهي مقاييس تحدد شكل المنحنى من حيث التماثل والالتواء Skewness ومن حيث جهة تدبب القمة وتفلطحها Kurtosis

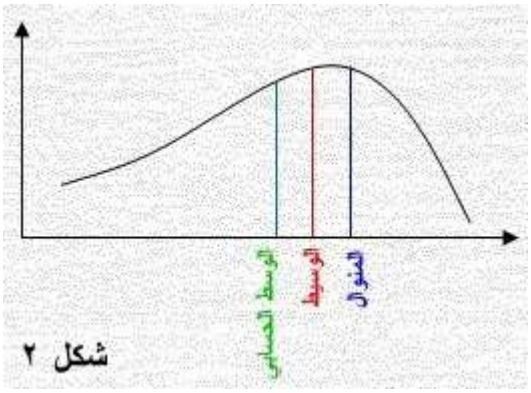
مقاييس الالتواء Measures of Skewness

التوزيع التكراري المتماثل ما كانت تكراراته موزعة توزيعاً متماثلاً حول وسطه الحسابي بما يعني قيم المتغير المتساوية البعد عن الوسط الحسابي لها نفس التكرار فالتواء توزيع يعني مدى بعده عن التماثل وقد يكون الالتواء جهة اليمين (موجب الالتواء) أو الالتواء جهة اليسار (سالب الالتواء) والأشكال الثلاثة التالية تبين ذلك



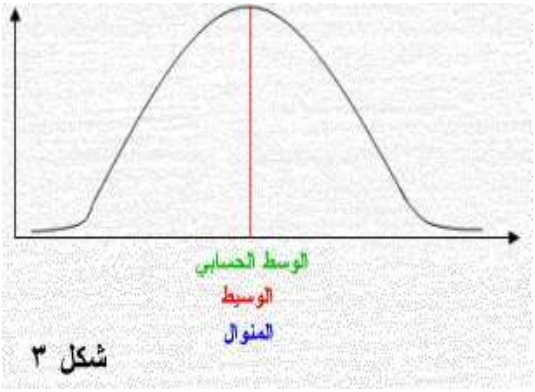
شكل رقم (١) منحني ملتوي التواء موجب نحو اليمين

اما منحني التوزيع ذو الالتواء السالب فأن مفرداته تتمركز في الجهة اليمنى (عند الفئات العليا)بينما طرفه يمتد الى اليسار كما في الشكل التالي :



شكل رقم (٢) منحني ملتوي التواء سالب نحو اليسار

اما الشكل الاتي يمثل التوزيع المتماثل



شكل رقم (٣) التوزيع المتماثل

وبافتراض أن هذه الأشكال تمثل درجات طلاب في امتحان ما، ففي الشكل ١ يعني عدد كبير من الطلاب حصلوا على درجات أقل من المتوسط بمعنى أن مستوى الطلاب أقل من مستوى الامتحان أو أن الامتحان صعب في حين العكس تماماً في الشكل ٢ حيث يبن أن عدد كبير من الطلاب حصلوا على درجات أكبر من المتوسط أي أن مستوى الطلاب أعلى من مستوى الامتحان أو أن الامتحان سهل في حين الشكل ٣ يعنى مستوى الطلاب مناسب لمستوى الامتحان
ومن اهم مقاييس الالتواء

اولا: باستخدام المنوال
ثانيا : باستخدام الوسيط
ثالثا: باستخدام العزوم

وهي مقاييس نسبية يمكن استخدامها للمقارنة بين التواء التوزيعات وهذا وفي جميع هذه الطرق يكون الالتواء موجبا عندما يكون معامل الالتواءموجبا وسالبا عندما يكون معامل الالتواء سالبا ومتماثلا عندما يكون معامل الالتواء صفرا وتتراوح قيمة معامل الالتواء بين (3-و3) .

اولا معامل الالتواء باستخدام الوسيط

$$\text{معامل الالتواء الاول} = \frac{\text{الحسابي الحسابي}-\text{النوال}}{\text{الانحراف القياسي}}$$

$$\alpha_1 = \frac{(\overline{Y} - \overline{MO})}{S}$$

ومن هذا يتضح اذا كانت قيمة الوسط الحسابي اكبر من قيمة المنوال كانت النتيجة موجبة وبالتالي فان الالتواء يكون موجبا كما في الشكل رقم ١ اماذا كانت قيمة الوسط الحسابي اقل من قيمة المنوال كانت النتيجة سالبة وعندئذ

فان الالتواء سيكون سالبا كما في الشكل رقم ٢ اما اذاكانت قيمة الوسط الحسابي تساوي قيممة المنوال فان قيمة معامل الالتواء تساوي صفر والتوزيع يكون متماثلا كما في الشكل رقم ٣

ثانيا معامل الالتواء باستخدام الوسيط

$$\text{معامل الالتواء الثاني} = \frac{\text{الحسابي الحسابي}-\text{الوسيط}}{\text{الانحراف القياسي}}$$

$$\alpha_2 = \frac{(\overline{Y} - \overline{Me})}{S}$$

ومن هذا يتضح اذا كانت قيمة الوسط الحسابي اكبر من قيمة الوسيط كانت النتيجة موجبة وبالتالي فان الالتواء يكون موجبا اما اذا كانت قيمة الوسط الحسابي اقل من قيمة الوسيط كانت النتيجة سالبة وعندئذ فان الالتواء سيكون سالبا اما اذا كانت قيمة الوسط الحسابي تساوي قيمة الوسيط فان قيمة معامل الالتواء تساوي صفر والتوزيع يكون متماثلا .

ثالثا معامل الالتواء باستخدام العزوم

$$\text{معامل الالتواء الثالث} = \frac{\text{العزم الثالث حول الوسط الحسابي}}{\text{مكعب الجذر التربيعي للعزم الثاني حول الوسط الحسابي}}$$

$$\alpha_3 = \frac{m_3}{(\sqrt{m_2})^3}$$

اذا كانت النتيجة موجبة وبالتالي فان الالتواء يكون موجبا اما اذا كانت النتيجة سالبة وعندئذ فان الالتواء سيكون سالبا اما اذا كانت قيمة الوسط الحسابي تساوي قيمة الوسيط فان قيمة معامل الالتواء تساوي صفر والتوزيع يكون متماثلا

مثال: احسب معامل الالتواء الاول والثاني لجدول توزع التكراري اذا علمت ان :

$$y = 79.76 \quad , Me = 79.06 \quad , Mo = 77.50 \ , S = 15.60$$

$$\alpha_1 = \frac{\bar{y}-\bar{M}o}{S} = \frac{79.76\ -77.50}{15.60} = 0.14$$

$$\alpha_2 = \frac{\bar{y}-\bar{M}e}{S} = \frac{79.76\ -79.06}{15.60} = 0.13$$

بما ان α_1 و α_2 موجبتان فان المنحني ملتوي التواء موجب

مثال لجدول التوزيع التكراري التالي احسب معامل الالتواء الثالث

الفئات	التكرار
62-60	5
65-63	18
68-66	42
71-69	27
74-72	8

الحل:

$$m_2 = 8.5275$$

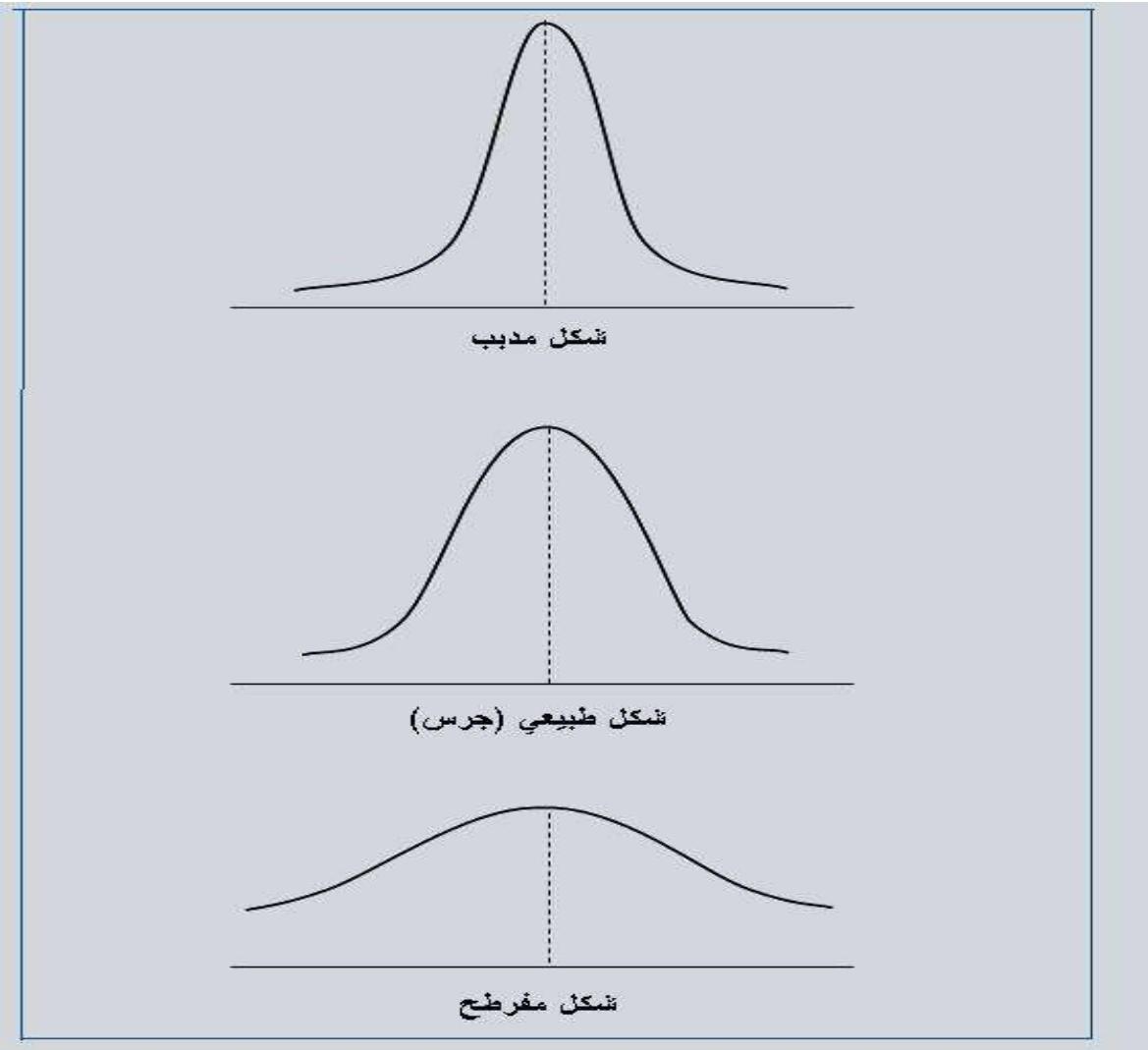
$$m_3 = -2.6932$$

$$\alpha_3 = \frac{m_3}{(\sqrt{m_2})^3} = \frac{-2.6932}{(\sqrt{8.5275})^3} = -0.14$$

بما ان قيمة α_3 سالبة فان المنحنى ملتوي التواء سالب

مقاييس التفلطح Measures of Kurtosis

لا حضنا في دراستنا لموضوع التشتت وجود مقاييس توضح لنا درجة تجانس مجموعة من القيم سميناها بمقاييس التشتت الهدف منها اعطاء صورة اكثر تكاملا عن طبيعة البيانات الى جانب مقاييس النزعة المركزية ان لتشتت قيم توزيع معين اثر كبير على هيئة المنحني التكراري لذلك التوزيع فكلما كان تشتت القيم كبير فان ذلك يعني تسطح المنحنى في حين كلما كان تشتت القيم قليلا فان ذلك يعني تحدب المنحنى التكراري . وعلى هذا الاساس اقترح العالم كارل بيرسون مقياس يبين مقدار تحدب او تسطح المنحنى التكراري اسماه مقياس التفلطح حيث يعرف التفلطح بانه مقدار تسطح او تدبب المنحنى التكراري لتوزيع معين



ان الصيغ التي اقترحها العالم كارل بيرسون في قياس التفلطح هي :

$$\beta_2 = \frac{m_4}{m_2^2} = \frac{(x_i - \overline{x})^4}{((x_i - \overline{x})^2)^2}$$

$$\gamma_2 = \beta_2 - 3$$

ان الشكل المثالي للتفلطح (اي ان المنحنى ليس كثير التسطح ولا كثير التحدب)يتحقق عندما $\gamma_2 = 0$ وهذا يحصل عندما $\beta_2 = 3$ اما اذا كانت $\beta_2 > 3$ فذلك يعني $\gamma_2 > 0$ وفي هذه الحالة يقال ان المنحنى مدبب وتزداد درجة تدبب المنحنى بزيادة γ_2نحو الجانب الموجب

للقيم اما اذا كانت $\beta_2 < 3$ فذلك يعني ان $\gamma_2 < 0$ في هذه الحالة يقال ان المنحنى التكراري مسطح وتزداد درجة التسطح بزيادة قيمة γ_2 نحو الجانب السالب للقيم

ان مقدار تدبب او تسطح المنحنى التكراري لتوزيع معين يرتبط ارتباطا وثيقا بمقياس تشتت قيم ذلك التوزيع فكلما كان التشتت عال فذلك دلالة على تسطح المنحنى التكراري في حين كلما كان التشتت واطيء فذلك دلالة على تدبب المنحنى وفيما يلي موجز مما تقدم

اذا كانت $\beta_2 = 0$ سميت القمة معتدلة

اما اذا كانت $\beta_2 > 0$ سميت القمة مدببة

اما اذا كانت $\beta_2 < 0$ سميت القمة مفلطحة

مثال:اذا علمت ان العزم المركزي الرابع في توزيع معين كان 170 والعزم المركزي الثاني كان 7 حدد شكل تفلطح المنحني التكراري لهذا التوزيع

الحل :

$$\beta_2 = \frac{m_4}{m_2^2} = \frac{170}{49} = 3.47$$

$$\gamma_2 = \beta_2 - 3 = 3.47 - 3 = 0.47$$

وهذا يعني ان منحنى هذا التوزيع مدبب

واجب بيتي

السؤال الاول : لجدول التوزيع التكراري الاتي

الفئات	التكرارات
126-118	3
135-127	5
144-136	9
153-145	12
162-154	5
171-163	4
180- 172	2

اوجد

اولا: معامل الالتواء الاول والثاني

ثانيا: معامل التفلطح

السؤال الثاني: اذا كان العزم الثاني حول الوسط الحسابي لتوزيعين هما 9 و 16 على التوالي بينما العزم الثالث حول الوسط الحسابي للتوزيعين هما (8.1-)و(12.8-) على التوالي فأي التوزيعين اكثر التواء لليسار اكثر من الاخر

الارتباط الخطي Linear correlation

ان مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين او اكثر ترتبط مع بعض بعلاقات خطية معينة على سبيل المثال العلاقة بين طول الطالب ووزنه او العلاقة بين شفاء المريض وكمية الجرعة الدوائية او العلاقة بين مستوى الطالب الدراسي ودخل الاسرة المعاشي وغيرها من الامثلة الاخرى . فعندما يكون المتغير يؤثر على متغير اخر او عدة متغيرات يقال ان المتغيرات مرتبطة فيما بينها . واذا كان المتغيرين المرتبطين(او مجموعة المتغيرات المرتبطة تتغير) بنفس الاتجاه اي زيادة او نقصان في الاخر عندئذ يقال ان الارتباط بينهما موجب على سبيل المثال انخفاض دخل الفرد يؤدي الى انخفاض في انفاقه على السلع اما ذا كانت المتغيرات او المتغيرين المرتبطين تتغير باتجاه معاكس اي زيادة في احدهما يؤدي الى نقصان في المتغير الاخر او المتغيرات الاخرى او نقصان في متغير يؤدي الى زيادة في المتغير الاخر عندئذ يقال ان الارتباط بينهما ارتباط سالب على سبيل المثال ان انخفاض درجات الحرارة يؤدي الى ارتفاع اسعار الوقود

ويتم حساب الارتباط بمعامل يدعى معامل الارتباط correlation coefficient وعليه يعرف معامل الارتباط بانه درجة او قيمة العلاقة التي تربط متغيرين او اكثر مع بعض وهي قيمة حقيقية خالية من وحدات قياس المتغيرات المرتبطة بهذه العلاقة . والارتباط على انواع عديدة وفيما يلي عرض لكل نوع من هذه الانواع .

اولا : الارتباط الخطي البسيط

يعرف الارتباط الخطي البسيط بانه القيمة العددية للعلاقة بين متغيرين ويحسب وفق معامل الارتباط الخطي البسيط

وفق الصيغ الاتية :

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

حيث يمكن اختصار الصيغة اعلاه بالصيغة الاتية :

$$r_{x,y} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2]\,[\;n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

مثال :للبيانات الاتية جد معامل الارتباط البسيط

xy	x^2	y^2	x	y
6	4	9	2	3
10	4	25	2	5
35	25	49	5	7
32	16	64	4	8
45	25	81	5	9
66	36	121	6	11
18	9	36	3	6
40	25	64	5	8
24	16	36	4	6
276	160	485	36	63

$$r_{x,y} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2]\,[\;n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

$$r_{x,y} = \frac{(9)(276) - (36)(63)}{\sqrt{[9(160) - (36)^2\,] * \;[9(485) - (63)^2\,]}}$$

$$r_{x,y} = \frac{24}{\sqrt{[16] * [44]}} = 0.905$$

خصائص معامل الارتباط البسيط:

اولا :قيمة معامل الارتباط البسيط تتراوح بين (-1,1)

ثانيا : ان اية تحويلات خطية على قيم المتغيرات x, y لا تؤثر على قيمة $r_{x,y}$

معامل الارتباط الجزئي Partial correlation coefficient

يلاحظ في بعض الاحيان وجود ارتباط بين متغيرين يعزى جزئيا الى ارتباط متغير ثالث مرتبط مع كلا المتغيرين لنفرض ان المتغيرين x_1وx_2 مرتبطين وهنالك متغير اخر مثل x_3 مرتبط مع كلا المتغيرين x_1وx_2 واننا نرغب في قياس درجة الارتباط بين x_1وx_2 باستبعاد اثر x_3 على كلا المتغيرين . ان المعامل الذي يقيس درجة ارتباط متغيرين باستبعاد اثر متغير ثالث يدعى بمعامل الارتباط الجزئي .

ويمكن ايجاد معامل الارتباط الجزئي وفق ما يلي

افرض ان x_1وx_2وx_3 تمثل ثلاث متغيرات عشوائية بحيث ان المتغيرين x_1, x_2مرتبطين وان x_3 مرتبط مع كل من x_1, x_2 وافرض انه على اساس عينة عشوائية من المفردات قوامها n وان معامل الارتباط البسيط بين متغيرين منها هو و$r_{1,2}$و$r_{1,3}$ و $r_{2,3}$ عند ئذ فان معامل الارتباط الجزئي بين المتغيرين x_1, x_2 باستبعاد المتغير x_3 هو :

$$r_{12.3} = \frac{r_{12} - r_{13} * r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

مثال: افرض ان X_1 يمثل دخل الاسرة الشهري وان X_2 يمثل انفاق الاسرة الشهري وان X_3 يمثل عدد افراد الاسرة افرض لعينة قوامها *20* اسرة تم الحصول على ما يلي .

$$r_{12} = 0.91, \quad r_{13} = 0.39 \quad , \quad r_{23} = 0.62$$

جد معامل الارتباط بين دخل الاسرة وانفاقها باستبعاد اثر عدد افراد الاسرة

الحل: ان المطلوب بالسؤال ايجاد $r_{12.3}$

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$r_{12.3} = \frac{0.91 - (0.39)\,(0.62)}{\sqrt{(1 - 0.39^2\,) * (1 - 0.62^2)}} = 0.925$$

مثال : اذا علمت ان $r_{12} = r_{13} = r_{23} = r$ برهن ان

$$r_{12.3} = r_{13.1} = r_{23.1} = \frac{r}{r+1}$$

الحل :

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$= \frac{r - r.r}{\sqrt{(1 - r^2)(1 - r^2)}}$$

$$= \frac{r(1-r)}{1 - r^2} = \frac{r(1-r)}{(1-r)(1+r)} = \frac{r}{r+1}$$

ملاحظة

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{32}}{\sqrt{(1 - r_{12}^2)(1 - r_{32}^2)}}$$

$$r_{23.1} = \frac{r_{23} - r_{21}\, r_{31}}{\sqrt{(1 - r_{21}^2)(1 - r_{31}^2)}}$$

واجب بيتي كمل حل

$$r_{23.1} = \frac{r}{r+1}$$

$$r_{13.2} = \frac{r}{r+1}$$

معامل الارتباط المتعدد Multiple correlation coefficient

يلاحظ في كثير من الاحيان ان التغير الذي يطرا على ظاهرة معينة ناجم عن تغير مجموعة ظواهر اخرى مجتمعة وليس ظاهرة واحدة فقط على سبيل المثال التغير في انتاجية الدونم الواحد من الحنطة يتأثر بعدة عوامل مثل نوع البذور او نوع السماد ونوع التربة وهكذا ان المعامل الذي يبين نوع العلاقة بين متغير واحد مع عدة متغيرات يسمى معامل الارتباط المتعدد

على افتراض ان X_1وX_2وX_3 ثلاثة متغيرات عشوائية نرغب في قياس درجة العلاقة بين X_1 مع X_2وX_3 فان ذلك يتم وفق الصيغتين وكالاتي :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}.r_{13}.r_{23}}{1 - r_{23}^2}} \qquad r_{23} \neq \mp 1$$

$$R_{1.23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13.2}^2)}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

مثال : افرض ان X_1 يمثل الانفاق الشهري للاسرة وX_2 يمثل دخل الاسرة الشهري X_3 يمثل عدد افراد الاسرة وعلى اساس عينة عشوائية من الاسر قوامها 20 تم الحصول على ما يلي

$$r_{12} = 0.91 \text{ و } r_{13} = 0.39 \text{ و } r_{23} = 0.62$$

يطلب حساب معامل الارتباط المتعدد بين X_1 مع X_2وX_3

الحل : باستخدام الصيغة الاولى

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}.r_{13}.r_{23}}{1 - r_{23}^2}}$$

$$R_{1.23} = \sqrt{\frac{(0.91)^2 + (0.39)^2 - 2 * 0.91 * 0.39 * 0.62}{1 - (0.62)^2}} = 0.937$$

باستخدام الصيغة الثانية :

$$R_{1.23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13.2}^2)}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$r_{13.2} = \frac{0.39 - 0.91 * 0.62}{\sqrt{(1 - (0.91)^2)(1 - (0.62)^2)}} = -0.536$$

$$R_{1.23} = \sqrt{1 - (1 - (0.91)^2)(1 - (-0.536)^2)} = 0.937$$

خصائص معامل الارتباط المتعدد

اولا: ان قيمته تتراوح بين الصفر والواحد $0 \le R \le 1$

ثانيا : ان قيمته لا تقل عن قيمة اي معامل ارتباط بسيط بين اي متغيرين دخلت في حساب قيمته اي ان, $R_{1.23} \ge r_{12}\, r_{13}\, r_{23}$

ثالثا:اذا كانت قيمته مساوية للصفر فذلك يعني ان كافة معاملات الارتباط البسيطة والجزئية التي يدخل المتغير x_1 في حسابها يساوي صفر.

رابعا: ان قيمته تتزايد بزيادة عدد المتغيرات التي تؤثر في المتغير x_1 اي انه اذا كانت $x_1, x_2, x_3, x_4 x_5,$ متغيرات تؤثر على x_1 فأن

$$R_{1.23} \le R_{1.234} \le R_{1.2345}$$

ارتباط الرتب *Rank Correlation*

ان الصيغ السابقة الخاصة لحساب معامل الارتباط (البسيط, الجزئي, المتعدد) تستند على بيانات من النوع الكمي (اي متغيرات ممكنة القياس بوحدات قياس مألوفة كمقياس الطول والوزن) الا انه من الناحية العملية هناك الكثير من الحالات التي تكون فيها المتغيرات من النوع الوصفي(اي متغيرات غير قابلة للقياس مثل الحالة الاجتماعية او المهنة وهكذا)

عندئذ وبهدف حساب الارتباط بين هذه المتغيرات من هذا النوع يتم اجراء تعديلات على الصيغ السابقة بالشكل الذي يمكننا من استخدامها في هذه الاحوال ان معامل الارتباط الذي يقيس درجة الارتباط بين صفتين يسمى يسمى معامل الارتباط لسبيرمان

معامل ارتباط الرتب لسبيرمان

افرض ان X, Y متغيرين من النوع الوصفي على اساس عينة عشوائية من المفردات قوامها n بهيئة صفات غير قابلة للقياس الكمي وهي ممكنة الترتيب تصاعديا او تنازليا عندئذ فان معامل ارتباط سبيرمان بين بين هاذين المتغيرين هو

$$r_{x,y} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad , d_i = (x_i - y_i)$$

مثال : كانت تقديرات ستة طلاب في مادتي الاحصاء والرياضيات هي:

الاحصاء (x) :جيد، متوسط، ضعيف ،مقبول، جيدجدا ، ممتاز

الرياضيات (y): متوسط، جيد، مقبول، ضعيف، ممتاز،جيد جدا،

جد معامل الارتباط البسيط بين بين تقدير الطالب وفي امتحان الاحصاء وتقديره في امتحان الرياضيات

الحل : نبدا بترتيب التقديرات وفق ترتيب تصاعدي او تنازلي وليكن ترتيب تصاعدي ثم تخصص رتبا تمثل قيم سلسلة اعداد طبيعية

التقديرات : ضعيف ، مقبول ، متوسط، جيد، جيد جدا ، ممتاز،

الرتب :1، 2، 3، 4، 5، 6

ثم نعود لتخصيص هذه الرتب للتقديرات الاصلية كل حسب موقعها كما موضح في الجدول التالي:

x_i	y_i	رتب x_i	رتب y_i	$d_i = x_i - y_i$	d_i^2
جيد	متوسط	4	3	1	1
متوسط	جيد	3	4	-1	1
ضعيف	مقبول	1	2	-1	1
مقبول	ضعيف	2	1	1	1
جيدجدا	ممتاز	5	6	-1	1
ممتاز	جيد جدا	6	5	1	1
المجموع				0	6

$$r_{x,y} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6*6}{6(6^2-1)} = 0.892$$

الواجب البيتي

السؤال الاول : اذا علمت ان $r_{12} = r_{13} = r_{23} = r$

برهن ان $R_{1.23} = \sqrt{\frac{2r^2}{1+r}}$

$$R_{2.13} = \sqrt{\frac{2r^2}{1+r}}$$

$$R_{3.12} = \sqrt{\frac{2r^2}{1+r}}$$

السؤال الثاني : الاتي تقديرات لعشرة عمال في احد المصانع من حيث ادارتهم في تشغيل نوعين من المكائن الحديثة يطلب حساب معامل الارتباط البسيط بين كفاءة العامل في ادارة هذين النوعين من المكائن

النوع الاول (x): جيد، متوسط، جيد جدا ،متوسط، ممتاز ، ضعيف، مقبول،متوسط ، جيد ، جيد

النوع الثاني (y): متوسط ، جيد ، جيد، مقبول، جيد جدا، ضعيف، مقبول، متوسط، ، متوسط، ممتاز

الانحدار الخطي البسيط simple linear regression

سوف نأخذ فكرة مختصرة جدا ومبسطة عن الانحدار الخطي البسيط حيث سيتم دراسة هذا الموضوع بشكل مفصل بالسنوات القادمة بصورة عامة ينقسم الانحدار الخطي البسيط الى

اولا: الانحدار البسيط simple regression ويشتمل على متغيرين احدهما تابع y والاخر مستقل x

ثانيا : الانحدار المتعدد multiple regression ويشتمل على متغير تابع y وعدة متغيرات مستقلة $x_1, \dots, x_n$

سوف يقتصر شرحنا فقط على الانحدار الخطي البسيط وبصورة مبسطة جدا لان هذا الموضوع سوف تتم دراسته في السنوات السابقة .

ان المعادلة التي تصف العلاقة بين x, y هي معادلة خطية وتمثل بيانيا بخط مستقيم . نفرض اننا نريد التنبؤ بالدرجة النهائية لطالب في مادة الاحصاء (y) معتمدين بذلك على درجته الفصلية x لذا فان (y, x) تمثل النتيجة لاي طالب في المجتمع هذا يعني ان هناك دالة رياضية تجمع هذه المتغيرات يمكن صياغتها بشكل خطي $y = a + bx$ حيث ان y هو المتغير التابع وx هو المتغير المستقل وان :

$$a, b$$

ثوابت حقيقية (معالم الدالة)تعبر عن خط مستقيم في المستوى y, x ويجب صياغة هذه المعادلة عشوائيا وحتى نتمكن من الصياغة العشوائية يجب ان يتم الاخذ بنظر الاعتبار جملة من المتغيرات الاخرى المجهولة فاذا رمزنا لهذه المتغيرات بالرمز e عندئذ يمكن صياغة هذه الدالة كالاتي

$$y = a + bx + e$$

يهدف الانحدار الخطي البسيط الى تقدير قيم عددية الى معالم النموذج اعلاه اي تقدير a, b

في دراستنا المبسطة هذه سوف نعطي هذه التقديرات مباشرة دون الخوض في التفاصيل حيث ان التفاصيل سيتم التطرق لها في الدراسات والمراحل المستقبلية .

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_X^2}$$

وهذا يعني ان قيمة $\hat{b}$ ما هي الا قيمة التباين المشترك بين x, y على تباين x وان $\hat{a} = \bar{y} - \hat{b}\bar{x}$ ويلاحظ ان كل من $\hat{a}, \hat{b}$ هي دالة بدلالة قياسات مفردات العينة x_i, y_i ويلاحظ ان قيمة $\hat{b}$ تكون سالبة او موجبة تبعا لقيمة S_{xy} اذا كانت موجبة او سالبة فقيمة$\hat{b}$ الموجبة تعني ان العلاقة بين yوx موجبة والعكس بالعكس وغالبا يقال لما سبق تم تقدير معادلة انحدار y/x اي ايجاد معادلة انحدار فيها المتغير التابع هو y والمتغير المستقل هو x

مثال: البيانات الاتية تمثل الدرجات الفصلية والدرجات النهائية في درس الاحصاء لـ 12طالب اوجد معادلة خط الانحدار y/x

الدرجة النهائية y	الدرجة الفصلية x
85	65
74	50
76	55
90	65
85	55
87	70
94	65

98	70
81	55
91	70
76	50
74	55

الحل: الجدول التالي يبين خطوات حساب معامل الانحدار

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$\hat{y}$
65	85	4225	7225	5525	88.4
50	74	2500	5476	3700	74.9
55	76	3025	5776	4180	79.4
65	90	4225	8100	5850	88.4
55	85	3025	7225	4675	79.4
70	87	4900	7569	6090	92.8
65	94	4225	8836	6110	88.4
70	98	4900	9604	6860	92.8
55	81	3025	6561	4455	79.4
70	91	4900	8281	6370	92.8
50	76	2500	5776	3800	74.9
55	74	3025	5476	4070	79.4
725	1011	44475	85905	61685	

$$\bar{x} = \frac{725}{12} = 60.4 \quad , \quad \bar{y} = \frac{1011}{12} = 84.25$$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{61685 - 10 * 60.4 * 84.25}{44475 - 10 * (60.4)^2} = 0.897$$

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

$$\hat{a} = 84.25 - 0.897 * 60.4$$

$$\hat{a} = 30.056$$

وبالتعويض بمعادلة خط الانحدار $y = a + bx + e$ نحصل على المعادلة المقدرة الاتية :

$$\hat{y} = 30.056 + 0.897x$$

والان لو عوضنا عن اي قيمتين من قيم x المعطاة بالسؤال ولتكونا 50و70 فأن :

$$\hat{y}_{50} = 30.05 + 0.897(50) = 74.9$$

$$\hat{y}_{70} = 30.05 + 0.897(70) = 92.8$$

وهذا وبالتعويض في هذه المعادلة عن كل قيمة من قيم x فانه يمكن حساب القيم التقديرية الـ اي y_x والتي تقع جميعها على خط الانحدار البسيط وهذه القيم موضحة في الجدول اعلاه.

واجب بيتي : للبيانات السابقة اوجد معادلة خط انحدار x/y

العلاقة بين معامل الانحدار ومعامل الارتباط البسيط

يمكن حساب معامل الانحدار اذا علمت معامل الارتباط البسيط بين x وy وفق الصيغة *(x/y)*

$$\hat{b} = \frac{S_x}{S_y} . r_{x,y}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \quad \text{و} \quad S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}$$

اما معامل الارتباط يستخرج وفق الصيغة الاتية :

$$r_{x,y} = \frac{S_x}{S_y} * \hat{b}$$

وحيث $\frac{S_x}{S_y}$ و $\frac{S_y}{S_x}$ كميات موجبة دائما لذا فأن اشارة معامل الانحدار تتحدد من خلال اشارة معامل الارتباط والعكس صحيح دلالة على طبيعة العلاقة بين y,x

مثال : اذا علمت ان $r_{x,y} = 0.905$ وان $S_x = 1.333$ و$S_y = 2.211$ جد معامل انحدار $y \ / \ x$ ومعامل انحدار $x \ / \ y$

الحل :

$$coefficient\ of(x \ /y) = \frac{S_x}{S_y}.r_{x,y}$$

$$coefficient\ of(x \ /y) = \frac{1.333}{2.211}.0.905$$

$$coefficient\ of(x \ /y) = 0.546$$

##

$$coefficient\ of(y \ /x) = \frac{S_y}{S_x}.r_{x,y}$$

$$coefficient\ of\ (y \ /x) = \frac{2.211}{1.333}.0.905$$

$$coefficient\ of(y \ /x) = 1.501$$